

BASES INDISPENSABLES DES PROBABILITÉS ET DES STATISTIQUES

JULIAN TUGAUT
Version du 13 mars 2024

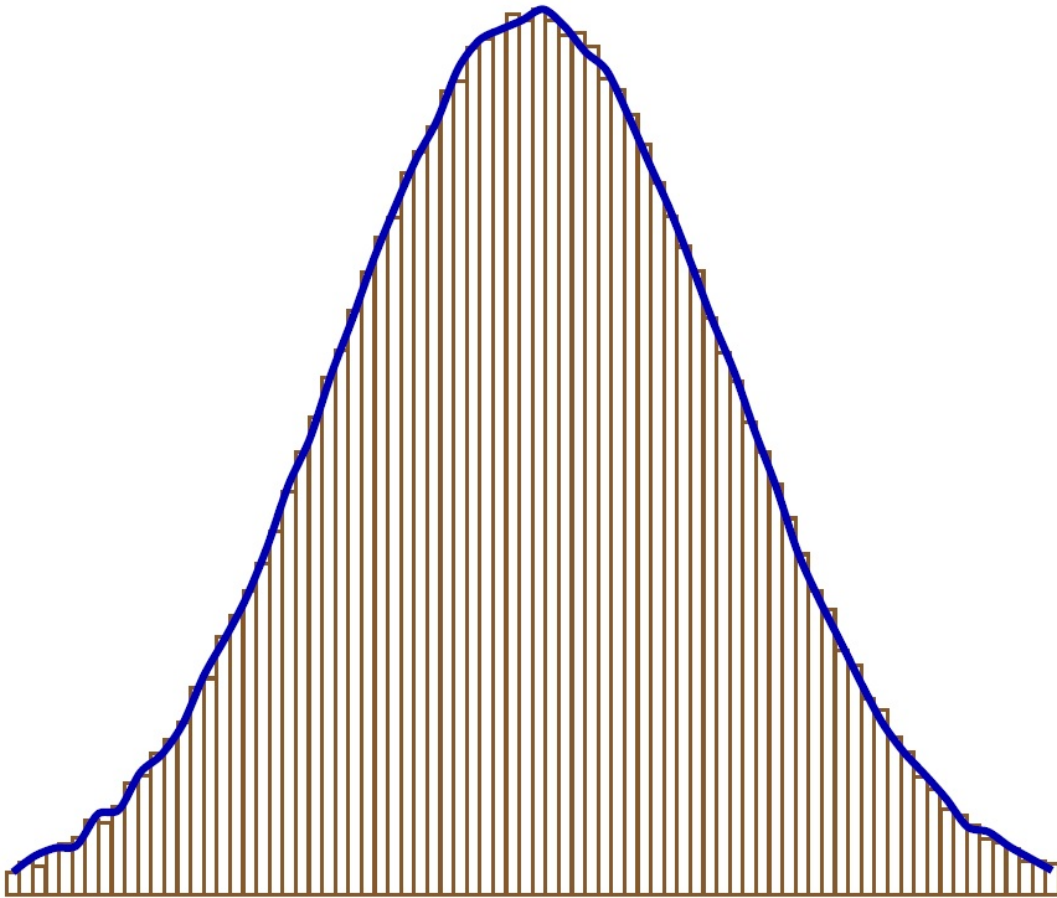


Table des matières

Table des matières	iii
Liste des figures	xiii
Liste des tableaux	xvii
Avant-propos	1
Notations	5
I Probabilités	7
1 Le langage des ensembles	9
1.1 Introduction	9
1.2 Définitions	9
1.2.1 Notion d'ensemble	9
1.2.2 Ensemble vide	10
1.2.3 Inclusion, Sous-ensembles	10
1.3 Opérations sur les ensembles	11
1.3.1 Intersection de deux ensembles	12
1.3.2 Réunion de deux ensembles	14
1.3.3 Propriétés de distributivité	15
1.3.4 Complémentaire d'un sous-ensemble	16
1.4 Partition d'un ensemble	18
1.5 Rappels sur la dénombrabilité	20

2	Le modèle probabiliste	21
2.1	Introduction	21
2.2	Modélisation d'une expérience aléatoire	21
2.2.1	L'espace fondamental	22
2.2.2	Les évènements	24
2.3	Définition et propriétés d'une probabilité	27
2.3.1	Définition	27
2.3.2	Propriétés	28
2.3.3	Espace fondamental dénombrable	32
2.4	Probabilité conditionnelle et indépendance	35
2.4.1	Probabilité conditionnelle	35
2.4.2	Indépendance d'évènements	38
2.4.3	Lemmes de Borel-Cantelli	41
3	Généralités sur les variables aléatoires	45
3.1	Introduction	45
3.2	Définition et Opérations	46
3.2.1	Définition	46
3.2.2	Opérations sur les variables aléatoires réelles	48
3.3	Loi d'une variable aléatoire réelle	49
3.3.1	Définition	50
3.3.2	Fonction de répartition	51
3.4	Indépendance des variables aléatoires réelles	55
3.4.1	Définition	56
3.4.2	Exemples	56
3.4.3	Contre-exemple	57
3.4.4	Propriétés	57
4	Variables aléatoires discrètes	59
4.1	Introduction	59
4.2	Caractérisation de la loi	59
4.2.1	Calcul de la probabilité d'un intervalle	60
4.3	Caractéristiques	62
4.3.1	Espérance d'une variable aléatoire réelle	62
4.3.2	Variance, Écart-type	68

5	Lois de probabilité discrètes classiques	77
5.1	Introduction	77
5.2	Loi de Bernoulli	78
5.3	Loi binomiale	79
5.3.1	Formulation	79
5.3.2	Espace fondamental	80
5.3.3	Solution	81
5.3.4	Définition	83
5.3.5	Décomposition de X	84
5.3.6	Caractéristiques	84
5.3.7	Somme de lois binomiales	87
5.3.8	Courbe représentative de la loi binomiale	87
5.3.9	Tables de la loi binomiale	89
5.4	Loi de Poisson	90
5.4.1	Définition	90
5.4.2	Courbe représentative de la loi de Poisson	91
5.4.3	Caractéristiques	91
5.4.4	Somme de lois de Poisson	92
5.4.5	Tables de la loi de Poisson	93
6	Autres lois discrètes classiques	95
6.1	Introduction	95
6.2	Loi de Rademacher	95
6.3	Loi uniforme discrète	96
6.4	Loi triangulaire discrète	97
6.5	Loi géométrique	99
6.6	Loi hypergéométrique	102
6.7	Loi de Zipf	103
6.8	Loi de Benford	104
6.9	Loi logarithmique	104
6.10	Loi zêta	105
7	Fonctions génératrices	109
7.1	Introduction	109
7.2	Définition et premières propriétés	109
7.3	Fonctions génératrices des lois classiques	110
7.3.1	Si $\mathbb{P}_X = \mathcal{B}(p)$	110
7.3.2	Si $\mathbb{P}_X = \mathcal{B}(n, p)$	111
7.3.3	Si $\mathbb{P}_X = \mathcal{P}(\lambda)$	111
7.3.4	Pour d'autres lois discrètes usuelles	111
7.4	Résultats importants	112
7.5	Extinction des grands noms	115

8	Variables aléatoires à densité	117
8.1	Introduction	117
8.2	Densité : définition et propriétés	118
8.2.1	Définition	118
8.2.2	Propriétés	119
8.3	Calcul de la probabilité d'un intervalle	121
8.4	Caractéristiques	121
8.4.1	Espérance d'une variable aléatoire réelle	122
8.4.2	Variance, Écart-type	126
8.4.3	Mélange de variables discrètes et à densité	129
8.4.4	Skewness et kurtosis	129
9	Lois de probabilité continues classiques	131
9.1	Introduction	131
9.2	Loi uniforme	132
9.2.1	Caractéristiques	134
9.3	Loi exponentielle	136
9.3.1	Caractéristiques	137
9.4	Loi normale (ou gaussienne)	138
9.4.1	Définition	138
9.4.2	Caractéristiques	141
9.4.3	Propriétés	142
10	Autres lois continues classiques	147
10.1	Introduction	147
10.2	Loi triangulaire	147
10.3	Loi logistique	149
10.4	Loi d'Erlang	151
10.5	Loi de Weibull	152
10.6	Loi Gamma	153
10.7	Loi du χ^2	155
10.8	Loi de Rayleigh	158
10.9	Loi log-normale	160
10.10	Loi de Student	162
10.11	Loi de Fisher	165
10.12	Loi de Pareto	166
10.13	Loi de Cauchy	167

11 Fonctions caractéristiques	171
11.1 Introduction	171
11.2 Définition et premières propriétés	172
11.3 Fonctions caractéristiques des lois usuelles	175
11.3.1 Lois discrètes	175
11.3.2 Lois à densité	177
11.4 Résultats importants	179
12 Vecteurs aléatoires	183
12.1 Introduction	183
12.2 Premières définitions	184
12.3 Loi d'un vecteur aléatoire	184
12.4 Fonction de répartition	185
12.5 Fonction caractéristique	186
12.6 Vecteurs aléatoires discrets	187
12.7 Vecteurs aléatoires à densité	188
12.8 Indépendance des composantes	189
12.9 Lois conditionnelles	190
13 Vecteurs gaussiens	193
13.1 Introduction	193
13.2 Rappels et compléments	193
13.3 Définition	194
13.4 Existence	195
13.5 Fonction caractéristique	196
13.6 Densité de probabilité	197
13.7 Théorème de Cochran	198
14 Convergences des variables aléatoires	199
14.1 Introduction	199
14.1.1 Paradoxe	200
14.1.2 Convergence simple	201
14.2 Convergence presque sûre	202
14.2.1 Définition et premières propriétés	202
14.2.2 Caractérisation	203
14.3 Convergence dans L^p	204
14.3.1 Définition	205
14.3.2 Différence avec l'espace L^p en analyse	206

14.4	Convergence en probabilité	206
14.4.1	Définition	206
14.4.2	Liens avec les deux autres convergences	207
14.4.3	Remarque	207
14.4.4	Réciproques partielles	208
14.5	Convergence en loi	209
14.5.1	Définition et première propriété	209
14.5.2	Remarque	209
14.5.3	Liens avec les convergences précédentes	210
14.5.4	Obtenir une convergence en loi	210
14.6	Retour sur le paradoxe	211
14.7	Synthèse	212
15	Grands nombres et théorème central	213
15.1	Introduction	213
15.2	Petit jeu	213
15.3	Lois des grands nombres	214
15.3.1	Motivation : méthodes de Monte Carlo	214
15.3.2	Loi faible des grands nombres	214
15.3.3	Loi forte des grands nombres	215
15.3.4	Loi des grands nombres de Kolmogorov	216
15.4	Théorème central de la limite	217
15.4.1	Énoncé du théorème central limite	217
15.4.2	Preuve du théorème central limite	217
15.4.3	Inégalité de Berry-Esseen	218
15.5	Retour sur le jeu	219
15.5.1	Premier raisonnement (faux)	220
15.5.2	Deuxième raisonnement (faux également)	220
15.5.3	Troisième raisonnement	220
15.5.4	Synthèse	220

II	Statistiques	223
16	Statistiques descriptives	225
16.1	Introduction	225
16.2	Vocabulaire	225
16.3	Caractère qualitatif	227
16.4	Séries statistiques simples : présentation	230
16.4.1	Série à valeurs isolées	230
16.4.2	Série à valeurs classées	232
16.5	Séries statistiques simples : caractéristiques	236
16.5.1	Moyenne arithmétique	236
16.5.2	Quantiles	239
16.5.3	Variance, Écart-type	242
16.5.4	Autres caractéristiques	246
16.6	Séries statistiques doubles	247
16.6.1	Définitions	247
16.6.2	Tableau de contingence	247
16.6.3	Indépendance de deux caractères	250
16.6.4	Nuage de points	253
16.6.5	Covariance	255
16.6.6	Description empirique de la dépendance	256
17	Estimation ponctuelle	257
17.1	Introduction	257
17.2	Préliminaires	257
17.3	Vocabulaire	258
17.4	Notion de statistique	259
17.5	Moyenne d'échantillon	260
17.6	Variance d'échantillon	262
17.7	Proportion d'échantillon	265
17.8	Estimateurs	266
17.8.1	Préliminaires	266
17.8.2	Qualité des estimateurs	266
17.8.3	Retour sur l'Exemple 17.8.3	271
18	Deux méthodes classiques d'estimation ponctuelle	275
18.1	Introduction	275
18.2	Méthode des moments	275
18.2.1	Exemple introductif	276
18.2.2	Méthode générale	276
18.2.3	Exemples	277
18.2.4	Limites	279

18.3	Méthode du maximum de vraisemblance	279
18.3.1	Exemple introductif	280
18.3.2	Méthode générale	280
18.3.3	Exemples	281
19	Intervalles de confiance	287
19.1	Introduction	287
19.2	Préliminaires	287
19.2.1	Philosophie des intervalles de confiance	287
19.2.2	Définitions	288
19.2.3	Premier intervalle de confiance de la moyenne	289
19.3	Cas gaussien	289
19.3.1	Estimation de μ si σ^2 est connue	290
19.3.2	Estimation de μ si σ^2 est inconnue	290
19.3.3	Estimation de σ^2 si μ est connue	291
19.3.4	Estimation de σ^2 si μ est inconnue	292
19.3.5	Différence de deux moyennes	292
19.4	Intervalles de confiance asymptotiques	293
19.4.1	Loi de Poisson	294
19.4.2	Loi de Bernoulli	294
19.5	Cas exponentiel	295
19.5.1	Intervalles asymptotiques	295
19.5.2	Intervalles non asymptotiques	295
19.6	Pour aller plus loin	295
20	Généralités sur les tests statistiques	299
20.1	Introduction	299
20.1.1	Motivations	299
20.1.2	Définition	300
20.1.3	Deux familles de tests	300
20.1.4	Quels tests peut-on effectuer?	301
20.2	Principe général	302
20.2.1	Formuler (H_0)	302
20.2.2	Définir (H_1)	302
20.2.3	Contrôler avec une statistique	303
20.2.4	Établir la distribution de la statistique de contrôle	303
20.2.5	Choisir α	303
20.2.6	Déterminer la région critique associée	303
20.2.7	Calculer la valeur empirique de la statistique	304
20.2.8	Décider	304
20.3	Erreur de première espèce	304
20.4	Erreur de seconde espèce	304

21 Tests d'hypothèse classiques	307
21.1 Introduction	307
21.2 Tests sur la moyenne μ	308
21.2.1 En supposant la variance σ^2 connue	308
21.2.2 En supposant la variance σ^2 inconnue	309
21.3 Tests sur la variance σ^2	311
21.3.1 En supposant la moyenne μ connue	311
21.3.2 En supposant la moyenne μ inconnue	312
21.4 Test sur une proportion f	314
21.5 Test de comparaison entre deux moyennes	315
21.5.1 En supposant les variances σ_X^2 et σ_Y^2 connues	316
21.5.2 En supposant les variances σ_X^2 et σ_Y^2 inconnues	316
21.6 Test de comparaison entre deux variances	317
21.7 Test de comparaison entre deux proportions	318
21.8 Test du χ^2 d'ajustement	319
21.8.1 Justification théorique du test	320
21.8.2 Pour une loi discrète	321
21.8.3 Pour une loi continue	323
21.8.4 Pour des lois qui dépendent d'un paramètre θ	323
21.9 Test du χ^2 d'indépendance	323
21.10 Test du χ^2 d'homogénéité	325
21.11 Tests de Kolmogorov-Smirnov	325
21.11.1 Test d'ajustement	326
21.11.2 Test d'homogénéité	327
22 Régression linéaire	329
22.1 Introduction	329
22.2 Équation de la droite de régression	329
22.3 Lien avec $r_{x,y}$	330
III Annexes	333
23 Rappels et compléments	335
23.1 Introduction	335
23.2 Ensembles	335
23.3 Suites	336
23.4 Fonctions	337
23.5 Topologie	338
23.6 Intégration de Riemann	338

24 Éléments de théorie de la mesure	341
24.1 Introduction	341
24.2 Tribus	341
24.3 Fonctions mesurables	344
24.4 Tribu borélienne	345
24.5 Mesures	346
24.6 Intégration de Lebesgue	349
24.7 Retour sur les chapitres précédents	352
25 Tables des lois de probabilité usuelles	355
Tables des lois discrètes	355
Tables de la loi binomiale	356
Tables de la loi de Poisson	361
Tables des lois continues	364
Tables de la loi normale	364
Tables de la loi de Student	366
Tables de la loi du χ^2	367
Tables de la loi de Fisher	368
Bibliographie	371
Index	373

Liste des figures

1.1	Élément d'un ensemble	10
1.2	Inclusion d'un ensemble	11
1.3	Intersection de deux ensembles	12
1.4	Ensembles disjoints	13
1.5	Réunion de deux ensembles	14
1.6	Complémentaire d'un sous-ensemble	17
1.7	Partition d'un ensemble	19
1.8	Partition particulière d'un ensemble	19
2.1	Croissance de la probabilité	31
3.1	Variable aléatoire réelle	47
3.2	Graphes de la fonction de répartition	53
4.1	Diagramme en bâtons de la loi de probabilité	60
4.2	Inégalité de Bienaymé-Tchebychev	71
5.1	Arbre de probabilité pour la loi de Bernoulli	78
5.2	Arbre de probabilité pour la loi binomiale	80
5.3	Diagramme en bâtons de la loi binomiale 1	88
5.4	Diagramme en bâtons de la loi binomiale 2	88
5.5	Diagramme en bâtons de la loi de Poisson 1	91
5.6	Diagramme en bâtons de la loi de Poisson 2	91
6.1	Arbre de probabilité pour la loi de Rademacher	95
6.2	Diagramme en bâtons de la loi de Rademacher	96
6.3	Diagramme en bâtons de la loi triangulaire discrète	99
6.4	Diagramme en bâtons de la loi géométrique 1	101
6.5	Diagramme en bâtons de la loi géométrique 2	102
6.6	Diagramme en bâtons de la loi hypergéométrique	103
6.7	Diagramme en bâtons de la loi de Zipf	103
6.8	Diagramme en bâtons de la loi de Benford	104
6.9	Diagramme en bâtons de la loi logarithmique	105

9.1	Densité de probabilité de la loi uniforme 1	132
9.2	Densité de probabilité de la loi uniforme 2	133
9.3	Fonction de répartition de la loi uniforme 1	133
9.4	Fonction de répartition de la loi uniforme 2	134
9.5	Densité de probabilité de la loi exponentielle	136
9.6	Fonction de répartition de la loi exponentielle	137
9.7	Densité de probabilité de la loi normale	139
9.8	Densité de probabilité de la loi normale centrée réduite	140
9.9	Densités de probabilité de différentes lois normales centrées	140
9.10	Fonction de répartition de la loi normale	141
9.11	Fonction de répartition de la loi normale centrée réduite	141
10.1	Densité de probabilité de la loi triangulaire	149
10.2	Fonction de répartition de la loi triangulaire	149
10.3	Densité de probabilité de la loi logistique standard	150
10.4	Fonction de répartition de la loi logistique standard	150
10.5	Densité de probabilité de la loi d'Erlang	151
10.6	Fonction de répartition de la loi d'Erlang	152
10.7	Densité de probabilité de la loi de Weibull	152
10.8	Fonction de répartition de la loi de Weibull	153
10.9	Densité de probabilité de la loi Gamma	153
10.10	Fonction de répartition de la loi Gamma	154
10.11	Densité de probabilité de la loi du Khi-deux	158
10.12	Fonction de répartition de la loi du Khi-deux	158
10.13	Densité de probabilité de la loi de Rayleigh	159
10.14	Fonction de répartition de la loi de Rayleigh	160
10.15	Densité de probabilité de la loi lognormale	161
10.16	Fonction de répartition de la loi lognormale	161
10.17	Densité de probabilité de la loi de Student	165
10.18	Densité de probabilité de la loi de Pareto	166
10.19	Fonction de répartition de la loi de Pareto	167
14.1	Liens entre les convergences	212
16.1	Avis des étudiants sur le cours magistral - 1	228
16.2	Avis des étudiants sur le cours magistral - 2	229
16.3	Avis des étudiants sur le cours magistral - 3	229
16.4	Effectifs d'une série statistique simple	231
16.5	Effectifs cumulés d'une série statistique simple	232
16.6	Histogramme des fréquences corrigées	234
16.7	Courbe des effectifs cumulés	235
16.8	Histogramme des fréquences	236
16.9	Exemple de box-plot (horizontale)	240
16.10	Exemple de box-plot (verticale)	240
16.11	Nuage de points - 1	253

16.12	Nuage de points - 2	253
16.13	Nuage de points - 3	254
16.14	Nuage de points - 4	254
21.1	Comparaison des effectifs réels (en bleu) et théoriques (en rouge)	322

Liste des tableaux

3.1	Définition de la probabilité \mathbb{P} sur l'univers Ω	49
3.2	Réalisations de la variable aléatoire réelle X	49
3.3	Loi de probabilité de la variable aléatoire réelle X	50
3.4	Loi de probabilité de la variable aléatoire réelle X	52
5.1	Loi de probabilité de la variable aléatoire réelle X	78
6.1	Loi de probabilité de la variable aléatoire réelle X	96
16.1	Avis des soixantes étudiants	228
16.2	Série à valeurs isolées	230
16.3	Exemple de série à valeurs isolées	230
16.4	Série à valeurs classées	232
16.5	Exemple de série à valeurs classées	233
16.6	Série à valeurs classées	234
16.7	Tableau de contingence - 1	248
16.8	Tableau de contingence - 2	249
16.9	Exemple de tableau de contingence	251
16.10	Tableau de contingence (poids/taille)	251
21.1	Dé électronique	321
21.2	Test du χ^2 sur le dé	322
25.1	Table de la loi binomiale avec $p = 0.05$	356
25.2	Table de la loi binomiale avec $p = 0.10$	356
25.3	Table de la loi binomiale avec $p = 0.20$	357
25.4	Table de la loi binomiale avec $p = 0.30$	358
25.5	Table de la loi binomiale avec $p = 0.40$	359
25.6	Table de la loi binomiale avec $p = 0.50$	360
25.7	Table de la loi de Poisson - 1	361
25.8	Table de la loi de Poisson - 2	362
25.9	Table de la loi de Poisson - 3	363
25.10	Table de la fonction de répartition de la loi normale	364
25.11	Cas des grandes valeurs de $t - 1$	365
25.12	Cas des grandes valeurs de $t - 2$	365

25.13	Quantile d'ordre $1 - \frac{\gamma}{2}$ de la loi normale centrée réduite	365
25.14	Quantiles d'ordre γ de la loi de Student à n degrés de liberté . . .	366
25.15	Quantiles d'ordre γ de la loi du χ^2 à n degrés de liberté	367
25.16	Quantiles d'ordre 0.975 de la loi $\mathcal{F}(d_1, d_2)$	368
25.17	Quantiles d'ordre 0.95 de la loi $\mathcal{F}(d_1, d_2)$	369

Avant-propos

Ce livre est issu du polycopié que nous avons rédigé pour les cours de Probabilités et de Statistiques en première année de cycle ingénieur à Télécom Saint-Étienne. Les étudiants qui suivent ce cursus sont issus de différents parcours comme ceux des classes préparatoires aux grandes écoles d'ingénieurs (MPI, MP, PSI, PC, PT, TSI), de Licence, de BUT (ex-DUT), de Bachelors...

In fine, les profils sont variés et il nous a vite apparu qu'il était nécessaire de nous adapter à cette hétérogénéité. De fait, les cours que nous avons donnés n'abordent pas du tout la théorie de la mesure, pourtant indispensable à toute étude fine des probabilités. Néanmoins, le public auquel était destiné le cours n'a pas vocation à suivre des études poussées en théorie. Les aspects pratiques dans la vie professionnelle sont alors le maître-mot. Ceci se ressent dans le présent ouvrage où les concepts ardués liés à la théorie de la mesure (tribu, mesure, intégrale au sens de Lebesgue générale...) sont passés sous silence autant que faire se peut.

La modélisation aléatoire tient aujourd'hui une grande part dans l'enseignement des mathématiques appliquées dans les écoles d'ingénieurs. On peut trouver cela curieux à premier abord. En effet, d'après le déterminisme laplacien, si l'on connaît les conditions initiales ainsi que les lois qui régissent le monde, il n'y a point d'aléa. Cela dit, connaître toutes les conditions initiales est impossible. Et, calculer la position de chaque particule du système étudié serait trop coûteux en temps de calcul. La raison en est la métastabilité sous-jacente à la dynamique moléculaire. D'ailleurs, certaines méthodes pour éviter le fléau de la métastabilité sont probabilistes... L'on dit souvent, à juste titre, que les mathématiques sont une science exacte. Néanmoins, l'application des mathématiques dans le monde réel l'est moins. En effet, la partie modélisation altère la réalité physique pour la faire correspondre à des modèles donnés qui, de fait, ne sont qu'une approximation plus ou moins grossière de la réalité. Ce phénomène est particulièrement palpable quand on traite des probabilités et une large part du travail consiste en la modélisation. Par conséquent, cette branche des mathématiques peut sembler différente aux étudiants qui l'abordent pour la première fois.

Dans cet ouvrage, nous avons choisi de ne pas aborder la théorie de la mesure autrement que dans une annexe à la fin. En effet, le public visé n'est pas dans l'optique de la théorie mais dans celle de l'utilisation dans la vie professionnelle. Des choix en découlent. Ainsi, la notion de tribu n'est pas abordée. Néanmoins,

il arrive que nous en parlions brièvement pour pointer du doigt les endroits où l'on triche ostensiblement. Pour aller plus loin et aborder les martingales ainsi que les processus de Markov, il serait nécessaire de faire de la théorie de la mesure. Ici, nous présentons surtout les bases indispensables des probabilités et des statistiques ; dont l'objectif est justement de gérer la modélisation aléatoire.

Les prérequis sont divers. Outre avoir un niveau solide de Terminale, il est crucial que le lecteur connaisse la trigonométrie et les nombres complexes (par exemple pour aborder sereinement les fonctions caractéristiques), l'algèbre linéaire (l'espérance est linéaire par exemple), les suites numériques, les séries numériques, les suites de fonctions, les séries de fonctions (la fonction génératrice étant en particulier une série entière) et bien sûr l'intégration au sens de Riemann (les probabilités relatives aux variables aléatoires à densité se traitant à coup d'intégrales). Notons toutefois que des éventuelles lacunes dans ces domaines ne seraient rédhibitoires. Plutôt qu'une compréhension fine des divers concepts, il est important que le lecteur soit familier avec eux. Cependant, les facultés calculatoires sont un plus indéniable. De même, connaître les équations différentielles serait un plus au même titre que de connaître les formes quadratiques.

Il convient de noter que des exercices (non corrigés) ont été ajoutés au sein du texte pour que le lecteur puisse se familiariser aux idées introduites. Par ailleurs, quasiment tous les chapitres comportent une section d'exercices corrigés. La correction desdits exercices est fournie dans un chapitre après les annexes.

Dans une première partie, nous présentons les bases des probabilités. Pour ce faire, nous abordons le langage des ensembles, clef de voûte des mathématiques. Nous présentons ensuite le modèle probabiliste. Puis, nous enchaînons sur les variables aléatoires en général et les variables aléatoires discrètes en particulier ainsi que les lois discrètes usuelles. Suite à cela, un chapitre parle des fonctions génératrices. Puis, nous attaquons les variables aléatoires à densité et les lois à densité usuelles.

Nous passons alors à des concepts plus ardues en probabilités : fonctions caractéristiques, vecteurs aléatoires, vecteurs gaussiens (dont le théorème de Cochran). Puis, nous présentons les différents modes de convergence pour les variables aléatoires et nous finissons avec un chapitre sur les lois des grands nombres et le théorème central de la limite ; lesquels sont les premiers résultats non élémentaires en théorie des probabilités.

Dans la seconde partie du manuscrit, nous donnons les bases indispensables des statistiques en commençant par les statistiques descriptives suite à quoi nous faisons un chapitre sur l'estimation ponctuelle puis un sur les deux méthodes les plus classiques d'estimation (la méthode des moments et le maximum de vraisemblance). Nous changeons alors de paradigme pour quantifier l'erreur commise avec les intervalles de confiance. Plus loin, nous parlons des tests statistiques. Enfin, nous donnons quelques rudiments sur la régression linéaire.

Pour terminer, nous avons écrit deux annexes : l'une concerne des rappels et

des compléments essentiels. L'autre traite des rudiments de la théorie de la mesure. Il convient de noter que l'on a rédigé une page de notations et nous donnons les tables des lois usuelles. Nous terminons enfin par la bibliographie et un index.

Il est maintenant temps de lister toutes les autres sources qui furent d'une aide plus ou moins grande dans l'écriture de cet ouvrage et qui ne sont pas listées dans la bibliographie : Wikipedia et Google, la précédente version du polycopié par Didier Vincent à Télécom Saint-Étienne, le cours de Jean-Marie Monnez à l'IUT Charlemagne à Nancy en 2009 et le polycopié de Jean-Michel Morel et Cédric Bernardin pour la préparation à l'agrégation de mathématiques à l'ÉNS de Cachan.

Ce livre a été écrit avec \LaTeX , via l'éditeur Texmaker. Les images ont intégralement été réalisées à partir de TikZ.

Nous achevons cet avant-propos en soulignant l'importance de nous signaler toute erreur, aussi infime soit-elle.

Notations

- $:=$ signifie que l'on introduit une quantité ou qu'on l'implémente.
- $a \approx b$ signifie que la quantité a est à peu près égale à la quantité b .
- $\bigcap_{i=1}^n A_i$ désigne l'intersection des ensembles A_1, \dots, A_n .
- $\bigcup_{i=1}^n A_i$ désigne la réunion des ensembles A_1, \dots, A_n .
- \log désigne le logarithme népérien.
- \log_{10} désigne le logarithme décimal.
- \forall désigne le quantificateur universel.
- \exists désigne le quantificateur existentiel.
- $\Re(z)$ désigne la partie réelle du complexe z .
- $\Im(z)$ désigne la partie imaginaire du complexe z .
- z^* désigne le conjugué du complexe z .
- $\binom{n}{k} := \frac{n!}{k!(n-k)!}$ désigne le coefficient binomial.
- H désigne la fonction de Heaviside c'est-à-dire que $H(x) = 1$ si $x \geq 0$ et $H(x) = 0$ sinon.
- $\mathbb{1}_A$ désigne la fonction indicatrice de l'ensemble A c'est-à-dire que $\mathbb{1}_A(x) = 1$ si $x \in A$ et $\mathbb{1}_A(x) = 0$ sinon.
- $\#A$ désigne le cardinal de l'ensemble fini A .
- 2^Ω désigne l'ensemble des parties de Ω .
- \mathbb{R}^Ω désigne l'ensemble des applications de Ω dans \mathbb{R} .
- $[a; b]$ désigne l'ensemble des réels entre a et b quand $a < b$.
- $\llbracket 1; n \rrbracket$ désigne l'ensemble des entiers compris entre 1 et n .
- $\llbracket a; b \rrbracket$ désigne l'ensemble des entiers compris entre a et b .
- \mathfrak{S}_n désigne l'ensemble des permutations de l'ensemble $\llbracket 1; n \rrbracket$.
- δ_a désigne la distribution de Dirac au point a .
- $\mathbb{P}(A)$, $\mathbb{P}[A]$ et $\mathbb{P}\{A\}$ désignent la probabilité de l'évènement A .
- $\mathbb{E}(X)$, $\mathbb{E}[X]$ et $\mathbb{E}\{X\}$ désignent l'espérance de la variable aléatoire X .
- $\text{Var}(X)$, $\text{Var}[X]$ et $\text{Var}\{X\}$ désignent la variance de la variable aléatoire X .
- $\mathcal{L}(X)$ désigne la loi de probabilité de la variable aléatoire X .
- $\mathcal{B}(p)$ désigne la loi de Bernoulli de paramètre p .
- $\mathcal{B}(n, p)$ désigne la loi binomiale de paramètres n et p .
- $\mathcal{P}(\lambda)$ désigne la loi de Poisson de paramètre λ .
- $\mathcal{U}_{[a; b]}$ désigne la loi uniforme sur l'intervalle $[a; b]$.
- $\mathcal{E}(\lambda)$ désigne la loi exponentielle de paramètre λ .
- $\mathcal{N}(m, \sigma^2)$ désigne la loi normale de paramètres m et σ^2 .

- Φ est la fonction de répartition de la loi normale centrée réduite.
- $X_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X$ signifie que la suite $(X_n)_n$ converge presque sûrement vers X .
- $X_n \xrightarrow[n \rightarrow +\infty]{L^p} X$ signifie que la suite $(X_n)_n$ converge vers X dans L^p .
- $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ signifie que la suite $(X_n)_n$ converge en probabilité vers X .
- $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$ signifie que la suite $(X_n)_n$ converge en loi vers X .

Première partie
Probabilités

Le langage des ensembles

1.1 Introduction

On commence le livre par l'étude du langage des ensembles. En effet, celui-ci nous servira subséquemment pour représenter et pour nous représenter de façon palpable ce à quoi correspondent les divers objets du modèle probabiliste.

Aucun pré-requis n'est nécessaire si ce n'est une base solide de Terminale. Il serait toutefois très positif de savoir ce qu'est une bijection et d'avoir déjà entendu parler d'ensembles dénombrables.

Les principaux objectifs de ce chapitre sont la découverte des diagrammes de Venn et des opérations sur les ensembles (inclusion, intersection, réunion et complémentation). Il est également attendu qu'à l'issue du chapitre, le lecteur ait bien compris la notion de distributivité ainsi que les lois de Morgan. Enfin, savoir identifier un ensemble dénombrable a un rôle clef dans les chapitres suivants. Il est donc fortement conseillé de porter une attention toute particulière à la section relative à la notion de dénombrabilité.

1.2 Définitions

1.2.1 Notion d'ensemble

On considère la notion intuitive suivante d'un ensemble Ω (qui désignera dans les prochains chapitres l'univers, que l'on appelle aussi l'espace fondamental) :

Définition 1.2.1. *Un ensemble Ω est une collection d'objets. Il est déterminé lorsque l'on peut dire si un objet ω lui appartient ou ne lui appartient pas.*

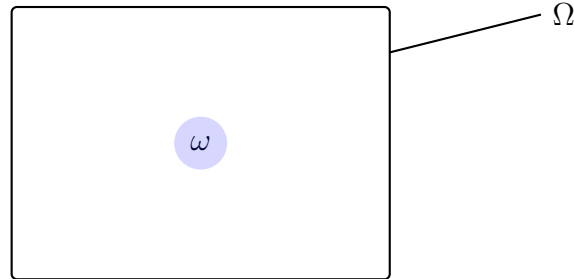
Notation 1.2.2. *Si l'objet ω appartient à l'ensemble Ω , on note : $\omega \in \Omega$.*

Si l'objet ω n'appartient pas à l'ensemble Ω , on note : $\omega \notin \Omega$.

Définition 1.2.3. *Un objet ω appartenant à l'ensemble Ω ($\omega \in \Omega$) est appelé un élément de Ω .*

Il peut être pratique d'utiliser un diagramme de Venn pour se représenter l'appartenance à un ensemble :

FIGURE 1.1 – Élément d'un ensemble



Il faut toutefois garder à l'esprit que le diagramme de Venn ne saurait se substituer au raisonnement.

1.2.2 Ensemble vide

Définition 1.2.4. On définit l'ensemble vide comme étant l'ensemble qui ne contient aucun élément.

Notation 1.2.5. L'ensemble vide est noté \emptyset .

Remarque 1.2.6. Il ne faut pas confondre l'ensemble vide (\emptyset) avec le zéro (0). On a coutume de dire : “Être nul, c'est déjà exister”.

1.2.3 Inclusion, Sous-ensembles

Définition 1.2.7. On dit qu'un ensemble A est inclus dans un ensemble B lorsque tout élément de A appartient à B . Plus formellement, A est inclus dans B lorsque

$$\forall \omega \in A, \omega \in B.$$

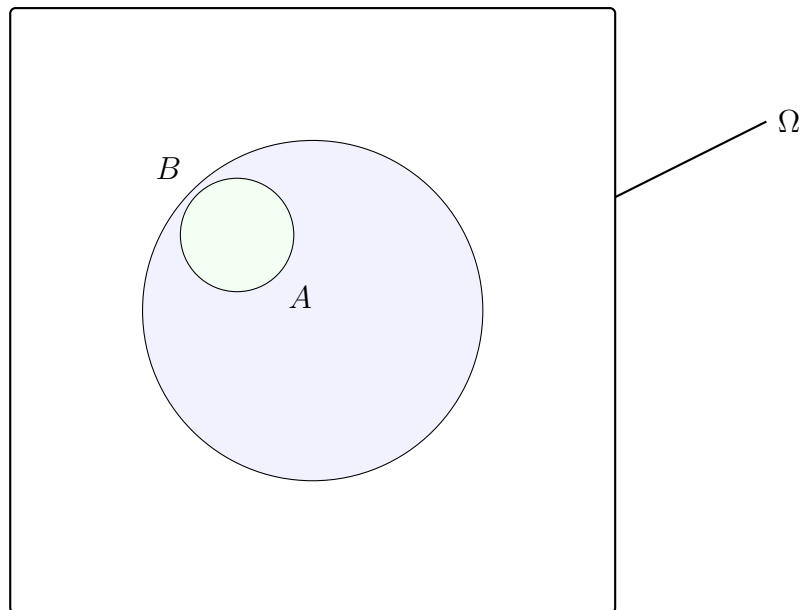
Remarque 1.2.8. On dit aussi que B contient A ou que A est un sous-ensemble de B .

Notation 1.2.9. Si A est un sous-ensemble de B , on note : $A \subset B$.

On peut aussi trouver la notation $B \supset A$.

Il peut être pratique d'utiliser un diagramme de Venn pour se représenter l'inclusion d'un ensemble dans un autre :

FIGURE 1.2 – Inclusion d'un ensemble



Exemple 1.2.10. Soit Ω l'ensemble des étudiants en France. Soit B l'ensemble des étudiants de Télécom Saint-Étienne. Soit A l'ensemble des FISE1 de Télécom Saint-Étienne.

Alors, A est inclus dans B : $A \subset B$.

Remarque 1.2.11. On peut remarquer dans l'exemple précédent que l'on a bien $A \subset \Omega$ et $B \subset \Omega$.

Théorème 1.2.12. Soient deux ensembles A et B . Alors, $A = B$ si et seulement si $A \subset B$ et $B \subset A$.

Remarque 1.2.13. Dans la pratique, démontrer que deux ensembles sont égaux n'est pas toujours évident. On peut alors, d'après le Théorème 1.2.12 procéder par double inclusion.

1.3 Opérations sur les ensembles

La mathématique est une science logico-formelle. Nous présentons donc maintenant les opérations logiques élémentaires, traduites en langage ensembliste.

Nous verrons ainsi l'intersection, la réunion et la complémentation.

Également, nous verrons les propriétés de ces différentes opérations.

1.3.1 Intersection de deux ensembles

Définition 1.3.1. On appelle *intersection* de deux ensembles A et B l'ensemble des éléments communs à A et à B .

Notation 1.3.2. L'intersection de deux ensembles A et B est notée $A \cap B$.

Plus formellement, on peut écrire :

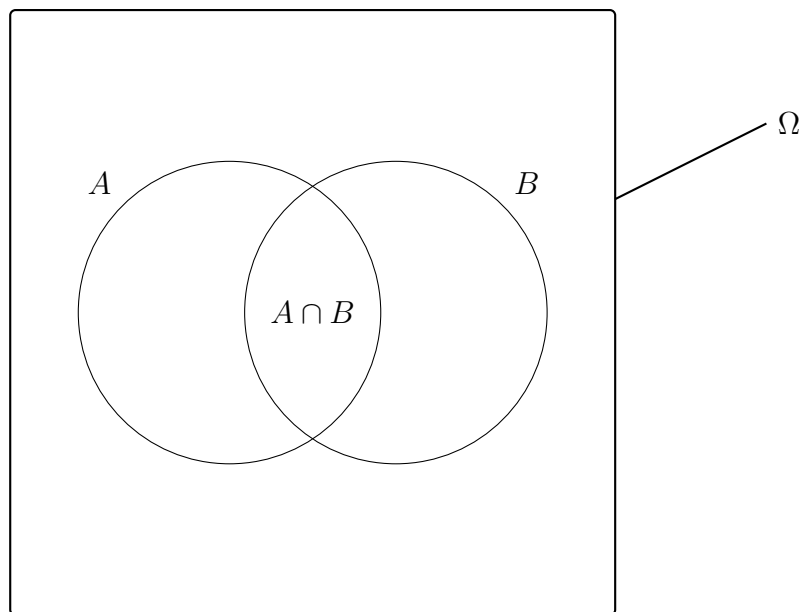
$$A \cap B = \{\omega : \omega \in A, \omega \in B\}$$

ou

$$\omega \in A \cap B \iff \omega \in A \text{ et } \omega \in B.$$

Avec un diagramme de Venn :

FIGURE 1.3 – Intersection de deux ensembles



Exemple 1.3.3 (Ensemble fini petit). Soit $\Omega := \{a, b, c, d, e, f, g, h, i\}$ un ensemble de lettres. Soient les deux sous-ensembles de Ω : $A := \{a, b, c, d, e\}$ et $B := \{c, d, e, f, g\}$. Alors, on a : $A \cap B = \{c, d, e\}$.

Exemple 1.3.4 (Ensemble fini grand). Soit Ω l'ensemble des ingénieurs formés en France. Soit A le sous-ensemble des ingénieurs exerçant dans l'industrie. Soit B l'ensemble des ingénieurs diplômés de Télécom Saint-Étienne (TSE). Alors, $A \cap B$ est l'ensemble des ingénieurs diplômés de TSE qui travaillent dans l'industrie.

Remarque 1.3.5. Ici, les mots “petit” et “grand” ne recouvrent aucune réalité mathématique et doivent être compris comme étant dans leur sens courant.

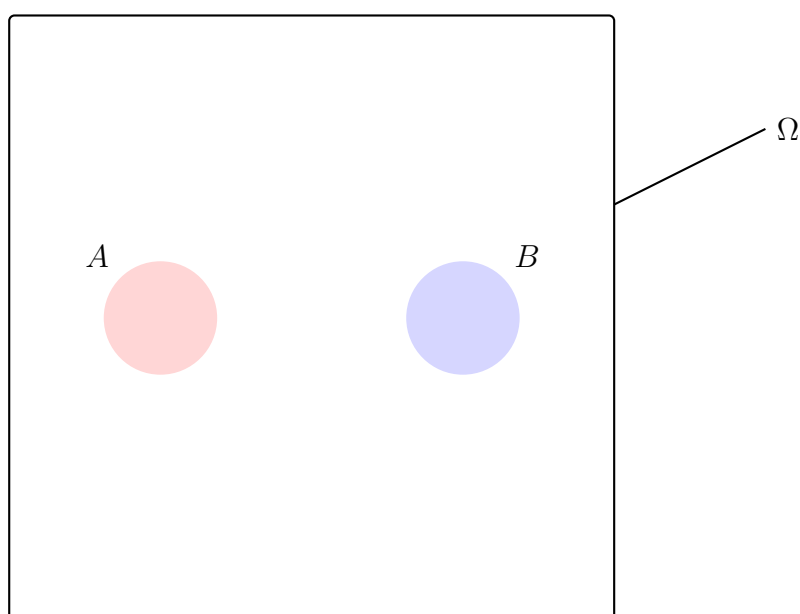
Exemple 1.3.6 (Ensemble infini dénombrable). Soit $\Omega := \mathbb{N}$ l'ensemble des entiers positifs ou nuls. Soit A le sous-ensemble des nombres premiers et soit B celui des nombres pairs. Alors, $A \cap B$ est le singleton $\{2\}$.

Exemple 1.3.7 (Ensemble infini non dénombrable). Soit Ω l'ensemble des réels strictement positifs. Soit A l'ensemble des réels strictement plus grands que 1. Soit $B :=]0; 4]$. Alors, $A \cap B$ est l'intervalle $]1; 4]$.

Définition 1.3.8 (Ensembles disjoints). On dit que deux ensembles A et B sont disjoints lorsque leur intersection est vide : ils n'ont aucun élément en commun. En d'autres termes, on dit que A et B sont disjoints si l'on a $A \cap B = \emptyset$.

Avec un diagramme de Venn :

FIGURE 1.4 – Ensembles disjoints



On présente maintenant les propriétés basiques de l'intersection.

Proposition 1.3.9 (Commutativité). Soient deux ensembles A et B .

Alors $A \cap B = B \cap A$.

Proposition 1.3.10 (Associativité). Soient trois ensembles A , B et C .

Alors :

$$A \cap (B \cap C) = (A \cap B) \cap C =: A \cap B \cap C.$$

Ainsi, s'il n'y a que de l'intersection entre les ensembles, les parenthèses sont superflues.

Proposition 1.3.11. Soit un ensemble A . Alors, on a $A \cap A = A$.

Proposition 1.3.12. Soit un ensemble A . Alors, on a $A \cap \emptyset = \emptyset$.

1.3.2 Réunion de deux ensembles

Définition 1.3.13. On appelle réunion de deux ensembles A et B l'ensemble des éléments qui sont dans A ou (au sens inclusif) qui sont dans B .

Notation 1.3.14. La réunion de deux ensembles A et B est notée $A \cup B$.

Plus formellement, on peut écrire :

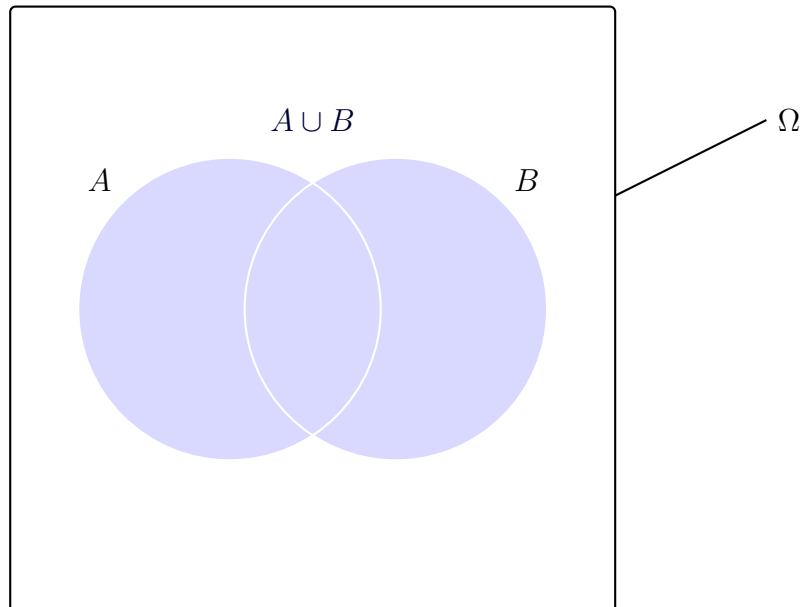
$$A \cup B = \{\omega : \omega \in A \text{ ou } \omega \in B\}$$

ou

$$\omega \in A \cup B \iff \omega \in A \text{ ou } \omega \in B.$$

Avec un diagramme de Venn :

FIGURE 1.5 – Réunion de deux ensembles



On remarque dans ce diagramme que l'on a

$$A \cap B \subset A \subset A \cup B \quad \text{et} \quad A \cap B \subset B \subset A \cup B.$$

Exemple 1.3.15 (Ensembles finis petits). Soit Ω l'alphabet français. Soient les deux ensembles de lettres : $A := \{a, b, c, d, e\}$ et $B := \{c, d, e, f, g\}$. Alors, on a : $A \cup B = \{a, b, c, d, e, f, g\}$.

Exemple 1.3.16 (Ensembles finis grands). Soit Ω l'ensemble des étudiants de Saint-Étienne. Soit A l'ensemble des étudiants de Télécom Saint-Étienne et soit B l'ensemble des étudiants de l'IUT de Saint-Étienne. Alors $A \cup B$ est l'ensemble des étudiants de TSE ou de l'IUT.

Remarque 1.3.17. À nouveau, les mots “petit” et “grand” ne recouvrent aucune réalité mathématique et doivent être compris comme étant dans leur sens courant.

Exemple 1.3.18 (Ensemble infini dénombrable). Soit $\Omega := \mathbb{Z}$ l'ensemble des entiers relatifs. Soit A le sous-ensemble des nombres pairs positifs et soit B celui des nombres impairs positifs. Alors, $A \cup B = \mathbb{N}$.

Exemple 1.3.19 (Ensemble infini non dénombrable). Soit Ω l'ensemble des réels strictement positifs. Soit $A := [1; 3]$ et soit $B :=]2; 4]$. Alors, $A \cup B$ est l'intervalle $[1; 4]$.

On présente maintenant les propriétés basiques de la réunion.

Proposition 1.3.20 (Commutativité). Soient deux ensembles A et B .

Alors $A \cup B = B \cup A$.

Proposition 1.3.21 (Associativité). Soient trois ensembles A , B et C .

Alors :

$$A \cup (B \cup C) = (A \cup B) \cup C =: A \cup B \cup C.$$

Ainsi, s'il n'y a que de la réunion entre les ensembles, les parenthèses sont superflues.

Proposition 1.3.22. Soit un ensemble A . Alors, on a $A \cup A = A$.

Proposition 1.3.23. Soit un ensemble A . Alors, on a $A \cup \emptyset = A$.

1.3.3 Propriétés de distributivité

L'associativité de la réunion ainsi que l'associativité de l'intersection permettent de s'affranchir de l'emploi de parenthèses ; à condition qu'il n'y ait que de la réunion ou que de l'intersection. Nous allons maintenant voir ce qu'il en est quand on a les deux opérations.

Nous avons vu les propriétés de la réunion et celles de l'intersection. Nous présentons maintenant les propriétés de l'intersection par rapport à la réunion et réciproquement les propriétés de la réunion par rapport à l'intersection. Considérons trois ensembles A , B et C quelconques, inclus dans Ω . Alors :

Proposition 1.3.24. L'intersection est distributive par rapport à la réunion :

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

Démonstration. Prouvons d'abord $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$. Soit $\omega \in A \cap (B \cup C)$. Par définition, ω appartient à A . De même, ω appartient à $B \cup C$. Ainsi, ω appartient à B ou il appartient à C . Si ω appartient à B , alors $\omega \in A \cap B$ d'où $\omega \in (A \cap B) \cup (A \cap C)$. Si ω appartient à C , alors $\omega \in A \cap C$ d'où $\omega \in (A \cap B) \cup (A \cap C)$.

Prouvons maintenant $A \cap (B \cup C) \supset (A \cap B) \cup (A \cap C)$. Soit $\omega \in (A \cap B) \cup (A \cap C)$. Alors, $\omega \in A \cap B$ ou $\omega \in A \cap C$. Si $\omega \in A \cap B$, alors $\omega \in A$. Mais aussi, $\omega \in B \subset B \cup C$. Donc $\omega \in A \cap (B \cup C)$. De même, si $\omega \in A \cap C$, alors $\omega \in A$. Mais aussi, $\omega \in C \subset B \cup C$. Donc $\omega \in A \cap (B \cup C)$.

La preuve est achevée en appliquant le Théorème 1.2.12. □

Proposition 1.3.25. *La réunion est distributive par rapport à l'intersection :*

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) .$$

Démonstration. Prouvons d'abord $A \cup (B \cap C) \subset (A \cup B) \cap (A \cup C)$. Soit $\omega \in A \cup (B \cap C)$. Par définition, ω appartient à A ou il appartient à $B \cap C$. Si $\omega \in A$, alors $\omega \in A \cup B$ et $\omega \in A \cup C$. Ainsi, on a $\omega \in (A \cup B) \cap (A \cup C)$. Si $\omega \in B \cap C$, alors $\omega \in B \subset A \cup B$ et $\omega \in C \subset A \cup C$. Ainsi, on a $\omega \in (A \cup B) \cap (A \cup C)$.

Prouvons maintenant $A \cup (B \cap C) \supset (A \cup B) \cap (A \cup C)$. Soit $\omega \in (A \cup B) \cap (A \cup C)$. Par définition, $\omega \in A \cup B$. Donc ω appartient à A ou il appartient à B . Si $\omega \in A$, alors $\omega \in A \cup (B \cap C)$. Si $\omega \notin A$, alors $\omega \in B$. Or, $\omega \in A \cup C$. Comme $\omega \notin A$, on a $\omega \in C$ d'où $\omega \in B \cap C \subset A \cup (B \cap C)$.

La preuve est achevée en appliquant le Théorème 1.2.12. □

Comme la réunion est distributive par rapport à l'intersection et comme l'intersection est distributive par rapport à la réunion, il faut *systématiquement* mettre des parenthèses quand on utilise l'intersection *et* la réunion.

1.3.4 Complémentaire d'un sous-ensemble

Définition 1.3.26. *Soit un ensemble Ω (l'univers des événements). Soit A un sous-ensemble de Ω . On appelle complémentaire de A dans Ω l'ensemble des éléments de Ω qui n'appartiennent pas à A .*

Notation 1.3.27. *Le complémentaire de A est noté \bar{A} ou A^c .*

Remarque 1.3.28. *Dans la plupart des cas, on ne précise pas quel est l'ensemble Ω ; ce qui transparait dans les deux notations.*

Plus formellement, on a :

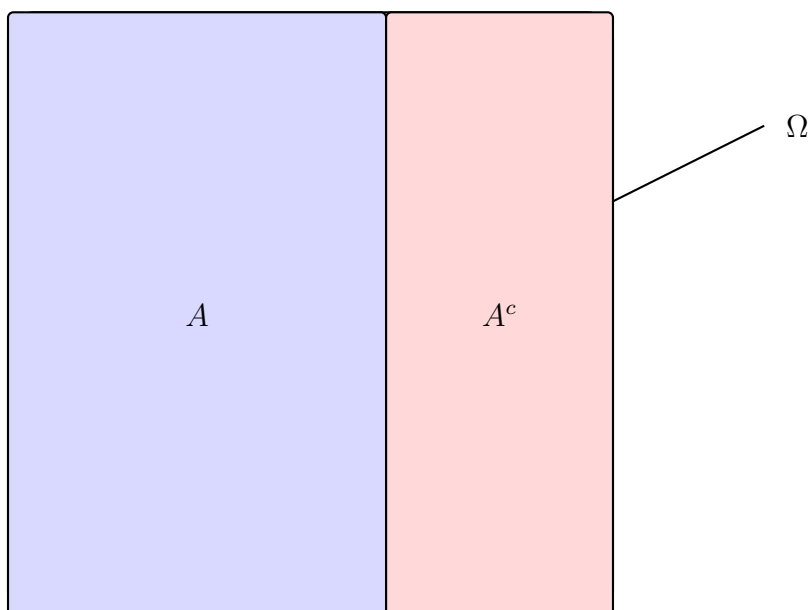
$$A^c = \{\omega \in \Omega : \omega \notin A\}$$

ou

$$\omega \in A^c \iff \omega \notin A .$$

Avec un diagramme de Venn :

FIGURE 1.6 – Complémentaire d'un sous-ensemble



On présente maintenant les propriétés classiques de la complémentation.

Proposition 1.3.29. *Soit un ensemble Ω et soit A un sous-ensemble de Ω .*

Alors, $(A^c)^c = A$.

Démonstration. Soit $\omega \in A$. Alors, $\omega \notin A^c$ donc $\omega \in (A^c)^c$. On a ainsi prouvé que l'on a $A \subset (A^c)^c$. De même, si $\omega \in (A^c)^c$, alors $\omega \notin A^c$ donc $\omega \in A$. \square

Remarque 1.3.30. *On dit que la complémentation est involutive.*

Proposition 1.3.31. *Soit un ensemble Ω . Alors, $\Omega^c = \emptyset$. De même, $\emptyset^c = \Omega$.*

Proposition 1.3.32. *Soit un ensemble Ω et soit A un sous-ensemble de Ω .*

Alors, $A \cap A^c = \emptyset$ et $A \cup A^c = \Omega$.

Théorème 1.3.33 (Lois de Morgan). *Soit un ensemble Ω et soient A et B deux sous-ensembles de Ω . Alors, on a*

$$(A \cap B)^c = A^c \cup B^c \quad (1.1)$$

et

$$(A \cup B)^c = A^c \cap B^c. \quad (1.2)$$

Les Relations (1.1) et (1.2) s'illustrent particulièrement bien en regardant les diagrammes de Venn associés.

Exercice 1.3.34. Prouver (1.2) avec des diagrammes de Venn.

Exercice 1.3.35. Démontrer rigoureusement (c'est-à-dire sans utiliser de diagramme de Venn) le Théorème 1.3.33.

Exemple 1.3.36 (Ensemble fini). Soit $\Omega := \{a, b, c, d, e, f, g, h, i\}$ un ensemble de lettres. Soient les deux sous-ensembles suivants de Ω : $A := \{a, b, c, d, e\}$ et $B := \{c, d, e, f, g\}$. Alors, on a $A^c = \{f, g, h, i\}$ et $B^c = \{a, b, h, i\}$. Mais aussi $A \cap B = \{c, d, e\}$ et $A \cup B = \{a, b, c, d, e, f, g\}$. Enfin, on a

$$A^c \cap B^c = \{h, i\} = (A \cup B)^c$$

ainsi que

$$A^c \cup B^c = \{a, b, f, g, h, i\} = (A \cap B)^c .$$

Exemple 1.3.37 (Ensemble infini dénombrable). Soit $\Omega := \mathbb{Z}$ l'ensemble des entiers relatifs. Soit A le sous-ensemble des entiers pairs et soit B le sous-ensemble des entiers divisibles par 3. Alors A^c est le sous-ensemble des entiers impairs et B^c est celui des entiers de la forme $3k + 1$ et $3k + 2$ où $k \in \mathbb{Z}$. Puis, $A^c \cap B^c$ est l'ensemble des entiers congrus à 1 modulo 6 ou à 5 modulo 6. Or, $A \cup B$ est l'ensemble des entiers congrus à 0, 2, 3 ou 4 modulo 6.

De même, $A^c \cup B^c$ est l'ensemble des entiers congrus à 1, 2, 3, 4 ou 5 modulo 6 tandis que $A \cap B$ est bien l'ensemble des entiers divisibles par 6.

Exemple 1.3.38 (Ensemble infini non dénombrable). Soit $\Omega := \mathbb{C}$. Soit A l'ensemble des complexes de partie imaginaire supérieure ou égale à 2 et soit B l'ensemble des complexes de partie imaginaire inférieure ou égale à 3. Alors $A^c = \{z \in \mathbb{C} : \Im(z) < 2\}$ et $B^c = \{z \in \mathbb{C} : \Im(z) > 3\}$. Puis,

$$\begin{aligned} A^c \cup B^c &= \{z \in \mathbb{C} : \Im(z) < 2 \text{ ou } \Im(z) > 3\} \\ &= \{z \in \mathbb{C} : 2 \leq \Im(z) \leq 3\}^c \\ &= (A \cap B)^c . \end{aligned}$$

De même :

$$\begin{aligned} A^c \cap B^c &= \{z \in \mathbb{C} : \Im(z) < 2 \text{ et } \Im(z) > 3\} \\ &= \emptyset \\ &= \Omega^c \\ &= (A \cup B)^c . \end{aligned}$$

1.4 Partition d'un ensemble

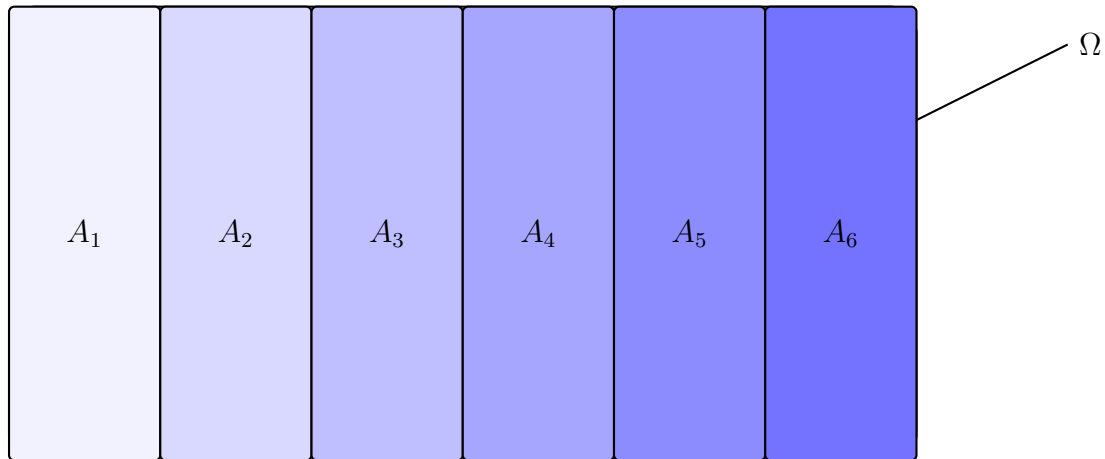
Définition 1.4.1. Soit Ω un ensemble et n sous-ensembles de Ω : A_1, \dots, A_n . On dit qu'ils forment une partition de Ω s'ils sont deux à deux disjoints et si leur réunion est égale à Ω .

Plus formellement, (A_1, \dots, A_n) est une partition de Ω si et seulement si

$$A_k \cap A_p = \emptyset \quad \text{si } 1 \leq k \neq p \leq n \quad \text{et} \quad A_1 \cup \dots \cup A_n = \Omega.$$

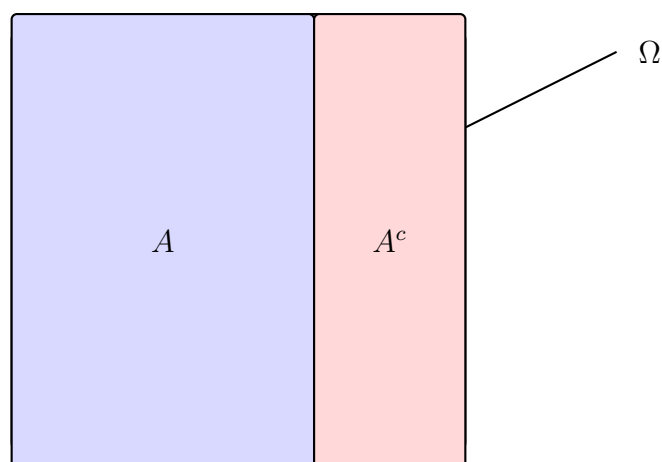
Avec un diagramme de Venn :

FIGURE 1.7 – Partition d'un ensemble



Exemple 1.4.2 (Cas particulier de partition). Soit Ω un ensemble. Soit A un sous-ensemble de Ω . Alors (A, A^c) est une partition de Ω . En effet, on a $A \cap A^c = \emptyset$ et $A \cup A^c = \Omega$ par définition. Regardons cela sur un diagramme de Venn :

FIGURE 1.8 – Partition particulière d'un ensemble



Exemple 1.4.3. Par exemple, un entier naturel est soit pair soit impair. Ainsi, $\mathbb{N} = A \cup A^c$ et $A \cap A^c = \emptyset$ avec A l'ensemble des entiers naturels pairs.

1.5 Rappels sur la dénombrabilité

Bien comprendre la dénombrabilité est essentiel en théorie des probabilités. En effet, les ensembles finis et les ensembles infinis dénombrables se comportent d'une manière très différente de celle des ensembles infinis non dénombrables. Les différences deviendront capitales quand on étudiera les variables aléatoires.

Définition 1.5.1. *On dit qu'un ensemble est de cardinal fini s'il contient un nombre fini d'éléments.*

Exemple 1.5.2. *Par exemple, l'ensemble $\llbracket 1; 17\ 000 \rrbracket$ est fini puisqu'il contient un nombre fini d'éléments (17 000).*

Définition 1.5.3. *Le cardinal d'un ensemble fini est le nombre d'éléments de cet ensemble.*

Définition 1.5.4. *Le cardinal d'un ensemble E est noté $\#E$.*

Exemple 1.5.5. *Le cardinal de l'ensemble $\llbracket 0; 4\ 815\ 162\ 342 \rrbracket$ est 4 815 162 343.*

Remarque 1.5.6. *L'élément 0 compte évidemment et il ne faut donc pas l'oublier.*

Définition 1.5.7. *Un ensemble qui n'est pas de cardinal fini est dit infini.*

Exemple 1.5.8. *L'ensemble \mathbb{N} est infini. De même, \mathbb{R} est infini.*

Définition 1.5.9. *On dit qu'un ensemble infini est dénombrable s'il est en bijection avec \mathbb{N} .*

Que deux ensembles soient en bijection signifie *grosso modo* qu'ils contiennent le même nombre d'éléments.

Exemple 1.5.10. *L'ensemble des rationnels, \mathbb{Q} , est dénombrable.*

Contre-exemple 1.5.11. *L'ensemble des réels, \mathbb{R} , n'est pas dénombrable.*

Proposition 1.5.12. *Une réunion finie ou dénombrable d'ensembles finis ou dénombrables est soit un ensemble fini soit un ensemble infini dénombrable.*

Remarque 1.5.13. *Il arrive parfois que le mot "dénombrable" signifie en fait "fini ou infini dénombrable".*

Le modèle probabiliste

2.1 Introduction

Dans ce chapitre, on donne la modélisation inhérente à toute expérience aléatoire, ce qui est la source des probabilités.

Pour pouvoir aborder ce chapitre, il est conseillé d'avoir quelques notions de dénombrement et de combinatoire. Également, avoir lu le chapitre précédent est primordial.

Les principaux objectifs sont les suivants. D'abord, à l'issue du chapitre, on attend du lecteur qu'il sache associer aussi bien les événements aux sous-ensembles de l'univers que les systèmes complets d'événements aux partitions de l'univers. Il est également primordial de comprendre que la disjonction et l'indépendance sont deux notions radicalement différentes. Nous verrons aussi ce qu'est une probabilité (les trois axiomes qui définissent ladite notion de probabilité) ainsi que ses propriétés principales (probabilité du complémentaire, formule des probabilités totales, croissance de la probabilité). Par la suite, il sera essentiel de bien comprendre ce qu'est une probabilité conditionnelle (la définition et aussi l'objet que celle-ci représente). Le lecteur sera enfin amené à découvrir la formule de Bayes et la définition de l'indépendance entre deux événements. Nous le rappelons : l'indépendance et la disjonction sont deux notions fondamentalement différentes.

2.2 Modélisation d'une expérience aléatoire

Le résultat d'une expérience est dit aléatoire lorsqu'on ne peut pas le prédire avec certitude. Plusieurs raisons peuvent expliquer cette impossibilité de prédiction : on ne connaît pas toutes les conditions initiales (les causes) ou on ne sait pas déterminer comment le système évolue précisément. L'expérience est alors dite aléatoire.

Exemple 2.2.1. *Jeter un dé à six faces est une expérience aléatoire. On peut énumérer quelques causes de variabilité du résultat :*

- *position exacte de la main (hauteur, angles...),*
- *position exacte du dé dans la main,*
- *vitesse à laquelle le dé est lancé de la main,*

- état exact de la surface de réception,
- état exact du dé (coins plus ou moins arrondis, poids du dé...),
- résistance de l'air,

et beaucoup d'autres. Il est difficile de fixer a priori les valeurs exactes des différents paramètres.

Cependant, on constate que si l'on jette un dé à six faces un grand nombre de fois, on obtient en moyenne :

- une fois sur six la face 1,
- une fois sur six la face 2,
- une fois sur six la face 3,
- une fois sur six la face 4,
- une fois sur six la face 5,
- une fois sur six la face 6.

À partir de là, on peut définir une loi de variabilité du résultat.

Exemple 2.2.2. On peut aussi penser à l'expérience consistant à faire un enfant : celui-ci sera-t-il un garçon ou une fille ? Au moment de la conception, il est difficile de prévoir le résultat de l'expérience.

La théorie des probabilités vise à fournir un modèle mathématique pour décrire ces phénomènes.

2.2.1 L'espace fondamental

On définit les résultats possibles (ω) d'une expérience aléatoire.

Exemple 2.2.3. *Expérience aléatoire : jeter un dé. Les résultats possibles sont alors*

- ω_1 := "on obtient la face 1",
- ω_2 := "on obtient la face 2",
- ω_3 := "on obtient la face 3",
- ω_4 := "on obtient la face 4",
- ω_5 := "on obtient la face 5",
- ω_6 := "on obtient la face 6".

Définition 2.2.4. *L'ensemble de tous les résultats possibles d'une expérience aléatoire est appelé l'espace fondamental. On parle aussi d'univers.*

Notation 2.2.5. *L'univers est noté Ω .*

Remarque 2.2.6. *La définition des résultats possibles (et donc de l'univers) n'est pas nécessairement unique. Reprenons l'Exemple 2.2.3. On pourrait considérer $\tilde{\Omega} := \{p, i\}$ où p est le résultat "la face obtenue est paire" et i est le résultat "la face obtenue est impaire". Mais alors, il n'est pas possible de connaître*

la valeur précise sur la face du dé. On pourrait également considérer l'univers $\widehat{\Omega} := \{(\omega_k, x) : 1 \leq k \leq 6, x \in \mathbb{R}^2\}$ où l'on se donnerait de plus la position sur la surface de réception du centre de la face du dessous... Mais est-ce utile ?

L'enjeu du choix de l'univers répond donc à deux contraintes. Il faut que l'univers soit suffisamment riche pour répondre aux questions que l'on se pose. Mais il faut également qu'il soit suffisamment simple pour que l'on puisse travailler dessus.

Remarque 2.2.7. *L'espace fondamental associé à une expérience aléatoire peut être fini, infini dénombrable ou infini non dénombrable.*

Exemple 2.2.8 (Univers fini). *On considère l'expérience aléatoire suivante : on tire au hasard un individu d'une population de N individus numérotés de 1 à N . Certains de ces individus vérifient une propriété (P).*

Il peut s'agir d'une population d'électeurs et la propriété (P) correspond alors à une intention de vote. Il peut aussi s'agir d'une population de pièces fabriquées par une machine et la propriété (P) correspond alors à la défectuosité de la pièce.

Première définition : Les résultats possibles sont $\omega_k :=$ "on obtient l'individu k ". L'espace fondamental est alors $\Omega := \{\omega_1, \dots, \omega_N\}$. Il est fini.

Seconde définition : Les résultats possibles sont $\omega_+ :=$ "l'individu obtenu a la propriété (P)" et $\omega_- :=$ "l'individu obtenu n'a pas la propriété (P)". L'espace fondamental est alors $\Omega := \{\omega_+, \omega_-\}$. Il est fini également.

Exemple 2.2.9 (Univers infini dénombrable). *On considère l'expérience aléatoire suivante : on observe le nombre d'octets échangés dans un système de pair à pair durant un intervalle de temps de durée fixée.*

Les résultats possibles sont ainsi $\omega_k :=$ "k octets sont échangés durant l'intervalle de temps" avec $k \in \mathbb{N}$. L'espace fondamental est alors

$$\Omega := \{\omega_k : k \in \mathbb{N}\} .$$

Il est infini dénombrable.

Exemple 2.2.10 (Univers infini non dénombrable). *On considère l'expérience aléatoire suivante : on observe la durée de vie, en heures, d'un composant électronique. Il s'agit d'une étude de fiabilité.*

Les résultats possibles sont ainsi $\omega_t :=$ "la durée de vie du composant est t " avec $t \geq 0$. L'espace fondamental est alors $\Omega := \{\omega_t : t \geq 0\}$. Il est infini non dénombrable.

2.2.2 Les évènements

2.2.2.1 Définition

Définition 2.2.11. *On appelle évènement associé à une expérience aléatoire toute proposition logique relative au résultat de cette expérience.*

Exemple 2.2.12. *Si l'expérience aléatoire est le lancer de dé, un évènement peut être $A :=$ "obtenir une face paire".*

Si l'expérience aléatoire est la quantité d'octets échangés dans un système de pair à pair durant un intervalle de temps de durée fixée, un évènement peut être $A :=$ "Plus de mille giga-octets ont été échangés en une heure".

Contre-exemple 2.2.13. *Si l'expérience aléatoire est le lancer de dé, $B :=$ "il va pleuvoir demain" n'est pas un évènement.*

On représente un évènement par l'ensemble des résultats possibles qui vérifient la proposition qui le définit : c'est un sous-ensemble de l'espace fondamental Ω . À partir de maintenant, on confond l'évènement et le sous-ensemble qui le représente.

Ainsi, un évènement n'est rien d'autre qu'un sous-ensemble de l'univers Ω .

Exemple 2.2.14. *Si l'expérience aléatoire est le lancer de dé, on pose $A :=$ "on obtient une face paire". L'espace fondamental est $\Omega := \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ et l'on écrit $A = \{\omega_2, \omega_4, \omega_6\}$.*

Remarque 2.2.15. *L'ensemble de tous les évènements que l'on peut lier à une variable aléatoire est par définition stable par intersection et par réunion. On peut également considérer le contraire d'un évènement ce qui revient au complémentaire du point de vue ensembliste. Cet ensemble d'évènements doit aussi contenir l'évènement certain (ce qui correspond à tout l'univers). De plus, l'ensemble des évènements doit être stable par réunion dénombrable, ce qui signifie du point de vue pratique que l'on peut faire un nombre dénombrable d'expériences aléatoires.*

Une telle classe d'ensembles, stable par réunion dénombrable, par intersection, par la complémentation et qui contient l'univers est appelé une tribu. On a choisi, dans ce livre, de ne pas entrer dans les détails de cette théorie, voir le Chapitre 24 à la page 341 pour plus d'informations.

2.2.2.2 Évènement certain, impossible

Définition 2.2.16 (Évènement certain). *On dit qu'un évènement A est certain si tous les résultats vérifient la proposition logique qui le définit : $A = \Omega$.*

Définition 2.2.17 (Évènement impossible). *On dit qu'un évènement A est impossible si aucun résultat possible ne vérifie la proposition logique qui le définit : $A = \emptyset$.*

Remarque 2.2.18. *Jusqu'à maintenant, on ne parle pas de probabilités.*

2.2.2.3 Notion d'implication

Définition 2.2.19. Soient deux évènements A et B associés à une même expérience aléatoire. On dit que A implique B lorsque à chaque fois que A est réalisé, B l'est également.

Notation 2.2.20. Si A implique B , on note $A \Rightarrow B$.

Exemple 2.2.21. On jette un dé. On pose $A :=$ "on obtient une face plus grande que trois". On pose $B :=$ "on obtient une face plus grande que deux". Alors, A implique $B : A \Rightarrow B$.

Remarque 2.2.22 (Traduction ensembliste). À chaque fois que A est réalisé, B l'est également. En d'autres termes, pour tout $\omega \in A$, on a $\omega \in B$. Ceci signifie $A \subset B$. On a donc

$$(A \Rightarrow B) \iff (A \subset B) .$$

Exemple 2.2.23. On se place dans le cadre de l'Exemple 2.2.21.

Alors, $B = \{\omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ et $A = \{\omega_3, \omega_4, \omega_5, \omega_6\}$. On a donc bien $A \subset B$.

Remarque 2.2.24. Pour en revenir à l'exemple précédent, il convient de toujours raisonner tranquillement. En effet, il est très facile d'obtenir un résultat en probabilités. Malheureusement, si on n'y prend pas garde, ce résultat risque d'être faux.

2.2.2.4 Opérations sur les évènements

Soient deux évènements A et B associés à une même expérience aléatoire.

Définition 2.2.25. On considère l'évènement E qui est réalisé si et seulement si A et B sont tous les deux réalisés. Traduction ensembliste : $E = A \cap B$.

Définition 2.2.26. On considère l'évènement E qui est réalisé si et seulement si A ou B est réalisé. Traduction ensembliste : $E = A \cup B$.

Définition 2.2.27 (Contraire d'un évènement). On considère l'évènement E qui est réalisé si et seulement si A n'est pas réalisé. Traduction ensembliste : $E = A^c = \bar{A}$.

2.2.2.5 Évènements incompatibles

Soient deux évènements A et B associés à une même expérience aléatoire.

Définition 2.2.28. Si les deux évènements A et B ne peuvent pas être réalisés simultanément, on dit qu'ils sont incompatibles. Traduction ensembliste : $A \cap B = \emptyset$. On dit aussi que A et B sont disjoints.

Il est essentiel de différencier la notion d'évènements disjoints de la notion d'indépendance d'évènements que nous verrons par la suite. Nous insistons donc pour que le lecteur soit précautionneux à ce sujet.

Exemple 2.2.29. On reprend l'Exemple 2.2.10. On considère la durée de vie d'un composant électronique. On pose $S_{t_1} :=$ "la durée de vie du composant est supérieure à t_1 " et $I_{t_2} :=$ "la durée de vie du composant est inférieure à t_2 ". La traduction ensembliste de ces deux évènements est donc la suivante :

$$S_{t_1} = \{\omega_t : t \geq t_1\}$$

$$\text{et } I_{t_2} = \{\omega_t : 0 \leq t \leq t_2\} .$$

Ainsi, si $t_1 \leq t_2$, on a

$$E := S_{t_1} \cap I_{t_2} = \{\omega_t : t_1 \leq t \leq t_2\} .$$

Mais, si $t_1 > t_2$, on a

$$E := S_{t_1} \cap I_{t_2} = \emptyset .$$

Les deux évènements sont alors incompatibles.

2.2.2.6 Système complet d'évènements

Définition 2.2.30. Soient A_1, \dots, A_n des évènements associés à une même expérience aléatoire. On dit qu'ils forment un système complet d'évènements si les deux hypothèses suivantes sont satisfaites :

- les évènements sont deux à deux incompatibles,
- à chaque répétition de l'expérience aléatoire, l'un des n évènements est réalisé.

Remarque 2.2.31 (Traduction ensembliste). (A_1, \dots, A_n) est un système complet d'évènements de l'expérience aléatoire si et seulement si (A_1, \dots, A_n) est une partition de l'espace fondamental Ω .

Exemple 2.2.32 (Cas particulier). Soit A un évènement quelconque d'une expérience aléatoire. Alors, (A, A^c) est un système complet d'évènements.

Exemple 2.2.33. L'expérience aléatoire que l'on considère est la suivante. On tire au hasard une carte dans un jeu de 32 cartes. On se donne les évènements :

- $A_1 :=$ "On obtient un carreau",
- $A_2 :=$ "On obtient un cœur",
- $A_3 :=$ "On obtient un pique",
- $A_4 :=$ "On obtient un trèfle".

Alors, (A_1, A_2, A_3, A_4) est un système complet d'évènements.

Exemple 2.2.34. L'expérience aléatoire que l'on considère est la suivante. On tire au hasard un nombre entier entre 1 et 100. On se donne les évènements :

- $A_0 :=$ "Le nombre est congru à 0 modulo 3",
- $A_1 :=$ "Le nombre est congru à 1 modulo 3",
- $A_2 :=$ "Le nombre est congru à 2 modulo 3".

Alors, (A_0, A_1, A_2) est un système complet d'évènements.

2.3 Définition et propriétés d'une probabilité

2.3.1 Définition

On commence d'abord par donner une vision intuitive de la notion de probabilité. Si l'on réalise une suite de jets d'un dé à six faces bien équilibré, on peut constater que la fréquence d'obtention de chaque face s'approche de $\frac{1}{6}$. Il s'agit d'une définition objective basée sur l'expérience.

On peut aussi opter pour un point de vue plus subjectif : en raison de la symétrie du dé, on a une chance sur six d'obtenir chaque face.

On peut faire plusieurs observations :

- la fréquence de réalisation est un nombre compris entre 0 et 1,
- la fréquence de réalisation de l'évènement certain est égale à 1,
- la fréquence de réalisation de $A \cup B$ est la somme des fréquences de réalisation de A et de B lorsque A et B sont incompatibles. Cette propriété peut être étendue au cas d'une suite finie ou infinie dénombrable d'évènements deux à deux incompatibles.

On axiomatise (définit) maintenant ce qu'est une probabilité.

Définition 2.3.1. Soit une expérience aléatoire e . On note Ω l'espace fondamental associé. Une probabilité est une application \mathbb{P} qui, à tout évènement A associé à e (c'est-à-dire à tout sous-ensemble de Ω), fait correspondre un nombre noté $\mathbb{P}(A)$ (que l'on appelle "probabilité de A ") et qui vérifie les trois axiomes suivants :

$$(A_1) \quad 0 \leq \mathbb{P}(A) \leq 1.$$

$$(A_2) \quad \mathbb{P}(\Omega) = 1.$$

(A₃) Pour toute famille $(A_k)_{k \in \mathcal{I}}$ d'évènements deux à deux incompatibles où \mathcal{I} est un ensemble fini ou infini dénombrable d'indices, on a

$$\mathbb{P} \left(\bigcup_{k \in \mathcal{I}} A_k \right) = \sum_{k \in \mathcal{I}} \mathbb{P}(A_k) .$$

Remarque 2.3.2. La troisième hypothèse porte aussi le nom d'axiome des probabilités totales.

Remarque 2.3.3. La probabilité \mathbb{P} n'est pas une application de Ω dans $[0; 1]$. Si Ω est fini ou infini dénombrable, nous verrons que la donnée de \mathbb{P} sur les singletons inclus dans Ω suffit à caractériser \mathbb{P} . De fait, vous avez peut-être l'habitude de raisonner avec \mathbb{P} comme application de l'univers dans $[0; 1]$. C'est une mauvaise habitude qu'il faudra perdre aussi vite que possible.

Remarque 2.3.4. En fait, on peut observer que \mathbb{P} est une application de la tribu des évènements \mathcal{F} (l'ensemble de tous les évènements liés à une expérience aléatoire) vers $[0; 1]$. On dit d'ailleurs que $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de probabilité.

En effet, en admettant l'axiome du choix, il peut être démontré que dans le cas d'un ensemble infini Ω , certaines probabilités ne sont pas définissables sur la tribu la plus riche, à savoir 2^Ω . Conséquemment, on doit se restreindre à une tribu plus adaptée.

Dans ce livre, nous faisons fi de la notion de tribu, et nous y revenons au Chapitre 24 à la page 341.

2.3.2 Propriétés

À partir de ces trois axiomes, on peut déduire un certain nombre de propriétés que l'on présente à présent.

2.3.2.1 Probabilité de l'évènement contraire

Soit un évènement A de probabilité $\mathbb{P}(A)$. Par définition du complémentaire (voir Définition 2.2.27), A et A^c sont incompatibles. Or, d'après l'axiome (A_2) , on a

$$1 = \mathbb{P}(\Omega).$$

On a aussi $\Omega = A \cup A^c$. Ainsi, il vient

$$\mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c),$$

d'après l'axiome des probabilités totales, vu que A et A^c sont incompatibles. Conséquemment, on a le résultat suivant.

Théorème 2.3.5. *Soit un évènement A . Alors, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.*

Remarque 2.3.6. *Comme $\Omega^c = \emptyset$, on en déduit $\mathbb{P}(\emptyset) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0$. La probabilité de l'évènement impossible est nulle.*

Il convient de souligner qu'un évènement de probabilité nulle n'est pas nécessairement impossible. De même, un évènement de probabilité égale à 1 n'est pas nécessairement l'évènement certain. On dira d'ailleurs plutôt qu'un tel évènement est presque sûr.

On peut généraliser le résultat du Théorème 2.3.5 à un système complet d'évènements. Si (A_1, \dots, A_n) est un système complet d'évènements, alors

$$\mathbb{P}(A_1) + \dots + \mathbb{P}(A_n) = 1.$$

2.3.2.2 Décomposition de tout évènement B

Soit une expérience aléatoire e . On note Ω l'espace fondamental associé. Soit un système complet d'évènements, (A_1, \dots, A_n) . Soit maintenant B un évènement. On a alors le résultat suivant

Théorème 2.3.7.

$$\mathbb{P}(B) = \mathbb{P}(B \cap A_1) + \cdots + \mathbb{P}(B \cap A_n) .$$

Démonstration. Étape 1. Prouvons d'abord que l'on a

$$B = (B \cap A_1) \cup \cdots \cup (B \cap A_n) .$$

Comme B est un évènement lié à l'expérience aléatoire e , on a $B \subset \Omega$. Consé-
quemment, on trouve $B = B \cap \Omega$ d'où

$$B = B \cap (A_1 \cup \cdots \cup A_n) .$$

D'après la propriété de distributivité de l'intersection par rapport à la réunion (Proposition 1.3.24 à la page 15), il vient

$$B = (B \cap A_1) \cup \cdots \cup (B \cap A_n) .$$

Étape 2. Prouvons maintenant que les évènements $B \cap A_k$ sont deux à deux incompatibles.

En effet, pour $k \neq p$, on a

$$(B \cap A_k) \cap (B \cap A_p) = B \cap A_k \cap B \cap A_p ,$$

par associativité de l'intersection. Puis, en utilisant la propriété de commutativité de l'intersection, on obtient

$$(B \cap A_k) \cap (B \cap A_p) = B \cap B \cap A_k \cap A_p .$$

En utilisant à nouveau l'associativité, on trouve

$$(B \cap A_k) \cap (B \cap A_p) = (B \cap B) \cap (A_k \cap A_p) .$$

Or, $B \cap B = B$ et $A_k \cap A_p = \emptyset$ vu que (A_1, \dots, A_n) est un système complet d'évènements. Par conséquent :

$$(B \cap A_k) \cap (B \cap A_p) = B \cap \emptyset .$$

Or, l'intersection de tout ensemble avec l'ensemble vide est l'ensemble vide. On a donc bien

$$(B \cap A_k) \cap (B \cap A_p) = \emptyset ,$$

comme annoncé au début de l'Étape 2.

Étape 3. On achève la preuve en appliquant l'axiome des probabilités totales. □

Cette décomposition est très utilisée dans les problèmes de dénombrement. Donnons un exemple simple.

Exemple 2.3.8. On considère l'expérience aléatoire suivante. On tire une carte au hasard dans un jeu de 32 cartes. On se donne les quatre évènements suivants :

- A_1 := "on obtient un cœur",
- A_2 := "on obtient un carreau",
- A_3 := "on obtient un pique",
- A_4 := "on obtient un trèfle".

Alors (A_1, A_2, A_3, A_4) est un système complet d'évènements. Soit B l'évènement "on obtient un as". On a alors

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(B \cap A_1) + \mathbb{P}(B \cap A_2) + \mathbb{P}(B \cap A_3) + \mathbb{P}(B \cap A_4) \\ &= \mathbb{P}(\{\text{on obtient l'as de cœur}\}) + \mathbb{P}(\{\text{on obtient l'as de carreau}\}) \\ &\quad + \mathbb{P}(\{\text{on obtient l'as de pique}\}) + \mathbb{P}(\{\text{on obtient l'as de trèfle}\}) . \end{aligned}$$

Chacun de ces quatre évènements est de probabilité $\frac{1}{32}$ d'où l'on a

$$\mathbb{P}(B) = 4 \frac{1}{32} = \frac{4}{32} = \frac{1}{8} .$$

À nouveau, on peut s'intéresser au cas particulier du système complet (A, A^c) . Alors, pour tout évènement B , on a

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c) . \quad (2.1)$$

2.3.2.3 Croissance de la probabilité

Soit une expérience aléatoire e . Soit Ω l'espace fondamental. On considère deux évènements A et B liés à cette expérience aléatoire. On a le résultat suivant de croissance de la probabilité

Théorème 2.3.9. Si A implique B c'est-à-dire si $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Démonstration. Réécrivons l'Équation (2.1) :

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c) .$$

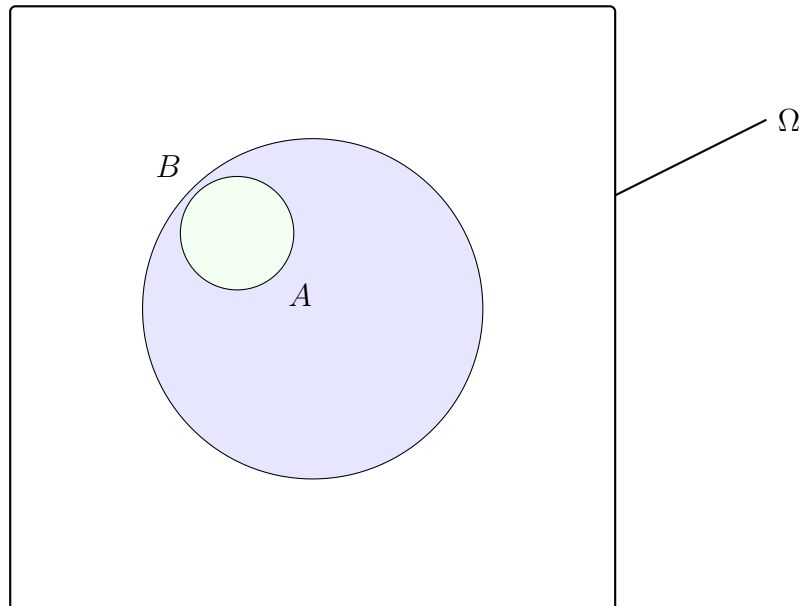
Comme $A \subset B$, on a $A \cap B = A$ d'où

$$\mathbb{P}(B) = \mathbb{P}(A) + \underbrace{\mathbb{P}(B \cap A^c)}_{\geq 0} \geq \mathbb{P}(A) ,$$

ce qui achève la preuve. □

On le voit bien aussi sur un diagramme de Venn :

FIGURE 2.1 – Croissance de la probabilité



2.3.2.4 Théorème des probabilités totales

L'axiome des probabilités totales concerne la réunion de deux événements incompatibles. L'objet de ce paragraphe est de regarder ce qu'il se passe lorsque les deux événements ne sont pas incompatibles.

Théorème 2.3.10. *Soit une expérience aléatoire e . On note Ω l'espace fondamental associé. Soient deux événements A et B . Alors, on a*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \quad (2.2)$$

Démonstration. On applique l'Égalité (2.1) à $A \cup B$ et (A, A^c) :

$$\mathbb{P}(A \cup B) = \mathbb{P}((A \cup B) \cap A) + \mathbb{P}((A \cup B) \cap A^c).$$

Or, $(A \cup B) \cap A = A$ car $A \subset A \cup B$. Et, en utilisant la distributivité de l'intersection par rapport à la réunion, on a

$$(A \cup B) \cap A^c = (A \cap A^c) \cup (B \cap A^c) = \emptyset \cup (B \cap A^c) = B \cap A^c. \quad (2.3)$$

On en déduit ainsi $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$. Or, en appliquant l'Égalité (2.1) à B et (A, A^c) , on obtient

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c),$$

d'où $\mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$. Ainsi, en appliquant ceci, on trouve l'Égalité (2.2). \square

Remarque 2.3.11. *Cette formule ressemble à la formule de Grassmann en algèbre linéaire.*

Exercice 2.3.12. *Prouver le Théorème 2.3.10 en utilisant un diagramme de Venn.*

Exercice 2.3.13. *Soient trois évènements A , B et C . Calculer $\mathbb{P}(A \cup B \cup C)$, sans utiliser les diagrammes de Venn.*

On peut généraliser ces formules à la réunion de n évènements. Il s'agit du lemme de Poincaré.

2.3.3 Espace fondamental dénombrable

Dans cette section, $\Omega := \{\omega_i, i \in I\}$ où I est un ensemble fini ou infini dénombrable d'indices.

2.3.3.1 Caractérisation de la probabilité

On se donne une probabilité \mathbb{P} sur l'univers Ω , c'est-à-dire une fonction de 2^Ω dans \mathbb{R} qui vérifie les trois axiomes de la Définition 2.3.1.

Théorème 2.3.14. *La probabilité \mathbb{P} sur l'espace fondamental dénombrable Ω est entièrement déterminée par la donnée des probabilités des évènements élémentaires (c'est-à-dire celles des singletons) : $\mathbb{P}(\{\omega_k\}) =: \mathbb{P}(\omega_k)$.*

Démonstration. On suppose connues les probabilités élémentaires $\mathbb{P}(\omega_k)$, pour tout $k \in I$. Montrons alors que l'on peut calculer la probabilité de tout évènement. Soit un évènement A quelconque. En tant que sous-ensemble de Ω , il est fini ou infini dénombrable. Donc, il existe un ensemble fini ou infini dénombrable d'indices $\mathcal{J} \subset \mathcal{I}$ tel que $A = \{\omega_j : j \in \mathcal{J}\}$. De fait, on a

$$A = \bigcup_{j \in \mathcal{J}} \{\omega_j\}.$$

Or, les singletons sont deux à deux incompatibles. Ainsi, d'après l'axiome des probabilités totales, on a

$$\mathbb{P}(A) = \sum_{j \in \mathcal{J}} \mathbb{P}(\omega_j).$$

La probabilité $\mathbb{P}(A)$ est donc entièrement déterminée pour tout $A \in 2^\Omega$. □

Remarque 2.3.15. *En appliquant ce raisonnement à $\Omega = \{\omega_i, i \in I\}$, on a*

$$\sum_{i \in I} \mathbb{P}(\omega_i) = 1. \tag{2.4}$$

Il est important de comprendre que le Théorème 2.3.14 n'est pas *a priori* évident. En effet, on a ici prouvé que l'on peut caractériser la probabilité \mathbb{P} qui est, par définition, définie sur 2^Ω , par sa restriction à $\{\{\omega\} : \omega \in \Omega\}$, qui est en bijection avec Ω . Or, on peut montrer à l'aide de la théorie des ensembles qu'il n'existe aucune bijection de Ω vers 2^Ω .

On peut se poser la question de la caractérisation d'une probabilité si l'espace fondamental n'est pas dénombrable.

Théorème 2.3.16. *Soit un univers Ω non dénombrable et muni d'une probabilité \mathbb{P} . On note $M \subset \Omega$ le sous-ensemble qui contient les éléments ω de probabilité strictement positive. Alors M est fini ou infini dénombrable.*

Démonstration. Pour tout $n \in \mathbb{N}$, on pose $M_n := \{\omega : \mathbb{P}(\omega) \geq \frac{1}{n}\}$. On observe que l'on a $M = \bigcup_{n=1}^{\infty} M_n$. Il suffit maintenant de prouver que l'ensemble M_n est fini ou infini dénombrable pour tout $n \in \mathbb{N}^*$. En effet, une réunion dénombrable d'ensembles finis ou infinis dénombrables est un ensemble fini ou infini dénombrable d'après la Proposition 1.5.12. Supposons par l'absurde que l'un des ensembles M_n est infini. En particulier, il contient au moins $2n$ éléments distincts que l'on note $\omega_1, \dots, \omega_{2n}$. Puis, il vient $\mathbb{P}(M_n) \geq \sum_{i=1}^{2n} \mathbb{P}(\omega_i) \geq 2n \times \frac{1}{n} = 2$ ce qui est impossible car \mathbb{P} est une probabilité et donc on a $\mathbb{P}(M_n) \leq \mathbb{P}(\Omega) = 1$. \square

À partir de là, deux cas sont possibles. Ou bien $\mathbb{P}(M) = 1$, ce qui signifie que l'on peut considérer M comme nouvel espace fondamental. Ou bien $\mathbb{P}(M) < 1$, ce qui signifie qu'une partie de la "masse" est cachée dans les éléments de probabilité égale à 0, lesquels sont en quantité non dénombrable. Dans ce dernier cas, on ne peut caractériser la probabilité \mathbb{P} par la seule donnée de la nullité de \mathbb{P} sur chaque événement élémentaire de M^c .

Exemple 2.3.17. *On pose $\Omega := [0; 1]$. On considère la probabilité \mathbb{P}_1 définie par*

$$\mathbb{P}_1(\mathcal{I}) := \int_{\mathcal{I}} dx,$$

où \mathcal{I} est un intervalle quelconque inclus dans $[0; 1]$. On considère la probabilité \mathbb{P}_2 définie par

$$\mathbb{P}_2(\mathcal{I}) := \int_{\mathcal{I}} 2x dx.$$

On peut vérifier facilement que \mathbb{P}_1 et \mathbb{P}_2 sont deux probabilités. Également, ces deux probabilités ne sont pas les mêmes puisque $\mathbb{P}_1([0; \frac{1}{2}]) = \frac{1}{2}$ alors que $\mathbb{P}_2([0; \frac{1}{2}]) = \frac{1}{4}$. Et pourtant, pour tout $\omega \in \Omega = [0; 1]$, on a $\mathbb{P}_1(\{\omega\}) = \mathbb{P}_2(\{\omega\}) = 0$.

2.3.3.2 Cas particulier de l'équiprobabilité

Soit un espace fondamental Ω fini : $\Omega = \{\omega_1, \dots, \omega_N\}$ avec $N < \infty$. On a donc $N = \#\Omega$.

Définition 2.3.18. Dans le cas particulier où $\mathbb{P}(\omega_1) = \dots = \mathbb{P}(\omega_N)$, on dit que l'espace fondamental fini Ω est muni de l'équiprobabilité.

Théorème 2.3.19. Si \mathbb{P} est l'équiprobabilité sur $\Omega = \{\omega_1, \dots, \omega_N\}$, on a $\mathbb{P}(\omega_k) = \frac{1}{N} = \frac{1}{\#\Omega}$ pour tout $1 \leq k \leq N$.

Démonstration. On note $p := \mathbb{P}(\omega_1) = \dots = \mathbb{P}(\omega_N)$. D'après l'Égalité (2.4), on a

$$\sum_{k=1}^N p = 1.$$

D'où $Np = 1$ ce qui implique $p = \frac{1}{N}$. □

Remarque 2.3.20. Une probabilité sur un univers Ω fini n'est pas nécessairement l'équiprobabilité.

Comme on a calculé les probabilités élémentaires, on a donc entièrement caractérisé la probabilité \mathbb{P} et l'on peut ainsi calculer $\mathbb{P}(A)$ pour tout évènement $A \subset \Omega$.

Théorème 2.3.21. Si \mathbb{P} est l'équiprobabilité sur $\Omega = \{\omega_1, \dots, \omega_N\}$, on a

$$\mathbb{P}(A) = \frac{\#\{\text{cas favorables}\}}{\#\{\text{cas possibles}\}} = \frac{\#A}{\#\Omega}.$$

pour tout évènement A .

Exemple 2.3.22. On jette un dé à six faces que l'on suppose non pipé. Les six résultats possibles $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5$ et ω_6 sont équiprobables. Ainsi, l'univers $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ est muni de l'équiprobabilité :

$$\mathbb{P}(\omega_k) = \frac{1}{6},$$

pour tout $1 \leq k \leq 6$. Soit l'évènement $A :=$ "on obtient une face paire". Alors, $A = \{\omega_2, \omega_4, \omega_6\}$ puis

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{3}{6} = \frac{1}{2}.$$

Exemple 2.3.23. On prend huit cartes dans un jeu de trente-deux cartes. On a exactement $\binom{32}{8} = \frac{32!}{8!24!} = 10\,518\,300$ façons de le faire. L'espace fondamental Ω est muni de l'équiprobabilité. On considère l'évènement $A :=$ "on ne prend aucun as". Il vient alors

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{\binom{28}{8}}{\binom{32}{8}} \approx 0.295.$$

Définition 2.3.24. On dit que le tirage d'un individu de la population P finie est au hasard lorsque tout individu a la même probabilité d'être tiré. En d'autres termes, un tirage au hasard est fait suivant l'équiprobabilité.

2.4 Probabilité conditionnelle et indépendance

2.4.1 Probabilité conditionnelle

2.4.1.1 Notion de probabilité conditionnelle

Soient A et B deux évènements associés à une même expérience aléatoire e . On note Ω l'espace fondamental associé. On suppose $\mathbb{P}(A) \neq 0$.

La probabilité conditionnelle de B par rapport à l'évènement A est la probabilité que B soit réalisé lorsque l'on sait que A est réalisé.

Notation 2.4.1. La probabilité conditionnelle de B en sachant A est notée $\mathbb{P}_A(B)$ ou $\mathbb{P}(B|A)$. On préfère la seconde notation. En effet, par la suite, \mathbb{P}_X représentera la loi de probabilité de toute variable aléatoire X . Néanmoins, si aucune confusion n'est possible, on s'autorisera de temps en temps la première notation.

Exemple 2.4.2. On jette un dé. Soient les évènements $A :=$ "on obtient une face paire" et $B :=$ "on obtient une face supérieure ou égale à trois". L'univers $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ est muni de l'équiprobabilité. Ici, $A = \{\omega_2, \omega_4, \omega_6\}$ d'où $\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}$. Et, $B = \{\omega_3, \omega_4, \omega_5, \omega_6\}$ d'où $\mathbb{P}(B) = \frac{4}{6} = \frac{2}{3}$. Puis, $A \cap B = \{\omega_4, \omega_6\}$ d'où $\mathbb{P}(A \cap B) = \frac{2}{6} = \frac{1}{3}$.

On jette le dé. Supposons que l'on sache que l'on a obtenu une face paire. (Pour ce faire, il suffit de relancer le dé jusqu'à ce que la face obtenue soit paire.) Quelle est la probabilité qu'on obtienne une face supérieure ou égale à trois ? Dans notre exemple, l'univers est A et la probabilité que B soit réalisé est donc la probabilité que $A \cap B$ soit réalisé pour l'univers A muni de l'équiprobabilité associée \mathbb{P}_A avec $\mathbb{P}_A(\omega_2) = \mathbb{P}_A(\omega_4) = \mathbb{P}_A(\omega_6) = \frac{1}{3}$. Il vient ainsi $\mathbb{P}_A(A \cap B) = \frac{\#A \cap B}{\#A} = \frac{2}{3} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$. Conséquemment, on a $\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$.

2.4.1.2 Définition

Soit A un évènement de probabilité $\mathbb{P}(A)$ non nulle. On définit l'application $\mathbb{P}(\cdot | A)$ qui, à tout évènement B fait correspondre le nombre

$$\mathbb{P}(B|A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

On commence par établir le résultat suivant :

Théorème 2.4.3. $\mathbb{P}(\cdot | A)$ est une probabilité sur Ω .

Démonstration. On vérifie les trois axiomes introduits dans la Définition 2.3.1.

Axiome 1. Pour tout évènement B , on a $\mathbb{P}(B|A) \in [0; 1]$.

En effet, \mathbb{P} est une probabilité donc $\mathbb{P}(A \cap B) \geq 0$. Or, $\mathbb{P}(A) > 0$ par hypothèse. Alors, on a $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \geq 0$. Puis, comme $A \cap B \subset A$, on a $\mathbb{P}(A \cap B) \leq \mathbb{P}(A)$ d'après le Théorème 2.3.9. Ainsi, il vient $0 \leq \mathbb{P}(B | A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \leq 1$.

Axiome 2. $\mathbb{P}(\Omega | A) = 1$.

En effet, on a

$$\mathbb{P}(\Omega | A) = \frac{\mathbb{P}(\Omega \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1.$$

Axiome 3. Soit $(B_k)_k$ une suite d'évènements deux à deux incompatibles. Alors $\mathbb{P}(\bigcup_{k=1}^{\infty} B_k | A) = \sum_{k=1}^{\infty} \mathbb{P}(B_k | A)$.

En effet, on a

$$A \cap \left(\bigcup_{k=1}^{\infty} B_k \right) = \bigcup_{k=1}^{\infty} (A \cap B_k),$$

par distributivité de l'intersection par rapport à la réunion. Puis, comme les évènements $A \cap B_k$ sont deux à deux incompatibles, l'axiome des probabilités totales vérifié par \mathbb{P} implique

$$\mathbb{P} \left(A \cap \left(\bigcup_{k=1}^{\infty} B_k \right) \right) = \sum_{k=1}^{\infty} \mathbb{P}(A \cap B_k).$$

Par conséquent, on a

$$\mathbb{P} \left(\bigcup_{k=1}^{\infty} B_k | A \right) = \frac{\sum_{k=1}^{\infty} \mathbb{P}(A \cap B_k)}{\mathbb{P}(A)} = \sum_{k=1}^{\infty} \mathbb{P}(B_k | A),$$

ce qui achève la preuve. □

Remarque 2.4.4. Tous les résultats portant sur les probabilités peuvent donc s'appliquer à l'application $B \mapsto \mathbb{P}(B | A)$. Notamment :

- la probabilité conditionnelle par rapport à A de l'évènement contraire de B est 1 moins la probabilité conditionnelle de l'évènement B ,
- la probabilité conditionnelle par rapport à A de l'ensemble vide est 0,
- la probabilité conditionnelle par rapport à A de B est inférieure à celle de C si $B \subset C$,
- le théorème des probabilités totales est encore valide.

Exercice 2.4.5. Vérifiez qu'en toute généralité, si A et B sont deux évènements associés à une même expérience aléatoire, alors $\mathbb{P}[A | B^c] \neq 1 - \mathbb{P}[A | B]$.

Remarque 2.4.6 (ATTENTION). On définit une probabilité conditionnelle mais en aucun cas on ne définit des évènements conditionnels. En effet, on ne considère que des éléments de la tribu. Ainsi, la probabilité conditionnelle de B sachant A est plutôt la probabilité conditionnelle, sachant A , de B .

Théorème 2.4.7 (Théorème des probabilités composées). *Soient deux évènements A et B de probabilités non nulles. Alors, on a*

$$\mathbb{P}(A \cap B) = \mathbb{P}(B | A) \mathbb{P}(A) = \mathbb{P}(A | B) \mathbb{P}(B).$$

Démonstration. Par définition,

$$\mathbb{P}(B | A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)},$$

d'où $\mathbb{P}(A \cap B) = \mathbb{P}(B | A) \mathbb{P}(A)$. On a de même $\mathbb{P}(A \cap B) = \mathbb{P}(A | B) \mathbb{P}(B)$. \square

2.4.1.3 Théorème de Bayes

Voyons maintenant le théorème de Bayes, lequel porte aussi le nom de théorème de la probabilité des causes.

Soit (A_1, \dots, A_n) un système complet d'évènements de Ω . Soit B un évènement de Ω de probabilité non nulle. On suppose que l'on connaît $\mathbb{P}(A_1), \dots, \mathbb{P}(A_n)$ et que ces n probabilités sont strictement positives. On suppose que l'on connaît également $\mathbb{P}(B | A_1), \dots, \mathbb{P}(B | A_n)$. On cherche à calculer $\mathbb{P}(A_k | B)$ pour tout $1 \leq k \leq n$.

On utilise alors le théorème des probabilités composées (Théorème 2.4.7) :

$$\mathbb{P}(A_k | B) = \frac{\mathbb{P}(B \cap A_k)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B | A_k) \mathbb{P}(A_k)}{\mathbb{P}(B)}.$$

Or, (A_1, \dots, A_n) est un système complet d'évènements de Ω . Ainsi, d'après le Théorème 2.3.7 (voir page 29), il vient

$$\mathbb{P}(B) = \sum_{k=1}^n \mathbb{P}(B \cap A_k).$$

On utilise à nouveau le théorème des probabilités composées (Théorème 2.4.7) et l'on a alors

$$\mathbb{P}(A_k | B) = \frac{\mathbb{P}(B | A_k) \mathbb{P}(A_k)}{\mathbb{P}(B | A_1) \mathbb{P}(A_1) + \dots + \mathbb{P}(B | A_n) \mathbb{P}(A_n)}.$$

Théorème 2.4.8 (Théorème de Bayes). *Soit (A_1, \dots, A_n) un système complet d'évènements de Ω . Soit B un évènement de Ω de probabilité non nulle. Alors :*

$$\mathbb{P}(A_i | B) = \frac{\mathbb{P}(B | A_i) \mathbb{P}(A_i)}{\sum_{j=1}^n \mathbb{P}(B | A_j) \mathbb{P}(A_j)}.$$

On a notamment le cas particulier avec (A, A^c) si $\mathbb{P}(A) \in]0; 1[$:

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(B | A) \mathbb{P}(A)}{\mathbb{P}(B | A) \mathbb{P}(A) + \mathbb{P}(B | A^c) \mathbb{P}(A^c)}.$$

Exercice 2.4.9. Une machine fabrique des composants électroniques. Une proportion $R = 5\%$ de ces composants est hors des tolérances. L'entreprise peut repérer les composants défectueux par un test. On suppose que si le composant est hors des tolérances alors le test est positif dans 95% des cas (sensibilité du test). Si le composant n'est pas défectueux, on suppose que le test est négatif dans 90% des cas (spécificité du test).

On suppose que tous les composants sont testés et que les composants testés positifs sont jetés. Quelle proportion de composants est jetée ? Parmi les composants jetés, quelle est la proportion de composants effectivement hors des tolérances ?

On suppose maintenant que l'on fait deux tests (indépendamment) et l'on jette les composants testés positifs deux fois. Quelle proportion de composants est jetée ? Parmi les composants jetés, quelle est la proportion de composants effectivement hors des tolérances ?

Réponses : 14.25% des composants sont jetés si l'on procède à un unique test et seulement un tiers de ces composants est effectivement hors des tolérances. En procédant à deux tests indépendants, on ne jette que 5.4625% des composants et environ 82.6% des composants jetés sont effectivement hors des tolérances.

2.4.2 Indépendance d'évènements

Soient A et B deux évènements associés à une même expérience aléatoire e à laquelle on associe l'univers Ω . Rappelons l'approche intuitive des probabilités par les fréquences. La fréquence de réalisation de l'évènement A lorsque l'on répète l'expérience e converge vers $\mathbb{P}(A)$, qui quantifie "les chances de voir A réalisé". Cette approche est validée par les lois des grands nombres, voir la Section 15.3 à la page 214.

Supposons maintenant qu'on sache que l'évènement B est réalisé. Les chances de voir A réalisé vont *a priori* changer et être quantifiées par le nombre $\mathbb{P}(A | B)$. Toutefois, il se peut que le fait de savoir que B est réalisé ne donne aucune information quant à la réalisation de A . On dit alors que A est indépendant de B . Dans cette partie, nous allons aborder rigoureusement la question de l'indépendance.

2.4.2.1 Indépendance de deux évènements

Soient deux évènements A et B de probabilités non nulles associés à une même expérience aléatoire e à laquelle on associe l'univers Ω .

Définition 2.4.10. On dit que A est indépendant de B si l'on a $\mathbb{P}(A | B) = \mathbb{P}(A)$.

Exemple 2.4.11. On jette deux dés à six faces. L'un est rouge et l'autre est bleu. On considère les évènements $A :=$ "le dé rouge indique une face paire" et $B :=$ "le dé bleu indique une face paire". Alors A est indépendant de B : $\mathbb{P}(A | B) = \mathbb{P}(A) = \frac{1}{2}$. De même, B est indépendant de A .

Contre-exemple 2.4.12. On tire deux cartes sans remise dans un jeu de trente-deux cartes. On considère les événements $A :=$ “la première carte est l’as de cœur” et $B :=$ “la deuxième carte est l’as de cœur”. Alors A n’est pas indépendant de B : $\mathbb{P}(A|B) = 0$ tandis que $\mathbb{P}(A) = \frac{1}{32}$. De même, B n’est pas indépendant de A .

Théorème 2.4.13. La relation d’indépendance est symétrique : si A est indépendant de B alors B est indépendant de A . De plus, A et B sont indépendants si et seulement si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Démonstration. A est indépendant de B si et seulement si $\mathbb{P}(A|B) = \mathbb{P}(A)$. Or, $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. On en déduit immédiatement que l’indépendance de A par rapport à B équivaut à

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Cette égalité étant symétrique, l’indépendance de A par rapport à B implique celle de B par rapport à A . \square

Remarque 2.4.14 (ATTENTION). Deux événements incompatibles et de probabilités non nulles ne sont pas indépendants. L’indépendance et l’incompatibilité sont deux notions radicalement différentes.

Proposition 2.4.15. On a équivalence entre les quatre assertions suivantes :

- A et B sont indépendants,
- A et B^c sont indépendants,
- A^c et B sont indépendants,
- A^c et B^c sont indépendants.

Démonstration. D’après la symétrie obtenue dans le Théorème 2.4.13, il suffit de montrer que la première proposition implique la deuxième. On suppose que A et B sont indépendants. On a alors

$$\mathbb{P}(B^c | A) = 1 - \mathbb{P}(B | A) = 1 - \mathbb{P}(B) = \mathbb{P}(B^c)$$

ce qui achève la preuve. \square

2.4.2.2 Indépendance mutuelle de n événements

On se propose de généraliser la notion d’indépendance à plus de deux événements ; dans un sens naturel. Examinons d’abord la situation suivante.

Exemple 2.4.16. Une urne contient quatre jetons : un bleu, un blanc, un rouge et un bleu-blanc-rouge. On en tire un au hasard. Considérons les trois événements

- $A :=$ {le jeton tiré contient du bleu},
- $B :=$ {le jeton tiré contient du blanc},
- $C :=$ {le jeton tiré contient du rouge}.

Il est clair que $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{2}{4} = \frac{1}{2}$. D'autre part, on a

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{\text{on tire le jeton tricolore}\}) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B).$$

Et, de même $\mathbb{P}(B \cap C) = \frac{1}{4} = \mathbb{P}(B)\mathbb{P}(C)$ et $\mathbb{P}(A \cap C) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(C)$. Ainsi, les événements A , B et C sont deux à deux indépendants.

Cependant, $\mathbb{P}(A | B \cap C) = 1$ car $B \cap C = \{\text{on tire le jeton tricolore}\}$. Donc la connaissance simultanée de B et C modifie notre information sur A .

La notion d'indépendance deux à deux n'est donc pas suffisante pour traduire l'idée intuitive d'indépendance de plusieurs événements. Ceci motive la définition suivante.

Définition 2.4.17. Soient n événements A_1, \dots, A_n associés à une même expérience aléatoire e à laquelle on associe l'univers Ω ; c'est-à-dire que A_1, \dots, A_n sont des sous-ensembles de Ω . On dit que les événements A_1, \dots, A_n sont mutuellement indépendants si pour toute sous-famille $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ avec $1 \leq i_1 < \dots < i_k \leq n$ où $k \in \llbracket 1; n \rrbracket$, on a :

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \times \dots \times \mathbb{P}(A_{i_k}).$$

L'indépendance mutuelle implique évidemment l'indépendance deux à deux (en considérant toutes les sous-familles de deux événements) et la réciproque est fautive comme le montre l'Exemple 2.4.16.

Remarque 2.4.18. Il convient de noter que l'égalité

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i)$$

n'est pas suffisante pour s'assurer que les événements sont mutuellement indépendants.

Proposition 2.4.19. Soient n événements A_1, \dots, A_n associés à une même expérience aléatoire e à laquelle on associe l'univers Ω . On suppose que les événements A_1, \dots, A_n sont mutuellement indépendants. Alors toute famille obtenue en remplaçant certains des A_i par leurs complémentaires est encore mutuellement indépendante.

Si les événements sont mutuellement indépendants, on peut calculer facilement la probabilité de leur réunion.

Proposition 2.4.20. Soient n événements A_1, \dots, A_n associés à une même expérience aléatoire e à laquelle on associe l'univers Ω . On suppose que les événements A_1, \dots, A_n sont mutuellement indépendants. Alors, on a

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = 1 - (1 - \mathbb{P}(A_1)) \times \dots \times (1 - \mathbb{P}(A_n)).$$

Démonstration. D'après le Théorème 2.3.5 (voir page 28), on a

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = 1 - \mathbb{P}(\overline{A_1 \cup \dots \cup A_n}) .$$

On applique ensuite les lois de Morgan (Théorème 1.3.33, voir page 17) et plus précisément (1.2) et l'on obtient

$$\mathbb{P}(\overline{A_1 \cup \dots \cup A_n}) = \mathbb{P}(\overline{A_1} \cap \dots \cap \overline{A_n}) .$$

Or, d'après la Proposition 2.4.19, on sait que les événements $\overline{A_1}, \dots, \overline{A_n}$ sont mutuellement indépendants. Il vient

$$\mathbb{P}(\overline{A_1} \cap \dots \cap \overline{A_n}) = \mathbb{P}(\overline{A_1}) \times \dots \times \mathbb{P}(\overline{A_n}) ,$$

d'où l'on obtient

$$\mathbb{P}(\overline{A_1} \cap \dots \cap \overline{A_n}) = (1 - \mathbb{P}(A_1)) \times \dots \times (1 - \mathbb{P}(A_n)) ,$$

en appliquant le Théorème 2.3.5 (voir page 28). Ceci achève la preuve. \square

2.4.3 Lemmes de Borel-Cantelli

Dans cette section, on donne trois résultats qui resserviront par la suite. Le premier est la loi du 0 – 1. Puis, l'on présente les deux lemmes de Borel-Cantelli.

Faisons d'abord quelques remarques. D'abord, si $A_n \subset A_{n+1}$ pour tout n , on dit que la suite $(A_n)_n$ est croissante et si $A_n \supset A_{n+1}$ pour tout n , on dit que la suite $(A_n)_n$ est décroissante.

Remarque 2.4.21. *Une suite croissante d'évènements $(A_n)_n$ est convergente de limite $\lim_{n \rightarrow \infty} A_n = \bigcup_{n \in \mathbb{N}} A_n$.*

On a de même avec une suite décroissante :

Remarque 2.4.22. *Une suite décroissante d'évènements $(A_n)_n$ est convergente de limite $\lim_{n \rightarrow \infty} A_n = \bigcap_{n \in \mathbb{N}} A_n$.*

On introduit maintenant la limite supérieure d'une suite d'évènements.

Définition 2.4.23. *Soit une suite quelconque d'évènements $(A_n)_n$. La suite d'évènements définie par*

$$B_n := \bigcup_{k \geq n} A_k$$

est décroissante. Sa limite est notée

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n \in \mathbb{N}} \left(\bigcup_{k \geq n} A_k \right) .$$

Plus intuitivement, $\omega \in \limsup_{n \rightarrow \infty} A_n$ si ω appartient à une infinité d'évènements A_n . En d'autres termes,

$$\omega \in \limsup_{n \rightarrow \infty} A_n \Leftrightarrow \forall p \in \mathbb{N} \quad \exists n(\omega, p) \geq p \text{ tel que } \omega \in A_{n(\omega, p)}.$$

On introduit maintenant la limite inférieure d'une suite d'évènements.

Définition 2.4.24. Soit une suite quelconque d'évènements $(A_n)_n$. La suite d'évènements définie par

$$B_n := \bigcap_{k \geq n} A_k$$

est croissante. Sa limite est notée

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n \in \mathbb{N}} \left(\bigcap_{k \geq n} A_k \right).$$

Plus intuitivement, $\omega \in \liminf_{n \rightarrow \infty} A_n$ s'il existe $p(\omega)$ tel que ω appartient à tous les évènements A_n pour $n \geq p(\omega)$. En d'autres termes,

$$\omega \in \liminf_{n \rightarrow \infty} A_n \Leftrightarrow \exists p(\omega) \in \mathbb{N} \text{ tel que } \forall n \geq p(\omega) \quad \omega \in A_n.$$

On présente ici la loi du 0 – 1, aussi appelée loi du tout ou rien. Le théorème est dû à Kolmogorov.

Théorème 2.4.25 (Loi du tout ou rien). Si les évènements A_n sont mutuellement indépendants, alors la probabilité de $A := \limsup_{n \rightarrow \infty} A_n$ vaut 0 ou 1.

Le théorème est admis.

L'idée du lemme suivant est de fournir une condition suffisante pour que la probabilité de la limite supérieure d'une suite d'évènements soit égale à 0, sans toutefois supposer l'indépendance.

Lemme 2.4.26 (Lemme de Borel-Cantelli 1). Si $(A_n)_n$ est une suite d'évènements telle que $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < \infty$ alors $\mathbb{P}(\limsup A_n) = 0$.

Démonstration. Par définition de la limite supérieure,

$$\mathbb{P}(\limsup A_n) = \mathbb{P} \left(\bigcap_{n \in \mathbb{N}} B_n \right),$$

où $B_n := \bigcup_{k \geq n} A_k$. Or, pour tout $p \in \mathbb{N}$, $\bigcap_{n \in \mathbb{N}} B_n \subset B_p$ donc par croissance de la probabilité, voir le Théorème 2.3.9, on obtient :

$$\mathbb{P}(\limsup A_n) \leq \mathbb{P}(B_p),$$

pour tout $p \in \mathbb{N}$. Or, on pourrait montrer par récurrence :

$$\mathbb{P} \left(\bigcup_{k \geq p} A_k \right) \leq \sum_{k=p}^{\infty} \mathbb{P}(A_k).$$

Cependant, la somme $\sum_{k \in \mathbb{N}} \mathbb{P}(A_k)$ est finie donc la suite des restes tend vers 0. Il s'ensuit : $\lim_{p \rightarrow \infty} \mathbb{P}(B_p) = 0$. Ainsi, $\mathbb{P}(\limsup A_n) = 0$.

□

Lemme 2.4.27 (Lemme de Borel-Cantelli 2). *Soit $(A_n)_n$ une suite d'évènements indépendants. On suppose : $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = +\infty$. Alors :*

$$\mathbb{P}(\limsup A_n) = 1.$$

Démonstration. Pour ce faire, on va prouver que le complémentaire de la limite supérieure est de probabilité égale à 0. Par définition :

$$\begin{aligned} \mathbb{P}((\limsup A_n)^c) &= \mathbb{P} \left(\left(\bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k \right)^c \right) \\ &= \mathbb{P} \left(\bigcup_{n \in \mathbb{N}} \bigcap_{k \geq n} A_k^c \right) \\ &= \mathbb{P}(\liminf A_n^c) \\ &= \lim_{n \rightarrow \infty} \mathbb{P} \left(\bigcap_{k \geq n} A_k^c \right) \end{aligned}$$

en utilisant le théorème de convergence monotone (que l'on admet) et la croissance de la suite d'évènements $(\bigcap_{k \geq n} A_k^c)_n$.

Or, à n fixé, on a

$$\mathbb{P} \left(\bigcap_{k \geq n} A_k^c \right) = \lim_{p \rightarrow \infty} \mathbb{P} \left(\bigcap_{p \geq k \geq n} A_k^c \right)$$

car la suite d'évènements $(\bigcap_{p \geq k \geq n} A_k^c)_p$ est décroissante.

Or, les évènements A_k sont indépendants donc il en est de même pour les complémentaires. Il vient :

$$\mathbb{P} \left(\bigcap_{p \geq k \geq n} A_k^c \right) = \prod_{k=n}^p \mathbb{P}(A_k^c) = \prod_{k=n}^p (1 - \mathbb{P}(A_k)).$$

Ensuite :

$$-\log \left[\mathbb{P} \left(\bigcap_{p \geq k \geq n} A_k^c \right) \right] = -\sum_{k=n}^p \log(1 - \mathbb{P}(A_k)).$$

Comme $-\log(1-x) \geq x$ pour tout $x \in [0; 1]$, il s'ensuit

$$-\log \left[\mathbb{P} \left(\bigcap_{p \geq k \geq n} A_k^c \right) \right] \geq \sum_{k=n}^p \mathbb{P}(A_k) \longrightarrow +\infty.$$

Conséquemment, $\mathbb{P} \left(\bigcap_{p \geq k \geq n} A_k^c \right)$ tend vers 0 quand p tend vers l'infini d'où

$$\mathbb{P} \left(\bigcap_{k \geq n} A_k^c \right) = 0.$$

La limite de cette quantité est 0 quand n tend vers l'infini. On a ainsi terminé la preuve du lemme. □

Exemple 2.4.28. *Un singe tape au hasard sur les touches d'un clavier de manière régulière produisant ainsi une suite de caractères $(x_n)_n$. On admet que le clavier contient un nombre fini de caractères dont chacun a une probabilité strictement positive d'être tiré. On admet également que pour tout $N \in \mathbb{N}$, pour tous les entiers i_1, \dots, i_N , et pour tous les caractères c_1, \dots, c_N , on a*

$$\mathbb{P}(x_{i_1} = c_{i_1}, \dots, x_{i_N} = c_{i_N}) = \prod_{k=1}^N \mathbb{P}(x_{i_k} = c_{i_k}).$$

Soit l'évènement $H :=$ "le singe produit une infinité de fois le scénario de Dark". Alors, $\mathbb{P}(H) = 1$.

On appelle ϵ le minimum, sur l'ensemble des caractères du clavier, des probabilités d'être tiré. Pour tout $i \in \mathbb{N}$, on note A_i l'évènement

$$A_i := \{\text{le singe produit le scénario de Dark durant l'intervalle } [iM + 1; iM + M]\}$$

où M est le nombre de caractères dans le scénario de Dark. On a clairement

$$\limsup_{i \in \mathbb{N}} A_i \subset H.$$

Il est donc suffisant de prouver $\mathbb{P}(\limsup A_i) = 1$. Or, les évènements $(A_i)_i$ sont indépendants et de plus, par indépendance mutuelle des évènements $\{x_k = c_k\}$, on a

$$\mathbb{P}(A_i) \geq \epsilon^M > 0.$$

Conséquemment, $\sum_{i=0}^{\infty} \mathbb{P}(A_i) = +\infty$. On applique le lemme de Borel-Cantelli-2 pour conclure.

Généralités sur les variables aléatoires

3.1 Introduction

Si l'on ne s'intéressait qu'à des événements, les probabilités seraient d'un intérêt réduit. De fait, on regarde ce que l'on appelle des variables aléatoires.

Les variables aléatoires sont à la source de très nombreuses applications rien que par le mouvement Brownien (qui est une collection de variables aléatoires vérifiant certaines propriétés) : finance, biologie, batteries de lithium, physique des plasmas mais aussi apprentissage profond par réseaux de neurones... Les variables aléatoires réelles sont l'objet de ce chapitre.

Une bonne connaissance de la notion de fonctions et d'applications est nécessaire à la compréhension de ce chapitre. Notons aussi que le lecteur familier du concept de distributions est légèrement avantagé. Néanmoins, il n'est pas rédhibitoire de ne pas savoir ce que représente cet objet. Pour finir, il sera important que le lecteur ait déjà été familiarisé avec la notion de fonction continue ainsi que celle d'intervalles.

Les principaux objectifs que nous souhaitons atteindre à l'issue de ce chapitre sont la définition d'une variable aléatoire (modulo un *à peu près* sur la rigueur vu que l'on n'utilise pas la notion de tribu), bien comprendre que l'on omet parfois l'aléa dans le modèle (préférant ainsi noter X plutôt que $X(\omega)$, savoir que l'ensemble des variables aléatoires réelles satisfait de bonnes propriétés de stabilité par la somme, le produit et de manière générale toute transformation un tant soit peu raisonnable. Certains exemples de variables aléatoires réelles seront donnés : ces exemples sont d'une importance capitale en vue de la compréhension de la notion. Nous verrons aussi que les variables en question sont de deux types : discrètes ou continues. Enfin, nous aborderons une notion compliquée mais cruciale : la loi d'une variable aléatoire réelle (laquelle n'est rien d'autre qu'une probabilité représentant la loi de variabilité de la variable sur l'ensemble des réalisations possibles). Cette loi peut d'ailleurs se caractériser simplement, en dimension un, en utilisant un outil des plus précieux que nous découvrirons au fil du chapitre : la fonction de répartition. Nous terminerons ce chapitre par l'indépendance entre les variables aléatoires réelles.

3.2 Définition et Opérations

On commence par considérer un exemple introductif.

Exemple 3.2.1. Soit une population de N composants électroniques numérotés de 1 à N . Soit z_i l'impédance de l'individu numéro i pour tout $i \in \llbracket 1; N \rrbracket$.

On considère l'expérience aléatoire e : tirer au hasard un individu de la population. On lui associe l'univers Ω . Les résultats possibles sont les $\omega_i :=$ "on obtient l'individu numéro i " pour $i \in \llbracket 1; N \rrbracket$. L'espace fondamental est alors $\Omega = \{\omega_1, \dots, \omega_N\}$. La probabilité que l'on définit sur Ω est l'équiprobabilité car le tirage est au hasard : $\mathbb{P}(\omega_1) = \dots = \mathbb{P}(\omega_N) = \frac{1}{N}$.

On considère l'application X de Ω dans \mathbb{R} avec $X(\omega_i) := z_i$. Si l'on tire l'individu numéro i , X prend la valeur z_i , c'est-à-dire l'impédance de l'individu tiré.

On dit que X a pour réalisation l'impédance de l'individu tiré. L'application X est appelée une variable aléatoire réelle.

On peut observer sur cet exemple que les réalisations possibles de la variable aléatoire X dépendent des résultats de l'expérience aléatoire sur laquelle l'espace fondamental est défini. De manière générale, une variable aléatoire réelle est une fonction de Ω dans \mathbb{R} telle que les réalisations de la fonction sont entièrement déterminées par les résultats de l'expérience aléatoire.

3.2.1 Définition

On appelle variable aléatoire réelle définie sur un espace fondamental Ω toute application de Ω dans \mathbb{R} :

$$\begin{aligned} X &: \Omega \rightarrow \mathbb{R} \\ \omega &\mapsto X(\omega). \end{aligned}$$

$X(\omega)$ est une réalisation possible de X .

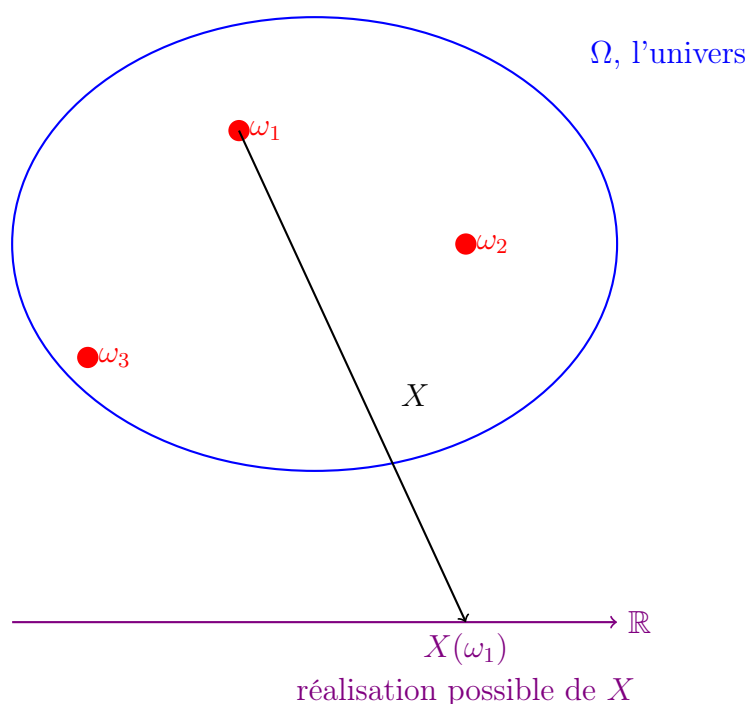
Remarque 3.2.2. Cette définition n'est pas tout à fait exacte. Une variable aléatoire réelle est en fait une application mesurable de Ω dans \mathbb{R} . En d'autres termes, l'image réciproque d'un borélien (élément de la tribu borélienne sur \mathbb{R}) est un élément de la tribu sur Ω .

On peut cependant noter que toute application de Ω dans \mathbb{R} peut être rendue mesurable si l'on munit Ω d'une tribu adéquate (la plus petite tribu rendant l'application X mesurable, $\sigma(X)$). Ceci nous permet de nous affranchir de la construction rigoureuse. Plus de détails se trouvent au Chapitre 24 à la page 341.

L'ensemble de toutes les réalisations possibles de X , $\{X(\omega) : \omega \in \Omega\}$, est noté $X(\Omega)$.

Illustrons avec le schéma suivant :

FIGURE 3.1 – Variable aléatoire réelle



Remarque 3.2.3. On peut classer les variables aléatoires (et subséquentement les lois de probabilité associées) en fonction de $X(\Omega)$. En fait, l'ensemble des réalisations possibles joue un grand rôle alors que l'espace fondamental lui-même a un rôle mineur par la suite ; ce qui est normal car nous n'y avons quasiment jamais accès.

Définition 3.2.4. Lorsque l'ensemble des réalisations possibles de la variable aléatoire réelle X est fini ou infini dénombrable, on dit que la variable aléatoire réelle X est discrète. Sinon, on dit que la variable aléatoire réelle X est continue.

Présentons quelques exemples.

Exemple 3.2.5 (Variable aléatoire réelle avec $X(\Omega)$ fini). Soit un évènement A associé à l'expérience aléatoire e (c'est-à-dire : $A \subset \Omega$ où Ω est l'univers associé). On définit la variable aléatoire $\mathbb{1}_A$ de la façon suivante :

$$\mathbb{1}_A : \Omega \rightarrow \{0, 1\}$$

$$\omega \mapsto \mathbb{1}_A(\omega) := \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases} .$$

Ici, $\mathbb{1}_A(\Omega) = \{0, 1\}$ est un ensemble fini donc la variable aléatoire réelle $\mathbb{1}_A$ est discrète.

Il convient de noter que dans l'exemple précédent, Ω n'est pas obligatoirement fini ou infini dénombrable.

Exemple 3.2.6 (Variable aléatoire réelle avec $X(\Omega)$ infini dénombrable). *On se donne l'expérience aléatoire suivante. On observe le nombre d'octets échangés sur un système de pair à pair durant un intervalle de temps de durée fixée. Les résultats possibles sont les $\omega_n :=$ "n octets sont échangés". L'espace fondamental est alors $\Omega = \{\omega_0, \omega_1, \dots, \omega_n, \dots\}$. On se donne la variable aléatoire réelle X définie par*

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{N} \\ \omega_n &\mapsto X(\omega_n) := n. \end{aligned}$$

L'ensemble des réalisations possibles est ainsi $X(\Omega) = \mathbb{N}$. La variable aléatoire réelle X est donc discrète.

Exemple 3.2.7 (Variable aléatoire réelle avec $X(\Omega)$ infini non dénombrable). *On se donne l'expérience aléatoire suivante : on observe la durée de vie d'un composant électronique. Les résultats possibles sont les $\omega_t :=$ "la durée de vie du composant est t", avec $t \in \mathbb{R}_+$. L'espace fondamental est alors $\Omega = \{\omega_t : t \geq 0\}$. On se donne la variable aléatoire réelle X définie par*

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R}_+ \\ \omega_t &\mapsto X(\omega_t) := t. \end{aligned}$$

L'ensemble des réalisations possibles est ainsi $X(\Omega) = \mathbb{R}_+$. La variable aléatoire réelle X est donc continue.

3.2.2 Opérations sur les variables aléatoires réelles

3.2.2.1 Fonction d'une variable aléatoire réelle

Soit f une fonction réelle de la variable réelle. On se donne une variable aléatoire réelle X de Ω dans \mathbb{R} . On définit alors la variable aléatoire réelle $f(X)$ par :

$$\begin{aligned} f(X) : \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto f[X(\omega)]. \end{aligned}$$

Remarque 3.2.8. *En fait, ceci n'est vrai qu'avec des fonctions boréliennes c'est-à-dire des fonctions mesurables par rapport à la tribu borélienne.*

Par exemple, avec $f_p(x) := x^p$:

$$\begin{aligned} f_p(X) = X^p : \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto X(\omega)^p, \end{aligned}$$

pour tout $p \in \mathbb{N}^*$. Cette variable aléatoire réelle est utilisée pour calculer le moment d'ordre p de la variable aléatoire réelle X , voir les pages 65 (variables aléatoires discrètes) et 123 (variables aléatoires à densité).

3.2.2.2 Opération sur deux variables aléatoires réelles

Soient deux variables aléatoires réelles X et Y définies sur un même espace fondamental Ω . Soit h une fonction de \mathbb{R}^2 dans \mathbb{R} . On définit la variable aléatoire réelle $h(X, Y)$:

$$\begin{aligned} h(X, Y) &: \Omega \rightarrow \mathbb{R} \\ \omega &\mapsto h[X(\omega), Y(\omega)] . \end{aligned}$$

Comme écrit précédemment, la fonction $h(X, Y)$ est une variable aléatoire réelle. On en déduit immédiatement que la somme de deux variables aléatoires réelles est une variable aléatoire réelle et de même avec le produit. Il suffit pour s'en convaincre de considérer les fonctions $h_1(x, y) := x + y$ et $h_2(x, y) := x \times y$.

Remarque 3.2.9. *En fait, ceci n'est vrai que si h est mesurable de $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ où $\mathcal{B}(\mathbb{R})$ est la tribu borélienne sur \mathbb{R} et $\mathcal{B}(\mathbb{R}^2)$ est celle sur \mathbb{R}^2 .*

3.3 Loi d'une variable aléatoire réelle

Avant de définir ce qu'est la loi de probabilité d'une variable aléatoire réelle, on regarde un exemple introductif.

Exemple 3.3.1. *Soit l'espace fondamental $\Omega := \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}$ muni de la probabilité \mathbb{P} :*

TABLE 3.1 – Définition de la probabilité \mathbb{P} sur l'univers Ω

ω	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8
$\mathbb{P}(\{\omega\})$	0.1	0.05	0.1	0.1	0.15	0.1	0.15	0.25

On note notamment que \mathbb{P} n'est pas l'équiprobabilité. On se donne maintenant la variable aléatoire réelle X de Ω dans \mathbb{R} définie par

TABLE 3.2 – Réalisations de la variable aléatoire réelle X

ω	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8
$X(\omega)$	2	4	1	2	5	4	5	3

L'ensemble des réalisations possibles de X est alors $X(\Omega) = \{1, 2, 3, 4, 5\}$. Calculons maintenant la probabilité que X ait la réalisation 4, notée $\mathbb{P}(X = 4)$,

c'est-à-dire la probabilité d'obtenir un résultat ω tel que $X(\omega) = 4$. Par une simple observation, on a

$$\{X = 4\} = \{\omega : X(\omega) = 4\} = \{\omega_2, \omega_6\} .$$

Conséquemment, la probabilité est égale à

$$\mathbb{P}(X = 4) = \mathbb{P}(\{\omega_2, \omega_6\}) = \mathbb{P}(\{\omega_2\}) + \mathbb{P}(\{\omega_6\}) = 0.05 + 0.1 = 0.15 .$$

De manière plus générale :

TABLE 3.3 – Loi de probabilité de la variable aléatoire réelle X

k	1	2	3	4	5
$\mathbb{P}(X = k)$	0.1	0.2	0.25	0.15	0.3

En posant $\mathbb{P}_X(k) := \mathbb{P}(X = k)$, on remarque :

$$\mathbb{P}_X(1) + \mathbb{P}_X(2) + \mathbb{P}_X(3) + \mathbb{P}_X(4) + \mathbb{P}_X(5) = 1 .$$

Ainsi, la fonction \mathbb{P}_X de $X(\Omega) = \{1, 2, 3, 4, 5\}$ dans $[0; 1]$ peut être étendue en une application additive sur $2^{X(\Omega)}$ qui vérifie $\mathbb{P}_X(X(\Omega)) = 1$, c'est-à-dire en une probabilité sur $X(\Omega)$. On continue de la noter \mathbb{P}_X . Cette probabilité est appelée la loi de probabilité de X . Notons qu'on peut la caractériser simplement à l'aide d'un tableau car l'ensemble des réalisations possibles est fini.

On peut aussi utiliser la notation $\mathbb{P}X^{-1}$ car

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbb{P}(X^{-1}(A)) .$$

Avec une écriture de type distribution, on a

$$\mathbb{P}_X = 0.1\delta_1 + 0.2\delta_2 + 0.25\delta_3 + 0.15\delta_4 + 0.3\delta_5 .$$

La notation avec les distributions de Dirac signifie ici qu'il y a une masse (une probabilité) 0.1 en 1, une masse de 0.2 en 2, une masse de 0.25 en 3, une masse de 0.15 en 4 et une masse de 0.3 en 5.

3.3.1 Définition

De façon générale, soit X une variable aléatoire réelle de Ω dans \mathbb{R} . Soit I une partie de \mathbb{R} . La probabilité que X ait une réalisation dans I , notée $\mathbb{P}(X \in I)$, est la probabilité d'obtenir un résultat $\omega \in \Omega$ tel que $X(\omega) \in I$:

$$\mathbb{P}_X(I) := \mathbb{P}(X \in I) = \mathbb{P}(\{\omega : X(\omega) \in I\}) .$$

Définition 3.3.2. *La loi de probabilité de la variable aléatoire réelle X est l'application \mathbb{P}_X aussi notée $\mathbb{P}X^{-1}$ qui, à toute partie I de \mathbb{R} , fait correspondre la probabilité que X ait une réalisation dans I , $\mathbb{P}(\{\omega : X(\omega) \in I\})$ notée $\mathbb{P}(X \in I)$.*

C'est l'image réciproque par l'application X de la probabilité \mathbb{P} .

On parle aussi de mesure de probabilité.

Remarque 3.3.3. *En fait, on ne regarde pas la probabilité que X ait une réalisation dans n'importe quelle partie de \mathbb{R} . On ne s'intéresse qu'aux parties "raisonnables" de \mathbb{R} . Il s'agit des ensembles boréliens (qui appartiennent à la tribu borélienne). Typiquement, il faut se représenter ces ensembles comme des réunions dénombrables d'intervalles ouverts et d'intervalles fermés. On choisit de ne pas entrer dans ces détails dans le présent ouvrage. Par ailleurs, il n'est pas simple de construire une partie non raisonnable puisqu'il faut faire appel à l'axiome du choix. Ainsi, dans la pratique professionnelle, supposer que toutes les parties de \mathbb{R} sont raisonnables n'est pas un trop gros sacrifice quant à l'utilisation mathématique dans le monde réel. Pour plus de rigueur, nous invitons le lecteur à consulter le Chapitre 24 à la page 341.*

3.3.2 Fonction de répartition

La fonction de répartition est un autre moyen de caractériser la loi de probabilité d'une variable aléatoire réelle. En effet, pour connaître la probabilité que X ait une réalisation dans une partie (borélienne) de \mathbb{R} , il suffit de connaître $\mathbb{P}(X \leq x)$ pour tout $x \in \mathbb{R}$.

Donnons une justification succincte de ce fait. On peut montrer que la tribu des boréliens (la tribu raisonnable que l'on considère) est engendrée par la classe des intervalles semi-ouverts $] -\infty; x]$ où x parcourt l'ensemble des réels. Ainsi, l'on peut caractériser la probabilité de tout borélien si l'on connaît la probabilité de tous ces intervalles semi-ouverts.

L'avantage de la fonction de répartition est qu'il s'agit d'une fonction vérifiant de bonnes propriétés (que l'on voit subséquentement).

3.3.2.1 Définition

Définition 3.3.4. *La fonction de répartition d'une variable aléatoire réelle X est la fonction réelle d'une variable réelle notée F_X (parfois F s'il n'y a pas d'ambiguïté) telle que*

$$F_X(t) := \mathbb{P}_X(]-\infty; t]) = \mathbb{P}(X \leq t) = \mathbb{P}(\{\omega : X(\omega) \leq t\}).$$

Exemple 3.3.5. *On reprend la variable aléatoire réelle X de l'Exemple 3.3.1. Il s'agit d'une variable aléatoire réelle dont l'ensemble des réalisations possibles est $\{1, 2, 3, 4, 5\}$. Et, on a :*

TABLE 3.4 – Loi de probabilité de la variable aléatoire réelle X

k	1	2	3	4	5
$\mathbb{P}(X = k)$	0.1	0.2	0.25	0.15	0.3

Calculons $F_X(t)$ pour tout $t \in \mathbb{R}$. On procède à une disjonction de cas :

1. Si $t < 1$, on a

$$\begin{aligned}
 F_X(t) &= \mathbb{P}(X \leq t) \\
 &= \mathbb{P}(X \in]-\infty; t] \cap \{1, 2, 3, 4, 5\}) \\
 &= \mathbb{P}(X \in \emptyset) \\
 &= \mathbb{P}(\emptyset) \\
 &= 0.
 \end{aligned}$$

2. Si $1 \leq t < 2$, on a

$$\begin{aligned}
 F_X(t) &= \mathbb{P}(X \leq t) \\
 &= \mathbb{P}(X \in]-\infty; t] \cap \{1, 2, 3, 4, 5\}) \\
 &= \mathbb{P}(X \in \{1\}) \\
 &= \mathbb{P}(X = 1) \\
 &= 0.1.
 \end{aligned}$$

3. Si $2 \leq t < 3$, on a

$$\begin{aligned}
 F_X(t) &= \mathbb{P}(X \leq t) \\
 &= \mathbb{P}(X \in]-\infty; t] \cap \{1, 2, 3, 4, 5\}) \\
 &= \mathbb{P}(X \in \{1, 2\}) \\
 &= \mathbb{P}(X = 1) + \mathbb{P}(X = 2) \\
 &= 0.3.
 \end{aligned}$$

4. Si $3 \leq t < 4$, on a

$$\begin{aligned}
 F_X(t) &= \mathbb{P}(X \leq t) \\
 &= \mathbb{P}(X \in]-\infty; t] \cap \{1, 2, 3, 4, 5\}) \\
 &= \mathbb{P}(X \in \{1, 2, 3\}) \\
 &= \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) \\
 &= 0.55.
 \end{aligned}$$

5. Si $4 \leq t < 5$, on a

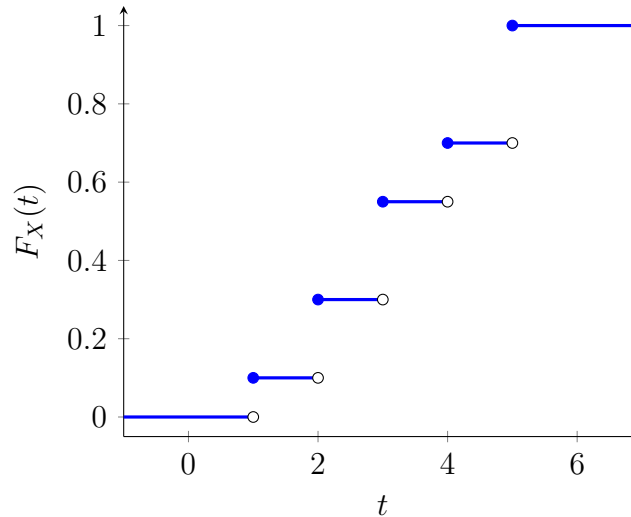
$$\begin{aligned}
 F_X(t) &= \mathbb{P}(X \leq t) \\
 &= \mathbb{P}(X \in]-\infty; t] \cap \{1, 2, 3, 4, 5\}) \\
 &= \mathbb{P}(X \in \{1, 2, 3, 4\}) \\
 &= \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \mathbb{P}(X = 4) \\
 &= 0.7.
 \end{aligned}$$

6. Si $5 \leq t$, on a

$$\begin{aligned}
 F_X(t) &= \mathbb{P}(X \leq t) \\
 &= \mathbb{P}(X \in]-\infty; t] \cap \{1, 2, 3, 4, 5\}) \\
 &= \mathbb{P}(X \in \{1, 2, 3, 4, 5\}) \\
 &= \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \mathbb{P}(X = 4) + \mathbb{P}(X = 5) \\
 &= 1.
 \end{aligned}$$

On connaît ainsi la représentation graphique de la fonction de répartition F_X :

FIGURE 3.2 – Graphe de la fonction de répartition



Remarque 3.3.6. La fonction F_X n'est pas continue dans le cas présent.

Remarque 3.3.7. La dérivée (au sens des distributions) de F_X est justement

$$0.1\delta_1 + 0.2\delta_2 + 0.25\delta_3 + 0.15\delta_4 + 0.3\delta_5,$$

d'après la formule des sauts.

3.3.2.2 Propriétés

Présentons les différentes propriétés de la fonction de répartition d'une variable aléatoire réelle X .

Proposition 3.3.8 (Bornitude). *Pour tout $t \in \mathbb{R}$, on a $0 \leq F_X(t) \leq 1$. De plus, $F_X(-\infty) := \lim_{t \rightarrow -\infty} F_X(t) = 0$ et $F_X(+\infty) := \lim_{t \rightarrow +\infty} F_X(t) = 1$.*

Démonstration. Par définition, la quantité $F_X(t)$ est égale à

$$F_X(t) := \mathbb{P}(X \leq t) ,$$

pour tout $t \in \mathbb{R}$. Or, par axiome, $\mathbb{P}(X \leq t) \in [0; 1]$. Puis, l'évènement $\{X \leq -\infty\}$ étant vide, il vient $F_X(-\infty) = 0$. De même, l'évènement $\{X \leq +\infty\}$ étant certain, il vient $F_X(+\infty) = 1$. \square

Proposition 3.3.9. *Pour tout $t \in \mathbb{R}$, on dispose de l'expression suivante de la probabilité de $]t; +\infty[$:*

$$\mathbb{P}(X > t) = 1 - F_X(t) .$$

Démonstration. En effet, l'évènement $\{X > t\}$ est le complémentaire de l'évènement $\{X \leq t\}$. On a ainsi

$$\mathbb{P}(X > t) = 1 - \mathbb{P}(X \leq t) = 1 - F_X(t) .$$

\square

Il est essentiel de faire attention à la manipulation des relations d'ordre. En effet, comme la fonction de répartition n'est pas nécessairement continue, il faut bien différencier $\mathbb{P}(X < t)$ et $\mathbb{P}(X \leq t)$.

Proposition 3.3.10. *Pour tous les réels t_1 et t_2 avec $t_2 > t_1$, l'égalité suivante est vraie :*

$$\mathbb{P}(t_1 < X \leq t_2) = F_X(t_2) - F_X(t_1) .$$

Démonstration. En effet, on peut écrire l'évènement $\{t_1 < X \leq t_2\}$ comme suit

$$\{t_1 < X \leq t_2\} = \{X \leq t_2\} \cap \{X \leq t_1\}^c .$$

Conséquemment, on a

$$\mathbb{P}(t_1 < X \leq t_2) = \mathbb{P}(X \leq t_2) - \mathbb{P}(\{X \leq t_1\} \cap \{X \leq t_2\}) .$$

Or, comme $t_1 < t_2$ on a $\{X \leq t_1\} \cap \{X \leq t_2\} = \{X \leq t_1\}$ d'où

$$\mathbb{P}(t_1 < X \leq t_2) = \mathbb{P}(X \leq t_2) - \mathbb{P}(X \leq t_1) ,$$

ce qui achève la preuve. \square

Remarque 3.3.11 (ATTENTION). *Il faut ici considérer une probabilité de la forme $\mathbb{P}(a < X \leq b)$. En effet, la formule ne serait alors pas vraie pour peu que F_X présente une discontinuité en a ou en b .*

Proposition 3.3.12. *La fonction F_X est croissante.*

Démonstration. D'après la Proposition 3.3.10, quels que soient $s < t$, on a

$$F_X(t) - F_X(s) = \mathbb{P}(s < X \leq t) \geq 0,$$

d'où $F_X(t) \geq F_X(s)$. La fonction est donc bien croissante. On peut aussi noter $\{X \leq t\} \supset \{X \leq s\}$ et le résultat est immédiat par la croissance de la probabilité, voir page 30. \square

Remarque 3.3.13. *La fonction F_X n'est pas nécessairement strictement croissante. Par exemple, la fonction F_X de l'Exemple 3.3.5 est constante par morceaux.*

Proposition 3.3.14. *La fonction F_X est càdlàg (continue à droite et limitée à gauche). En d'autres termes, pour tout $t \in \mathbb{R}$, $\lim_{s \rightarrow t^-} F_X(s)$ existe et $\lim_{s \rightarrow t^+} F_X(s) = F_X(t)$.*

La preuve de cette proposition est plus délicate donc elle n'est pas donnée. Notons toutefois que la fonction n'est pas nécessairement continue, voir l'Exemple 3.3.5.

Notation 3.3.15. *La limite à gauche de F en t , à savoir $\lim_{s \rightarrow t^-} F_X(s)$ est notée $F_X(t^-)$.*

Théorème 3.3.16. *Soit une fonction F qui vérifie les Propositions 3.3.8, 3.3.12 et 3.3.14, c'est-à-dire une fonction croissante et càdlàg de \mathbb{R} dans $[0; 1]$ telle que $F(-\infty) = 0$ et $F(+\infty) = 1$. Alors, il existe un espace fondamental Ω , muni d'une probabilité \mathbb{P} et une variable aléatoire réelle X de Ω dans \mathbb{R} telle que F est la fonction de répartition de la variable aléatoire réelle X .*

Ainsi, à l'étude des lois de probabilité des variables aléatoires réelles, on peut lui substituer l'étude des fonctions croissantes et càdlàg à valeurs dans $[0; 1]$ qui tendent vers 0 en $-\infty$ et vers 1 en $+\infty$.

3.4 Indépendance des variables aléatoires réelles

La notion d'indépendance est probablement la base des probabilités. Pour l'aborder de manière rigoureuse, il faut utiliser la théorie plus abstraite des algèbres de Boole et des tribus, voir le Chapitre 24 à la page 341. Cependant, on procède autrement pour l'instant.

3.4.1 Définition

Définition 3.4.1. Soient n variables aléatoires réelles X_1, \dots, X_n définies sur un même espace fondamental Ω muni d'une probabilité \mathbb{P} . On dit que X_1, \dots, X_n sont mutuellement indépendantes si pour toute partie S_1, \dots, S_n de \mathbb{R} , les événements $\{X_1 \in S_1\}, \dots, \{X_n \in S_n\}$ sont mutuellement indépendants (voir la Définition 2.4.17, page 40). En d'autres termes :

$$\mathbb{P}(X_1 \in S_1, \dots, X_n \in S_n) = \mathbb{P}(X_1 \in S_1) \times \dots \times \mathbb{P}(X_n \in S_n) .$$

Remarque 3.4.2. On pourrait s'étonner de ne pas considérer des sous-familles comme on l'a fait dans la Définition 2.4.17. En effet, pour se restreindre à toute sous-famille $(X_{i_1}, \dots, X_{i_k})$ avec $1 \leq i_1 \leq \dots \leq i_k \leq n$, il suffit de prendre $S_p = \mathbb{R}$ pour tout p n'appartenant pas à la sous-famille et alors $\{X_p \in S_p\} = \Omega$ et l'on retrouve donc bien

$$\mathbb{P}(X_{i_1} \in S_{i_1}, \dots, X_{i_k} \in S_{i_k}) = \mathbb{P}(X_{i_1} \in S_{i_1}) \times \dots \times \mathbb{P}(X_{i_k} \in S_{i_k}) .$$

Remarque 3.4.3. Comme déjà mentionné, cette définition porte en fait uniquement sur les parties boréliennes de \mathbb{R} .

Remarque 3.4.4. De manière équivalente, l'indépendance mutuelle de n variables aléatoires réelles signifie que les tribus engendrées par les variables aléatoires réelles sont mutuellement indépendantes (en d'autres termes que les événements des tribus sont mutuellement indépendants).

En particulier, dans le cas où $n = 2$, deux variables aléatoires réelles X_1 et X_2 sont indépendantes si

$$\mathbb{P}(X_1 \in S_1, X_2 \in S_2) = \mathbb{P}(X_1 \in S_1) \mathbb{P}(X_2 \in S_2) ,$$

pour toutes les parties S_1 et S_2 de \mathbb{R} ou de façon équivalente :

$$\mathbb{P}(X_1 \in S_1 \mid X_2 \in S_2) = \mathbb{P}(X_1 \in S_1) ,$$

si $\mathbb{P}(X_2 \in S_2) > 0$.

3.4.2 Exemples

Exemple 3.4.5. On jette deux dés à six faces, un rouge et un bleu. Soient X_1 la variable aléatoire réelle qui a pour réalisation la face obtenue du dé rouge et X_2 la variable aléatoire réelle qui a pour réalisation la face obtenue du dé bleu. On a ici $\Omega = \{(i, j) : i \in \llbracket 1; 6 \rrbracket, j \in \llbracket 1; 6 \rrbracket\}$. Et, $X_1(\Omega) = \{1, 2, 3, 4, 5, 6\}$ et $X_2(\Omega) = \{1, 2, 3, 4, 5, 6\}$. Alors les deux variables aléatoires réelles X_1 et X_2 sont indépendantes.

Exemple 3.4.6. Soit une population de N composants électroniques. Soit l'expérience aléatoire e qui consiste à tirer au hasard (avec remise) n individus de la population. Comme le tirage est avec remise, les n tirages au hasard d'un individu sont des expériences aléatoires mutuellement indépendantes.

Pour $1 \leq i \leq n$, soit X_i la variable aléatoire réelle qui a pour réalisation l'impédance du i -ème composant tiré. Alors les n variables aléatoires réelles X_1, \dots, X_n sont mutuellement indépendantes.

Remarque 3.4.7. On peut constater que dans les deux exemples que l'on a donnés, ces variables aléatoires réelles ont la même loi de probabilité. Elles constituent donc un échantillon *i.i.d.* (c'est-à-dire de variables aléatoires réelles indépendantes et identiquement distribuées). Cependant, il est très important de comprendre que l'indépendance n'implique pas le caractère *i.i.d.* des variables aléatoires. Un exemple de cette assertion : le temps qu'une diffusion d'Itô met pour sortir d'un domaine donné et le lieu par lequel cette diffusion sort sont deux variables aléatoires indépendantes. Mais les lois sont fondamentalement différentes (la première est une loi à valeurs dans \mathbb{R}_+ et l'autre a pour support la frontière dudit domaine donné).

3.4.3 Contre-exemple

On présente maintenant un contre-exemple.

Contre-exemple 3.4.8. Soit une population de N hommes adultes vivant en France. Soit l'expérience aléatoire e qui consiste à tirer au hasard un individu de la population.

Soit X la variable aléatoire réelle qui a pour réalisation la taille de l'individu tiré, en centimètres. Soit Y la variable aléatoire réelle qui a pour réalisation le poids de l'individu tiré, en kilos. On devine aisément :

$$\mathbb{P}(Y \geq 90 \mid X \leq 150) \neq \mathbb{P}(Y \geq 90 \mid X \geq 200) .$$

Ainsi, les variables aléatoires réelles X et Y ne sont pas indépendantes.

Évidemment, il est important de ne pas se contenter de l'intuition et de faire des statistiques rigoureuses pour établir la véracité de cette assertion. On pense notamment au test du Khi-deux d'indépendance, voir la Section 21.9.

3.4.4 Propriétés

Proposition 3.4.9. Soient n variables aléatoires réelles X_1, \dots, X_n définies sur un même espace fondamental Ω . On suppose que les variables aléatoires réelles sont mutuellement indépendantes. Soient n fonctions f_1, \dots, f_n . Alors, les n variables aléatoires réelles $f_1(X_1), \dots, f_n(X_n)$ sont mutuellement indépendantes.

Remarque 3.4.10. *Pour que cette proposition soit totalement vraie, il faudrait préciser que les fonctions doivent être mesurables pour la tribu borélienne. Les fonctions continues et les fonctions continues par morceaux le sont.*

Remarque 3.4.11. *Cette proposition se prouve immédiatement une fois que l'on a formalisé. En effet, les variables aléatoires réelles sont mutuellement indépendantes si les tribus qu'elles engendrent sont mutuellement indépendantes. Or, si X est une variable aléatoire réelle et si f est une fonction mesurable, la variable aléatoire réelle $f(X)$ est une variable aléatoire réelle dont la tribu engendrée est incluse dans celle engendrée par X .*

Donnons un exemple illustrant la Proposition 3.4.9.

Exemple 3.4.12. *Soient X_1, X_2, X_3 et X_4 quatre variables aléatoires réelles mutuellement indépendantes. Alors, les variables aléatoires réelles $X_1^2, |X_2|, \log |X_3|$ et e^{X_4} sont mutuellement indépendantes.*

Variables aléatoires discrètes

4.1 Introduction

Dans tout le chapitre, on suppose l'ensemble des réalisations possibles $X(\Omega)$ fini ou infini dénombrable. En d'autres termes, on suppose que l'on a $X(\Omega) = \{x_i : i \in I\}$ où I est un ensemble au plus dénombrable d'indices.

Bien que n'étant pas un pré-requis indispensable au chapitre, savoir ce qu'est une distribution et plus exactement comprendre ce qu'est la distribution de Dirac est un plus indéniable. Quelques éléments de statistiques descriptives peuvent également être un plus ; tout comme la notion de forme quadratique et de forme bilinéaire symétrique. En revanche, savoir ce qu'est le symbole \sum est indispensable.

L'objectif principal du chapitre est la notion d'espérance dont beaucoup de concepts découlent. Il est attendu du lecteur à la fin de son apprentissage de ce chapitre qu'il soit au courant de l'existence de conditions pour que cette espérance existe mais aussi de la définition de ladite espérance. Bien sûr, connaître les propriétés les plus basiques et classiques est un pré-requis à la poursuite de l'étude des probabilités : l'espérance d'une constante, la formule de transfert, la linéarité, la positivité de l'espérance d'une variable aléatoire positive, l'espérance d'un produit de variables aléatoires réelles indépendantes et enfin l'inégalité de Markov. De l'espérance, nous introduirons la notion de variance (puis d'écart-type) qui mesure la dispersion d'une loi de probabilité. Et, nous verrons l'inégalité de Bienaymé-Tchebychev dont l'utilité n'est pas à démontrer, que ce soit pour la loi faible des grands nombres ou pour la qualité des estimateurs en estimation ponctuelle. Enfin, la notion de covariance étant essentielle dès lors que l'on traite de vecteurs gaussiens ou de processus gaussiens, nous attendons qu'elle soit bien maîtrisée. De même, savoir centrer et réduire une variable aléatoire réelle discrète n'est pas une option.

4.2 Caractérisation de la loi

Théorème 4.2.1. *La loi de probabilité d'une variable aléatoire réelle X est caractérisée par la donnée des probabilités $\mathbb{P}(X = x_i)$. En d'autres termes, pour toute*

partie S de \mathbb{R} , on peut calculer $\mathbb{P}(X \in S)$ si l'on connaît les probabilités $\mathbb{P}(X = x_i)$ pour tout $i \in I$.

Démonstration. En effet, \mathbb{P}_X est une probabilité sur un univers dénombrable à savoir $X(\Omega)$. Il suffit alors d'appliquer le Théorème 2.3.14. On étend ensuite en une probabilité sur \mathbb{R} en mettant une masse égale à 0 à tout ensemble qui est d'intersection vide avec $X(\Omega)$. \square

En particulier, en prenant $S = \mathbb{R}$, on a

$$\mathbb{P}(S) = \sum_{i \in I} \mathbb{P}(X = x_i) = 1.$$

Le Théorème 4.2.1 signifie que la connaissance de la probabilité de chacune des réalisations possibles suffit pour caractériser complètement la loi de X . On peut mettre tous ces résultats sous la forme d'un tableau comme dans l'Exemple 3.3.5 (page 51).

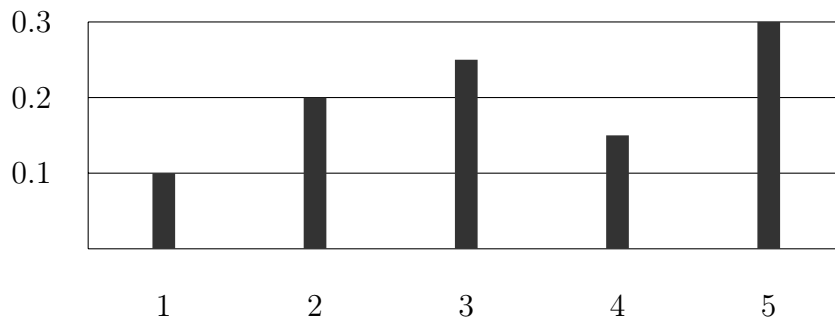
On peut aussi écrire :

$$\mathbb{P}_X = 0.1\delta_1 + 0.2\delta_2 + 0.25\delta_3 + 0.15\delta_4 + 0.3\delta_5.$$

On reprend ici une écriture de type distribution.

On peut également utiliser un diagramme en bâtons :

FIGURE 4.1 – Diagramme en bâtons de la loi de probabilité



4.2.1 Calcul de la probabilité d'un intervalle

On se donne un espace fondamental Ω muni d'une probabilité \mathbb{P} . Soit X une variable aléatoire réelle de Ω dans \mathbb{R} . On peut calculer la probabilité d'un intervalle \mathcal{K} de deux manières différentes : soit en utilisant la loi de probabilité soit en utilisant la fonction de répartition.

4.2.1.1 En utilisant la loi de probabilité

Soit $J \subset I$ l'ensemble des indices i tels que $x_i \in \mathcal{K}$. On a alors

$$\mathbb{P}(X \in \mathcal{K}) = \sum_{i \in J} \mathbb{P}(X = x_i) .$$

On connaît ainsi $\mathbb{P}(X \in \mathcal{K})$.

4.2.1.2 En utilisant la fonction de répartition

Considérons tous les types possibles d'intervalles.

- Si $\mathcal{K} =]a; b]$ avec $a, b \in \mathbb{R}$. Alors :

$$\mathbb{P}(X \in]a; b]) = \mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) ,$$

d'après la Proposition 3.3.10 de la page 54.

- Si $\mathcal{K} = [a; b]$ avec $a, b \in \mathbb{R}$. Alors :

$$\begin{aligned} \mathbb{P}(X \in [a; b]) &= \mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a^-) \\ &= F_X(b) - (F_X(a) - \mathbb{P}_X(\{a\})) . \end{aligned}$$

- Si $\mathcal{K} = [a; b[$ avec $a, b \in \mathbb{R}$. Alors :

$$\begin{aligned} \mathbb{P}(X \in [a; b[) &= \mathbb{P}(a \leq X < b) = F_X(b^-) - F_X(a^-) \\ &= (F_X(b) - \mathbb{P}_X(\{b\})) - (F_X(a) - \mathbb{P}_X(\{a\})) . \end{aligned}$$

- Si $\mathcal{K} =]a; b[$ avec $a, b \in \mathbb{R}$. Alors :

$$\begin{aligned} \mathbb{P}(X \in]a; b[) &= \mathbb{P}(a < X < b) = F_X(b^-) - F_X(a) \\ &= (F_X(b) - \mathbb{P}_X(\{b\})) - F_X(a) . \end{aligned}$$

- Si $\mathcal{K} =]-\infty; b]$ avec $b \in \mathbb{R}$. Alors :

$$\mathbb{P}(X \in]-\infty; b]) = \mathbb{P}(X \leq b) = F_X(b) .$$

- Si $\mathcal{K} =]-\infty; b[$ avec $b \in \mathbb{R}$. Alors :

$$\mathbb{P}(X \in]-\infty; b[) = \mathbb{P}(X < b) = F_X(b^-) = (F_X(b) - \mathbb{P}_X(\{b\})) .$$

- Si $\mathcal{K} =]a; +\infty[$ avec $a \in \mathbb{R}$. Alors :

$$\mathbb{P}(X \in]a; +\infty[) = \mathbb{P}(a < X) = 1 - F_X(a) .$$

- Si $\mathcal{K} = [a; +\infty[$ avec $a \in \mathbb{R}$. Alors :

$$\mathbb{P}(X \in [a; +\infty[) = \mathbb{P}(a \leq X) = 1 - F_X(a^-) = 1 - (F_X(a) - \mathbb{P}_X(\{a\})) .$$

C'est pourquoi il est crucial d'être précautionneux avec les **bornes** des intervalles que l'on manipule.

4.3 Caractéristiques

Dans cette section, on regarde les grandeurs habituelles des variables aléatoires réelles. Notons que la plupart des définitions et des résultats s'étendent pour des variables aléatoires réelles continues, voir le Chapitre 8 à la page 117.

4.3.1 Espérance d'une variable aléatoire réelle

On commence par présenter l'espérance mathématique. On parle également de moyenne. Introduisons cette dernière par un exemple.

Exemple 4.3.1. *On se donne un jeu de hasard tel qu'à chaque partie, on ait la probabilité*

- 0.2 de perdre deux euros,
- 0.3 de perdre un euro,
- 0.1 de ne rien perdre ni gagner,
- 0.2 de gagner un euro,
- 0.2 de gagner deux euros.

Quel est le gain moyen par partie ? (On parle aussi d'espérance mathématique).

Supposons que l'on joue un grand nombre N de parties. On note N_{-2} le nombre de parties où l'on a perdu 2 euros, N_{-1} le nombre de parties où l'on a perdu 1 euro, N_0 le nombre de parties où l'on n'a ni perdu ni gagné. On note également N_1 le nombre de parties où l'on a gagné 1 euro et N_2 le nombre de parties où l'on a gagné 2 euros. Par définition, on a

$$N_{-2} + N_{-1} + N_0 + N_1 + N_2 = N.$$

Et, la somme totale que l'on a gagnée au bout des N parties est

$$S_N := (-2) \times N_{-2} + (-1) \times N_{-1} + 0 \times N_0 + 1 \times N_1 + 2 \times N_2.$$

Ainsi, en moyenne, la somme que l'on a gagnée à chaque partie est

$$m_N := \frac{S_N}{N} = (-2) \times \frac{N_{-2}}{N} + (-1) \times \frac{N_{-1}}{N} + 0 \times \frac{N_0}{N} + 1 \times \frac{N_1}{N} + 2 \times \frac{N_2}{N}.$$

Or, par définition de la probabilité vue en tant que limite de la fréquence de réalisation d'un résultat aléatoire, on dispose des limites suivantes :

$$\lim_{N \rightarrow +\infty} \frac{N_{-2}}{N} = 0.2, \quad \lim_{N \rightarrow +\infty} \frac{N_{-1}}{N} = 0.3, \quad \lim_{N \rightarrow +\infty} \frac{N_0}{N} = 0.1,$$

$$\lim_{N \rightarrow +\infty} \frac{N_1}{N} = 0.2 \text{ et } \lim_{N \rightarrow +\infty} \frac{N_2}{N} = 0.2.$$

Ainsi, en faisant tendre le nombre de parties vers l'infini, le gain moyen par partie est égal à

$$m_\infty = (-2) \times 0.2 + (-1) \times 0.3 + 0 \times 0.1 + 1 \times 0.2 + 2 \times 0.2 = -0.1.$$

En d'autres termes, en moyenne, on perd dix centimes à chaque partie. Ce nombre m_∞ est appelé l'espérance de la variable aléatoire réelle X correspondant au gain obtenu durant une partie.

4.3.1.1 Définition

On présente maintenant la définition précise de l'espérance dans le cas des variables aléatoires réelles discrètes.

Définition 4.3.2. Soit une variable aléatoire réelle discrète X . On pose $X(\Omega) = \{x_i : i \in I\}$ où I est un ensemble fini ou infini dénombrable d'indices. On dit que X admet une espérance si l'on a

$$\sum_{i \in I} |x_i| \mathbb{P}(X = x_i) < \infty.$$

Dans ce cas, l'espérance de X est égale à

$$\mathbb{E}[X] := \sum_{i \in I} x_i \mathbb{P}(X = x_i). \quad (4.1)$$

Remarque 4.3.3. On doit supposer la finitude de la quantité $\sum_{i \in I} |x_i| \mathbb{P}(X = x_i)$ sans quoi l'espérance telle qu'elle est définie dans (4.1) n'a pas forcément de sens (voir théorème de réarrangement de Riemann). En effet, considérons par exemple une variable aléatoire discrète X telle que

$$\mathbb{P}_X = \sum_{n=1}^{\infty} \frac{6}{\pi^2} \frac{1}{n^2} \delta_{(-1)^n n}.$$

D'abord, comme $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$, \mathbb{P}_X est bien une probabilité sur \mathbb{Z} . Ensuite, $\sum_{n=1}^{\infty} |(-1)^n n| \times \mathbb{P}(X = (-1)^n n) = \sum_{n=1}^{\infty} \frac{1}{n} = +\infty$. On a ici un exemple de variable aléatoire n'admettant pas d'espérance bien que la quantité $\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{(-1)^k}{k} = \log(2)$ soit finie.

Illustrons l'espérance avec l'exemple suivant :

Exemple 4.3.4. On considère l'expérience aléatoire e qui consiste à jeter un dé. Soit X la variable aléatoire réelle qui a pour réalisation la face obtenue sur le dé. Calculons l'espérance de X :

$$\begin{aligned} \mathbb{E}[X] &= 1 \times \mathbb{P}(X = 1) + 2 \times \mathbb{P}(X = 2) + 3 \times \mathbb{P}(X = 3) + 4 \times \mathbb{P}(X = 4) \\ &\quad + 5 \times \mathbb{P}(X = 5) + 6 \times \mathbb{P}(X = 6) \\ &= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} \\ &= \frac{21}{6} \\ &= 3.5. \end{aligned}$$

On peut regarder des cas plus compliqués.

Exercice 4.3.5. *On considère l'expérience aléatoire e qui consiste à jeter deux dés, un bleu et un rouge. Soit X la variable aléatoire réelle qui a pour réalisation la somme des deux faces obtenues sur les dés. Calculer l'espérance de X sans utiliser de résultats sur l'espérance autres que ceux déjà énoncés.*

Cet exercice se résout très facilement en utilisant la linéarité de l'espérance mathématique, que l'on n'a pas encore vue. Donnons un exemple très important de variable aléatoire :

Exemple 4.3.6. *On considère un évènement A de probabilité $\mathbb{P}(A)$. On peut lui associer la variable aléatoire $\mathbb{1}_A$ définie comme suit. $\mathbb{1}_A(\omega) = 1$ si $\omega \in A$ et $\mathbb{1}_A(\omega) = 0$ si $\omega \notin A$. Alors, la variable aléatoire $\mathbb{1}_A$ admet une espérance et cette dernière vaut $\mathbb{P}(A)$.*

Démonstration. La variable aléatoire $\mathbb{1}_A$ prend deux valeurs : 0 et 1. Donc $\mathbb{E}[\mathbb{1}_A] = 1 \times \mathbb{P}(\mathbb{1}_A = 1) + 0 \times \mathbb{P}(\mathbb{1}_A = 0) = \mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(\{\omega \in \Omega : \mathbb{1}_A(\omega) = 1\}) = \mathbb{P}(\{\omega \in \Omega : \omega \in A\}) = \mathbb{P}(A)$. \square

4.3.1.2 Propriétés

Présentons ici les propriétés classiques de l'espérance mathématique.

4.3.1.2.1 Espérance d'une variable aléatoire réelle constante Soit une variable aléatoire réelle X qui a toujours la même réalisation a . En d'autres termes, pour tout $\omega \in \Omega$, on a $X(\omega) = a$. L'ensemble des réalisations de X est donc $X(\Omega) = \{a\}$. Et, la loi de probabilité de X est $\mathbb{P}_X = \delta_a$ c'est-à-dire que l'on a $\mathbb{P}(X = a) = 1$. Par définition, on a donc

$$\mathbb{E}[X] = a \times \mathbb{P}(X = a) = a \times 1 = a.$$

Si a est une constante réelle, on a donc

$$\mathbb{E}[a] = a.$$

En particulier, pour toute variable aléatoire réelle X , on a $\mathbb{E}\{\mathbb{E}[X]\} = \mathbb{E}[X]$.

Remarque 4.3.7. *Pour être totalement rigoureux, il faudrait supposer que l'on a défini une variable aléatoire \tilde{X} qui à tout $\omega \in \Omega$ associe la valeur $\mathbb{E}[X]$. On ne fera plus la distinction dorénavant.*

Remarque 4.3.8. *Ainsi, en admettant la linéarité de l'espérance, on peut se représenter celle-ci comme une projection sur l'ensemble des variables aléatoires constantes. C'est aussi la raison pour laquelle l'espérance d'une variable aléatoire contient beaucoup moins d'informations que la variable aléatoire (ou même sa loi). En effet, une variable aléatoire vit dans un espace vectoriel de dimension infinie alors que \mathbb{R} est de dimension un.*

4.3.1.2.2 Espérance d'une fonction de variable aléatoire réelle Soit maintenant une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . Soit une fonction réelle (et mesurable) ψ de \mathbb{R} dans \mathbb{R} . Alors, $\psi(X)$ est une variable aléatoire réelle. On cherche à calculer l'espérance de $\psi(X)$.

Théorème 4.3.9 (Formule de transfert). *Supposons que X soit une variable aléatoire réelle discrète avec $X(\Omega) = \{x_i : i \in I\}$ où I est un ensemble au plus dénombrable d'indices. Alors, sous la finitude de la quantité $\sum_{i \in I} |\psi(x_i)| \mathbb{P}(X = x_i)$, la variable aléatoire $\psi(X)$ admet une espérance et de plus, on a*

$$\mathbb{E}[\psi(X)] = \sum_{i \in I} \psi(x_i) \mathbb{P}(X = x_i) .$$

On peut prouver ce résultat de deux manières différentes. D'abord, on peut construire cette espérance de la même manière que l'on a construit $\mathbb{E}[X]$ dans l'exemple introductif en utilisant la limite des fréquences. Donnons ici une autre preuve.

Procédons à celle-ci dans le cas où la fonction ψ est bijective. L'idée est la même avec une fonction quelconque mais choisir une fonction bijective permet de s'affranchir de certaines considérations techniques.

Démonstration. Par définition, l'ensemble des réalisations de la variable aléatoire réelle $\psi(X)$ est $\{\psi(x_i) : i \in I\}$. Et, par définition, on a

$$\begin{aligned} \mathbb{E}[\psi(X)] &= \sum_{i \in I} \psi(x_i) \mathbb{P}(\psi(X) = \psi(x_i)) \\ &= \sum_{i \in I} \psi(x_i) \mathbb{P}(X = x_i) , \end{aligned}$$

ce qui achève la preuve. □

Donnons un exemple classique de ce résultat.

Exemple 4.3.10 (Moment à l'ordre n). *Soit $n \in \mathbb{N}^*$. Prenons $\psi_n(x) := x^n$. On peut alors calculer, si elle existe, la quantité*

$$\mathbb{E}[X^n] = \sum_{i \in I} x_i^n \mathbb{P}(X = x_i) .$$

Cette quantité est appelée "moment d'ordre n de la variable aléatoire réelle X ".

4.3.1.2.3 Linéarité de l'espérance Soient deux variables aléatoires réelles X et Y définies sur le même espace fondamental. Soit $\lambda \in \mathbb{R}$. On suppose que les deux variables aléatoires réelles X et Y admettent chacune une espérance. On a alors

$$\mathbb{E}[\lambda X + Y] = \lambda \mathbb{E}[X] + \mathbb{E}[Y] . \tag{4.2}$$

L'égalité

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

est admise dans ce cours. L'idée sous-jacente est la suivante :

$$\begin{aligned} \mathbb{E}[X + Y] &= \int_{\Omega} (X(\omega) + Y(\omega))\mathbb{P}(d\omega) \\ &= \int_{\Omega} X(\omega)\mathbb{P}(d\omega) + \int_{\Omega} Y(\omega)\mathbb{P}(d\omega) \\ &= \mathbb{E}[X] + \mathbb{E}[Y]. \end{aligned}$$

Quant à l'égalité

$$\mathbb{E}[\lambda X] = \lambda\mathbb{E}[X],$$

elle découle immédiatement du Théorème 4.3.9 en prenant la fonction $\psi(x) := \lambda \times x$. On peut généraliser ce résultat.

Théorème 4.3.11. *Soient n variables aléatoires réelles X_1, \dots, X_n définies sur un même espace fondamental Ω muni d'une probabilité \mathbb{P} . Soient $\lambda_0, \lambda_1, \dots, \lambda_n$ des réels. On a alors*

$$\mathbb{E}[\lambda_0 + \lambda_1 X_1 + \dots + \lambda_n X_n] = \lambda_0 + \lambda_1 \mathbb{E}[X_1] + \dots + \lambda_n \mathbb{E}[X_n].$$

Ce résultat de linéarité permet de résoudre immédiatement l'Exercice 4.3.5.

4.3.1.2.4 Espérance d'une variable aléatoire réelle de signe constant

Proposition 4.3.12. *Soit une variable aléatoire réelle X qui a toutes ses réalisations possibles positives. Alors son espérance est positive. De même, si la variable aléatoire réelle X est négative alors son espérance est négative :*

$$\begin{aligned} X \geq 0 &\implies \mathbb{E}[X] \geq 0 \\ \text{et } X \leq 0 &\implies \mathbb{E}[X] \leq 0. \end{aligned}$$

Démonstration. On considère uniquement le cas où X est positive. En effet, si la positivité de X implique la positivité de son espérance, alors on utilise (4.2) avec $\lambda := -1$ et $Y = 0$.

On a, par définition :

$$\mathbb{E}[X] = \sum_{i \in I} x_i \mathbb{P}(X = x_i).$$

Or, pour tout $i \in I$, $x_i \geq 0$. De même, $\mathbb{P}(X = x_i) \geq 0$. On en déduit que $x_i \mathbb{P}(X = x_i) \geq 0$ pour tout $i \in I$. Une somme de termes positifs étant positive, la preuve est achevée.

□

4.3.1.2.5 Espérance d'un produit de variables aléatoires réelles mutuellement indépendantes On présente dans ce paragraphe un résultat important.

Proposition 4.3.13. *Soient n variables aléatoires réelles X_1, \dots, X_n définies sur un même espace fondamental Ω muni d'une probabilité \mathbb{P} . Supposons que les n variables aléatoires réelles sont mutuellement indépendantes, voir la Définition 3.4.1 à la page 56. On suppose également que chacune de ces variables aléatoires réelles admet une espérance. Alors, la variable aléatoire réelle $X_1 \times \dots \times X_n$ admet une espérance et de plus :*

$$\mathbb{E}[X_1 \times \dots \times X_n] = \mathbb{E}[X_1] \times \dots \times \mathbb{E}[X_n].$$

La preuve est admise.

Remarque 4.3.14. *Cette égalité est nécessaire mais elle n'est pas suffisante pour que l'on ait l'indépendance mutuelle des variables aléatoires.*

On peut aller plus loin.

Théorème 4.3.15. *Soient n variables aléatoires réelles X_1, \dots, X_n définies sur un même espace fondamental Ω muni d'une probabilité \mathbb{P} . Supposons que les n variables aléatoires réelles sont mutuellement indépendantes, voir la Définition 3.4.1 à la page 56. Soient n fonctions bornées f_1, \dots, f_n . Alors, la variable aléatoire réelle $f_1(X_1) \times \dots \times f_n(X_n)$ admet une espérance et de plus :*

$$\mathbb{E}[f_1(X_1) \times \dots \times f_n(X_n)] = \mathbb{E}[f_1(X_1)] \times \dots \times \mathbb{E}[f_n(X_n)]. \quad (4.3)$$

Notons que la réciproque du Théorème 4.3.15 est vraie.

Proposition 4.3.16. *Soient n variables aléatoires réelles X_1, \dots, X_n définies sur un même espace fondamental Ω muni d'une probabilité \mathbb{P} . Supposons que l'Égalité (4.3) est vérifiée pour toutes les fonctions bornées, alors les n variables aléatoires réelles X_1, \dots, X_n sont mutuellement indépendantes.*

4.3.1.2.6 Inégalité de Markov On donne maintenant une inégalité importante concernant l'espérance mathématique pour les queues de distribution.

Théorème 4.3.17 (Inégalité de Markov). *Soit une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . Supposons qu'elle admette une espérance finie c'est-à-dire que l'on ait $\mathbb{E}[|X|] < \infty$. Alors, pour tout $R > 0$, on a*

$$\mathbb{P}(|X| \geq R) \leq \frac{\mathbb{E}[|X|]}{R}. \quad (4.4)$$

Démonstration. D'abord, en posant la fonction bornée f_R définie par $f_R(x) := R\mathbb{1}_{|x| \geq R}$, on observe

$$R\mathbb{P}(|X| \geq R) = \mathbb{E}[f_R(X)],$$

d'après l'Exemple 4.3.6. Puis, on remarque l'inégalité

$$\begin{aligned} |X| - f_R(X) &= |X| - R\mathbb{1}_{|X| \geq R} \\ &= |X| - R\mathbb{1}_{R \leq |X|} \\ &\geq |X| - |X|\mathbb{1}_{R \leq |X|} \\ &\geq |X|\mathbb{1}_{|X| < R} \\ &\geq 0. \end{aligned}$$

Or, la linéarité de l'espérance nous donne :

$$\mathbb{E}[f_R(X)] = \mathbb{E}[|X|] - \mathbb{E}[|X| - f_R(X)].$$

Et, la variable aléatoire $|X| - f_R(X)$ étant positive, son espérance l'est aussi (voir Proposition 4.3.12) d'où l'on a

$$R\mathbb{P}(|X| \geq R) = \mathbb{E}[f_R(X)] \leq \mathbb{E}[|X|].$$

En divisant par R , on obtient l'Inégalité de Markov (4.4). \square

4.3.1.3 Variable aléatoire réelle centrée

Définition 4.3.18. On dit qu'une variable aléatoire réelle X est centrée lorsque son espérance mathématique est nulle.

Centrer une variable aléatoire réelle X , c'est lui retrancher son espérance, c'est-à-dire considérer la variable aléatoire réelle $X - \mathbb{E}[X]$. En effet, d'après la linéarité de l'espérance, on a

$$\mathbb{E}\{X - \mathbb{E}[X]\} = \mathbb{E}[X] - \mathbb{E}[\mathbb{E}[X]].$$

Or, l'espérance d'une variable aléatoire constante est cette même constante donc on en déduit que l'on a $\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]$ si bien que

$$\mathbb{E}\{X - \mathbb{E}[X]\} = 0.$$

4.3.2 Variance, Écart-type

La connaissance de l'espérance est insuffisante pour donner des informations générales précises sur une loi de probabilité. En effet, soient X_1 et X_2 deux variables aléatoires réelles dont les lois de probabilité sont les suivantes :

$$\mathbb{P}_{X_1} = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$$

et

$$\mathbb{P}_{X_2} = \frac{1}{100}\delta_{-99} + \frac{99}{100}\delta_1$$

c'est-à-dire

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_2 = -1) = \frac{1}{2}$$

et $\mathbb{P}(X_2 = -99) = 1 - \mathbb{P}(X_2 = 1) = \frac{1}{100}$.

On note que les deux variables aléatoires réelles ont la même espérance. En effet :

$$\mathbb{E}[X_1] = \frac{1}{2} \times 1 + \frac{1}{2} \times (-1) = \frac{1}{2}(1 - 1) = 0$$

et

$$\mathbb{E}[X_2] = \frac{99}{100} \times 1 + \frac{1}{100} \times (-99) = \frac{1}{100}(99 - 99) = 0 = \mathbb{E}[X_1].$$

Bien que les deux variables aléatoires réelles aient la même espérance, la seconde est plus dispersée. On mesure cette dispersion avec la variance et l'écart-type.

4.3.2.1 Définition

Définition 4.3.19. *Soit une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance. Alors, par définition, la variance de la variable aléatoire réelle X est*

$$\sigma^2(X) = \text{Var}(X) := \mathbb{E}\{(X - \mathbb{E}[X])^2\}.$$

D'après la positivité de l'espérance d'une variable aléatoire réelle positive, on en déduit que $\text{Var}(X)$ est positive dès qu'elle est définie. En fait, cette quantité est toujours définie mais elle peut éventuellement être égale à $+\infty$. Si elle est finie, on dit que la variance de X est finie.

Définition 4.3.20. *Soit une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance. Alors, par définition, l'écart-type de la variable aléatoire réelle X est*

$$\sigma(X) := \sqrt{\text{Var}(X)} = \sqrt{\mathbb{E}\{(X - \mathbb{E}[X])^2\}}.$$

La variance et l'écart-type sont des indicateurs de la dispersion de la loi de X autour de son espérance. Reprenons les deux variables aléatoires réelles mentionnées plus haut :

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_2 = -1) = \frac{1}{2}$$

et $\mathbb{P}(X_2 = -99) = 1 - \mathbb{P}(X_2 = 1) = \frac{1}{100}$.

On a $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$. Calculons maintenant les variances des deux variables aléatoires réelles :

$$\begin{aligned}\text{Var}(X_1) &= \frac{1}{2}(1-0)^2 + \frac{1}{2}(-1-0)^2 = 1 \\ \text{et } \text{Var}(X_2) &= \frac{1}{100}(-99-0)^2 + \frac{99}{100}(1-0)^2 = 99.\end{aligned}$$

On observe bien que la variable aléatoire réelle X_2 est plus dispersée que la variable aléatoire réelle X_1 .

Proposition 4.3.21. *Soit une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance ainsi qu'une variance finie. Alors, $\mathbb{E}[X^2]$ est finie et de plus*

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \quad (4.5)$$

Démonstration. On développe la variable aléatoire réelle

$$(X - \mathbb{E}[X])^2 = X^2 - 2\mathbb{E}[X]X + \mathbb{E}[X]^2.$$

Par linéarité de l'espérance, on obtient

$$\text{Var}(X) = \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}\{\mathbb{E}[X]^2\}.$$

Comme $\mathbb{E}[X]^2$ est une constante, le calcul donne

$$\text{Var}(X) = \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2,$$

ce qui achève la preuve. □

4.3.2.2 Inégalité de Bienaymé-Tchebychev

Théorème 4.3.22. *Soit une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance ainsi qu'une variance finie. Alors, pour tout $t > 0$, on a l'inégalité*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}. \quad (4.6)$$

Démonstration. On applique l'inégalité de Markov (4.4) à la variable aléatoire réelle $Y := |X - \mathbb{E}[X]|^2$ avec $R := t^2$. En effet :

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) = \mathbb{P}(|X - \mathbb{E}[X]|^2 \geq t^2),$$

d'après la bijectivité de la fonction $x \mapsto x^2$ sur \mathbb{R}_+ . □

On peut présenter l'inégalité de Bienaymé-Tchebychev d'une autre manière. En effet, comme dit plus haut, l'écart-type mesure la dispersion d'une variable aléatoire réelle X .

Théorème 4.3.23. *Soit une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance ainsi qu'une variance finie. Alors, pour tout $s > 0$, on a l'inégalité*

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq s\sqrt{\text{Var}(X)}\right) \leq \frac{1}{s^2}.$$

Démonstration. On applique l'inégalité de Bienaymé-Tchebychev (4.6) avec $t := s\sqrt{\text{Var}(X)}$. \square

Cette inégalité est importante pour établir la loi faible des grands nombres, voir la page 214.

Remarque 4.3.24. *Ainsi, cette inégalité nous donne une information sur la queue de distribution. Néanmoins, cette information n'est pas toujours optimale et il est parfois préférable d'utiliser les propriétés propres à la loi de la variable. Comme pour l'inégalité de Markov, ces inégalités n'ont d'intérêt que si l'on s'intéresse à un paramètre tendant vers l'infini. Mais, dès lors que l'on souhaite obtenir une estimation précise, il faut utiliser d'autres outils.*

Exemple 4.3.25. *Soit une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance ($\mathbb{E}[X] = m$) ainsi qu'une variance finie ($\text{Var}(X) = \sigma^2$ avec $\sigma > 0$). Alors, on en déduit*

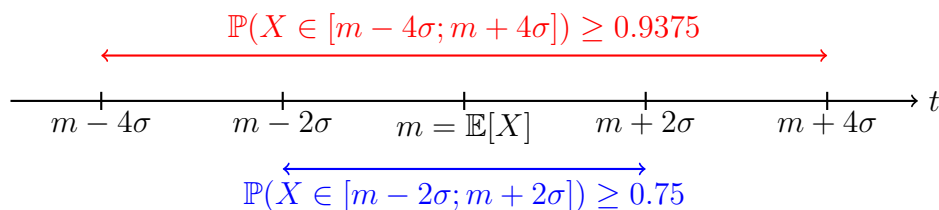
$$\mathbb{P}(m - 2\sigma < X < m + 2\sigma) \geq \frac{3}{4}.$$

De même :

$$\mathbb{P}(m - 4\sigma < X < m + 4\sigma) \geq \frac{15}{16}.$$

Cet exemple peut s'illustrer ainsi :

FIGURE 4.2 – Inégalité de Bienaymé-Tchebychev



Exercice 4.3.26. Soit une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance ($\mathbb{E}[X] = m$) ainsi qu'une variance finie ($\text{Var}(X) = \sigma^2$ avec $\sigma > 0$). Soit $\delta > 0$. Trouver un réel positif t_δ tel que l'on ait

$$\mathbb{P}(m - t_\delta\sigma < X < m + t_\delta\sigma) \geq 1 - \delta.$$

Remarque 4.3.27. La constante t_δ que l'on obtient ici n'est pas nécessairement optimale.

4.3.2.3 Propriétés

On présente maintenant les propriétés classiques sur la variance d'une variable aléatoire réelle.

4.3.2.3.1 Variance d'une variable aléatoire réelle constante Soit une variable aléatoire réelle X qui a toujours la même réalisation a : $\mathbb{P}(X = a) = 1$. Alors, on a

$$\text{Var}(X) = 0.$$

En effet, $\mathbb{E}[X^2] = \mathbb{E}[a^2] = a^2$ et $\mathbb{E}[X]^2 = \mathbb{E}[a]^2 = a^2$. Ainsi, la variance d'une variable aléatoire réelle qui a toujours la même réalisation est égale à 0. En d'autres termes :

$$\forall a \in \mathbb{R}, \text{ on a : } \text{Var}(a) = 0. \quad (4.7)$$

Remarque 4.3.28. Il est assez normal de constater qu'une variable aléatoire constante est sans dispersion.

4.3.2.3.2 Calcul de $\text{Var}[aX]$ avec $a \in \mathbb{R}$ En utilisant les diverses hypothèses sur l'espérance, on a

$$\begin{aligned} \text{Var}[aX] &= \mathbb{E}[(aX - \mathbb{E}[aX])^2] \\ &= \mathbb{E}[(aX - a\mathbb{E}[X])^2] \\ &= \mathbb{E}[a^2(X - \mathbb{E}[X])^2] \\ &= a^2\mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= a^2\text{Var}[X]. \end{aligned}$$

Ainsi, la variance de aX est égale à a^2 fois la variance de X :

$$\text{Var}[aX] = a^2\text{Var}[X].$$

Donc, l'écart-type de aX est égal à $|a|$ fois l'écart-type de X :

$$\sqrt{\text{Var}[aX]} = |a|\sqrt{\text{Var}[X]}.$$

Remarque 4.3.29. Il est essentiel de ne pas oublier la valeur absolue.

4.3.2.3.3 Calcul de $\text{Var}[X + Y]$ Soient deux variables aléatoires réelles X et Y définies sur un même espace fondamental Ω muni d'une probabilité \mathbb{P} . Alors, on a :

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] . \quad (4.8)$$

Exercice 4.3.30. Prouver l'Égalité (4.8).

Contrairement à l'espérance, la variance n'est pas linéaire. En effet, un terme supplémentaire intervient dans la somme.

Définition 4.3.31. Soient deux variables aléatoires réelles X et Y définies sur un même espace fondamental Ω muni d'une probabilité \mathbb{P} . On appelle covariance de X et Y , et l'on note $\text{Cov}(X, Y)$ la quantité

$$\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] .$$

Remarque 4.3.32. On peut observer : $\text{Cov}(X, X) = \text{Var}[X]$.

Ainsi, l'Égalité (4.8) devient

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}(X, Y) . \quad (4.9)$$

On peut alors remarquer

$$\text{Var}[X - Y] = \text{Var}[X] + \text{Var}[Y] - 2\text{Cov}(X, Y) .$$

Si la covariance des variables aléatoires réelles X et Y est nulle, alors la variance de la somme est égale à la somme des variances.

Théorème 4.3.33. Soient n variables aléatoires réelles X_1, \dots, X_n définies sur un même univers Ω muni d'une probabilité \mathbb{P} . On suppose que les n variables aléatoires réelles admettent une espérance et une variance finie. On suppose également qu'elles sont indépendantes deux à deux. Alors, la variance de leur somme est égale à la somme de leurs variances :

$$\text{Var}[X_1 + \dots + X_n] = \text{Var}[X_1] + \dots + \text{Var}[X_n] .$$

Observons que l'on n'a pas besoin de supposer l'indépendance mutuelle mais seulement l'indépendance deux à deux.

Démonstration. On procède au développement comme précédemment et l'on obtient

$$\text{Var}[X_1 + \dots + X_n] = \text{Var}[X_1] + \dots + \text{Var}[X_n] + \sum_{1 \leq i \neq j \leq n} \text{Cov}(X_i, X_j) .$$

Or, si $i \neq j$, l'indépendance des variables aléatoires réelles X_i et X_j donne

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] \\ &= \mathbb{E}[X_i - \mathbb{E}[X_i]] \times \mathbb{E}[X_j - \mathbb{E}[X_j]] \\ &= 0 \times 0 \\ &= 0 . \end{aligned}$$

Ceci achève la preuve. □

Remarque 4.3.34. Si $\text{Cov}(X, Y) = 0$, on en déduit que X et Y ne sont pas corrélées linéairement. Mais, on n'en déduit pas l'indépendance. Ainsi, l'indépendance est une propriété plus forte que la décorrélation.

En particulier, si X_1, \dots, X_n sont n variables aléatoires réelles indépendantes et identiquement distribuées, on a

$$\text{Var}[X_1 + \dots + X_n] = n\text{Var}[X_1].$$

Proposition 4.3.35. Soit une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance ainsi qu'une variance finie. Soit $a \in \mathbb{R}$. Alors, on a

$$\text{Var}[X + a] = \text{Var}[X].$$

Démonstration. On utilise l'Égalité (4.9) :

$$\text{Var}[X + a] = \text{Var}[X] + \text{Var}[a] + 2\text{Cov}(X, a).$$

Or, d'après (4.7), $\text{Var}[a] = 0$. On calcule maintenant la covariance :

$$\begin{aligned} \text{Cov}(X, a) &= \mathbb{E}[(X - \mathbb{E}[X])(a - \mathbb{E}[a])] \\ &= \mathbb{E}[(X - \mathbb{E}[X])(a - a)] \\ &= \mathbb{E}[0] \\ &= 0. \end{aligned}$$

Conséquemment, on obtient $\text{Var}[X + a] = \text{Var}[X]$. □

Ce résultat est naturel puisque l'ajout d'une constante ne fait que translater la variable aléatoire réelle X . Elle ne participe pas à sa dispersion.

On pouvait prouver la Proposition 4.3.35 en procédant comme suit :

$$\begin{aligned} \text{Var}[X + a] &= \mathbb{E}[(X + a - \mathbb{E}[X + a])^2] \\ &= \mathbb{E}[(X + a - \mathbb{E}[X] - a)^2] \\ &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \text{Var}[X]. \end{aligned}$$

4.3.2.4 Variable aléatoire réelle réduite

Définition 4.3.36. On dit qu'une variable aléatoire réelle X est réduite lorsque sa variance est égale à 1. Ainsi, réduire une variable aléatoire réelle consiste à la diviser par son écart-type.

Ainsi, une variable aléatoire réelle X est dite centrée, réduite si l'on a $\mathbb{E}[X] = 0$ et $\text{Var}[X] = 1$. Soit une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X a une espérance et admet une variance finie. Alors, la variable aléatoire réelle

$$Y := \frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}[X]}}$$

est centrée et réduite.

Exercice 4.3.37. *Montrer que Y est bien centrée et réduite.*

Remarque 4.3.38. *Centrer et réduire est crucial pour l'obtention d'intervalles de confiance, voir à la page 287.*

Lois de probabilité discrètes classiques

5.1 Introduction

Dans ce chapitre, nous avons choisi de ne présenter en détail que trois lois discrètes. Néanmoins, d'autres lois seront aussi étudiées dans le chapitre subséquent à la page 95.

Bien qu'une compréhension fine du produit de convolution et de la notion de distribution soit un plus, celle-ci n'est pas nécessaire à la bonne compréhension du chapitre.

Les objectifs relatifs à ce chapitre sont simples. Il s'agit de connaître la définition et les principales caractéristiques (espérance et variance par exemple) de la loi de Bernoulli, de la loi binomiale et de la loi de Poisson. À l'issue du chapitre, il conviendra de savoir lire les tables et d'être capable d'identifier les différentes lois. En particulier, il sera crucial que le lecteur sache qu'une somme de variables aléatoires indépendantes et suivant chacune une loi binomiale de même paramètre p est une variable aléatoire réelle suivant une loi binomiale. Le même phénomène de stabilité est vérifié pour la loi de Poisson.

On donne un résultat très utile :

Proposition 5.1.1. *Soit un espace fondamental Ω muni d'une probabilité \mathbb{P} . On considère deux variables aléatoires réelles X_1 et X_2 . On suppose que X_1 et X_2 sont discrètes et indépendantes. Alors, la variable aléatoire $X_1 + X_2$ est discrète et sa loi de probabilité est le produit de convolution des lois de probabilité de X_1 et de X_2 : $\mathbb{P}_{X_1+X_2} = \mathbb{P}_{X_1} * \mathbb{P}_{X_2}$.*

Rappel 5.1.2. *Pour tous les réels a et b , on a $\delta_a * \delta_b = \delta_{a+b}$ où δ_{x_0} est la distribution de Dirac en x_0 .*

Rappel 5.1.3. *Le produit de convolution est distributif par rapport à l'addition.*

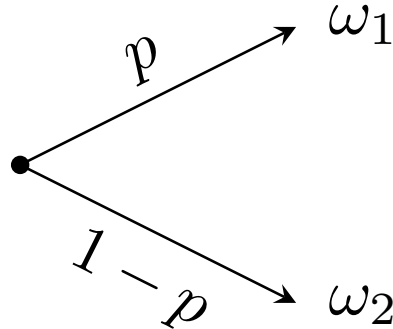
Exemple 5.1.4. *On peut ainsi écrire :*

$$\left(\frac{1}{4}\delta_0 + \frac{3}{4}\delta_1\right) * \left(\frac{1}{3}\delta_0 + \frac{2}{3}\delta_1\right) = \frac{1}{12}\delta_0 + \frac{5}{12}\delta_1 + \frac{6}{12}\delta_2.$$

5.2 Loi de Bernoulli

Soit une expérience aléatoire e à deux résultats possibles ω_1 et ω_2 , de probabilités respectives $\mathbb{P}(\omega_1) = p$ et $\mathbb{P}(\omega_2) = 1 - p$.

FIGURE 5.1 – Arbre de probabilité pour la loi de Bernoulli



L'espace fondamental est donc $\Omega = \{\omega_1, \omega_2\}$. On définit la variable aléatoire réelle X comme suit :

$$X(\omega_1) = 1 \quad \text{et} \quad X(\omega_2) = 0.$$

En d'autres termes, X est la variable aléatoire indicatrice de l'évènement $\{\omega_1\}$. L'ensemble des réalisations de la variable aléatoire réelle X est $X(\Omega) = \{0, 1\}$. La loi de probabilité de X est ici $\mathbb{P}_X = (1 - p)\delta_0 + p\delta_1$ ce qui donne le tableau :

TABLE 5.1 – Loi de probabilité de la variable aléatoire réelle X

k	0	1
$\mathbb{P}(X = k)$	$1 - p$	p

Définition 5.2.1. *On dit que la variable aléatoire réelle X suit la loi de Bernoulli de paramètre p si sa loi de probabilité correspond au Tableau 5.1.*

Notation 5.2.2. *La loi de Bernoulli de paramètre p est notée $\mathcal{B}(p)$.*

On étudie maintenant les caractéristiques de la loi X . Le calcul des moments d'une variable aléatoire réelle suivant $\mathcal{B}(p)$ donne

$$\mathbb{E}[X^n] = p \tag{5.1}$$

pour tout $n \in \mathbb{N}^*$ et l'on a également

$$\text{Var}[X] = p(1 - p). \tag{5.2}$$

Exercice 5.2.3. Démontrer les Égalités (5.1) et (5.2).

Remarque 5.2.4. L'univers Ω n'est pas nécessairement dénombrable. Il suffit pour s'en convaincre de considérer $\Omega := [0; 1]$ et $A := [0; p]$. On considère la probabilité sur $[0; 1]$ qui correspond à la "mesure de Haar". En d'autres termes, la probabilité d'un intervalle est la longueur de celui-ci. On admet qu'une telle mesure de probabilité existe bien, modulo la définition rigoureuse de la tribu Lebesguienne.

Alors, la variable aléatoire $X := \mathbb{1}_A$ prend deux valeurs : 0 et 1. Et, $\mathbb{P}(X = 1) = p$ bien que Ω ne soit pas dénombrable. Ainsi, $\mathbb{P}_X = (1 - p)\delta_0 + p\delta_1$.

5.3 Loi binomiale

On présente maintenant la loi binomiale associée au modèle binomial.

5.3.1 Formulation

On suppose vérifiées les trois hypothèses (\mathcal{H}_1) , (\mathcal{H}_2) et (\mathcal{H}_3) suivantes :

(\mathcal{H}_1) Soit une expérience aléatoire e à deux résultats possibles r_1 et r_2 de probabilités respectives p et $1 - p$.

(\mathcal{H}_2) On effectue n répétitions de l'expérience aléatoire e . Ainsi, l'expérience aléatoire est $E = (e_1, \dots, e_n)$.

(\mathcal{H}_3) Les expériences e_1, \dots, e_n sont mutuellement indépendantes (les événements associés à ces expériences sont donc mutuellement indépendants).

On se pose alors la question suivante.

Problème : Sous les trois hypothèses (\mathcal{H}_1) , (\mathcal{H}_2) et (\mathcal{H}_3) , quelle est la probabilité d'obtenir k fois le résultat r_1 (et donc $n - k$ fois le résultat r_2) pour tout $0 \leq k \leq n$?

Donnons un exemple pour justifier ce modèle.

Exemple 5.3.1. Dans un atelier, fonctionnent de façon indépendante trois machines. La probabilité de panne en une journée d'une de ces machines est 10%. Quelle est la probabilité pour qu'en une journée deux machines exactement tombent en panne ?

Vérifions que cette situation entre dans le cadre du modèle binomial décrit plus haut avec $n = 3$.

(\mathcal{H}_1) : On considère l'expérience aléatoire e qui consiste à observer le fonctionnement d'une machine en une journée. On a deux résultats possibles : $r_1 :=$ "la machine tombe en panne" et $r_2 :=$ "la machine ne tombe pas en panne". Alors, $\mathbb{P}(r_1) = 0.10$ et $\mathbb{P}(r_2) = 0.90$.

(\mathcal{H}_2) : On pose e_i l'expérience aléatoire consistant à observer la machine numéro i en une journée. L'expérience aléatoire est alors $E = (e_1, e_2, e_3)$.

(\mathcal{H}_3) : Les trois expériences aléatoires sont indépendantes car les trois machines fonctionnent indépendamment.

La question devient ici : quelle est la probabilité d'obtenir deux fois le résultat r_1 et une fois le résultat r_2 ?

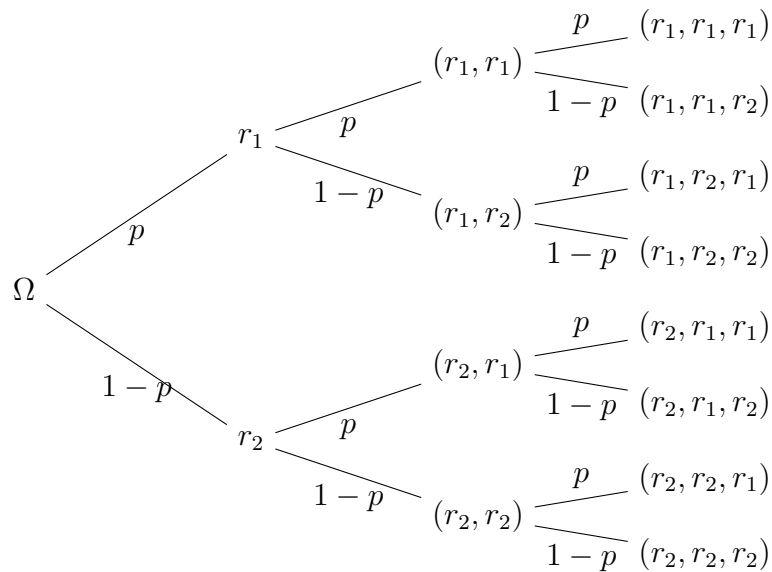
5.3.2 Espace fondamental

5.3.2.1 Cas où $n = 3$

On commence par résoudre le cas $n = 3$ pour simplifier et pour coller à l'Exemple 5.3.1.

Comme le nombre $n = 3$ est petit, il peut être pertinent de faire un arbre de probabilité :

FIGURE 5.2 – Arbre de probabilité pour la loi binomiale



Remarque 5.3.2. L'arbre de probabilités peut aider à comprendre le problème mais il ne doit pas se substituer au raisonnement.

Pour tout $(i, j) \in \{1; 2\}^2$, on note (r_i, r_j) l'obtention de r_i à l'expérience e_1 et celle de r_j à l'expérience e_2 .

De même, pour tout $(i, j, k) \in \{1; 2\}^3$, on note (r_i, r_j, r_k) l'obtention de r_i à l'expérience e_1 , celle de r_j à l'expérience e_2 et celle de r_k à l'expérience e_3 .

À partir de la Figure 5.2 à la page 80, on voit que l'on peut écrire l'espace fondamental comme suit :

$$\Omega := \left\{ (r_1, r_1, r_1), (r_1, r_1, r_2), (r_1, r_2, r_1), (r_1, r_2, r_2), \right. \\ \left. (r_2, r_1, r_1), (r_2, r_1, r_2), (r_2, r_2, r_1), (r_2, r_2, r_2) \right\}.$$

On remarque $\#\Omega = 8 = 2^3$. En effet, on a deux résultats possibles pour chacune des trois expériences aléatoires.

5.3.2.2 Cas général

On peut décrire Ω comme l'ensemble des suites ordonnées de n éléments de l'ensemble $\{r_1, r_2\}$. On remarque donc bien $\#\Omega = 2^n$.

5.3.3 Solution

5.3.3.1 Cas particulier où $n = 3$ et $k = 2$

Ici, $E = (e_1, e_2, e_3)$.

5.3.3.1.1 Traduction de l'énoncé On définit les évènements $A_i :=$ “on obtient le résultat r_i à l'expérience i ” pour $i = 1, 2, 3$. D'après l'hypothèse (\mathcal{H}_1) , on a $\mathbb{P}(A_i) = p$. Et, d'après l'hypothèse (\mathcal{H}_3) , les trois évènements A_1, A_2 et A_3 sont mutuellement indépendants.

5.3.3.1.2 Traduction de la question Soit l'évènement $B_2 :=$ “on obtient deux fois r_1 et une fois r_2 ”. Quelle est la probabilité de $\mathbb{P}(B_2)$?

5.3.3.1.3 Expression de B_2 en fonction des A_i L'écriture ensembliste de B_2 est

$$B_2 = \{(r_1, r_1, r_2), (r_1, r_2, r_1), (r_2, r_1, r_1)\}.$$

On a donc

$$B_2 = (A_1 \cap A_2 \cap A_3^c) \cup (A_1 \cap A_2^c \cap A_3) \cup (A_1^c \cap A_2 \cap A_3).$$

5.3.3.1.4 Calcul de $\mathbb{P}(B_2)$ Les trois évènements $A_1 \cap A_2 \cap A_3^c$, $A_1 \cap A_2^c \cap A_3$ et $A_1^c \cap A_2 \cap A_3$ sont deux à deux incompatibles. En effet, $A_1 \cap A_2 \cap A_3^c$ est inclus dans A_1 alors que $A_1^c \cap A_2 \cap A_3$ est inclus dans A_1^c . Et, $A_1 \cap A_2 \cap A_3^c$ est inclus dans A_2 tandis que $A_1 \cap A_2^c \cap A_3$ est inclus dans A_2^c .

Or, d'après l'axiome des probabilités totales (voir Définition 2.3.1 page 27), on a

$$\mathbb{P}(B_2) = \mathbb{P}(A_1 \cap A_2 \cap A_3^c) + \mathbb{P}(A_1 \cap A_2^c \cap A_3) + \mathbb{P}(A_1^c \cap A_2 \cap A_3).$$

Puis, comme les évènements sont mutuellement indépendants, on trouve

$$\mathbb{P}(A_1 \cap A_2 \cap A_3^c) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3^c) = p \times p \times (1 - p) = p^2(1 - p).$$

Conséquemment, on trouve $\mathbb{P}(B_2) = 3p^2(1 - p)$.

Exemple 5.3.3. *Pour revenir à l'Exemple 5.3.1, la probabilité que deux machines tombent en panne est égale à $3p^2(1 - p)$ avec $p = 10\%$. Ainsi, cette probabilité vaut à 2.7%.*

5.3.3.2 Cas général

On résout maintenant le cas général en procédant comme dans le cas particulier du paragraphe 5.3.3.1. Ici, $E = (e_1, \dots, e_n)$.

5.3.3.2.1 Traduction de l'énoncé On définit les évènements $A_i :=$ “on obtient le résultat r_1 à l'expérience i ” pour $1 \leq i \leq n$. D'après l'hypothèse (\mathcal{H}_1) , on a $\mathbb{P}(A_i) = p$. Et, d'après l'hypothèse (\mathcal{H}_3) , les n évènements A_1, \dots, A_n sont mutuellement indépendants.

5.3.3.2.2 Traduction de la question Soit l'évènement $B_k :=$ “on obtient k fois r_1 et $n - k$ fois r_2 ”. Quelle est la probabilité de $\mathbb{P}(B_k)$?

5.3.3.2.3 Expression de B_k en fonction des A_i L'évènement B_k est réalisé si le résultat effectif est un élément de l'espace fondamental tel que l'intersection de k évènements A_i et de $n - k$ évènements A_i^c est réalisée. On peut donc écrire

$$B_k = \bigcup_{\sigma \in \mathfrak{S}_n} (A_{\sigma(1)} \cap \dots \cap A_{\sigma(k)} \cap A_{\sigma(k+1)}^c \cap \dots \cap A_{\sigma(n)}^c),$$

où la réunion est prise sur l'ensemble \mathfrak{S}_n des permutations de l'ensemble $\llbracket 1; n \rrbracket$. En utilisant du dénombrement, on note que le nombre total d'ensembles différents parmi les $n!$ évènements de la forme $A_{\sigma(1)} \cap \dots \cap A_{\sigma(k)} \cap A_{\sigma(k+1)}^c \cap \dots \cap A_{\sigma(n)}^c$ est égal à $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, c'est-à-dire au coefficient binomial. On note $\widetilde{\mathfrak{S}}_n$ un sous-ensemble de \mathfrak{S}_n tel qu'il soit en bijection avec l'ensemble des $\binom{n}{k}$ évènements différents de la forme $A_{\sigma(1)} \cap \dots \cap A_{\sigma(k)} \cap A_{\sigma(k+1)}^c \cap \dots \cap A_{\sigma(n)}^c$.

5.3.3.2.4 Calcul de $\mathbb{P}(B_k)$ Les évènements

$$A_{\sigma(1)} \cap \dots \cap A_{\sigma(k)} \cap A_{\sigma(k+1)}^c \cap \dots \cap A_{\sigma(n)}^c$$

sont deux à deux disjoints de par leur construction. Or, d'après l'axiome des probabilités totales (voir Définition 2.3.1 page 27), on a

$$\mathbb{P}(B_k) = \sum_{\sigma \in \widetilde{\mathfrak{S}}_n} \mathbb{P}(A_{\sigma(1)} \cap \dots \cap A_{\sigma(k)} \cap A_{\sigma(k+1)}^c \cap \dots \cap A_{\sigma(n)}^c).$$

Puis, comme les évènements $A_{\sigma(i)}$, $i \in \llbracket 1; n \rrbracket$, sont mutuellement indépendants, on trouve

$$\begin{aligned} & \mathbb{P}(A_{\sigma(1)} \cap \cdots \cap A_{\sigma(k)} \cap A_{\sigma(k+1)}^c \cap \cdots \cap A_{\sigma(n)}^c) \\ &= \underbrace{\mathbb{P}(A_{\sigma(1)}) \times \cdots \times \mathbb{P}(A_{\sigma(k)})}_{p^k} \times \underbrace{\mathbb{P}(A_{\sigma(k+1)}^c) \times \cdots \times \mathbb{P}(A_{\sigma(n)}^c)}_{(1-p)^{n-k}} \\ &= p^k (1-p)^{n-k}. \end{aligned}$$

Conséquemment, on trouve $\mathbb{P}(B_k) = \binom{n}{k} p^k (1-p)^{n-k}$ vu qu'il y a $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ combinaisons. On peut maintenant écrire :

Théorème 5.3.4. *Sous les hypothèses (\mathcal{H}_1) , (\mathcal{H}_2) et (\mathcal{H}_3) , la probabilité d'obtenir k fois le résultat r_1 et $n - k$ fois le résultat r_2 est égale à $\binom{n}{k} p^k (1-p)^{n-k}$.*

On peut appliquer le modèle binomial aux sondages au hasard.

5.3.4 Définition

On suppose les trois conditions suivantes réalisées :

(\mathcal{H}_1) Soit une expérience aléatoire e à deux résultats possibles ω_1 et ω_2 , de probabilités respectives p et $1 - p$, voir Figure 5.2.

(\mathcal{H}_2) On effectue n fois l'expérience e . En d'autres termes, on effectue l'expérience aléatoire $E = (e_1, \dots, e_n)$.

(\mathcal{H}_3) Les n expériences aléatoires e_1, \dots, e_n sont mutuellement indépendantes.

Soit X la variable aléatoire réelle qui a pour réalisation le nombre d'obtentions du résultat ω_1 au cours des n expériences. On caractérise maintenant la loi de probabilité de X . L'ensemble des résultats possibles est $X(\Omega) = \llbracket 0; n \rrbracket$. Pour tout $k \in \llbracket 0; n \rrbracket$, on a

$$\mathbb{P}(X = k) = \mathbb{P}(\{\text{on obtient } k \text{ fois } \omega_1 \text{ et } n - k \text{ fois } \omega_2\}).$$

Sous les trois hypothèses (\mathcal{H}_1) , (\mathcal{H}_2) et (\mathcal{H}_3) , d'après le théorème du modèle binomial (Théorème 5.3.4), il vient

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (5.3)$$

En d'autres termes, on a $\mathbb{P}_X = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \delta_k$.

On peut observer que l'on a bien une probabilité. En effet, d'après la formule du binôme de Newton :

$$\sum_{k=0}^n \mathbb{P}(X = k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1.$$

Définition 5.3.5. *On dit que la variable aléatoire réelle X suit la loi binomiale de paramètres n et p lorsque l'Égalité (5.3) est satisfaite pour tout $k \in \llbracket 0; n \rrbracket$.*

Notation 5.3.6. *La loi binomiale de paramètres n et p est notée $\mathcal{B}(n, p)$.*

5.3.5 Décomposition de X

Avant d'étudier les caractéristiques d'une variable aléatoire réelle X qui suit la loi $\mathcal{B}(n, p)$, on se ramène au cas des variables aléatoires réelles qui suivent la loi $\mathcal{B}(p)$. On constate en effet :

$$X = X_1 + \cdots + X_n,$$

où les variables aléatoires réelles X_i suivent la loi de Bernoulli de paramètre p et sont mutuellement indépendantes.

En effet, l'idée fondamentale de la loi binomiale est de répéter *indépendamment* des expériences de Bernoulli de même paramètre p .

Remarque 5.3.7. *En particulier, on remarque :*

$$\mathbb{P}_X = \underbrace{((1-p)\delta_0 + p\delta_1) * \cdots * ((1-p)\delta_0 + p\delta_1)}_{n \text{ fois}}.$$

Remarque 5.3.8. *On peut alors écrire $\mathcal{B}(1, p) = \mathcal{B}(p)$.*

5.3.6 Caractéristiques

5.3.6.1 Espérance

Ici, $X(\Omega) = \{0, \dots, n\}$ et $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. On peut donc calculer l'espérance de X directement :

$$\mathbb{E}[X] = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}.$$

On procède comme suit.

$$\begin{aligned} \xi(x) &:= \sum_{k=0}^n k \binom{n}{k} x^k \\ &= x \times \frac{d}{dx} \sum_{k=0}^n \binom{n}{k} x^k \\ &= x \times \frac{d}{dx} (x+1)^n \\ &= nx(x+1)^{n-1}. \end{aligned}$$

En prenant $x := \frac{p}{1-p}$, il vient

$$\begin{aligned} \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} &= (1-p)^n \sum_{k=0}^n k \binom{n}{k} \left(\frac{p}{1-p}\right)^k \\ &= (1-p)^n n \frac{p}{1-p} \left(\frac{p}{1-p} + 1\right)^{n-1} \\ &= np. \end{aligned}$$

On peut éviter tous ces calculs en utilisant la décomposition de X . En effet, la linéarité de l'espérance nous donne

$$\mathbb{E}[X] = \mathbb{E}[X_1 + \dots + X_n] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n] = n \times p.$$

Une troisième méthode peut être utilisée en remarquant l'égalité suivante :

$$\begin{aligned} k \binom{n}{k} &= k \frac{n!}{k!(n-k)!} \\ &= \frac{n!}{(k-1)!((n-1)-(k-1))!} \\ &= n \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} \\ &= n \binom{n-1}{k-1}, \end{aligned}$$

pour $k \geq 1$.

On obtient alors :

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \\ &= np (p + (1-p))^{n-1} \\ &= np. \end{aligned}$$

5.3.6.2 Variance

On peut obtenir la variance de X en procédant au calcul direct. D'abord :

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2] - n^2 p^2 \\ &= \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} - n^2 p^2. \end{aligned}$$

Puis, on introduit la fonction :

$$\begin{aligned}\zeta(x) &:= \sum_{k=0}^n k^2 \binom{n}{k} x^k \\ &= \sum_{k=0}^n k(k-1) \binom{n}{k} x^k + \sum_{k=0}^n k \binom{n}{k} x^k \\ &= \sum_{k=0}^n k(k-1) \binom{n}{k} x^k + \xi(x),\end{aligned}$$

où $\xi(x) := \sum_{k=0}^n k \binom{n}{k} x^k$.

Or, on connaît l'expression de $\xi(x)$. On obtient donc :

$$\begin{aligned}\zeta(x) &= x^2 \times \frac{d^2}{dx^2} \sum_{k=0}^n \binom{n}{k} x^k + \xi(x) \\ &= x^2 \times \frac{d^2}{dx^2} (1+x)^n + nx(1+x)^{n-1} \\ &= n(n-1)x^2(1+x)^{n-2} + nx(1+x)^{n-1}.\end{aligned}$$

En prenant $x := \frac{p}{1-p}$, il vient

$$\begin{aligned}& \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} \\ &= (1-p)^n \sum_{k=0}^n k^2 \binom{n}{k} x^k \\ &= (1-p)^n \left[n(n-1) \frac{p^2}{(1-p)^2} \left(1 + \frac{p}{1-p}\right)^{n-2} + n \frac{p}{1-p} \left(1 + \frac{p}{1-p}\right)^{n-1} \right] \\ &= n(n-1)p^2 + np = n^2p^2 - np^2 + np.\end{aligned}$$

Il s'ensuit :

$$\text{Var}[X] = n^2p^2 - np^2 + np - n^2p^2 = np(1-p).$$

Toutefois, il est préférable d'utiliser la décomposition de X en somme de variables aléatoires réelles indépendantes suivant la loi de Bernoulli. En effet, d'après le Théorème 4.3.33, page 73, on a

$$\text{Var}[X] = \text{Var}[X_1 + \cdots + X_n] = \text{Var}[X_1] + \cdots + \text{Var}[X_n] = n \times p(1-p).$$

5.3.7 Somme de lois binomiales

La décomposition de X en somme de variables aléatoires réelles indépendantes et identiquement distribuées suivant la loi de Bernoulli signifie qu'une somme de n variables aléatoires réelles indépendantes et identiquement distribuées suivant la loi $\mathcal{B}(1, p)$ est une variable aléatoire réelle suivant la loi $\mathcal{B}(n, p)$. On étend maintenant ce résultat.

Théorème 5.3.9. *Soient r variables aléatoires réelles mutuellement indépendantes X_1, \dots, X_r . On suppose que pour tout $1 \leq i \leq r$, la variable aléatoire réelle X_i suit la loi $\mathcal{B}(n_i, p)$. Alors, la variable aléatoire réelle $X := X_1 + \dots + X_r$ suit la loi binomiale de paramètres $n := \sum_{i=1}^r n_i$ et p , $\mathcal{B}(\sum_{i=1}^r n_i, p)$. En d'autres termes :*

$$\mathcal{L}(X_1 + \dots + X_r) = \mathcal{B}(n_1 + \dots + n_r, p) .$$

Démonstration. Pour tout $1 \leq i \leq r$, X_i suit la loi binomiale de paramètres n_i et p . Conséquemment, on peut écrire :

$$X_i = X_i^1 + \dots + X_i^{n_i} ,$$

où les variables aléatoires réelles X_i^s , $1 \leq s \leq n_i$ sont mutuellement indépendantes et suivent la loi de Bernoulli de paramètre p . Conséquemment, on a

$$X := X_1 + \dots + X_r = \sum_{i=1}^r \sum_{s=1}^{n_i} X_i^s .$$

Ainsi, X est la somme de $n_1 + \dots + n_r$ variables aléatoires réelles mutuellement indépendantes et qui suivent la loi de Bernoulli de paramètre p . Ceci achève la preuve. \square

Exercice 5.3.10. *En utilisant les distributions et le produit de convolution, donner une autre preuve de ce théorème.*

Donnons un corollaire immédiat du Théorème 5.3.9.

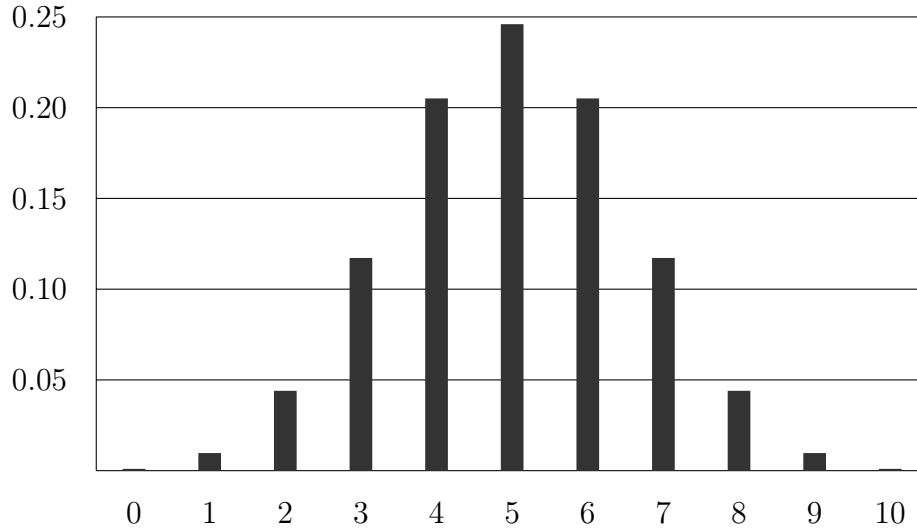
Corollaire 5.3.11. *Soit X de loi $\mathcal{B}(m, p)$ et Y de loi $\mathcal{B}(n, p)$. On suppose que X et Y sont indépendantes. Alors $X + Y$ suit la loi $\mathcal{B}(m + n, p)$.*

Remarque 5.3.12. *Si X_1 suit une loi binomiale de paramètres n_1 et p_1 et si X_2 en suit une de paramètres n_2 et p_2 et si de plus X_1 et X_2 sont indépendantes, alors on connaît la loi de $X_1 + X_2$ bien que ça ne soit pas une loi binomiale.*

5.3.8 Courbe représentative de la loi binomiale

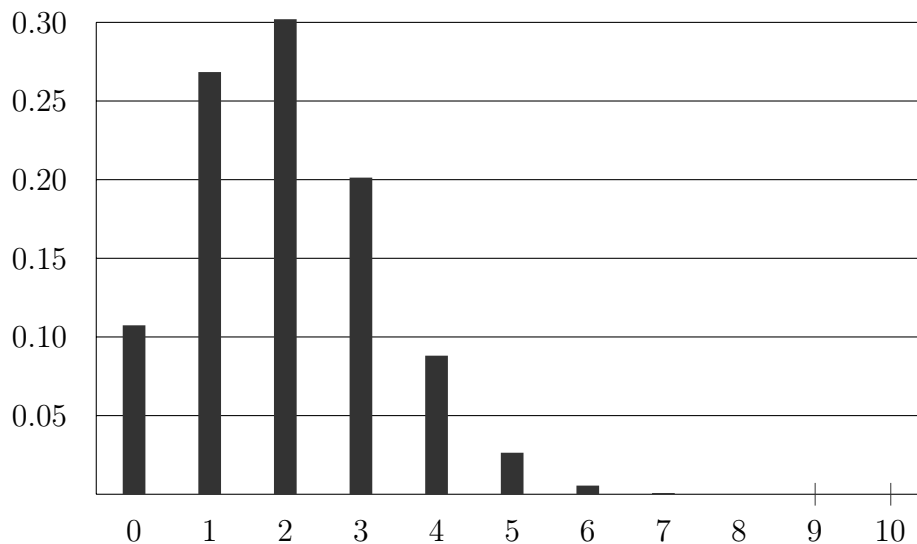
On trace le diagramme en bâtons de la loi (avec $n = 10$ et $p = 0.5$) :

FIGURE 5.3 – Diagramme en bâtons de la loi binomiale 1



On remarque que la loi binomiale est symétrique par rapport à $\frac{n}{2}$ pour tout $n \in \mathbb{N}^*$ lorsque $p = 0.5$. On trace le diagramme en bâtons avec d'autres paramètres (avec $n = 10$ et $p = 0.2$) :

FIGURE 5.4 – Diagramme en bâtons de la loi binomiale 2



On constate qu'il n'y a pas de symétrie dans le cas où $p \neq \frac{1}{2}$.

5.3.9 Tables de la loi binomiale

5.3.9.0.1 Pour $p \leq 0.5$ Les valeurs

$$p_k := \binom{n}{k} p^k (1-p)^{n-k}$$

sont tabulées pour $p = 0.05$, $p = 0.1$, $p = 0.2$, $p = 0.3$, $p = 0.4$ et $p = 0.5$ et pour $n = 10$, $n = 20$, $n = 30$, $n = 40$ et $n = 50$. Les valeurs

$$\sum_{i=0}^k p_i = \mathbb{P}(X \leq k) = F_X(k)$$

sont également tabulées.

Voir page 356.

5.3.9.1 Si $p > 0.5$

Soit une variable aléatoire réelle X qui suit la loi $\mathcal{B}(n, p)$ avec $p > 0.5$. On considère alors la variable aléatoire réelle $Y := n - X$. Calculons la loi de Y .

Les réalisations possibles de X sont $0, \dots, n$. De même, les réalisations possibles de Y sont $n, \dots, 0$. Et :

$$\begin{aligned} \mathbb{P}(Y = k) &= \mathbb{P}(n - X = k) \\ &= \mathbb{P}(X = n - k) \\ &= \binom{n}{n-k} p^{n-k} (1-p)^{n-(n-k)} \\ &= \binom{n}{k} (1-p)^k (1 - (1-p))^{n-k}. \end{aligned}$$

Ainsi, Y suit la loi binomiale de paramètres n et $1 - p \leq 0.5$.

Proposition 5.3.13. Si $\mathcal{L}(X) = \mathcal{B}(n, p)$ alors $\mathcal{L}(n - X) = \mathcal{B}(n, 1 - p)$.

On peut donc utiliser les tables fournies à la page 356.

Exemple 5.3.14. Soit une variable aléatoire réelle X qui suit la loi $\mathcal{B}(40, 0.9)$. Quelle est la probabilité que X soit égale à 30 ?

Soit $Y := 40 - X$. Alors $\mathcal{L}(Y) = \mathcal{B}(40, 0.1)$. On peut ainsi calculer

$$\mathbb{P}(X = 30) = \mathbb{P}(40 - Y = 30) = \mathbb{P}(Y = 10) = 0.0036,$$

d'après les tables.

5.3.9.2 Approximations de la loi binomiale pour n grand

Ce paragraphe prend tout son sens à condition que la section sur la convergence en loi, voir page 209 ait été lue avant.

En pratique, $n > 50$. On distingue suivant deux cas.

1. Si np est petit (typiquement, $np \leq 16$), on approxime la loi binomiale $\mathcal{B}(n, p)$ par la loi de Poisson de paramètre np , $\mathcal{P}(np)$, voir Section 5.4.
2. Si np et $n(1 - p)$ sont grands (en pratique, $np > 16$ et $n(1 - p) > 16$), on approche la loi binomiale $\mathcal{B}(n, p)$ par la loi normale $\mathcal{N}(np, np(1 - p))$, voir Section 9.4 à la page 138.

Soit X une variable aléatoire réelle qui suit la loi $\mathcal{B}(n, p)$. Alors, on a $\mathbb{P}(X = k) \approx \xi(k + \frac{1}{2}) - \xi(k - \frac{1}{2})$ où la fonction ξ est la fonction de répartition de la loi normale $\mathcal{N}(np, np(1 - p))$.

La justification théorique de la seconde approximation provient du Théorème central de la limite, voir la Section 15.4 à la page 217.

5.4 Loi de Poisson

La loi de Poisson sert à définir les processus de Poisson (que l'on ne présente pas dans ce livre). Ceux-ci peuvent être utilisés pour modéliser la radioactivité ou les incendies de forêt.

5.4.1 Définition

Définition 5.4.1. Soit λ un paramètre réel strictement positif. On dit que la variable aléatoire réelle X suit la loi de Poisson de paramètre λ lorsque l'ensemble des réalisations de X est l'ensemble des entiers; $X(\Omega) = \mathbb{N}$ et lorsque l'on a de plus

$$\mathbb{P}[X = n] = \frac{\lambda^n}{n!} e^{-\lambda}.$$

En d'autres termes, $\mathbb{P}_X = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \delta_n$.

Exercice 5.4.2. Montrer que l'on a bien défini une loi de probabilité sur \mathbb{N} .

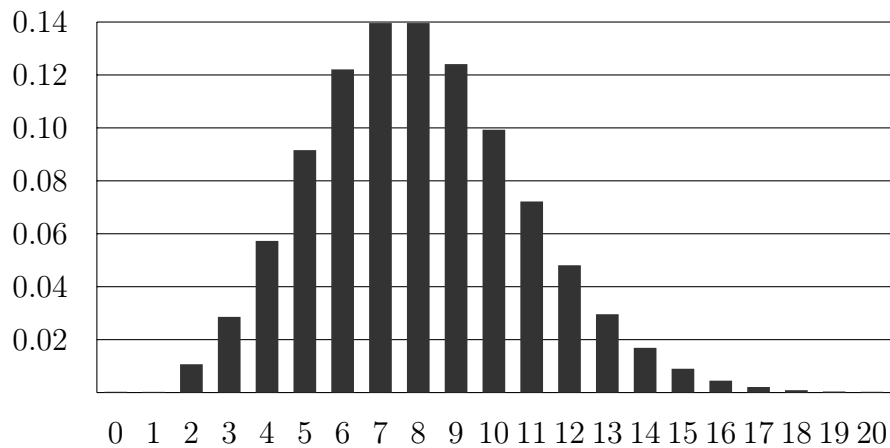
Notation 5.4.3. La loi de Poisson est notée $\mathcal{P}(\lambda)$.

Exemple 5.4.4. Sous certaines hypothèses, la variable aléatoire réelle qui a pour réalisation le nombre d'arrivées dans une file d'attente durant un intervalle de temps d'une heure suit une loi de Poisson de paramètre λ , où λ est le nombre moyen d'arrivées par heure.

5.4.2 Courbe représentative de la loi de Poisson

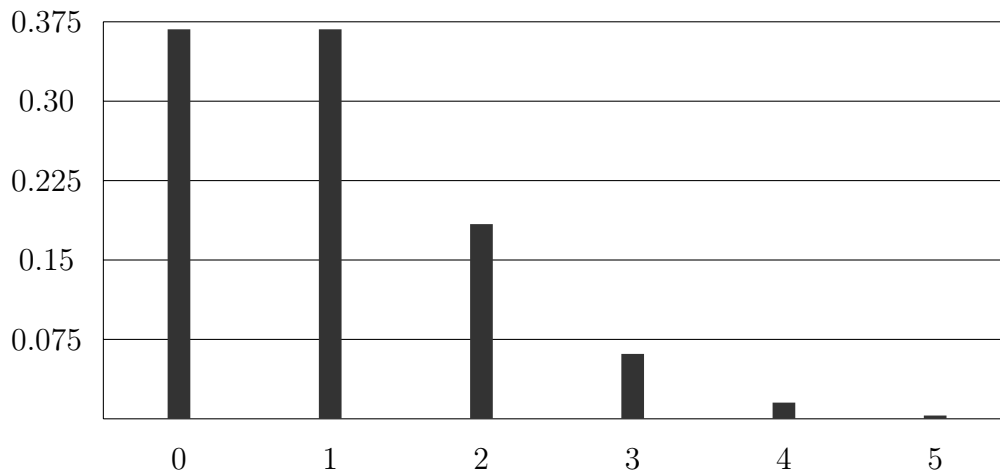
On trace le diagramme en bâtons de la loi (avec $\lambda = 8$) :

FIGURE 5.5 – Diagramme en bâtons de la loi de Poisson 1



On le trace également avec un autre paramètre (avec $\lambda = 1$) :

FIGURE 5.6 – Diagramme en bâtons de la loi de Poisson 2



5.4.3 Caractéristiques

Étudions maintenant ses caractéristiques principales. Soit X une variable aléatoire réelle suivant la loi de Poisson de paramètre $\lambda > 0$, $\mathcal{P}(\lambda)$. On a alors

$$\mathbb{E}[X] = \lambda \quad (5.4)$$

et

$$\text{Var}[X] = \lambda.$$

Démonstration. Démontrons d'abord (5.4) :

$$\begin{aligned} \mathbb{E}[X] &= \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} \sum_{n=1}^{\infty} n \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} \\ &= \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = \lambda. \end{aligned}$$

On calcule maintenant le moment d'ordre 2 de X :

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \sum_{n=1}^{\infty} n^2 \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \sum_{n=1}^{\infty} n \frac{\lambda^n}{(n-1)!} e^{-\lambda}. \end{aligned}$$

On utilise alors l'astuce $n = (n-1) + 1$:

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{n=1}^{\infty} [(n-1) + 1] \frac{\lambda^n}{(n-1)!} e^{-\lambda} \\ &= \sum_{n=1}^{\infty} (n-1) \frac{\lambda^n}{(n-1)!} e^{-\lambda} + \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} e^{-\lambda} \\ &= \sum_{n=2}^{\infty} (n-1) \frac{\lambda^n}{(n-1)!} e^{-\lambda} + \lambda \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \lambda^2 + \lambda. \end{aligned}$$

D'où $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda$. □

5.4.4 Somme de lois de Poisson

Dans le Théorème 5.3.9, on a établi qu'une somme de variables aléatoires réelles mutuellement indépendantes et suivant une loi binomiale de même paramètre p est une variable aléatoire réelle suivant une loi binomiale. Nous donnons ici un résultat similaire.

Théorème 5.4.5. *Soient n variables aléatoires réelles mutuellement indépendantes X_1, \dots, X_n telles que la variable aléatoire réelle X_i suit la loi de Poisson de paramètre $\lambda_i > 0$ pour tout $i \in \llbracket 1; n \rrbracket$. Alors, la variable aléatoire réelle*

$X_1 + \dots + X_n$ suit la loi de Poisson de paramètre $\lambda_1 + \dots + \lambda_n$. En d'autres termes :

$$\mathcal{L}(X_1 + \dots + X_n) = \mathcal{P}(\lambda_1 + \dots + \lambda_n).$$

Démonstration. On prouve le théorème pour $n = 2$. Le cas où n est quelconque s'obtient par récurrence.

D'abord, l'ensemble des réalisations possibles de $X_1 + X_2$ est \mathbb{N} . Calculons maintenant $\mathbb{P}(X_1 + X_2 = n)$ pour tout $n \in \mathbb{N}$:

$$\mathbb{P}(X_1 + X_2 = n) = \mathbb{P}\left(\bigcup_{k=0}^n \{X_1 = k\} \cap \{X_2 = n - k\}\right).$$

Les évènements $A_k := \{X_1 = k\} \cap \{X_2 = n - k\}$ sont incompatibles deux à deux. Conséquemment, on a

$$\mathbb{P}(X_1 + X_2 = n) = \sum_{k=0}^n \mathbb{P}(X_1 = k, X_2 = n - k).$$

Comme les variables aléatoires réelles sont mutuellement indépendantes, il vient

$$\begin{aligned} \mathbb{P}(X_1 + X_2 = n) &= \sum_{k=0}^n \mathbb{P}(X_1 = k) \mathbb{P}(X_2 = n - k) \\ &= \sum_{k=0}^n \frac{\lambda_1^k}{k!} \frac{\lambda_2^{n-k}}{(n-k)!} e^{-\lambda_1 - \lambda_2} \\ &= e^{-\lambda_1 - \lambda_2} \frac{1}{n!} \underbrace{\sum_{k=0}^n \binom{n}{k} \lambda_1^k \lambda_2^{n-k}}_{=(\lambda_1 + \lambda_2)^n} \\ &= \frac{(\lambda_1 + \lambda_2)^n}{n!} e^{-\lambda_1 - \lambda_2}, \end{aligned}$$

ce qui achève la preuve. □

Remarque 5.4.6. En particulier, on peut noter : $\mathbb{E}[X_1 + \dots + X_n] = \sum_{i=1}^n \lambda_i$ d'après la linéarité de l'espérance.

Exercice 5.4.7. En utilisant les distributions et le produit de convolution, donner une autre preuve du Théorème 5.4.5.

5.4.5 Tables de la loi de Poisson

5.4.5.0.1 Pour λ petit. Les valeurs $p_k := \mathbb{P}(X = k)$ et $\sum_{i=0}^k p_i$ sont tabulées pour $\lambda \in \{0.1, \dots, 0.9, 1, \dots, 16\}$.

Voir page 361.

5.4.5.1 Approximation de la loi de Poisson pour n grand

Ce paragraphe prend tout son sens à condition que la section sur la convergence en loi, voir page 209 ait été lue avant.

En pratique, pour $\lambda > 16$, on approche la loi de Poisson de paramètre λ , $\mathcal{P}(\lambda)$, par la loi normale $\mathcal{N}(\lambda, \lambda)$, voir la Section 9.4 à la page 138.

Soit X une variable aléatoire réelle qui suit la loi $\mathcal{P}(\lambda)$. Alors, $\mathbb{P}(X = n) \approx \xi(n + \frac{1}{2}) - \xi(n - \frac{1}{2})$ où ξ est la fonction de répartition de la loi $\mathcal{N}(\lambda, \lambda)$.

La justification théorique de cette approximation provient du Théorème central de la limite, voir la Section 15.4 à la page 217.

Autres lois discrètes classiques

6.1 Introduction

Dans ce chapitre, nous présentons d'autres lois de probabilité discrètes.

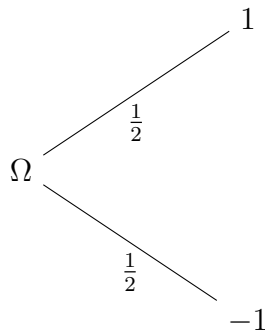
À nouveau, convolution et distribution peuvent aider à comprendre. Néanmoins, elles ne sont pas nécessaires.

L'objectif principal du chapitre n'est pas que le lecteur connaisse toutes ces lois mais qu'il les ait vues, qu'il les ait étudiées un peu. En effet, certaines seront mentionnées par la suite. Il est notamment attendu de bien comprendre qu'il existe une très grande variété de lois discrètes. Certaines sont théoriques et d'autres sont empiriques. Certaines dérivent de la loi binomiale. Également, l'une d'entre elles joue un rôle prépondérant en théorie probabiliste des nombres : la loi zêta. Plus qu'une compréhension fine de ces lois, nous espérons donner une culture de celles-ci au lecteur. Notamment, savoir que la loi géométrique est sans mémoire ne nous semble pas anodin.

6.2 Loi de Rademacher

La loi de Rademacher est une légère modification de la loi de Bernoulli de paramètre $\frac{1}{2}$. Ici, l'échec est modélisé par un -1 :

FIGURE 6.1 – Arbre de probabilité pour la loi de Rademacher



En d'autres termes, si X suit la loi de Rademacher, on a $\mathbb{P}_X = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. Ceci nous donne le tableau suivant :

TABLE 6.1 – Loi de probabilité de la variable aléatoire réelle X

k	-1	1
$\mathbb{P}(X = k)$	$\frac{1}{2}$	$\frac{1}{2}$

On étudie maintenant les caractéristiques de la loi X . Le calcul des moments d'une variable aléatoire réelle suivant la loi de Rademacher donne

$$\mathbb{E}[X^{2p}] = 1 \quad \text{et} \quad \mathbb{E}[X^{2p+1}] = 0 \quad (6.1)$$

pour tout $p \in \mathbb{N}$ et l'on a également

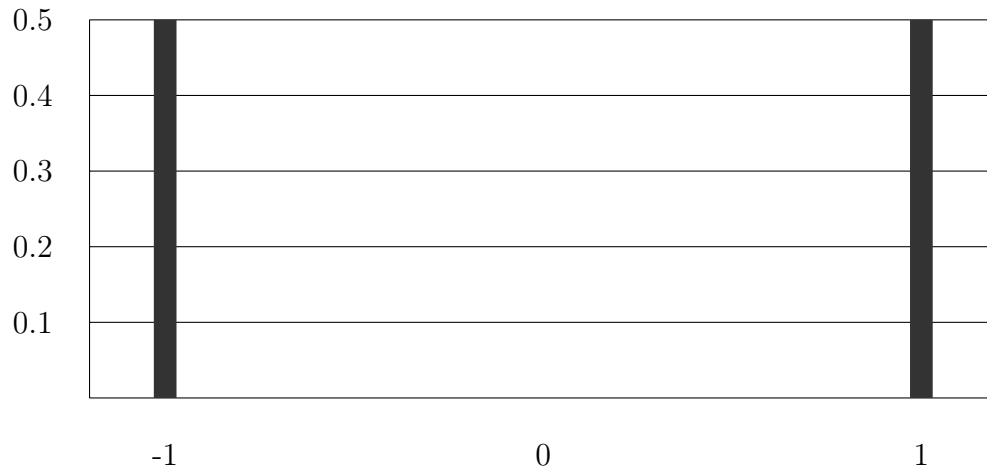
$$\text{Var}[X] = 1. \quad (6.2)$$

Exercice 6.2.1. *Démontrer les Égalités (6.1) et (6.2).*

Notons que la loi de Rademacher est particulièrement bien adaptée au jeu du pile ou face.

On trace le diagramme en bâtons de la loi :

FIGURE 6.2 – Diagramme en bâtons de la loi de Rademacher



6.3 Loi uniforme discrète

La loi uniforme discrète apparaît naturellement lorsque l'on joue avec un dé à six faces (un unique dé non pipé). Il s'agit de l'équiprobabilité sur un nombre n

de valeurs : $\alpha_1 < \dots < \alpha_n$. En d'autres termes, si X suit une loi uniforme discrète sur l'ensemble $\{\alpha_1, \dots, \alpha_n\}$, on a $\mathbb{P}_X = \frac{1}{n}\delta_{\alpha_1} + \dots + \frac{1}{n}\delta_{\alpha_n}$.

En particulier, si $n = 6$ et si $\alpha_i = i$, on retrouve le résultat du lancer de dé à six faces :

$$\mathbb{P}_X = \frac{1}{6} (\delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_6) .$$

Notons qu'une loi uniforme discrète sur l'ensemble $\{\frac{k}{N} : k \in \llbracket 0; N \rrbracket\}$ converge en loi vers la loi uniforme continue sur $[0; 1]$ (voir la page 132) quand N tend vers l'infini.

On peut par ailleurs montrer facilement que l'espérance d'une variable aléatoire X suivant la loi uniforme discrète sur l'ensemble $\llbracket 1; n \rrbracket$ est

$$\mathbb{E}[X] = \frac{1}{n} \sum_{k=1}^n k = \frac{1}{n} \times \frac{n(n+1)}{2} = \frac{n+1}{2} .$$

Et, la variance de X est

$$\begin{aligned} \text{Var}[X] &= \frac{1}{n} \sum_{k=1}^n k^2 - \left(\frac{n+1}{2}\right)^2 \\ &= \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\ &= \frac{n+1}{12} (4n+2-3n-3) \\ &= \frac{n^2-1}{12} . \end{aligned}$$

6.4 Loi triangulaire discrète

La loi triangulaire discrète apparaît naturellement lorsque l'on joue avec deux dés à six faces non pipés et que l'on regarde la somme des faces obtenues sur les deux dés.

Plus généralement, on se donne deux variables aléatoires indépendantes X et Y qui suivent la même loi uniforme discrète sur l'ensemble $\{1, 2, \dots, n\}$ (avec $n = 6$, on retrouve l'exemple du dé).

Alors, on dit que $Z := X + Y$ suit la loi triangulaire discrète sur $\{2, 3, \dots, 2n\}$.

Comme son nom l'indique, le diagramme en bâtons de la loi est triangulaire.

La loi a un unique mode (maximum local) : $n + 1$ et l'on a

$$\begin{aligned}\mathbb{P}_Z &:= \frac{1}{n^2}\delta_2 + \frac{2}{n^2}\delta_3 + \cdots + \frac{n-1}{n^2}\delta_n \\ &\quad + \frac{n}{n^2}\delta_{n+1} \\ &\quad + \frac{n-1}{n^2}\delta_{n+2} + \cdots + \frac{2}{n^2}\delta_{2n-1} + \frac{1}{n^2}\delta_{2n}.\end{aligned}$$

On peut vérifier que \mathbb{P}_Z est bien une loi de probabilité. En effet, chacun des coefficients est positif strictement. De plus, la somme de ceux-ci est égale à :

$$\begin{aligned}\frac{1}{n^2} \left(\sum_{k=2}^n (k-1) + n + \sum_{k=n+2}^{2n} (2n+1-k) \right) &= \frac{1}{n^2} \left(\sum_{k=1}^{n-1} k + n + \sum_{k=1}^{n-1} k \right) \\ &= \frac{1}{n^2} \left(2 \sum_{k=1}^{n-1} k + n \right) = \frac{1}{n^2} (n(n-1) + n) = \frac{1}{n^2} \times n^2 = 1.\end{aligned}$$

On peut calculer son espérance

$$\mathbb{E}[Z] = \mathbb{E}[X] + \mathbb{E}[Y] = n + 1,$$

ainsi que sa variance

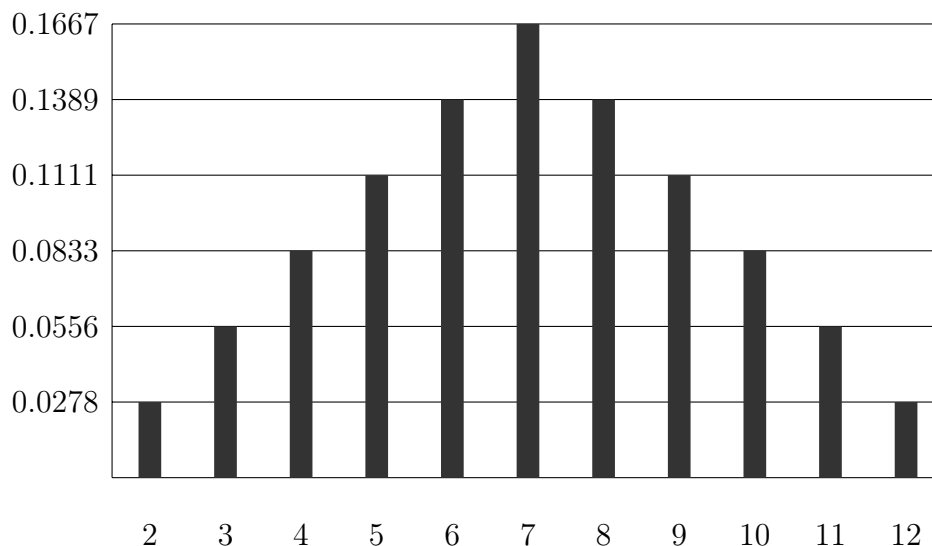
$$\text{Var}[Z] = \text{Var}[X] + \text{Var}[Y] = \frac{n^2 - 1}{6}.$$

Vérifions que la loi triangulaire discrète est bien celle que l'on a annoncée lorsque $n = 6$ en utilisant le produit de convolution :

$$\begin{aligned}\mathbb{P}_Z &= \mathbb{P}_X * \mathbb{P}_Y \\ &= \frac{1}{36} (\delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_6) * (\delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_6) \\ &= \frac{1}{36} \left\{ \delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_6 + \delta_7 \right. \\ &\quad + \delta_3 + \delta_4 + \delta_5 + \delta_6 + \delta_7 + \delta_8 \\ &\quad + \delta_4 + \delta_5 + \delta_6 + \delta_7 + \delta_8 + \delta_9 \\ &\quad + \delta_5 + \delta_6 + \delta_7 + \delta_8 + \delta_9 + \delta_{10} \\ &\quad + \delta_6 + \delta_7 + \delta_8 + \delta_9 + \delta_{10} + \delta_{11} \\ &\quad \left. + \delta_7 + \delta_8 + \delta_9 + \delta_{10} + \delta_{11} + \delta_{12} \right\} \\ &= \frac{1}{36} (\delta_2 + 2\delta_3 + 3\delta_4 + 4\delta_5 + 5\delta_6 + 6\delta_7 + 5\delta_8 + 4\delta_9 + 3\delta_{10} + 2\delta_{11} + \delta_{12}).\end{aligned}$$

On trace le diagramme en bâtons de la loi avec support égal à $\llbracket 2; 12 \rrbracket$, ce qui correspond à la somme de deux dés à six faces non pipés :

FIGURE 6.3 – Diagramme en bâtons de la loi triangulaire discrète



Dans ce cas de figure, l'espérance est simplement 7 et la variance est $\frac{35}{6}$.

Notons que l'on a ici choisi de présenter la loi triangulaire discrète dans un cas simple mais on pourrait considérer un cas plus général où le support de la loi uniforme discrète sous-jacente ne serait pas de la forme $\llbracket 1; n \rrbracket$.

6.5 Loi géométrique

On se donne $p \in]0; 1[$ et on considère une suite de variables aléatoires indépendantes et identiquement distribuées suivant la loi $\mathcal{B}(p) : (X_n)_{n \geq 1}$. On se demande combien d'essais sont nécessaires pour obtenir 1.

En d'autres termes, on introduit la variable aléatoire

$$N(\omega) := \inf \{n \geq 1 : X_n(\omega) = 1\}.$$

On dit alors que N suit la loi géométrique de paramètre p , $\mathcal{G}(p)$.

Calculons maintenant la loi de probabilité de N .

D'abord, on remarque $\mathbb{P}(N = 1) = \mathbb{P}(X_1 = 1) = p$.

Ensuite, avec un raisonnement similaire, il vient $\mathbb{P}(N = 2) = \mathbb{P}(X_1 = 0, X_2 = 1) = \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 1) = (1 - p) \times p = p(1 - p)$.

De manière générale, pour $n \geq 2$, on a

$$\begin{aligned}\mathbb{P}(N = n) &= \mathbb{P}(X_1 = 0, \dots, X_{n-1} = 0, X_n = 1) \\ &= \left(\prod_{i=1}^{n-1} \mathbb{P}(X_i = 0) \right) \times \mathbb{P}(X_n = 1) \\ &= (1-p)^{n-1} \times p \\ &= p(1-p)^{n-1}.\end{aligned}$$

Il vient ainsi $\mathbb{P}_N = \sum_{n=1}^{\infty} p(1-p)^{n-1} \delta_n$.

On s'intéresse par ailleurs à la fonction de survie de la loi N :

$$\mathbb{P}(N > n) = \sum_{k=n+1}^{\infty} p(1-p)^{k-1} = p(1-p)^n \frac{1}{1-(1-p)} = (1-p)^n.$$

On peut facilement calculer le premier moment de la loi géométrique comme suit :

$$\begin{aligned}\mathbb{E}[N] &= \sum_{n=1}^{\infty} p(1-p)^{n-1} n \\ &= p \sum_{n=1}^{\infty} n(1-p)^{n-1} \\ &= p\psi'(1-p),\end{aligned}$$

où $\psi(x) := \sum_{n=1}^{\infty} x^n = \frac{x}{1-x} = \frac{1}{1-x} - 1$ d'où $\psi'(x) = \frac{1}{(1-x)^2}$. Il vient $\psi'(1-p) = \frac{1}{p^2}$.
Conséquemment :

$$\mathbb{E}[N] = \frac{1}{p}.$$

On calcule maintenant son deuxième moment :

$$\begin{aligned}\mathbb{E}[N^2] &= \sum_{n=1}^{\infty} p(1-p)^{n-1} n^2 \\ &= p \sum_{n=1}^{\infty} n^2 (1-p)^{n-1} \\ &= p\xi'(1-p),\end{aligned}$$

où $\xi(x) := \sum_{n=1}^{\infty} nx^n = x\psi'(x) = \frac{x}{(1-x)^2}$ d'où $\xi'(x) = \frac{1+x}{(1-x)^3}$. Il vient $\xi'(1-p) = \frac{2-p}{p^3}$. Ainsi, $\mathbb{E}[N^2] = \frac{2-p}{p^2}$. Conséquemment :

$$\text{Var}[N] = \frac{1-p}{p^2}.$$

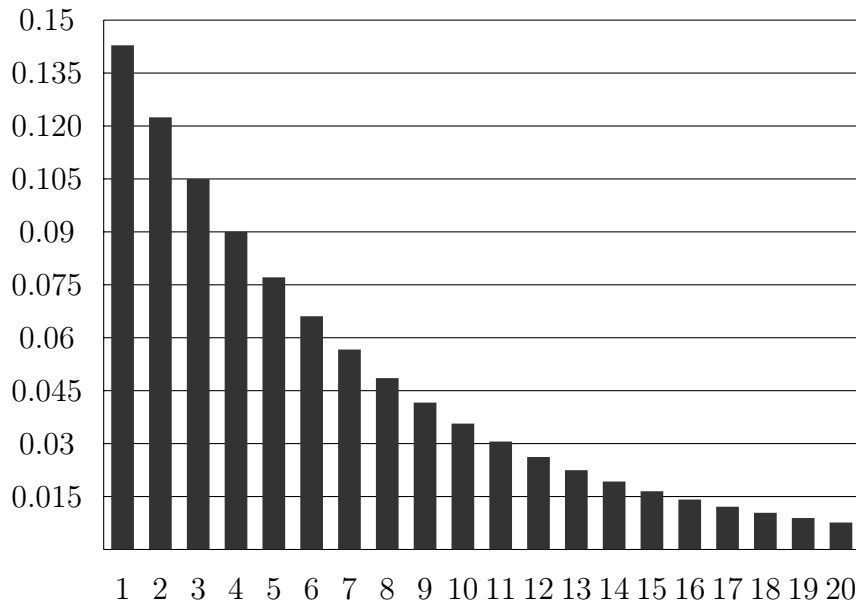
Notons que la loi géométrique peut être utilisée dans la modélisation du premier instant de désintégration d'une particule radioactive (dans le cas discret).

Ce qui justifie son utilisation pour de tels phénomènes est qu'il s'agit d'une loi sans mémoire discrète (là où la loi exponentielle est continue et sans mémoire) vu que pour tous les entiers strictement positifs m et n , on a $\mathbb{P}(N > m+n \mid N > n) = \mathbb{P}(N > m)$. En effet,

$$\begin{aligned} \mathbb{P}(N > m+n \mid N > n) &= \frac{\mathbb{P}(N > m+n, N > n)}{\mathbb{P}(N > n)} \\ &= \frac{\mathbb{P}(N > m+n)}{\mathbb{P}(N > n)} \\ &= \frac{(1-p)^{m+n}}{(1-p)^n} \\ &= (1-p)^m \\ &= \mathbb{P}(N > m). \end{aligned}$$

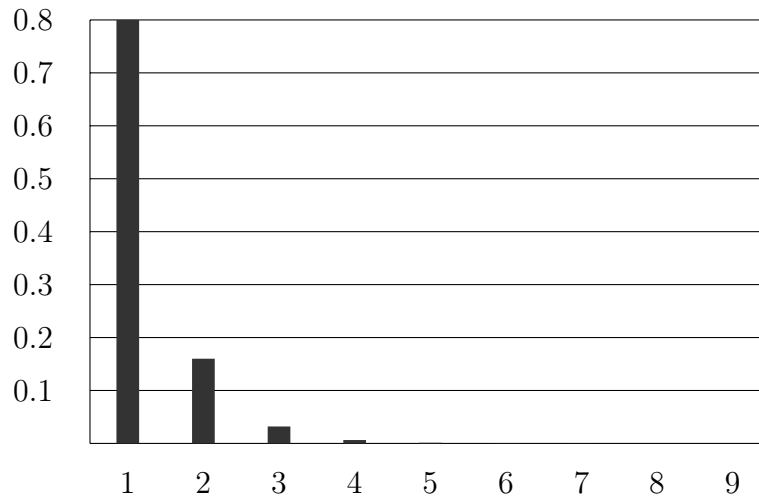
On trace le diagramme en bâtons de la loi géométrique avec $p = \frac{1}{7}$:

FIGURE 6.4 – Diagramme en bâtons de la loi géométrique 1



On le trace maintenant avec $p = 0.8$:

FIGURE 6.5 – Diagramme en bâtons de la loi géométrique 2



6.6 Loi hypergéométrique

Imaginons l'expérience suivante. On considère N individus d'une population. On suppose que n individus de cette population font plus de 100 kilogrammes. On suppose que $N - n$ individus font moins de 100 kilogrammes.

On tire ensuite k individus de la population et l'on se demande combien parmi ces k individus ont l'honneur *et le plaisir* de dépasser les 100 kilogrammes. Il s'agit ainsi d'une loi à support dans $\llbracket 0; \min(k; n) \rrbracket$.

De manière générale, on a une population de N individus composée de deux types d'individus : n vérifiant une propriété P et $N - n$ ne la vérifiant pas. On prend k individus au hasard parmi cette population et l'on se demande combien d'individus tirés vérifient la propriété P.

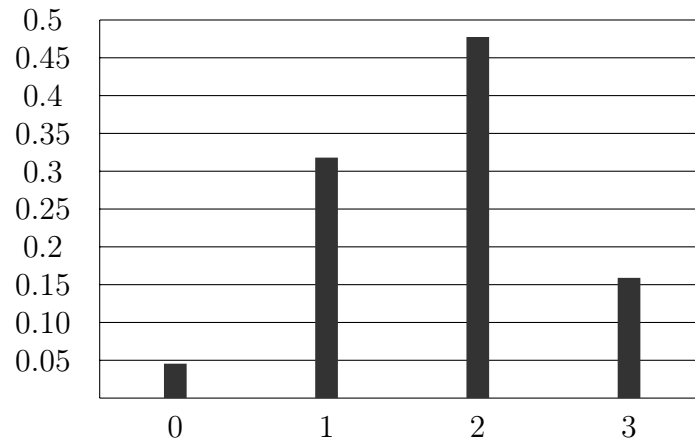
On peut calculer explicitement la loi de la variable aléatoire X :

$$\mathbb{P}_X = \sum_{i=0}^{\min(k;n)} \frac{\binom{n}{i} \binom{N-n}{k-i}}{\binom{N}{k}} \delta_i.$$

Il suffit pour s'en convaincre de dénombrer. Le nombre de cas possibles est $\binom{N}{k}$. Et, dire que X est égale à i est équivalent à dire que l'on a tiré i individus vérifiant la propriété P parmi les n la vérifiant au total puis que l'on a tiré $k - i$ individus qui ne la vérifient pas parmi les $N - n$ ne la vérifiant pas au total.

On trace le diagramme en bâtons de la loi hypergéométrique avec $N = 12$, $n = 3$ et $k = 7$:

FIGURE 6.6 – Diagramme en bâtons de la loi hypergéométrique



6.7 Loi de Zipf

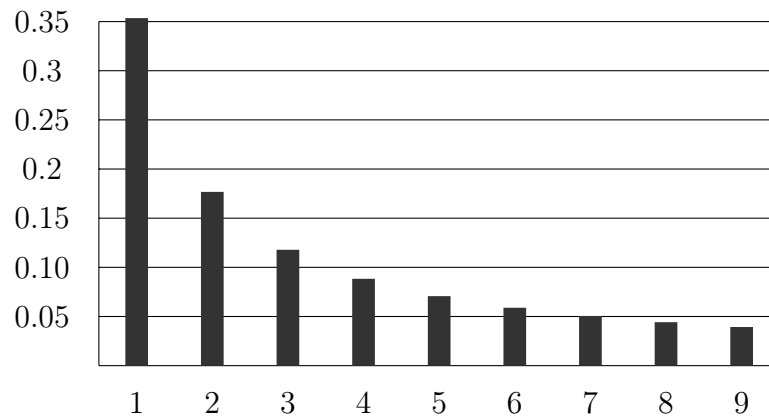
La loi de Zipf correspond à une restriction à un support fini de la loi zêta (voir Section 6.10 à la page 105). On dit que X suit la loi de Zipf de paramètres N et $s > 0$ si l'on a

$$\mathbb{P}(X = n) = \frac{n^{-s}}{\sum_{k=1}^N k^{-s}},$$

si $n \in \llbracket 1; N \rrbracket$ et $\mathbb{P}(X = n) = 0$ sinon. Cette loi de probabilité a été obtenue empiriquement par Zipf dans son étude des mots les plus courants dans un livre.

On trace le diagramme en bâtons de la loi de Zipf avec $N = 9$ et $s = 1$:

FIGURE 6.7 – Diagramme en bâtons de la loi de Zipf



6.8 Loi de Benford

La loi de Benford est également une loi empirique qui donne la distribution dans la vie de tous les jours des premiers chiffres des nombres. Contrairement à ce que l'intuition pourrait nous suggérer, les chiffres que l'on rencontre en début de nombres ne sont pas distribués suivant la loi uniforme discrète sur $\llbracket 1; 9 \rrbracket$.

En fait, si X suit la loi de Benford, on a :

$$\mathbb{P}(X = n) = \log_{10} \left(1 + \frac{1}{n} \right),$$

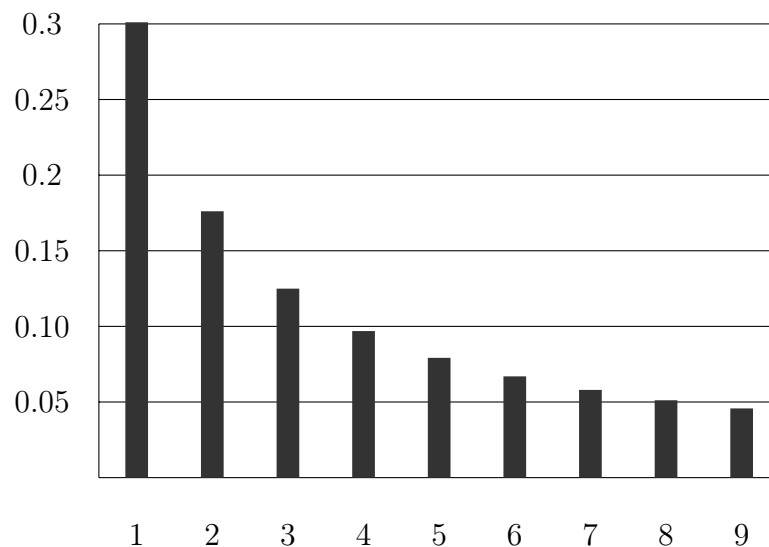
pour tout $n \in \llbracket 1; 9 \rrbracket$.

Cette loi est utilisée notamment pour la détection de fraudes fiscales.

Notons aussi qu'elle apparait d'un point de vue théorique (et non pas seulement empirique) lorsque l'on regarde les premiers chiffres des puissances de deux (ceci se démontre en utilisant les séries de Fourier).

On trace le diagramme en bâtons de la loi de Benford :

FIGURE 6.8 – Diagramme en bâtons de la loi de Benford



6.9 Loi logarithmique

La loi logarithmique découle du développement en série entière de la fonction $x \mapsto -\log(1-x)$. En effet, pour tout $p \in]0; 1[$, on a :

$$-\log(1-p) = \sum_{n=1}^{\infty} \frac{p^n}{n}.$$

Il s'ensuit :

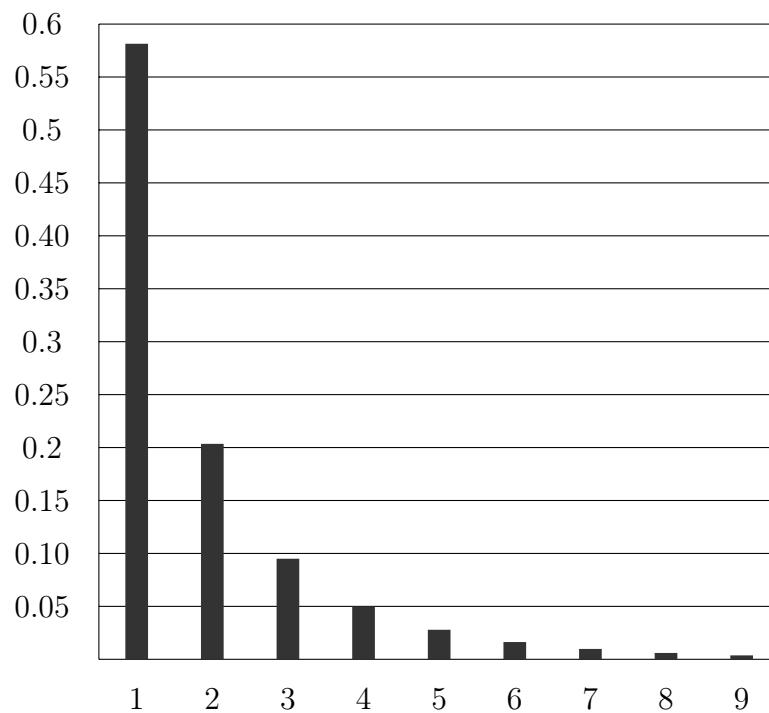
$$\sum_{n=1}^{\infty} \frac{1}{-\log(1-p)} \frac{p^n}{n} = 1,$$

ce qui - vu que chacun des termes est positif - nous donne une probabilité immédiatement :

$$\sum_{n=1}^{\infty} \frac{p^n}{n} \frac{1}{-\log(1-p)} \delta_n.$$

On trace le diagramme en bâtons de la loi logarithmique avec $p = 0.7$:

FIGURE 6.9 – Diagramme en bâtons de la loi logarithmique



6.10 Loi zêta

Rappelons que la fonction zêta de Riemann est définie comme suit (pour $s > 1$) :

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Contrairement à la loi de Zipf, le support est infini et pour des raisons de convergence, on doit imposer $s > 1$.

On dit alors que X suit la loi zêta de paramètre $s > 1$ si l'on a

$$\mathbb{P}_X = \sum_{n=1}^{\infty} \frac{n^{-s}}{\zeta(s)} \delta_n.$$

Cette loi a une grande importance en théorie probabiliste des nombres. Parlons-en un petit peu (nous n'exigeons pas du lecteur qu'il lise la suite de cette section mais nous l'y encourageons).

On introduit l'évènement $E_p := \{X \text{ est divisible par } p\}$ pour tout p premier.

Proposition 6.10.1. *Les évènements $(E_p)_p$ où p est premier sont mutuellement indépendants.*

Démonstration. Soient p_1, \dots, p_n nombres premiers. On va montrer que

$$\mathbb{P}(E_{p_1} \cap \dots \cap E_{p_n}) = \mathbb{P}(E_{p_1}) \times \dots \times \mathbb{P}(E_{p_n}).$$

Le calcul nous donne $\mathbb{P}(E_p) = p^{-s}$ pour tout p premier. En effet :

$$\mathbb{P}(E_p) = \sum_{\substack{n \in \mathbb{N}^* \\ p|n}} \frac{n^{-s}}{\zeta(s)} = \sum_{k=1}^{\infty} \frac{p^{-s} k^{-s}}{\zeta(s)} = p^{-s}.$$

Or, on sait qu'un nombre n'est divisible par p_1 , par p_2 , ..., par p_n que s'il est divisible par $p_1 \times \dots \times p_n$. On a donc

$$\mathbb{P}(E_{p_1} \cap \dots \cap E_{p_n}) = \sum_{k=1}^{\infty} \frac{(p_1 \times \dots \times p_n)^{-s} k^{-s}}{\zeta(s)} = p_1^{-s} \times \dots \times p_n^{-s},$$

ce qui achève la preuve. □

Proposition 6.10.2. *On dispose de la formule d'Euler :*

$$\frac{1}{\zeta(s)} = \prod_p \left(1 - \frac{1}{p^s}\right),$$

où le produit se fait sur les nombres premiers.

Démonstration. On calcule $\mathbb{P}(X = 1)$ de deux façons différentes. D'abord :

$$\mathbb{P}(X = 1) = \frac{1}{\zeta(s)}.$$

Ensuite, $X = 1$ si et seulement si X n'est divisible par aucun nombre premier. En d'autres termes, on a

$$\mathbb{P}(X = 1) = \mathbb{P}\left(\bigcap_p E_p^c\right) = \prod_p \mathbb{P}(E_p^c) = \prod_p \left(1 - \frac{1}{p^s}\right),$$

ce qui achève la preuve. □

Proposition 6.10.3. *On a : $\mathbb{P}(X \text{ n'a pas de facteurs carrés}) = \frac{1}{\zeta(2s)}$.*

Démonstration. On note E_{p^2} l'évènement $E_{p^2} := \{p^2 \text{ divise } X\}$. On peut prouver facilement que l'on a $\mathbb{P}(E_{p^2}) = p^{-2s}$ et que les évènements E_{p^2} sont mutuellement indépendants. Il vient

$$\mathbb{P}(X \text{ n'a pas de facteurs carrés}) = \prod_p \left(1 - \frac{1}{p^{2s}}\right) = \frac{1}{\zeta(2s)},$$

d'après la Proposition 6.10.2. □

Fonctions génératrices

7.1 Introduction

La fonction caractéristique d'une loi correspond à la transformée de Fourier de celle-ci. En particulier, la fonction caractéristique caractérise la loi. De même, la fonction génératrice caractérise la loi d'une variable aléatoire à valeurs dans \mathbb{N} . Les fonctions génératrices sont un outil de calcul commode. Elles se prêtent bien au calcul de loi de sommes de variables aléatoires indépendantes, ainsi qu'à l'étude de la convergence en loi. En particulier, on verra qu'elle joue un rôle non négligeable dans le problème de l'extinction des grands noms.

Pour attaquer ce chapitre, il est important que le lecteur ait des bases solides concernant la notion de séries entières ou tout du moins sur les séries de fonctions. Dit autrement, un lecteur qui n'aurait jamais vu la notion de série entière ou au moins celle de série de fonctions serait fortement pénalisé dans la compréhension des preuves et des théorèmes. De la même manière, savoir ce qu'est la dérivée d'une fonction et bien comprendre la notion de limite sont des pré-requis indispensables.

Comprendre que la fonction génératrice (aussi appelée fonction génératrice des moments) génère les moments (dans un sens à préciser subséquentement) est très important. Dit autrement, si l'on connaît la fonction génératrice d'une loi, on en connaît automatiquement les moments. Il faudra également que le lecteur arrive à jongler avec les différentes propriétés : fonction génératrice de la somme, dérivées successives de la fonction génératrice, fonction génératrice d'une somme aléatoire de variables aléatoires... Le chapitre sera conclu par l'étude d'un phénomène qui se résout bien dès lors que l'on utilise les fonctions génératrices : celui de l'extinction des grands noms.

7.2 Définition et premières propriétés

Dans toute cette section, X est une variable aléatoire réelle à valeurs dans \mathbb{N} .

Définition 7.2.1. *La fonction génératrice de la variable aléatoire X est définie comme étant la fonction de la variable réelle :*

$$G_X(s) := \mathbb{E}(s^X) .$$

Proposition 7.2.2. *Le domaine de définition de G_X contient l'intervalle $[-1; 1]$. De plus, on a $G_X(1) = 1$ et pour tout $s \in [-1; 1]$:*

$$|G_X(s)| \leq 1.$$

Démonstration. Pour tout $s \in [-1; 1]$, comme X est à valeurs dans \mathbb{N} , pour tout ω , on a : $|s^{X(\omega)}| \leq 1$. Le théorème de convergence dominée de Lebesgue implique que s^X est intégrable, c'est-à-dire que $G_X(s)$ existe. Pour établir l'inégalité, on procède ainsi :

$$|G_X(s)| = |\mathbb{E}(s^X)| \leq \mathbb{E}(|s^X|) \leq \mathbb{E}(1) = 1.$$

Enfin, si $s = 1$, $s^{X(\omega)} = 1$ pour tout $\omega \in \Omega$ d'où $\mathbb{E}[s^X] = \mathbb{E}[1] = 1$. □

Proposition 7.2.3. *Pour tout $s \in [-1; 1]$, on a :*

$$G_X(s) = \sum_{k=0}^{\infty} p_k s^k,$$

où $p_k := \mathbb{P}(X = k)$ et où on adopte la convention $0^0 = 1$.

Démonstration. On utilise pour cela la formule de transfert. Il faut néanmoins vérifier que l'on a bien $\sum_{k=0}^{\infty} p_k |s^k| < \infty$. Or, on remarque :

$$\sum_{k=0}^{\infty} p_k |s^k| = p_0 + \sum_{k=1}^{\infty} p_k |s|^k \leq p_0 + \sum_{k=1}^{\infty} p_k = 1,$$

ce qui achève la preuve. □

Proposition 7.2.4. *La fonction génératrice G_X est continue sur $[-1; 1]$ et de classe \mathcal{C}^∞ sur $] -1; 1[$.*

Démonstration. Il suffit de remarquer que G_X est une série entière de rayon de convergence supérieur ou égal à 1. En effet, par définition, $\sum_{k=0}^{\infty} p_k = 1 < \infty$. □

Proposition 7.2.5. *La fonction génératrice G_X restreinte à l'intervalle $[-1; 1]$ caractérise la loi de X . En effet, pour tout $n \in \mathbb{N}$, on a :*

$$\mathbb{P}(X = n) = \frac{G_X^{(n)}(0)}{n!}.$$

Exercice 7.2.6. *Démontrer la Proposition 7.2.5.*

7.3 Fonctions génératrices des lois classiques

7.3.1 Si $\mathbb{P}_X = \mathcal{B}(p)$

Si X suit une loi de Bernoulli de paramètre p , on a :

$$G_X(s) = ps + (1 - p). \tag{7.1}$$

Exercice 7.3.1. *Démontrer la Formule (7.1).*

7.3.2 Si $\mathbb{P}_X = \mathcal{B}(n, p)$

Si X suit une loi binomiale de paramètres n et p , on a :

$$G_X(s) = (ps + (1 - p))^n . \quad (7.2)$$

En effet :

$$\begin{aligned} G_X(s) &= \sum_{k=0}^{\infty} \mathbb{P}(X = k) s^k \\ &= \sum_{k=0}^n \binom{n}{k} (ps)^k (1 - p)^{n-k} \\ &= (ps + (1 - p))^n . \end{aligned}$$

On peut observer que la fonction génératrice de la loi binomiale de paramètres n et p est la puissance n -ième de la fonction génératrice de la loi de Bernoulli de paramètre p .

7.3.3 Si $\mathbb{P}_X = \mathcal{P}(\lambda)$

Si X suit la loi de Poisson de paramètre $\lambda > 0$, on a :

$$G_X(s) = \exp \{ \lambda(s - 1) \} . \quad (7.3)$$

En effet :

$$\begin{aligned} G_X(s) &= \sum_{k=0}^{\infty} \mathbb{P}(X = k) s^k \\ &= \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} s^k \\ &= \exp \{ -\lambda \} \exp \{ \lambda s \} \\ &= \exp \{ \lambda(s - 1) \} . \end{aligned}$$

7.3.4 Pour d'autres lois discrètes usuelles

Dans ce paragraphe, nous listons les fonctions génératrices des autres lois usuelles présentes dans le Chapitre 6.

- Si X suit la loi de Rademacher, alors comme $X(\Omega)$ n'est pas inclus dans \mathbb{N} , on ne considère pas sa fonction génératrice.
- Si X suit la loi uniforme discrète sur $\llbracket 1; N \rrbracket$, alors $G_X(s) = \frac{s}{N} \frac{1-s^N}{1-s}$ pour $s \neq 1$ et $G_X(1) = 1$.
- Si X suit la loi triangulaire discrète sur $\llbracket 2; 12 \rrbracket$, alors $G_X(s) = \frac{s^2}{36} \left(\frac{1-s^6}{1-s} \right)^2$ pour $s \neq 1$ et $G_X(1) = 1$.

- Si X suit la loi géométrique de paramètre p , alors $G_X(s) = \frac{ps}{1-(1-p)s}$ pour tout $s \in [-1; 1]$.
- Si X suit la loi hypergéométrique de paramètres $N = 12$, $n = 3$ et $k = 7$, alors $G_X(s) = \frac{1}{44} (2 + 14s + 21s^2 + 7s^3)$ pour tout s .
- Si X suit la loi de Zipf de paramètres N et $\alpha > 0$, alors $G_X(s) = \frac{\sum_{k=1}^N n^{-\alpha} s^n}{\sum_{k=1}^N n^{-\alpha}}$ pour tout s .
- Si X suit la loi de Benford, alors $G_X(s) = \sum_{k=1}^9 s^k \log\left(\frac{k+1}{k}\right)$ pour tout s .
- Si X suit la loi logarithmique, alors $G_X(s) = \frac{\log(1-ps)}{\log(1-p)}$ pour tout $s \in [-1; 1]$.
- Si X suit la loi zêta de paramètre $\alpha > 1$, alors $G_X(s) = \frac{\sum_{k=1}^{\infty} n^{-\alpha} s^n}{\zeta(\alpha)}$ pour tout $s \in [-1; 1]$.

7.4 Résultats importants

Théorème 7.4.1. *Soient deux variables aléatoires réelles X et Y . On suppose que X et Y sont indépendantes. On a alors*

$$G_{X+Y}(s) = G_X(s)G_Y(s), \quad (7.4)$$

pour tout $s \in [-1; 1]$.

Démonstration. On applique la définition de $G_{X+Y}(s)$ comme suit

$$G_{X+Y}(s) = \mathbb{E} \{s^{X+Y}\} = \mathbb{E} \{s^X s^Y\}.$$

Comme X et Y sont indépendantes, s^X et s^Y le sont également et donc on obtient

$$\mathbb{E} \{s^X s^Y\} = \mathbb{E} [s^X] \times \mathbb{E} [s^Y] = G_X(s)G_Y(s).$$

□

La fonction génératrice d'une variable aléatoire déterminant sa loi, il est naturel qu'elle donne ses moments, lorsque ceux-ci existent.

Par la suite, nous noterons, pour $r \in \mathbb{N}^*$, $G_X^{(r)}(1^-)$ la r -ième dérivée à gauche de G_X en 1 lorsqu'elle existe. En d'autres termes, on pose

$$G_X^{(r)}(1^-) := \lim_{\substack{s \rightarrow 1 \\ s < 1}} G_X^{(r)}(s).$$

Proposition 7.4.2. *Pour que X admette un moment d'ordre $r \in \mathbb{N}^*$, il faut et il suffit que sa fonction génératrice G_X soit r fois dérivable à gauche en 1, et dans ce cas, on a :*

$$G_X^{(r)}(1^-) = \sum_{k=r}^{\infty} k(k-1) \cdots (k-r+1) \mathbb{P}(X = k), \quad (7.5)$$

ce qui s'écrit encore :

$$\mathbb{E}[X(X-1)\cdots(X-r+1)] = G_X^{(r)}(1^-). \quad (7.6)$$

En particulier, pour $r = 1$, il s'ensuit $\mathbb{E}(X) = G_X'(1^-)$.

L'idée sous-jacente consiste à appliquer le théorème de dérivation sous une somme finie (ou sous une somme infinie si $X(\Omega)$ est infini dénombrable).

En effet, intuitivement (cela mérite d'être vérifié rigoureusement cela dit) :

$$\frac{d}{ds}\mathbb{E}[s^X] = \mathbb{E}\left[\frac{d}{ds}s^X\right] = \mathbb{E}[Xs^{X-1}],$$

et donc en faisant tendre s vers 1 par valeurs inférieures, on retrouverait $\mathbb{E}(X)$.
De même :

$$\frac{d^2}{ds^2}\mathbb{E}[s^X] = \mathbb{E}[X(X-1)s^{X-2}],$$

et donc en faisant tendre s vers 1 par valeurs inférieures, on retrouverait $\mathbb{E}(X(X-1))$.

Parfois, dans les cas pratiques, nous pouvons rencontrer des situations où nous avons à étudier la loi d'une somme de variables aléatoires indépendantes de longueur aléatoire. Les fonctions génératrices peuvent être un précieux outil pour ce faire.

Théorème 7.4.3. *Soient $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes de même loi à valeurs dans \mathbb{N} . Soit T une variable aléatoire à valeurs dans \mathbb{N} . On suppose que les variables aléatoires T et X_n , $n \in \mathbb{N}^*$, sont indépendantes. On définit, pour tout $\omega \in \Omega$, la variable aléatoire*

$$S(\omega) := \sum_{j=1}^{T(\omega)} X_j(\omega).$$

Si G_T et G_{X_1} désignent respectivement les fonctions génératrices de T et de X_1 , la fonction génératrice de S est donnée par $G_S(s) = G_T(G_{X_1}(s))$.

Démonstration. Puisque T prend ses valeurs dans \mathbb{N} , on a, pour tout $s \in [-1; 1]$:

$$\begin{aligned} G_S(s) &= \sum_{k=0}^{\infty} s^k \mathbb{P}(S = k) \\ &= \sum_{k=0}^{\infty} s^k \mathbb{P}\left(\{S = k\} \cap \left(\bigcup_{n \in \mathbb{N}} \{T = n\}\right)\right) \\ &= \sum_{k=0}^{\infty} s^k \left[\sum_{n \in \mathbb{N}} \mathbb{P}(S = k, T = n) \right], \end{aligned}$$

après avoir utilisé le système complet d'évènements $(A_n)_{n \in \mathbb{N}}$ où $A_n := \{T = n\}$.

Mais, on a

$$\sum_{k=0}^{\infty} |s|^k \left[\sum_{n \in \mathbb{N}} \mathbb{P}(S = k, T = n) \right] \leq \sum_{k=0}^{\infty} \left[\sum_{n \in \mathbb{N}} \mathbb{P}(S = k, T = n) \right] = 1 < \infty.$$

La famille $\{s^k \mathbb{P}(S = k, T = n)\}_{(k,n) \in \mathbb{N} \times \mathbb{N}}$ est donc sommable et l'on peut utiliser la propriété de Fubini. On note que les évènements A_0 et $\{S > 0\}$ sont incompatibles. Ainsi, on peut écrire :

$$G_S(s) = \mathbb{P}(T = 0) + \sum_{n=1}^{\infty} \sum_{k=0}^{\infty} s^k \mathbb{P}(S = k, T = n).$$

On a ainsi en utilisant l'indépendance des variables aléatoires $\sum_{j=1}^n X_j$ et T :

$$\begin{aligned} G_S(s) &= \mathbb{P}(T = 0) + \sum_{n \in \mathbb{N}^*} \sum_{k=0}^{\infty} s^k \mathbb{P}\left(\sum_{j=0}^n X_j = k, T = n\right) \\ &= \mathbb{P}(T = 0) + \sum_{n \in \mathbb{N}^*} \sum_{k=0}^{\infty} s^k \mathbb{P}\left(\sum_{j=0}^n X_j = k\right) \mathbb{P}(T = n) \\ &= \mathbb{P}(T = 0) + \sum_{n \in \mathbb{N}^*} G_{X_1 + \dots + X_n}(s) \mathbb{P}(T = n) \\ &= (G_{X_1}(s))^0 \mathbb{P}(T = 0) + \sum_{n \in \mathbb{N}^*} (G_{X_1}(s))^n \mathbb{P}(T = n) \\ &= \sum_{n \in \mathbb{N}} (G_{X_1}(s))^n \mathbb{P}(T = n) \\ &= G_T(G_{X_1}(s)). \end{aligned}$$

C'est le résultat annoncé. □

Nous donnons dans ce même contexte des relations sur la moyenne et sur la variance de S .

Corollaire 7.4.4 (Identité de Wald). *Sous les mêmes hypothèses que le Théorème 7.4.3, on a les deux résultats suivants.*

1. Si X_1 et T admettent une moyenne, S admet une moyenne donnée par $\mathbb{E}(S) = \mathbb{E}(X_1)\mathbb{E}(T)$.
2. Si X_1 et T admettent un moment d'ordre deux, il en est de même pour S et sa variance est donnée par

$$\text{Var}[S] = \mathbb{E}(T)\text{Var}[X_1] + (\mathbb{E}(X_1))^2 \text{Var}[T].$$

Exercice 7.4.5. *Démontrer l'identité de Wald.*

7.5 Extinction des grands noms

Donnons un exemple de processus de branchement (ou processus de Galton-Watson).

On suppose qu'une population évolue à chaque génération $n = 0, 1, \dots$. On note Z_n le nombre d'individus de la population à la génération n . Chaque membre de la génération n donne naissance à un certain nombre d'enfants qui sont de fait membres de la $(n+1)$ -ème génération. La taille de la famille est donc une variable aléatoire.

On prend les hypothèses suivantes.

- Si on note $X_i^{(n)}$ le nombre d'enfants que l'individu i de la génération n a, alors les variables aléatoires $X_i^{(n)}$ sont mutuellement indépendantes pour $n \in \mathbb{N}$ et $1 \leq i \leq Z_n$.
- Les variables aléatoires $X_i^{(n)}$ suivent une loi commune. On note G la fonction génératrice de cette loi.
- $Z_0 = 1$.

On pose ici $G_n(s) := \mathbb{E}[s^{Z_n}]$.

On peut montrer facilement, avec le Théorème 7.4.3 que l'on a : $G_n(s) = G(G(\dots(G(s))))$. En effet, par récurrence, on obtient $G_n = G_{n-1} \circ G$.

En posant $\mu := \mathbb{E}[Z_1]$ et $\sigma^2 := \text{Var}[Z_1]$ puis $\mu_n := \mathbb{E}[Z_n]$ et $\sigma_n^2 := \text{Var}[Z_n]$ et en utilisant la Proposition 7.4.2, alors : $\mathbb{E}[Z_n] = \mu^n$ et

$$\sigma_n^2 = \frac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1},$$

si $\mu \neq 1$ et $\sigma_n^2 = n\sigma^2$ si $\mu = 1$.

En supposant ensuite que la taille des familles suit une loi *de type* géométrique avec $\mathbb{P}(X = k) = (1-p)p^k$ pour tout $k \in \mathbb{N}$ où $p \in]0; 1[$, alors $G(s) = \frac{1-p}{1-ps}$. Par récurrence, on aboutit à

$$G_n(s) = \frac{n - (n-1)s}{n+1 - ns} \text{ si } p = \frac{1}{2},$$

et

$$G_n(s) = \frac{(1-p)(p^n - (1-p)^n - ps(p^{n-1} - (1-p)^{n-1}))}{p^{n+1} - (1-p)^{n+1} - ps(p^n - (1-p)^n)} \text{ si } p \neq \frac{1}{2}.$$

On peut alors en déduire la probabilité d'extinction à la génération n :

$$\mathbb{P}[Z_n = 0] = \frac{n}{n+1} \text{ si } p = \frac{1}{2},$$

et

$$\mathbb{P}[Z_n = 0] = \frac{1-p}{p} \frac{1 - \left(\frac{1-p}{p}\right)^n}{1 - \left(\frac{1-p}{p}\right)^{n+1}} \text{ si } p \neq \frac{1}{2}.$$

On introduit maintenant l'évènement suivant :

$$\mathcal{E} := \{\omega \in \Omega : \exists n \in \mathbb{N}^* \text{ tel que } Z_n = 0\}.$$

Il s'ensuit :

$$\mathbb{P}(\mathcal{E}) = \lim_{n \rightarrow \infty} \mathbb{P}[Z_n = 0] = 1 \text{ si } p \leq \frac{1}{2},$$

et

$$\mathbb{P}(\mathcal{E}) = \lim_{n \rightarrow \infty} \mathbb{P}[Z_n = 0] = \frac{1-p}{p} \text{ si } p > \frac{1}{2}.$$

Variables aléatoires à densité

8.1 Introduction

L'autre grande famille des variables aléatoires réelles est celle des variables aléatoires réelles à densité. La différence avec les variables aléatoires discrètes est que les sommes (quand on considère l'espérance) deviennent des intégrales.

Il convient de noter que si parfois le cas à densité est plus complexe à étudier, il est parfois plus simple. Cela dépend également du goût de chacun.

Remarque 8.1.1. *Il existe des variables aléatoires réelles qui ne sont ni discrètes ni à densité. Ainsi, une variable aléatoire dont la fonction de répartition est l'escalier du diable sur $[0; 1]$ n'est pas discrète et bien qu'elle soit donc continue, elle n'est pas à densité.*

Les pré-requis sont l'intégration (de préférence celle de Lebesgue quoique celle de Riemann est amplement suffisante si l'on ne se destine pas à étudier les cas les plus tordus) et la dérivation.

Les objectifs du chapitre sont la découverte de la notion de densité de probabilité (à savoir la dérivée, quand elle existe, de la fonction de répartition) et de ses propriétés (positive et d'intégrale sur \mathbb{R} égale à un). La notion d'espérance que nous avons vue pour les variables aléatoires discrètes sera étendue aux variables aléatoires à densité. Nous verrons notamment qu'il existe des conditions pour que cette espérance existe et nous verrons les propriétés classiques de celle-ci dont la formule de transfert et l'inégalité de Markov.

Une définition alternative de la densité de probabilité sera fournie et il est cruciale de la connaître pour aborder sereinement le chapitre sur les vecteurs aléatoires à la page 183. De plus, nous verrons la notion de variance et l'inégalité de Bienaymé-Tchebychev. Enfin, nous verrons ce qu'il se produit quand on mélange le monde du discret et celui du continu (en faisant une somme ou un produit) et nous verrons que l'opération qui consiste à centrer puis à réduire ne diffère en rien dans l'un ou l'autre des deux mondes. Pour terminer, nous aborderons très brièvement d'autres quantités caractéristiques : le skewness et le kurtosis.

8.2 Densité : définition et propriétés

8.2.1 Définition

Définition 8.2.1. Soit une variable aléatoire réelle X . On suppose que sa fonction de répartition F_X est continue sur \mathbb{R} . On suppose aussi qu'elle est dérivable sauf sur un ensemble dénombrable de points. Alors, cette fonction dérivée, $f_X := F'_X$ est appelée la densité de probabilité de la variable aléatoire réelle X . Et, on dit que la variable aléatoire réelle X est à densité.

On peut écrire ici

$$\mathbb{P}_X = f_X dx .$$

Le dx sert ici à signaler que la variable est absolument continue par rapport à la mesure de Lebesgue.

Remarque 8.2.2 (Conséquence). Une variable aléatoire réelle discrète n'admet pas de densité car sa fonction de répartition n'est pas continue.

Remarque 8.2.3. Une variable aléatoire réelle discrète admet une densité au sens des distributions. Il convient donc de garder à l'esprit que les deux grands types de variables aléatoires réelles ne sont pas si différents.

Remarque 8.2.4 (Précision). La véritable définition d'une densité est un peu plus subtile que celle donnée en Définition 8.2.1. En effet, il faut en fait que la fonction F_X soit dérivable presque partout pour la mesure de Lebesgue. Toutefois, dans ce cours, on a décidé de ne pas parler de théorie de la mesure. Par ailleurs, les lois de probabilité continues usuelles (voir Chapitre 9, page 131 et Chapitre 10 à la page 147) satisfont la Définition 8.2.1. Le lecteur qui souhaite plus de rigueur est invité à consulter le Chapitre 24 à la page 341.

Définition 8.2.5. Une variable aléatoire réelle X qui admet une densité de probabilité est dite à densité.

Présentons un exemple simple de variable aléatoire réelle à densité.

Exemple 8.2.6 (Loi uniforme sur $[0; 1]$). On dit qu'une variable aléatoire réelle X suit la loi uniforme sur l'intervalle $[0; 1]$, $\mathcal{U}_{[0;1]}$, lorsqu'elle admet la fonction de répartition suivante :

$$F_X(t) = \mathbb{P}(X \leq t) = \begin{cases} 0 & \text{si } t \leq 0 \\ t & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } 1 \leq t \end{cases} .$$

On remarque que cette fonction est continue sur \mathbb{R} et qu'elle est dérivable sauf en $t = 0$ et en $t = 1$. Et, la dérivée de F_X est

$$f_X(t) = F'_X(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 & \text{si } 0 \leq t \leq 1 \\ 0 & \text{si } 1 < t \end{cases} .$$

Ainsi, une variable aléatoire réelle qui suit la loi $\mathcal{U}_{[0;1]}$ est à densité et sa densité est égale à $\mathbb{1}_{[0;1]}$.

La loi uniforme est étudiée plus en détails dans le Chapitre 9 à la page 131.

Notons que l'on peut donner une définition alternative de la densité de probabilité à partir de la proposition (admise) suivante :

Proposition 8.2.7 (Définition alternative). *Soit une variable aléatoire réelle X . On suppose qu'il existe une fonction f positive et d'intégrale sur \mathbb{R} égale à 1 telle que pour tout ensemble (borélien) E , on ait l'égalité suivante :*

$$\mathbb{P}(X \in E) = \int_E f(x) dx. \quad (8.1)$$

Alors, la variable aléatoire X est à densité et $f_X = f$.

Remarque 8.2.8. *Cette définition alternative est généralisée pour le cas des vecteurs aléatoires, voir la Définition 12.7.1 à la page 188.*

Remarque 8.2.9. *La tribu borélienne étant générée par les intervalles ouverts, il suffit que l'Égalité (8.1) soit vérifiée pour tout intervalle ouvert.*

8.2.1.1 Fonction de répartition et densité de probabilité

Par définition, on a $F'_X = f_X$ presque partout. Donc F_X est une primitive de f_X . Or, $F_X(-\infty) = 0$. Conséquemment, il vient

$$F_X(t) = \int_{-\infty}^t f_X(s) ds. \quad (8.2)$$

Exercice 8.2.10. *Montrer que pour tout $t \in \mathbb{R}$, l'Égalité (8.2) est vérifiée pour la loi uniforme présentée dans l'Exemple 8.2.6.*

8.2.2 Propriétés

Présentons maintenant les propriétés de la densité de probabilité d'une variable aléatoire réelle à densité.

Proposition 8.2.11. *Soit une variable aléatoire réelle X à densité. Soit f_X sa densité de probabilité. Alors, la fonction f_X est positive : $f_X(t) \geq 0$ pour tout $t \in \mathbb{R}$.*

Démonstration. En effet, la fonction de répartition F_X est croissante (voir Proposition 3.3.12, page 55). Donc sa dérivée est positive ou nulle. Or, la dérivée de F_X est égale à f_X presque partout. \square

Proposition 8.2.12. *Soit une variable aléatoire réelle X à densité. Soit f_X sa densité de probabilité. Alors, on a*

$$\int_{-\infty}^{+\infty} f_X(s) ds = 1.$$

Démonstration. En effet, on peut lier la fonction de répartition F_X et la densité de probabilité f_X comme suit d'après l'Égalité (8.2)

$$F_X(t) = \int_{-\infty}^t f_X(s) ds.$$

Or, on sait que $\lim_{t \rightarrow +\infty} F_X(t) = 1$. Ainsi, on en déduit

$$1 = \lim_{t \rightarrow +\infty} F_X(t) = \lim_{t \rightarrow +\infty} \int_{-\infty}^t f_X(s) ds = \int_{-\infty}^{+\infty} f_X(s) ds,$$

ce qui achève la preuve. □

Proposition 8.2.13. *Soit une variable aléatoire réelle X à densité. Soit f_X sa densité de probabilité. Alors, pour tout $t \in \mathbb{R}$, on a*

$$\mathbb{P}(X > t) = \int_t^{+\infty} f_X(s) ds.$$

Démonstration. En effet, on peut écrire

$$\begin{aligned} \mathbb{P}(X > t) &= 1 - \mathbb{P}(X \leq t) \\ &= 1 - F_X(t) \\ &= 1 - \int_{-\infty}^t f_X(s) ds \\ &= \int_{-\infty}^{+\infty} f_X(s) ds - \int_{-\infty}^t f_X(s) ds \\ &= \int_t^{+\infty} f_X(s) ds, \end{aligned}$$

par la relation de Chasles. □

Théorème 8.2.14. *Soit une fonction f de \mathbb{R} dans \mathbb{R} qui satisfait la Proposition 8.2.11 et la Proposition 8.2.12, c'est-à-dire telle que f est positive ou nulle et d'intégrale sur \mathbb{R} égale à 1. Alors, il existe un espace fondamental Ω , muni d'une probabilité \mathbb{P} et une variable aléatoire réelle à densité X de Ω dans \mathbb{R} telle que f est la densité de probabilité de X .*

Exercice 8.2.15. *Prouver le Théorème 8.2.14 en utilisant le Théorème 3.3.16 à la page 55.*

8.3 Calcul de la probabilité d'un intervalle

Théorème 8.3.1. *Soit une variable aléatoire réelle X . On suppose que la fonction de répartition F_X est continue au point a . Alors, $\mathbb{P}(X = a) = 0$.*

Démonstration. Soit $\delta > 0$ quelconque. On sait d'après la Proposition 3.3.10 :

$$\mathbb{P}(a - \delta < X \leq a + \delta) = F_X(a + \delta) - F_X(a - \delta).$$

D'après l'hypothèse de continuité de F_X en a , le membre de droite tend vers 0 quand δ tend vers 0. Or, l'évènement $\{X = a\}$ est inclus dans l'évènement $\{a - \delta < X \leq a + \delta\}$ pour tout $\delta > 0$. Il s'ensuit

$$\mathbb{P}(X = a) \leq \mathbb{P}(a - \delta < X \leq a + \delta) \longrightarrow 0,$$

quand δ tend vers 0. La preuve est ainsi achevée. \square

Remarque 8.3.2 (Conséquence). *Soit une variable aléatoire réelle à densité X . Alors, pour tout $t \in \mathbb{R}$, on a $\mathbb{P}(X = t) = 0$. En effet, comme la variable aléatoire réelle X est à densité, sa fonction de répartition est continue en tout point de \mathbb{R} .*

Ainsi, contrairement aux variables aléatoires réelles discrètes dont l'ensemble des réalisations est dénombrable, on ne peut pas caractériser la loi de probabilité d'une variable aléatoire réelle à densité en connaissant uniquement les probabilités élémentaires $\mathbb{P}(X = t)$. Ceci vient du fait que l'ensemble $\{t\}$ est de mesure de Lebesgue nulle. On doit donc regarder la probabilité que X soit dans un intervalle de mesure de Lebesgue strictement positive.

Soient deux réels a et b tels que $a < b$. On pose $I_1 := [a; b]$, $I_2 :=]a; b]$, $I_3 := [a; b[$, $I_4 :=]a; b[$. Alors, on a

$$\mathbb{P}(X \in I_1) = \mathbb{P}(X \in I_2) = \mathbb{P}(X \in I_3) = \mathbb{P}(X \in I_4) = \int_a^b f_X(t) dt. \quad (8.3)$$

On sait en effet d'après la Proposition 3.3.10 : $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$. Et, comme $\mathbb{P}(X = a) = \mathbb{P}(X = b) = 0$, l'Égalité (8.3) est vérifiée.

Exercice 8.3.3. *Soit X une variable aléatoire réelle à densité de loi $\mathcal{U}_{[0;1]}$. Calculer la quantité $\mathbb{P}\left(\frac{1}{2} < X < \frac{3}{2}\right)$.*

8.4 Caractéristiques

Dans cette section, on regarde les grandeurs habituelles des variables aléatoires réelles à densité. Les concepts sont très similaires à ceux présentés dans le cas des variables aléatoires réelles discrètes, voir la page 62.

8.4.1 Espérance d'une variable aléatoire réelle

Définition 8.4.1. Soit une variable aléatoire réelle à densité X . Soit f_X la densité de probabilité de la variable aléatoire réelle X . On dit que X admet une espérance si l'on a

$$\int_{\mathbb{R}} |x| f_X(x) dx < \infty.$$

Dans ce cas, l'espérance de X est égale à

$$\mathbb{E}[X] := \int_{\mathbb{R}} x f_X(x) dx. \quad (8.4)$$

Remarque 8.4.2. On doit supposer la finitude de la quantité $\int_{\mathbb{R}} |x| f_X(x) dx$ sans quoi l'espérance telle qu'elle est définie dans (8.4) n'a pas de sens du point de vue de la convergence dans L^1 . Par exemple, si l'on considère une variable aléatoire X suivant une loi de Cauchy c'est-à-dire telle que $f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}$, alors $\int_{\mathbb{R}} \frac{1}{\pi} \frac{x}{1+x^2} dx$ ne converge pas et l'on en déduit que la variable aléatoire X n'admet pas d'espérance.

Illustrons l'espérance avec l'exemple suivant

Exemple 8.4.3. Soit X une variable aléatoire réelle à densité de loi uniforme sur l'intervalle $[0; 1]$, $\mathcal{U}_{[0;1]}$. Ici, on a $f_X(x) = \mathbf{1}_{[0;1]}(x)$. Calculons l'espérance de X :

$$\begin{aligned} \mathbb{E}[X] &= \int_{\mathbb{R}} x f_X(x) dx \\ &= \int_0^1 x dx \\ &= \frac{1}{2}. \end{aligned}$$

8.4.1.1 Propriétés

Présentons ici les propriétés classiques de l'espérance mathématique d'une variable aléatoire à densité.

8.4.1.1.1 Espérance d'une fonction de variable aléatoire réelle Soit maintenant une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . Soit une fonction réelle (et mesurable) ψ de \mathbb{R} dans \mathbb{R} . Alors, $\psi(X)$ est une variable aléatoire réelle. On cherche à calculer l'espérance de $\psi(X)$.

Théorème 8.4.4 (Formule de transfert). Supposons que X soit une variable aléatoire réelle continue de densité de probabilité f_X . Alors, on a

$$\mathbb{E}[\psi(X)] = \int_{\mathbb{R}} \psi(x) f_X(x) dx, \quad (8.5)$$

à condition que l'intégrale $\int_{\mathbb{R}} |\psi(x)| f_X(x) dx$ converge.

Procédons à la preuve dans le cas où la fonction ψ est une bijection de \mathbb{R} dans \mathbb{R} telle que ψ' ne s'annule pas. L'idée est la même avec une fonction quelconque mais choisir une telle fonction permet de s'affranchir de certaines considérations techniques.

Démonstration. Par définition, l'espérance de la variable aléatoire réelle à densité $\psi(X)$ est égale à

$$\mathbb{E}[\psi(X)] = \int_{\mathbb{R}} x f_{\psi(X)}(x) dx,$$

où $f_{\psi(X)}$ est la densité de probabilité de la variable aléatoire réelle $\psi(X)$. Pour la calculer, on regarde d'abord la fonction de répartition de $\psi(X)$:

$$\begin{aligned} F_{\psi(X)}(t) &= \mathbb{P}(\psi(X) \leq t) \\ &= \mathbb{P}(X \leq \psi^{-1}(t)) \\ &= F_X(\psi^{-1}(t)), \end{aligned}$$

car la fonction ψ est strictement croissante. On dérive maintenant et l'on obtient

$$\begin{aligned} f_{\psi(X)}(t) &= \frac{d}{dt} F_{\psi(X)}(t) \\ &= \frac{d}{dt} F_X(\psi^{-1}(t)) \\ &= (\psi^{-1})'(t) F_X'(\psi^{-1}(t)) \\ &= (\psi^{-1})'(t) f_X(\psi^{-1}(t)). \end{aligned}$$

On en déduit

$$\mathbb{E}[\psi(X)] = \int_{\mathbb{R}} t (\psi^{-1})'(t) f_X(\psi^{-1}(t)) dt.$$

On procède au changement de variable $t := \psi(x)$ et la preuve de l'Égalité (8.5) est terminée. \square

Donnons un exemple classique de ce résultat.

Exemple 8.4.5 (Moment à l'ordre n). Soit $n \in \mathbb{N}^*$. Prenons $g_n(x) := x^n$. On peut alors calculer, si elle existe, la quantité

$$\mathbb{E}[X^n] = \int_{\mathbb{R}} x^n f_X(x) dx.$$

Cette quantité est appelée "moment d'ordre n de la variable aléatoire réelle X ".

Exercice 8.4.6. Soit $n \geq 1$. Soit X une variable aléatoire réelle à densité dont la loi de probabilité est la loi uniforme sur l'intervalle $[0; 1]$, $\mathcal{U}_{[0;1]}$. Calculer le moment d'ordre n de X .

8.4.1.1.2 Linéarité de l'espérance Soient deux variables aléatoires réelles X et Y définies sur le même espace fondamental. Soit $\lambda \in \mathbb{R}$. On suppose que les deux variables aléatoires réelles X et Y admettent chacune une espérance. On a alors

$$\mathbb{E}[\lambda X + Y] = \lambda \mathbb{E}[X] + \mathbb{E}[Y]. \quad (8.6)$$

La preuve de la linéarité lorsque les variables aléatoires réelles sont à densité nécessite de reconstruire l'intégrale de Lebesgue, ce que nous ne faisons pas ici. En d'autres termes, l'égalité

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

est admise dans ce livre. L'idée sous-jacente est la suivante :

$$\begin{aligned} \mathbb{E}[X + Y] &= \int_{\Omega} (X(\omega) + Y(\omega)) \mathbb{P}(d\omega) \\ &= \int_{\Omega} X(\omega) \mathbb{P}(d\omega) + \int_{\Omega} Y(\omega) \mathbb{P}(d\omega) \\ &= \mathbb{E}[X] + \mathbb{E}[Y]. \end{aligned}$$

Quant à l'égalité

$$\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X],$$

elle découle immédiatement du Théorème 8.4.4 en prenant la fonction $g(x) := \lambda \times x$. On peut généraliser ce résultat.

Théorème 8.4.7. *Soient n variables aléatoires réelles à densité X_1, \dots, X_n définies sur un même univers Ω muni d'une probabilité \mathbb{P} . Soient $\lambda_0, \lambda_1, \dots, \lambda_n$ des réels. On a alors*

$$\mathbb{E}[\lambda_0 + \lambda_1 X_1 + \dots + \lambda_n X_n] = \lambda_0 + \lambda_1 \mathbb{E}[X_1] + \dots + \lambda_n \mathbb{E}[X_n].$$

8.4.1.1.3 Espérance d'une variable aléatoire réelle de signe constant

Soit une variable aléatoire réelle X qui a toutes ses réalisations possibles positives. Alors son espérance est positive. De même, si la variable aléatoire réelle X est négative alors son espérance est négative :

$$\begin{aligned} X \geq 0 &\implies \mathbb{E}[X] \geq 0 \\ \text{et } X \leq 0 &\implies \mathbb{E}[X] \leq 0. \end{aligned}$$

Démonstration. On considère uniquement le cas où X est positive. En effet, si la positivité de X implique la positivité de son espérance, alors on utilise (8.6) avec $\lambda := -1$ et $Y = 0$.

Si X est à densité et si $X \geq 0$, alors $F_X(0) = 0$ et donc $F_X(t) = 0$ pour tout $t \leq 0$ d'où la densité f_X est nulle sur $] -\infty; 0[$. Ainsi, on a

$$\mathbb{E}[X] = \int_0^{\infty} x f_X(x) dx \geq 0,$$

en utilisant la positivité de l'intégrale d'une fonction positive sur un intervalle. \square

8.4.1.1.4 Espérance d'un produit de variables aléatoires réelles mutuellement indépendantes On présente dans ce paragraphe un résultat important.

Proposition 8.4.8. *Soient n variables aléatoires réelles à densité X_1, \dots, X_n définies sur un même espace fondamental Ω muni d'une probabilité \mathbb{P} . Supposons que les n variables aléatoires réelles sont mutuellement indépendantes, voir la Définition 3.4.1 à la page 56. On suppose également que chacune de ces variables aléatoires réelles admet une espérance. Alors, la variable aléatoire réelle $X_1 \times \dots \times X_n$ admet une espérance et de plus :*

$$\mathbb{E}[X_1 \times \dots \times X_n] = \mathbb{E}[X_1] \times \dots \times \mathbb{E}[X_n].$$

Remarque 8.4.9. *Cette égalité est nécessaire mais elle n'est pas suffisante pour que l'on ait l'indépendance mutuelle des variables aléatoires.*

On peut aller plus loin.

Théorème 8.4.10. *Soient n variables aléatoires réelles à densité X_1, \dots, X_n définies sur un même espace fondamental Ω muni d'une probabilité \mathbb{P} . Supposons que les n variables aléatoires réelles sont mutuellement indépendantes, voir la Définition 3.4.1 à la page 56. Soient n fonctions bornées f_1, \dots, f_n . Alors, la variable aléatoire réelle $f_1(X_1) \times \dots \times f_n(X_n)$ admet une espérance et de plus :*

$$\mathbb{E}[f_1(X_1) \times \dots \times f_n(X_n)] = \mathbb{E}[f_1(X_1)] \times \dots \times \mathbb{E}[f_n(X_n)]. \quad (8.7)$$

Notons que la réciproque est vraie.

Proposition 8.4.11. *Soient n variables aléatoires réelles à densité X_1, \dots, X_n définies sur un même espace fondamental Ω muni d'une probabilité \mathbb{P} . Supposons que l'Égalité (8.7) est vérifiée pour toutes les fonctions bornées, alors les n variables aléatoires réelles X_1, \dots, X_n sont mutuellement indépendantes.*

8.4.1.1.5 Inégalité de Markov On donne maintenant une inégalité importante concernant l'espérance mathématique, que l'on a déjà vue dans le cas discret.

Théorème 8.4.12 (Inégalité de Markov). *Soit une variable aléatoire réelle à densité X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . Supposons qu'elle admette une espérance finie c'est-à-dire que l'on ait $\mathbb{E}[|X|] < \infty$. Alors, pour tout $R > 0$, on a*

$$\mathbb{P}(|X| \geq R) \leq \frac{\mathbb{E}[|X|]}{R}. \quad (8.8)$$

La preuve est exactement la même que celle du Théorème 4.3.17 à la page 67.

8.4.1.2 Variable aléatoire réelle centrée

Définition 8.4.13. *On dit qu'une variable aléatoire réelle X à densité est centrée lorsque son espérance mathématique est nulle.*

Centrer une variable aléatoire réelle X , c'est lui retrancher son espérance, c'est-à-dire considérer la variable aléatoire réelle $X - \mathbb{E}[X]$. En effet, d'après la linéarité de l'espérance, on a

$$\begin{aligned} \mathbb{E}\{X - \mathbb{E}[X]\} &= \mathbb{E}[X] - \underbrace{\mathbb{E}[\mathbb{E}[X]]}_{\mathbb{E}[X], \mathbb{E}[X] \text{ étant constant}} = 0. \end{aligned}$$

8.4.2 Variance, Écart-type

La connaissance de l'espérance est insuffisante pour caractériser une loi de probabilité.

Par exemple, si X suit la loi normale centrée réduite et si Y suit la loi normale centrée de variance $\sigma^2 = 16$, la seconde est plus dispersée que la première.

8.4.2.1 Définition

Définition 8.4.14. *Soit une variable aléatoire réelle à densité X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance. Alors, par définition, la variance de la variable aléatoire réelle X est*

$$\sigma^2(X) = \text{Var}(X) := \mathbb{E}\{(X - \mathbb{E}[X])^2\}.$$

Cette quantité est toujours définie mais elle peut éventuellement être égale à l'infini. Si elle est finie, on dit que la variance de X est finie. D'après la positivité de l'espérance d'une variable aléatoire réelle positive, on en déduit que $\text{Var}(X)$ est positive.

Définition 8.4.15. *Soit une variable aléatoire réelle à densité X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance. Alors, par définition, l'écart-type de la variable aléatoire réelle X est*

$$\sigma(X) := \sqrt{\text{Var}(X)} = \sqrt{\mathbb{E}\{(X - \mathbb{E}[X])^2\}}.$$

La variance et l'écart-type sont des indicateurs de la dispersion de la loi de X autour de son espérance.

Exercice 8.4.16. *Soit une variable aléatoire réelle à densité X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X suit la loi de Laplace c'est-à-dire que sa densité f_X est :*

$$f_X(x) := \frac{1}{2}e^{-|x|}.$$

Calculer la variance de X .

Proposition 8.4.17. *Soit une variable aléatoire réelle à densité X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance ainsi qu'une variance finie. Alors, $\mathbb{E}[X^2]$ est finie et de plus*

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \quad (8.9)$$

La preuve est exactement la même que celle de la Proposition 4.3.21 à la page 70.

8.4.2.2 Inégalité de Bienaymé-Tchebychev

Théorème 8.4.18. *Soit une variable aléatoire réelle à densité X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance ainsi qu'une variance finie. Alors, pour tout $t > 0$, on a l'inégalité*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}. \quad (8.10)$$

La preuve est exactement la même que celle du Théorème 4.3.22 à la page 70.

On peut présenter l'inégalité de Bienaymé-Tchebychev d'une autre manière. En effet, comme dit plus haut, la variance et l'écart-type mesurent la dispersion d'une variable aléatoire réelle X .

Théorème 8.4.19. *Soit une variable aléatoire réelle à densité X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance ainsi qu'une variance finie. Alors, pour tout $s > 0$, on a l'inégalité*

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq s\sqrt{\text{Var}(X)}\right) \leq \frac{1}{s^2}.$$

8.4.2.3 Propriétés

On présente maintenant les propriétés classiques sur la variance d'une variable aléatoire réelle à densité.

8.4.2.3.1 Calcul de $\text{Var}[aX]$ avec $a \in \mathbb{R}$ En utilisant les diverses hypothèses sur l'espérance, comme dans le cas discret, on a

$$\text{Var}[aX] = a^2\text{Var}[X] \text{ et } \sqrt{\text{Var}[aX]} = |a|\sqrt{\text{Var}[X]}.$$

8.4.2.3.2 Calcul de $\text{Var}[X + Y]$ Soient deux variables aléatoires réelles à densité X et Y définies sur un même espace fondamental Ω muni d'une probabilité \mathbb{P} . Alors, on a :

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]. \quad (8.11)$$

Définition 8.4.20. Soient deux variables aléatoires réelles X et Y définies sur un même espace fondamental Ω muni d'une probabilité \mathbb{P} . On appelle covariance de X et Y , et l'on note $\text{Cov}(X, Y)$ la quantité

$$\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] .$$

Remarque 8.4.21. On peut observer : $\text{Cov}(X, X) = \text{Var}[X]$.

Ainsi, l'Égalité (8.11) devient

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}(X, Y) . \quad (8.12)$$

On peut alors remarquer

$$\text{Var}[X - Y] = \text{Var}[X] + \text{Var}[Y] - 2\text{Cov}(X, Y) .$$

Si la covariance des variables aléatoires réelles X et Y est nulle, alors la variance de la somme est égale à la somme des variances.

Théorème 8.4.22. Soient n variables aléatoires réelles X_1, \dots, X_n définies sur un même univers Ω muni d'une probabilité \mathbb{P} . On suppose que les n variables aléatoires réelles admettent une espérance et une variance finie. On suppose également qu'elles sont indépendantes deux à deux. Alors, la variance de leur somme est égale à la somme de leurs variances :

$$\text{Var}[X_1 + \dots + X_n] = \text{Var}[X_1] + \dots + \text{Var}[X_n] .$$

Observons que l'on n'a pas besoin de supposer l'indépendance mutuelle mais seulement l'indépendance deux à deux.

La preuve est exactement la même que celle du Théorème 4.3.33 à la page 73.

En particulier, si X_1, \dots, X_n sont n variables aléatoires réelles indépendantes et identiquement distribuées, on a

$$\text{Var}[X_1 + \dots + X_n] = n\text{Var}[X_1] .$$

Comme dans le cas discret, on a la Proposition suivante.

Proposition 8.4.23. Soit une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X admet une espérance ainsi qu'une variance finie. Soit $a \in \mathbb{R}$. Alors, on a

$$\text{Var}[X + a] = \text{Var}[X] .$$

Ce résultat est naturel puisque l'ajout d'une constante ne fait que translater la variable aléatoire réelle X . Elle ne participe pas à sa dispersion.

8.4.2.4 Variable aléatoire réelle réduite

Définition 8.4.24. *On dit qu'une variable aléatoire réelle à densité X est réduite lorsque sa variance est égale à 1. Ainsi, réduire une variable aléatoire réelle consiste à la diviser par son écart-type.*

Ainsi, une variable aléatoire réelle X est dite centrée, réduite si l'on a $\mathbb{E}[X] = 0$ et $\text{Var}[X] = 1$. Soit une variable aléatoire réelle X définie sur un espace fondamental Ω muni d'une probabilité \mathbb{P} . On suppose que X a une espérance et admet une variance finie. Alors, la variable aléatoire réelle

$$Y := \frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}[X]}}$$

est centrée et réduite.

8.4.3 Mélange de variables discrètes et à densité

On peut se demander ce qu'il se passe lorsque l'on considère des variables aléatoires discrètes et des variables aléatoires à densité ensemble. Par exemple, quelle est l'espérance de la somme d'une variable aléatoire réelle à densité et d'une variable aléatoire discrète ? Et l'espérance du produit si elles sont indépendantes ?

Tout est similaire. Ainsi,

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y],$$

quelle que soit la nature des variables aléatoires. De même, si X est indépendante de Y , alors l'espérance du produit est le produit des espérances. De fait, la variance de la somme est aussi la somme des variances.

Il convient de garder à l'esprit qu'au moyen de la théorie de la mesure, les deux grandes familles de variables aléatoires ne sont pas si différentes qu'il n'y paraît.

8.4.4 Skewness et kurtosis

On peut aussi considérer des statistiques (moments) d'ordre supérieur.

En particulier, on s'intéresse au coefficient d'asymétrie (aussi appelé skewness) :

$$\gamma_1 := \mathbb{E} \left[\left(\frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}[X]}} \right)^3 \right].$$

On regarde également le kurtosis :

$$\beta_2 := \mathbb{E} \left[\left(\frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}[X]}} \right)^4 \right].$$

Cette dernière quantité intervient naturellement lorsque l'on s'intéresse à la variance de la variance d'échantillon, voir page 264.

Il convient de noter que le skewness et le kurtosis peuvent aussi être définis dans le cas discret.

Lois de probabilité continues classiques

9.1 Introduction

On regarde maintenant quelques lois de probabilité admettant une densité. Comme pour le cas des lois discrètes usuelles, on choisit de n'en présenter que trois dans ce chapitre. D'autres seront étudiées en détail dans le Chapitre 10. Par définition, ces dernières admettent une densité de probabilité.

L'intégration et la dérivation sont un pré-requis indispensable à maîtriser avant de lire ce chapitre. La convolution serait un plus indéniable.

On rappelle justement ce résultat très utile portant sur la convolution des fonctions :

Proposition 9.1.1. *Soit un espace fondamental Ω muni d'une probabilité \mathbb{P} . On considère deux variables aléatoires réelles X_1 et X_2 . On suppose que X_1 et X_2 sont à densité et indépendantes. Alors, la variable aléatoire $X_1 + X_2$ est à densité et sa densité de probabilité est le produit de convolution des densités de probabilité de X_1 et de X_2 : $f_{X_1+X_2} = f_{X_1} * f_{X_2}$.*

Rappel 9.1.2. *Pour rappel, si f et g sont deux fonctions dans L^1 , alors la fonction $f * g$ est définie par $f * g(x) := \int_{\mathbb{R}} f(x-y)g(y)dy$. De plus, $f * g$ est elle-même dans L^1 .*

Les objectifs de ce chapitre sont de connaître sur le bout des doigts les trois lois que nous allons présenter : la loi uniforme continue, la loi exponentielle et surtout la loi normale (aussi appelée gaussienne). Savoir reconnaître les graphes ou même savoir les tracer approximativement est crucial. Il est également important de connaître l'espérance et la variance de la loi uniforme tandis qu'il faudrait dans l'idéal savoir tous les moments de la loi exponentielle. Connaître espérance et variance de la loi normale centrée réduite est à connaître absolument car comme son nom l'indique, la loi normale centrée réduite est centrée et réduite. À propos de la loi exponentielle, il serait préférable de saisir qu'elle est sans mémoire et idéalement de comprendre que c'est ce qui la rend si omniprésente dans les lois régissant le monde comme celle d'Arrhenius. Concernant la loi normale, celle-ci est stable par combinaison linéaire (sous une hypothèse d'indépendance mutuelle)

et il est si rare qu'une telle stabilité soit vérifiée que de fait, nous attendons du lecteur qu'il s'en souvienne. Enfin, lire les tables et connaître les propriétés de la fonction de répartition de la loi normale centrée réduite est essentiel. Notamment, il faudra être en mesure de pouvoir lire les tables et de savoir se servir du théorème de Thalès (ou d'un développement limité) pour approcher la valeur de la fonction de répartition en un point qui n'est pas répertorié dans les tables.

9.2 Loi uniforme

Définition 9.2.1. On dit que la variable aléatoire réelle X suit la loi uniforme sur $[a; b]$ avec $b > a$ lorsque X admet pour densité de probabilité la fonction

$$f_X(t) = \begin{cases} 0 & \text{si } t < a \\ \frac{1}{b-a} & \text{si } a \leq t \leq b \\ 0 & \text{si } t > b \end{cases} .$$

En d'autres termes,

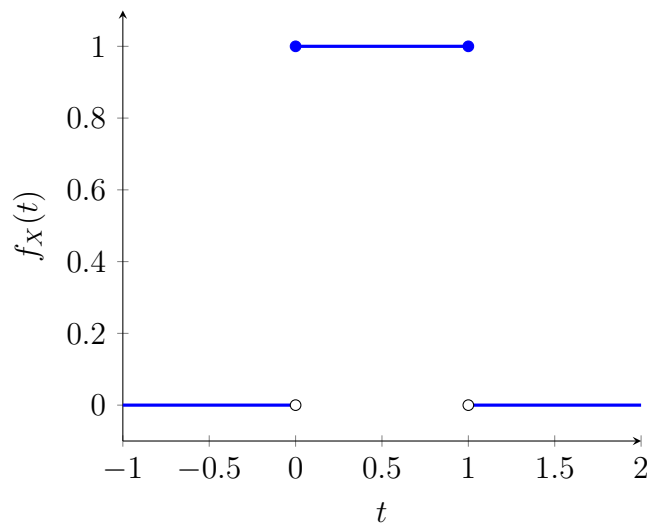
$$f_X(t) = \frac{1}{b-a} \mathbb{1}_{[a;b]}(t) .$$

Notation 9.2.2. La loi uniforme sur $[a; b]$ est notée $\mathcal{U}_{[a;b]}$.

Remarque 9.2.3. En cas d'oubli de la densité de probabilité de la loi uniforme, il suffit d'exploiter le caractère "uniforme" si bien que la densité de probabilité est constante sur l'intervalle. Puis, il suffit de diviser par la longueur de l'intervalle.

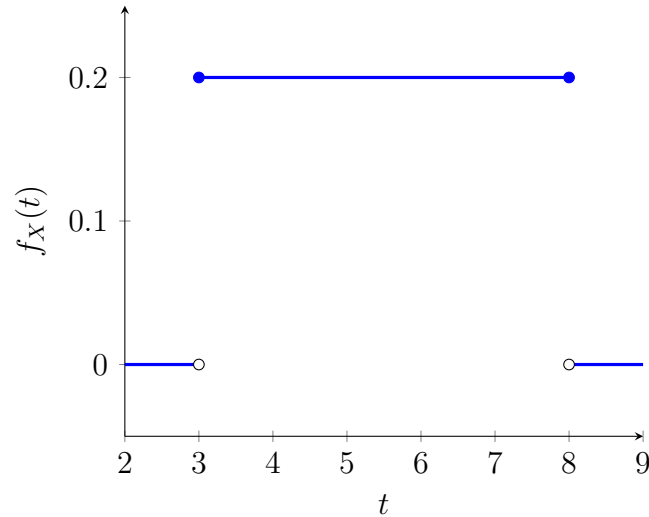
On trace sa courbe représentative avec $a = 0$ et $b = 1$:

FIGURE 9.1 – Densité de probabilité de la loi uniforme 1



On la trace également pour $a = 3$ et $b = 8$:

FIGURE 9.2 – Densité de probabilité de la loi uniforme 2

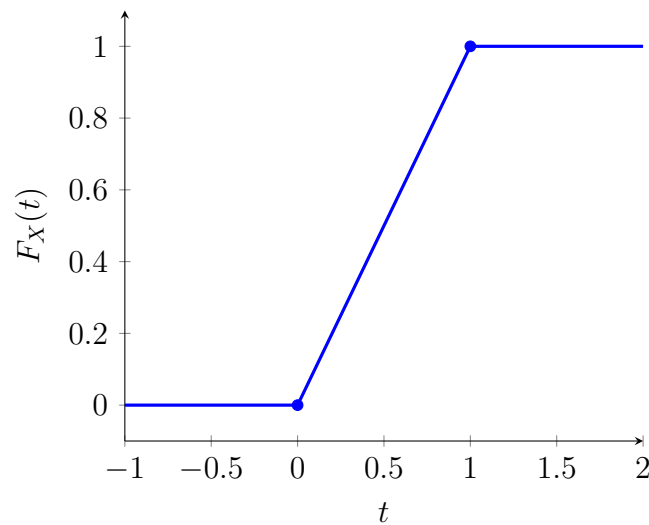


On peut maintenant calculer la fonction de répartition F_X avec $F_X(t) = \mathbb{P}(X \leq t)$:

$$F_X(t) = \begin{cases} 0 & \text{si } t < a \\ \frac{t-a}{b-a} & \text{si } a \leq t \leq b \\ 1 & \text{si } t > b \end{cases} .$$

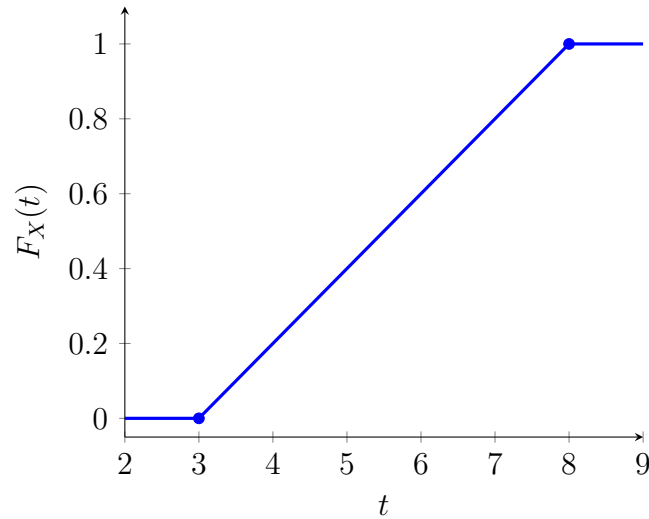
Cette fonction est continue et voici sa courbe représentative pour $a = 0$ et $b = 1$:

FIGURE 9.3 – Fonction de répartition de la loi uniforme 1



On la trace également pour $a = 3$ et $b = 8$:

FIGURE 9.4 – Fonction de répartition de la loi uniforme 2



9.2.1 Caractéristiques

On calcule ici les moments d'ordre n de la loi uniforme pour tout $n \in \mathbb{N}$:

$$\mathbb{E}[X^n] = \frac{1}{n+1} \frac{b^{n+1} - a^{n+1}}{b-a}. \quad (9.1)$$

Démonstration. On rappelle la formule de transfert :

$$\mathbb{E}[\psi(X)] = \int_{\mathbb{R}} \psi(t) f_X(t) dt,$$

pour toute fonction ψ telle que $\int_{\mathbb{R}} |\psi(t)| f_X(t) dt < \infty$.

On utilise la formule avec $\psi(t) := t^n$:

$$\begin{aligned} \mathbb{E}[X^n] &= \int_{-\infty}^{\infty} t^n f_X(t) dt \\ &= \int_{-\infty}^a t^n \times 0 dt + \int_a^b t^n \times \frac{1}{b-a} dt + \int_b^{\infty} t^n \times 0 dt \\ &= 0 + \frac{1}{b-a} \left[\frac{t^{n+1}}{n+1} \right]_a^b + 0 \\ &= \frac{1}{n+1} \frac{b^{n+1} - a^{n+1}}{b-a}. \end{aligned}$$

□

Remarque 9.2.4. On remarque notamment que la quantité $\mathbb{E}[X^0] = 1$ est bien égale à $\frac{1}{0+1} \frac{b-a}{b-a}$.

En particulier, on peut calculer facilement l'espérance et la variance :

$$\mathbb{E}[X] = \frac{a+b}{2} \quad (9.2)$$

et

$$\text{Var}[X] = \frac{(b-a)^2}{12}. \quad (9.3)$$

Pour montrer les deux égalités précédentes, il suffit d'utiliser l'Égalité (9.1) comme suit. Avec $n = 1$, on a immédiatement :

$$\mathbb{E}[X] = \frac{1}{1+1} \frac{b^2 - a^2}{b-a} = \frac{b+a}{2},$$

en utilisant une identité remarquable. Puis, pour le moment centré d'ordre deux :

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \frac{1}{3} \frac{b^3 - a^3}{b-a} - \frac{(a+b)^2}{4} \\ &= \frac{b^2 + ab + a^2}{3} - \frac{a^2 + 2ab + b^2}{4} \\ &= \frac{4b^2 + 4ab + a^2 - 3a^2 - 6ab - 3b^2}{12} \\ &= \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}. \end{aligned}$$

Remarque 9.2.5. La symétrie de la densité de probabilité par rapport à l'axe d'équation $x = \frac{a+b}{2}$ nous laissait deviner l'espérance sans que l'on ait besoin de procéder au calcul.

Il convient également de noter qu'une somme de variables aléatoires indépendantes mutuellement et suivant des lois uniformes ne suit pas une loi uniforme.

En effet, prenons le cas de n variables aléatoires X_1, \dots, X_n indépendantes et identiquement distribuées de loi uniforme sur $[0; 1]$. Alors $Y := X_1 + \dots + X_n$ est bien une variable à densité et la densité de celle-ci est le produit de convolution de $\mathbb{1}_{[0;1]}$ avec elle-même n fois. Toutefois, on observe :

$$\mathbb{1}_{[0;1]}(t) = H(t) - H(t-1) = (\delta_0 - \delta_1) * H,$$

où H est la fonction de Heaviside.

Puis, en utilisant l'associativité et la commutativité du produit de convolution dans l'algèbre de convolution des distributions causales, il vient :

$$f_Y(y) = (\delta_0 - \delta_1)^{*n} * \left(\underbrace{H * \dots * H}_n \text{ fois} \right) (y).$$

La partie avec les distributions de Dirac nous donne, en utilisant le binôme de Newton : $\sum_{k=0}^n \binom{n}{k} (-1)^k \delta_k$. Quant au produit de convolution de H avec elle-même à n reprises, on peut montrer par récurrence qu'il s'agit de la fonction $y \mapsto \frac{y^{n-1}}{(n-1)!} H(y)$. Par conséquent, la densité de Y est

$$f_Y(y) = \sum_{k=0}^n \binom{n}{k} (-1)^k \frac{(y-k)^{n-1}}{(n-1)!} H(y-k).$$

9.3 Loi exponentielle

L'un des intérêts de la loi exponentielle est qu'elle est sans mémoire (comme la loi géométrique $\mathcal{G}(p)$ dans le cadre discret), ce que nous verrons par la suite.

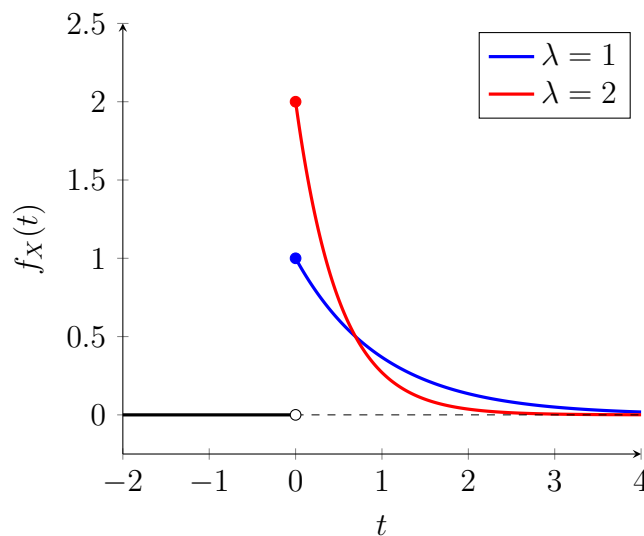
Définition 9.3.1. Soit λ un paramètre réel strictement positif. On dit que la variable aléatoire réelle X suit la loi exponentielle de paramètre λ lorsque X admet pour densité de probabilité la fonction

$$f_X(t) = \lambda e^{-\lambda t} H(t) = \begin{cases} 0 & \text{si } t < 0 \\ \lambda e^{-\lambda t} & \text{si } t \geq 0 \end{cases}.$$

Notation 9.3.2. La loi exponentielle de paramètre $\lambda > 0$ est notée $\mathcal{E}(\lambda)$.

Traçons la courbe de la densité de probabilité pour $\lambda = 1$ et pour $\lambda = 2$:

FIGURE 9.5 – Densité de probabilité de la loi exponentielle

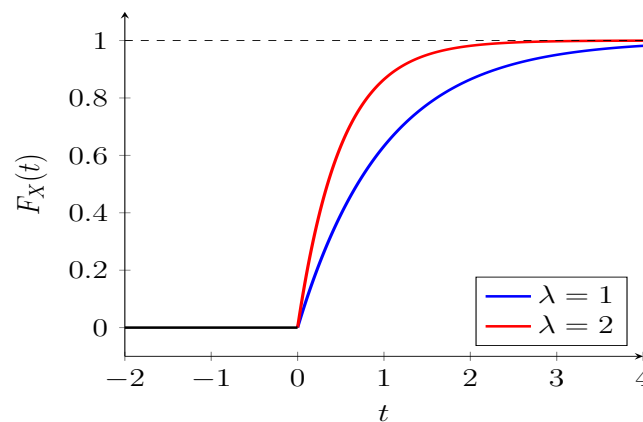


On peut maintenant calculer la fonction de répartition F_X à partir de la formule $F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^t f_X(s) ds$:

$$F_X(t) = (1 - e^{-\lambda t})H(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ 1 - e^{-\lambda t} & \text{si } t \geq 0 \end{cases} .$$

Cette fonction est continue et voici sa courbe représentative pour $\lambda = 1$ et pour $\lambda = 2$:

FIGURE 9.6 – Fonction de répartition de la loi exponentielle



En particulier, on constate que pour tout $t > 0$, on a $\mathbb{P}(X > t) = 1 - F_X(t) = e^{-\lambda t}$ d'où

$$\mathbb{P}(X > t + s \mid X > t) = \mathbb{P}(X > s),$$

pour tous les réels strictement positifs t et s . La réciproque est vraie : si une loi à densité est sans mémoire et prend ses valeurs dans \mathbb{R}_+ , alors c'est une loi exponentielle. On peut donc comparer la loi exponentielle à une loi géométrique.

Intuitivement, cela signifie que si vous allez pêcher, alors ne pas avoir attrapé une truite au bout de trois heures n'augmente pas vos chances d'en avoir une dans les trois heures qui suivent.

9.3.1 Caractéristiques

On calcule ici l'espérance et la variance de la variable aléatoire réelle à densité X :

$$\mathbb{E}[X] = \frac{1}{\lambda} .$$

Démonstration. On procède au calcul direct :

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} t f_X(t) dt \\ &= \int_{-\infty}^0 t \times 0 dt + \int_0^{\infty} t \times \lambda e^{-\lambda t} dt \\ &= 0 + \left[-\left(t + \frac{1}{\lambda}\right) e^{-\lambda t} \right]_0^{\infty} \\ &= \frac{1}{\lambda}.\end{aligned}$$

□

En procédant par récurrence, on peut prouver

$$\mathbb{E}[X^n] = \frac{n!}{\lambda^n},$$

pour tout $n \in \mathbb{N}$. En particulier, la variance de X est égale à

$$\text{Var}[X] = \frac{1}{\lambda^2}.$$

Exemple 9.3.3. *Sous certaines hypothèses, la variable aléatoire réelle qui a pour réalisation la durée de temps qui s'écoule entre deux arrivées dans une file d'attente suit la loi exponentielle de paramètre λ , λ étant le nombre moyen d'arrivées par unité de temps.*

Il convient de noter qu'une somme de variables aléatoires indépendantes et identiquement distribuées suivant la loi exponentielle de paramètre λ ne suit pas une loi exponentielle. Elle suit ce que l'on appelle une loi d'Erlang, voir la Section 10.4 à la page 151.

9.4 Loi normale (ou gaussienne)

La loi normale joue un rôle particulier en probabilité puisqu'elle est au centre du théorème central de la limite, voir la Section 15.4 à la page 217, lequel est vérifié sous des hypothèses "raisonnables".

9.4.1 Définition

Définition 9.4.1. *Soient deux réels m et σ^2 . On suppose $\sigma > 0$. On dit que la variable aléatoire réelle continue X suit la loi normale de paramètres m et σ^2 lorsqu'elle admet pour densité de probabilité la fonction*

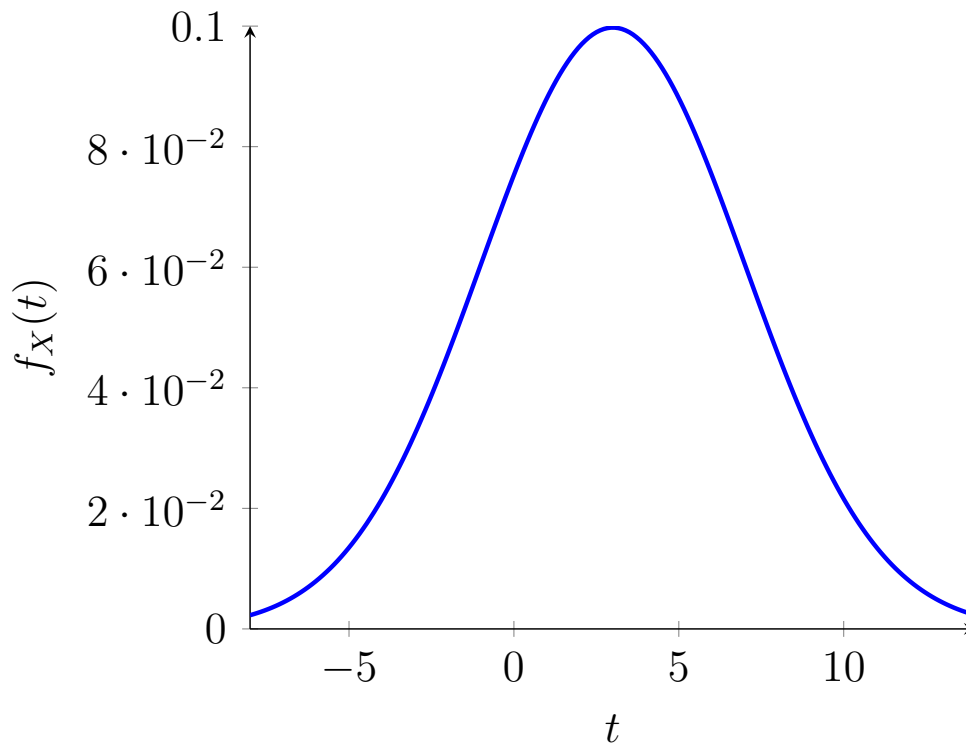
$$f_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \frac{(t-m)^2}{\sigma^2}\right\}.$$

Il peut sembler curieux que l'on note σ^2 le paramètre plutôt que σ . Il s'avère en fait qu'en dimension supérieure, au paramètre σ^2 , on substitue la matrice de covariance.

Travailler avec σ^2 est ainsi nettement plus naturel que de travailler avec sa racine carrée.

Cette fonction est continue et voici sa courbe représentative pour $m = 3$ et $\sigma^2 = 16$:

FIGURE 9.7 – Densité de probabilité de la loi normale



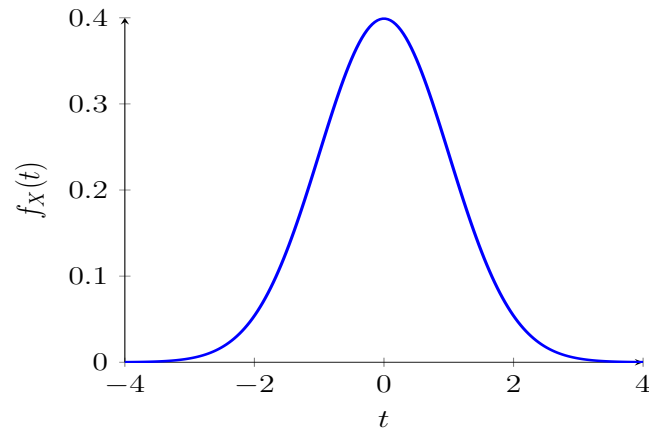
Notation 9.4.2. La loi normale de paramètres m et σ^2 est notée $\mathcal{N}(m, \sigma^2)$.

Définition 9.4.3. On dit que la variable aléatoire réelle continue X suit la loi normale centrée réduite si elle admet pour densité de probabilité la fonction

$$f_X(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.$$

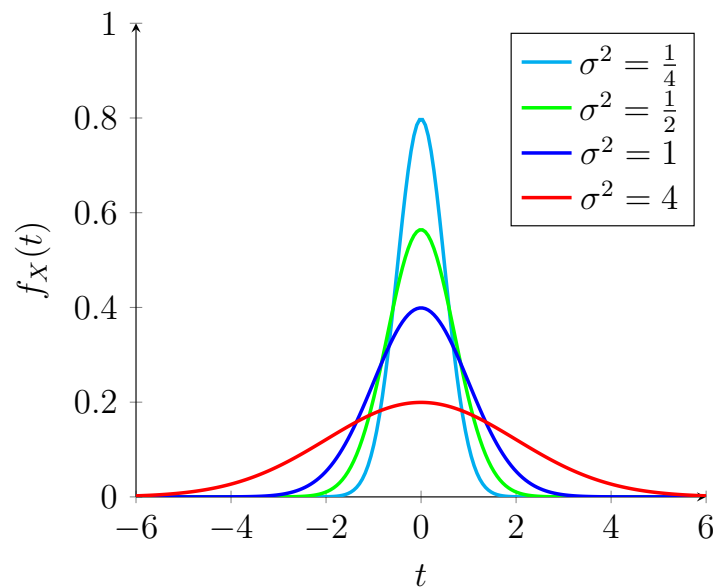
Cette fonction est continue et voici sa courbe représentative :

FIGURE 9.8 – Densité de probabilité de la loi normale centrée réduite



Superposons maintenant les densités de probabilités pour différentes lois normales centrées :

FIGURE 9.9 – Densités de probabilité de différentes lois normales centrées

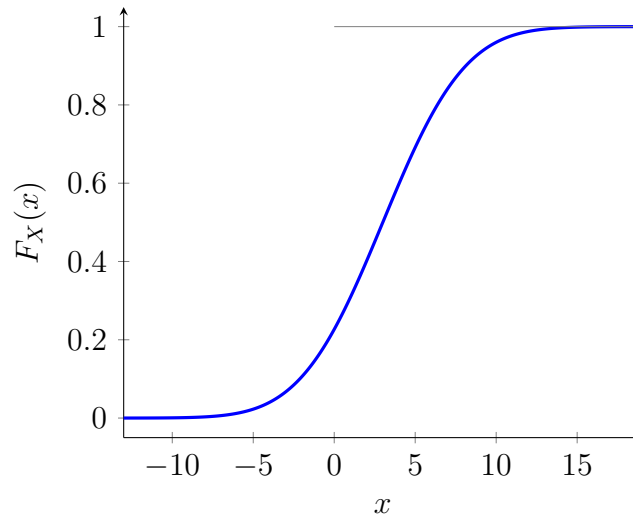


Notons que l'on ne peut pas calculer formellement la probabilité d'un intervalle. En effet, la fonction de répartition est

$$F_X(t) = \int_{-\infty}^t f_X(u) du = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^t \exp\left\{-\frac{1}{2} \frac{(u-m)^2}{\sigma^2}\right\} du.$$

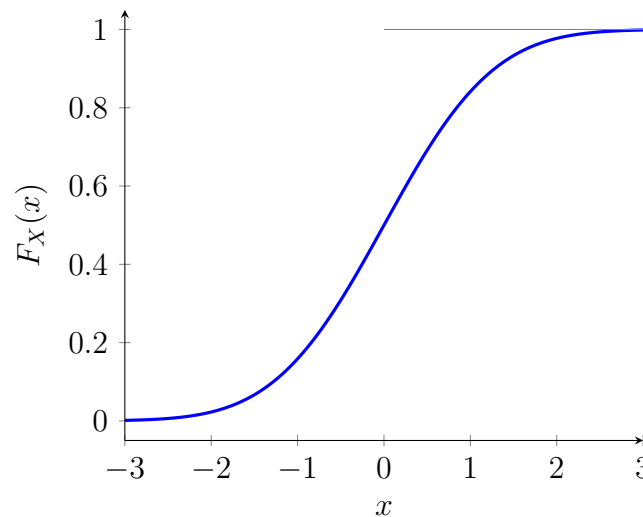
Voici sa courbe représentative pour $m = 3$ et $\sigma^2 = 16$:

FIGURE 9.10 – Fonction de répartition de la loi normale



On trace également la fonction de répartition de la loi normale centrée réduite :

FIGURE 9.11 – Fonction de répartition de la loi normale centrée réduite



9.4.2 Caractéristiques

On calcule les moments d'ordre n de la loi $\mathcal{N}(0, 1)$. On verra par la suite que l'on peut relier la loi normale $\mathcal{N}(m, \sigma^2)$ à $\mathcal{N}(0, 1)$. Soit X une variable aléatoire

réelle suivant la loi $\mathcal{N}(0, 1)$. Si n est impair ($n = 2p + 1$), on a

$$\mathbb{E}[X^{2p+1}] = 0,$$

vu que la loi normale centrée réduite est symétrique. Et, on peut prouver par récurrence :

$$\mathbb{E}[X^{2p}] = \frac{(2p)!}{2^p p!},$$

pour tout $p \in \mathbb{N}$. En particulier, on a

$$\mathbb{E}[X] = 0 \quad \text{et} \quad \text{Var}[X] = 1.$$

De manière générale, soit $m \in \mathbb{R}$ et soit $\sigma > 0$. Alors, si X est une variable aléatoire réelle suivant la loi $\mathcal{N}(m, \sigma^2)$, son espérance et sa variance sont

$$\mathbb{E}[X] = m \quad \text{et} \quad \text{Var}[X] = \sigma^2. \quad (9.4)$$

On verra dans la section suivante que si X suit une loi normale de paramètres m et σ^2 , alors $X = m + \sigma Y$ où Y suit une loi normale centrée réduite. De fait, on peut retrouver tous les moments de X :

$$\mathbb{E}[X^{2p}] = \sum_{k=0}^p \frac{(2p)! m^{2k} \sigma^{2p-2k}}{(2k)! (p-k)! 2^{p-k}},$$

et

$$\mathbb{E}[X^{2p+1}] = \sum_{k=0}^p \frac{(2p+1)! m^{2k+1} \sigma^{2p-2k}}{(2k+1)! (p-k)! 2^{p-k}}.$$

Il est bien sûr inutile d'apprendre les deux dernières formules ; vu qu'il est plus simple et plus rapide de les retrouver.

9.4.3 Propriétés

On donne ici les propriétés classiques de la loi normale.

Théorème 9.4.4. *Soient n variables aléatoires réelles mutuellement indépendantes X_1, \dots, X_n . On suppose :*

$$\mathcal{L}(X_k) = \mathcal{N}(m_k, \sigma_k^2).$$

Soient n réels $\alpha_1, \dots, \alpha_n$. Soit un autre réel β . Alors, la variable aléatoire réelle $\mathcal{X} := \alpha_1 X_1 + \dots + \alpha_n X_n + \beta$ suit une loi normale. Plus exactement :

$$\mathcal{L}(\mathcal{X}) = \mathcal{N}(\alpha_1 m_1 + \dots + \alpha_n m_n + \beta, \alpha_1^2 \sigma_1^2 + \dots + \alpha_n^2 \sigma_n^2). \quad (9.5)$$

Démonstration. Soit X_1 une variable aléatoire qui suit la loi $\mathcal{N}(m_1, \sigma_1^2)$ et soit X_2 indépendante de X_1 qui suit la loi $\mathcal{N}(m_2, \sigma_2^2)$. On suppose $\sigma_1^2 \sigma_2^2 > 0$. Alors, $Y := X_1 + X_2$ est à densité et

$$\begin{aligned} f_Y(y) &= f_{X_1} * f_{X_2}(y) \\ &= \int_{x \in \mathbb{R}} \frac{1}{2\pi \sqrt{\sigma_1^2 \sigma_2^2}} \exp \left\{ -\frac{(y-x-m_1)^2}{2\sigma_1^2} - \frac{(x-m_2)^2}{2\sigma_2^2} \right\} dx \\ &= \int_{x \in \mathbb{R}} \frac{1}{2\pi \sqrt{\sigma_1^2 \sigma_2^2}} \exp \left\{ -\frac{1}{2} [\alpha x^2 - 2\beta(y)x + \gamma(y)] \right\} dx. \end{aligned}$$

où $\alpha := \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$, $\beta(y) := \frac{y-m_1}{\sigma_1^2} + \frac{m_2}{\sigma_2^2}$ et $\gamma(y) := \frac{(y-m_1)^2}{\sigma_1^2} + \frac{m_2^2}{\sigma_2^2}$. En utilisant la forme canonique, il vient :

$$\begin{aligned} f_Y(y) &= \int_{x \in \mathbb{R}} \frac{1}{2\pi \sqrt{\sigma_1^2 \sigma_2^2}} \exp \left\{ -\frac{\alpha}{2} \left[x^2 - 2\frac{\beta(y)}{\alpha}x + \frac{\gamma(y)}{\alpha} \right] \right\} dx \\ &= \frac{\exp \left\{ -\frac{\gamma(y)}{2} \right\}}{2\pi \sqrt{\sigma_1^2 \sigma_2^2}} \int_{x \in \mathbb{R}} \exp \left\{ -\frac{\alpha}{2} \left[\left(x - \frac{\beta(y)}{\alpha} \right)^2 - \frac{\beta(y)^2}{\alpha^2} \right] \right\} dx \\ &= \frac{\exp \left\{ -\frac{\gamma(y)}{2} + \frac{\beta(y)^2}{2\alpha} \right\}}{2\pi \sqrt{\sigma_1^2 \sigma_2^2}} \int_{x \in \mathbb{R}} \exp \left\{ -\frac{\alpha}{2} \left[\left(x - \frac{\beta(y)}{\alpha} \right)^2 \right] \right\} dx \\ &= \frac{1}{2\pi \sqrt{\sigma_1^2 \sigma_2^2}} \sqrt{\frac{2\pi}{\alpha}} \exp \left\{ -\frac{\gamma(y)}{2} + \frac{\beta(y)^2}{2\alpha} \right\} \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp \left\{ -\gamma(y) + \frac{\beta(y)^2}{2\alpha} \right\}. \end{aligned}$$

$$\text{Or, } \frac{\gamma(y)}{2} - \frac{\beta(y)^2}{2\alpha} = \frac{1}{2(\sigma_1^2 + \sigma_2^2)} y^2 - \frac{m_1 + m_2}{\sigma_1^2 + \sigma_2^2} y + \frac{m_1^2 + m_2^2 + 2m_1 m_2}{2(\sigma_1^2 + \sigma_2^2)}.$$

Il s'ensuit que Y suit une loi normale de paramètres $m_1 + m_2$ et $\sigma_1^2 + \sigma_2^2$.

On peut ensuite prouver facilement que αX est une loi normale de paramètres αm et $\alpha^2 \sigma^2$ si X suit la loi normale de paramètres m et σ^2 .

Par récurrence, on prouve ensuite la stabilité par combinaison linéaire de variables aléatoires indépendantes suivant chacune une loi normale. Enfin, la somme d'une variable aléatoire réelle suivant une loi normale et d'une constante suit bien une loi normale. En effet, si X suit une loi normale de paramètres m et $\sigma^2 > 0$ et si a est une constante réelle, $Y := X + a$ est une variable aléatoire à densité et $f_Y(y) = (\delta_a * f_X)(y) = f_X(y - a)$; ce qui est bien la densité d'une loi normale de paramètres $m + a$ et σ^2 .

□

Exercice 9.4.5. On admet dorénavant que la variable aléatoire réelle $\alpha_1 X_1 + \dots + \alpha_n X_n + \beta$ suit une loi normale. Montrer (en utilisant les résultats classiques sur l'espérance et la variance ainsi que (9.4)) l'Égalité (9.5).

Le Théorème 9.4.4 s'applique en particulier pour passer d'une loi normale quelconque $\mathcal{N}(m, \sigma^2)$ à la loi normale centrée réduite $\mathcal{N}(0, 1)$.

Corollaire 9.4.6. Soit X une variable aléatoire réelle continue de densité de probabilité $\mathcal{N}(m, \sigma^2)$ avec $\sigma > 0$ et $m \in \mathbb{R}$. On pose $Y := \frac{X-m}{\sigma}$. Alors, la variable Y suit la loi de probabilité $\mathcal{N}(0, 1)$.

Réciproquement, soit Y une variable aléatoire réelle continue qui suit la loi de probabilité $\mathcal{N}(0, 1)$. Soit $m \in \mathbb{R}$ et $\sigma \neq 0$. Alors, la variable aléatoire réelle continue $\sigma Y + m$ suit la loi de probabilité $\mathcal{N}(m, \sigma^2)$.

Remarque 9.4.7. On constate en particulier que la loi normale centrée réduite $\mathcal{N}(0, 1)$ est bien la version centrée et réduite de toute loi normale; ce qui est cohérent.

Exemple 9.4.8. Soient deux variables aléatoires réelles indépendantes X et Y telles que $\mathcal{L}(X) = \mathcal{N}(2, 9)$ et $\mathcal{L}(Y) = \mathcal{N}(1, 4)$. Quelle est la loi de $X + 2Y + 4$?

D'après le Théorème 9.4.4, $\mathcal{L}(X + 2Y + 4) = \mathcal{N}(m, \sigma^2)$. Pour déterminer m et σ^2 , on écrit :

$$m = \mathbb{E}(X + 2Y + 4) = \mathbb{E}(X) + 2\mathbb{E}(Y) + 4 = 8,$$

ainsi que

$$\sigma^2 = \text{Var}(X + 2Y + 4) = \text{Var}(X + 2Y) = \text{Var}(X) + 4\text{Var}(Y) = 9 + 4 \times 4 = 25.$$

Ainsi, on obtient $\mathcal{L}(X + 2Y + 4) = \mathcal{N}(8, 25)$.

9.4.3.1 Tables

On se donne un intervalle $[a; b]$ et l'on cherche à connaître la probabilité de réalisation dans cet intervalle d'une variable aléatoire réelle X suivant la loi de probabilité $\mathcal{N}(m, \sigma^2)$ avec $\sigma > 0$,

$$\mathbb{P}(a \leq X \leq b).$$

La première étape pour ce calcul est de se ramener à une variable aléatoire réelle Y qui suit la loi $\mathcal{N}(0, 1)$. On utilise pour cela le Corollaire 9.4.6. On pose $Y := \frac{X-m}{\sigma}$. La variable aléatoire réelle Y suit la loi $\mathcal{N}(0, 1)$ puis :

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}\left(\frac{a-m}{\sigma} \leq Y \leq \frac{b-m}{\sigma}\right).$$

En notant Φ la fonction de répartition de la loi normale centrée réduite, il vient ainsi

$$\mathbb{P}(a \leq X \leq b) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right),$$

car la fonction Φ est continue. Conséquemment, la connaissance de la fonction Φ est suffisante. Étudions des propriétés de cette fonction.

Proposition 9.4.9. $\Phi(0) = \frac{1}{2}$.

Démonstration. La loi $\mathcal{N}(0, 1)$ est symétrique d'où $\mathbb{P}(X > 0) = \mathbb{P}(X < 0)$. \square

Proposition 9.4.10. Pour tout $t \in \mathbb{R}$, l'égalité suivante est vérifiée : $\Phi(-t) = 1 - \Phi(t)$.

Démonstration. Par définition,

$$\Phi(-t) = \mathbb{P}(X \leq -t) = \mathbb{P}(-X \geq t).$$

Or, d'après le Théorème 9.4.4, $-X$ suit la même loi que X . Ainsi :

$$\Phi(-t) = \mathbb{P}(X \geq t) = 1 - \mathbb{P}(X \leq t) = 1 - \Phi(t),$$

car la fonction Φ est continue. \square

Remarque 9.4.11. Les propriétés de la Propositions 9.4.9 et de la Propositions 9.4.10 sont en fait satisfaites dès que la loi est symétrique.

Or, un calcul approché nous donne

$$\Phi(3) = 0.99865,$$

à la cinquième décimale près. Ainsi, on se donne les valeurs approchées de $\Phi(t)$ pour t allant de 0 à 2.99. La cellule correspondant à la colonne 0.06 et à la ligne 1.5 est ainsi la valeur approchée de $\Phi(1.56)$. Voir Tableau 25.10 à la page 364.

Exemple 9.4.12. Soit une variable aléatoire réelle X suivant la loi $\mathcal{N}(5, 4)$. Calculons $\mathbb{P}(1 < X < 7)$. On considère d'abord $Y = \frac{X-5}{2}$. Alors $\mathcal{L}(Y) = \mathcal{N}(0, 1)$ puis :

$$\mathbb{P}(1 < X < 7) = \mathbb{P}(-2 < Y < 1) = \Phi(1) - \Phi(-2) = \Phi(1) + \Phi(2) - 1.$$

Or, d'après la table, $\Phi(1) \approx 0.8413$ et $\Phi(2) \approx 0.9772$ d'où $\mathbb{P}(1 < X < 7) \approx 0.8185$.

Donnons maintenant un exemple où l'on a besoin d'une valeur qui n'est pas dans la table.

Exemple 9.4.13. Soit une variable aléatoire réelle X suivant la loi $\mathcal{N}(5, 256)$. Calculons $\mathbb{P}(1 < X < 7)$. On considère d'abord $Y = \frac{X-5}{16}$. Alors $\mathcal{L}(Y) = \mathcal{N}(0, 1)$ puis :

$$\begin{aligned}\mathbb{P}(1 < X < 7) &= \mathbb{P}(-0.25 < Y < 0.125) \\ &= \Phi(0.125) - \Phi(-0.25) = \Phi(0.125) + \Phi(0.25) - 1.\end{aligned}$$

On connaît $\Phi(0.25) \approx 0.5987$. Mais on ne connaît pas $\Phi(0.125)$. On utilise alors l'approximation linéaire suivante :

$$\begin{aligned}\Phi(0.125) &= \Phi\left(\frac{1}{2} \times 0.12 + \frac{1}{2} \times 0.13\right) \\ &\approx \frac{\Phi(0.12) + \Phi(0.13)}{2} \approx \frac{0.5478 + 0.5517}{2} = 0.5498.\end{aligned}$$

Il vient ainsi $\mathbb{P}(1 < X < 7) \approx 0.1485$.

L'idée de l'approximation linéaire est la suivante. Comme Φ est de classe \mathcal{C}^1 , on peut procéder à un développement de Taylor à l'ordre un. Ainsi :

$$\Phi(x_0 + \Delta x) \approx \Phi(x_0) + \Delta x \Phi'(x_0),$$

où x_0 est dans les tables. On connaît $\Phi'(x_0)$ mais on ne peut pas l'évaluer si facilement sans calculatrice. On procède donc à l'approximation linéaire suivante :

$$\Phi'(x_0) \approx \frac{\Phi(x_0 + 10^{-2}) - \Phi(x_0)}{(x_0 + 10^{-2}) - x_0} = 10^2 (\Phi(x_0 + 10^{-2}) - \Phi(x_0)).$$

Par exemple, si l'on cherche $\Phi(0.127)$ au lieu de $\Phi(0.125)$, on a :

$$\begin{aligned}\Phi(0.127) &\approx \Phi(0.12) + 10^2 \times 0.007 \times (\Phi(0.13) - \Phi(0.12)) \\ &= 0.3\Phi(0.12) + 0.7\Phi(0.13) \\ &\approx 0.5505.\end{aligned}$$

D'autres tables sont fournies. Ainsi, à la page 365, on donne la fonction de répartition Φ en des *grandes valeurs* de t .

Et, on donne également le quantile d'ordre $1 - \frac{\gamma}{2}$ dans le Tableau 25.13 à la page 365.

Autres lois continues classiques

10.1 Introduction

Dans ce chapitre, nous présentons d'autres lois de probabilité continues classiques. Notons par ailleurs que la liste des lois présentées ici n'est pas exhaustive.

Par exemple, nous n'aborderons pas la loi de Gumbel, laquelle est utilisée en hydrologie pour prévoir le niveau des crues d'un fleuve, pour peu que l'on dispose du relevé des débits sur dix ans. De même, la loi logistique standard est abordée mais les autres lois logistiques ne le sont pas. La loi de Fréchet et la loi de l'extremum généralisée sont aussi passées sous silence ; de même que la loi bêta.

Les pré-requis de ce chapitre sont l'intégration, la dérivation mais aussi le produit de convolution de fonctions sommables. Par ailleurs, lors de la présentation de la loi de Student, nous utiliserons des résultats du Chapitre 12. Il est donc préférable, quoique non nécessaire, de transiger pour une fois avec la lecture linéaire de cet ouvrage.

L'objectif principal de ce chapitre est de découvrir de nombreuses autres lois à densité. Certaines sont utilisées très clairement en physique, d'autres en statistiques inférentielles.

En particulier, nous découvrirons la loi triangulaire, la loi logistique standard, la loi d'Erlang, la loi de Weibull, la loi Gamma, la loi du khi-deux, la loi de Rayleigh, la loi log-normale, la loi de Student, la loi de Fisher, la loi de Pareto et la loi de Cauchy.

10.2 Loi triangulaire

Comme pour la loi triangulaire discrète, la loi triangulaire est obtenue en faisant le produit de convolution de deux lois uniformes (continues) de même support (ce qui correspond à la loi de la somme de deux variables aléatoires réelles indépendantes suivant une même loi uniforme). On peut donc la calculer facilement.

Soient X et Y deux variables aléatoires suivant la loi $\mathcal{U}([a; b])$ avec $b > a$. On les suppose indépendantes. On a

$$f_X(x) = f_Y(x) = \frac{1}{b-a} \mathbf{1}_{[a;b]}(x).$$

On pose $Z := X + Y$. Z suit la loi triangulaire sur $[2a; 2b]$. En effet, comme X et Y sont indépendantes, alors Z est à densité et de plus :

$$f_Z(z) = (f_X * f_Y)(z) = \int_{\mathbb{R}} f_X(z-t) f_Y(t) dt.$$

On remarque que l'on a :

$$f_X(x) = \frac{1}{b-a} (H(t-a) - H(t-b)) = \frac{1}{b-a} ((\delta_a - \delta_b) * H)(t).$$

Ainsi, en utilisant la commutativité et l'associativité du produit de convolution pour des distributions à support compact :

$$f_Z(z) = \frac{1}{(b-a)^2} (\delta_{2a} - 2\delta_{a+b} + \delta_{2b}) * (H * H)(z).$$

Or, $(H * H)(z) = zH(z)$. Il vient :

$$f_Z(z) = \frac{1}{(b-a)^2} \left\{ (z-2a)H(z-2a) - 2(z-(a+b))H(z-(a+b)) + (z-2b)H(z-2b) \right\}.$$

On peut ensuite simplifier comme suit :

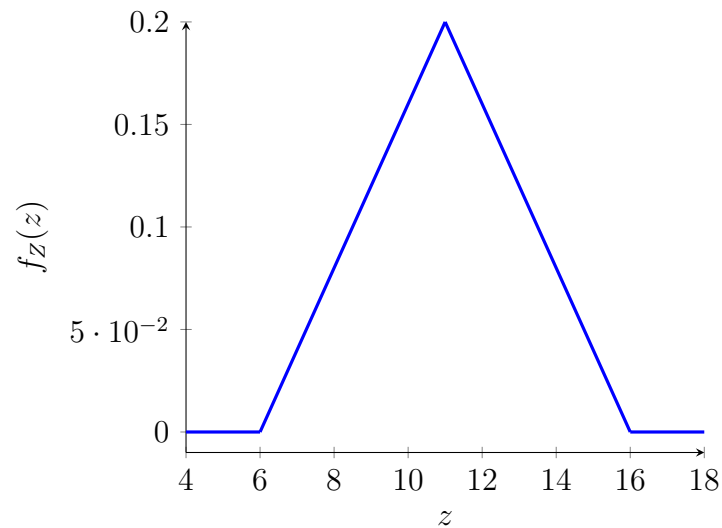
$$f_Z(z) = \frac{1}{(b-a)^2} \begin{cases} 0 & \text{si } z \leq 2a \\ z-2a & \text{si } 2a \leq z \leq a+b \\ 2b-z & \text{si } a+b \leq z \leq 2b \\ 0 & \text{si } z \geq 2b \end{cases}.$$

In fine, on a

$$f_Z(z) = \frac{(b-a) - |z - (a+b)|}{(b-a)^2} \mathbf{1}_{[2a; 2b]}(z).$$

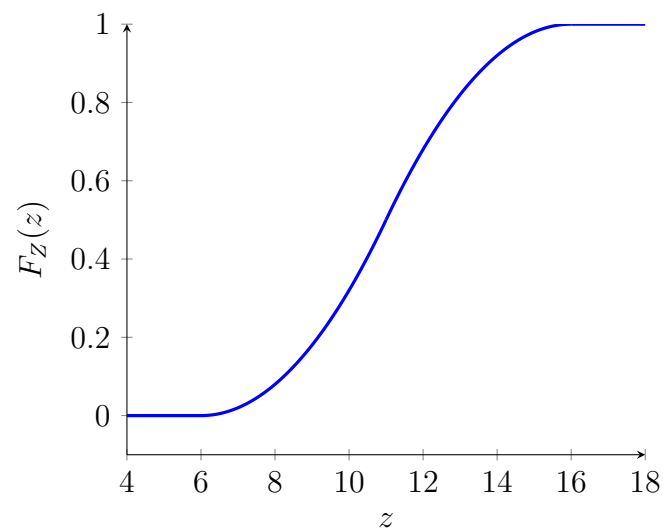
On trace cette densité de probabilité pour $a = 3$ et $b = 8$:

FIGURE 10.1 – Densité de probabilité de la loi triangulaire



On constate que la densité a bien une forme de triangle. On donne également la fonction de répartition, pour les mêmes valeurs :

FIGURE 10.2 – Fonction de répartition de la loi triangulaire



10.3 Loi logistique

Dans cette section, nous ne parlerons que de la loi logistique standard.

On dit que X suit la loi logistique standard si sa fonction de répartition est

$$F_X(x) = \frac{1}{1 + e^{-x}}.$$

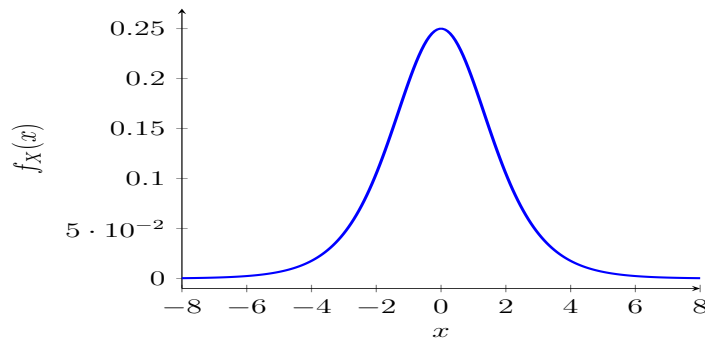
Il s'agit d'une fonction de classe \mathcal{C}^∞ donc en particulier, elle est continue sur \mathbb{R} et dérivable partout donc presque partout. La loi est donc à densité ; densité que l'on trouve en dérivant F_X :

$$f_X(x) = F'_X(x) = \frac{e^{-x}}{(1 + e^{-x})^2}.$$

Cette loi est utilisée en particulier pour le classement Elo, lequel sert à comparer le niveau de jeu des joueurs d'échecs, de go... Elle sert également dans les réseaux profonds de neurones.

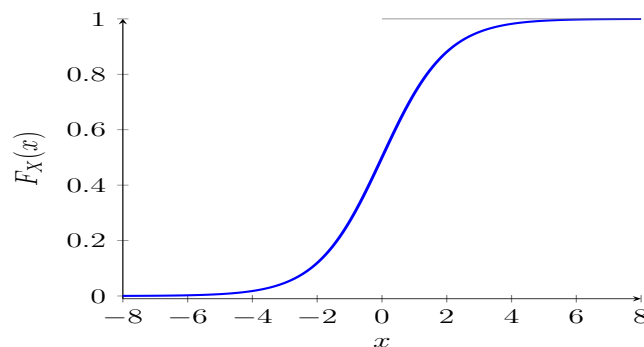
On trace cette densité de probabilité :

FIGURE 10.3 – Densité de probabilité de la loi logistique standard



On donne également la fonction de répartition :

FIGURE 10.4 – Fonction de répartition de la loi logistique standard



10.4 Loi d'Erlang

La loi d'Erlang dérive de la loi exponentielle. En effet, soient n variables aléatoires X_1, \dots, X_n indépendantes et identiquement distribuées suivant la loi exponentielle de paramètre $\lambda > 0$. Alors, la somme $Y := X_1 + \dots + X_n$ suit la loi d'Erlang de paramètres $n \in \mathbb{N}^*$ et $\lambda > 0$.

On peut prouver par récurrence, en utilisant le produit de convolution, que la loi de probabilité de Y est la suivante :

$$f_Y(y) = \lambda^n \frac{y^{n-1}}{(n-1)!} e^{-\lambda y} H(y).$$

On dit que n est le paramètre de forme et que λ est le paramètre d'intensité. Cette loi sert notamment à modéliser le nombre d'appels téléphoniques simultanés.

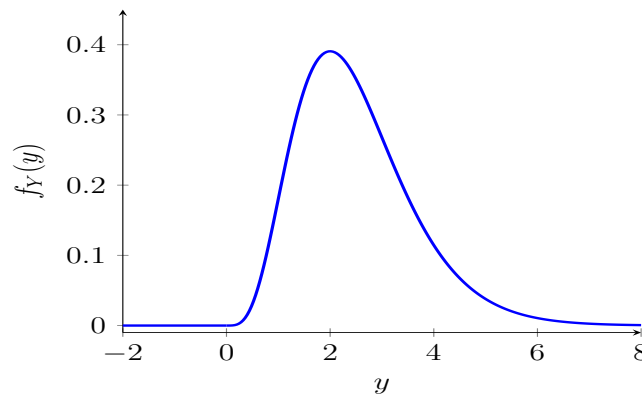
Notons que l'on peut en calculer facilement l'espérance et la variance : $\mathbb{E}[Y] = \frac{n}{\lambda}$ et $\text{Var}[Y] = \frac{n}{\lambda^2}$. En effet, Y est la somme de n variables aléatoires indépendantes suivant une loi exponentielle de paramètre λ . De manière alternative, on remarque :

$$\begin{aligned} \mathbb{E}[Y] &= \int_{\mathbb{R}_+} \lambda^n y^n \frac{1}{(n-1)!} e^{-\lambda y} dy \\ &= \frac{\lambda^{n-1}}{(n-1)!} \int_{\mathbb{R}_+} y^n e^{-\lambda y} dy \\ &= \frac{\lambda^{n-1}}{(n-1)!} \mathbb{E}[\tilde{X}^n], \end{aligned}$$

où \tilde{X} suit la loi exponentielle de paramètre $\lambda > 0$. Ainsi, $\mathbb{E}[Y] = \frac{\lambda^{n-1}}{(n-1)!} \times \frac{n!}{\lambda^n} = \frac{n}{\lambda}$.

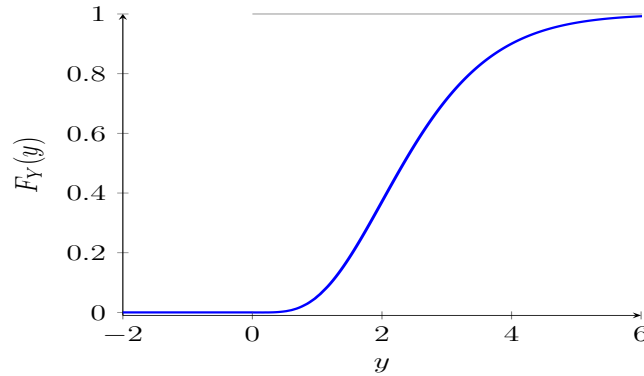
On trace sa densité de probabilité pour $n = 5$ et $\lambda = 2$:

FIGURE 10.5 – Densité de probabilité de la loi d'Erlang



On donne également la fonction de répartition, pour les mêmes valeurs :

FIGURE 10.6 – Fonction de répartition de la loi d'Erlang



10.5 Loi de Weibull

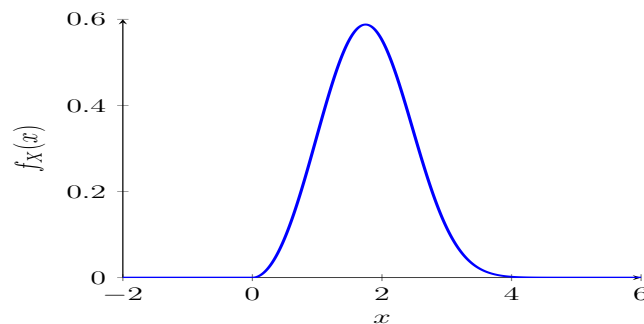
La loi de Weibull est une généralisation de la loi exponentielle. Elle est paramétrée par $k \in \mathbb{N}^*$ (le paramètre de forme) et par $\lambda > 0$ (le paramètre d'échelle). Par définition, on dit que X suit une telle loi si elle est à densité et si

$$f_X(x) = \frac{kx^{k-1}}{\lambda^k} e^{-\left(\frac{x}{\lambda}\right)^k} H(x).$$

Elle est notamment utilisée pour des analyses de défaillance en science des matériaux. Elle est aussi utilisée pour étudier le vent sur un site précis.

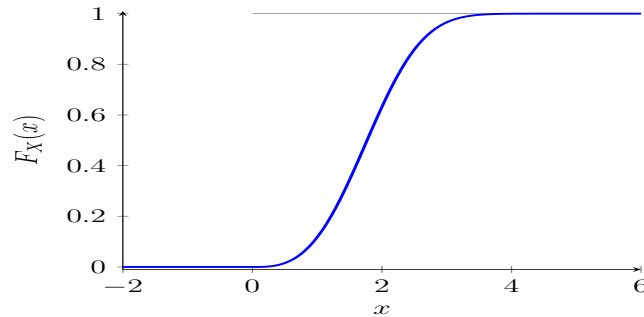
On trace sa densité de probabilité pour $k = 3$ et $\lambda = 2$:

FIGURE 10.7 – Densité de probabilité de la loi de Weibull



On donne également la fonction de répartition, pour les mêmes valeurs :

FIGURE 10.8 – Fonction de répartition de la loi de Weibull



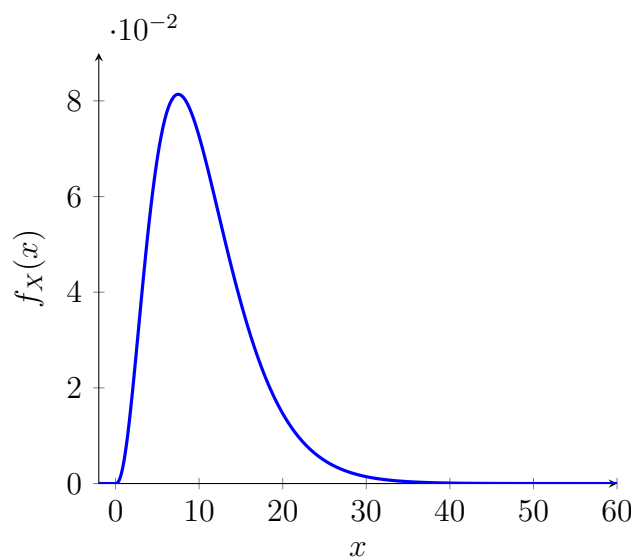
10.6 Loi Gamma

La loi Gamma découle de la fonction Gamma d'Euler que nous rappelons ici : $\Gamma(\alpha) := \int_0^\infty t^{\alpha-1} e^{-t} dt$ où $\alpha > 0$. On obtient par ailleurs (par un simple changement de variable) : $\int_0^\infty t^{\alpha-1} e^{-\frac{t}{\theta}} dt = \theta^\alpha \Gamma(\alpha)$. La loi Gamma est ainsi paramétrée par α et par θ et la densité de probabilité d'une variable aléatoire X qui la suit est

$$f_X(x) := \frac{x^{\alpha-1} e^{-\frac{x}{\theta}}}{\Gamma(\alpha) \theta^\alpha} H(x).$$

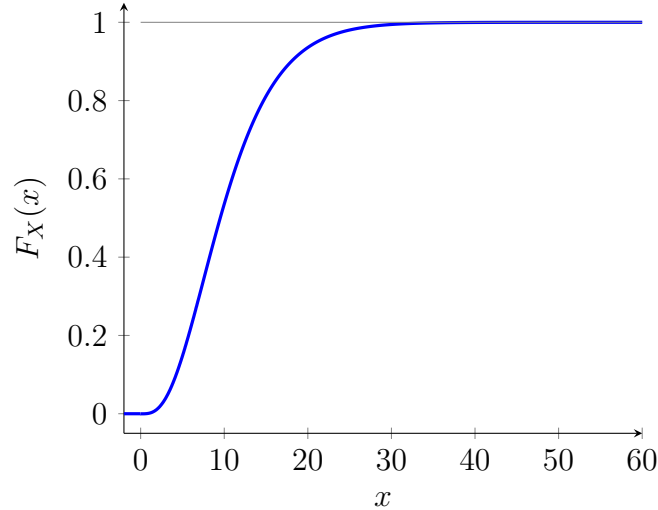
Notons que α n'est pas nécessairement entier (dans le cas où α est entier, on retrouve la loi d'Erlang). On trace sa densité de probabilité pour $\alpha = \frac{7}{2}$ et $\theta = 3$:

FIGURE 10.9 – Densité de probabilité de la loi Gamma



On donne également la fonction de répartition, pour les mêmes valeurs :

FIGURE 10.10 – Fonction de répartition de la loi Gamma



Théorème 10.6.1 (Stabilité par la somme). *Soient n variables aléatoires réelles mutuellement indépendantes X_1, \dots, X_n . On suppose :*

$$\mathcal{L}(X_k) = \Gamma(\alpha_k; \theta) .$$

Alors, la variable aléatoire réelle $\mathcal{X} := X_1 + \dots + X_n$ suit une loi Gamma. Plus exactement :

$$\mathcal{L}(\mathcal{X}) = \Gamma(\alpha_1 + \dots + \alpha_n; \theta) . \quad (10.1)$$

Démonstration. Il suffit de procéder par récurrence en montrant que $Y := X_1 + X_2$ suit la loi $\Gamma(\alpha_1 + \alpha_2; \theta)$. Pour ce faire, on utilise le produit de convolution :

$$f_Y(y) = f_{X_1} * f_{X_2}(y) = H(y) \int_0^y \frac{(y-t)^{\alpha_1-1} e^{-\frac{y-t}{\theta}}}{\Gamma(\alpha_1)\theta^{\alpha_1}} \frac{t^{\alpha_2-1} e^{-\frac{t}{\theta}}}{\Gamma(\alpha_2)\theta^{\alpha_2}} dt .$$

Puis, en faisant le changement de variable $t := zy$, il vient :

$$f_Y(y) = H(y) e^{-\frac{y}{\theta}} y^{\alpha_1+\alpha_2-1} \frac{\int_0^1 \frac{(1-z)^{\alpha_1-1} z^{\alpha_2-1}}{dz} z}{\theta^{\alpha_1+\alpha_2} \Gamma(\alpha_1)\Gamma(\alpha_2)} .$$

Il s'ensuit : $f_Y(y) = \frac{y^{\alpha-1} e^{-\frac{y}{\theta}}}{\Gamma(\alpha)\theta^\alpha} H(y) \times C$ où $\alpha := \alpha_1 + \alpha_2$ et où C est une constante. Comme $\int_{\mathbb{R}} f_Y(y) dy = 1$, on en déduit $C = 1$ et donc Y suit bien une loi Gamma de paramètres $\alpha_1 + \alpha_2$ et θ .

□

10.7 Loi du χ^2

Les lois du χ^2 à n degrés de liberté (où $n \in \mathbb{N}^*$) sont très importantes en statistiques dans le cadre du modèle gaussien.

Définition 10.7.1. Une variable aléatoire réelle X suit la loi $\chi^2(n)$ si elle est à densité et si

$$f_X(x) = \frac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}} \mathbb{1}_{]0;+\infty[}(x)$$

où Γ est la fonction Gamma d'Euler.

Remarque 10.7.2. La loi $\chi^2(n)$ est une instance particulière de la loi Gamma, avec $\alpha = \frac{n}{2}$ et $\theta = 2$.

Remarque 10.7.3. La loi $\chi^2(2)$ est en fait une loi exponentielle de paramètre $\lambda = \frac{1}{2}$ et de même la loi $\chi^2(2n)$ est une loi d'Erlang.

Proposition 10.7.4. Soit X une variable aléatoire réelle qui suit la loi du $\chi^2(n)$. Alors $\mathbb{E}[X] = n$.

Démonstration. Par définition : $\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) dx = \int_0^{\infty} \frac{x^{\frac{n}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}} dx$ d'où

$$\begin{aligned} \mathbb{E}[X] &= \frac{1}{\Gamma\left(\frac{n}{2}\right)} \int_0^{\infty} \left(\frac{x}{2}\right)^{\frac{n}{2}} e^{-\frac{x}{2}} dx \\ &= \frac{2}{\Gamma\left(\frac{n}{2}\right)} \int_0^{\infty} t^{\left(\frac{n}{2}+1\right)-1} e^{-t} dt \\ &= \frac{2}{\Gamma\left(\frac{n}{2}\right)} \Gamma\left(\frac{n}{2} + 1\right) \\ &= \frac{\frac{n}{2} 2 \Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \\ &= n, \end{aligned}$$

en utilisant les propriétés de la fonction Gamma d'Euler. □

Remarque 10.7.5. On rappelle que pour tout $t > 0$, on a

$$\Gamma(t+1) = t\Gamma(t).$$

On peut aussi calculer la variance en remarquant :

$$\begin{aligned}
\mathbb{E}[X^2] &= \int_{\mathbb{R}} x^2 f_X(x) dx \\
&= \int_0^{\infty} \frac{x^{\frac{n}{2}+1}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}} dx \\
&= \frac{4}{\Gamma\left(\frac{n}{2}\right)} \int_0^{\infty} t^{\frac{n}{2}+1} e^{-t} dt \\
&= \frac{4}{\Gamma\left(\frac{n}{2}\right)} \Gamma\left(\frac{n}{2} + 2\right) \\
&= \frac{4}{\Gamma\left(\frac{n}{2}\right)} \times \left(\frac{n}{2} + 1\right) \frac{n}{2} \Gamma\left(\frac{n}{2}\right) \\
&= n(n+2),
\end{aligned}$$

en utilisant les propriétés de la fonction Gamma d'Euler. Ceci nous amène au résultat suivant.

Proposition 10.7.6. *Soit X une variable aléatoire réelle qui suit la loi du $\chi^2(n)$. Alors $\text{Var}[X] = 2n$.*

Cette loi tire son origine de la loi normale comme l'atteste le théorème suivant.

Théorème 10.7.7. *Soient X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées suivant la loi normale centrée réduite. Alors $Y := X_1^2 + \dots + X_n^2$ suit la loi $\chi^2(n)$.*

Démonstration. D'après le Théorème 10.6.1 à la page 154, il est suffisant de montrer que X_i^2 suit la loi $\chi^2(1)$. En effet, la loi du $\chi^2(n)$ est une instance de la loi Gamma, comme noté dans la Remarque 10.7.2. On regarde donc d'abord la fonction de répartition de $Y := X_1^2$:

$$F_Y(y) := \mathbb{P}(Y \leq y) = \mathbb{P}(X_1^2 \leq y) .$$

D'une part, si $y < 0$, l'évènement $\{Y \leq y\}$ est impossible donc de probabilité nulle. Et, $\mathbb{P}(X_1^2 \leq 0) = \mathbb{P}(X_1 = 0) = 0$ car X_1 est une variable aléatoire à densité. Ainsi, $F_Y(y) = 0$ si $y \leq 0$. D'autre part, si $y > 0$.

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(X_1^2 \leq y) \\
&= \mathbb{P}(-\sqrt{y} \leq X_1 \leq \sqrt{y}) \\
&= \mathbb{P}(X_1 \leq \sqrt{y}) - \mathbb{P}(X_1 < -\sqrt{y}) \\
&= F_{X_1}(\sqrt{y}) - F_{X_1}(-\sqrt{y}) ,
\end{aligned}$$

vu que la variable X_1 est à densité. La fonction F_Y s'écrit donc

$$F_Y(y) = (F_{X_1}(\sqrt{y}) - F_{X_1}(-\sqrt{y})) \mathbb{1}_{]0;+\infty[}(y) .$$

La continuité sur $] - \infty; 0[$ est immédiate. De même, comme F_Y et $y \mapsto \pm\sqrt{y}$ sont continues, F_Y est continue sur $]0; +\infty[$. Finalement, $F_Y(0^-) = 0$ et $F_Y(0^+) = \frac{1}{2} - \frac{1}{2} = 0$ donc la fonction F_Y est continue sur \mathbb{R} .

On peut ensuite montrer facilement qu'elle est dérivable en tout point non nul de \mathbb{R} . Il s'ensuit que Y est bien une variable aléatoire à densité. Il nous suffit maintenant de dériver pour obtenir sa densité.

On considère $y > 0$:

$$\begin{aligned} f_Y(y) &:= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} (F_{X_1}(\sqrt{y}) - F_{X_1}(-\sqrt{y})) \\ &= \frac{d}{dy} \sqrt{y} \times f_{X_1}(\sqrt{y}) - \frac{d}{dy} (-\sqrt{y}) \times f_{X_1}(-\sqrt{y}) \\ &= \frac{1}{2\sqrt{y}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(\sqrt{y})^2}{2}\right] + \frac{1}{2\sqrt{y}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(-\sqrt{y})^2}{2}\right] \\ &= \frac{1}{\sqrt{2\pi y}} \exp\left[-\frac{y}{2}\right] \\ &= \frac{1}{2^{\frac{1}{2}} \Gamma\left(\frac{1}{2}\right)} y^{\frac{1}{2}-1} e^{-\frac{y}{2}}. \end{aligned}$$

Et, quand $y \leq 0$,

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} 0 \\ &= 0. \end{aligned}$$

Par conséquent, pour tout $y \in \mathbb{R}$, on a

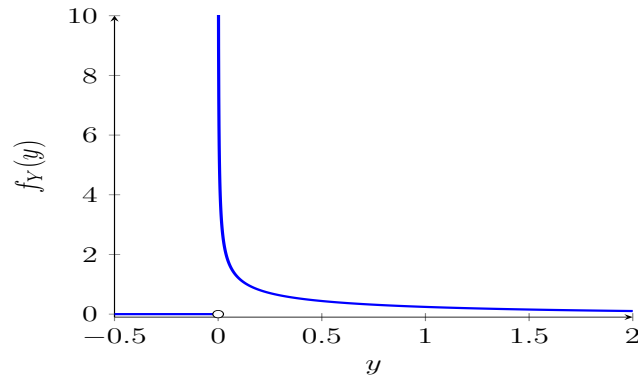
$$f_Y(y) = \frac{1}{2^{\frac{1}{2}} \Gamma\left(\frac{1}{2}\right)} y^{\frac{1}{2}-1} e^{-\frac{y}{2}} \mathbf{1}_{]0; +\infty[}(y),$$

si bien que Y suit la loi $\chi^2(1)$ et la preuve s'ensuit. □

On en déduit notamment que la somme de deux variables aléatoires indépendantes X_1 et X_2 , suivant respectivement les lois $\chi^2(n_1)$ et $\chi^2(n_2)$, suit la loi $\chi^2(n_1 + n_2)$.

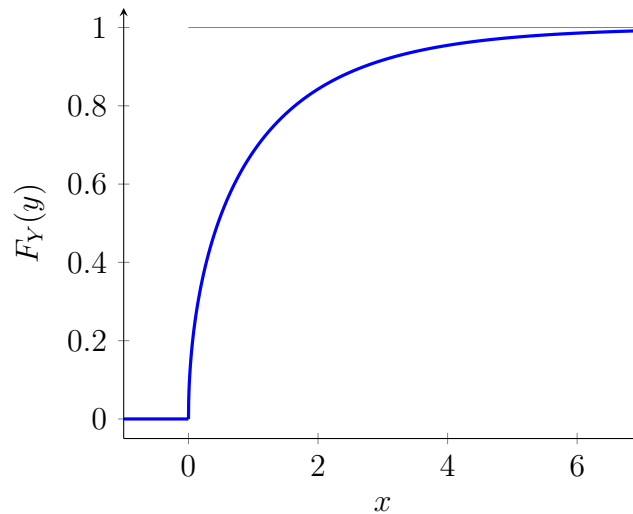
On trace la densité de probabilité pour $n = 1$:

FIGURE 10.11 – Densité de probabilité de la loi du Khi-deux



On donne également la fonction de répartition :

FIGURE 10.12 – Fonction de répartition de la loi du Khi-deux



On note que les tables des quantiles de la loi du χ^2 sont à la page 367.

10.8 Loi de Rayleigh

La loi de Rayleigh dérive de la loi normale. En effet, soient deux variables aléatoires indépendantes et identiquement distribuées suivant la loi $\mathcal{N}(0, \sigma^2)$ où $\sigma > 0$.

Alors, la variable aléatoire réelle $R := \sqrt{X^2 + Y^2}$ suit une loi de Rayleigh. En d'autres termes, la loi de Rayleigh correspond à $\sigma\sqrt{Z}$ où Z suit la loi du Khi-deux à deux degrés de liberté. On peut ainsi établir facilement sa densité de probabilité.

D'abord, $F_R(r) = 0$ si $r \leq 0$. Puis, si $r > 0$, on a :

$$F_R(r) = \mathbb{P}(R \leq r) = \mathbb{P}\left(\sqrt{Z} \leq \frac{r}{\sigma}\right) = \mathbb{P}\left(Z \leq \frac{r^2}{\sigma^2}\right) = F_Z\left(\frac{r^2}{\sigma^2}\right),$$

où Z suit une loi $\chi^2(2)$.

On en déduit pour $r > 0$:

$$f_R(r) = \frac{d}{dr}F_R(r) = \frac{2r}{\sigma^2}f_Z\left(\frac{r^2}{\sigma^2}\right) = \frac{2r}{\sigma^2} \frac{1}{2\Gamma(1)} \left(\frac{r^2}{\sigma^2}\right)^{\frac{2}{2}-1} e^{-\frac{r^2}{2\sigma^2}} = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}.$$

On aboutit ainsi à

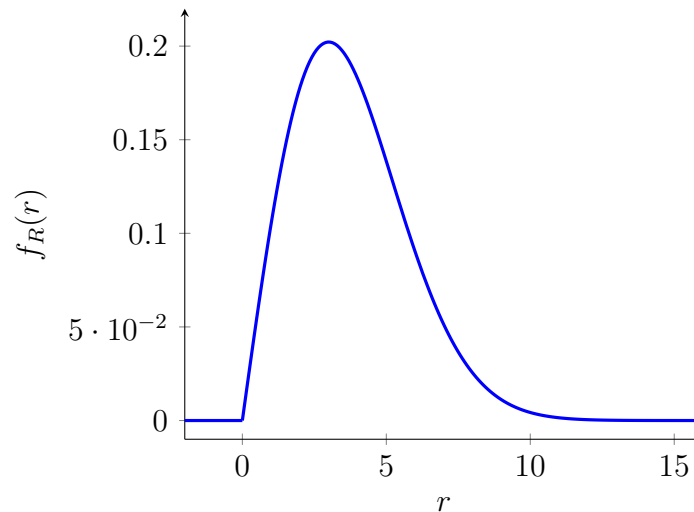
$$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} H(r),$$

pour tout $r \in \mathbb{R}$.

Notons que la loi de Rayleigh est un cas particulier de la loi de Weibull, voir la page 152.

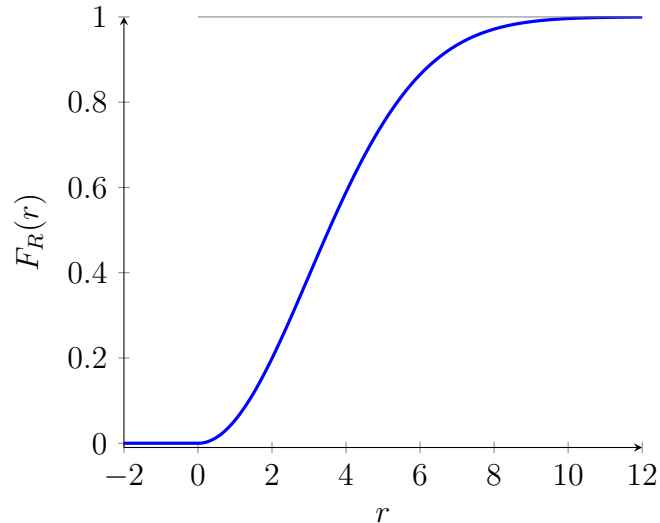
On trace sa densité de probabilité pour $\sigma^2 = 9$:

FIGURE 10.13 – Densité de probabilité de la loi de Rayleigh



On donne également la fonction de répartition pour la même valeur :

FIGURE 10.14 – Fonction de répartition de la loi de Rayleigh



10.9 Loi log-normale

Comme son nom l'indique, la loi log-normale découle de la loi normale. Soit Y une variable aléatoire qui suit la loi $\mathcal{N}(m, \sigma^2)$. On pose $L := e^Y$. On dit que L suit la loi log-normale de paramètres m et σ^2 . On peut calculer la densité de probabilité de L . Pour ce faire, on regarde la fonction de répartition de L :

$$\begin{aligned} F_L(\ell) &:= \mathbb{P}(L \leq \ell) \\ &= \mathbb{P}(e^Y \leq \ell) . \end{aligned}$$

D'une part, si $\ell \leq 0$, l'évènement $\{e^Y \leq \ell\}$ est impossible d'où $F_L(\ell) = 0$. D'autre part, si ℓ est strictement positif :

$$\begin{aligned} F_L(\ell) &= \mathbb{P}(e^Y \leq \ell) \\ &= \mathbb{P}(Y \leq \log(\ell)) \\ &= F_Y(\log(\ell)) . \end{aligned}$$

Ainsi, $F_L(\ell) = F_Y(\log(\ell)) \mathbf{1}_{]0; +\infty[}(\ell)$. La continuité sur $]0; +\infty[$ et sur $] -\infty; 0[$ est immédiate. La continuité en 0 découle de :

$$\lim_{\ell \rightarrow 0^+} F_L(\ell) = \lim_{\ell \rightarrow 0^+} F_Y(\log(\ell)) = F_Y(-\infty) = 0 = F_L(0) .$$

De même, on peut montrer la dérivabilité en tout point non nul de \mathbb{R} . On dérive maintenant et on obtient pour $\ell > 0$:

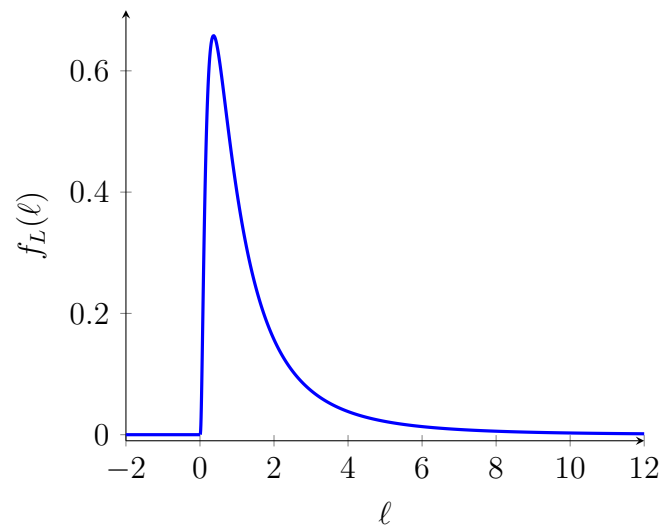
$$f_L(\ell) = \frac{1}{\ell} f_Y(\log(\ell)) = \frac{1}{\sqrt{2\pi\sigma^2\ell^2}} e^{-\frac{(\log(\ell)-m)^2}{2\sigma^2}} .$$

Par conséquent, on a pour tout $\ell \in \mathbb{R}$:

$$f_L(\ell) = \frac{1}{\sqrt{2\pi\sigma^2\ell^2}} e^{-\frac{(\log(\ell)-m)^2}{2\sigma^2}} \mathbb{1}_{]0;+\infty[}(\ell).$$

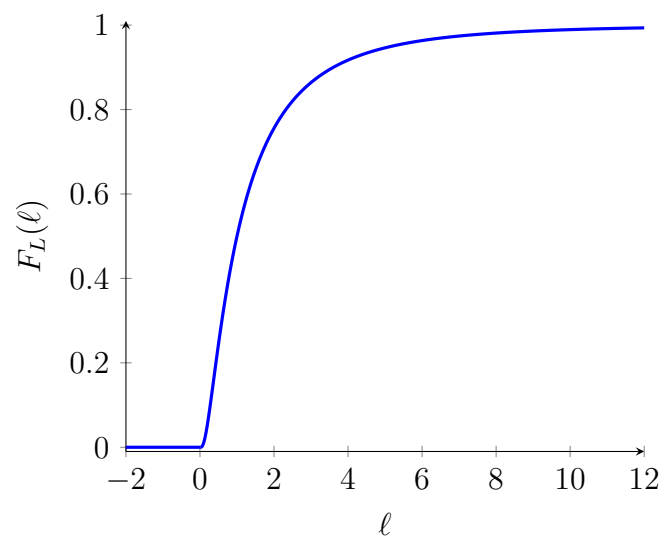
Cette loi trouve son utilité en biologie comme en analyse quantitative. On trace la densité de probabilité pour $m = 0$ et $\sigma^2 = 1$:

FIGURE 10.15 – Densité de probabilité de la loi lognormale



On donne également la fonction de répartition avec les mêmes paramètres :

FIGURE 10.16 – Fonction de répartition de la loi lognormale



10.10 Loi de Student

Les lois de Student sont très utilisées pour les intervalles de confiance en statistiques, dans le cas du modèle gaussien. Ainsi, la loi dérive de la loi gaussienne, comme dans le cas de la loi $\chi^2(n)$. Et, nous allons d'abord présenter sa densité avant de montrer le lien qu'elle entretient avec la loi normale (et la loi $\chi^2(n)$).

Dans la suite de cette section, n est un entier naturel strictement positif.

Définition 10.10.1. Une variable aléatoire réelle T suit la loi $\mathcal{T}(n)$ si elle est à densité et si

$$f_T(t) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}},$$

où Γ est la fonction Gamma d'Euler.

La parité de f_T implique la nullité de tous les moments d'ordres impairs, pour peu que ceux-ci soient bien définis. On peut calculer facilement sa variance si $n \geq 3$.

Proposition 10.10.2. Soit T une variable aléatoire réelle qui suit la loi $\mathcal{T}(n)$ avec $n \geq 3$. Alors $\text{Var}[T] = \frac{n}{n-2}$.

Démonstration. D'abord, $\mathbb{E}[T] = 0$. Il suffit pour s'en convaincre de vérifier que l'espérance est bien définie. Or, $n \geq 3$ donc la fonction $t \mapsto |t|(1 + \frac{t^2}{n})^{-\frac{n+1}{2}} = O(|t|^{-3})$. Cette fonction est bien intégrable et la nullité de l'espérance de T s'ensuit.

Ensuite :

$$\mathbb{E}[T^2] = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \int_{-\infty}^{\infty} t^2 \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt$$

Focalisons-nous sur l'intégrale :

$$\begin{aligned} & \int_{-\infty}^{\infty} t^2 \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt \\ &= 2 \int_0^{\infty} t^2 \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt \\ &= \int_0^{\infty} nt \frac{2t}{n} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt \\ &= \left[nt \frac{\left(1 + \frac{t^2}{n}\right)^{-\frac{n-1}{2}}}{-\frac{n-1}{2}} \right]_0^{\infty} - \int_0^{\infty} n \frac{\left(1 + \frac{t^2}{n}\right)^{-\frac{n-1}{2}}}{-\frac{n-1}{2}} dt \\ &= \frac{2n}{n-1} \int_0^{\infty} \left(1 + \frac{t^2}{n}\right)^{-\frac{n-1}{2}} dt, \end{aligned}$$

après intégration par parties. On procède ensuite au changement de variable $t := \sqrt{\frac{n}{n-2}}u$ et l'on obtient

$$\begin{aligned} & \int_{-\infty}^{\infty} t^2 \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt \\ &= \frac{2n}{n-1} \sqrt{\frac{n}{n-2}} \int_0^{\infty} \left(1 + \frac{u^2}{n-2}\right)^{-\frac{n-2+1}{2}} du. \end{aligned}$$

On reconnaît la loi $\mathcal{T}(n-2)$ d'où l'intégrale est égale à

$$\frac{n\sqrt{n}}{(n-1)\sqrt{n-2}} \times \frac{\sqrt{(n-2)\pi}\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)},$$

et par conséquent le second moment de T est

$$\begin{aligned} \mathbb{E}[T^2] &= \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{n\sqrt{n}}{(n-1)\sqrt{n-2}} \times \frac{\sqrt{(n-2)\pi}\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \\ &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{n}{(n-1)} \times \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \\ &= \frac{\frac{n-1}{2}\Gamma\left(\frac{n-1}{2}\right)}{\left(\frac{n}{2}-1\right)\Gamma\left(\frac{n-2}{2}\right)} \frac{n}{n-1} \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \\ &= \frac{n}{n-2}, \end{aligned}$$

ce qui achève la preuve, vu que $\text{Var}[T] = \mathbb{E}[T^2] - \mathbb{E}[T]^2$. □

On va maintenant montrer le lien que la loi de Student entretient avec la loi normale.

Théorème 10.10.3. *Soit $n \geq 1$. Soit Z une variable aléatoire qui suit la loi normale centrée réduite. Soit U une variable aléatoire indépendante de Z et telle que $\mathcal{L}(U) = \chi^2(n)$. Alors, la variable aléatoire*

$$T := \frac{Z}{\sqrt{\frac{U}{n}}}$$

suit la loi de Student à n degrés de liberté.

Démonstration. On pose $S := \sqrt{\frac{U}{n}}$. Alors, on peut calculer sa fonction de répartition :

$$F_S(s) = F_U(ns^2).$$

Puis, en dérivant, on obtient $f_S(s) = 2nsf_U(ns^2) = \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} s^{n-1} e^{-\frac{ns^2}{2}} H(s)$.

On regarde maintenant la fonction de répartition de T :

$$\begin{aligned} F_T(t) &= \mathbb{P}(T \leq t) \\ &= \mathbb{P}\left(\frac{Z}{S} \leq t\right) \\ &= \mathbb{P}(Z \leq tS), \end{aligned}$$

car $S > 0$ presque sûrement. On utilise ensuite les résultats sur les vecteurs aléatoires du chapitre qui s'y rapporte et que nous n'avons pas encore vus. Plus précisément, on utilise d'abord le Théorème 12.8.1 ainsi que la Définition 12.7.1. On obtient :

$$\mathbb{P}(T \leq t) = \int_{s=0}^{s=+\infty} \left(\int_{z=-\infty}^{z=ts} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \right) \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} s^{n-1} e^{-\frac{ns^2}{2}} ds.$$

On dérive ensuite par rapport à t (en utilisant le théorème de dérivation sous le signe somme) et l'on trouve :

$$f_T(t) = \int_0^{+\infty} \frac{s}{\sqrt{2\pi}} e^{-\frac{t^2s^2}{2}} \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} s^{n-1} e^{-\frac{ns^2}{2}} ds.$$

Ainsi, on a

$$f_T(t) = \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})\sqrt{2\pi}} I(t),$$

où $I(t) := \int_0^\infty s^n e^{-\frac{1}{2}(n+t^2)s^2} ds$. On procède au changement de variable $s := \sqrt{\frac{u}{n+t^2}}$ et l'on aboutit à

$$I(t) = \frac{1}{2(n+t^2)^{\frac{n+1}{2}}} \int_0^\infty u^{\frac{n+1}{2}-1} e^{-\frac{u}{2}} du.$$

On reconnaît la loi $\chi^2(n+1)$ à savoir la loi du Khi-deux à $n+1$ degrés de liberté dans l'intégrale d'où

$$I(t) = \frac{1}{2(n+t^2)^{\frac{n+1}{2}}} 2^{\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right).$$

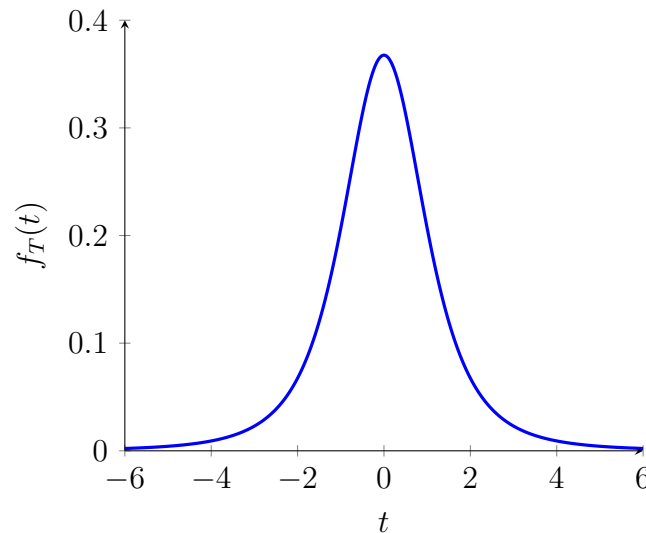
Conséquemment, on a

$$\begin{aligned}
 f_T(t) &= \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1} \Gamma\left(\frac{n}{2}\right) \sqrt{2\pi}} \frac{1}{2(n+t^2)^{\frac{n+1}{2}}} 2^{\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right) \\
 &= \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}},
 \end{aligned}$$

ce qui achève de prouver que T suit la loi $\mathcal{T}(n)$. □

On trace la densité de probabilité pour $n = 3$:

FIGURE 10.17 – Densité de probabilité de la loi de Student



Les quantiles de la loi de Student ont été tabulés, voir page 366

10.11 Loi de Fisher

La loi de Fisher est utilisée pour des tests statistiques, voir page 317.

Elle se définit comme suit.

Soient d_1 et d_2 deux entiers naturels strictement positifs. On suppose que X_1 suit la loi $\chi^2(d_1)$ et que X_2 suit la loi $\chi^2(d_2)$. On les suppose indépendantes.

Alors, la variable aléatoire

$$X := \frac{X_1/d_1}{X_2/d_2}$$

suit la loi de Fisher avec paramètres d_1 et d_2 . Et, la densité de probabilité de X est la suivante :

$$f_X(x) = \frac{1}{xB\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \left(\frac{d_1x}{d_1x + d_2}\right)^{\frac{d_1}{2}} \left(\frac{d_2}{d_1x + d_2}\right)^{\frac{d_2}{2}},$$

où la fonction B est la fonction bêta définie par $B(x, y) := \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$. Les quantiles d'ordres 0.975 (voir page 368) et 0.95 (voir page 369) ont été tabulés.

10.12 Loi de Pareto

La distribution de Pareto est utilisée pour modéliser la répartition des salaires au sein d'une société. Elle l'est également en réassurance. Elle se définit par sa fonction de survie, laquelle a deux paramètres : $x_0 > 0$ et $\alpha > 0$. Ainsi, si X suit la loi de Pareto de paramètres x_0 et α , on a pour tout $x \geq x_0$:

$$\mathbb{P}(X > x) = \left(\frac{x_0}{x}\right)^\alpha.$$

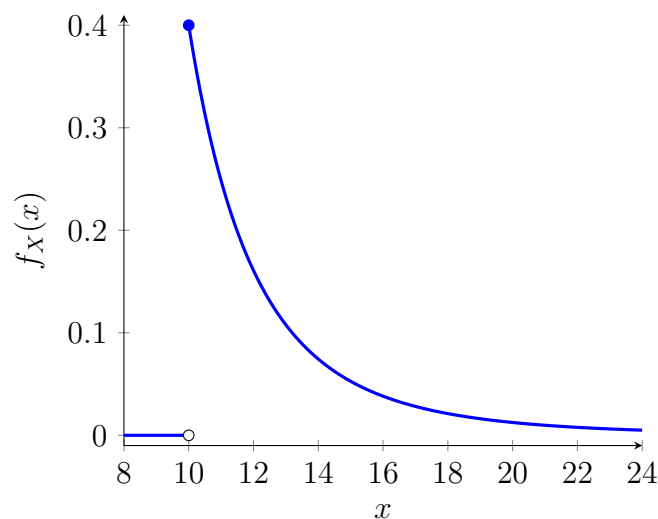
On en déduit immédiatement $\mathbb{P}(X > x_0) = 1$ d'où $F_X(x_0) = 0$. Puis, l'on peut obtenir la valeur de $F_X(x)$ pour tout $x \in \mathbb{R}$: $F_X(x) = \left(1 - \frac{x_0^\alpha}{x^\alpha}\right) \mathbb{1}_{[x_0; +\infty[}(x)$.

En procédant ensuite à une dérivation, il vient :

$$f_X(x) = \alpha \frac{x_0^\alpha}{x^{\alpha+1}} \mathbb{1}_{[x_0; +\infty[}(x).$$

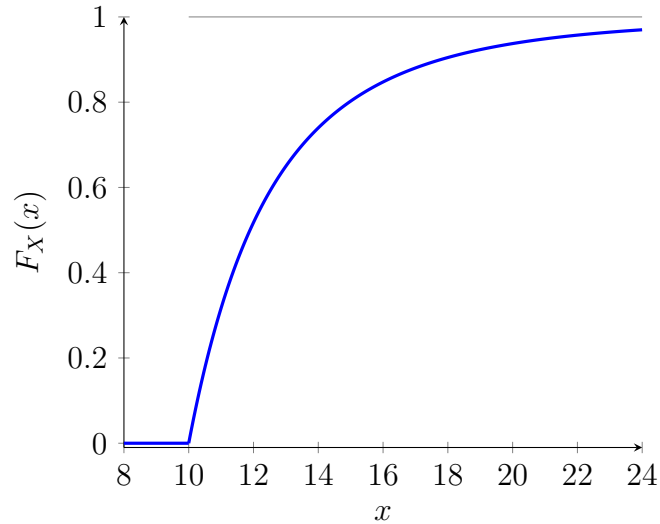
On trace la densité de probabilité pour $x_0 = 10$ et $\alpha = 4$:

FIGURE 10.18 – Densité de probabilité de la loi de Pareto



On donne également la fonction de répartition avec les mêmes paramètres :

FIGURE 10.19 – Fonction de répartition de la loi de Pareto



10.13 Loi de Cauchy

On peut parler de la loi de Cauchy sur toute la droite des réels. Soient $a \in \mathbb{R}$ et $b > 0$. La loi de Cauchy de paramètres a et b , $\mathcal{C}(a, b)$ admet pour densité de probabilité :

$$f(x) := \frac{1}{\pi} \frac{b}{b^2 + (x - a)^2}.$$

Il convient de noter que la loi de Cauchy est en fait la loi de Student à un degré de liberté si $a = 0$ et $b = 1$.

Cette loi n'admet pas d'espérance, ni de variance finie. Elle intervient naturellement quand on considère l'exemple suivant.

Exemple 10.13.1. *Un gyrophare envoie un flash lumineux dans une direction aléatoire θ . On place un écran de longueur infinie à une distance 1 du gyrophare. On cherche la distribution de l'abscisse du point d'impact du rayon lumineux sur l'écran.*

On sait que l'angle θ est une variable aléatoire uniformément distribuée sur l'intervalle $[-\frac{\pi}{2}; \frac{\pi}{2}]$. L'abscisse est donnée par $X := \tan(\theta)$.

On a donc

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(\tan(\theta) \leq x) \\ &= \mathbb{P}(\theta \leq \arctan(x)) \\ &= F_\theta(\arctan(x)) . \end{aligned}$$

Or,

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) \\ &= \frac{d}{dx} F_\theta(\arctan(x)) \\ &= \frac{1}{1+x^2} f_\theta(\arctan(x)) \\ &= \frac{1}{\pi} \frac{1}{1+x^2} . \end{aligned}$$

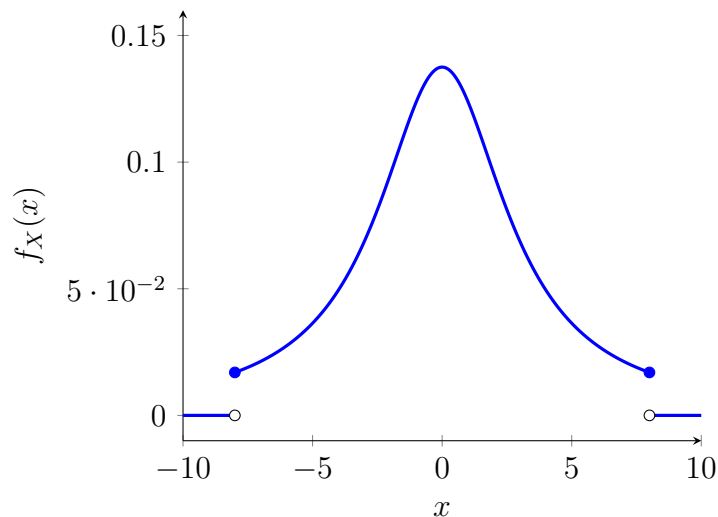
Il s'agit de la loi $\mathcal{C}(0, 1)$.

On peut par ailleurs restreindre la loi de Cauchy à un intervalle borné.

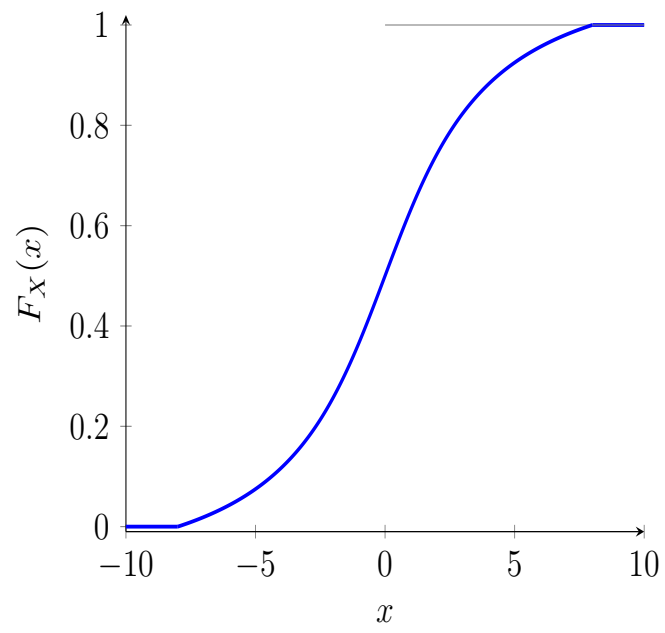
Par exemple, on peut considérer la loi ayant la densité de probabilité suivante :

$$f_X(x) = \frac{1}{2 \arctan\left(\frac{a}{b}\right)} \frac{b}{x^2 + b^2} \mathbb{1}_{[-a; a]}(x) .$$

Voici par ailleurs le graphe de cette dernière fonction pour $a = 8$ et $b = 3$:



Et, la fonction de répartition associée est :



Fonctions caractéristiques

11.1 Introduction

L'objet du présent chapitre est de présenter la fonction caractéristique d'une variable aléatoire, ou plutôt d'une mesure de probabilité d'une variable aléatoire.

Il s'agit de la transformée de Fourier d'une loi. Celle-ci est essentielle en traitement du signal car elle permet une analyse fréquentielle plutôt qu'une analyse temporelle. Dit autrement, cela consiste à appliquer une transformation pour se placer dans un nouvel espace afin d'adopter un nouveau point de vue.

Certaines transformations facilitent ainsi l'obtention de résultats d'une façon plus élégante. Ainsi, la transformation de Laplace ou la transformation de Fourier sont des outils puissants. Elles le sont également en probabilités. Par exemple, la transformation de Laplace joue un grand rôle quand on étudie les grandes déviations, ce que nous ne ferons pas dans ce livre.

Dans tout ce chapitre, X est une variable aléatoire réelle. On ne présuppose pas qu'elle soit discrète ni qu'elle soit à densité.

Les pré-requis de ce chapitre sont : séries entières (et donc séries de fonctions), intégration, dérivation. Il est également préférable de connaître la transformation de Fourier, quoique ce ne soit pas nécessaire. De l'analyse complexe serait un plus indéniable de même que la théorie des distributions.

Les objectifs du chapitre sont la découverte des fonctions caractéristiques. Notamment, il est attendu du lecteur qu'il connaisse la définition d'une fonction caractéristique, qu'il soit en mesure de la calculer si la loi est discrète ou si elle est à densité. Également, le lecteur doit être en mesure de pouvoir calculer les moments (à condition que la transformée de Fourier soit suffisamment régulière) de la variable aléatoire X . Par ailleurs, certaines propriétés sont à maîtriser : notamment la fonction caractéristique de la somme de deux variables indépendantes (qui en est le produit vu que la transformée de Fourier transforme le produit de convolution en produit simple), le fait que la fonction caractéristique caractérise la loi ainsi que la possibilité d'inverser la transformée de Fourier pour peu que la fonction caractéristique soit sommable. Enfin, il est plus que souhaitable que les fonctions caractéristiques des lois usuelles soient connues à l'issue de la lecture de ce chapitre ; en particulier les fonctions caractéristiques de la loi de Bernoulli, de la loi binomiale, de la loi de Poisson, de la loi uniforme continue, de la loi

exponentielle et de la loi normale (centrée et réduite ou non).

11.2 Définition et premières propriétés

Définition 11.2.1. *La fonction caractéristique de la variable aléatoire X est la fonction φ_X de la variable réelle à valeurs complexes définie par*

$$\varphi_X(u) := \mathbb{E}(e^{iuX}) .$$

Il s'agit en fait de la transformée de Fourier *inverse* de la distribution associée à X , \mathbb{P}_X . On admet l'existence de cette fonction. En effet, il faudrait utiliser la théorie de la mesure pour prouver que la fonction caractéristique existe quelle que soit la mesure de probabilité μ de la variable aléatoire réelle X .

Il convient également de noter que la transformée de Fourier (inverse ou non) en traitement du signal est parfois prise avec la convention suivante :

$$\hat{f}(u) = \int_{\mathbb{R}} e^{-2i\pi ut} f(t) dt ,$$

pour $f \in L^1$. Dans le cadre probabiliste, on ne met pas le facteur 2π dans l'exponentielle complexe. En revanche, il faudra alors renormaliser par 2π pour retrouver la mesure de probabilité à partir de la fonction caractéristique, voir le Théorème d'inversion à la page 181.

On peut montrer facilement l'existence de la fonction φ_X si X est discrète ou si X est à densité. Faisons-le d'abord avec X discrète.

Proposition 11.2.2. *Si la variable aléatoire X est discrète, on peut écrire*

$$\varphi_X(u) = \sum_{k \in I} e^{iux_k} \mathbb{P}(X = x_k) ,$$

où $\{x_k, k \in I\} = X(\Omega)$ est l'ensemble des réalisations possibles de X .

Il s'agit de la transformée de Fourier inverse de la distribution

$$\sum_{k \in I} \mathbb{P}(X = x_k) \delta_{x_k} .$$

L'existence de la fonction caractéristique est ici prouvée puisque l'on a la convergence absolue de la série de terme général $u \mapsto g_k(u) := e^{iux_k} \mathbb{P}(X = x_k)$. En effet, pour tout k , $|e^{iux_k}| = 1$. Puis, comme \mathbb{P}_X est une probabilité sur $X(\Omega)$, la convergence de la série s'ensuit.

On donne maintenant l'écriture lorsque X est à densité.

Proposition 11.2.3. *Si la variable aléatoire X admet une densité f_X , on peut écrire*

$$\varphi_X(u) = \int_{\mathbb{R}} e^{iux} f_X(x) dx.$$

Il s'agit de la transformée de Fourier inverse de la distribution régulière associée à la fonction f_X . L'existence de la fonction caractéristique est ici prouvée puisque l'on a la convergence absolue de l'intégrale de la fonction $x \mapsto \psi_u(x) := e^{iux} f_X(x)$. En effet, $|\psi_u(x)| = f_X(x)$ pour tous $u, x \in \mathbb{R}$.

L'un des intérêts de la fonction caractéristique est sa régularité. On suppose maintenant que X est discrète ou à densité.

Théorème 11.2.4. *Soit X une variable aléatoire réelle. On la suppose discrète ou à densité. Alors la fonction caractéristique de X , φ_X , est bornée de module inférieur à 1, continue et de plus $\varphi_X(0) = 1$.*

Démonstration. On effectue d'abord la preuve dans le cas où X est à densité. Comme le module de e^{iux} est égal à 1, l'inégalité triangulaire nous donne

$$|\varphi_X(u)| = \left| \int_{\mathbb{R}} e^{iux} f_X(x) dx \right| \leq \int_{\mathbb{R}} |e^{iux}| f_X(x) dx = 1.$$

On a donc bien la bornitude de $|\varphi_X|$ par 1.

On calcule maintenant $\varphi_X(0)$ comme suit :

$$\varphi_X(0) = \mathbb{E}(e^{i \times 0 \times X}) = \mathbb{E}(1) = 1.$$

Enfin, on prouve la continuité en utilisant le théorème de convergence dominée de Lebesgue. Soit $(u_n)_n$ une suite réelle qui converge vers $u \in \mathbb{R}$. La suite de fonctions $(x \mapsto e^{iu_n x} f_X(x))_n$ converge simplement vers la fonction $x \mapsto e^{iux} f_X(x)$. De plus, pour tout $n \in \mathbb{N}$, la fonction $x \mapsto e^{iu_n x} f_X(x)$ est bornée par la fonction f_X qui est intégrable sur \mathbb{R} . On en déduit la convergence de

$$\varphi_X(u_n) = \int_{\mathbb{R}} e^{iu_n x} f_X(x) dx$$

vers

$$\int_{\mathbb{R}} e^{iux} f_X(x) dx = \varphi_X(u).$$

Ceci correspond à la continuité de la fonction φ_X .

On va maintenant établir le théorème dans le cas où X est discrète, que $X(\Omega)$ soit fini ou infini dénombrable. D'abord, pour tout $k \in I$, on pose $g_k(u) := e^{iux_k} \mathbb{P}(X = x_k)$. Alors, $\|g_k\|_{\infty} = \mathbb{P}(X = x_k)$. La série $\sum_{k \in I} \|g_k\|_{\infty}$ converge donc on en déduit immédiatement, par application de l'inégalité triangulaire, que $|\varphi_X(u)| \leq 1$. De plus, $\varphi_X(0) = \sum_{k \in I} \mathbb{P}(X = x_k) = 1$. La continuité est immédiate après application du théorème de continuité des séries absolument convergentes pour la norme infinie.

□

Remarque 11.2.5. *Un des avantages de la théorie de la mesure est que l'on peut regrouper les variables aléatoires réelles discrètes, les variables aléatoires réelles à densité et les autres variables aléatoires réelles en une seule classe puis prouver le théorème de manière générale en écrivant*

$$\varphi_X(u) = \int_{\mathbb{R}} e^{iux} \mu(dx).$$

Il s'agit toutefois d'une théorie difficile que nous n'abordons pas ici. Quelques éléments sont donnés au Chapitre 24 à la page 341.

En fait, on pourrait même montrer que la fonction φ_X est uniformément continue.

La fonction caractéristique est dérivable si l'on suppose de plus que la variable aléatoire réelle X admet un moment d'ordre un. De manière plus générale, on a le théorème suivant.

Théorème 11.2.6. *Soit X une variable aléatoire réelle admettant des moments jusqu'à l'ordre m , c'est-à-dire telle que $\max_{1 \leq k \leq m} \mathbb{E}[|X|^k] < \infty$. On la suppose discrète ou à densité. La fonction caractéristique associée, φ_X , est alors de classe \mathcal{C}^m et de plus, pour tout $k \in \llbracket 1; m \rrbracket$, on a*

$$\frac{d^k}{du^k} \varphi_X(u) = i^k \mathbb{E}[X^k e^{iuX}].$$

Démonstration. On va d'abord prouver ce résultat dans le cas où X est à densité. La fonction $u \mapsto e^{iux} f_X(x)$ est de classe \mathcal{C}^∞ et sa dérivée k -ième est égale à

$$u \mapsto i^k x^k e^{iux} f_X(x)$$

pour tout $k \in \mathbb{N}$. Par ailleurs, la fonction $x \mapsto i^k x^k e^{iux} f_X(x)$ est bornée par la fonction $x \mapsto |x|^k f_X(x)$. Or, comme $\max_{1 \leq k \leq m} \mathbb{E}[|X|^k] < \infty$, on en déduit que les intégrales $\int_{\mathbb{R}} |x|^k f_X(x) dx$ sont convergentes pour $k \in \llbracket 1; m \rrbracket$. On applique alors le théorème de dérivation sous le signe somme et l'on en déduit que la fonction φ_X est dérivable k fois de dérivée égale à $\int_{\mathbb{R}} i^k x^k e^{iux} f_X(x) dx$. Par définition de la densité de probabilité, on en déduit que la dérivée k -ième de φ_X , pour $k \in \llbracket 1; m \rrbracket$, est égale à

$$i^k \mathbb{E}[X^k e^{iuX}].$$

On va maintenant établir le théorème dans le cas où X est discrète. Comme X admet un moment d'ordre m , $\mathbb{E}[|X|^m] < \infty$. En d'autres termes, on a

$$\sum_{k \in I} |x_k|^m \mathbb{P}(X = x_k) < \infty.$$

On en déduit immédiatement que la série de fonctions de terme général $u \mapsto i^m x_k^m e^{iux} \mathbb{P}(X = x_k)$ converge absolument pour la norme infinie. De fait, la série

converge. On en déduit immédiatement que la série de fonctions qui définit la fonction caractéristique est de classe \mathcal{C}^m et de plus $\varphi_X^{(p)}(u) = i^p \mathbb{E}[X^p e^{iuX}]$ pour tout $p \in \llbracket 1; m \rrbracket$.

□

Une application immédiate du Théorème 11.2.6 est l'utilisation de la fonction caractéristique pour calculer les moments d'une variable aléatoire réelle. En effet, si les quantités suivantes ont un sens, il vient

$$\varphi'_X(0) = i\mathbb{E}[X]$$

ainsi que

$$\varphi''_X(0) = -\mathbb{E}[X^2].$$

Il en découle $\text{Var}(X) = -\varphi''_X(0) + (\varphi'_X(0))^2$.

11.3 Fonctions caractéristiques des lois usuelles

Rappelons que l'un des intérêts de la fonction caractéristique est sa continuité, même si la variable aléatoire est discrète.

On donne ici les fonctions caractéristiques de la plupart des lois usuelles présentées dans les Chapitres 5, 6, 9 et 10.

11.3.1 Lois discrètes

11.3.1.1 Loi de Bernoulli

On suppose que X_1 suit la loi de Bernoulli de paramètre p c'est-à-dire que l'on a $\mathbb{P}_{X_1} = (1-p)\delta_0 + p\delta_1$. La fonction caractéristique associée est donc

$$\varphi_{X_1}(u) = \mathbb{E}[e^{iuX_1}] = (1-p)e^{iu \times 0} + pe^{iu \times 1} = (1-p) + pe^{iu}.$$

11.3.1.2 Loi binomiale

On suppose que X_2 suit la loi binomiale de paramètres n et p c'est-à-dire que l'on a

$$\mathbb{P}_{X_2} = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \delta_k.$$

Le calcul de la fonction caractéristique donne donc

$$\begin{aligned} \varphi_{X_2}(u) &= \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} e^{iuk} \\ &= \sum_{k=0}^n \binom{n}{k} (pe^{iu})^k (1-p)^{n-k} \\ &= ((1-p) + pe^{iu})^n. \end{aligned}$$

Remarquons que l'on dispose de l'égalité

$$\varphi_{X_2}(u) = (\varphi_{X_1}(u))^n,$$

qui traduit le fait que X_2 suit la loi binomiale de paramètres n et p si et seulement si X_2 est la somme de n variable aléatoires réelles indépendantes suivant la loi de Bernoulli de paramètre p .

11.3.1.3 Loi de Poisson

On suppose que X_3 suit la loi de Poisson de paramètre $\lambda > 0$ c'est-à-dire que l'on a

$$\mathbb{P}_{X_3} = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \delta_n.$$

On peut facilement montrer :

$$\varphi_{X_3}(u) = e^{\lambda(e^{iu}-1)}.$$

11.3.1.4 Loi de Rademacher

Ici, $\mathbb{P}_{X_4} = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ d'où $\varphi_{X_4}(u) = \frac{1}{2}(e^{-iu} + e^{iu}) = \cos(u)$.

On observe en particulier que la loi de Rademacher étant symétrique, sa fonction caractéristique est réelle.

11.3.1.5 Loi uniforme discrète

Prenons le cas où X_5 suit la loi uniforme discrète sur $\llbracket 1; N \rrbracket$. Alors :

$$\varphi_{X_5}(u) = \sum_{k=1}^N \frac{1}{N} e^{iuk} = \frac{1}{N} \frac{e^{iu} - e^{iuN}}{1 - e^{iu}},$$

pour $u \neq 0$.

11.3.1.6 Loi triangulaire discrète

Dans le cas où X_6 suit la loi triangulaire discrète sur $\llbracket 2; 2N \rrbracket$, on peut montrer facilement que l'on a :

$$\varphi_{X_6}(u) = \frac{1}{N^2} \frac{(e^{iu} - e^{iuN})^2}{(1 - e^{iu})^2},$$

pour $u \neq 0$.

11.3.1.7 Loi géométrique

Ici, $\mathbb{P}_{X_7} = \sum_{n=1}^{\infty} p(1-p)^{n-1} \delta_n$. D'où

$$\varphi_{X_7}(u) = \frac{pe^{iu}}{1 - (1-p)e^{iu}}.$$

11.3.2 Lois à densité

11.3.2.1 Loi uniforme

Dans ce paragraphe, X_8 est une variable aléatoire réelle qui suit la loi uniforme sur l'intervalle $[a; b]$ ($a < b$) c'est-à-dire que l'on a

$$f_{X_8}(x) = \frac{1}{b-a} \mathbb{1}_{[a;b]}(x).$$

Le calcul nous donne alors

$$\varphi_{X_8}(u) = \frac{1}{b-a} \int_a^b e^{iux} dx = \frac{1}{b-a} \frac{e^{iub} - e^{iua}}{iu}$$

si $u \neq 0$ et $\varphi_{X_8}(0) = 1$. En particulier, si on considère la loi uniforme sur l'intervalle $[-a; a]$ avec $a > 0$, il vient

$$\varphi_{X_8}(u) = \frac{\sin(ua)}{ua},$$

pour $u \neq 0$. À nouveau, si l'intervalle est symétrique alors la loi est symétrique d'où la fonction caractéristique est réelle.

11.3.2.2 Loi exponentielle

Ici, la variable aléatoire X_9 suit la loi exponentielle de paramètre $\lambda > 0$. En d'autres termes, on a

$$f_{X_9}(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0; +\infty[}(x).$$

On procède maintenant au calcul :

$$\varphi_{X_9}(u) = \lambda \int_0^{\infty} e^{iux} e^{-\lambda x} dx = \frac{\lambda}{\lambda - iu}.$$

11.3.2.3 Loi normale centrée réduite

Dans ce paragraphe, on se donne une variable aléatoire réelle X_{10} qui suit la loi normale centrée réduite c'est-à-dire que l'on a

$$f_{X_{10}}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

On peut procéder au calcul direct en utilisant l'analyse complexe. On va plutôt utiliser les propriétés sur la fonction caractéristique. Pour tout $u \in \mathbb{R}$, d'après le Théorème 11.2.6, φ_X est dérivable et l'on a de plus

$$\varphi'_{X_{10}}(u) = i\mathbb{E}[X_{10}e^{iuX_{10}}] = \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{iux} e^{-\frac{x^2}{2}} dx.$$

On procède à une intégration par parties et l'on obtient

$$\begin{aligned} \varphi'_{X_{10}}(u) &= \frac{i}{\sqrt{2\pi}} \left[-e^{-\frac{x^2}{2}} e^{iux} \right]_{-\infty}^{\infty} + \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} iue^{iux} e^{-\frac{x^2}{2}} dx \\ &= 0 - u\varphi_{X_{10}}(u). \end{aligned}$$

Conséquemment, on a

$$\varphi_{X_{10}}(u) = Ce^{-\frac{u^2}{2}},$$

où C est une constante que l'on détermine facilement vu que $\varphi_{X_{10}}(0) = 1$. On trouve ainsi

$$\varphi_{X_{10}}(u) = e^{-\frac{u^2}{2}}.$$

Remarque 11.3.1. *On constate que la distribution normale est un vecteur propre associée à la transformation de Fourier. Ceci explique peut-être que la loi normale soit si importante et si omniprésente.*

11.3.2.4 Loi normale quelconque

Soit X_{11} une variable aléatoire suivant la loi normale de paramètres m et σ^2 avec $\sigma > 0$. Alors, il existe X_{10} suivant la loi normale centrée réduite telle que $X_{11} = m + \sigma X_{10}$. Il s'ensuit

$$\begin{aligned} \varphi_{X_{11}}(u) &= \mathbb{E}[\exp(iuX_{11})] \\ &= \mathbb{E}[\exp(ium + iu\sigma X_{10})] \\ &= e^{ium} \varphi_{X_{10}}(u\sigma) \\ &= \exp\left(ium - \frac{\sigma^2}{2}u^2\right). \end{aligned}$$

Remarque 11.3.2. *Il est important de remarquer que la variance est au dénominateur dans le cas de la densité de probabilité mais elle est au numérateur dans le cas de la fonction caractéristique. Ceci est à rapprocher du principe d'Heisenberg.*

11.3.2.5 Loi triangulaire

Dans le cas où X_{12} suit la loi triangulaire sur $[-2; 2]$, on a donc

$$f_{X_{12}}(x) = \frac{|4-x|}{16} \mathbb{1}_{[-2;2]}(x).$$

Alors, on peut montrer $\varphi_{X_{12}}(u) = \left(\frac{\sin(2u)}{2u}\right)^2$ pour $u \neq 0$.

11.3.2.6 Loi d'Erlang

Dans le cas où X_{13} suit la loi d'Erlang de paramètres $n \in \mathbb{N}^*$ et $\lambda > 0$, alors

$$\varphi_{X_{13}}(u) = \left(\frac{\lambda}{\lambda - iu} \right)^n.$$

11.3.2.7 Loi de Laplace

Dans le cas où X_{14} suit la loi de Laplace, c'est-à-dire où l'on a $f_{X_{14}}(x) = \frac{1}{2}e^{-|x|}$, alors :

$$\varphi_{X_{14}}(u) = \frac{1}{1 + u^2}.$$

Comme la loi est symétrique, la fonction caractéristique est réelle.

11.3.2.8 Loi de Cauchy

Ici, X_{15} a pour densité de probabilité $x \mapsto \frac{1}{\pi} \frac{1}{1+x^2}$. On peut montrer, en utilisant le théorème des résidus (en analyse complexe), que l'on a

$$\varphi_{X_{15}}(u) = e^{-|u|}.$$

Comme la loi est symétrique, la fonction caractéristique est réelle.

11.4 Résultats importants

Nous avons vu les fonctions caractéristiques sur les lois de probabilité usuelles. Pour que ce soit réellement utile, il faut que la transformée de Fourier inverse d'une probabilité caractérise ladite probabilité.

Nous donnons donc le résultat suivant d'unicité.

Théorème 11.4.1. *Soient deux variables aléatoires réelles X_1 et X_2 . On suppose que l'on a*

$$\varphi_{X_1} = \varphi_{X_2}.$$

Alors, les variables aléatoires réelles X_1 et X_2 suivent la même loi.

La preuve étant technique et faisant appel à des notions de théorie de la mesure ainsi qu'à de nombreuses utilisations cachées de la valeur principale de Cauchy, elle est omise.

Ce résultat d'unicité, conjointement avec le Théorème 11.4.2 et la Proposition 11.4.3, nous donne immédiatement une preuve des Propositions 5.1.1 et 9.1.1.

Théorème 11.4.2. *Soient deux variables aléatoires réelles X et Y . On suppose que X et Y sont indépendantes. On a alors*

$$\varphi_{X+Y}(u) = \varphi_X(u)\varphi_Y(u), \quad (11.1)$$

pour tout $u \in \mathbb{R}$.

Démonstration. On applique la définition de $\varphi_{X+Y}(u)$ comme suit

$$\varphi_{X+Y}(u) = \mathbb{E} \{ e^{iu(X+Y)} \} = \mathbb{E} \{ e^{iuX} e^{iuY} \}.$$

Comme X et Y sont indépendantes, les variables aléatoires e^{iuX} et e^{iuY} le sont aussi et l'on obtient

$$\mathbb{E} \{ e^{iuX} e^{iuY} \} = \mathbb{E} [e^{iuX}] \times \mathbb{E} [e^{iuY}] = \varphi_X(u)\varphi_Y(u).$$

□

On rappelle maintenant une propriété bien connue de la transformée de Fourier.

Proposition 11.4.3. *Soient deux fonctions positives f et g . On note \hat{f} et \hat{g} les transformées de Fourier respectives de f et g . Alors, si $*$ désigne le produit de convolution, il vient $\widehat{f * g} = \hat{f} \times \hat{g}$.*

Cette proposition est vraie pour les mesures de probabilités. Immédiatement, on a le résultat suivant.

Théorème 11.4.4. *Soient deux variables aléatoires réelles X et Y . On suppose qu'elles sont indépendantes. Alors, on a*

$$\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y.$$

Démonstration. D'après le Théorème 11.4.2, on dispose de l'Égalité (11.1) :

$$\varphi_{X+Y}(u) = \varphi_X(u)\varphi_Y(u).$$

Soit maintenant une variable aléatoire Z qui suit la loi $\mathbb{P}_X * \mathbb{P}_Y$. On a, d'après la Proposition 11.4.3 :

$$\varphi_Z(u) = \varphi_X(u)\varphi_Y(u).$$

D'où Z et $X + Y$ ont même loi d'après le Théorème 11.4.1, ce qui achève la preuve. □

On peut préciser le théorème d'injectivité 11.4.1 à la page 179. Il existe en effet une formule qui donne la densité de probabilité en fonction de la fonction caractéristique, lorsque la variable aléatoire réelle admet une densité de probabilité.

Théorème 11.4.5. *Soit X une variable aléatoire réelle. On suppose que sa fonction caractéristique φ_X est intégrable au sens de Lebesgue. Alors X est une variable à densité et sa densité est donnée par la formule suivante*

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(u) e^{-ixu} du. \quad (11.2)$$

La preuve est omise.

Exemple 11.4.6. *Soit X une variable aléatoire réelle qui suit la loi normale centrée et réduite. Sa fonction caractéristique φ_X est intégrable au sens de Lebesgue vu que $\varphi_X(u) = e^{-\frac{u^2}{2}} = \sqrt{2\pi} f_X(u)$. Et comme φ_X est réelle ainsi que paire, on en déduit bien*

$$\begin{aligned} \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(u) e^{-ixu} du &= \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(u) e^{ixu} du \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f_X(u) e^{ixu} du \\ &= \frac{1}{\sqrt{2\pi}} \varphi_X(x) \\ &= f_X(x). \end{aligned}$$

Contre-exemple 11.4.7. *La fonction caractéristique φ_X peut ne pas être intégrable bien que X est à densité. C'est le cas de la loi exponentielle de paramètre $\lambda > 0$.*

Le contre-exemple précédent nous indique donc que les hypothèses du Théorème 11.4.5 sont suffisantes sans pour autant être nécessaires.

En fait, on peut aller plus loin, en faisant usage de la valeur principale de Cauchy.

Théorème 11.4.8. *Soit une variable aléatoire X de densité f_X . On suppose qu'il existe un nombre fini de réels $a_1 < \dots < a_n$ tels que f_X est de classe \mathcal{C}^1 sur chacun des intervalles $]a_{k-1}; a_k[$ pour tout $k \in \llbracket 1; n+1 \rrbracket$ où l'on a posé $a_0 := -\infty$ et $a_{n+1} := +\infty$. On suppose également que f_X' est sommable. Alors :*

$$\lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^T e^{ixu} \varphi_X(u) du = \frac{1}{2} (f_X(x^-) + f_X(x^+)).$$

On note que la loi exponentielle satisfait les hypothèses du Théorème 11.4.8.

Vecteurs aléatoires

12.1 Introduction

Bien que cela n'ait pas été mentionné explicitement dans les chapitres précédents (à l'exception de la caractérisation de la loi de Student à la page 163), nous avons déjà abordé les vecteurs aléatoires.

En effet, parler d'indépendance de variables aléatoires revient en fait à parler d'indépendance des composantes d'un vecteur aléatoire. Par ailleurs, par essence, bien que cela soit plus technique que dans le cas de la dimension un, la dimension supérieure n'est pas fondamentalement différente.

Il convient aussi de noter que rien n'interdit *a priori* l'étude des variables aléatoires dans des espaces de dimension infinie. Notamment, les processus stochastiques sont des variables aléatoires en dimension infinie. De plus, en traitement du signal, les signaux aléatoires (bruités) sont intéressants dès que la dimension dépasse un.

Les pré-requis de ce chapitre sont les chapitres précédents mais aussi la connaissance de ce qu'est un vecteur dans \mathbb{R}^n . Une connaissance fine de l'algèbre linéaire est un plus indéniable mais elle n'est pas requise pour aborder sereinement ce chapitre.

Dans ce chapitre, nous allons étendre au cas multidimensionnel avec un intérêt particulier pour les cas d'indépendance. Plus précisément, les objectifs sont de maîtriser le vocabulaire (savoir ce qu'est une marginale, une loi marginale, une loi conjointe), de savoir ce qu'est la loi d'un vecteur aléatoire et notamment de connaître sa densité dans le cas où le vecteur est à densité. Également, on attend du lecteur qu'il comprenne parfaitement ce qu'est la fonction caractéristique du vecteur puis de savoir identifier en un coup d'œil les cas où il y a indépendance (que ce soit via la densité, la probabilité, la fonction de répartition ou la fonction caractéristique).

Il conviendra également de garder à l'esprit que la fonction de répartition a un intérêt modéré en dimension supérieure mais aussi que la loi conjointe contient généralement plus d'informations que les lois marginales.

12.2 Premières définitions

Définition 12.2.1. On appelle *vecteur aléatoire* toute application (mesurable) X de Ω dans \mathbb{R}^n .

Soit un vecteur aléatoire X . Pour tout $\omega \in \Omega$, $X(\omega)$ est un élément de \mathbb{R}^n donc il s'écrit sous la forme $X(\omega) = (x_1, \dots, x_n)$. Comme x_i dépend *a priori* de $\omega \in \Omega$, on a donc $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$ où X_i est une application de Ω dans \mathbb{R} . En d'autres termes, X_i est une variable aléatoire pour tout $i \in \llbracket 1; n \rrbracket$.

Définition 12.2.2. Pour tout $i \in \llbracket 1; n \rrbracket$, X_i est la i -ème composante du vecteur aléatoire $X = (X_1, \dots, X_n)$.

Théorème 12.2.3. La donnée d'un vecteur aléatoire X dans \mathbb{R}^n équivaut à celle de ses n composantes.

12.3 Loi d'un vecteur aléatoire

Soit X un vecteur aléatoire de Ω dans \mathbb{R}^n . Soit \mathcal{B} une partie (borélienne) de \mathbb{R}^n . La probabilité que X ait une réalisation dans \mathcal{B} , notée $\mathbb{P}(X \in \mathcal{B})$, est la probabilité d'obtenir un résultat $\omega \in \Omega$ tel que $X(\omega) \in \mathcal{B}$:

$$\mathbb{P}_X(\mathcal{B}) := \mathbb{P}(X \in \mathcal{B}) = \mathbb{P}(\{\omega : X(\omega) \in \mathcal{B}\}) .$$

Définition 12.3.1. La loi de probabilité du vecteur aléatoire X est l'application $\mathbb{P}_X = \mathbb{P}X^{-1}$ qui à toute partie \mathcal{B} de \mathbb{R}^n fait correspondre la probabilité que X ait une réalisation dans \mathcal{B} , $\mathbb{P}(\{\omega : X(\omega) \in \mathcal{B}\})$ notée $\mathbb{P}(X \in \mathcal{B})$.

C'est l'image réciproque par l'application X de la probabilité \mathbb{P} .

On parle aussi de mesure de probabilité sur \mathbb{R}^n .

Théorème 12.3.2. Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire sur Ω à valeurs dans \mathbb{R}^n . Alors, pour tout $i \in \llbracket 1; n \rrbracket$, on a

$$\mathbb{P}_{X_i}(E) = \mathbb{P}_X(\mathbb{R}^{i-1} \times E \times \mathbb{R}^{n-i}) .$$

Démonstration. Par définition, pour tout $j \in \llbracket 1; n \rrbracket$ avec $j \neq i$, on a $\mathbb{P}(X_j \in \mathbb{R}) = \mathbb{P}(\Omega)$. De fait :

$$\begin{aligned}
\mathbb{P}_{X_i}(E) &= \mathbb{P}(X_i \in E) \\
&= \mathbb{P}(\{\omega \in \Omega : X_i(\omega) \in E\}) \\
&= \mathbb{P}(\Omega \cap \dots \cap \Omega \cap \{\omega \in \Omega : X_i(\omega) \in E\} \cap \Omega \cap \dots \cap \Omega) \\
&= \mathbb{P}\left[\left(\bigcap_{j=1}^{i-1} \{X_j \in \mathbb{R}\}\right) \cap \{X_i \in E\} \cap \left(\bigcap_{j=i+1}^n \{X_j \in \mathbb{R}\}\right)\right] \\
&= \mathbb{P}(X_1 \in \mathbb{R}, \dots, X_{i-1} \in \mathbb{R}, X_i \in E, X_{i+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}) \\
&= \mathbb{P}(X \in \mathbb{R}^{i-1} \times E \times \mathbb{R}^{n-i}) \\
&= \mathbb{P}_X(\mathbb{R}^{i-1} \times E \times \mathbb{R}^{n-i}).
\end{aligned}$$

□

Remarque 12.3.3. La loi \mathbb{P}_X du vecteur aléatoire X est appelée loi conjointe des variables aléatoires $(X_i)_{1 \leq i \leq n}$ qui sont ses composantes. Quant aux lois des composantes, on les appelle les lois marginales.

Remarque 12.3.4. La loi du vecteur X détermine les lois marginales. Néanmoins, la réciproque est en général fautive.

12.4 Fonction de répartition

Comme dans le cas des variables aléatoires réelles, on peut considérer la fonction de répartition d'un vecteur aléatoire. Néanmoins, elle est alors d'un intérêt assez modéré.

Définition 12.4.1. Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^n . Soit \mathbb{P}_X sa loi de probabilité. On appelle alors fonction de répartition de X la fonction F_X de \mathbb{R}^n dans $[0; 1]$ définie pour tout $t_1, \dots, t_n \in \mathbb{R}$ par

$$F_X(t_1, \dots, t_n) := \mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n).$$

Remarque 12.4.2. La tribu borélienne sur \mathbb{R}^n étant engendrée par les pavés $]-\infty; t_1] \times \dots \times]-\infty; t_n]$, la loi \mathbb{P}_X d'un vecteur aléatoire X est complètement caractérisée par sa fonction de répartition F_X .

Théorème 12.4.3. Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^n , de loi de probabilité \mathbb{P}_X et de fonction de répartition F_X . Alors F_X est croissante et continue à droite par rapport à chaque variable t_i . De plus, pour tout $i \in \llbracket 1; n \rrbracket$, on a

$$\lim_{t_i \rightarrow -\infty} F_X(t_1, \dots, t_i, t_{i+1}, \dots, t_n) = 0$$

et

$$\lim_{t_1, \dots, t_n \rightarrow +\infty} F_X(t_1, \dots, t_n) = 1.$$

La croissance en chaque variable est immédiate comme ce fut le cas en dimension un. Quant aux deux limites, on a

$$F_X(t_1, \dots, t_i, t_{i+1}, \dots, t_n) \leq F_{X_i}(t_i) \longrightarrow 0,$$

si t_i tend vers $-\infty$. Puis, si pour tout $i \in \llbracket 1; n \rrbracket$, on a $t_i \geq R > 0$, alors :

$$\begin{aligned} F_X(t_1, \dots, t_n) &\geq F_X(R, \dots, R) \\ &\geq 1 - \mathbb{P}\left(\bigcup_{i=1}^n \{\omega \in \Omega : X_i(\omega) > R\}\right) \\ &\geq 1 - \sum_{i=1}^n \mathbb{P}(X_i > R) \longrightarrow 1, \end{aligned}$$

quand R tend vers l'infini.

Théorème 12.4.4. *Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^n , de loi de probabilité \mathbb{P}_X et de fonction de répartition F_X . Alors, si F_{X_i} est la fonction de répartition de la composante X_i , on a*

$$F_{X_i}(t_i) = \lim_{t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n \rightarrow +\infty} F_X(t_1, \dots, t_{i-1}, t_i, t_{i+1}, \dots, t_n).$$

Remarque 12.4.5. *Ainsi la fonction de répartition du vecteur aléatoire X détermine celles de ses composantes. La réciproque est évidemment fautive dans le cas général.*

12.5 Fonction caractéristique

Contrairement à la fonction de répartition dont l'intérêt est limité (bien que non nul) en dimension supérieure ou égale à deux, la fonction caractéristique est particulièrement utile.

Définition 12.5.1. *Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^n , de loi de probabilité \mathbb{P}_X . On appelle alors fonction caractéristique de X la fonction complexe φ_X de n variables réelles $(t_1, \dots, t_n) =: t$ définie par*

$$\varphi_X(t) := \mathbb{E}[\exp(i\langle t, X \rangle)] = \mathbb{E}[\exp(i(t_1 X_1 + \dots + t_n X_n))].$$

Théorème 12.5.2. *Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^n , de loi de probabilité \mathbb{P}_X et de fonction caractéristique φ_X . Alors, on a les deux résultats suivants.*

1. φ_X est continue sur \mathbb{R}^n et $|\varphi_X(t)| \leq \varphi_X(0, \dots, 0) = 1$.

2. Pour toute matrice carrée et symétrique A de taille n et pour tout $b \in \mathbb{R}^n$, on a

$$\varphi_{AX+b}(t) = e^{i\langle t, b \rangle} \varphi_X(tA).$$

Bien que cela ne soit pas l'objet de cet ouvrage, il est important de noter que la fonction caractéristique tire son nom des caractères, c'est-à-dire des morphismes du groupe abélien $(\mathbb{R}^n, +)$ dans le groupe abélien des complexes de module 1 et muni de la multiplication.

Si, en dimension 1, ceux-ci sont les fonctions de la forme $x \mapsto e^{iux}$, en dimension n , les caractères sont en bijection avec \mathbb{R}^n .

12.6 Vecteurs aléatoires discrets

Définition 12.6.1. Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^n . On dit que ce vecteur aléatoire est discret si l'ensemble de ses réalisations possibles $X(\Omega)$ est fini ou infini dénombrable.

Théorème 12.6.2. Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^n . On pose $S := X(\Omega)$ et on suppose que S est fini ou infini dénombrable. Alors :

$$\mathbb{P}_X = \sum_{s \in S} \mathbb{P}(X = s) \delta_s.$$

Ici, δ_s est une distribution multidimensionnelle.

Théorème 12.6.3. Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire dont les valeurs sont dans \mathbb{R}^n . On suppose que S est fini ou infini dénombrable. Alors pour tout $x \in \mathbb{R}$ et pour tout $i \in \llbracket 1; n \rrbracket$, on a

$$\mathbb{P}_{X_i}(x) = \sum_{\substack{s \in X(\Omega) \\ s_i = x}} \mathbb{P}_X(s).$$

En d'autres termes,

$$\mathbb{P}_{X_i} = \sum_{x \in X_i(\Omega)} \left(\sum_{\substack{s \in X(\Omega) \\ s_i = x}} \mathbb{P}_X(s) \right) \delta_x.$$

Regardons maintenant le cas particulier d'un couple de variables aléatoires discrètes à valeurs dans \mathbb{N} .

Théorème 12.6.4. Soit $X = (X_1, X_2)$ un couple de variables aléatoires à valeurs dans \mathbb{N}^2 (c'est-à-dire que X_1 et X_2 sont à valeurs dans \mathbb{N}). Soit $S := X_1 + X_2$. Alors, pour tout $n \in \mathbb{N}$, on a

$$\mathbb{P}(S = n) = \sum_{k=0}^n \mathbb{P}(X_1 = k, X_2 = n - k).$$

Démonstration. Par définition :

$$\begin{aligned}
 \mathbb{P}(S = n) &= \mathbb{P}(X_1 + X_2 = n) \\
 &= \sum_{k=0}^{\infty} \mathbb{P}(X_1 + X_2 = n, X_1 = k) \\
 &= \sum_{k=0}^n \mathbb{P}(X_1 + X_2 = n, X_1 = k) \\
 &= \sum_{k=0}^n \mathbb{P}(k + X_2 = n, X_1 = k) \\
 &= \sum_{k=0}^n \mathbb{P}(X_1 = k, X_2 = n - k),
 \end{aligned}$$

ce qui achève la preuve. □

Remarque 12.6.5. On peut noter que ce théorème implique (si l'on suppose de plus l'indépendance de X_1 et de X_2) la Proposition 5.1.1.

12.7 Vecteurs aléatoires à densité

Définition 12.7.1. Soit un vecteur aléatoire X à valeurs dans \mathbb{R}^n . On dit que X est à densité s'il existe une fonction f_X de n variables réelles à valeurs dans \mathbb{R}_+ telle que pour toute partie $\mathcal{B} \subset \mathbb{R}^n$, on ait

$$\mathbb{P}_X(\mathcal{B}) = \int_{\mathcal{B}} f_X(x) dx = \int_{\mathcal{B}} f_X(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

On peut alors écrire

$$\mathbb{P}_X = f_X dx.$$

Le dx sert ici à signaler que la variable est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^n .

Il convient de noter que l'ensemble \mathcal{B} doit être borélien.

Remarque 12.7.2. Comme dans le cas des variables aléatoires réelles, l'intégrale sur tout l'espace \mathbb{R}^n de f_X vaut 1.

Théorème 12.7.3. Soit un vecteur aléatoire X à valeurs dans \mathbb{R}^n , de loi \mathbb{P}_X . On suppose que X est à densité. Alors, chacune de ses composantes X_i est à densité et de plus

$$f_{X_i}(x) = \int_{\tilde{x}_i \in \mathbb{R}^{n-1}} f_X(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) d\tilde{x}_i,$$

où $\tilde{x}_i := (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$.

Exemple 12.7.4. Dans le cas d'un couple de variables aléatoires (X, Y) , on a

$$f_X(x) = \int_{y \in \mathbb{R}} f_{(X,Y)}(x, y) dy,$$

et

$$f_Y(y) = \int_{x \in \mathbb{R}} f_{(X,Y)}(x, y) dx.$$

On donne maintenant un résultat similaire au Théorème 12.6.4 pour un couple de variables aléatoires à densité.

Théorème 12.7.5. Soit $X = (X_1, X_2)$ un couple de variables aléatoires à densité dans \mathbb{R}^2 . Soit $S := X_1 + X_2$. Alors, S est une variable aléatoire à densité. De plus, sa densité f_S satisfait :

$$f_S(x) = \int_{u \in \mathbb{R}} f_{(X_1, X_2)}(u, x - u) du.$$

Remarque 12.7.6. On peut noter que ce théorème implique (si l'on suppose de plus l'indépendance de X_1 et de X_2) la Proposition 9.1.1.

12.8 Indépendance des composantes

Dans cette section, nous donnons des conditions pour que les composantes d'un vecteur aléatoire soient indépendantes entre elles.

Théorème 12.8.1. Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^n et soient X_1, \dots, X_n ses composantes. On note \mathbb{P}_X la loi de probabilité de X , F_X sa fonction de répartition et φ_X sa fonction caractéristique. De même, on note \mathbb{P}_{X_i} la loi de probabilité de la composante X_i , F_{X_i} sa fonction de répartition et φ_{X_i} sa fonction caractéristique. De plus, si X est à densité, on note f_X sa densité et f_{X_i} celle de la composante X_i . Les propositions suivantes sont alors équivalentes :

- les variables aléatoires X_1, \dots, X_n sont indépendantes,
- pour tous les intervalles $I_i \subset \mathbb{R}$, on a

$$\mathbb{P}(X_1 \in I_1, \dots, X_n \in I_n) = \prod_{i=1}^n \mathbb{P}(X_i \in I_i),$$

- pour tout $x_1, \dots, x_n \in \mathbb{R}$, on a

$$F_X(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i),$$

- pour tout $u_1, \dots, u_n \in \mathbb{R}$, on a

$$\varphi_X(u_1, \dots, u_n) = \prod_{i=1}^n \varphi_{X_i}(u_i).$$

- si les variables aléatoires sont discrètes, pour tout $x_1, \dots, x_n \in \mathbb{R}$, on a

$$\mathbb{P}(X = (x_1, \dots, x_n)) = \prod_{i=1}^n \mathbb{P}(X_i = x_i),$$

- si les variables aléatoires sont à densité, pour tout $x_1, \dots, x_n \in \mathbb{R}$, on a

$$f_X(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i),$$

12.9 Lois conditionnelles

La connaissance de la loi du couple permet de retrouver les lois marginales. Toutefois, il n'est pas possible de déterminer la loi d'un couple si l'on ne connaît que les lois marginales. En effet, deux couples peuvent avoir les mêmes lois marginales sans avoir les mêmes lois conjointes.

Par exemple, si X et Y sont deux variables aléatoires réelles indépendantes, si X suit la loi normale centrée réduite et si Y suit la loi de Rademacher, alors $Z := XY$ suit aussi la loi normale centrée réduite. Néanmoins, X et Z ne sont pas indépendantes. Par conséquent, la loi du couple (X, Z) n'est pas la même que celle du couple (X, X') où X' est une variable aléatoire indépendante de X et de même loi.

Définition 12.9.1 (Lois conditionnelles). Soit $Z = (X, Y)$ un couple de variables aléatoires discrètes. Pour $y \in Y(\Omega)$, on appelle loi conditionnelle de X sachant $\{Y = y\}$ l'application qui à $x \in X(\Omega)$ associe

$$\frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} =: \mathbb{P}(X = x \mid Y = y).$$

Remarque 12.9.2. La connaissance de la loi de Y et des lois conditionnelles de X sachant $\{Y = y\}$ pour chacun des $y \in Y(\Omega)$ est suffisante pour déterminer la loi conjointe du couple (X, Y) . En particulier, on peut alors retrouver la loi de X .

Définition 12.9.3 (Lois conditionnelles). Soit $Z = (X, Y)$ un couple de variables aléatoires discrètes. Pour $x \in X(\Omega)$, on appelle loi conditionnelle de Y sachant $\{X = x\}$ l'application qui à $y \in Y(\Omega)$ associe

$$\frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)} =: \mathbb{P}(Y = y \mid X = x).$$

Remarque 12.9.4. La connaissance de la loi de X et des lois conditionnelles de Y sachant $\{X = x\}$ pour chacun des $x \in X(\Omega)$ est suffisante pour déterminer la loi conjointe du couple (X, Y) . En particulier, on peut alors retrouver la loi de Y .

L'équivalent continu de ces probabilités conditionnelles est la densité conditionnelle, que l'on présente maintenant.

Définition 12.9.5. Soit $Z = (X, Y)$ un couple de variables aléatoires à densité. Pour $y \in \mathbb{R}$ tel que $f_Y(y) > 0$, on appelle densité conditionnelle de X sachant $\{Y = y\}$ l'application $f_{X|Y=y}$ qui à $x \in \mathbb{R}$ associe

$$f_{X|Y=y}(x) := \frac{f_{(X,Y)}(x, y)}{f_Y(y)}.$$

Vecteurs gaussiens

13.1 Introduction

Les pré-requis à ce chapitre sont tous les chapitres précédents et notamment celui sur les vecteurs gaussiens ainsi que celui sur les lois de probabilités usuelles. Dit autrement, pour aborder les vecteurs gaussiens, il faut absolument maîtriser les vecteurs aléatoires ainsi que les gaussiennes. Le calcul matriciel est également un préalable à la compréhension du chapitre. Par ailleurs, savoir ce que sont une forme bilinéaire symétrique et sa forme quadratique associée est préférable bien que cela ne soit pas rédhibitoire de ne pas savoir de quoi il s'agit.

Les objectifs du chapitre sont la capacité à reconnaître facilement un vecteur gaussien et être sensibilisé au fait que certains vecteurs composés de gaussiennes ne sont pas des vecteurs gaussiens. Il est également crucial à l'issue de l'étude du chapitre de savoir que la donnée d'une espérance (vecteur dans \mathbb{R}^n) et d'une matrice de covariance de taille $n \times n$ caractérise entièrement la loi d'un vecteur gaussien ; mais aussi que ledit vecteur gaussien existe toujours. Une propriété intéressante de ces vecteurs gaussiens est que l'indépendance des composantes équivaut à la décorrélation de celles-ci. De plus, il faut bien comprendre que si la densité de probabilité est à peu près inutile dans le cas général (de par sa possible non-existence), la fonction caractéristique associée à un vecteur gaussien est toujours utile. Enfin, nous aborderons le théorème de Cochran ; lequel est d'une importance capitale en statistiques dès que l'on effectue un test du khi-deux mais aussi pour les intervalles de confiance pour lesquels on dispose d'un modèle gaussien à variance inconnue.

13.2 Rappels et compléments

On commence par donner quelques rappels.

Rappel 13.2.1. Soit $m \in \mathbb{R}$ et soit $\sigma^2 > 0$. On dit que X suit la loi normale de paramètres m et σ^2 si X est à densité et si celle-ci est égale à

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x-m)^2}{2\sigma^2} \right\} .$$

Complétons la définition d'une gaussienne comme suit.

Définition 13.2.2. Soit $m \in \mathbb{R}$. Alors, on dit que X suit la loi normale de paramètres m et 0 si $X = m$ presque sûrement.

Notons que la somme de deux gaussiennes indépendantes est une gaussienne comme on l'a vu dans le Théorème 9.4.4.

Également, si X suit la loi normale de paramètres m et $\sigma^2 < \infty$, alors $\text{Var}[X] < \infty$. Il s'ensuit que la covariance de deux variables aléatoires gaussiennes est toujours définie d'après l'inégalité de Cauchy-Schwarz. En effet, si X et Y sont deux telles variables aléatoires, alors :

$$\begin{aligned} \text{Cov}[X, Y] &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &\leq \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2] \mathbb{E}[(Y - \mathbb{E}[Y])^2]} \\ &= \sqrt{\text{Var}[X] \text{Var}[Y]}. \end{aligned}$$

13.3 Définition

On donne maintenant la définition d'un vecteur aléatoire gaussien.

Définition 13.3.1. Un vecteur aléatoire $X := (X_1, \dots, X_n)$ est un vecteur gaussien si et seulement si pour tout $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, la variable aléatoire

$$\alpha_1 X_1 + \dots + \alpha_n X_n$$

est une gaussienne.

Exemple 13.3.2. Le Théorème 9.4.4 nous fournit immédiatement un exemple de tel vecteur gaussien. Soient X_1, \dots, X_n des gaussiennes mutuellement indépendantes. Alors $\alpha_1 X_1 + \dots + \alpha_n X_n$ suit une loi normale pour tous les coefficients réels $\alpha_1, \dots, \alpha_n$.

Notons que l'indépendance des composantes du vecteur gaussien n'est pas nécessaire.

Exemple 13.3.3. Soit X une variable aléatoire réelle suivant la loi normale de paramètres m et σ^2 . Alors, le vecteur aléatoire (X, X) est bien gaussien vu que $\alpha_1 X + \alpha_2 X = (\alpha_1 + \alpha_2)X$ suit la loi normale de paramètres $m(\alpha_1 + \alpha_2)$ et $\sigma^2(\alpha_1 + \alpha_2)^2$ pour tous les coefficients réels α_1, α_2 .

On peut noter que si X est un vecteur gaussien alors chacune de ses composantes est une gaussienne. La réciproque est fautive en général.

Contre-exemple 13.3.4. Soit X une variable aléatoire réelle suivant la loi normale centrée réduite. Soit Y indépendante de X et suivant la loi de Rademacher. Alors, $Z := XY$ suit une loi normale centrée réduite. Ainsi, (X, Z) est un vecteur aléatoire dont chaque composante est une gaussienne. Néanmoins, $X + Z$ ne suit pas une gaussienne vu que $\mathbb{P}(X + Z = 0) = \frac{1}{2}$.

Définition 13.3.5. Soit $X := (X_1, \dots, X_n)$ un vecteur gaussien. On définit son espérance comme suit : $m := \mathbb{E}[X] := (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$.

On définit de même la matrice dite de covariance.

Définition 13.3.6. Soit $X := (X_1, \dots, X_n)$ un vecteur gaussien. On définit sa matrice de covariance Σ comme étant la matrice symétrique positive dont le coefficient $\Sigma_{i,j}$ est égal à

$$\Sigma_{i,j} := \text{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \mathbb{E}[X_j].$$

13.4 Existence

On a montré qu'à tout vecteur gaussien, on peut associer une espérance et une matrice de covariance.

Le théorème suivant, dit d'existence, donne la réciproque.

Théorème 13.4.1. Soit $m \in \mathbb{R}^n$ et soit Σ une matrice symétrique positive de taille $n \times n$. Alors, on peut construire un vecteur gaussien admettant m comme espérance et Σ comme matrice de covariance.

Démonstration. La matrice Σ est symétrique et positive donc on peut la diagonaliser dans une base orthonormale et chacune de ses valeurs propres est positive. Ainsi : $C = ODO^T$ où O est une matrice orthonormale et où D est la matrice diagonale formée de coefficients $D_{i,i} \geq 0$ pour tout $i \in \llbracket 1; n \rrbracket$. Posons la matrice $A := O\tilde{D}O^T$ où \tilde{D} est une matrice diagonale telle que $\tilde{D}_{i,i} = \sqrt{D_{i,i}}$ pour tout $i \in \llbracket 1; n \rrbracket$. La matrice A est telle que $A^2 = \Sigma$.

On considère ensuite le vecteur gaussien Y constitué de variables aléatoires indépendantes et identiquement distribuées suivant la loi normale centrée réduite. Puis, on pose $X := m + AY$.

Il n'est alors pas difficile de montrer que X est un vecteur gaussien et que X admet m comme espérance et Σ comme matrice de covariance. □

Remarque 13.4.2. On peut montrer que si X et Y sont deux vecteurs gaussiens ayant même espérance et même matrice de covariance, alors ils ont même loi. On en verra une preuve dans la section suivante.

13.5 Fonction caractéristique

On rappelle que la fonction caractéristique de X est l'application qui associe $\mathbb{E} \left[e^{i \sum_{k=1}^n u_k X_k} \right]$ à tout vecteur $u := (u_1, \dots, u_n) \in \mathbb{R}^n$.

Proposition 13.5.1. *Soit X un vecteur gaussien d'espérance m et de matrice de covariance Σ . Alors*

$$\varphi_X(u_1, \dots, u_n) = \exp \left\{ i \sum_{k=1}^n u_k m_k - \frac{1}{2} Q(u_1, \dots, u_n) \right\},$$

où Q est la forme quadratique associée à la matrice symétrique Σ à savoir : $Q(u_1, \dots, u_n) = \sum_{k=1}^n \sum_{l=1}^n u_k \Sigma_{k,l} u_l$.

La preuve est omise. Il suffit de noter que $\sum_{k=1}^n u_k X_k$ est une gaussienne et ainsi $\varphi_X(u_1, \dots, u_n) = \varphi_Y(1)$.

Comme la fonction caractéristique caractérise la loi de probabilité, il s'ensuit que l'espérance et la matrice de covariance suffisent à caractériser entièrement la loi d'un vecteur gaussien.

De la proposition 13.5.1, on en déduit le résultat suivant sur la stabilité des vecteurs gaussiens par combinaison linéaire.

Théorème 13.5.2. *Soient deux vecteurs gaussiens X et Y , en dimension n , d'espérances respectives m_X et m_Y et de matrices de covariance respectives Σ_X et Σ_Y . Alors si X et Y sont indépendants, $X+Y$ est un vecteur gaussien d'espérance $m_X + m_Y$ et de matrice de covariance $\Sigma_X + \Sigma_Y$.*

Démonstration. Comme X et Y sont indépendants, on en déduit que pour tous les réels u_1, \dots, u_n , $\tilde{X} := \sum_{k=1}^n u_k X_k$ et $\tilde{Y} := \sum_{k=1}^n u_k Y_k$ sont indépendantes et de fait $e^{i\tilde{X}}$ et $e^{i\tilde{Y}}$ le sont aussi. Ceci nous donne

$$\varphi_{X+Y}(u_1, \dots, u_n) = \varphi_X(u_1, \dots, u_n) \varphi_Y(u_1, \dots, u_n).$$

On utilise ensuite l'expression obtenue dans la Proposition 13.5.1 et il vient

$$\varphi_{X+Y}(u_1, \dots, u_n) = \exp \left\{ i \sum_{k=1}^n u_k (m_X(k) + m_Y(k)) - \frac{1}{2} \hat{Q}(u_1, \dots, u_n) \right\},$$

où \hat{Q} est la forme quadratique associée à la matrice symétrique et positive $\hat{\Sigma} := \Sigma_X + \Sigma_Y$. □

Une question que l'on peut se poser naturellement est de savoir si les composantes du vecteur gaussien sont indépendantes ou non. En effet, on a présenté un exemple où il y a indépendance mais l'on en a aussi présenté un où cette condition était violée.

Cette question admet une réponse simple avec le théorème suivant.

Théorème 13.5.3. *Soit X un vecteur gaussien de \mathbb{R}^n . Alors ses composantes sont mutuellement indépendantes si et seulement si elles sont décorrélées.*

Démonstration. Si les composantes sont indépendantes, elles sont décorrélées vu que l'indépendance de deux variables aléatoires implique immédiatement que leur covariance est nulle.

Supposons maintenant que X est un vecteur gaussien dont les composantes X_1, \dots, X_n sont décorrélées. La matrice Σ est de fait diagonale. Alors, pour tous les réels u_1, \dots, u_n , on a

$$\begin{aligned} \varphi_X(u_1, \dots, u_n) &= \exp \left\{ i \sum_{k=1}^n u_k m_k - \frac{1}{2} Q(u_1, \dots, u_n) \right\} \\ &= \prod_{k=1}^n e^{iu_k m_k} e^{-\frac{\Sigma_{k,k} u_k^2}{2}} \\ &= \prod_{k=1}^n \varphi_{X_k}(u_k). \end{aligned}$$

La preuve s'ensuit du Théorème 12.8.1. □

Il convient de bien comprendre que ceci n'est vrai que pour les vecteurs gaussiens. En particulier, le vecteur aléatoire du Contre-exemple 13.3.4 n'était pas gaussien.

13.6 Densité de probabilité

Il est naturel de se demander pourquoi nous n'avons pas encore abordé la notion de densité de probabilité. Cela vient de sa non-existence dans le cas général.

Théorème 13.6.1. *Soit X un vecteur gaussien d'espérance m et de matrice de covariance Σ . Alors, X admet une densité si et seulement si la matrice Σ est inversible ; et le cas échéant, l'on a de plus :*

$$f_X(x_1, \dots, x_n) = \frac{1}{\sqrt{2\pi}^n} \frac{1}{\sqrt{\text{Dét}(\Sigma)}} \exp \left\{ -\frac{1}{2} \bar{Q}(x_1 - m_1, \dots, x_n - m_n) \right\},$$

où \bar{Q} est la forme quadratique associée à la matrice définie positive $\Sigma^{-1} =: A = (A_{k,l})_{1 \leq k, l \leq n}$. En d'autres termes :

$$f_X(x_1, \dots, x_n) = \frac{1}{\sqrt{2\pi}^n} \frac{1}{\sqrt{\text{Dét}(\Sigma)}} \exp \left\{ -\frac{1}{2} \sum_{k=1}^n \sum_{l=1}^n (x_k - m_k) A_{k,l} (x_l - m_l) \right\}.$$

La preuve est omise.

13.7 Théorème de Cochran

Le théorème suivant est d'une importance capitale en statistiques et plus précisément pour les intervalles de confiance et pour les tests du χ^2 . Il nécessite de bien connaître la notion de projection orthogonale dans un espace vectoriel.

Théorème 13.7.1. *Soit un vecteur gaussien $X = (X_1, \dots, X_n)$. On suppose que son espérance est $m_X = (0, \dots, 0)$ et que sa matrice de covariance Σ_X est l'identité. En d'autres termes, on considère des variables aléatoires indépendantes et identiquement distribuées suivant la loi normale centrée réduite. Soit $\mathbb{F} \subset \mathbb{R}^n$ un espace vectoriel de dimension $p \in \llbracket 1; n-1 \rrbracket$. On note $p_{\mathbb{F}}$ la projection orthogonale sur \mathbb{F} du vecteur gaussien X . Alors, $\|p_{\mathbb{F}}(X)\|^2$ et $\|X - p_{\mathbb{F}}(X)\|^2$ sont indépendantes et suivent respectivement la loi $\chi^2(p)$ et la loi $\chi^2(n-p)$.*

La preuve est omise. On en donne cependant un corollaire immédiat et fort utile pour le test du χ^2 et pour les intervalles de confiance.

Corollaire 13.7.2. *Soit $n \geq 2$ et soient X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées suivant la loi normale centrée réduite. Alors, si l'on pose $\overline{X}_n := \frac{1}{n} \sum_{k=1}^n X_k$, la variable aléatoire $S_n^2 := \sum_{k=1}^n (X_k - \overline{X}_n)^2$ suit la loi $\chi^2(n-1)$. Et de plus, \overline{X}_n est indépendante de S_n^2 .*

Il suffit en effet de remarquer que le vecteur $\overline{X}_n (1, \dots, 1)$ est la projection orthogonale de $X := (X_1, \dots, X_n)$ sur $\mathbb{F} := \text{Vect}(\overrightarrow{u}_0)$ avec $\overrightarrow{u}_0 := (1, \dots, 1)$. Par conséquent, S_n^2 est la norme au carré de $X - \overline{X}_n \overrightarrow{u}_0$. On applique alors le Théorème 13.7.1.

Convergences des variables aléatoires

14.1 Introduction

La définition même de ce qu'est une probabilité contient en son sein l'idée d'infini. En effet, le troisième axiome, à savoir l'axiome des probabilités totales signifie que l'on va s'intéresser tôt ou tard à la notion de limite donc à celle de convergence.

La notion de convergence dépend directement de la topologie que l'on associe à l'espace que l'on étudie. Par exemple, quand on regarde les suites réelles, la topologie principale (en terme d'utilisation) est celle associée à la valeur absolue, en guise de norme. Toutefois, si l'on s'intéresse aux suites de fonctions, on dispose de plusieurs convergences : la convergence simple et la convergence uniforme. Avec les séries de fonctions, on a la convergence simple, la convergence uniforme mais aussi la convergence absolue.

Quand on utilise la notion d'infini, il convient de toujours préciser quelle topologie on utilise. En particulier, l'ensemble \mathbb{R}^Ω est un espace fonctionnel donc il faut être précautionneux avec la périlleuse notion de convergence.

Les notions qui doivent avoir été vues au préalable sont celles des suites numériques, des suites de fonctions. En particulier, comprendre de quoi il retourne dans les convergences de variables aléatoires nécessite de bien maîtriser le fameux "pour tout $\epsilon > 0$, il existe..." Les notions de topologie sont aussi un vrai plus de même que la connaissance de la loi géométrique. En effet, cette dernière est utilisée dans l'exemple qui sert de fil rouge.

Les objectifs de ce chapitre sont de sensibiliser le lecteur à la non-unicité de la convergence. Il est aussi important de bien saisir que ce qui est intuitif pour une convergence ne l'est pas forcément pour une autre. Pire, ce qui est vrai pour une convergence ne l'est pas forcément pour une autre. Il est notamment attendu du lecteur qu'il soit désormais des plus précautionneux dès qu'il utilise la notion de convergence. Les convergences que nous verrons sont : la convergence presque sûre, la convergence dans L^p pour $p \in [1; +\infty[$ et la convergence en probabilité. Nous ferons également l'étude de la convergence en loi, laquelle est fondamentalement différente des autres. Très important : être capable d'avoir une convergence à partir

d'une autre. Également, un objectif est d'être en mesure de donner des réciproques partielles aux implications en question. Enfin, l'étudiant est fortement invité à savoir parfaitement obtenir une convergence en loi (par la fonction de répartition, par la densité de probabilité ou par la fonction caractéristique).

14.1.1 Paradoxe

On effectue plusieurs lancers indépendants d'une pièce équilibrée. Quand la pièce fait pile, on perd la somme mise. Quand elle fait face, on récupère le double de ce qu'on a misé.

De fait, la modélisation par des expériences indépendantes d'une loi de Rademacher se prête bien à ce cadre. On note m_n la mise du lancer numéro n . On pose $X_n := -1$ si la pièce fait pile au lancer numéro n et $X_n := 1$ sinon.

Le gain G_n du joueur après n lancers est donc égal à $G_n = \sum_{k=1}^n m_k X_k$. Si la mise est déterministe, la linéarité de l'espérance implique $\mathbb{E}[G_n] = 0$ pour tout $n \in \mathbb{N}^*$. Néanmoins, l'idée est ici que le joueur va adapter sa mise en fonction des résultats aux lancers *précédents*. En effet, on part du principe que le joueur veut optimiser ses gains. On prend aussi comme hypothèse qu'il ne dispose pas de pouvoirs lui permettant de deviner les lancers futurs.

Nous allons opter pour la "martingale" (dans le sens commun du mot) classique consistant à miser un euro au premier lancer puis à doubler la mise jusqu'à gagner une fois. L'idée est en fait d'introduire ce que l'on appelle un temps d'arrêt, lequel suit en fait la loi géométrique de paramètre $\frac{1}{2}$:

$$\tau := \inf \{k \geq 1 : X_k = 1\} ,$$

avec la convention $\inf \emptyset = +\infty$.

On utilise alors la stratégie de mise **non déterministe** suivante. On pose : $m_1 := 1$. Puis, $m_{n+1} := 2m_n \mathbb{1}_{n < \tau}$. En d'autres termes, on double la mise jusqu'à tomber sur face puis on arrête le jeu. On en déduit que pour tout $k \leq \tau$, on a $m_k = 2^{k-1}$. Le gain au lancer numéro $\tau(\omega)$ est ainsi :

$$G_{\tau(\omega)} = \sum_{k=1}^{\tau(\omega)} m_k X_k = \sum_{k=1}^{\tau(\omega)} 2^{k-1} X_k .$$

Or, $X_{\tau(\omega)} = +1$ et $X_k = -1$ pour tout $k \in \llbracket 1; \tau(\omega) - 1 \rrbracket$, si $\tau(\omega) \geq 2$. Le cas où $\tau(\omega) = 1$ est passé sous silence vu que cela signifie qu'on a gagné au premier essai et donc il n'y a aucune stratégie à avoir. Il vient :

$$G_{\tau(\omega)} = - \sum_{k=1}^{\tau(\omega)-1} 2^{k-1} + 2^{\tau(\omega)-1} = - \sum_{k=0}^{\tau(\omega)-2} 2^k + 2^{\tau(\omega)-1} = 1 - 2^{\tau(\omega)-1} + 2^{\tau(\omega)-1} = 1 .$$

Par conséquent, si $n \geq \tau(\omega)$, $G_n = 1$. Et, si $n < \tau(\omega)$, $G_n = -\sum_{k=1}^n 2^{k-1} = 1 - 2^n$.

On remarque :

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} G_n = 1\right) = 1.$$

En effet, $\{\omega \in \Omega : \lim_{n \rightarrow \infty} G_n(\omega) \neq 1\} = \bigcap_{k=1}^{\infty} \{X_k = -1\}$. Or, cet évènement est inclus dans $\bigcap_{k=1}^N \{X_k = -1\}$ pour tout $N \in \mathbb{N}^*$. Cet évènement est de probabilité 2^{-N} . Il s'ensuit le résultat annoncé.

Pourtant, en rappelant que $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$ pour tout évènement A , on remarque aussi :

$$\begin{aligned} \mathbb{E}[G_n] &= \mathbb{E}[G_n \mathbb{1}_{n < \tau}] + \mathbb{E}[G_n \mathbb{1}_{n \geq \tau}] \\ &= \mathbb{E}[(1 - 2^n) \times \mathbb{1}_{n < \tau}] + \mathbb{E}[1 \times \mathbb{1}_{n \geq \tau}] \\ &= (1 - 2^n)\mathbb{P}(\tau > n) + \mathbb{P}(\tau \leq n) \\ &= (1 - 2^n)2^{-n} + (1 - 2^{-n}) \\ &= 0. \end{aligned}$$

En effet, τ suit une loi géométrique de paramètre $\frac{1}{2}$, $\mathcal{G}\left(\frac{1}{2}\right)$. Ainsi, $\mathbb{P}(\tau > n) = 2^{-n}$.

Ce résultat semble surprenant. En effet, il semblerait qu'il suffise de refaire la procédure un grand nombre de fois pour dégager un bénéfice moyen.

Comment concilier ces deux résultats justes mais qui semblent incompatibles ?

14.1.2 Convergence simple

Étudions pour l'instant un exemple simple de topologie : celui de la convergence simple.

En effet, on rappelle qu'une variable aléatoire réelle est une fonction (mesurable) de l'espace fondamental Ω vers \mathbb{R} . On peut donc naturellement munir l'ensemble des variables aléatoires de cette topologie. En d'autres termes, on dit que la suite de variables aléatoires réelles $(X_n)_n$ converge vers X si $X_n(\omega)$ converge vers $X(\omega)$ pour tout $\omega \in \Omega$.

Cette convergence est à peu près inutile en probabilités vu qu'elle n'exploite pas la mesure de probabilité \mathbb{P} dont on a muni l'espace fondamental. Justifions maintenant cette assertion selon laquelle cette convergence est inutile.

Exemple 14.1.1. Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées suivant la loi $\mathbb{P}_{X_1} = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. En pratique, on lance une pièce équilibrée. Alors, X_n vaut 1 si la pièce donne "face" et X_n vaut -1 si la pièce donne "pile".

Lorsque le nombre n de lancers est grand, on s'attend à ce que la proportion de "faces" soit à peu près égale à $\frac{1}{2}$ au vu de l'approche fréquentiste de la notion de probabilité. Ainsi, on s'attend à ce que l'on ait

$$\lim_{n \rightarrow \infty} \frac{X_1 + \cdots + X_n}{n} = 0.$$

Or, cette limite est fautive pour certaines valeurs de ω . Posons par exemple

$$A := \{\omega \in \Omega : \# \{\text{nombre de faces}\} < \infty\}.$$

On a alors immédiatement

$$\lim_{n \rightarrow \infty} \frac{X_1(\omega) + \cdots + X_n(\omega)}{n} = -1$$

pour tout $\omega \in A$.

Comme la convergence uniforme est encore plus restrictive que la convergence simple, cette dernière non plus n'a pas d'utilité en probabilités.

On pourra objecter à l'exemple ci-dessus que l'évènement A est peu probable. On peut d'ailleurs facilement vérifier, en utilisant l'indépendance ainsi que le Lemme 2.4.26 à la page 42 que sa probabilité est égale à 0. Or, seuls les évènements de probabilités strictement positives nous intéressent.

On pourrait montrer en utilisant la loi forte des grands nombres (voir Section 15.3 à la page 214) que l'on dispose de la limite suivante :

$$\mathbb{P} \left(\left\{ \omega : \lim_{n \rightarrow \infty} \frac{X_1(\omega) + \cdots + X_n(\omega)}{n} = 0 \right\} \right) = 1.$$

En d'autres termes, on n'a pas convergence pour tout ω mais uniquement pour presque tout ω , ou dit plus rigoureusement, pour \mathbb{P} -presque tout ω . C'est typiquement cette convergence presque sûre qui nous intéresse.

14.2 Convergence presque sûre

Détaillons maintenant cette convergence presque sûre.

14.2.1 Définition et premières propriétés

Définition 14.2.1 (Convergence presque sûre). Soit $(X_n)_n$ une suite de variables aléatoires définies sur un espace (Ω, \mathbb{P}) . Soit X une autre variable aléatoire définie sur (Ω, \mathbb{P}) . On dit que la suite $(X_n)_n$ converge presque sûrement vers X si $\mathbb{P}(\Lambda) = 1$ où l'on a

$$\Lambda := \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\}.$$

Notation 14.2.2. Si $(X_n)_n$ converge presque sûrement vers X , alors on note

$$X_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X.$$

On dispose de quelques premières propriétés.

Si $X_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X$ et si $Y_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} Y$ alors $X_n + Y_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X + Y$. Et pour tout $\lambda \in \mathbb{R}$, $\lambda X_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \lambda X$.

Également, si f est une fonction continue de la variable réelle, alors si $X_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X$, on a $f(X_n) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} f(X)$.

14.2.2 Caractérisation

On commence par une première proposition.

Proposition 14.2.3. *L'intersection d'une famille indexée par un ensemble infini dénombrable d'indices I d'évènements $(A_i)_{i \in I}$ est de probabilité 1 si et seulement si chacun des A_i est de probabilité 1.*

Démonstration. D'abord, $\mathbb{P}(\bigcap_{i \in I} A_i) \leq \mathbb{P}(A_k)$ pour tout $k \in I$. Donc si l'intersection est de probabilité 1, immédiatement il en est de même de chacun des A_k .

Inversement, si chacun des A_i est de probabilité 1, alors

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = 1 - \mathbb{P}\left(\bigcup_{i \in I} A_i^c\right) \geq 1 - \sum_{i \in I} \mathbb{P}(A_i^c) = 1 - \sum_{i \in I} 0 = 1 - 0 = 1,$$

la somme des zéros valant bien zéro puisque l'ensemble I est au plus dénombrable. \square

Définition 14.2.4. *La limite supérieure d'une suite majorée de réels est par définition la plus grande de ses valeurs d'adhérence. Et, si la suite n'est pas majorée, sa limite supérieure est $+\infty$. On la note $\limsup_{n \rightarrow \infty} x_n$.*

Remarque 14.2.5. *On a l'équivalence entre les deux assertions suivantes :*

- $\limsup_{n \rightarrow \infty} x_n \leq M$,
- pour tout $\epsilon > 0$, $\{n \in \mathbb{N} : x_n \geq M + \epsilon\}$ est fini.

Remarque 14.2.6. *On a également l'équivalence entre les deux assertions suivantes :*

- $\limsup_{n \rightarrow \infty} x_n \geq M$,
- pour tout $\epsilon > 0$, $\{n \in \mathbb{N} : x_n \geq M - \epsilon\}$ est infini.

De la même manière, pour prouver que

$$\limsup_{n \rightarrow \infty} X_n \geq M \text{ presque sûrement,}$$

il suffit de prouver que pour tout $a < M$, on a $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$ où $A_n := \{X_n \geq a\}$. Inversement, pour prouver que

$$\limsup_{n \rightarrow \infty} X_n \leq M \text{ presque sûrement,}$$

il suffit de prouver que pour tout $a > M$, on a $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$ où $A_n := \{X_n \geq a\}$.

On en déduit le théorème suivant.

Théorème 14.2.7. *Soit $(X_n)_n$ une suite de variables aléatoires définies sur un espace (Ω, \mathbb{P}) et soit $M \in \mathbb{R}$. On suppose que pour tout $a < M$, on a*

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \{X_n \geq a\}\right) = 1$$

et pour tout $a > M$, on a

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \{X_n \geq a\}\right) = 0.$$

Alors, on en déduit

$$\limsup_{n \rightarrow \infty} X_n = M \text{ presque sûrement.}$$

Ceci nous amène au critère fondamental pour obtenir la convergence presque sûre à savoir le corollaire suivant.

Corollaire 14.2.8. *Soit $(X_n)_n$ une suite de variables aléatoires définies sur un espace (Ω, \mathbb{P}) et soit X une variable aléatoire définie sur le même espace. On a alors l'équivalence entre les deux assertions qui suivent.*

- $X_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X$.
- Pour tout $\epsilon > 0$, $\mathbb{P}(\limsup_{n \rightarrow \infty} \{|X_n - X| \geq \epsilon\}) = 0$.

Comme on le verra par la suite, cette convergence est toutefois très forte et on devra considérer, par la force des choses, des convergences moins restrictives mais qui satisfont de bonnes propriétés et s'avèrent utiles en pratique. Nous pensons en particulier à la convergence en probabilité.

14.3 Convergence dans L^p

On présente maintenant un mode de convergence, dite en moyenne.

14.3.1 Définition

Définition 14.3.1 (Convergence dans L^p ($p \in [1; +\infty[$)). Soit $(X_n)_n$ une suite de variables aléatoires définies sur un espace (Ω, \mathbb{P}) . Soit X une autre variable aléatoire définie sur (Ω, \mathbb{P}) . On dit que la suite $(X_n)_n$ converge vers X dans L^p si les variables aléatoires X_n et X sont dans L^p (c'est-à-dire que l'on a $\mathbb{E}[|X_n|^p] < \infty$ et $\mathbb{E}[|X|^p] < \infty$) et si de plus

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0.$$

On dit aussi que X_n converge en moyenne d'ordre p vers X . Les convergences qui nous intéressent le plus sont pour $p = 1$ et pour $p = 2$.

Notation 14.3.2. Si $(X_n)_n$ converge dans L^p vers X , alors on note

$$X_n \xrightarrow[n \rightarrow +\infty]{L^p} X.$$

On dispose de quelques premières propriétés.

Si $X_n \xrightarrow[n \rightarrow +\infty]{L^p} X$ et si $Y_n \xrightarrow[n \rightarrow +\infty]{L^p} Y$ alors $X_n + Y_n \xrightarrow[n \rightarrow +\infty]{L^p} X + Y$. Et pour tout $\lambda \in \mathbb{R}$, $\lambda X_n \xrightarrow[n \rightarrow +\infty]{L^p} \lambda X$.

Remarque 14.3.3. L'espace L^p est complet pour $p \in [1; \infty[$ pour la norme $\|\cdot\|_p$.

Nous ne détaillerons pas les propriétés des espaces complets. Ce qu'il faut surtout retenir, c'est que les espaces non complets sont beaucoup plus capricieux.

Remarque 14.3.4. Si $1 \leq p \leq q < +\infty$, alors l'inégalité de Jensen nous assure que pour toute variable aléatoire réelle Y , on a l'inégalité $\mathbb{E}[|Y|^p] \leq \mathbb{E}[|Y|^q]^{\frac{p}{q}}$. Ainsi, si une suite $(X_n)_n$ converge vers X dans L^q alors elle converge également vers X dans L^p . En effet, $\mathbb{E}[|X_n - X|^p] \leq (\mathbb{E}[|X_n - X|^q])^{\frac{p}{q}}$.

Lemme 14.3.5. Si la suite de terme général X_n converge vers X dans L^1 , alors la suite $(\mathbb{E}[X_n])_n$ converge vers $\mathbb{E}[X]$.

Démonstration. On a d'après l'inégalité triangulaire :

$$|\mathbb{E}[X_n] - \mathbb{E}[X]| \leq \mathbb{E}[|X_n - X|] \longrightarrow 0,$$

quand n tend vers l'infini.

□

14.3.2 Différence avec l'espace L^p en analyse

En analyse, et pour être plus exact, quand on travaille avec l'intégrale de Lebesgue, l'espace L^p est l'espace des fonctions (ou plutôt des classes d'équivalence de fonctions) telles que $\int_{\mathbb{R}} |f(t)|^p dt$.

Ici, l'espace est très différent. En effet, dans le cas où X est à densité, alors sa norme p est

$$\mathbb{E}[|X|^p] := \int_{\mathbb{R}} |t|^p f_X(t) dt,$$

ce qui n'a rien à voir avec $\int_{\mathbb{R}} |f_X(t)|^p dt$. Pour réconcilier les deux points de vue, il faudrait faire de la théorie de la mesure mais l'on a ici choisi de ne pas en faire.

On peut néanmoins faire un pont entre les deux espaces en se représentant le moment d'ordre p d'une variable aléatoire réelle X comme étant

$$\mathbb{E}[X^p] = \int_{\Omega} (X(\omega))^p \mathbb{P}(d\omega).$$

La différence est donc qu'on intègre sur un espace plus subtile et que ce n'est pas nécessairement la mesure de Lebesgue que l'on utilise pour intégrer. Nous invitons le lecteur curieux à consulter le Chapitre 24 à la page 341.

14.4 Convergence en probabilité

On présente ici la convergence dite faible.

14.4.1 Définition

Définition 14.4.1 (Convergence en probabilité). *Soit $(X_n)_n$ une suite de variables aléatoires définies sur un espace (Ω, \mathbb{P}) . Soit X une autre variable aléatoire définie sur (Ω, \mathbb{P}) . On dit que la suite $(X_n)_n$ converge en probabilité vers X si pour tout $\epsilon > 0$, on a la limite suivante :*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}) = 0.$$

Notation 14.4.2. *Si $(X_n)_n$ converge en probabilité vers X , alors on note*

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X.$$

On dispose de quelques premières propriétés.

Si $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ et si $Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} Y$ alors $X_n + Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X + Y$. Et pour tout $\lambda \in \mathbb{R}$, $\lambda X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \lambda X$.

14.4.2 Liens avec les deux autres convergences

Comme mentionné plus haut, la convergence en probabilité est dite faible. Ainsi, on dispose des deux théorèmes suivants.

Théorème 14.4.3. *Pour tout $p \in [1; +\infty[$, la convergence dans L^p implique la convergence en probabilité.*

Démonstration. On remarque :

$$\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\} = \{\omega : |X_n(\omega) - X(\omega)|^p > \epsilon^p\} .$$

On applique alors l'inégalité de Markov et l'on aboutit à

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{1}{\epsilon^p} \mathbb{E}(|X_n - X|^p) \longrightarrow 0 .$$

□

Théorème 14.4.4. *La convergence presque sûre implique celle en probabilité.*

Démonstration. D'après le Corollaire 14.2.8, la convergence presque sûre de la suite de terme général X_n vers X implique :

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \{|X_n - X| \geq \epsilon\}\right) = 0 ,$$

pour tout $\epsilon > 0$. Par définition de la limite supérieure et par décroissance de la suite d'évènements $B_n := \bigcup_{k=n}^{\infty} \{|X_k - X| \geq \epsilon\}$, on en déduit

$$0 = \mathbb{P}\left(\limsup_{n \rightarrow \infty} \{|X_n - X| \geq \epsilon\}\right) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n) .$$

Or, $\{|X_n - X| \geq \epsilon\} \subset B_n$ donc $\lim_{n \rightarrow \infty} \mathbb{P}(\{|X_n - X| \geq \epsilon\}) = 0$, ce qui achève la preuve de la convergence de la suite de terme général X_n vers X , en probabilité.

□

14.4.3 Remarque

La convergence presque sûre dénote un caractère trajectorien. Ainsi, pour presque tout $\omega \in \Omega$, $X_n(\omega) - X(\omega)$ tend vers 0.

La convergence dans L^p signifie qu'en moyennant, à n fixé sur toutes les trajectoires de $\omega \mapsto |X_n(\omega) - X(\omega)|$, alors cette moyenne tend vers 0.

La convergence en probabilité, quant à elle, signifie que pour $\epsilon > 0$ fixé, la proportion de trajectoires de $\omega \mapsto X_n(\omega) - X(\omega)$ à l'extérieur de la bande $[-\epsilon; \epsilon]$ tend vers 0.

14.4.4 Réciproques partielles

Théorème 14.4.5. *Soit X_n une suite de variables aléatoires définies sur un espace (Ω, \mathbb{P}) . Soit X une autre variable aléatoire définie sur (Ω, \mathbb{P}) . On suppose $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$. Alors, il existe une sous-suite de $(X_n)_n$ qui converge presque sûrement vers X .*

Démonstration. On pose $n_0 := 0$. Puis, pour $k \geq 1$, on pose

$$n_k := \inf \left\{ n > n_{k-1} : \mathbb{P} \left(|X_n - X| \geq \frac{1}{k} \right) \leq 2^{-k} \right\},$$

avec la convention $\inf \emptyset = +\infty$. Comme $\mathbb{P} \left(|X_n - X| \geq \frac{1}{k} \right)$ tend vers 0 à k fixé, on en déduit la finitude de n_k .

Puis, la série de terme général $\mathbb{P} \left(|X_{n_k} - X| \geq \frac{1}{k} \right)$ étant convergente, le lemme de Borel-Cantelli 2.4.26 nous assure que l'on a

$$\mathbb{P} \left(\limsup_{k \rightarrow \infty} \left\{ |X_{n_k} - X| \geq \frac{1}{k} \right\} \right) = 0,$$

ce qui équivaut à

$$\mathbb{P} \left(\liminf_{k \rightarrow \infty} \left\{ |X_{n_k} - X| < \frac{1}{k} \right\} \right) = 1.$$

En d'autres termes, pour presque tout $\omega \in \Omega$, il existe $k_0(\omega)$ tel que pour tout $k \geq k_0(\omega)$, on a

$$|X_{n_k}(\omega) - X(\omega)| < \frac{1}{k}.$$

Ceci traduit la convergence presque sûre de la suite de terme général X_{n_k} vers X . □

Proposition 14.4.6. *Si $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ et si f est continue de \mathbb{R} dans \mathbb{R} alors $f(X_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} f(X)$.*

On donne maintenant une condition suffisante pour que la convergence en probabilité implique la convergence dans L^p .

Théorème 14.4.7. *On suppose que la suite de variables aléatoires $(X_n)_n$ converge en probabilité vers X . On suppose de plus que X_n est uniformément bornée par une variable $Y \in L^p$ avec $p \in [1; \infty[$. Alors, $X \in L^p$ et X_n converge vers X dans L^p .*

La preuve est omise.

14.5 Convergence en loi

Les trois convergences que l'on a vues dans les deux sections précédentes peuvent être conçues comme des variantes des convergences habituelles. En probabilités, on s'intéresse à un autre type de convergence, très utile mais de nature bien différente. Il s'agit de la convergence en loi. Ce mode de convergence est encore plus faible que celui de la convergence en probabilité. Ce qui rend inhabituelle cette convergence est que la suite de variables aléatoires $(X_n)_n$ peut converger vers X sans que les valeurs prises par X_n ne se rapprochent de celles prises par X . À vrai dire, il serait plus rigoureux de parler de convergence de la mesure de probabilité.

14.5.1 Définition et première propriété

Définition 14.5.1 (Convergence en loi). *Soit $(X_n)_n$ une suite de variables aléatoires définies sur un espace (Ω, \mathbb{P}) . Soit X une autre variable aléatoire définie sur (Ω, \mathbb{P}) . On dit que la suite $(X_n)_n$ converge en loi vers X si pour toute fonction réelle, continue et bornée sur \mathbb{R} , on a la limite*

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] .$$

Notation 14.5.2. *Si $(X_n)_n$ converge en loi vers X , alors on note*

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X .$$

Proposition 14.5.3. *Soit $(X_n)_n$ une suite de variables aléatoires définies sur un espace (Ω, \mathbb{P}) . Soit X une autre variable aléatoire définie sur (Ω, \mathbb{P}) . On suppose que $(X_n)_n$ converge en loi vers X . Alors, pour toute fonction continue f , $(f(X_n))_n$ converge en loi vers $f(X)$.*

Démonstration. Il suffit de remarquer que pour toute fonction f continue et pour toute fonction continue et bornée ψ , alors $\xi := \psi \circ f$ est une fonction continue et bornée. De fait, la convergence de la suite de terme général $\mathbb{E}[\psi(f(X_n))] = \mathbb{E}[\xi(X_n)]$ vers $\mathbb{E}[\xi(X)] = \mathbb{E}[\psi(f(X))]$ est immédiate. \square

14.5.2 Remarque

Soit X une variable aléatoire qui suit la loi de Rademacher, voir page 95. On considère $X_n := -X$. Alors, X_n suit aussi la loi de Rademacher. On a immédiatement la convergence en loi de X_n vers X . Pourtant, on a $X_n = -X$ et donc on pourrait écrire que $-X$ converge vers X . Ainsi, cette convergence ne porte pas sur les variables aléatoires elles-mêmes mais bien sur les lois des variables aléatoires.

14.5.3 Liens avec les convergences précédentes

Théorème 14.5.4. *Soit $(X_n)_n$ une suite de variables aléatoires définies sur un espace (Ω, \mathbb{P}) . Soit X une autre variable aléatoire définie sur (Ω, \mathbb{P}) . On suppose que $(X_n)_n$ converge en probabilité vers X . Alors $(X_n)_n$ converge en loi vers X .*

Démonstration. On se donne une fonction f continue et bornée. On pose $x_n := \mathbb{E}[f(X_n)]$. Comme f est bornée, la suite $(x_n)_n$ est aussi bornée. Elle admet donc une valeur d'adhérence d'après le théorème de Bolzano-Weierstrass. Soit justement x_∞ une valeur d'adhérence de la suite. Alors, il existe une sous-suite $(x_{n_k})_k$ telle que $\lim_{k \rightarrow \infty} x_{n_k} = x_\infty$. Or, $X_{n_k} \xrightarrow[k \rightarrow +\infty]{\mathbb{P}} X$. On peut donc extraire une sous-suite de $(X_{n_k})_k$ qui converge presque sûrement vers X , d'après le Théorème 14.4.5. Notons $(X_{m_p})_p$ cette sous-suite. La continuité de la fonction f implique

$$f(X_{m_p}) \xrightarrow[p \rightarrow +\infty]{\text{p.s.}} f(X).$$

La fonction f étant bornée, on peut appliquer le théorème de convergence dominée de Lebesgue. Il vient $\lim_{p \rightarrow \infty} \mathbb{E}[f(X_{m_p})] = \mathbb{E}[f(X)]$. Or, $\lim_{p \rightarrow \infty} x_{m_p} = x_\infty$ donc $x_\infty = \mathbb{E}[f(X)]$. On conclut en rappelant que la suite $(x_n)_n$ est bornée puis comme elle a une unique valeur d'adhérence qui est $\mathbb{E}[f(X)]$, alors $x_n \rightarrow \mathbb{E}[f(X)]$. Ceci montre que X_n converge en loi vers X quand n tend vers l'infini. □

Il s'ensuit que la convergence presque sûre implique la convergence en loi. Et, il en est de même avec la convergence dans L^p . Présentons maintenant un résultat essentiel en vue des tests du khi-deux, à savoir le lemme de Slutsky.

Théorème 14.5.5. *Soit $(X_n)_n$ et $(Y_n)_n$ deux suites de variables aléatoires définies sur un espace (Ω, \mathbb{P}) . On suppose que $(X_n)_n$ converge en loi vers une variable aléatoire X . On suppose également que $(Y_n)_n$ converge en loi vers une constante $\lambda \in \mathbb{R}$. Alors, si l'on pose $Z_n := X_n + Y_n$ et $T_n := X_n Y_n$ pour tout $n \in \mathbb{N}$, on en déduit la convergence en loi de la suite $(Z_n)_n$ vers $X + \lambda$ ainsi que celle de la suite $(T_n)_n$ vers λX .*

14.5.4 Obtenir une convergence en loi

14.5.4.1 Convergence en loi et fonction de répartition

Théorème 14.5.6. *On suppose que $(X_n)_n$ converge en loi vers X . Alors, on dispose de la limite*

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

pour tout x tel que $F_X(x) = F_X(x^-)$.

14.5.4.2 Convergence en loi et fonction caractéristique

Théorème 14.5.7 (Premier théorème de Lévy). *Soit une suite de variables aléatoires $(X_n)_n$ et soit une variable aléatoire X . On note φ_n la fonction caractéristique de X_n et φ celle de X . Alors $(X_n)_n$ converge en loi vers X si et seulement si pour tout $t \in \mathbb{R}$, $\varphi_n(t) \rightarrow \varphi(t)$ quand n tend vers l'infini.*

Remarque 14.5.8. *L'énoncé précédent est en fait valable pour les vecteurs aléatoires en dimension d , avec $t \in \mathbb{R}^d$.*

On dispose également du théorème suivant.

Théorème 14.5.9. *Soit une suite de variables aléatoires $(X_n)_n$ et soit φ une fonction donnée. On suppose que pour tout $t \in \mathbb{R}$, $\varphi_{X_n}(t)$ converge vers $\varphi(t)$ quand n tend vers l'infini. On suppose de plus que la fonction φ est continue en 0. Alors, il existe une variable aléatoire X dont la loi est entièrement déterminée par φ telle que $\varphi_X = \varphi$ et de plus $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$.*

Remarque 14.5.10. *L'énoncé précédent est en fait valable pour les vecteurs aléatoires en dimension d , avec $t \in \mathbb{R}^d$.*

14.5.4.3 Convergence en loi et densité de probabilité

Dans le cas où X_n et X sont à densité, on peut caractériser la convergence en loi avec la densité de probabilité.

Théorème 14.5.11. *Soit une suite de variables aléatoires $(X_n)_n$ et soit X une variable aléatoire. On suppose que X_n admet une densité f_{X_n} et que X admet une densité f_X . Si la suite de fonctions $(f_{X_n})_n$ converge simplement vers f_X , alors X_n converge en loi vers X quand n tend vers l'infini.*

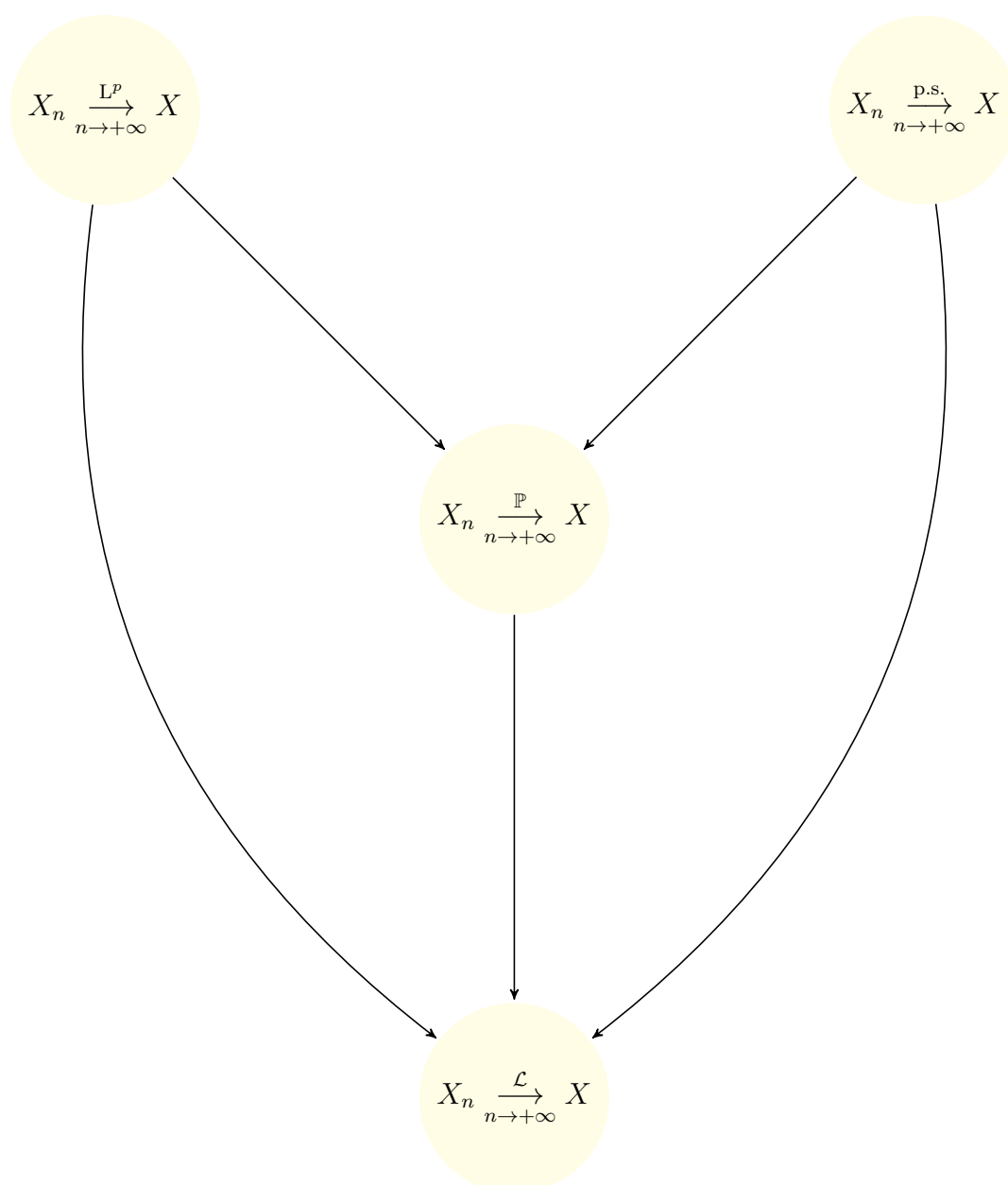
14.6 Retour sur le paradoxe

Dans l'exemple que l'on a donné à la page 200, l'on a exhibé une convergence presque sûre. Toutefois, on n'obtient pas un tel résultat en moyenne dans la mesure où les variables aléatoires $m_n X_n$ (qui représentent le gain au lancer n) ne sont pas bornées uniformément. En effet, $|m_n X_n| = |m_n| = 2^{n-1}$. Mieux, on verra avec la loi forte des grands nombres à la page 216 que le gain moyen tend vers 0 presque sûrement. En fait, on a simplement exhibé l'existence d'une suite croissante de temps d'arrêt $(\tau_n)_n$ à valeurs dans \mathbb{N} telle que le gain au temps τ_n est égal à n . Toute la question est de savoir si τ_n sera atteint avant la mort du joueur. Également, le joueur dispose-t-il d'assez de fonds pour doubler la mise à chaque fois ?

14.7 Synthèse

Dans la dernière section de ce chapitre, on donne un diagramme simple et qu'il faut connaître. Dans celui-ci, on peut voir les liens entre les différents modes de convergence.

FIGURE 14.1 – Liens entre les convergences



Grands nombres et théorème central

15.1 Introduction

Il est essentiel de savoir intégrer pour aborder ce chapitre. Évidemment, la non-maîtrise des treize chapitres précédents serait rédhibitoire.

Dans ce chapitre, nous présentons les deux principaux résultats impliquant une convergence de variables aléatoires : la loi des grands nombres ainsi que le théorème central limite.

Les objectifs sont donc les suivants : comprendre l'intérêt des lois des grands nombres mais aussi celui du théorème central de la limite en statistiques inférentielles. Il faut aussi savoir quand ces deux résultats ne peuvent pas s'appliquer. Notamment, on verra qu'il y a différentes lois des grands nombres avec différents jeux d'hypothèses. Connaître l'énoncé de la loi forte et de la loi faible des grands nombres ainsi que l'idée de la preuve de la loi faible (Inégalité de Bienaymé-Tchebychev) est essentiel. Il en est de même pour le théorème central de la limite : il faudra à l'issue du chapitre connaître son énoncé ainsi que l'idée de sa preuve à savoir l'utilisation de la fonction caractéristique.

15.2 Petit jeu

Avant tout, introduisons l'expérience aléatoire suivante.

Un casino propose un nouveau jeu à ses clients. Ce jeu est fondé sur celui du pile ou face. Aucune mise pour jouer mais le client s'engage sur l'honneur à payer ce qu'il doit en cas de défaite. On considère que le croupier commence avec un euro. À chaque tour, le croupier lance la pièce. Si elle fait pile, la fortune du croupier est divisée par quatre. Si au contraire, la pièce donne face, elle est multipliée par deux. Le jeu comporte cent tours.

Qui veut jouer ?

15.3 Lois des grands nombres

15.3.1 Motivation : méthodes de Monte Carlo

Un intérêt de la loi des grands nombres réside dans l'utilisation des méthodes de simulation des variables aléatoires pour l'intégration numérique. Soit f une fonction borélienne sur $[0; 1]$. On la suppose intégrable par rapport à la mesure de Lebesgue. Il n'est pas forcément simple de trouver la valeur formelle de l'intégrale

$$I_\infty := \int_0^1 f(x) dx.$$

La méthode des rectangles et celle des trapèzes sont bien connues mais elles ont le désavantage de dépendre de la dimension du support d'intégration. De plus, il faudrait, pour les utiliser, supposer de plus que f est suffisamment régulière (typiquement il faut qu'elle soit lipschitzienne).

Une autre méthode de calcul est de considérer une suite de variables aléatoires $(U_j)_{j \in \mathbb{N}^*}$ indépendantes et identiquement distribuées suivant la loi uniforme sur $[0; 1]$, $\mathcal{U}_{[0;1]}$. On pose alors :

$$I_n := \frac{1}{n} \sum_{j=1}^n f(U_j).$$

Il s'agit d'une variable aléatoire. Néanmoins, si l'on arrive à prouver une loi des grands nombres, on a alors la convergence, dans un sens à préciser, vers la quantité

$$I_\infty = \mathbb{E}[f(U)] = \int_0^1 f(x) dx.$$

On peut ensuite essayer de détendre les hypothèses sur l'indépendance afin d'optimiser la vitesse de convergence. Mais, c'est une autre histoire.

15.3.2 Loi faible des grands nombres

On commence par présenter la loi faible des grands nombres.

Théorème 15.3.1. *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées. On suppose $\mathbb{E}[X_1] = \mu$ et $\text{Var}[X_1] = \sigma^2 < \infty$. Soit $\overline{X}_n := \frac{\sum_{i=1}^n X_i}{n}$. Alors \overline{X}_n converge en probabilité vers μ .*

Démonstration. On a : $\mathbb{E}[\overline{X}_n] = \mu$ en utilisant la linéarité de l'espérance. Puis, d'après l'inégalité de Bienaymé-Tchebychev, il vient

$$\begin{aligned} \mathbb{P}(|\overline{X}_n - \mu| > \epsilon) &= \mathbb{P}(|\overline{X}_n - \mathbb{E}[\overline{X}_n]| > \epsilon) \\ &\leq \frac{\text{Var}[\overline{X}_n]}{\epsilon^2}. \end{aligned}$$

Or, on peut faire le calcul explicite suivant :

$$\begin{aligned}\text{Var} [\overline{X}_n] &= \frac{1}{n^2} \text{Var} \left[\sum_{i=1}^n X_i \right] \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var} [X_i] \\ &= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 \\ &= \frac{1}{n^2} n \sigma^2 \\ &= \frac{\sigma^2}{n}.\end{aligned}$$

Conséquemment, il vient :

$$\mathbb{P} (|\overline{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0,$$

ce qui achève la preuve. □

15.3.3 Loi forte des grands nombres

On peut toutefois aller plus loin, sous les mêmes hypothèses.

Théorème 15.3.2. *Sous les hypothèses du Théorème 15.3.1, \overline{X}_n converge dans L^2 vers μ .*

Démonstration. Quitte à considérer $X_n - \mu$ au lieu de X_n , on peut supposer sans perte de généralité que la moyenne μ est égale à 0.

La convergence dans L^2 de \overline{X}_n équivaut donc à la convergence vers 0 de $\mathbb{E} [(\overline{X}_n)^2]$ quand n tend vers l'infini. On peut calculer explicitement :

$$\mathbb{E} [(\overline{X}_n)^2] = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} [X_i X_j].$$

Or, les variables aléatoires sont indépendantes. Ainsi, $\mathbb{E} [X_i X_j] = \mathbb{E} [X_i] \mathbb{E} [X_j] = 0$ si $i \neq j$. Par conséquent :

$$\mathbb{E} [(\overline{X}_n)^2] = \frac{1}{n^2} \sum_{i=1}^n \mathbb{E} [X_i^2] = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n} \longrightarrow 0,$$

ce qui achève de prouver la convergence dans L^2 . □

Théorème 15.3.3. *Sous les hypothèses du Théorème 15.3.1 et si de plus $\mathbb{E}[X_1^4] = m_4 < \infty$, alors \overline{X}_n converge presque sûrement vers μ .*

Démonstration. On se place d'abord dans le cas où $\mu = 0$. On a alors :

$$\begin{aligned} \mathbb{E} \left[n^4 \overline{X}_n^4 \right] &= \mathbb{E} \left[\left(\sum_{i=1}^n X_i \right)^4 \right] \\ &= \sum_{i=1}^n \mathbb{E} [X_i^4] + 6 \sum_{1 \leq i < j \leq n} \mathbb{E} [X_i^2] \mathbb{E} [X_j^2] \\ &= nm_4 + 6 \frac{n(n-1)}{2} \sigma^4. \end{aligned}$$

En effet, les variables aléatoires étant centrées et indépendantes, les termes ne comprenant pas de carré sont nuls. Par conséquent :

$$\mathbb{E} \left[n^4 \overline{X}_n^4 \right] = nm_4 + 3n(n-1)\sigma^4.$$

De fait, $\mathbb{E} \left[\overline{X}_n^4 \right] \leq \frac{C}{n^2}$ où $C > 0$. Par conséquent, la série $\sum_{n=1}^{\infty} \mathbb{E} \left[\overline{X}_n^4 \right]$ converge d'où la variable aléatoire $Y := \sum_{n=1}^{\infty} \overline{X}_n^4$ admet un moment d'ordre 1. Ainsi, Y est presque sûrement finie et donc $\overline{X}_n^4 \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0$ ce qui implique la convergence presque sûre de \overline{X}_n vers 0 quand n tend vers l'infini.

Si $\mu \neq 0$, alors on applique la même méthode à $X'_i := X_i - \mu$. Le moment d'ordre 4 de X'_i est dans ce cas fini, ce qui achève la preuve. En effet :

$$\mathbb{E}[(X'_i)^4] = \mathbb{E}[(X_i - \mu)^4] \leq 16(\mathbb{E}[X_i^4] + \mu^4),$$

en utilisant $(a+b)^n \leq 2^n(|a|^n + |b|^n)$, résultat que l'on peut obtenir pour tout n par récurrence. □

15.3.4 Loi des grands nombres de Kolmogorov

On termine ce chapitre en donnant les hypothèses minimales assurant la validité de la loi des grands nombres pour les sommes de variables aléatoires indépendantes. Nous ne donnons pas la preuve car elle nécessite des connaissances supplémentaires, notamment sur les martingales.

Théorème 15.3.4. *Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées. Soit $\mu \in \mathbb{R}$. On pose $\overline{X}_n := \frac{\sum_{k=1}^n X_k}{n}$. Alors, la suite de terme général \overline{X}_n converge presque sûrement vers μ si et seulement si $\mathbb{E}[X_1] = \mu$. Dans ce cas, la convergence a aussi lieu dans L^1 .*

15.4 Théorème central de la limite

Comme pour la loi des grands nombres, on va commencer par donner une motivation du théorème central limite. Nous venons de voir les lois des grands nombres. On sait ainsi que \overline{X}_n converge presque sûrement, dans L^2 et en probabilité vers $\mathbb{E}[X_1]$ sous certaines hypothèses. Néanmoins, dans la pratique, obtenir une limite n'est pas suffisant. Il faut aussi quantifier la vitesse de convergence. C'est notamment l'objet du théorème central de la limite. Ce dernier stipule que la vitesse est de la forme $O\left(n^{-\frac{1}{2}}\right)$. En particulier, on peut obtenir des intervalles de confiance en se ramenant au cas du modèle gaussien.

15.4.1 Énoncé du théorème central limite

Théorème 15.4.1 (Théorème central de la limite). *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées avec $\mathbb{E}[X_1] = \mu$ et $\text{Var}[X_1] = \sigma^2$. On suppose $\sigma \in]0; \infty[$. On pose $Y_n := \frac{\sum_{k=1}^n X_k - n\mu}{\sigma\sqrt{n}}$. Alors, la suite de variables aléatoires $(Y_n)_n$ converge en loi vers la loi normale centrée et réduite. De manière équivalente :*

$$\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1) ,$$

$$\text{où } \overline{X}_n := \frac{1}{n} \sum_{i=1}^n X_i .$$

Remarque 15.4.2. *En pratique, $n > 30$ est suffisant.*

Remarque 15.4.3. *Si X_i suit la loi normale de paramètres μ et σ^2 , le résultat est immédiat de par la stabilité par la somme pour les lois normales.*

15.4.2 Preuve du théorème central limite

Soit φ la fonction caractéristique de $X_1 - \mu$. On note ψ_n la fonction caractéristique de Y_n . On a :

$$\begin{aligned} \psi_n(u) &:= \mathbb{E} \left\{ e^{iuY_n} \right\} \\ &= \mathbb{E} \left\{ \exp \left[iu \frac{S_n - n\mu}{\sigma\sqrt{n}} \right] \right\} = \mathbb{E} \left\{ \exp \left[i \frac{u}{\sigma\sqrt{n}} (S_n - n\mu) \right] \right\} \\ &= \mathbb{E} \left\{ \exp \left[i \frac{u}{\sigma\sqrt{n}} \sum_{k=1}^n (X_k - \mu) \right] \right\} = \prod_{k=1}^n \mathbb{E} \left\{ \exp \left[i \frac{u}{\sigma\sqrt{n}} (X_k - \mu) \right] \right\} \\ &= \prod_{k=1}^n \varphi \left(\frac{u}{\sigma\sqrt{n}} \right) = \left[\varphi \left(\frac{u}{\sigma\sqrt{n}} \right) \right]^n . \end{aligned}$$

Or, $\mathbb{E}[X_1 - \mu] = 0$ et $\text{Var}[X_1 - \mu] = \sigma^2 < \infty$. On en déduit que φ est de classe \mathcal{C}^2 et telle que

$$\varphi'(0) = 0 \quad \text{et} \quad \varphi''(0) = -\sigma^2.$$

On procède ensuite au développement de Taylor de la fonction φ à l'ordre deux au voisinage de 0 :

$$\varphi(u) = 1 - \frac{\sigma^2}{2}u^2 + u^2h(u),$$

avec $h(u) \rightarrow 0$ quand $u \rightarrow 0$. Puis, l'on a :

$$\begin{aligned} \psi_n(u) &= \left[\varphi\left(\frac{u}{\sigma\sqrt{n}}\right) \right]^n = \exp \left\{ n \log \left[\varphi\left(\frac{u}{\sigma\sqrt{n}}\right) \right] \right\} \\ &= \exp \left\{ n \log \left[1 - \frac{1}{2n}u^2 + \frac{1}{n\sigma^2}u^2h\left(\frac{u}{\sigma\sqrt{n}}\right) \right] \right\}, \end{aligned}$$

où \log désigne la valeur principale du logarithme complexe (valant 0 au point 1 et définie partout sauf sur la demi-droite négative). Quand n tend vers l'infini, il vient

$$\psi_n(u) \rightarrow e^{-\frac{u^2}{2}}.$$

Le Théorème 14.5.9 nous assure alors que la suite de variables aléatoires $(Y_n)_n$ converge en loi vers Z quand n tend vers l'infini où $\varphi_Z(u) = e^{-\frac{u^2}{2}}$. Il s'avère ainsi que Z suit de plus la loi normale centrée réduite.

15.4.3 Inégalité de Berry-Esseen

Se pose maintenant la question de la vitesse à laquelle converge la suite de variables aléatoires $(Y_n)_n$ vers la loi normale centrée réduite. Le théorème de Berry-Esseen fournit une réponse à cette question.

Théorème 15.4.4. *Il existe une constante universelle c telle que si $(X_n)_n$ est une suite de variables aléatoires indépendantes et identiquement distribuées satisfaisant $\mathbb{E}[|X_1|^3] < \infty$ et $\sigma^2 := \text{Var}[X_1] > 0$ avec $\sigma > 0$, alors :*

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq c \frac{\mathbb{E}[|X_1|^3]}{\sigma^3 \sqrt{n}}, \quad (15.1)$$

où $F_n(x) := \mathbb{P}\left(\frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right)$ avec $\mu := \mathbb{E}[X_1]$ et $\overline{X_n} := \frac{1}{n} \sum_{k=1}^n X_k$ et où Φ est la fonction de répartition de la loi normale centrée réduite.

Dans la pratique, la constante c peut être prise égale à 0.7975. La recherche d'un majorant le plus petit possible est donc l'objet de travaux de nombreux chercheurs. Il convient de noter que cette constante universelle c est supérieure à $\frac{3+\sqrt{10}}{6\sqrt{2\pi}} \approx 0.4097$.

Notons que l'on peut avoir une meilleure inégalité si l'on considère des variables aléatoires qui suivent la loi de Bernoulli :

Théorème 15.4.5. Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées suivant la loi de Bernoulli de paramètre p . On pose $F_n(x) := \mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\right)$ où $S_n := \sum_{k=1}^n X_k$. On a alors l'existence d'une constante universelle c telle que :

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq c \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}}.$$

Démonstration. En effet, on applique le Théorème 15.4.4 à $X'_n := X_n - p$ et l'on obtient

$$\mathbb{E}[|X'_n|^3] = (1-p) \times |p|^3 + p \times |1-p|^3 = p(1-p)(p^2 + (1-p)^2).$$

Et, $\sigma^2 = p(1-p)$. Ainsi, le terme de droite dans l'inégalité (15.1) devient

$$c \frac{\mathbb{E}[|X'_1|^3]}{\sigma^3 \sqrt{n}} = c \frac{p(1-p)(p^2 + (1-p)^2)}{\sqrt{n} \sqrt{p(1-p)}^3} = c \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}},$$

ce qui achève la preuve. □

Il est intéressant de noter que si $p = \frac{1}{2}$, alors l'inégalité devient

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{c}{\sqrt{n}}.$$

Comme $c \approx 0.7975$, prendre $n = 10^4$ suffit pour que $\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)|$ soit inférieur à 8.10^{-3} .

En revanche, si $p = 10^{-2}$, l'inégalité de Berry-Esseen devient :

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{78.96}{\sqrt{n}}.$$

Ainsi, pour s'assurer une précision de 10^{-2} , il faut $n \geq 6.10^7$, ce qui est énorme. On voit donc que le théorème central de la limite ne peut être utilisé à la légère pour la loi binomiale. C'est aussi pourquoi il arrive que la loi binomiale puisse être approchée par la loi de Poisson, voir la page 90.

15.5 Retour sur le jeu

Revenons-en maintenant au jeu de la page 213.

15.5.1 Premier raisonnement (faux)

Un premier raisonnement (faux) est de dire que si l'on joue 100 fois, l'on fera pile 50 fois et face 50 fois. Ainsi, la somme que possède le croupier au début du jeu (un euro) sera multipliée par $(\frac{1}{4})^{50} \times 2^{50} = (\frac{1}{2})^{50} \approx 0$.

Il s'ensuit que le joueur gagne 1 euro à coup sûr ; si l'on en croit ce premier raisonnement (faux).

15.5.2 Deuxième raisonnement (faux également)

Un deuxième raisonnement est d'utiliser la loi des grands nombres comme suit. On note X_i la variable aléatoire de loi $\frac{1}{2}\delta_{\frac{1}{4}} + \frac{1}{2}\delta_2$. Les variables aléatoires X_i sont indépendantes. Or, la moyenne de X_i est $\frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times 2 = \frac{9}{8} > 1$. Ainsi, si l'on joue 100 fois, en moyenne, la fortune du croupier est de $1 \times (\frac{9}{8})^{100} \approx 130\,392$ euros.

Le joueur n'a donc aucun intérêt à s'engager à perdre une telle somme en moyenne.

15.5.3 Troisième raisonnement

La fortune que le croupier possède au n -ième lancer de la pièce est $X_1 \times \dots \times X_n$ où les variables aléatoires X_i sont indépendantes et identiquement distribuées suivant la loi $\frac{1}{2}\delta_{\frac{1}{4}} + \frac{1}{2}\delta_2$.

Conséquemment, les variables aléatoires $\log(X_i)$ sont indépendantes mutuellement et suivent la loi $\frac{1}{2}\delta_{-2\log(2)} + \frac{1}{2}\delta_{\log(2)}$. Ainsi, $\mathbb{E}[\log(X_1)] = -\frac{\log(2)}{2}$. On en déduit la convergence presque sûre de $\frac{1}{n} \sum_{i=1}^n \log(X_i)$ vers $-\frac{\log(2)}{2}$ si bien que $\sum_{i=1}^n \log(X_i)$ tend presque sûrement vers $-\infty$ d'où $X_1 \times \dots \times X_n = \exp(\sum_{i=1}^n \log(X_i))$ tend presque sûrement vers 0.

Ainsi, presque sûrement, le joueur gagne un euro. Ce raisonnement est en quelque sorte le premier, lequel est justifié par la loi des grands nombres.

15.5.4 Synthèse

Comment concilier le deuxième raisonnement et le troisième ?

En fait, l'espérance ne prend un intérêt que si l'on répète l'expérience un grand nombre de fois. Or, ici, l'expérience consiste à considérer $X_1 \times \dots \times X_n$ une fois et une fois seulement. Par conséquent, on ne se donne qu'une trajectoire par joueur. La loi des grands nombres ne peut donc pas s'appliquer pour le joueur.

Cela dit, s'il y a un grand nombre de joueurs et de croupiers, la plupart des croupiers vont perdre un euro et feront donc gagner la somme de un euro à de nombreux joueurs. Toutefois, certains joueurs malchanceux vont perdre beaucoup d'argent. Le casino va donc s'y retrouver car là, on peut appliquer la loi des

grands nombres en considérant $\frac{1}{N} \sum_{i=1}^N Y_i$ où $Y_i := \prod_{k=1}^{100} X_{i,k}$ avec N le nombre de joueurs.

À la question de savoir si ce jeu en vaut la peine, nous répondrons que le joueur gagne dans la plupart des cas un euro. Mais s'il perd, il perd gros.

Deuxième partie
Statistiques

Statistiques descriptives

16.1 Introduction

Les statistiques descriptives correspondent au premier chapitre que nous verrons sur les statistiques. L'objet principal des statistiques descriptives est de décrire comme leur nom l'indique. En d'autres termes, on dispose de données et l'on cherche à dégager certaines tendances à partir de celles-ci. C'est une première étape avant de pouvoir inférer les lois sous-jacentes qui régissent lesdites données.

Il n'y a pas de pré-requis à proprement parler à ce chapitre. Toutefois, avoir survolé la première partie du livre sur les probabilités peut être un plus pour faire le lien entre les statistiques et les probabilités.

D'ailleurs, l'un des objectifs dudit chapitre est de sentir le lien profond qui unit les deux domaines. Ainsi, les fréquences correspondent aux probabilités, les fréquences cumulées sont une discrétisation de la fonction de répartition, les séries à valeurs isolées sont des réalisations de variables aléatoires discrètes tandis que les séries à valeurs classées sont modélisées par des variables aléatoires à densité. De même, la fréquence conditionnelle correspond à une probabilité conditionnelle. Il est aussi attendu lecteur qu'il se familiarise avec le vocabulaire propre aux statistiques et qu'il sache à tout moment quelle est la représentation graphique adaptée pour représenter de façon palpable les séries (camembert, diagramme en barres, diagramme en bâtons, histogramme, boîte à moustaches...). Un autre objectif est de savoir calculer une moyenne ou une variance (ainsi que la variance corrigée) ainsi que la covariance et les quantiles. Par ailleurs, l'indépendance des caractères est cruciale ; bien qu'elle ne soit quasiment jamais vérifiée au sens strict. Enfin, la lecture instantanée d'un tableau de contingence quand on est face à une série statistique double est essentiel.

16.2 Vocabulaire

Dans cette section, on introduit le vocabulaire propre aux statistiques et plus précisément aux statistiques descriptives. Ces dernières ont pour objet de décrire des caractéristiques dans une population donnée. On parle d'enquête statistique.

Une enquête statistique consiste à effectuer des observations sur des éléments

d'un ensemble (la population) que l'on appelle individus de la population. Par exemple, on peut considérer l'ensemble des pièces fabriquées par une machine et l'on souhaite savoir si elles sont défectueuses. Il peut aussi s'agir de la population des actifs en France et l'on veut décrire la répartition des différentes catégories socio-professionnelles.

On peut distinguer deux types d'enquêtes statistiques.

- On peut faire des observations sur tous les individus de la population. On dit alors que l'on fait un recensement. C'est ce qui est fait en démographie. Les premières études statistiques ont été les recensements de la population par les états. Ce type d'enquêtes est le plus naturel et il ne nécessite aucune inférence statistique puisque l'échantillon est la population elle-même. Néanmoins, on ne peut pas toujours effectuer de recensement.
- On peut faire des observations sur des individus tirés de la population qui constituent un échantillon. On dit alors que l'on fait un sondage. Par exemple, certains contrôles qualité détruisent le produit. Dans ce cas, on ne les pratique pas sur tous les produits. Également, il peut être coûteux en temps de faire un recensement. Ainsi, un sondage pré-électoral ne peut pas être fait sur tous les individus de la population.

En général, on observe sur des individus des variables. En statistiques, on parle plutôt de caractères.

Il y a plusieurs types de caractères. Le premier type auquel on peut penser est celui des caractères dits qualitatifs. Ici, les possibilités (on parle plutôt de modalités) que peut avoir le caractère qualitatif ne sont pas des nombres. C'est ainsi le cas dans une enquête d'opinion. On peut donner l'exemple de l'avis des étudiants sur un cours magistral à Télécom Saint-Étienne ; les modalités pouvant être "Très Bien", "Bien", "Moyen", "Mauvais", "Très Mauvais" ou "Absent en CM". On peut objecter à cet exemple qu'un ordre naturel apparaît dans les modalités. Ainsi, bien que le caractère soit qualitatif, il peut être vu comme s'il était à valeurs numériques. Ici, "Très Bien" serait 20, "Bien" serait 16, "Moyen" serait 12, "Mauvais" serait 8 et "Très Mauvais" serait 4 tandis que la modalité "Absent en CM" pourrait être associée à 0.

Dans ce cas, on dira que le caractère est qualitatif ordinal. Un autre exemple est celui de la qualité gustative dans la population des vins : on peut établir un ordre naturel (très bonne, bonne, moyenne, médiocre, mauvaise, très mauvaise).

Au contraire, si l'on ne peut pas établir un ordre naturel, on parlera de caractère qualitatif nominal. Ainsi, la catégorie socio-professionnelle de la population des actifs est un caractère qualitatif nominal. De même, le genre d'un enfant parmi la population des nouveau-nés en France en 2022 est un caractère qualitatif nominal.

Enfin, si les modalités sont des nombres, on dit que le caractère est quantitatif. Ainsi, la taille des nouveau-nés en France en 2022 est un caractère quantitatif. Il en est de même du poids. Si on prend l'exemple des actifs en France, l'âge est un caractère quantitatif également. Globalement, un caractère quantitatif peut être vu comme la réalisation d'une variable aléatoire.

Comme pour les variables aléatoires, on distingue les caractères quantitatifs en fonction de l'ensemble des modalités possibles. Si l'ensemble est fini ou infini dénombrable (par exemple, la durée de vie en années d'une ampoule), on dit que le caractère est discret. Si l'ensemble des modalités que peut prendre un caractère quantitatif est infini non dénombrable, on dit que le caractère est continu.

En réalité, en statistiques, la simplicité de la lecture étant le maître-mot, la différence est plus ténue. Par exemple, la taille en cm est un caractère quantitatif discret sur la population des actifs. Néanmoins, le nombre de valeurs étant très élevé, il n'est pas envisageable de représenter ce caractère quantitatif comme s'il s'agissait du nombre d'enfants (que l'on peut, sans trop de restrictions, majorer par 10) au sein des ménages français. Il est alors pertinent de le considérer comme étant continu. Il est crucial de comprendre que c'est l'usage qui déterminera souvent si le caractère sera à considérer comme discret ou comme continu.

Enfin, les données peuvent être vues de deux manières. On peut les garder telles quelles ou on peut les grouper. Plutôt que d'essayer de théoriser, donnons un exemple. On s'intéresse ainsi aux notes obtenues sur 20 par 15 étudiants. Voici les données brutes :

10, 20, 12, 10, 20, 15, 16, 20, 10, 3, 4, 15, 20, 12, 12.

Lire des données est faisable car le nombre 15 n'est pas trop élevé. Néanmoins, dessiner des tendances n'est pas immédiat. On les groupe donc comme suit pour obtenir les données dites groupées :

3(1), 4(1), 10(3), 12(3), 15(2), 16(1), 20(4).

Le nombre entre parenthèses correspond au nombre d'occurrences de la modalité qui la précède. On peut ensuite présenter de manière plus adéquate pour vraiment étudier cette série.

16.3 Caractère qualitatif

Dans le cas où le caractère sous-jacent est qualitatif, les notions de médiane, de moyenne ou même d'écart-type (que nous verrons par la suite) n'ont ni légitimité ni sens. De fait, dans cette section, nous verrons exclusivement un moyen de représenter graphiquement les séries statistiques ayant des modalités non numériques.

La représentation utilisée est celle du diagramme circulaire, aussi appelé diagramme en secteurs ou camembert. Ce diagramme se prête bien au cas où il y a un nombre fini et peu élevé de modalités r . Alors, on trace un disque partitionné en différents secteurs. Chaque secteur représente l'une des modalités. Et, l'angle du secteur est proportionnel à la fréquence f_k de chaque modalité a_k pour $k \in \llbracket 1; r \rrbracket$. De fait, il en est de même pour l'aire dudit secteur. Ici, la fréquence est $f_k := \frac{n_k}{n}$ où n_k est le nombre d'individus ayant la modalité a_k tandis que n est le nombre total d'individus de la population. Pour simplifier, on utilise plus facilement les

degrés que les radians. En effet, cela ne change rien vu que l'on passe de l'un à l'autre linéairement.

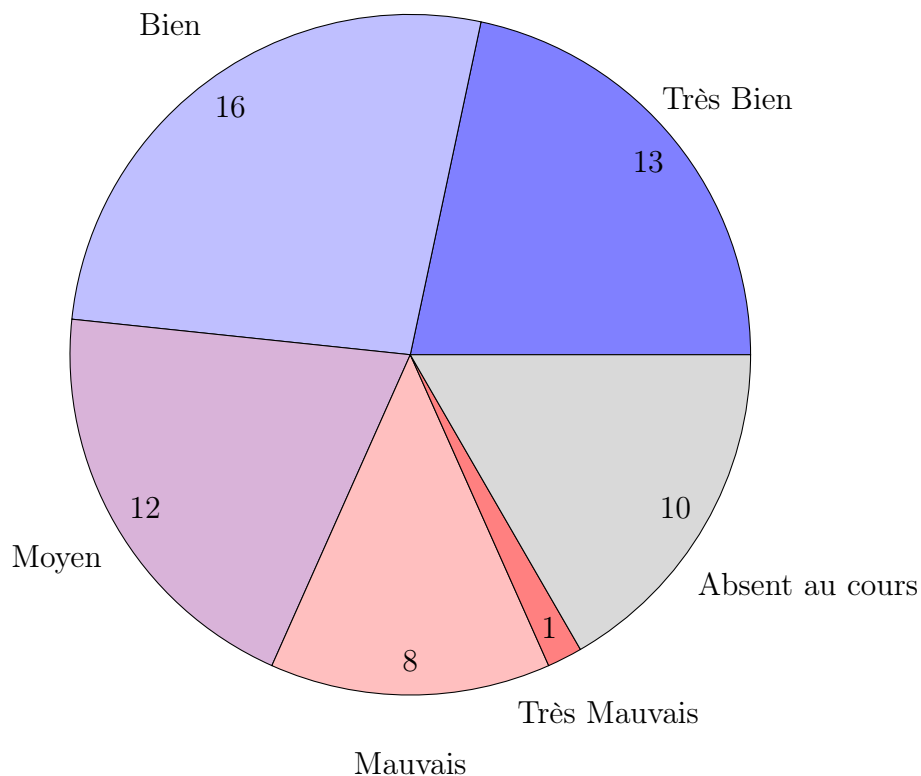
Exemple 16.3.1. *Pour un bilan de bloc à Télécom Saint-Étienne, les délégués réalisent un sondage sur l'appréciation d'un cours magistral. Les modalités sont : Très Bien (TB); Bien (B); Moyen; Mauvais (M); Très Mauvais (TM); Absent au cours (A). Sur les cent dix étudiants, soixante ont répondu. Et, voici la répartition des avis :*

TABLE 16.1 – Avis des soixantes étudiants

Avis	TB	B	Moyen	M	TM	A
Nombre d'étudiants	13	16	12	8	1	10

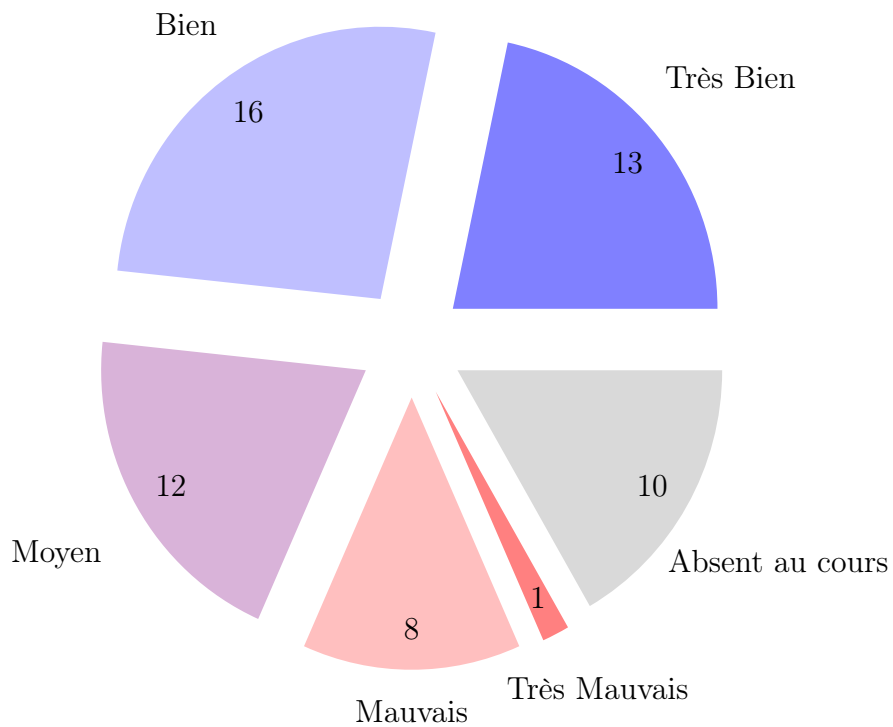
On obtient ainsi le diagramme circulaire suivant :

FIGURE 16.1 – Avis des étudiants sur le cours magistral - 1



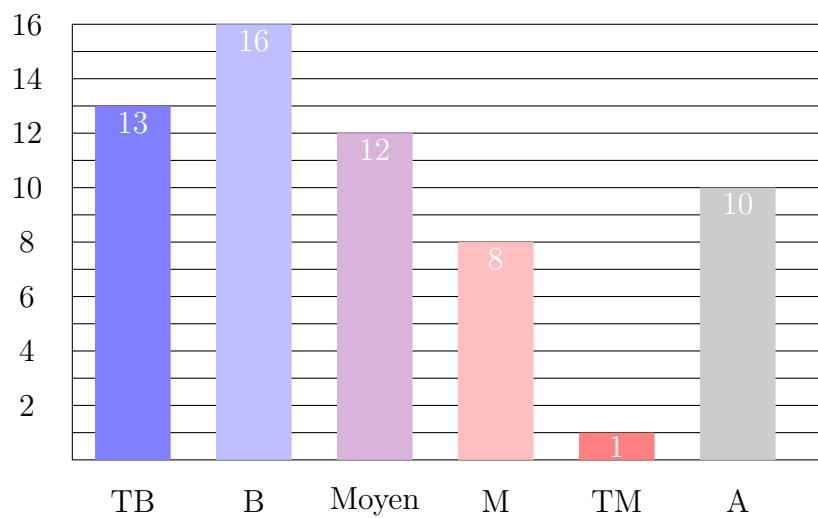
On peut aussi “explorer” le diagramme pour le rendre plus lisible :

FIGURE 16.2 – Avis des étudiants sur le cours magistral - 2



Il convient de noter que l'on peut aussi utiliser un diagramme en barres, lequel est, la plupart du temps, plus facile à lire.

FIGURE 16.3 – Avis des étudiants sur le cours magistral - 3



16.4 Séries statistiques simples : présentation

Passons aux caractères quantitatifs. On a observé sur n individus d'une population I un caractère quantitatif x dont on note les valeurs observées : $x[1], \dots, x[n]$. La valeur observée de l'individu i est $x[i]$.

Définition 16.4.1. *La suite de valeurs $(x[1], \dots, x[n])$ est appelée une série statistique simple.*

16.4.1 Série à valeurs isolées

Les séries statistiques simples à valeurs isolées sont utilisées pour les caractères quantitatifs discrets. Plus exactement, on les utilise pour les caractères quantitatifs dont le nombre de modalités possibles n'est pas trop élevé. Évidemment, la notion de "nombre pas trop élevé" est subjective et dépend de l'usage que l'on en fait.

On regroupe les valeurs égales de la série, on note l'effectif de chaque valeur isolée et on les range par ordre croissant. On suppose qu'il y a r valeurs différentes : $y_1 < \dots < y_k < \dots < y_r$.

TABLE 16.2 – Série à valeurs isolées

Valeurs isolées	y_1	y_2	\dots	y_k	\dots	y_r
Effectifs	n_1	n_2	\dots	n_k	\dots	n_r
Fréquences	$f_1 = \frac{n_1}{n}$	$f_2 = \frac{n_2}{n}$	\dots	$f_k = \frac{n_k}{n}$	\dots	$f_r = \frac{n_r}{n}$

Proposition 16.4.2. *D'abord, $r \leq n$. Ensuite : $\sum_{k=1}^r n_k = n$ et $\sum_{k=1}^r f_k = 1$.*

Ici, f_k est la proportion d'éléments de la série égaux à y_k .

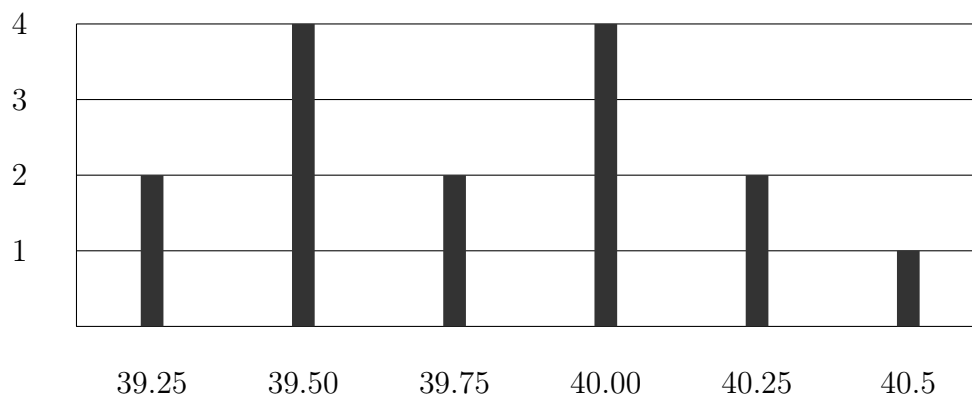
On considère un exemple avec $n = 15$ et $r = 6$:

TABLE 16.3 – Exemple de série à valeurs isolées

y_i	n_i (effectifs)	f_i (fréquence)	N_i	F_i
39.25	2	$\frac{2}{15}$	2	$\frac{2}{15}$
39.50	4	$\frac{4}{15}$	6	$\frac{6}{15}$
39.75	2	$\frac{2}{15}$	8	$\frac{8}{15}$
40.00	4	$\frac{4}{15}$	12	$\frac{12}{15}$
40.25	2	$\frac{2}{15}$	14	$\frac{14}{15}$
40.50	1	$\frac{1}{15}$	15	$\frac{15}{15}$

On représente cette série par le diagramme en bâtons des effectifs :

FIGURE 16.4 – Effectifs d’une série statistique simple



On pourrait aussi utiliser le diagramme en bâtons des fréquences. Cela revient au même pour peu que l'on connaisse le nombre n d'individus de la population.

Définition 16.4.3. Les points où l'on a un maximum relatif (local) de l'effectif (ou de la fréquence) sont appelés des modes.

Remarque 16.4.4. La plurimodalité, c'est-à-dire le fait d'avoir plusieurs modes, signifie que la population que l'on regarde n'est pas homogène. Ainsi, il est judicieux de la scinder pour récupérer des sous-populations homogènes.

Dans le Tableau 16.3, les modes sont en 39.50 et 40.

On calcule les effectifs cumulés et les fréquences cumulées (l'équivalent de la fonction de répartition) :

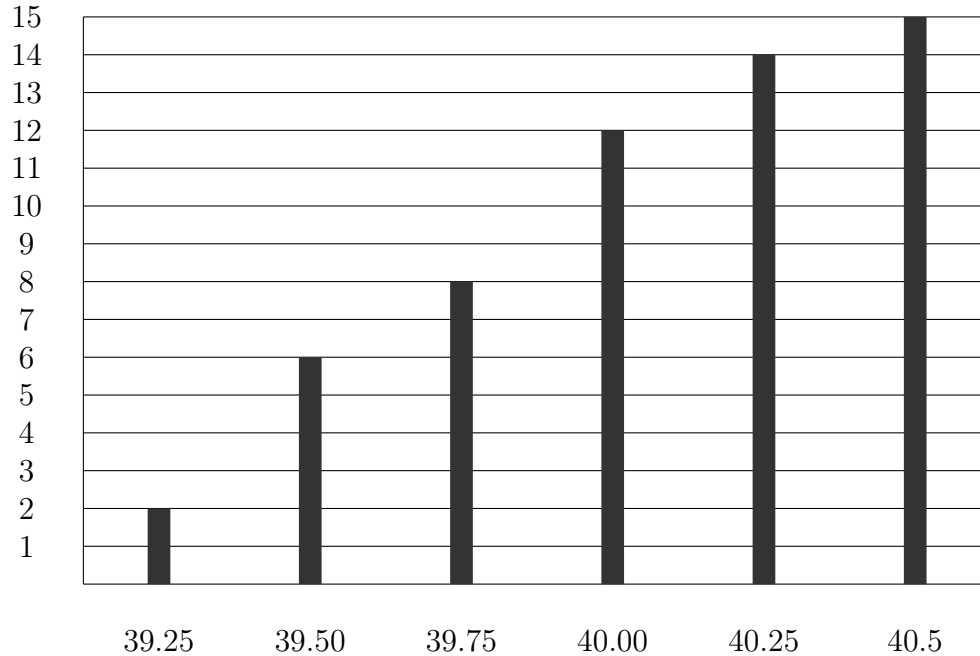
$$N_k := n_1 + \cdots + n_k = \sum_{l=1}^k n_l,$$

et

$$F_k := f_1 + \cdots + f_k = \sum_{l=1}^k f_l = \frac{N_k}{n}.$$

N_k est le nombre d'éléments de la série qui sont inférieurs ou égaux à y_k . Et, F_k est la proportion de tels éléments. Ainsi, douze éléments de la série sont inférieurs ou égaux à 40. On construit ensuite le diagramme en bâtons des effectifs cumulés (ou celui des fréquences cumulées; ce qui revient au même pour peu que l'on connaisse n) :

FIGURE 16.5 – Effectifs cumulés d’une série statistique simple



16.4.2 Série à valeurs classées

Les séries à valeurs classées sont utilisées pour les caractères quantitatifs continus ou pour les caractères quantitatifs discrets qui prennent un grand nombre de valeurs possibles. Comme précédemment, la notion de “grand nombre de valeurs possibles” est à l’appréciation de celui qui étudie la série statistique.

On regroupe les éléments de la série dans des intervalles semi-ouverts $[z_k; z_{k+1}[$ appelés classes; où $z_{k-1} < z_k$ pour tout $k \in \llbracket 1; s \rrbracket$. Et, on note l’effectif (et la fréquence de chaque classe). On considère s classes.

TABLE 16.4 – Série à valeurs classées

Classes	$[z_0; z_1[$	$[z_1; z_2[$	\cdots	$[z_{k-1}; z_k[$	\cdots	$[z_{s-1}; z_s[$
Effectifs	n_1	n_2	\cdots	n_k	\cdots	n_s
Fréquences	$f_1 = \frac{n_1}{n}$	$f_2 = \frac{n_2}{n}$	\cdots	$f_k = \frac{n_k}{n}$	\cdots	$f_s = \frac{n_s}{n}$

La quantité f_k représente la proportion d’éléments de la série qui sont dans la classe $[z_{k-1}; z_k[$.

Proposition 16.4.5. *D'abord, $s \leq n$. Ensuite :*

$$\sum_{k=1}^s n_k = n \quad \text{et} \quad \sum_{k=1}^s f_k = 1.$$

On donne un exemple avec $n = 80$:

TABLE 16.5 – Exemple de série à valeurs classées

$[z_{k-1}; z_k[$	n_k	f_k	N_k	F_k
$[100; 120[$	10	$\frac{10}{80}$	10	$\frac{10}{80}$
$[120; 130[$	15	$\frac{15}{80}$	25	$\frac{25}{80}$
$[130; 140[$	25	$\frac{25}{80}$	50	$\frac{50}{80}$
$[140; 150[$	20	$\frac{20}{80}$	70	$\frac{70}{80}$
$[150; 170[$	10	$\frac{10}{80}$	80	$\frac{80}{80} = 1$

Remarque 16.4.6. *La répartition des éléments de la série à l'intérieur d'une même classe ne doit pas être trop éloignée d'une répartition uniforme. Cette hypothèse de distribution uniforme au sein des classes est nécessaire pour effectuer des calculs sur la moyenne et la variance. De même, le calcul des quantiles nécessite que la courbe des fréquences cumulées soit affine par morceaux et donc qu'au sein de chaque classe la répartition soit uniforme. De fait, il est important que les classes ne soient pas trop larges non plus.*

On regarde maintenant l'histogramme des effectifs (ou des fréquences). En abscisse, on met les bornes des classes et en ordonnée les effectifs par unité de longueur de classe (effectifs corrigés). Ainsi, il est essentiel de bien comprendre que l'on ne représente pas simplement l'histogramme tel qu'il a pu être vu au lycée.

L'objectif principal de cette renormalisation est d'approximer la densité de probabilité de la variable aléatoire continue sous-jacente.

La première chose à faire est de choisir une unité adéquate. Par exemple, ici, les valeurs de z_k vont de 100 à 170. Ainsi, on pose $u = 10$.

Proposition 16.4.7. *L'aire du rectangle correspondant à $[z_{k-1}; z_k[$ dans l'histogramme des effectifs corrigés est proportionnelle à l'effectif de la classe. En effet, il vaut $A_k = (z_k - z_{k-1}) \times \frac{n_k \times u}{z_k - z_{k-1}} = un_k$. On en déduit par ailleurs que l'aire totale est égale à $un_1 + \dots + un_k + \dots + un_s = un$.*

On peut noter que pour l'histogramme des fréquences corrigées, en prenant $u = 1$, l'aire totale vaut 1 ; comme c'est le cas pour toute densité de probabilité.

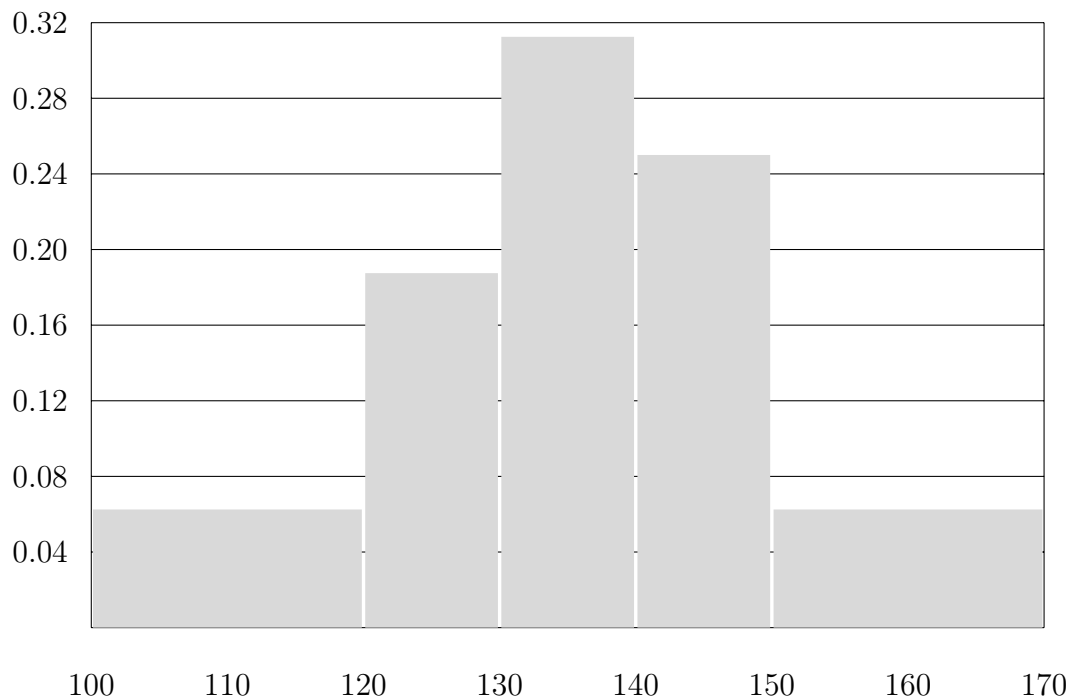
Pour l'effectuer, on reprend le tableau 16.5 :

TABLE 16.6 – Série à valeurs classées

$[z_{k-1}; z_k[$	n_k	f_k	$z_k - z_{k-1}$	$\frac{z_k - z_{k-1}}{u}$	$\frac{f_k \times u}{z_k - z_{k-1}}$ (fréquences corrigées)
$[100; 120[$	10	$\frac{10}{80}$	20	2	$\frac{5}{80}$
$[120; 130[$	15	$\frac{15}{80}$	10	1	$\frac{15}{80}$
$[130; 140[$	25	$\frac{25}{80}$	10	1	$\frac{25}{80}$
$[140; 150[$	20	$\frac{20}{80}$	10	1	$\frac{20}{80}$
$[150; 170[$	10	$\frac{10}{80}$	20	2	$\frac{5}{80}$

On obtient ainsi

FIGURE 16.6 – Histogramme des fréquences corrigées



Définition 16.4.8. Les classes pour lesquelles on a un maximum relatif ou local de l'effectif corrigé sont appelées "classes modales". On prend pour modes les milieux des classes modales.

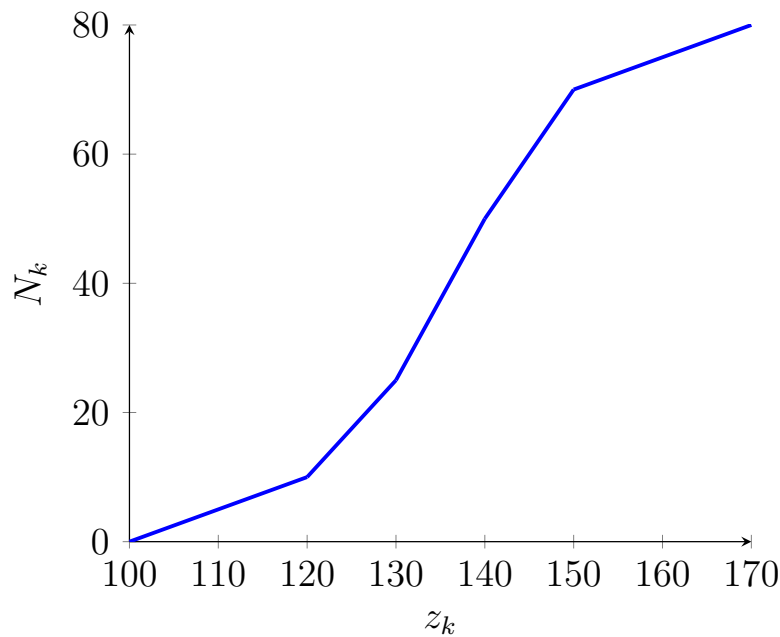
Dans l'exemple plus haut, la classe modale est $[130; 140[$ et le mode est donc 135.

Les effectifs cumulés et les fréquences cumulées sont calculés de la même manière que pour les séries statistiques simples à valeurs isolées. Ainsi, cinquante éléments sont strictement inférieurs à 140. N_k (respectivement F_k) représente le nombre (respectivement la proportion) d'éléments de la série strictement inférieurs à z_k .

On obtient alors la courbe des effectifs cumulés (ou celle des fréquences cumulées) en joignant par des segments de droite les points (z_k, N_k) (ou (z_k, F_k)). En effet, comme on suppose que la répartition au sein d'une classe est uniforme, la fonction de répartition sous-jacente est affine par morceaux.

Avec le Tableau 16.5, on a :

FIGURE 16.7 – Courbe des effectifs cumulés

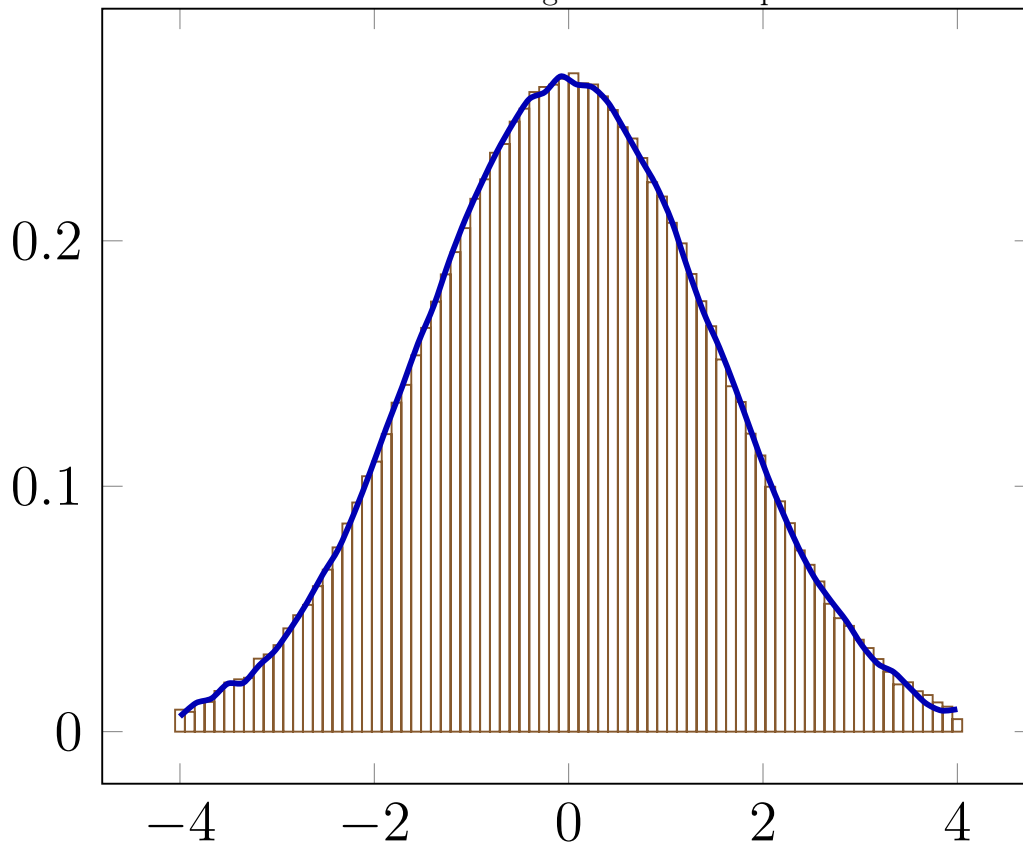


Par convention, on pose $N_0 = F_0 = 0$.

Remarque 16.4.9. *La courbe des fréquences cumulées peut s'assimiler à la fonction de répartition.*

Remarque 16.4.10. *Il est généralement plus simple de considérer que la longueur des classes est constante. Alors, il n'y a pas besoin de corriger les effectifs, ou les fréquences. Voici un exemple d'histogramme où chaque classe est de longueur 0.1 :*

FIGURE 16.8 – Histogramme des fréquences



Par ailleurs, la densité de probabilité empirique a été superposée, en bleu.

16.5 Séries statistiques simples : caractéristiques

On se donne une série statistique simple $(x[1], \dots, x[i], \dots, x[n])$.

16.5.1 Moyenne arithmétique

Définition 16.5.1. On appelle *moyenne arithmétique* de cette série le nombre égal à la somme des éléments de la série divisé par l'effectif n . On note \bar{x} cette moyenne arithmétique :

$$\bar{x} = \frac{\sum_{i=1}^n x[i]}{n}.$$

Cette quantité correspond à l'espérance dans le cas d'une variable aléatoire réelle discrète.

Remarque 16.5.2. La moyenne est sensible aux valeurs extrêmes (qui peuvent être aberrantes) de la série.

16.5.1.1 Calcul dans le cas d'une série à valeurs isolées

On suppose que l'on dispose de r modalités $y_1, \dots, y_k, \dots, y_r$ et l'effectif associé à la modalité y_k est n_k . La fréquence est ainsi $f_k = \frac{n_k}{n}$.

On peut montrer facilement, en utilisant la commutativité de la somme, la proposition suivante :

Proposition 16.5.3. *Dans ce cas, la moyenne arithmétique est :*

$$\bar{x} = \frac{\sum_{k=1}^r n_k y_k}{n} = \sum_{k=1}^r f_k y_k.$$

Exemple 16.5.4. *On reprend le Tableau 16.3 :*

y_k	n_k (effectifs)	f_k (fréquence)	N_k	F_k
39.25	2	$\frac{2}{15}$	2	$\frac{2}{15}$
39.50	4	$\frac{4}{15}$	6	$\frac{6}{15}$
39.75	2	$\frac{2}{15}$	8	$\frac{8}{15}$
40.00	4	$\frac{4}{15}$	12	$\frac{12}{15}$
40.25	2	$\frac{2}{15}$	14	$\frac{14}{15}$
40.50	1	$\frac{1}{15}$	15	$\frac{15}{15}$

Alors, la moyenne arithmétique est égale à

$$\begin{aligned} \bar{x} &= \frac{\sum_{k=1}^r n_k y_k}{n} \\ &= \frac{2 \times 39.25 + 4 \times 39.50 + 2 \times 39.75 + 4 \times 40.00 + 2 \times 40.25 + 1 \times 40.50}{15} \\ &= 39.80. \end{aligned}$$

Remarque 16.5.5. *Dans l'exemple précédent, le calcul de la moyenne pouvait sembler difficile à premier abord à la vue des valeurs. Néanmoins, on peut accélérer les calculs (pour les effectuer à la main) de la façon suivante.*

On remarque que la moyenne tourne autour de 40. Puis, on écrit :

$$\begin{aligned} \bar{x} - 40 &= \frac{2 \times (-0.75) + 4 \times (-0.5) + 2 \times (-0.25) + 4 \times 0 + 2 \times 0.25 + 1 \times 0.5}{15} \\ &= \frac{-1.5 + (1 - 4) \times 0.5}{15} \\ &= \frac{-1.5 - 1.5}{15} \\ &= -0.2. \end{aligned}$$

Conséquemment, on a $\bar{x} = 40 - 0.2 = 39.8$.

16.5.1.2 Calcul dans le cas d'une série à valeurs classées

On suppose que l'on dispose de s classes $[z_0; z_1[$, \dots , $[z_{k-1}; z_k[$, \dots , $[z_{s-1}; z_s[$ et l'effectif associé à la classe numéro k est n_k . La fréquence est ainsi $f_k = \frac{n_k}{n}$.

Ici, il est vraiment crucial que les répartitions au sein des différentes classes soient uniformes. En effet, on fait l'approximation consistant à supposer que pour toute classe $[z_{k-1}; z_k[$, les éléments de la série à l'intérieur de cette classe sont égaux à $\frac{z_{k-1} + z_k}{2}$. On est alors ramené au cas précédent.

Proposition 16.5.6. *Si la répartition au sein de chaque classe est uniforme, alors la moyenne $\bar{x} = \frac{\sum_{i=1}^n x^{[i]}}{n}$ peut être approchée par*

$$\frac{1}{n} \sum_{k=1}^s n_k \frac{z_{k-1} + z_k}{2} = \sum_{k=1}^s f_k \frac{z_{k-1} + z_k}{2}.$$

On ne peut pas donner une preuve à proprement parler. En effet, il s'agit ici d'une approximation et les deux valeurs ne peuvent être égales à moins que la répartition soit exactement uniforme, ce qui suppose que n soit égal à l'infini.

On va donc donner l'idée sous-jacente.

On note $U_1, \dots, U_k, \dots, U_s$ s variables aléatoires indépendantes. On suppose que la loi de U_k est la loi uniforme sur l'intervalle $[z_{k-1}; z_k[$. Il est important de les supposer indépendantes car *a priori*, il n'y a aucune raison pour que les fluctuations d'un intervalle influent sur celles d'un autre.

On suppose également que l'on a une variable aléatoire \mathcal{K} , indépendante des $(U_k)_{k \in [1; s]}$ et telle que $\mathbb{P}(\mathcal{K} = k) = f_k$ pour tout $k \in [1; s]$. Il n'est pas difficile de voir que la variable aléatoire sous-jacente est $U_{\mathcal{K}}$. Et, son espérance (laquelle est la limite pour n tendant vers l'infini de la moyenne arithmétique, voir Section 15.3 à la page 214) vaut, en utilisant le système complet d'évènements naturel associé à \mathcal{K} :

$$\begin{aligned} \mathbb{E}[U_{\mathcal{K}}] &= \sum_{k=1}^s \mathbb{E}[U_{\mathcal{K}} \mathbf{1}_{\mathcal{K}=k}] \\ &= \sum_{k=1}^s \mathbb{E}[U_k \mathbf{1}_{\mathcal{K}=k}] \\ &= \sum_{k=1}^s \mathbb{E}[U_k] \mathbb{P}(\mathcal{K} = k) \\ &= \sum_{k=1}^s f_k \mathbb{E}[U_k], \end{aligned}$$

en utilisant l'indépendance entre \mathcal{K} et les variables aléatoires U_1, \dots, U_s . Or, l'espérance de U_k est $\frac{z_k + z_{k-1}}{2}$; ce qui achève de prouver que la variable aléatoire sous-jacente a bien $\sum_{k=1}^s f_k \frac{z_{k-1} + z_k}{2}$ comme espérance.

On reprend le Tableau 16.5 :

$[z_{k-1}; z_k[$	n_k	f_k	N_k	F_k
$[100; 120[$	10	$\frac{10}{80}$	10	$\frac{10}{80}$
$[120; 130[$	15	$\frac{15}{80}$	25	$\frac{25}{80}$
$[130; 140[$	25	$\frac{25}{80}$	50	$\frac{50}{80}$
$[140; 150[$	20	$\frac{20}{80}$	70	$\frac{70}{80}$
$[150; 170[$	10	$\frac{10}{80}$	80	$\frac{80}{80} = 1$

Alors, on peut calculer comme suit

$$\bar{x} = \frac{10 \times 110 + 15 \times 125 + 25 \times 135 + 20 \times 145 + 10 \times 160}{80} = 135.625.$$

À nouveau, on peut utiliser une “valeur centrale” (le mode 135 typiquement) pour calculer sans faire appel à des outils informatiques :

$$\begin{aligned} \bar{x} - 135 &= \frac{10 \times (-25) + 15 \times (-10) + 25 \times 0 + 20 \times 10 + 10 \times 25}{80} \\ &= \frac{(10 - 10) \times 25 + (20 - 15) \times 10}{80} \\ &= \frac{5}{8} \\ &= 0.625. \end{aligned}$$

On a donc bien $\bar{x} = 135.625$.

Remarque 16.5.7. Pour vérifier que l'approximation n'est pas mauvaise, bien que le nombre 80 ne soit pas l'infini, on a simulé (avec le logiciel R) 10 variables suivant la loi uniforme entre 100 et 120, 15 entre 120 et 130, 25 entre 130 et 140, 20 entre 140 et 150 et 10 entre 150 et 170. Puis, l'on a calculé la moyenne arithmétique en suivant la Définition 16.5.1. On obtient : 135.5849 ; ce qui fait une erreur d'à peine 2.96%.

16.5.2 Quantiles

Définition 16.5.8. Soit $0 < p < 1$. Le quantile (ou fractile) d'ordre p est le plus petit nombre $q_p(x)$ tel que la proportion des éléments de la série x strictement inférieurs à $q_p(x)$ soit supérieure ou égale à p .

Notation 16.5.9. Pour $p = \frac{1}{2}$, la quantité $q_{\frac{1}{2}}(x)$ porte le nom de médiane.

Pour $p = \frac{1}{4}$, on parle de premier quartile. Pour $p = \frac{3}{4}$, on parle de troisième quartile.

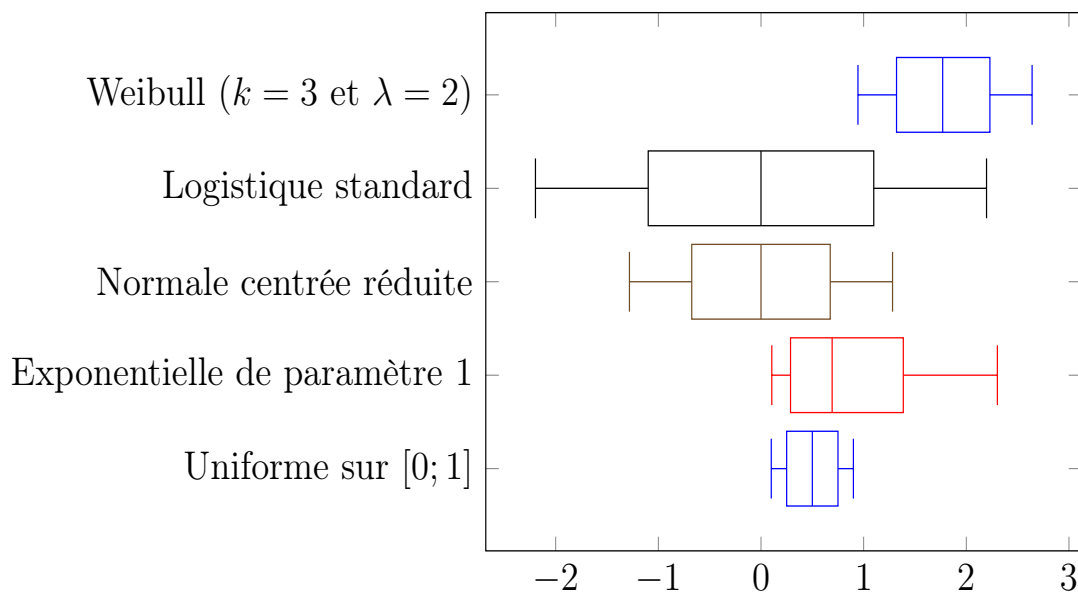
Pour $p = \frac{k}{10}$ (avec $k \in \llbracket 1; 9 \rrbracket$), on parle de *déciles*.

Pour $p = \frac{k}{100}$ (avec $k \in \llbracket 1; 99 \rrbracket$), on parle de *centiles*.

Une représentation graphique fort utile pour visualiser d'un coup d'œil les quantiles est le box-plot aussi appelée le diagramme à moustaches. Ce diagramme représente un rectangle dont les extrémités sont le premier quartile et le troisième quartile. Ce rectangle est coupé d'une barre qui représente la médiane. Enfin, la boîte a deux "moustaches" qui représente le premier décile et le neuvième décile.

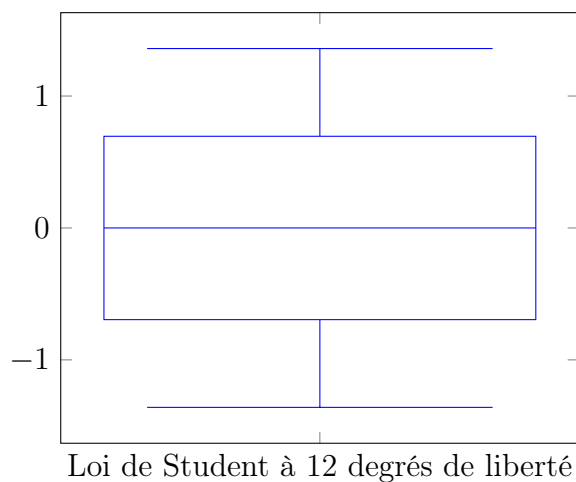
Voici un exemple pour diverses lois :

FIGURE 16.9 – Exemple de box-plot (horizontale)



On peut parfois la représenter verticalement :

FIGURE 16.10 – Exemple de box-plot (verticale)



Cette représentation peut servir à comparer des séries statistiques pour identifier si une hypothèse d'homogénéité sous-jacente est plausible ou non.

16.5.2.1 Calcul dans le cas d'une série à valeurs isolées

On reprend à nouveau le Tableau 16.3 :

y_i	n_i (effectifs)	f_i (fréquence)	N_i	F_i
39.25	2	$\frac{2}{15}$	2	$\frac{2}{15}$
39.50	4	$\frac{4}{15}$	6	$\frac{6}{15}$
39.75	2	$\frac{2}{15}$	8	$\frac{8}{15}$
40.00	4	$\frac{4}{15}$	12	$\frac{12}{15}$
40.25	2	$\frac{2}{15}$	14	$\frac{14}{15}$
40.50	1	$\frac{1}{15}$	15	$\frac{15}{15}$

On calcule la médiane (quantile d'ordre $p = \frac{1}{2}$). Ici, $n = 15$ donc $n_p = \frac{1}{2} \times n = 7.5 \notin \mathbb{N}$. On regarde la plus petite valeur telle que les effectifs cumulés sont supérieurs ou égaux à 7.5 donc à 8. Ainsi, on a $q_{\frac{1}{2}}(x) = 39.75$.

Exercice 16.5.10. Donner le premier quartile et le troisième quartile de la série statistique simple.

On peut aussi déterminer les quantiles en utilisant le diagramme en bâtons des fréquences cumulées et en traçant les droites d'équations ($y = \frac{1}{4}$) et ($y = \frac{3}{4}$).

16.5.2.2 Calcul dans le cas d'une série à valeurs classées

On reprend à nouveau le Tableau 16.5 :

$[z_{k-1}; z_k[$	n_k	f_k	N_k	F_k
[100; 120[10	$\frac{10}{80}$	10	$\frac{10}{80}$
[120; 130[15	$\frac{15}{80}$	25	$\frac{25}{80}$
[130; 140[25	$\frac{25}{80}$	50	$\frac{50}{80}$
[140; 150[20	$\frac{20}{80}$	70	$\frac{70}{80}$
[150; 170[10	$\frac{10}{80}$	80	$\frac{80}{80} = 1$

On regarde le premier quartile (quantile d'ordre $\frac{1}{4}$) : $q_{\frac{1}{4}}(x)$. Ici, $n = 80$ donc $n_p = p \times n = 20 \in \mathbb{N}$. L'intervalle de premier quartile est donc [120; 130[. En d'autres termes, $q_{\frac{1}{4}}(x) \in [120; 130[$. Pour calculer le premier quartile, on procède alors à une interpolation linéaire. On fait implicitement l'hypothèse que les éléments de la série dans la classe [120; 130[sont répartis uniformément.

On suppose ainsi que la courbe des effectifs cumulés a pour équation :

$$y = ax + b.$$

Il reste à déterminer a et b . On sait que l'on a

$$10 = 120a + b \quad \text{et} \quad 25 = 130a + b.$$

On en déduit immédiatement $a = \frac{15}{10}$ et $b = 10 - 120 \times a = -170$. Par conséquent, on obtient

$$20 = \frac{15}{10}q_{\frac{1}{4}}(x) - 170.$$

Ceci nous amène directement à $q_{\frac{1}{4}}(x) = \frac{20+170}{\frac{15}{10}} = \frac{380}{3} \approx 126.67$.

Exercice 16.5.11. *Calculer le septième décile de la série x définie dans le Tableau 16.5.*

16.5.3 Variance, Écart-type

Ces caractéristiques mesurent la dispersion des éléments de la série par rapport à une valeur centrale.

On suppose que l'on a observé un caractère quantitatif x sur une population de n individus : $x[1], \dots, x[i], \dots, x[n]$.

Définition 16.5.12. *La variance de la série statistique simple x est la moyenne des carrés des écarts des éléments de la série à leur moyenne arithmétique \bar{x} :*

$$s_x^2 := \frac{1}{n} \sum_{i=1}^n (x[i] - \bar{x})^2 \quad (16.1)$$

Notation 16.5.13. *La variance de la série x est notée s_x^2 . Le s provient de l'anglais "standard deviation".*

Définition 16.5.14. *On appelle écart-type de la série la racine carrée de la variance :*

$$s_x := \sqrt{\frac{1}{n} \sum_{i=1}^n (x[i] - \bar{x})^2}.$$

Exemple 16.5.15. *On considère la série $x := (9, 11)$. Sa moyenne est 10 et sa variance est*

$$s_x^2 = \frac{1}{2} [(9 - 10)^2 + (11 - 10)^2] = 1.$$

Exemple 16.5.16. *On considère la série $x := (5, 15)$. Sa moyenne est 10 et sa variance est*

$$s_x^2 = \frac{1}{2} [(5 - 10)^2 + (15 - 10)^2] = 25.$$

Proposition 16.5.17 (Formule de calcul). *Soit une série statistique simple $x := (x[1], \dots, x[n])$. On note x^2 la série statistique simple : $x^2 := (x[1]^2, \dots, x[n]^2)$. On a alors*

$$s_x^2 = \overline{x^2} - \bar{x}^2.$$

Cette formule est à rapprocher de $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

16.5.3.1 Calcul dans le cas d'une série à valeurs isolées

Proposition 16.5.18. *Si l'on a une série statistique simple à valeurs isolées, la variance est égale à*

$$s_x^2 = \frac{1}{n} \sum_{k=1}^r n_k (y_k - \bar{x})^2.$$

Ainsi, l'écart-type est :

$$s_x = \sqrt{\frac{1}{n} \sum_{k=1}^r n_k (y_k - \bar{x})^2}.$$

Définition 16.5.19. *Parfois, on introduit aussi la variance dite corrigée pour $n \geq 2$:*

$$\tilde{s}_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x[i] - \bar{x})^2 = \frac{n}{n-1} s_x^2.$$

L'intérêt de cette définition alternative prend tout son sens quand on fait de l'estimation ponctuelle (voir Chapitre 17) ou par intervalles de confiance (voir Chapitre 19).

On peut par exemple reprendre à nouveau le Tableau 16.3 :

y_i	n_i (effectifs)	f_i (fréquence)	N_i	F_i
39.25	2	$\frac{2}{15}$	2	$\frac{2}{15}$
39.50	4	$\frac{4}{15}$	6	$\frac{6}{15}$
39.75	2	$\frac{2}{15}$	8	$\frac{8}{15}$
40.00	4	$\frac{4}{15}$	12	$\frac{12}{15}$
40.25	2	$\frac{2}{15}$	14	$\frac{14}{15}$
40.50	1	$\frac{1}{15}$	15	$\frac{15}{15}$

Ici, $\bar{x} = 39.80$. Et, la variance est

$$\begin{aligned} s_x^2 &= \frac{1}{15} [2 \times (39.25 - 39.80)^2 + 4 \times (39.50 - 39.80)^2 + 2 \times (39.75 - 39.80)^2 \\ &\quad + 4 \times (40.00 - 39.80)^2 + 2 \times (40.25 - 39.80)^2 + 1 \times (40.50 - 39.80)^2] \\ &= 0.135. \end{aligned}$$

Et, l'écart-type est $s_x \approx 0.3674$.

Quant à la variance corrigée, elle vaut $\tilde{s}_x^2 = \frac{80}{79} \times 0.135 \approx 0.1367$ et l'écart-type associé est alors $\tilde{s}_x \approx 0.3697$.

16.5.3.2 Calcul dans le cas d'une série à valeurs classées

Dans le cas d'une série à valeurs classées, on pourrait être tenté de reprendre la même méthode que pour la moyenne.

On a en effet envie de dire que la variance est $\sum_{k=1}^s f_k \left(\frac{z_{k-1} + z_k}{2}\right)^2 - \bar{x}^2$. Néanmoins, ce serait négliger la dispersion au sein des classes.

La bonne approximation est donc la suivante.

Proposition 16.5.20. *Pour une série à valeurs classées, la formule*

$$s_x^2 = \sum_{k=1}^s f_k \left(\frac{z_{k-1} + z_k}{2}\right)^2 - \left(\sum_{k=1}^s f_k \frac{z_{k-1} + z_k}{2}\right)^2 + \sum_{k=1}^s f_k \frac{(z_k - z_{k-1})^2}{12},$$

est une bonne approximation de la réalité.

Démonstration. On utilise la même modélisation que pour la moyenne arithmétique.

On note $U_1, \dots, U_k, \dots, U_s$ s variables aléatoires indépendantes. On suppose que la loi de U_k est la loi uniforme sur l'intervalle $[z_{k-1}; z_k[$. Il est important de les supposer indépendantes car *a priori*, il n'y a aucune raison pour que les fluctuations d'un intervalle influent sur celles d'un autre.

On suppose également que l'on a une variable aléatoire \mathcal{K} , indépendante des $(U_k)_{k \in [1; s]}$ et telle que $\mathbb{P}(\mathcal{K} = k) = f_k$ pour tout $k \in [1; s]$. La moyenne arithmétique de $U_{\mathcal{K}}$ n'a pas changé et elle vaut $\sum_{k=1}^s f_k \frac{z_{k-1} + z_k}{2}$. Calculons son moment d'ordre deux :

$$\begin{aligned}
\mathbb{E}[U_{\mathcal{K}}^2] &= \sum_{k=1}^s \mathbb{E}[U_{\mathcal{K}}^2 \mathbf{1}_{\mathcal{K}=k}] \\
&= \sum_{k=1}^s \mathbb{E}[U_k^2 \mathbf{1}_{\mathcal{K}=k}] \\
&= \sum_{k=1}^s \mathbb{E}[U_k^2] \mathbb{P}(\mathcal{K} = k) \\
&= \sum_{k=1}^s f_k \mathbb{E}[U_k^2] \\
&= \sum_{k=1}^s f_k (\text{Var}[U_k] + \mathbb{E}[U_k]^2) \\
&= \sum_{k=1}^s f_k \left(\frac{(z_k - z_{k-1})^2}{12} + \left(\frac{z_k + z_{k-1}}{2} \right)^2 \right),
\end{aligned}$$

en utilisant l'indépendance entre \mathcal{K} et les variables aléatoires U_1, \dots, U_s et les propriétés sur les moments d'ordre 1 et 2 de la loi uniforme.

On en déduit le résultat annoncé. □

Ainsi, on a la variance de la série statistique simple à valeurs isolées prises aux milieux des classes à laquelle on ajoute la variance dite intra-classe.

Il convient de noter que dans le cas où les classes sont toutes de même longueur L , cette variance intra-classe est simplement

$$\sum_{k=1}^s f_k \frac{L^2}{12} = \frac{L^2}{12}.$$

On reprend à nouveau le Tableau 16.5 :

$[z_{k-1}; z_k[$	n_k	f_k	N_k	F_k
$[100; 120[$	10	$\frac{10}{80}$	10	$\frac{10}{80}$
$[120; 130[$	15	$\frac{15}{80}$	25	$\frac{25}{80}$
$[130; 140[$	25	$\frac{25}{80}$	50	$\frac{50}{80}$
$[140; 150[$	20	$\frac{20}{80}$	70	$\frac{70}{80}$
$[150; 170[$	10	$\frac{10}{80}$	80	$\frac{80}{80} = 1$

On a déjà vu que l'on avait $\bar{x} = 135.625$. Alors, la variance est

$$\begin{aligned} s_x^2 &= \frac{1}{80} [10 \times 110^2 + 15 \times 125^2 + 25 \times 135^2 + 20 \times 145^2 + 10 \times 160^2] - 135.625^2 \\ &+ \frac{1}{80} \left[10 \times \frac{(120 - 100)^2}{12} + \dots + 10 \times \frac{(170 - 150)^2}{12} \right] \\ &= 214.1927. \end{aligned}$$

Ainsi, l'écart-type vaut $s_x \approx 14.6353$. Ici, la variance intra-classe vaut 14.58. Elle représente ainsi plus de 6.8% de la variance totale.

Quant à la variance corrigée, elle vaut $\tilde{s}_x^2 \approx 216.9040$ et l'écart-type associé est $\tilde{s}_x \approx 14.7277$.

16.5.4 Autres caractéristiques

On utilise parfois d'autres caractéristiques. Présentons-en quelques unes.

16.5.4.1 Coefficient de dispersion

Le coefficient de dispersion mesure la dispersion de la série en tenant compte de sa moyenne arithmétique. Il s'agit de

$$\frac{s_x}{\bar{x}}.$$

Ce coefficient ne dépend pas de l'unité de mesure choisie.

L'intérêt est qu'en renormalisant par la moyenne, l'ordre de grandeur n'influera pas. En effet, un écart-type de 1 pour une série ayant une moyenne de 10 correspond à un écart de 10% tandis que pour une moyenne de 1000, ça ne correspond qu'à 0.1%.

Néanmoins, quand la moyenne est proche de zéro, ce coefficient tend à être très grand et donc d'autant plus sensible aux variations de la moyenne. De plus, il ne sert pas pour la construction des intervalles de confiance.

16.5.4.2 L'étendue

Définition 16.5.21. On appelle *étendue* ("range" en anglais) de la série statistique simple x la différence entre le plus grand élément de la série et le plus petit.

Exemple 16.5.22 (Série à valeurs isolées). Dans le cas de la série définie dans le Tableau 16.3, l'étendue est $40.50 - 39.25 = 1.25$.

Exemple 16.5.23 (Série à valeurs classées). Dans le cas de la série définie dans le Tableau 16.5, l'étendue est $170 - 110 = 60$.

Remarque 16.5.24. L'étendue est sensible aux valeurs aberrantes.

16.5.4.3 L'interquartile

Définition 16.5.25 (Intervalle interquartile). *On appelle intervalle interquartile de la série statistique simple $x := (x[1], \dots, x[i], \dots, x[n])$ l'intervalle semi-ouvert $[q_{\frac{1}{4}}(x); q_{\frac{3}{4}}(x)[$ où l'on rappelle que $q_{\frac{1}{4}}(x)$ (respectivement $q_{\frac{3}{4}}(x)$) est le premier quartile (respectivement le troisième quartile) de la série.*

Définition 16.5.26 (Interquartile). *On appelle interquartile de la série x la longueur de l'intervalle interquartile.*

L'intervalle interquartile contient 50% des éléments de la série.

16.6 Séries statistiques doubles

16.6.1 Définitions

On mesure deux caractères quantitatifs x et y sur un ensemble de n individus.

Soit $x[i]$ la mesure de x pour l'individu i et $y[i]$ celle de y .

Définition 16.6.1. *La suite $((x[1], y[1]), \dots, (x[i], y[i]), \dots, (x[n], y[n]))$ est appelée une série statistique double.*

Définition 16.6.2 (Marginales). *Les deux séries statistiques simples x et y avec $x := (x[1], \dots, x[n])$ et $y := (y[1], \dots, y[n])$ sont appelées les séries statistiques marginales de la série statistique double. On dit que x est la première marginale et que y est la deuxième.*

Exemple 16.6.3. *On considère l'ensemble des habitants de la France. Alors, si $x[i]$ est la taille de l'individu i et si $y[i]$ est le poids de ce même individu, la série x est la première marginale de (x, y) tandis que y en est la deuxième marginale.*

16.6.2 Tableau de contingence

On présente la série statistique double dans un tableau de contingence.

Pour ce faire, on commence par grouper les données.

On donne chaque série marginale sous la forme d'une série à valeurs isolées ou une série à valeurs classées. On suppose que $a_1, \dots, a_k, \dots, a_r$ sont les r modalités de la série x . Et, $b_1, \dots, b_l, \dots, b_s$ sont les s modalités de la série y . Ici, les modalités a_k et b_l représentent une valeur isolée ou une classe. On dit qu'un individu a la modalité a_k pour le caractère x lorsque :

- la valeur de x pour cet individu est égale à la valeur isolée représentée par a_k ,
- la valeur de x appartient à la classe représentée par a_k .

On fait de même avec y et b_l pour tout $l \in \llbracket 1; s \rrbracket$.

À l'intersection de la ligne a_k et de la colonne b_l , on porte le nombre total $n_{k,l}$ d'individus qui ont la modalité a_k pour x et la modalité b_l pour y . On l'appelle effectif de (a_k, b_l) .

TABLE 16.7 – Tableau de contingence - 1

$x \backslash y$	b_1	\dots	b_l	\dots	b_s
a_1	$n_{1,1}$	\dots	$n_{1,l}$	\dots	$n_{1,s}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
a_k	$n_{k,1}$	\dots	$n_{k,l}$	\dots	$n_{k,s}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
a_r	$n_{r,1}$	\dots	$n_{r,l}$	\dots	$n_{r,s}$

La somme des effectifs sur une ligne est importante :

$$n_{k,\bullet} := \sum_{l=1}^s n_{k,l}.$$

Alors, $n_{k,\bullet}$ représente le nombre d'individus qui ont la modalité a_k de x . On l'appelle effectif marginal de la modalité a_k .

De même, on somme sur une colonne :

$$n_{\bullet,l} := \sum_{k=1}^r n_{k,l}.$$

Alors, $n_{\bullet,l}$ représente le nombre d'individus qui ont la modalité b_l de y . On l'appelle effectif marginal de la modalité b_l . On a les égalités suivantes :

$$\sum_{k=1}^r n_{k,\bullet} = \sum_{l=1}^s n_{\bullet,l} = \sum_{k=1}^r \sum_{l=1}^s n_{k,l} = \sum_{l=1}^s \sum_{k=1}^r n_{k,l} = n.$$

On peut définir la fréquence du couple (a_k, b_l) comme on le fit pour les séries statistiques simples :

$$f_{k,l} := \frac{n_{k,l}}{n}.$$

C'est la proportion d'individus de la population qui ont la modalité a_k de x et la modalité b_l de y . On introduit également

$$f_{k,\bullet} := \frac{n_{k,\bullet}}{n} \quad \text{et} \quad f_{\bullet,l} := \frac{n_{\bullet,l}}{n},$$

pour tout $k \in \llbracket 1; r \rrbracket$ et pour tout $l \in \llbracket 1; s \rrbracket$. $f_{k,\bullet}$ est la fréquence marginale de la modalité a_k : c'est la proportion d'éléments de la population qui ont la modalité a_k de x . De même, $f_{\bullet,l}$ est la fréquence marginale de la modalité b_l : c'est la proportion d'éléments de la population qui ont la modalité b_l de y . Or, $f_{k,\bullet} = \sum_{l=1}^s n_{k,l}$ donc

$$f_{k,\bullet} := \sum_{l=1}^s f_{k,l} \quad \text{et de même} \quad f_{\bullet,l} := \sum_{k=1}^r f_{k,l}.$$

Puis, l'on a

$$\sum_{k=1}^r f_{k,\bullet} = \sum_{l=1}^s f_{\bullet,l} = \sum_{k=1}^r \sum_{l=1}^s f_{k,l} = \sum_{l=1}^s \sum_{k=1}^r f_{k,l} = 1.$$

On obtient ainsi le tableau de contingence suivant :

TABLE 16.8 – Tableau de contingence - 2

$x \backslash y$	b_1	\dots	b_l	\dots	b_s	
a_1	$n_{1,1}$	\dots	$n_{1,l}$	\dots	$n_{1,s}$	$n_{1,\bullet}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
a_k	$n_{k,1}$	\dots	$n_{k,l}$	\dots	$n_{k,s}$	$n_{k,\bullet}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
a_r	$n_{r,1}$	\dots	$n_{r,l}$	\dots	$n_{r,s}$	$n_{r,\bullet}$
	$n_{\bullet,1}$	\dots	$n_{\bullet,l}$	\dots	$n_{\bullet,s}$	n

On va maintenant définir la fréquence conditionnelle.

Définition 16.6.4. On pose : $f(b_l | a_k) := \frac{f_{k,l}}{f_{k,\bullet}} = \frac{n_{k,l}}{n_{k,\bullet}}$. Cette quantité est appelée fréquence conditionnelle de la modalité b_l par rapport à la modalité a_k .

C'est aussi la proportion d'individus qui ont la modalité b_l parmi les $n_{k,\bullet}$ individus qui ont la modalité a_k . C'est à rapprocher de $\mathbb{P}(Y = b_l | X = a_k)$.

Remarque 16.6.5. On a ainsi une suite de s valeurs positives telles que leur somme vaut

$$\sum_{l=1}^s \frac{f_{k,l}}{f_{k,\bullet}} = \frac{\sum_{l=1}^s f_{k,l}}{f_{k,\bullet}} = \frac{f_{k,\bullet}}{f_{k,\bullet}} = 1.$$

Il y a donc bien une loi de probabilité sous-jacente.

Définition 16.6.6. L'ensemble des couples $(b_l, f(b_l | a_k))$ s'appelle la distribution conditionnelle du caractère y par rapport à la modalité a_k de x .

Remarque 16.6.7. Pour simplifier l'écriture, on pourra faire l'abus de notation suivant :

$$f(b_1 | a_k) \delta_{b_1} + \cdots + f(b_l | a_k) \delta_{b_l} + \cdots + f(b_s | a_k) \delta_{b_s}.$$

Dans le cas où les modalités $b_1, \dots, b_l, \dots, b_s$ sont des valeurs isolées, la distribution précédente a du sens.

En revanche, si ce sont des classes ou même si le caractère y est qualitatif, la distribution de Dirac en b_l n'a aucun sens.

Il faut donc bien comprendre qu'on commet une erreur.

De la même manière, on définit ensuite $f(a_k | b_l)$ comme étant le rapport $\frac{f_{k,l}}{f_{\bullet,l}} = \frac{n_{k,l}}{n_{\bullet,l}}$, appelée fréquence conditionnelle de la modalité a_k par rapport à la modalité b_l : Il s'agit de la proportion d'individus qui ont la modalité a_k parmi les $n_{\bullet,l}$ individus qui ont la modalité b_l .

Définition 16.6.8. L'ensemble des couples $(a_k, f(a_k | b_l))$ s'appelle la distribution conditionnelle du caractère x par rapport à la modalité b_l de y .

16.6.3 Indépendance de deux caractères

Définition 16.6.9. On dit que le caractère y est indépendant du caractère x dans l'ensemble I lorsque la distribution conditionnelle du caractère y par rapport à la modalité a_k de x est la même pour tout $k \in \llbracket 1; r \rrbracket$.

En d'autres termes, pour tout $k \in \llbracket 1; r \rrbracket$, cette distribution conditionnelle est égale à celle de y par rapport à la modalité a_1 de x :

$$\sum_{l=1}^s f(b_l | a_k) \delta_{b_l} = \sum_{l=1}^s f(b_l | a_1) \delta_{b_l}.$$

Par conséquent, l'indépendance du caractère y par rapport au caractère x peut se traduire par l'égalité $f(b_l | a_k) = f(b_l | a_1)$ pour tout $k \in \llbracket 1; r \rrbracket$ et pour tout $l \in \llbracket 1; s \rrbracket$.

Exemple 16.6.10. On se donne le tableau de contingence suivant :

TABLE 16.9 – Exemple de tableau de contingence

$x \backslash y$	b_1	b_2	b_3	b_4	
a_1	2	4	5	1	12
a_2	4	8	10	2	24
	6	12	15	3	36

Déterminons les distributions conditionnelles de y par rapport à la modalité a_1 puis par rapport à la modalité a_2 .

La distribution conditionnelle de y par rapport à la modalité a_1 de x est

$$\frac{2}{12}\delta_{b_1} + \frac{4}{12}\delta_{b_2} + \frac{5}{12}\delta_{b_3} + \frac{1}{12}\delta_{b_4} = \frac{1}{6}\delta_{b_1} + \frac{1}{3}\delta_{b_2} + \frac{5}{12}\delta_{b_3} + \frac{1}{12}\delta_{b_4}.$$

Celle de y par rapport à la modalité a_2 de x est

$$\frac{4}{24}\delta_{b_1} + \frac{8}{24}\delta_{b_2} + \frac{10}{24}\delta_{b_3} + \frac{2}{24}\delta_{b_4} = \frac{1}{6}\delta_{b_1} + \frac{1}{3}\delta_{b_2} + \frac{5}{12}\delta_{b_3} + \frac{1}{12}\delta_{b_4}.$$

L'égalité des deux distributions conditionnelles s'ensuit. On en déduit que y est indépendante de x dans cette population.

Contre-exemple 16.6.11. Le poids d'un individu dépend statistiquement de sa taille dans la population observée sur le tableau de contingence suivant :

TABLE 16.10 – Tableau de contingence (poids/taille)

poids \ taille	[150; 160[[160; 170[[170; 180[[180; 190[[190; 200[
	[45; 55[2	1	0	0	
[55; 65[7	8	4	2	0	21
[65; 75[5	15	22	7	1	50
[75; 85[2	12	63	19	5	101
[85; 95[0	7	28	32	12	79
[95; 105[0	2	10	20	7	39
[105; 115[0	0	1	4	2	7
	16	45	128	84	27	300

Il suffit pour s'en convaincre de vérifier que l'on a

$$f([190; 200[| [45; 55]) \neq f([190; 200[| [105; 115]) .$$

Or, $f([190; 200[| [45; 55]) = \frac{0}{3} = 0$ tandis que $f([190; 200[| [105; 115]) = \frac{2}{7} \neq 0$.

Théorème 16.6.12. *La relation d'indépendance est symétrique. Si y est indépendant de x alors x est indépendant de y . On dit d'ailleurs que x et y sont indépendants.*

La preuve du Théorème 16.6.12 est immédiate à partir du résultat du théorème qui suit.

Théorème 16.6.13. *Les caractères x et y sont indépendants si et seulement si $f_{k,l} = f_{k,\bullet} f_{\bullet,l}$ pour tout $k \in \llbracket 1; r \rrbracket$ et pour tout $l \in \llbracket 1; s \rrbracket$.*

Démonstration. On a déjà vu que l'indépendance de y par rapport à x signifie que l'on a

$$f(b_l | a_k) = f(b_l | a_1) ,$$

pour tout $k \in \llbracket 1; r \rrbracket$ et pour tout $l \in \llbracket 1; s \rrbracket$. On en déduit l'égalité

$$\frac{f_{k,l}}{f_{k,\bullet}} = \frac{f_{1,l}}{f_{1,\bullet}} .$$

Puis, l'on a :

$$\frac{f_{k,l}}{f_{k,\bullet}} = \frac{f_{k',l}}{f_{k',\bullet}} ,$$

pour tout $1 \leq k, k' \leq r$ et tout $1 \leq l \leq s$. Ainsi :

$$f_{k,l} f_{k',\bullet} = f_{k,\bullet} f_{k',l} .$$

On fait la somme pour k' allant de 1 à r et l'on trouve :

$$f_{k,l} \sum_{k'=1}^r f_{k',\bullet} = f_{k,\bullet} \sum_{k'=1}^r f_{k',l} .$$

Or, $\sum_{k'=1}^r f_{k',\bullet} = 1$ et $\sum_{k'=1}^r f_{k',l} = f_{\bullet,l}$; ce qui achève la preuve. □

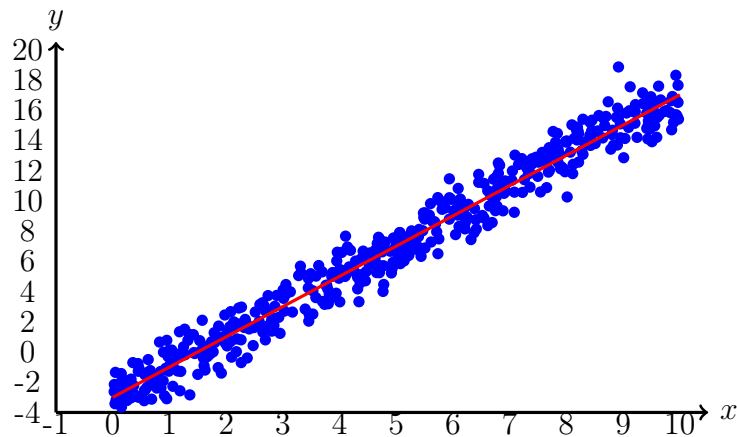
Remarque 16.6.14. *L'indépendance des caractères x et y peut aussi s'obtenir à partir de l'égalité $n_{k,l} = \frac{n_{k,\bullet} n_{\bullet,l}}{n}$.*

Il est important de comprendre que cette notion d'indépendance est extrêmement restrictive. En effet, si l'on a un échantillon de taille n finie, même si les deux variables aléatoires sous-jacentes sont bien indépendantes, il est quasiment impossible que les égalités soient vérifiées. De fait, dans la pratique, il faut effectuer un test d'indépendance pour vérifier si l'on peut conclure ou non sur l'indépendance des variables aléatoires sous-jacentes. Voir la Section 21.9 à la page 323.

16.6.4 Nuage de points

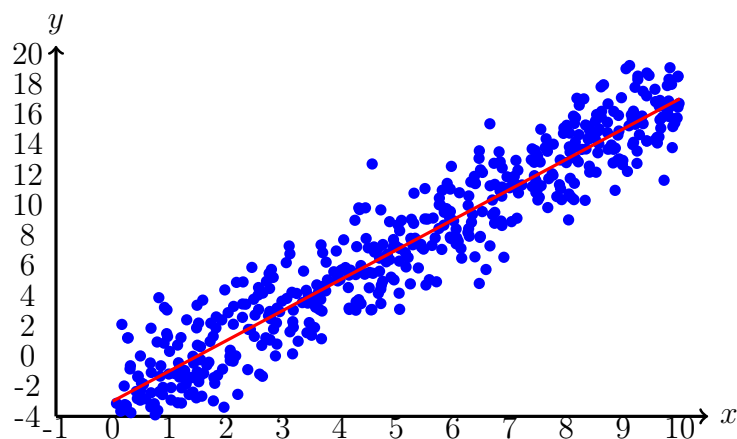
Soit une série statistique double $((x[1], y[1]), \dots, (x[n], y[n]))$ sur une population de n individus. On porte dans un système d'axes orthogonaux les points dont les coordonnées sont $(x[i], y[i])$. L'un des intérêts de ce nuage de points est de vérifier, à l'œil nu s'il est pertinent ou non d'effectuer une régression linéaire (voir Chapitre 22). Voici ainsi un premier nuage de points :

FIGURE 16.11 – Nuage de points - 1



Il est assez clair qu'une direction se précise et que ce nuage peut être approché par une droite. Ainsi, la régression linéaire sera pertinente. En voici un deuxième :

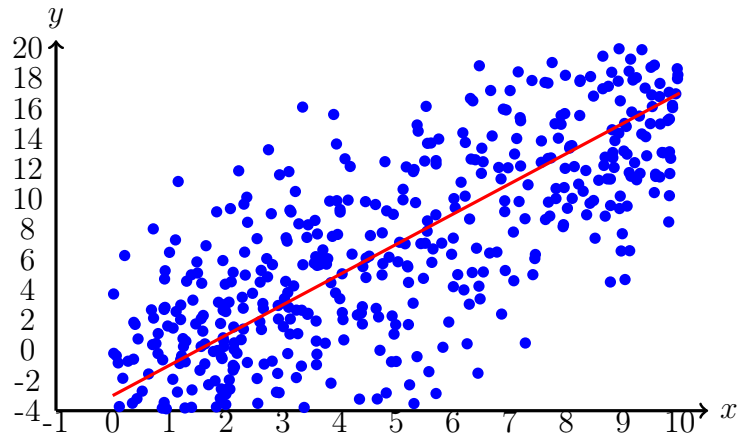
FIGURE 16.12 – Nuage de points - 2



Dans ce nuage, la direction est moins nette mais l'on peut distinguer une certaine droite qui, par ailleurs, aurait la même équation que la précédente. Voici

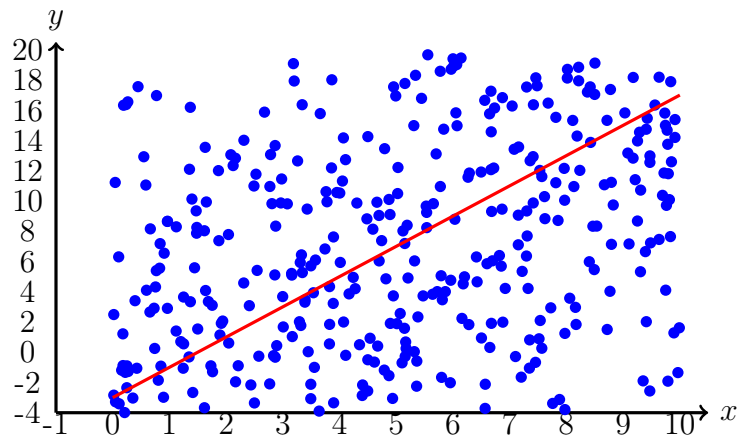
un troisième nuage de points :

FIGURE 16.13 – Nuage de points - 3



L'existence ou non d'une direction privilégiée est encore moins perceptible. Voici un quatrième nuage de points :

FIGURE 16.14 – Nuage de points - 4



Cette fois, il est difficilement envisageable qu'une régression linéaire apporte quoi que ce soit de pertinent.

Remarque 16.6.15. *Pour réaliser ces nuages, on a simulé 500 variables aléatoires suivant la loi uniforme sur $[0; 10]$. Ceci nous donne la série x . Puis, pour y , on a posé :*

$$y[i] := ax[i] + b + \rho\epsilon[i],$$

où $a = 2$, $b = -3$, $(\epsilon[i])_i$ est une suite de variables aléatoires indépendantes et identiquement distribuées suivant la loi normale centrée réduite. Quant à ρ , la valeur du bruit, elle a été prise égale à 1 dans le premier nuage, à 2 dans le deuxième, à 4 dans le troisième et à 8 dans le quatrième. On a par ailleurs superposé la droite d'équation $y = 2x - 3$ au nuage.

16.6.5 Covariance

La covariance est une quantité essentielle pour mesurer la corrélation linéaire entre deux caractères quantitatifs mesurés sur une même population. Elle est essentielle dans le Chapitre 22.

Définition 16.6.16. La covariance entre x et y est

$$s_{xy} := \frac{1}{n} \sum_{i=1}^n (x[i] - \bar{x})(y[i] - \bar{y}) = \frac{1}{n} \sum_{i=1}^n x[i]y[i] - \bar{x}\bar{y}.$$

En ce qui concerne le calcul pratique, si x et y sont deux séries à valeurs isolées, le calcul donne simplement $s_{xy} = \sum_{k=1}^r \sum_{l=1}^s f_{k,l} a_k b_l - \bar{x}\bar{y}$.

Si x est à valeurs classées alors que y est à valeurs isolées, il faut reprendre la modélisation déjà utilisée précédemment. Ainsi, x est une mesure de la variable aléatoire $X := U_{\mathcal{K}}$ et a_k doit être pris égal à $\frac{z_{k-1} + z_k}{2}$. On ne suppose évidemment pas l'indépendance entre X et Y . Néanmoins, les erreurs de fluctuation propres à chaque intervalle sont supposées indépendantes de la variable aléatoire Y . De fait :

$$\begin{aligned} \mathbb{E}[XY] &= \mathbb{E}[U_{\mathcal{K}}Y] \\ &= \sum_{k=1}^r \mathbb{E}[U_{\mathcal{K}}Y \mathbf{1}_{\mathcal{K}=k}] \\ &= \sum_{k=1}^r \mathbb{E}[U_k Y \mathbf{1}_{\mathcal{K}=k}] \\ &= \sum_{k=1}^r \mathbb{E}[U_k] \mathbb{E}[Y \mathbf{1}_{\mathcal{K}=k}] \\ &= \sum_{k=1}^r \frac{z_{k-1} + z_k}{2} \mathbb{P}(\mathcal{K} = k) \sum_{l=1}^s b_l f(b_l | [z_{k-1}; z_k]) \\ &= \sum_{k=1}^r \frac{z_{k-1} + z_k}{2} f_{k,\bullet} \sum_{l=1}^s b_l \frac{f_{k,l}}{f_{k,\bullet}} \\ &= \sum_{k=1}^r \sum_{l=1}^s \frac{z_{k-1} + z_k}{2} f_{k,l} b_l. \end{aligned}$$

Par conséquent, la covariance sera approchée par

$$s_{xy} = \sum_{k=1}^r \sum_{l=1}^s f_{k,l} \frac{z_{k-1} + z_k}{2} b_l - \overline{xy}.$$

Enfin, dans le cas où x et y sont à valeurs classées, supposer l'indépendance des perturbations de x avec celles de y conduit au résultat suivant :

$$s_{xy} = \sum_{k=1}^r \sum_{l=1}^s f_{k,l} \frac{z_{k-1} + z_k}{2} \frac{\widetilde{z}_{l-1} + \widetilde{z}_l}{2} - \overline{xy}.$$

16.6.6 Description empirique de la dépendance

On s'intéresse ici à la courbe des moyennes conditionnelles. On suppose que le caractère x a une action sur le caractère y .

Pour les individus ayant la modalité a_k du caractère x , on calcule la moyenne des valeurs du caractère y , notée \overline{y}_k , appelée moyenne conditionnelle de y pour x ayant la modalité a_k .

Exemple 16.6.17. *On reprend le Tableau 16.10.*

On calcule la moyenne conditionnelle du poids pour les individus de taille dans [150; 160[:

taille \ poids	[45; 55[[55; 65[[65; 75[[75; 85[
[150; 160[2	7	5	2

On a alors $\overline{y}_1 = \frac{2 \times 50 + 7 \times 60 + 5 \times 70 + 2 \times 80}{16} = 64.375$. Pour obtenir la courbe des moyennes conditionnelles, on joint par des segments de droite les points de coordonnées (ξ_k, \overline{y}_k) où ξ_k est la valeur isolée en x ou le milieu de la classe de x . Dans le cas présent, on joint les points $(155, 64.375)$, $(165, 74.89)$, $(175, 81.64)$, $(185, 88.69)$ et $(195, 91.48)$.

Estimation ponctuelle

17.1 Introduction

Dans ce chapitre, il s'agit d'estimer certaines caractéristiques statistiques de la loi (moyenne, variance, fonction de répartition) au travers d'une série d'observations.

Les pré-requis sont les chapitres de probabilités notamment sur les convergences des variables aléatoires et les lois des grands nombres. Il est aussi préférable d'avoir des connaissances sur les distributions et sur le produit de convolution des distributions afin de pouvoir saisir l'exemple qui sert de fil rouge : celui sur l'usine d'embouteillage.

Les objectifs principaux du chapitre sont le vocabulaire d'un côté (échantillons, statistiques...) et la compréhension du concept d'estimateur de l'autre. Également, il est attendu du lecteur qu'à l'issue de ce chapitre, il connaisse quelques exemples d'estimateurs très classiques comme la moyenne d'échantillon, la proportion d'échantillon, la variance d'échantillon, la variance d'échantillon corrigée ainsi que leurs caractéristiques statistiques (espérance et variance). Par ailleurs, nous insistons pour que l'étude du chapitre (et du précédent) permette de faire instantanément le lien entre les notions du chapitre précédent et de celui-ci. Enfin, il est crucial de bien saisir qu'il n'y a pas qu'un critère pour évaluer les estimateurs puisque ceux-ci peuvent avoir différentes qualités comme l'absence de biais, la convergence, la normalité asymptotique, l'efficacité ou la robustesse.

17.2 Préliminaires

Le problème de l'estimation peut s'énoncer de la façon suivante : disposant d'observations $x[1], x[2], \dots, x[n]$ d'une variable aléatoire X , obtenues à partir d'un échantillonnage aléatoire (n -échantillon de la variable aléatoire X), quelle loi théorique inconnue peut-on retenir comme loi de X ? (loi parente)

Si ce choix devait être fait parmi l'ensemble de toutes les lois de probabilité existantes, on conçoit que le problème serait difficilement résolu sans un échantillon de taille très grande. Mais, dans ce chapitre, nous allons voir que l'on peut le résoudre dès lors que l'on restreint le choix de la loi parente à une famille de lois de

probabilité parfaitement déterminées par la donnée d'un ou plusieurs paramètres.

Le problème devient alors de choisir une valeur de chaque paramètre (estimation ponctuelle).

L'estimation consiste à donner des valeurs approximatives aux paramètres d'une population à l'aide d'un échantillon de n observations issues de cette population. On peut se tromper sur la valeur exacte, mais on donne la "meilleure valeur" possible que l'on peut supposer.

Exemple 17.2.1. *Avant de choisir un véhicule automobile, on se fixe un critère de choix basé sur le nombre X de pannes par an que l'on est susceptible d'avoir avec un modèle donné. Ayant la possibilité de faire une étude statistique chez un concessionnaire donné, on prélève au hasard n dossiers de véhicules et l'on note $x[1], x[2], \dots, x[n]$ le nombre de pannes subies la première année de mise en circulation de ces véhicules.*

La loi de Poisson est bien adaptée à la modélisation du nombre de pannes. Conséquemment, le choix de la loi parente est fait parmi la famille de lois $\{\mathcal{P}(\lambda); \lambda > 0\}$.

L'unique paramètre déterminant la loi est ici λ . Or, on sait que λ est l'espérance de la loi :

$$\lambda = \mathbb{E}[X].$$

On estime donc la valeur de ce paramètre par la moyenne des valeurs observées sur l'échantillon :

$$\bar{x} := \frac{1}{n} \sum_{i=1}^n x[i].$$

17.3 Vocabulaire

On commence par donner les définitions basiques en statistiques inférentielles. Dans tout le reste du cours, Ω désigne une population de taille N . On suppose que N est grand si bien qu'un caractère C peut être vu comme une variable aléatoire sur l'univers Ω .

Définition 17.3.1 (Échantillon). *On appelle échantillon un sous-ensemble de la population Ω .*

Définition 17.3.2 (Taille de l'échantillon). *Un échantillon de taille n est une liste de n individus $\{\omega_1, \dots, \omega_n\}$ extraits de la population Ω .*

On dit que Ω est la population mère de l'échantillon.

Exemple 17.3.3. *On considère une population constituée de cinquante étudiants. On a ainsi $\Omega = \{\omega_1, \dots, \omega_{50}\}$ et $N = 50$. On prend un échantillon de six étudiants, $\{\omega_4, \omega_8, \omega_{15}, \omega_{16}, \omega_{23}, \omega_{42}\}$ et $n = 6$.*

Définition 17.3.4 (Taux d'échantillonnage). *Le rapport de l'effectif n de l'échantillon sur l'effectif N de la population mère dans laquelle il a été prélevé est appelé taux d'échantillonnage : $t = \frac{n}{N}$.*

On parle aussi de fraction de sondage.

Exemple 17.3.5. *Dans l'Exemple 17.3.3, le taux d'échantillonnage est égal à $\frac{6}{50} = 0.12 = 12\%$.*

On cherche à décrire un caractère qualitatif ou quantitatif (qu'il soit discret ou continu) C dans une population mère Ω à travers l'étude des résultats obtenus sur un échantillon de taille n .

Remarque 17.3.6. *Le taux d'échantillonnage doit répondre à deux critères : il faut qu'il soit suffisamment élevé pour rendre compte de la population mère et il faut qu'il soit suffisamment petit pour être simple à étudier.*

Exemple 17.3.7. *Étant donné une population d'étudiants, on peut s'intéresser à un caractère quantitatif discret comme la note à un partiel (0, 5, 15, 20...).*

Étant donné une population d'étudiants, on peut s'intéresser à un caractère quantitatif continu comme la taille.

Étant donné une population d'étudiants, on peut s'intéresser à un caractère qualitatif comme la matière préférée ("probabilités", "statistiques", "signaux et systèmes discrets", "traitement des signaux déterministes", "signaux aléatoires", "la sieste"...).

Définition 17.3.8 (n -échantillon de valeurs de X). *Soit C un caractère quantitatif défini sur une population mère Ω . C est la réalisation d'une variable aléatoire X définie sur Ω par $X(\omega_i) = x[i]$. On appelle n -échantillon de valeurs de X la liste des valeurs (x_1, \dots, x_n) observées prises par X sur un échantillon $\{\omega_1, \dots, \omega_n\}$ de la population Ω .*

Remarque 17.3.9. *Les coordonnées peuvent être considérées comme les valeurs des réalisations d'un vecteur aléatoire (X_1, \dots, X_n) appelé n -échantillon de X où les variables aléatoires réelles X_i sont indépendantes et identiquement distribuées (de même loi).*

17.4 Notion de statistique

Définition 17.4.1 (Statistique). *On appelle statistique toute variable aléatoire qui s'écrit à l'aide des variables aléatoires X_1, \dots, X_n .*

Remarque 17.4.2. *Une statistique est donc une variable aléatoire qui est mesurable par rapport à la tribu engendrée par les n variables aléatoires X_1, \dots, X_n .*

Remarque 17.4.3. *En d'autres termes, une statistique est une variable aléatoire de la forme $S := \varphi(X_1, \dots, X_n)$ où φ est une application mesurable de \mathbb{R}^n dans \mathbb{R} .*

Cette notion est intuitive. En effet, on part du principe que l'on ne dispose d'aucune autre information que celles obtenues par les observations de l'échantillon.

Exemple 17.4.4. *La variable aléatoire*

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$$

est une statistique.

Exemple 17.4.5. *La variable aléatoire*

$$S_n^2 := \frac{1}{n} \sum_{i=1}^n \left(X_i - \frac{1}{n} \sum_{j=1}^n X_j \right)^2$$

est aussi une statistique.

Contre-exemple 17.4.6. *La variable aléatoire X_{n+1} n'est pas une statistique.*

Si on extrait plusieurs échantillons de taille n fixée, les résultats que l'on va pouvoir déduire sont variables car ils dépendent de l'échantillon considéré. On parle de "fluctuations d'échantillonnage".

17.5 Moyenne d'échantillon

On considère une population Ω dont les éléments possèdent un caractère quantitatif (discret ou continu) C qui est la réalisation d'une variable aléatoire X suivant une loi de probabilité d'espérance égale à μ et de variance égale à σ^2 (donc d'écart-type égal à σ).

Définition 17.5.1 (Moyenne d'échantillon). *On définit la variable aléatoire \overline{X}_n , appelée moyenne d'échantillon par :*

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^n X_i.$$

Comme \overline{X}_n est une variable aléatoire réelle, il peut être judicieux d'étudier ses caractéristiques (espérance et variance).

Théorème 17.5.2 (Distribution d'échantillonnage de la moyenne). *Quelle que soit la loi de la variable aléatoire X d'espérance μ et de variance σ^2 avec $\sigma > 0$ et quel que soit n , on a*

$$\mathbb{E}[\overline{X}_n] = \mu = \mathbb{E}[X],$$

ainsi que

$$\text{Var}[\overline{X}_n] = \frac{\sigma^2}{n} = \frac{\text{Var}[X]}{n}.$$

Démonstration. La linéarité de l'espérance nous donne

$$\begin{aligned}\mathbb{E}[\overline{X}_n] &= \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^n X_i\right\} \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X] \\ &= \frac{1}{n}n\mathbb{E}[X] \\ &= \mathbb{E}[X],\end{aligned}$$

ce qui achève la première partie de la preuve.

On calcule désormais la variance de la moyenne d'échantillon. Par définition de la variance, il vient

$$\begin{aligned}\text{Var}[\overline{X}_n] &= \text{Var}\left\{\frac{1}{n}\sum_{i=1}^n X_i\right\} \\ &= \frac{1}{n^2}\text{Var}\left[\sum_{i=1}^n X_i\right] \\ &= \frac{1}{n^2}\sum_{i=1}^n \text{Var}[X_i] \\ &= \frac{1}{n^2}\sum_{i=1}^n \text{Var}[X] \\ &= \frac{1}{n^2}n\text{Var}[X] \\ &= \frac{\text{Var}[X]}{n},\end{aligned}$$

ce qui achève la preuve. □

La convergence vers 0 de la variance signifie que la variable aléatoire \overline{X}_n est de plus en plus concentrée autour de sa moyenne, c'est-à-dire autour de μ , quand n augmente. Il vient

Proposition 17.5.3. *La variable aléatoire \overline{X}_n converge en probabilité vers $\mu = \mathbb{E}[X]$.*

Démonstration. Par définition, la convergence en probabilité de \overline{X}_n vers μ signifie que pour tout $\epsilon > 0$, on a

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\overline{X}_n - \mu| > \epsilon) = 0.$$

On utilise l'inégalité de Bienaymé-Tchebychev :

$$\mathbb{P}(|\overline{X}_n - \mu| > \epsilon) \leq \frac{\text{Var}[\overline{X}_n]}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2} \frac{1}{n} \longrightarrow 0.$$

La preuve est ainsi achevée. \square

Le prochain résultat s'intéresse au cas particulier où X suit une loi normale.

Théorème 17.5.4. *Si la variable aléatoire X suit la loi normale $\mathcal{N}(\mu, \sigma^2)$ avec $\sigma > 0$, alors la moyenne empirique \overline{X}_n suit la loi normale $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. Puis, la variable aléatoire $\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$ suit la loi normale centrée réduite.*

Démonstration. Comme les X_i ont même loi que X , les variables aléatoires X_i suivent la loi normale $\mathcal{N}(\mu, \sigma^2)$. Conséquemment, la variable aléatoire \overline{X}_n suit une loi normale. Or, d'après le Théorème 17.5.2, son espérance vaut μ et sa variance vaut $\frac{\sigma^2}{n}$ ce qui achève de prouver que \overline{X}_n suit la loi normale $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

Puis, en lui retranchant son espérance et en la divisant par son écart-type, on obtient une variable aléatoire qui suit la loi normale centrée réduite. \square

17.6 Variance d'échantillon

On considère à nouveau une population Ω dont les éléments possèdent un caractère quantitatif (discret ou continu) C qui est la réalisation d'une variable aléatoire X suivant une loi de probabilité d'espérance égale à μ et de variance égale à σ^2 (donc d'écart-type égal à σ).

Définition 17.6.1 (Variance d'échantillon). *On définit la variable aléatoire S_n^2 , appelée variance d'échantillon par :*

$$S_n^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n)^2.$$

Comme S_n^2 est une variable aléatoire réelle, il peut être judicieux d'étudier ses caractéristiques (espérance et variance).

Remarque 17.6.2 (ATTENTION). *Ne pas confondre la variance d'échantillon S_n^2 avec la variance de la moyenne d'échantillon $\text{Var}[\overline{X}_n]$.*

Théorème 17.6.3 (Distribution d'échantillonnage de la variance). *Quelle que soit la loi de la variable aléatoire X d'espérance μ et de variance σ^2 avec $\sigma > 0$ et quel que soit n , on a*

$$\mathbb{E}[S_n^2] = \frac{n-1}{n} \sigma^2 = \frac{n-1}{n} \text{Var}[X].$$

Démonstration. Par définition,

$$\begin{aligned}
 S_n^2 &:= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \\
 &= \frac{1}{n} \sum_{i=1}^n \{(X_i - \mu) - (\bar{X}_n - \mu)\}^2 \\
 &= \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 + \frac{1}{n} \sum_{i=1}^n (\bar{X}_n - \mu)^2 - \frac{2}{n} (\bar{X}_n - \mu) \sum_{i=1}^n (X_i - \mu) \\
 &= \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 + (\bar{X}_n - \mu)^2 - 2 (\bar{X}_n - \mu)^2 \\
 &= \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - (\bar{X}_n - \mu)^2 .
 \end{aligned}$$

La linéarité de l'espérance donne alors

$$\begin{aligned}
 \mathbb{E} [S_n^2] &= \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - (\bar{X}_n - \mu)^2 \right\} \\
 &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \{(X_i - \mu)^2\} - \mathbb{E} \{(\bar{X}_n - \mu)^2\} .
 \end{aligned}$$

Le second terme est égal à la variance de \bar{X}_n . Conséquemment, il vient

$$\begin{aligned}
 \mathbb{E} [S_n^2] &= \frac{1}{n} \sum_{i=1}^n \text{Var}[X_i] - \text{Var} [\bar{X}_n] \\
 &= \frac{1}{n} \sum_{i=1}^n \text{Var}[X] - \frac{\sigma^2}{n} \\
 &= \text{Var}[X] - \frac{\sigma^2}{n} \\
 &= \sigma^2 - \frac{\sigma^2}{n} \\
 &= \frac{n-1}{n} \sigma^2 ,
 \end{aligned}$$

ce qui achève la preuve du théorème. \square

Le calcul de la *variance de la variance d'échantillon* nécessite que la variable aléatoire X admette un moment d'ordre quatre fini.

Proposition 17.6.4. *Quelle que soit la loi de la variable aléatoire X d'espérance μ , de variance σ^2 avec $\sigma > 0$ et de moment centré d'ordre quatre égal à μ_4 (c'est-à-dire tel que $\mathbb{E} [(X - \mu)^4] = \mu_4$) et quel que soit n , on a*

$$\text{Var} [S_n^2] = \frac{n-1}{n^3} ((n-1)\mu_4 - (n-3)\sigma^4) .$$

Exercice 17.6.5. Démontrer la Proposition.

Remarque 17.6.6. On peut aussi écrire le résultat comme suit :

$$\text{Var}[S_n^2] = \sigma^4 \frac{n-1}{n^3} ((n-1)\beta_2 - (n-3)) ,$$

où β_2 est le kurtosis, voir la page 129.

Remarque 17.6.7. On dispose de l'équivalence $\text{Var}[S_n^2] \sim \sigma^4 \frac{\beta_2-1}{n}$ pour n grand. Conséquemment, l'espérance de S_n^2 tend vers la variance de X et la variance de S_n^2 tend vers 0. On peut ainsi montrer la convergence en probabilité de S_n^2 vers $\sigma^2 = \text{Var}[X]$.

On peut noter que la moyenne de la variance d'échantillon n'est pas exactement égale à la variance de X . C'est pourquoi on introduit la variance corrigée (ou modifiée) où l'on ne divise pas par le nombre de termes de la somme, mais par le nombre de termes indépendants, à savoir $n-1$. En effet, les termes sont liés par la relation $\sum_{i=1}^n (X_i - \bar{X}_n) = 0$.

Définition 17.6.8 (Variance d'échantillon corrigée). On définit la variable aléatoire \widetilde{S}_n^2 pour $n \geq 2$, appelée variance d'échantillon corrigée, par

$$\widetilde{S}_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 .$$

On a donc

$$\widetilde{S}_n^2 = \frac{n}{n-1} S_n^2 .$$

Conséquemment, il vient immédiatement :

Proposition 17.6.9. Quelle que soit la loi de la variable aléatoire X d'espérance μ et de variance σ^2 avec $\sigma \in]0; +\infty[$ et quel que soit n , on a

$$\mathbb{E}[\widetilde{S}_n^2] = \sigma^2 = \text{Var}[X] .$$

On dira par la suite que \widetilde{S}_n^2 est un estimateur sans biais tandis que S_n^2 est un estimateur asymptotiquement sans biais.

Le prochain théorème est admis.

Théorème 17.6.10. Si la variable aléatoire X suit la loi normale $\mathcal{N}(\mu, \sigma^2)$ avec $\sigma > 0$, alors la variable aléatoire $(n-1) \frac{\widetilde{S}_n^2}{\sigma^2}$ suit la loi du Khi-deux à $n-1$ degrés de liberté, $\chi^2(n-1)$.

Remarque 17.6.11. Le Théorème 17.6.10 se prouve en utilisant le théorème de Cochran (Théorème 13.7.1) sur les vecteurs gaussiens, voir la page 198.

Remarque 17.6.12. La variable aléatoire $(n-1) \frac{\widetilde{S}_n^2}{\sigma^2}$ est en fait égale à la variable aléatoire $n \frac{S_n^2}{\sigma^2}$.

17.7 Proportion d'échantillon

Il arrive que le caractère C à estimer ne soit pas quantitatif mais qualitatif. Soit p la proportion d'individus présentant le caractère étudié dans la population mère Ω .

La proportion d'individus ayant le caractère C obtenue dans un n -échantillon est la valeur observée d'une variable aléatoire F_n , fréquence d'apparition dans un échantillon de taille n , appelée proportion d'échantillon. On parle aussi de fréquence statistique.

On peut lier la fréquence statistique à la fréquence d'une série statistique simple.

Définition 17.7.1. *On définit la variable aléatoire F_n , que l'on appelle proportion d'échantillon ou fréquence statistique par*

$$F_n := \frac{K_n}{n},$$

où K_n est la variable aléatoire qui compte le nombre d'apparitions du caractère considéré dans un échantillon de taille n .

On peut écrire $K_n(\omega_1, \dots, \omega_n) := \sum_{i=1}^n \mathbb{1}_C(\omega_i) = \sum_{i=1}^n \mathbb{1}_{C_i}(\omega)$ où l'univers Ω est lié au tirage de n individus dans la population. Il s'agit ainsi d'une somme de n variables aléatoires indépendantes suivant la loi de Bernoulli $\mathcal{B}(p)$. Conséquemment, la variable aléatoire K_n suit la loi binomiale $\mathcal{B}(n, p)$. On en déduit donc

$$\mathbb{E}[K_n] = np,$$

ainsi que

$$\text{Var}[K_n] = np(1 - p).$$

On en déduit le théorème suivant.

Théorème 17.7.2 (Distribution d'échantillonnage des proportions). *Pour tout entier strictement positif n , on a*

$$\mathbb{E}[F_n] = p,$$

ainsi que

$$\text{Var}[F_n] = \frac{p(1 - p)}{n}.$$

Exercice 17.7.3. *Démontrer le théorème.*

Ainsi, la variance de la variable aléatoire réelle F_n tend vers 0. Il s'ensuit que F_n converge en probabilité vers p . On dira que l'estimateur F_n est convergent.

17.8 Estimateurs

Dans la suite de ce chapitre, nous nous intéressons à l'estimation d'un paramètre $\theta^* \in \Theta \subset \mathbb{R}^d$ d'une loi parente, à partir d'un n -échantillon. Ce paramètre θ^* peut être la moyenne μ , la variance σ^2 , une proportion p ou même le couple (μ, σ^2) .

17.8.1 Préliminaires

Définition 17.8.1. *Un estimateur $\hat{\theta}_n$ est une statistique permettant d'évaluer un paramètre **inconnu** θ^* relatif à la loi de probabilité parente.*

Le paramètre θ^* étant inconnu, la fonction φ telle que $\hat{\theta}_n := \varphi(X_1, \dots, X_n)$ ne doit pas dépendre de θ^* .

Définition 17.8.2. *On parle d'estimation de θ^* associée à cet estimateur la valeur observée lors de l'expérience, c'est-à-dire la valeur prise par la fonction au point observé $(x[1], \dots, x[n])$.*

Exemple 17.8.3. *Une usine d'embouteillage possède une machine qui remplit des bouteilles de un litre. Cette machine n'est pas très précise et le volume de boisson versé dans une bouteille est toujours supérieur à un litre. On supposera par la suite que le surplus sur une journée suit une loi uniforme entre 0 et θ , θ représentant le nombre maximal de litres que peut verser en trop la machine en une journée.*

On aimerait avoir une estimation de ce nombre. Pour cela, un technicien regarde la machine pendant n jours et note pour chaque $i \in \{1, \dots, n\}$, le nombre de litres X_i que la machine a versé en trop durant la journée numéro i .

On rappelle que si X suit une loi uniforme sur $[0; \theta]$, alors : $\mathbb{E}[X] = \frac{\theta}{2}$ et $\text{Var}[X] = \frac{\theta^2}{12}$.

On propose deux stratégies pour estimer θ . La première consiste à considérer l'estimateur

$$\hat{\theta}_1(n) := 2\overline{X}_n.$$

La deuxième consiste à définir $\hat{\theta}_2(n)$ comme étant la plus grande des variables aléatoires parmi X_1, \dots, X_n .

Quel est le meilleur estimateur ?

17.8.2 Qualité des estimateurs

Il semble logique de prendre comme estimation d'un paramètre de la population totale Ω celui de l'échantillon, mais est-ce là la meilleure estimation ?

Comme on le verra par la suite, le deuxième estimateur de l'Exemple 17.8.3 est, sur certains points, de meilleure qualité que le premier.

Dans ce paragraphe, on va détailler certaines des propriétés souhaitées pour un estimateur : l'absence de biais, la convergence, le caractère asymptotiquement normal, l'efficacité et la robustesse.

17.8.2.1 Absence de biais

Lorsqu'on évalue les qualités d'un appareil de mesure, on considère essentiellement la justesse (évaluée *a contrario* par l'erreur systématique du réglage).

Un estimateur nous fournit, comme un appareil de mesure, une valeur numérique : d'une manière analogue, nous évaluons les qualités d'un estimateur par sa justesse.

Pour ce faire, on regardera l'espérance de l'estimateur.

Remarque 17.8.4. La loi de probabilité d'un estimateur $\hat{\theta}_n$ dépend de la valeur du paramètre θ^* . C'est pourquoi, dans la suite de ce cours, on note $\mathbb{E}_{\theta^*}[\hat{\theta}_n]$ (respectivement $\text{Var}_{\theta^*}[\hat{\theta}_n]$) l'espérance (respectivement la variance) de la statistique $\hat{\theta}_n$.

Définition 17.8.5 (Biais d'un estimateur). On appelle biais de $\hat{\theta}_n$ pour θ^* la valeur

$$b_{\theta^*}(\hat{\theta}_n) := \mathbb{E}_{\theta^*}[\hat{\theta}_n] - \theta^* .$$

Définition 17.8.6. Un estimateur est dit sans biais si $b_{\theta^*}(\hat{\theta}_n) = 0$ c'est-à-dire si

$$\mathbb{E}_{\theta^*}[\hat{\theta}_n] = \theta^* .$$

On dit qu'un estimateur est sans biais si en moyenne, on ne fait pas d'erreur systématique.

Exemple 17.8.7. Si le paramètre à estimer est l'espérance de la loi $\theta^* = \mathbb{E}[X]$, l'estimateur naturel est la moyenne d'échantillon :

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i .$$

Alors, on a $\mathbb{E}_{\theta^*}[\bar{X}_n] = \mathbb{E}[X] = \theta^*$. Il s'ensuit que \bar{X}_n est un estimateur sans biais de l'espérance.

Pour certains estimateurs, la propriété précédente ne peut pas être strictement vérifiée, le biais diminuant seulement quand la taille n de l'échantillon augmente. Ceci correspond à la définition suivante.

Définition 17.8.8. On dit que $\hat{\theta}_n$ est un estimateur asymptotiquement sans biais de θ^* si pour tout $\theta \in \Theta$, on a $\lim_{n \rightarrow \infty} b_{\theta}[\hat{\theta}_n] = 0$ c'est-à-dire si

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\theta}[\hat{\theta}_n] = \theta .$$

Exemple 17.8.9. Si le paramètre à estimer est la variance de la loi, $\theta^* = \text{Var}[X]$, l'estimateur naturel est la variance d'échantillon :

$$S_n^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 .$$

Or, $\mathbb{E}_{\theta^*}[S_n^2] = \frac{n-1}{n} \text{Var}[X] = \frac{n-1}{n} \theta^*$. Il s'ensuit que S_n^2 n'est pas un estimateur sans biais de la variance. Mais, c'est un estimateur asymptotiquement sans biais.

D'autre part, un autre estimateur de la variance est la variance d'échantillon corrigée

$$\widetilde{S}_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 .$$

Puis, l'on rappelle que $\mathbb{E}_{\theta^*}[\widetilde{S}_n^2] = \text{Var}[X] = \theta^*$. Il s'ensuit que la variance d'échantillon corrigée, \widetilde{S}_n^2 , est un estimateur sans biais de la variance.

17.8.2.2 Convergence

L'absence de biais n'est pas suffisante pour s'assurer de l'efficacité d'un estimateur.

Définition 17.8.10 (Estimateur convergent). $\widehat{\theta}_n$ est un estimateur convergent si la suite de variables aléatoires $(\widehat{\theta}_n)_n$ converge en probabilité vers θ^* .

En anglais, on dit "consistent".

En d'autres termes, un estimateur est convergent si pour tout $\epsilon > 0$, la suite numérique de terme général

$$\mathbb{P} \left(\left| \widehat{\theta}_n - \theta^* \right| > \epsilon \right)$$

converge vers 0.

Remarque 17.8.11. La distribution d'un estimateur convergent tend à se concentrer autour de la valeur θ^* du paramètre à estimer quand la taille de l'échantillon augmente.

Démontrer la convergence en probabilité n'est pas facile. On peut toutefois montrer qu'un estimateur est convergent en utilisant le théorème suivant.

Théorème 17.8.12. Un estimateur asymptotiquement sans biais dont la variance tend vers zéro est convergent.

Démonstration. Fixons $\epsilon > 0$ arbitrairement petit. Montrons que

$$\mathbb{P} \left(\left| \widehat{\theta}_n - \theta^* \right| > \epsilon \right)$$

tend vers 0 quand n tend vers l'infini.

L'inégalité triangulaire nous donne :

$$\left| \widehat{\theta}_n - \mathbb{E}_{\theta^*}[\widehat{\theta}_n] \right| \geq \left| \widehat{\theta}_n - \theta^* \right| - \left| \mathbb{E}_{\theta^*}[\widehat{\theta}_n] - \theta^* \right|.$$

Conséquemment, si $\left| \widehat{\theta}_n - \theta^* \right| > \epsilon$, alors on a $\left| \widehat{\theta}_n - \mathbb{E}_{\theta^*}[\widehat{\theta}_n] \right| > \epsilon - \left| \mathbb{E}_{\theta^*}[\widehat{\theta}_n] - \theta^* \right|$. Il s'ensuit l'inégalité

$$\mathbb{P} \left(\left| \widehat{\theta}_n - \theta^* \right| > \epsilon \right) \leq \mathbb{P} \left(\left| \widehat{\theta}_n - \mathbb{E}_{\theta^*}[\widehat{\theta}_n] \right| > \epsilon - \left| \mathbb{E}_{\theta^*}[\widehat{\theta}_n] - \theta^* \right| \right).$$

Or, l'estimateur est asymptotiquement sans biais donc la quantité $\left| \mathbb{E}_{\theta^*}[\widehat{\theta}_n] - \theta^* \right|$ est plus petite que $\frac{\epsilon}{2}$ si n est assez grand. Donc, pour n assez grand, il vient

$$\mathbb{P} \left(\left| \widehat{\theta}_n - \theta^* \right| > \epsilon \right) \leq \mathbb{P} \left(\left| \widehat{\theta}_n - \mathbb{E}_{\theta^*}[\widehat{\theta}_n] \right| > \frac{\epsilon}{2} \right).$$

On utilise ensuite l'inégalité de Bienaymé-Tchebychev :

$$\mathbb{P} \left(\left| \widehat{\theta}_n - \theta^* \right| > \epsilon \right) \leq \frac{4}{\epsilon^2} \mathbb{E}_{\theta^*} \left\{ \left(\widehat{\theta}_n - \mathbb{E}_{\theta^*}[\widehat{\theta}_n] \right)^2 \right\} = \frac{4}{\epsilon^2} \text{Var}_{\theta^*}[\widehat{\theta}_n],$$

ce qui converge vers 0 quand n tend vers l'infini. La preuve est ainsi achevée. \square

Exemple 17.8.13. Si le paramètre à estimer est l'espérance de la loi, $\theta = \mathbb{E}[X]$, ce paramètre est estimé sans biais par \overline{X}_n . Or, on sait que l'on a

$$\text{Var}_{\theta}[\overline{X}_n] = \frac{\text{Var}[X]}{n} \longrightarrow 0.$$

Il s'ensuit que la moyenne d'échantillon est un estimateur convergent de l'espérance.

De la même manière, on sait, sous certaines hypothèses, que la variance de S_n^2 tend vers 0 quand n tend vers l'infini. Par conséquent, S_n^2 (ainsi que \widetilde{S}_n^2) est un estimateur convergent de la variance.

Remarque 17.8.14. Certaines calculatrices fournissent deux valeurs pour la variance d'un échantillon, σ_n^2 et $\widetilde{\sigma}_n^2$. La première valeur correspond à la valeur observée de S_n^2 et la seconde correspond à la valeur observée de $\widetilde{S}_n^2 = \frac{n}{n-1} S_n^2$. Ces deux valeurs sont peu différentes lorsque n est grand mais la seconde est une estimation non biaisée de la variance de la population.

Définition 17.8.15. Un estimateur $\widehat{\theta}_n$ est dit absolument correct si $\widehat{\theta}_n$ est sans biais et si sa variance tend vers 0 quand n tend vers l'infini.

Exemple 17.8.16. La moyenne d'échantillon est un estimateur absolument correct.

17.8.2.3 Normalité asymptotique

En vu des intervalles de confiance, il est parfois crucial que la loi de l'estimateur puisse être approchée par une loi normale quand n est grand.

Définition 17.8.17. *Un estimateur réel $\widehat{\theta}_n$ d'un paramètre réel θ^* est dit asymptotiquement normal s'il existe deux suites de réels $(m_n(\theta^*))_{n \in \mathbb{N}^*}$ et $(v_n(\theta^*))_{n \in \mathbb{N}^*}$ telles que $v_n(\theta^*) > 0$ pour tout $n \geq 1$ et*

$$\frac{\widehat{\theta}_n - m_n(\theta^*)}{v_n(\theta^*)} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Remarque 17.8.18. *En d'autres termes, pour tout $u \in \mathbb{R}$, on a*

$$\mathbb{P}_{\theta^*} \left(\frac{\widehat{\theta}_n - m_n(\theta^*)}{v_n(\theta^*)} \leq u \right) \longrightarrow \Phi(u),$$

quand n tend vers l'infini. Ici, $\Phi(u)$ est la fonction de répartition de la loi normale centrée réduite.

Pour établir ce genre de résultats, on s'adosse la plupart du temps au théorème central de la limite.

Exemple 17.8.19. *La proportion d'échantillon est asymptotiquement normale. Il en est de même avec la moyenne d'échantillon. Enfin, si le moment d'ordre quatre de la variable aléatoire X est fini, la variance d'échantillon est également asymptotiquement normale.*

17.8.2.4 Efficacité

La qualité d'un estimateur se mesure aussi par l'erreur quadratique moyenne.

Définition 17.8.20. *L'erreur quadratique moyenne d'un estimateur $\widehat{\theta}_n$ est définie comme étant $\mathbb{E}_{\theta^*} \left[\left(\widehat{\theta}_n - \theta^* \right)^2 \right]$.*

Proposition 17.8.21. *Soit $\widehat{\theta}_n$ un estimateur du paramètre θ^* à étudier. On a alors*

$$\mathbb{E}_{\theta^*} \left[\left(\widehat{\theta}_n - \theta^* \right)^2 \right] = \text{Var}_{\theta^*}[\widehat{\theta}_n] + \left(\mathbb{E}_{\theta^*}[\widehat{\theta}_n] - \theta^* \right)^2.$$

Exercice 17.8.22. *Démontrer la proposition.*

Remarque 17.8.23. *Entre deux estimateurs sans biais, le meilleur est celui dont la variance est minimale. On parle d'efficacité.*

Remarque 17.8.24. *Le critère d'erreur quadratique moyenne n'est pas parfait mais il est préféré à d'autres critères qui semblent plus naturels comme l'erreur absolue moyenne $\mathbb{E}_{\theta^*} \left[\left| \widehat{\theta}_n - \theta^* \right| \right]$ car il s'exprime en fonction de notions simples comme le biais et la variance et est relativement facile à manipuler analytiquement.*

17.8.2.5 Robustesse

La dernière propriété que l'on aime voir satisfaite par un estimateur est la robustesse. Cela signifie que l'ajout d'une nouvelle donnée ne perturbe pas complètement l'estimateur. On pense ici à des données extrémales et qui peuvent être aberrantes.

Ainsi, la moyenne d'échantillon n'est pas toujours robuste. Il en est de même avec l'estimateur du maximum de vraisemblance que nous verrons par la suite, voir page 279. Par conséquent, de nombreux statisticiens travaillent à fournir une version plus robuste de ce dernier.

17.8.3 Retour sur l'Exemple 17.8.3

17.8.3.1 Étude de $\hat{\theta}_1(n)$

On commence par étudier l'estimateur $\hat{\theta}_1(n) := 2\overline{X}_n = \frac{2}{n} \sum_{k=1}^n X_k$. D'abord, on sait que l'on a $\mathbb{E}_\theta [\hat{\theta}_1(n)] = 2\mathbb{E}_\theta[\overline{X}_n] = 2\frac{\theta}{2} = \theta$. Ainsi, l'estimateur est sans biais. On peut ensuite prouver facilement qu'il est convergent en utilisant la loi faible des grands nombres. En effet, la variance de la loi uniforme sur $[0; \theta]$ est $\frac{\theta^2}{12} < +\infty$. En fait, on sait même que $\hat{\theta}_1(n)$ converge presque sûrement et dans L^2 vers θ d'après la loi forte.

Le Théorème central de la limite implique immédiatement le caractère asymptotiquement normal de l'estimateur.

On peut par ailleurs calculer explicitement la loi de l'estimateur. En effet, la densité de probabilité de la variable aléatoire X est $f_X(x) := \frac{1}{\theta} \mathbb{1}_{[0; \theta]}(x)$. Or, les variables $(X_i)_i$ sont indépendantes et de même loi que X . De fait, la densité de probabilité de $Z := \sum_{k=1}^n X_k$ est

$$f_Z = f_{X_1} * \cdots * f_{X_n}.$$

Puis, on remarque $f_X(x) = \frac{1}{\theta} ((\delta_0 - \delta_\theta) * H)(x)$ où H est la fonction de Heaviside. Par associativité et commutativité dans l'algèbre de convolution des distributions causales, il vient :

$$f_Z(z) = \frac{1}{\theta^n} \left[\left((\delta_0 - \delta_\theta)^{*n} \right) * \left(H * \cdots * H \right) \right] (z).$$

On peut prouver par récurrence que l'on a $(H * \cdots * H)(z) = \frac{z^{n-1}}{(n-1)!} H(z) =: \xi_n(z)$. Donc :

$$f_Z(z) = \frac{\theta^{-n}}{(n-1)!} \left(\left(\sum_{k=0}^n \binom{n}{k} (-1)^k \delta_{k\theta} \right) * \xi \right) (z).$$

On obtient ainsi :

$$f_Z(z) = \frac{\theta^{-n}}{(n-1)!} \sum_{k=0}^n \binom{n}{k} (-1)^k (z - k\theta)^{n-1} H(z - k\theta).$$

Puis, la densité de probabilité de $\hat{\theta}_1(n)$ est

$$\begin{aligned} f_{\hat{\theta}_1(n)}(x) &= \frac{\theta^{-n}}{(n-1)!} \sum_{k=0}^n \binom{n}{k} (-1)^k \left(\frac{n}{2}x - k\theta\right)^{n-1} H\left(\frac{n}{2}x - k\theta\right) \\ &= \frac{\theta^{-n} n^{n-1}}{2^{n-1}(n-1)!} \sum_{k=0}^n \binom{n}{k} (-1)^k \left(x - \frac{2k\theta}{n}\right)^{n-1} H\left(x - \frac{2k\theta}{n}\right). \end{aligned}$$

Évidemment, le calcul explicite n'est pas aisé et il est donc préférable de se servir du théorème central de la limite pour attester de la normalité asymptotique.

Terminons l'étude du premier estimateur par le calcul de son erreur quadratique moyenne. Comme l'estimateur est sans biais, la Proposition 17.8.21 implique que l'erreur quadratique moyenne est simplement sa variance :

$$\text{Var} \left[\hat{\theta}_1(n) \right] = \text{Var} [2\bar{X}_n] = 4\text{Var} [\bar{X}_n] = \frac{4\text{Var} [X]}{n}.$$

17.8.3.2 Étude de $\hat{\theta}_2(n)$

On étudie maintenant l'estimateur $\hat{\theta}_2(n) := \max(X_1, \dots, X_n)$. Pour calculer son biais, son caractère asymptotiquement normal, son erreur quadratique moyenne et sa convergence en probabilité, on va calculer la densité de probabilité de $\hat{\theta}_2(n)$.

Pour ce faire, on regarde d'abord sa fonction de répartition :

$$\begin{aligned} F_{\hat{\theta}_2(n)}(x) &= \mathbb{P} \left(\hat{\theta}_2(n) \leq x \right) \\ &= \mathbb{P} \left(\max(X_1, \dots, X_n) \leq x \right) \\ &= \mathbb{P} \left(X_1 \leq x, \dots, X_n \leq x \right) \\ &= \mathbb{P}(X \leq x)^n. \end{aligned}$$

Ainsi, si $x \geq \theta$, $F_{\hat{\theta}_2(n)}(x) = 1^n = 1$. Si $x \leq 0$, $F_{\hat{\theta}_2(n)}(x) = 0^n = 0$. Enfin, si $x \in [0; \theta]$, on a

$$F_{\hat{\theta}_2(n)}(x) = \left(\frac{x}{\theta}\right)^n.$$

Il s'ensuit $f_{\hat{\theta}_2(n)}(x) = \frac{nx^{n-1}}{\theta^n} \mathbb{1}_{[0; \theta]}(x)$.

On peut alors calculer l'espérance comme suit :

$$\begin{aligned}
\mathbb{E}_\theta \left[\widehat{\theta}_2(n) \right] &= \int_{\mathbb{R}} x f_{\widehat{\theta}_2(n)}(x) dx \\
&= \int_0^\theta x \frac{nx^{n-1}}{\theta^n} dx \\
&= \frac{n}{\theta^n} \int_0^\theta x^n dx \\
&= \frac{n}{\theta^n} \frac{1}{n+1} \theta^{n+1} \\
&= \frac{n}{n+1} \theta.
\end{aligned}$$

Immédiatement, on en déduit que le deuxième estimateur est asymptotiquement sans biais. Néanmoins, il n'est pas sans biais. Il semble donc, au premier abord, préférable de prendre le premier estimateur puisque l'on ne commet pas d'erreur systématique dans l'estimation.

Calculons maintenant sa variance :

$$\begin{aligned}
\text{Var}_\theta \left[\widehat{\theta}_2(n) \right] &= \int_{\mathbb{R}} x^2 f_{\widehat{\theta}_2(n)}(x) dx - \left(\frac{n}{n+1} \theta \right)^2 \\
&= \frac{n}{\theta^n} \int_0^\theta x^{n+1} dx - \theta^2 \frac{n^2}{(n+1)^2} \\
&= \left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2} \right) \theta^2 \\
&= \frac{n}{(n+1)^2(n+2)} \theta^2.
\end{aligned}$$

Par conséquent, l'erreur quadratique moyenne du deuxième estimateur est

$$\frac{n}{(n+1)^2(n+2)} \theta^2 + \left(\frac{n}{n+1} \theta - \theta \right)^2 = \frac{2}{(n+1)(n+2)} \theta^2.$$

On constate cette fois que l'erreur quadratique moyenne décroît en $O\left(\frac{1}{n^2}\right)$ alors que celle du premier estimateur est seulement en $O\left(\frac{1}{n}\right)$.

Conséquemment, pour n assez grand, le deuxième estimateur est plus efficace.

17.8.3.3 Remarque sur le modèle étudié

Ici, nous nous sommes intéressés à une loi uniforme sur un intervalle dont on avait fixé la borne inférieure à 0. Mais, si l'on regardait une loi uniforme sur $[-\theta; \theta]$, alors la moyenne d'échantillon convergerait vers 0, qui ne dépend pas de θ . De fait, le premier estimateur perd tout intérêt et il faut regarder le moment d'ordre deux, c'est-à-dire la variance d'échantillon.

On verra dans le chapitre suivant que ce premier estimateur correspond en fait à la méthode dite des moments. Le second correspond à celle du maximum de vraisemblance, voir page 279.

Enfin, si on regardait un intervalle de la forme $[a; b]$, il faudrait regarder conjointement \bar{X}_n et S_n^2 . Plus exactement, l'estimateur de la méthode des moments de a serait $\bar{X}_n - \sqrt{3}S_n$ et celui de b serait $\bar{X}_n + \sqrt{3}S_n$. L'étude de ces deux estimateurs en serait plus complexe. Concernant le maximum de vraisemblance, on estimerait a par $\min(X_1, \dots, X_n)$ et b par $\max(X_1, \dots, X_n)$.

Deux méthodes classiques d'estimation ponctuelle

18.1 Introduction

Dans ce chapitre, on va présenter deux méthodes classiquement utilisées pour obtenir des estimateurs. Ces deux méthodes populaires sont la méthode des moments et celle du maximum de vraisemblance. La première est simple à mettre en œuvre dans la plupart des cas. Néanmoins, toutes les propriétés attendues que l'on a énumérées précédemment ne sont pas forcément vérifiées. La deuxième est tout aussi intuitive que la première mais elle est plus difficile à appliquer. Son intérêt principal est de fournir des propriétés très appréciables sur les estimateurs qu'elle fournit. Globalement, les deux peuvent nécessiter l'utilisation d'outils numériques pour fournir une valeur au moins approchée dans le cas de données concrètes.

Chacune a ses avantages et ses inconvénients si bien qu'il faut avoir vu les deux et faire le bon choix en fonction de l'usage que l'on en fera. Dans certains cas, l'estimation par la méthode des moments est moins bonne que l'estimation par le maximum de vraisemblance. Néanmoins, dans le cas de la loi Gamma par exemple, le calcul de la fonction de vraisemblance peut poser des problèmes (l'utilisation de l'ordinateur et d'algorithmes numériques est indispensable) tandis que l'estimation des moments est très facilement accessible.

Tous les chapitres précédents sont des pré-requis au présent chapitre.

Les objectifs principaux sont de sensibiliser le lecteur à la méthode des moments et à la méthode du maximum de vraisemblance. Notamment, nous fournissons des exemples et ceux-ci doivent absolument être étudiés en profondeur.

18.2 Méthode des moments

On commence par présenter la méthode des moments.

Soit un modèle statistique indexé par un ensemble Θ . Supposons donné un n -échantillon (X_1, \dots, X_n) d'une loi intégrable sous \mathbb{P}_θ pour chaque $\theta \in \Theta$. On cherche un estimateur $\hat{\theta}_n$ de θ^* .

Supposons dans un premier temps que Θ est une partie de \mathbb{R} .

L'idée de la méthode des moments est de prendre comme estimateur $\widehat{\theta}_n$ une valeur de θ telle que $\mathbb{E}_{\widehat{\theta}_n} =: m(\widehat{\theta}_n)$ coïncide avec la moyenne empirique observée.

18.2.1 Exemple introductif

Présentons cette méthode intuitive sur un exemple simple. On suppose que X suit la loi de Poisson de paramètre $\theta^* > 0$, $\mathcal{P}(\theta^*)$. Ici, $\theta^* \in]0; +\infty[$. Or, $\Theta :=]0; +\infty[$ est bien un intervalle de \mathbb{R} .

Un estimateur par la méthode des moments est alors une solution de l'équation

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = m(\widehat{\theta}_n).$$

Si m réalise une bijection de Θ dans $m(\Theta)$, alors $\widehat{\theta}_n = m^{-1}(\overline{X}_n)$.

Pour la loi de Poisson de paramètre θ^* , la moyenne est égale à θ^* . De fait, on prend $\widehat{\theta}_n := \overline{X}_n$.

Ainsi, si le nombre d'observations n est assez grand, on prend comme estimation du paramètre la valeur $\frac{1}{n} \sum_{i=1}^n x[i]$.

18.2.2 Méthode générale

Si l'on cherchait à appliquer cette méthode à la loi uniforme sur l'intervalle centré $[-\theta; \theta]$, nous n'y arriverions pas vu que la moyenne est égale à 0 et de fait elle ne dépend pas de θ .

Formalisons maintenant la méthode générale.

D'abord, on suppose que Θ est une partie de \mathbb{R}^d avec $d \geq 1$.

On suppose également que le moment d'ordre p de la variable aléatoire X est fini.

Enfin, on admet que parmi les p premiers moments, il y en a au moins d qui dépendent du paramètre θ . Dit plus rigoureusement, l'application de Θ dans \mathbb{R}^d qui à θ associe $(\mathbb{E}_\theta(X^{k_1}), \dots, \mathbb{E}_\theta(X^{k_d}))$ est injective; où $1 \leq k_1 < \dots < k_d \leq p$.

Exemple 18.2.1. Si l'on cherche le paramètre $\theta \in \mathbb{R}_+^*$ d'une loi uniforme sur l'intervalle centré $[-\theta; \theta]$, on doit donc aller jusqu'au moment d'ordre deux vu que le moment d'ordre un est constant et égal à 0.

Alors, l'estimateur par la méthode des moments est la solution (unique) du système d'équations

$$\frac{1}{n} \sum_{i=1}^n X_i^{k_l} = m_{k_l}(\widehat{\theta}_n),$$

pour tout $l \in \llbracket 1; d \rrbracket$ où $m_k(\widehat{\theta}_n) := \mathbb{E}_{\widehat{\theta}_n}(X^k)$ pour tout $k \geq 1$.

18.2.3 Exemples

Dans cette section, on présente quelques exemples pour que le lecteur se familiarise bien à la méthode des moments.

18.2.3.1 Loi de Bernoulli

On suppose que X suit la loi $\mathcal{B}(p)$ avec $p \in]0; 1[$. Alors, on sait que $\mathbb{E}[X] = p$. D'où l'on prendra $\widehat{p}_n := \frac{1}{n} \sum_{k=1}^n X_k$, la proportion d'échantillon.

18.2.3.2 Loi géométrique

On suppose que X suit la loi géométrique de paramètre $p \in]0; 1[$. Alors,

$$\mathbb{E}[X] = \frac{1}{p}.$$

D'où l'on prendra $\widehat{p}_n := \frac{1}{\overline{X}_n} = \frac{n}{\sum_{k=1}^n X_k}$.

18.2.3.3 Loi uniforme

On suppose que X suit la loi uniforme sur l'intervalle $[a; b]$ avec $a < b$. On a ici deux paramètres donc il faudra deux équations.

Or, $\mathbb{E}[X] = \frac{a+b}{2}$ et $\text{Var}[X] = \frac{(b-a)^2}{12}$. Ainsi, on peut écrire

$$a = \mathbb{E}[X] - \sqrt{3}\sqrt{\text{Var}[X]} \quad \text{et} \quad b = \mathbb{E}[X] + \sqrt{3}\sqrt{\text{Var}[X]}.$$

Par conséquent, on prendra les estimateurs de la méthode des moments sont $\widehat{a}_n := \overline{X}_n - \sqrt{3}S_n$ et $\widehat{b}_n := \overline{X}_n + \sqrt{3}S_n$.

18.2.3.4 Loi exponentielle

On suppose que X suit la loi exponentielle de paramètre $\lambda > 0$. On sait que l'on a $\mathbb{E}[X] = \frac{1}{\lambda}$. On prendra donc $\widehat{\lambda}_n := \frac{1}{\overline{X}_n} = \frac{n}{\sum_{k=1}^n X_k}$.

18.2.3.5 Loi normale

On suppose que X suit la gaussienne de paramètres $m \in \mathbb{R}$ et $\sigma^2 > 0$. On sait que m correspond à l'espérance tandis que σ^2 correspond à la variance. Ainsi, on prendra $\widehat{m}_n := \overline{X}_n$ et $\widehat{\sigma}_n^2 := S_n^2$.

18.2.3.6 Loi d'Erlang

On suppose que X suit la loi d'Erlang de paramètres $r \in \mathbb{N}^*$ et $\lambda > 0$. Cette loi correspond à la loi de la somme de r variables aléatoires indépendantes et identiquement distribuées suivant la loi $\mathcal{E}(\lambda)$. On suppose r connu. On a donc $\mathbb{E}[X] = \frac{r}{\lambda}$. On prendra ainsi $\widehat{\lambda}_n := \frac{r}{\overline{X}_n} = \frac{nr}{\sum_{k=1}^n X_k}$.

18.2.3.7 Loi de Weibull

On suppose que X suit la loi de Weibull de paramètres $k \in \mathbb{N}^*$ et $\lambda > 0$. On suppose k connu. On peut montrer que son espérance est $\frac{\lambda}{k} \Gamma\left(\frac{1}{k}\right)$. On prendra donc comme estimateur : $\widehat{\lambda}_n := \frac{k}{\Gamma\left(\frac{1}{k}\right)} \overline{X}_n$.

On peut noter que le moment d'ordre k est égal à λ^k d'où l'on pourrait aussi prendre comme estimateur $\widehat{\lambda}_n := \left(\frac{1}{n} \sum_{i=1}^n x[i]^k\right)^{\frac{1}{k}}$.

18.2.3.8 Loi de Pareto

On suppose que X suit la loi de Pareto de paramètres $x_0 > 0$ et $\alpha > 0$. On suppose que l'on connaît x_0 . On suppose également que α est strictement plus grand que 1. En effet, cette condition est nécessaire pour avoir une espérance. Et, cette espérance vaut

$$\int_{x_0}^{\infty} \alpha \left(\frac{x_0}{x}\right)^{\alpha} dx = \frac{\alpha}{\alpha - 1} x_0.$$

Alors, on prendra comme estimateur $\widehat{\alpha}_n := \frac{\overline{X}_n}{\overline{X}_n - x_0}$.

Si $\alpha > 2$, on peut aussi estimer l'autre paramètre à savoir x_0 en utilisant le second moment :

$$\mathbb{E}[X^2] = \frac{\alpha}{\alpha - 2} x_0^2,$$

d'où

$$\text{Var}[X] = \frac{\alpha x_0^2}{(\alpha - 1)^2 (\alpha - 2)}.$$

On est ensuite amené à résoudre

$$\frac{\alpha}{\alpha - 1} x_0 = \overline{X}_n \quad \text{et} \quad \frac{\alpha x_0^2}{(\alpha - 1)^2 (\alpha - 2)} = S_n^2.$$

18.2.3.9 Loi de Cauchy

On suppose que X suit la loi de Cauchy de densité de probabilité

$$f_X(x) = \frac{1}{\pi} \frac{b}{b^2 + (x - a)^2}.$$

X n'a aucun moment donc la méthode ne peut pas être utilisée.

18.2.4 Limites

Il convient de noter que la méthode des moments ne peut être appliquée qu'à des modèles paramétriques admettant un nombre suffisant de moments. Ceci est vérifié pour la plupart des lois que l'on a étudiées. Mais, dans la pratique, on peut se retrouver confronté à d'autres types de lois. Par exemple, la loi de Cauchy est un bon contre-exemple à l'application de la méthode des moments.

De plus, les équations à résoudre sont, par essence, non linéaires si bien qu'il est difficile, en toute généralité, de les résoudre formellement. Il faut alors faire appel à des outils numériques.

Enfin, si la taille de l'échantillon n'est pas suffisamment grande, la loi des grands nombres ne s'applique pas et par conséquent, les moments empiriques n'approchent pas suffisamment les moments théoriques.

18.3 Méthode du maximum de vraisemblance

Présentons d'abord la *philosophie* de la méthode. Pour ce faire, il convient de signaler que le paragraphe suivant n'est pas des plus rigoureux et trahit clairement la vision de l'auteur qui est un probabiliste et non pas un statisticien.

On suppose que l'on dispose d'un modèle paramétrique où le paramètre $\theta \subset \mathbb{R}^d$ peut prendre un nombre fini r de valeurs : $\theta_1, \dots, \theta_r$.

On dispose d'un n -échantillon (X_1, \dots, X_n) dont les valeurs que l'on a observées sont $(x[1], \dots, x[n])$. On suppose enfin que la variable aléatoire X est discrète quelle que soit la valeur de θ .

On admet également que l'on ne dispose d'aucun *a priori* (aussi appelé *prior* en statistiques bayésiennes) sur le choix entre $\theta_1, \dots, \theta_k$. En d'autres termes, $\mathbb{P}(\theta = \theta_1) = \dots = \mathbb{P}(\theta = \theta_r) = \frac{1}{r}$.

Alors, d'après la formule de Bayes (Théorème 2.4.8), on a :

$$\mathbb{P}(\theta = \theta_l | X_1 = x[1], \dots, X_n = x[n]) = \frac{\prod_{i=1}^n \mathbb{P}(X = x[i] | \theta = \theta_l)}{\sum_{k=1}^r \prod_{i=1}^n \mathbb{P}(X = x[i] | \theta = \theta_k)}.$$

Il est alors raisonnable de choisir le mode, c'est-à-dire dans le cas présent la valeur l telle que la probabilité conditionnelle $\mathbb{P}(\theta = \theta_l | X_1 = x[1], \dots, X_n = x[n])$

soit la plus élevée. Pour ce faire, il suffit de maximiser la quantité $\prod_{i=1}^n \mathbb{P}(X = x[i] \mid \theta = \theta_l)$.

Cette quantité sera appelée vraisemblance. On cherche ainsi à maximiser la vraisemblance.

18.3.1 Exemple introductif

Plutôt que de théoriser tout de suite, on considère un exemple introductif, ce dernier étant le même que celui que l'on a donné pour la méthode des moments.

On suppose que X suit la loi de Poisson de paramètre $\theta^* > 0$, $\mathcal{P}(\theta^*)$. Ici, $\theta^* \in]0; +\infty[$. On ne dispose donc plus d'un nombre fini de modalités possibles pour θ .

La vraisemblance est ici :

$$L_{(x[1], \dots, x[n])}(\theta) := \prod_{i=1}^n \frac{\theta^{x[i]}}{x[i]!} e^{-\theta} = \frac{\exp[-(n\theta - \sum_{i=1}^n x[i] \log(\theta))]}{\prod_{i=1}^n x[i]!}.$$

Pour maximiser cette fonction, on minimise $\theta \mapsto -\log(L_{(x[1], \dots, x[n])}(\theta))$. Pour ce faire, on dérive et l'on a un unique point critique qui est atteint en

$$\hat{\theta}_n := \frac{1}{n} \sum_{i=1}^n x[i].$$

On peut ensuite vérifier que ce minimum critique correspond bien à un maximum global.

On note que cet estimateur n'est rien d'autre que l'estimateur obtenu avec la méthode des moments.

18.3.2 Méthode générale

On présente ici la méthode générale.

Soit X une variable aléatoire réelle de loi paramétrique (discrète ou continue), dont on veut estimer le paramètre θ . Alors, on définit une fonction f comme suit :

$$f(x, \theta) := f_{\theta}(x)$$

si X est une variable aléatoire continue de densité f_{θ} et

$$f(x, \theta) := \mathbb{P}_{\theta}(X = x)$$

si X est une variable aléatoire discrète.

Définition 18.3.1. On appelle fonction de vraisemblance de θ pour une réalisation $(x[1], \dots, x[n])$ d'un échantillon, la fonction de θ :

$$L_{(x[1], \dots, x[n])}(\theta) := \prod_{i=1}^n f(x[i], \theta).$$

La méthode qui consiste à estimer θ par la valeur qui maximise $L_{(x[1], \dots, x[n])}$ (la vraisemblance) s'appelle "méthode du maximum de vraisemblance".

On appelle $\hat{\theta}$ l'estimateur associé :

$$\hat{\theta} := \left\{ \theta : L_{(x[1], \dots, x[n])}(\hat{\theta}) = \sup_{\theta \in \Theta} L_{(x[1], \dots, x[n])}(\theta) \right\}.$$

Ceci est un problème d'optimisation. On utilise généralement le fait que si $L_{(x[1], \dots, x[n])}$ est dérivable et si $L_{(x[1], \dots, x[n])}$ admet un maximum global en une valeur, alors la dérivée première de $L_{(x[1], \dots, x[n])}$ s'y annule et sa dérivée seconde y est négative.

Réciproquement, si la dérivée première de $L_{(x[1], \dots, x[n])}$ s'annule en $\theta_0 := \theta^*$, et si sa dérivée seconde est négative strictement en θ^* , alors θ^* est un point où $L_{(x[1], \dots, x[n])}$ admet un maximum local (et non global). Il est alors nécessaire de vérifier que le maximum est global.

La vraisemblance étant positive et le logarithme népérien, \log , étant une fonction croissante, il est équivalent et souvent plus simple de maximiser le logarithme népérien de la vraisemblance (le produit se transforme en somme, ce qui est plus simple à dériver) et plus facile à calculer numériquement.

18.3.3 Exemples

Dans cette section, on présente quelques exemples pour que le lecteur se familiarise bien à la méthode du maximum de vraisemblance.

18.3.3.1 Loi de Bernoulli

On suppose que X suit la loi de Bernoulli de paramètre $p \in]0; 1[$. On observe n valeurs $x[1], \dots, x[n]$. Si $x = 0$, $\mathbb{P}(X = x) = 1 - p$ sinon $\mathbb{P}(X = x) = p$. Alors, si q est la proportion d'éléments égaux à 1, la vraisemblance est

$$L_{(x[1], \dots, x[n])}(p) := (1 - p)^{n(1-q)} p^{nq}.$$

On prend le logarithme népérien et on dérive par rapport à p pour trouver l'unique point critique qui est

$$\hat{p}_n = q.$$

Or, $q = \frac{1}{n} \sum_{i=1}^n X_i$ donc l'estimateur du maximum de vraisemblance coïncide avec celui de la méthode des moments dans ce cas précis.

18.3.3.2 Loi géométrique

On suppose que X suit la loi géométrique de paramètre $p \in]0; 1[$. Alors $\mathbb{P}_p(X = k) = (1 - p)^{k-1}p$ d'où la vraisemblance est

$$L_{(x[1], \dots, x[n])}(p) = \prod_{i=1}^n p(1-p)^{x[i]-1} = \left(\frac{p}{1-p}\right)^n (1-p)^{\sum_{i=1}^n x[i]}.$$

On prend le logarithme népérien, on dérive et l'on aboutit à l'unique point critique suivant :

$$p = \frac{n}{\sum_{i=1}^n x[i]}.$$

Ainsi, à nouveau l'estimateur du maximum de vraisemblance est le même que celui de la méthode des moments.

18.3.3.3 Loi uniforme

On suppose que X suit la loi uniforme sur $[a; b]$. On a alors $f_X(x) = \frac{1}{b-a} \mathbb{1}_{[a; b]}(x)$. Ainsi, la vraisemblance est

$$L_{(x[1], \dots, x[n])}(a, b) = \frac{1}{(b-a)^n} \prod_{i=1}^n \mathbb{1}_{[a; b]}(x[i]).$$

Pour maximiser, il faut d'abord s'assurer que $x[i] \in [a; b]$ pour tout $i \in \llbracket 1; n \rrbracket$. On a donc $\hat{a}_n \leq \min(X_1, \dots, X_n)$ et $\max(X_1, \dots, X_n) \leq \hat{b}_n$. Puis, il faut minimiser $b - a$ en respectant cette contrainte. On aboutit donc à

$$\hat{a}_n := \min(X_1, \dots, X_n) \quad \text{et} \quad \hat{b}_n = \max(X_1, \dots, X_n).$$

Cette fois, le maximum de vraisemblance ne correspond pas à la méthode des moments. On peut vérifier que ces estimateurs ont un biais mais que leur erreur quadratique moyenne est plus faible que celle de la méthode des moments, voir l'Exemple 17.8.3.

18.3.3.4 Loi exponentielle

On suppose que X suit la loi exponentielle de paramètre $\lambda > 0$. Alors, la vraisemblance est

$$L_{(x[1], \dots, x[n])}(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n x[i]} \prod_{i=1}^n H(x[i]).$$

On prend le logarithme népérien, on dérive et l'on trouve un unique point critique :

$$\lambda_0 = \frac{n}{\sum_{i=1}^n x[i]}.$$

Ainsi, à nouveau l'estimateur du maximum de vraisemblance est le même que celui de la méthode des moments.

18.3.3.5 Loi normale

On suppose que X suit la loi normale de paramètres m et $\sigma^2 > 0$. Alors, la vraisemblance est :

$$L_{(x[1], \dots, x[n])}(m, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}} \right)^n (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x[i] - m)^2 \right\}.$$

Après avoir pris le logarithme népérien et avoir pris les dérivées partielles, on aboutit à

$$\widehat{m}_n := \frac{1}{n} \sum_{i=1}^n x[i]$$

et

$$\widehat{\sigma}_n^2 := \frac{1}{n} \sum_{i=1}^n \left(x[i] - \frac{1}{n} \sum_{j=1}^n x[j] \right)^2.$$

Ainsi, les deux estimateurs sont respectivement \overline{X}_n et S_n^2 à savoir la moyenne d'échantillon et la variance d'échantillon non corrigée.

Il convient de noter que si l'on suppose m connu, alors l'estimation de σ^2 n'est plus biaisée.

18.3.3.6 Loi d'Erlang

On suppose que X suit la loi d'Erlang de paramètres $r \in \mathbb{N}^*$ et $\lambda > 0$. On suppose r connu. Alors, la densité de probabilité est $f_X(x) := \lambda^r \frac{x^{r-1}}{(r-1)!} e^{-\lambda x} H(x)$. La vraisemblance est donc

$$L_{(x[1], \dots, x[n])} = \lambda^{nr} \frac{\prod_{i=1}^n x[i]^{r-1}}{((r-1)!)^n} e^{-\lambda \sum_{i=1}^n x[i]} \prod_{i=1}^n H(x[i]).$$

Le calcul fournit immédiatement $\widehat{\lambda}_n = \frac{r}{\overline{X}_n}$, ce qui correspond à la méthode des moments.

18.3.3.7 Loi de Weibull

On suppose que X suit la loi de Weibull de paramètres $k \in \mathbb{N}^*$ et $\lambda > 0$. On suppose k connu. La densité de probabilité de X est $f_X(x) := kx^{k-1}\lambda^{-k}e^{-\left(\frac{x}{\lambda}\right)^k}H(x)$.

Ainsi, la vraisemblance est

$$L_{(x[1], \dots, x[n])} = k^n \prod_{i=1}^n x[i]^{k-1} \exp \left\{ -nk \log(\lambda) - \frac{1}{\lambda^k} \sum_{i=1}^n x[i]^k \right\}.$$

On prend le logarithme népérien puis l'on dérive par rapport à λ et l'on obtient un unique point critique ; qui est

$$\lambda_0 = \left(\frac{1}{n} \sum_{i=1}^n x[i]^k \right)^{\frac{1}{k}}.$$

L'estimateur fourni par le maximum de vraisemblance n'est donc pas à proprement parler celui de la méthode des moments (si l'on prend le premier moment). Néanmoins, si l'on prend le moment d'ordre k , cela coïncide bien.

18.3.3.8 Loi de Pareto

On suppose que X suit la loi de Pareto de paramètres $x_0 > 0$ et $\alpha > 0$. On suppose que l'on connaît x_0 . On ne suppose ni $\alpha > 1$ ni $\alpha > 2$. On a alors :

$$L_{(x[1], \dots, x[n])}(\alpha) = \alpha^n x_0^{n\alpha} \left(\prod_{i=1}^n x[i] \right)^{-\alpha-1} \prod_{i=1}^n \mathbb{1}_{[x_0; +\infty[}(x[i]).$$

On prend le logarithme népérien et l'on dérive par rapport à α et l'on obtient :

$$\frac{\partial}{\partial \alpha} \log(L_{(x[1], \dots, x[n])}(\alpha)) = \frac{n}{\alpha} + n \log(x_0) - \sum_{i=1}^n \log(x[i]).$$

Ainsi, l'estimateur associé est

$$\widehat{\alpha}_n := \frac{1}{\frac{1}{n} \sum_{i=1}^n \log\left(\frac{x[i]}{x_0}\right)}.$$

Si on ne connaît pas x_0 , on l'estime naturellement par

$$\widehat{x_0(n)} := \min(X_1, \dots, X_n)$$

et de fait

$$\widehat{\alpha}_n := \frac{1}{\frac{1}{n} \sum_{i=1}^n \log\left(\frac{x[i]}{\min(x[1], \dots, x[n])}\right)}.$$

Ainsi, on pouvait définir l'estimateur du maximum de vraisemblance sans hypothèse supplémentaire ; contrairement à la méthode des moments.

18.3.3.9 Loi de Cauchy

Comme pour la loi de Pareto avec $\alpha \leq 1$, on peut définir l'estimateur du maximum de vraisemblance alors que la méthode des moments était inapplicable.

On trouve ici une vraisemblance égale à

$$L_{(x[1], \dots, x[n])}(a, b) = \frac{1}{\pi^n} \frac{b^n}{\prod_{i=1}^n (b^2 + (x[i] - a)^2)}.$$

Dérivons la log-vraisemblance par rapport à b :

$$\frac{\partial}{\partial b} \log (L_{(x[1], \dots, x[n])}(a, b)) = \frac{n}{b} - 2b \sum_{i=1}^n \frac{1}{b^2 + (x[i] - a)^2}.$$

On dérive maintenant par rapport à a :

$$\frac{\partial}{\partial a} \log (L_{(x[1], \dots, x[n])}(a, b)) = 2 \sum_{i=1}^n \frac{x[i] - a}{b^2 + (x[i] - a)^2}.$$

Le calcul explicite nécessite d'utiliser les outils numériques pour trouver a_0 et b_0 satisfaisant $\frac{\partial}{\partial a} \log (L_{(x[1], \dots, x[n])}(a_0, b_0)) = 0$ et $\frac{\partial}{\partial b} \log (L_{(x[1], \dots, x[n])}(a_0, b_0)) = 0$.

Intervalles de confiance

19.1 Introduction

Comprendre ce chapitre requiert de maîtriser la plupart des chapitres précédents.

L'objectif principal du présent chapitre est d'abord de bien saisir que la définition d'un intervalle de confiance n'est pas unique, qu'elle dépend du contexte et qu'il faut faire preuve d'adaptabilité. Évidemment, il est aussi attendu de savoir construire un tel intervalle adapté à certaines situations comme le cas du modèle gaussien, de l'homoscédasticité. Par ailleurs, parfois, on ne peut pas fournir un tel intervalle et il nous faut nous restreindre à des intervalles dits asymptotiques. Enfin, on attend du lecteur qu'il sache quelles sont les lois à utiliser : normales, Khi-deux, Student, Erlang, Fisher... Notamment, il verra ainsi une utilisation pratique du théorème de Cochran (du chapitre sur les vecteurs gaussiens).

19.2 Préliminaires

Les estimations ponctuelles fournies dans les deux chapitres précédents (voir page 257 et page 275) sont totalement dépendantes de l'échantillon prélevé au hasard. En effet, estimer la moyenne μ par \overline{X}_n , c'est faire la "meilleure estimation possible". La meilleure estimation **ponctuelle** possible du moins.

La variable aléatoire \overline{X}_n n'étant pas une constante, selon l'aléa sous-jacent ω , on obtiendra ainsi un résultat donné ou un autre. En tout cas, il est très improbable que l'on ait $\overline{X}_n(\omega) = \mu$. Pire, si le modèle paramétrique sous-jacent est à densité, alors la probabilité que \overline{X}_n soit exactement égal à μ est 0.

Le même phénomène a lieu avec S_n^2 et avec \widetilde{S}_n^2 .

De fait, on peut changer de paradigme.

19.2.1 Philosophie des intervalles de confiance

En général, on ne peut pas évaluer exactement l'erreur que l'on commet en faisant une estimation ponctuelle. Néanmoins, en étudiant la loi de variabilité

de cette erreur, on peut calculer une marge d'erreur pour laquelle il existe une certitude suffisante qu'elle ne sera pas dépassée.

Typiquement, on se donne un risque α que l'on est prêt à accepter. Le risque en question dépend des usages que l'on en fait. Il peut s'agir de 10%, 5%, 1% voire 0.5% selon les sciences dans lesquelles on applique les intervalles de confiance en question.

Toute la difficulté est maintenant de fournir un intervalle de \mathbb{R} dans lequel le paramètre à estimer appartient avec une probabilité d'au moins $1 - \alpha$. Cela peut être pour une estimation de la moyenne, de la variance...

Il convient de noter que l'intervalle n'a aucune raison d'être unique. Également, plus le risque est faible, plus l'intervalle est large. Enfin, les intervalles de confiance que l'on va construire n'ont aucune raison d'être symétriques par rapport à l'estimateur sur lequel on se fonde. Il est cependant vrai que si l'estimateur est sans biais comme c'est le cas pour la moyenne dans le modèle gaussien, alors on prendra plutôt un intervalle symétrique. Néanmoins, nous verrons que dans le cas de la variance (même dans le modèle gaussien), l'intervalle est construit de manière très différente.

Le degré de confiance à savoir $1 - \alpha$ est aussi appelé le niveau de confiance ou le seuil de confiance.

Pour la moyenne μ , en notant par γ le degré de confiance et par d la marge d'erreur, on doit avoir

$$\gamma = \mathbb{P}(|\bar{X}_n - \mu| \leq d) , \quad (19.1)$$

où \bar{X}_n est la moyenne d'échantillon. C'est à partir de cette équation de base que sont déterminées toutes les quantités qui entrent en jeu dans un intervalle de confiance ; pour la moyenne.

19.2.2 Définitions

On présente maintenant la notion d'intervalle de confiance de façon plus rigoureuse.

Définition 19.2.1 (Intervalle de confiance). *Soit X une variable aléatoire suivant une loi qui dépend d'un paramètre inconnu $\theta^* \in \Theta \subset \mathbb{R}$ de la loi parente. Les intervalles de confiance de θ^* au seuil de risque $\alpha \in]0; 1[$ sont des intervalles*

$$\mathcal{I}_\alpha := [a(X_1, \dots, X_n, \alpha); b(X_1, \dots, X_n, \alpha)] ,$$

construits d'une façon telle qu'a priori une proportion $1 - \alpha$ de ces intervalles contienne la valeur de θ^ .*

Définition 19.2.2. *La quantité $1 - \alpha$ est ici appelée niveau de confiance de l'intervalle.*

Remarque 19.2.3. *Il est également nécessaire que l'intervalle \mathcal{I}_α ne dépende pas d'autres paramètres inconnus. En effet, il ne serait pas raisonnable de construire un intervalle de confiance pour un paramètre inconnu en utilisant pour cela un paramètre que l'on ne connaît même pas.*

19.2.3 Premier intervalle de confiance de la moyenne

Une première méthode naturelle est d'utiliser des inégalités classiques. Toutefois, nous verrons que les intervalles fournis ne sont pas optimaux.

La première inégalité à laquelle on pense pour obtenir un intervalle de confiance symétrique est l'inégalité de Bienaymé-Tchebychev. En effet, si l'on connaît la variance σ^2 , alors on a :

$$\mathbb{P}(|X - \mathbb{E}[X]| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

En prenant $\epsilon_\alpha := \sqrt{\frac{\sigma^2}{\alpha}}$, on aboutit donc à

$$\mathbb{P}(|X - \mathbb{E}[X]| > \epsilon_\alpha) \leq \alpha.$$

Ainsi, si l'on souhaite estimer $\theta := \mathbb{E}[X]$ par \overline{X}_n , comme $\sigma_n^2 = \frac{\sigma^2}{n}$, on en déduit que l'on a

$$\mathbb{P}(\theta \in \mathcal{I}_\alpha^{(n)}) \geq 1 - \alpha,$$

$$\text{où } \mathcal{I}_\alpha^{(n)} := \left[\overline{X}_n - \sqrt{\frac{\sigma^2}{n\alpha}}; \overline{X}_n + \sqrt{\frac{\sigma^2}{n\alpha}} \right].$$

Néanmoins, si α est petit, l'intervalle est de longueur $O(\alpha^{-\frac{1}{2}})$. Mais, comme nous le verrons par la suite, dans le cas gaussien par exemple (lequel est vérifié dès que la taille des échantillons est assez élevée de par le théorème central de la limite), le quantile d'ordre $1 - \alpha$ est de la forme $O(\sqrt{-\log(\alpha)})$.

19.3 Cas gaussien

Dans cette section, on suppose que l'on dispose d'un modèle paramétrique gaussien. En d'autres termes, les variables aléatoires sous-jacentes suivent une loi normale de paramètres μ et σ^2 .

L'objectif est ainsi d'estimer par intervalle de confiance la moyenne μ et la variance σ^2 .

19.3.1 Estimation de μ si σ^2 est connue

On commence par estimer la moyenne μ . Pour ce faire, on fait l'hypothèse peu raisonnable (mais qui fournit un cadre simple pour s'entraîner à construire des intervalles de confiance) que la variance σ^2 est connue.

On sait alors, d'après le Théorème 9.4.4 que la variable aléatoire \overline{X}_n suit une loi normale de paramètres μ et $\frac{\sigma^2}{n}$. De fait, la variable aléatoire $Y_n := \frac{\sqrt{n}}{\sigma} (\overline{X}_n - \mu)$ suit la loi normale centrée réduite dont les tables sont à la page 364.

En notant $q_{1-\frac{\alpha}{2}}$ le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite, il vient

$$\begin{aligned} \mathbb{P}(|Y_n| \leq q_{1-\frac{\alpha}{2}}) &= \mathbb{P}(-q_{1-\frac{\alpha}{2}} \leq Y_n \leq q_{1-\frac{\alpha}{2}}) \\ &= \Phi(q_{1-\frac{\alpha}{2}}) - \Phi(-q_{1-\frac{\alpha}{2}}) \\ &= \left(1 - \frac{\alpha}{2}\right) - \left(1 - \left(1 - \frac{\alpha}{2}\right)\right) \\ &= 1 - \alpha. \end{aligned}$$

Ainsi, on a

$$\mathbb{P}\left(-q_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}}{\sigma} (\overline{X}_n - \mu) \leq q_{1-\frac{\alpha}{2}}\right) = 1 - \alpha,$$

ce qui donne

$$\mathbb{P}\left(\mu - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \leq \overline{X}_n \leq \mu + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

On en déduit

$$\mathbb{P}(\mu \in \mathcal{I}_\alpha(n)) = 1 - \alpha,$$

où

$$\mathcal{I}_\alpha(n) := \left[\overline{X}_n - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}; \overline{X}_n + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \right],$$

est l'intervalle de confiance de μ de risque α .

19.3.2 Estimation de μ si σ^2 est inconnue

Dans un cas plus général, on ne dispose pas de la variance. En effet, comment pourrait-on connaître la variance alors qu'on ignore la moyenne ?

Dans ce cas, on estime σ^2 par la variance d'échantillon corrigée à savoir la variable aléatoire $\widehat{S}_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$.

On va maintenant se ramener à une loi connue qui ne dépend ni de σ^2 ni de μ . On sait, pour commencer que la variable aléatoire $Y_n := \frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu)$ suit la loi normale centrée réduite, voir le paragraphe précédent. On sait également que $Z_n := \frac{n-1}{\sigma^2} \widetilde{S}_n^2$ suit la loi du Khi-deux à $n - 1$ degrés de liberté, voir le Théorème 17.6.10. Puis, en admettant que Z_n et Y_n sont indépendantes (ce qui se démontre avec le Théorème 13.7.1), on peut appliquer le Théorème 10.10.3. Il s'ensuit que la variable aléatoire

$$T_n := \frac{Y_n}{\sqrt{\frac{Z_n}{n-1}}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\widetilde{S}_n^2}}$$

suit la loi de Student à $n - 1$ degrés de liberté.

Le nouvel intervalle de confiance est donc

$$\mathcal{I}_\alpha(n) := \left[\bar{X}_n - \frac{\widetilde{S}_n}{\sqrt{n}} t_{1-\frac{\alpha}{2}}; \bar{X}_n + \frac{\widetilde{S}_n}{\sqrt{n}} t_{1-\frac{\alpha}{2}} \right],$$

où $t_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de Student à $n - 1$ degrés de liberté.

19.3.3 Estimation de σ^2 si μ est connue

Ici, on cherche à estimer σ^2 . On suppose que l'on connaît la moyenne μ .

On sait alors que les variables aléatoires $\frac{X_i - \mu}{\sigma}$ suivent une loi normale centrée réduite. De plus, elles sont mutuellement indépendantes. De fait,

$$\chi^2 := \sum_{k=1}^n \left(\frac{X_k - \mu}{\sigma} \right)^2$$

suit une loi du Khi-deux à n degrés de liberté. Par conséquent,

$$\mathbb{P}(k_1 \leq \chi^2 \leq k_2) = \left(1 - \frac{\alpha}{2}\right) - \frac{\alpha}{2} = 1 - \alpha,$$

où k_1 et k_2 sont respectivement les quantiles d'ordre $1 - \frac{\alpha}{2}$ et $\frac{\alpha}{2}$ de la loi du khi-deux à n degrés de liberté.

On en déduit

$$\mathbb{P}(\sigma^2 \in \mathcal{I}_\alpha) = 1 - \alpha,$$

où l'intervalle de confiance de la variance \mathcal{I}_α au risque α est

$$\mathcal{I}_\alpha := \left[\frac{1}{k_2} \sum_{i=1}^n (X_i - \mu)^2; \frac{1}{k_1} \sum_{i=1}^n (X_i - \mu)^2 \right].$$

Remarque 19.3.1. *Il peut sembler curieux à première lecture que l'on ne divise pas par n et on pourrait croire que l'intervalle de confiance ne devient pas meilleur pour peu que n augmente. Il faut en fait se rappeler que k_1 comme k_2 dépendent de n vu qu'ils sont des quantiles de la loi du khi-deux à n degrés de liberté.*

19.3.4 Estimation de σ^2 si μ est inconnue

On suppose maintenant que la moyenne est également inconnue. Il est naturel de remplacer cette moyenne μ par son estimation ponctuelle \overline{X}_n . Néanmoins, on perd alors un degré de liberté d'après le Théorème de Cochran, voir page 198.

Ainsi, un intervalle de confiance de la variance \mathcal{I}_α au risque α est désormais :

$$\mathcal{I}_\alpha := \left[\frac{1}{k'_2} \sum_{i=1}^n (X_i - \overline{X}_n)^2 ; \frac{1}{k'_1} \sum_{i=1}^n (X_i - \overline{X}_n)^2 \right],$$

où k'_1 et k'_2 sont respectivement les quantiles d'ordre $1 - \frac{\alpha}{2}$ et $\frac{\alpha}{2}$ de la loi du khi-deux à $n - 1$ degrés de liberté.

19.3.5 Différence de deux moyennes

Dans ce paragraphe, on suppose que l'on dispose de deux populations et donc de deux suites de variables aléatoires gaussiennes. $(X_n)_n$ est une suite de variables aléatoires indépendantes et identiquement distribuées suivant la loi normale $\mathcal{N}(\mu_1, \sigma_1^2)$. Et, $(Y_n)_n$ est une suite de variables aléatoires indépendantes et identiquement distribuées suivant la loi normale $\mathcal{N}(\mu_2, \sigma_2^2)$. On suppose de plus que les deux suites sont indépendantes.

On cherche à estimer $\mu_1 - \mu_2$. Pour ce faire, on pose $\overline{X}_N := \frac{1}{N} \sum_{n=1}^N X_n$ et $\overline{Y}_M := \frac{1}{M} \sum_{m=1}^M Y_m$. Enfin, on introduit l'estimateur sans biais de $\mu_1 - \mu_2$:

$$\overline{D}_{N,M} := \overline{X}_N - \overline{Y}_M.$$

19.3.5.1 Si σ_1^2 et σ_2^2 sont connues

On suppose pour commencer que σ_1^2 et σ_2^2 sont connues.

Alors, la variable aléatoire $\overline{D}_{N,M}$ suit la loi normale de paramètres $\mu_1 - \mu_2$ et $\frac{\sigma_1^2}{N} + \frac{\sigma_2^2}{M}$. On a donc immédiatement un intervalle de confiance au seuil de risque α pour $\mu_1 - \mu_2$:

$$\mathcal{I}_\alpha := \left[\overline{D}_{N,M} - q_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{N} + \frac{\sigma_2^2}{M}} ; \overline{D}_{N,M} + q_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{N} + \frac{\sigma_2^2}{M}} \right].$$

19.3.5.2 Si σ_1^2 et σ_2^2 sont supposées égales

On suppose dorénavant que l'on ne connaît pas les deux variances σ_1^2 et σ_2^2 . Néanmoins, on suppose l'hypothèse d'homoscédasticité, c'est-à-dire que $\sigma_1^2 = \sigma_2^2$.

Dans ce cas,

$$\widetilde{S}_{X,N}^2 := \frac{1}{N-1} \sum_{n=1}^N (X_n - \overline{X}_N)^2$$

et

$$\widetilde{S}_{Y,M}^2 := \frac{1}{M-1} \sum_{m=1}^M (Y_m - \overline{Y}_M)^2$$

sont des estimateurs de $\sigma^2 := \sigma_1^2 = \sigma_2^2$.

On pose alors

$$S_{(X,Y),(N,M)}^2 := \frac{(N-1)\widetilde{S}_{X,N}^2 + (M-1)\widetilde{S}_{Y,M}^2}{N+M-2}.$$

Or, $\frac{\widetilde{S}_{X,N}^2}{\sigma^2}$ suit une loi du khi-deux à $N-1$ degrés de liberté. De même, $\frac{\widetilde{S}_{Y,M}^2}{\sigma^2}$ suit une loi du khi-deux à $M-1$ degrés de liberté. De fait, la variable aléatoire $(N+M-2)S_{(X,Y),(N,M)}^2$ suit une loi du khi-deux à $N+M-2$ degrés de liberté.

On peut ainsi utiliser la loi de Student comme précédemment. Il vient qu'un intervalle de confiance de $\mu_1 - \mu_2$ au risque α est

$$\left[\overline{D}_{N,M} - t_{1-\frac{\alpha}{2}} \sqrt{\frac{1}{N} + \frac{1}{M}} S_{(X,Y),(N,M)}; \overline{D}_{N,M} + t_{1-\frac{\alpha}{2}} \sqrt{\frac{1}{N} + \frac{1}{M}} S_{(X,Y),(N,M)} \right],$$

où $t_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de Student à $N+M-2$ degrés de liberté.

19.4 Intervalles de confiance asymptotiques

Dans le paragraphe précédent, on a établi des intervalles de confiance lorsque les variances sont supposées égales ou même lorsqu'elles sont connues. On peut se demander ce qu'il se passe lorsqu'elles ne sont ni connues ni supposées égales. Néanmoins, dans ce cas de figure, l'intervalle de confiance sera asymptotique, c'est-à-dire que la probabilité pour que $\mu_1 - \mu_2$ soit dans ledit intervalle tend vers $1 - \alpha$ quand N et M tendent vers l'infini. Nous avons choisi de ne pas détailler ce cas.

À la place, nous allons présenter les intervalles de confiance asymptotiques dans des cas simples. D'abord, nous étudierons le cas de la loi de Poisson puis nous nous intéresserons à la loi de Bernoulli pour la proportion d'échantillon.

19.4.1 Loi de Poisson

Supposons que les variables aléatoires soient indépendantes et identiquement distribuées suivant une loi de Poisson de paramètre $\theta > 0$ inconnu. On sait, d'après le Théorème central de la limite que $Y_n := \sqrt{n}(\overline{X}_n - \theta)$ converge en loi vers la loi normale centrée et de variance $\theta > 0$.

Or, \overline{X}_n estime le paramètre de variance θ . En effet, \overline{X}_n converge presque sûrement vers θ d'après la loi forte des grands nombres. Puis, cela implique que $\frac{1}{\sqrt{\overline{X}_n}}$ converge presque sûrement et donc en loi vers $\frac{1}{\sqrt{\theta}}$. Puis, comme Y_n converge en loi vers la loi normale centrée de variance θ , il s'ensuit que $\frac{Y_n}{\sqrt{\overline{X}_n}}$ converge en loi vers la gaussienne centrée réduite, d'après le lemme de Slutsky, voir page 210.

On obtient ainsi un intervalle de confiance asymptotique :

$$\mathcal{I}_\alpha := \left[\overline{X}_n - q_{1-\frac{\alpha}{2}} \frac{\sqrt{\overline{X}_n}}{\sqrt{n}}; \overline{X}_n + q_{1-\frac{\alpha}{2}} \frac{\sqrt{\overline{X}_n}}{\sqrt{n}} \right],$$

où $q_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite.

19.4.2 Loi de Bernoulli

On s'intéresse ici à la proportion d'échantillon. En d'autres termes, les variables aléatoires $(X_n)_n$ sont indépendantes et identiquement distribuées suivant la loi de Bernoulli de paramètre p .

Alors, $n\overline{X}_n$ suit une loi binomiale de paramètres n et p donc l'espérance de \overline{X}_n est p tandis que sa variance est $\frac{p(1-p)}{n}$.

En utilisant le Théorème central de la limite, on aboutit à l'intervalle de confiance asymptotique suivant au seuil de risque α :

$$\mathcal{I}_\alpha := \left[\overline{X}_n - q_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}; \overline{X}_n + q_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}} \right],$$

où $q_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite.

Cet intervalle est dû à la méthode de Wald. Il existe d'autres méthodes (lesquelles présentent certains avantages) comme la méthode du score de Wilson. On peut également appliquer ce que l'on appelle la correction de continuité de Yates.

Néanmoins, dans cet ouvrage, nous ne présentons que les bases des statistiques. Il faut donc simplement être conscient que cela existe.

19.5 Cas exponentiel

Ici, on présente le cas où les variables aléatoires sous-jacentes suivent la loi exponentielle de paramètre $\theta > 0$.

On a donc $\mathbb{E}[X_i] = \frac{1}{\theta}$ et $\text{Var}[X_i] = \frac{1}{\theta^2}$. Ainsi :

$$\mathbb{E}[\bar{X}_n] = \frac{1}{\theta} \quad \text{et} \quad \text{Var}[\bar{X}_n] = \frac{1}{n\theta^2}.$$

On va ici chercher un intervalle de confiance pour θ .

19.5.1 Intervalles asymptotiques

On commence avec un intervalle asymptotique. On applique le théorème central de la limite et l'on aboutit à la convergence en loi de $\frac{\sqrt{n}(\bar{X}_n - \frac{1}{\theta})}{\frac{1}{\theta}} = \sqrt{n}(\theta\bar{X}_n - 1)$ vers la loi normale centrée réduite.

De fait, un intervalle de confiance asymptotique immédiat de θ est

$$\mathcal{I}_\alpha := \left[\frac{1}{\bar{X}_n} - \frac{1}{\sqrt{n}} \frac{q_{1-\frac{\alpha}{2}}}{\bar{X}_n}; \frac{1}{\bar{X}_n} + \frac{1}{\sqrt{n}} \frac{q_{1-\frac{\alpha}{2}}}{\bar{X}_n} \right].$$

Néanmoins, il est essentiel de bien comprendre que cet intervalle n'a d'intérêt que pour des grandes valeurs de n .

19.5.2 Intervalles non asymptotiques

On rappelle que la somme de n variables aléatoires indépendantes et suivant une même loi exponentielle suit la loi d'Erlang (voir la Section 10.4), laquelle est une instance particulière de la loi Gamma (voir la Section 10.6).

Ainsi, $n\theta\bar{X}_n$ suit la loi d'Erlang de paramètres n et 1.

Il s'ensuit qu'un intervalle de confiance (non asymptotique) du paramètre θ est

$$\mathcal{I}_\alpha := \left[\frac{q_{\frac{\alpha}{2}}}{n\bar{X}_n}; \frac{q_{1-\frac{\alpha}{2}}}{n\bar{X}_n} \right],$$

où $q_{\frac{\alpha}{2}}$ et $q_{1-\frac{\alpha}{2}}$ sont les quantiles d'ordre respectifs $\frac{\alpha}{2}$ et $1 - \frac{\alpha}{2}$ de la loi d'Erlang de paramètres n et 1.

19.6 Pour aller plus loin

On suppose que les variables aléatoires sous-jacentes $(X_n)_{n \in \mathbb{N}^*}$ sont indépendantes et identiquement distribuées de loi $\mathcal{L}(X)$.

Pour tout $\lambda \in \mathbb{R}$, on introduit la fonction de log-Laplace suivante :

$$\Lambda(\lambda) := \log [\mathbb{E} (e^{\lambda X})] ,$$

avec la convention $\Lambda(\lambda) := +\infty$ si l'espérance n'est pas définie.

Ensuite, pour tout $x \in \mathbb{R}$, on introduit la transformée de Legendre de la fonction Λ :

$$\Lambda^*(x) := \sup_{\lambda \in \mathbb{R}} (\lambda x - \Lambda(\lambda)) .$$

On peut alors montrer que pour tout $\epsilon > 0$, on a :

$$\mathbb{P} (|\overline{X}_n - \mu| > \epsilon) \leq \exp(-n\Lambda^*(\mu + \epsilon)) + \exp(-n\Lambda^*(\mu - \epsilon)) .$$

En effet, pour $\lambda > 0$:

$$\begin{aligned} \mathbb{P} (\overline{X}_n - \mu > \epsilon) &= \mathbb{P} (e^{\lambda(\overline{X}_n - \mu)} > e^{\lambda\epsilon}) \\ &\leq \frac{\mathbb{E} [e^{\lambda(\overline{X}_n - \mu)}]}{e^{\lambda\epsilon}} \\ &\leq \mathbb{E} \left[e^{\frac{\lambda}{n} \sum_{k=1}^n (X_k - \mu)} \right] e^{-\lambda\epsilon} \\ &\leq \exp \left\{ -n \left(\frac{\lambda}{n} (\mu + \epsilon) - \Lambda \left(\frac{\lambda}{n} \right) \right) \lambda \right\} . \end{aligned}$$

On peut ensuite vérifier que pour $x > \mu$, alors $\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}_+} (\lambda x - \Lambda(\lambda))$. Ainsi, en prenant l'infimum pour $\frac{\lambda}{n}$ sur \mathbb{R}_+ , on a

$$\mathbb{P} (\overline{X}_n - \mu > \epsilon) \leq \exp \{-n\Lambda^*(\mu + \epsilon)\} .$$

On peut prouver de même que l'on a

$$\mathbb{P} (\overline{X}_n - \mu < -\epsilon) \leq \exp \{-n\Lambda^*(\mu - \epsilon)\} ,$$

ce qui achève la preuve.

Par conséquent, pour avoir un intervalle de confiance au seuil de risque α , il "suffit" de résoudre l'équation $\exp(-n\Lambda^*(\mu + \epsilon)) + \exp(-n\Lambda^*(\mu - \epsilon)) = \alpha$ en ϵ et l'on a un intervalle de confiance.

Néanmoins, la résolution n'est pas immédiate car la fonction Λ^* n'a aucune raison d'être simple et de plus, cette fonction dépend *a priori* des différents paramètres de la loi.

Par exemple, pour une loi de Poisson de paramètre θ , on a $\Lambda^*(x) = x \log \left(\frac{x}{\theta} \right) + \theta - x$ pour $x > 0$. Et, comme $\mu = \theta$, il vient :

$$\Lambda^*(\theta - \epsilon) = \epsilon + (\theta - \epsilon) \log \left(1 - \frac{\epsilon}{\theta}\right) = \frac{\epsilon^2}{\theta} + o(\epsilon^2),$$

et

$$\Lambda^*(\theta + \epsilon) = -\epsilon + (\theta + \epsilon) \log \left(1 + \frac{\epsilon}{\theta}\right) = \frac{\epsilon^2}{\theta} + o(\epsilon^2).$$

Ainsi, pour $\epsilon > 0$ petit, on a

$$\mathbb{P}(|\overline{X}_n - \mu| > \epsilon) \leq 2 \exp \left(-n \frac{\epsilon^2}{\theta}\right).$$

On en déduit que l'intervalle de confiance s'obtient avec $\epsilon = \sqrt{-\frac{\theta}{n} \log \left(\frac{\alpha}{2}\right)}$. Malheureusement, il dépend de θ .

Restreignons-nous au cas où la loi est à support compact. C'est notamment vrai pour des lois de Bernoulli de paramètre p . Dans ce cas, $\Lambda(\lambda) = \log((1-p) + pe^\lambda)$ et donc $\Lambda^*(x) = x \log \left(\frac{x}{p}\right) + (1-x) \log \left(\frac{1-x}{1-p}\right)$. Ici, $\mu = p$. Il vient donc

$$\Lambda^*(p + \epsilon) = \frac{\epsilon^2}{p} + \frac{\epsilon^2}{1-p} + o(\epsilon^2),$$

ainsi que

$$\Lambda^*(p - \epsilon) = \frac{\epsilon^2}{p} + \frac{\epsilon^2}{1-p} + o(\epsilon^2).$$

En supposant que l'on sait $p \in [\alpha; \beta]$ où $0 < \alpha < \beta < 1$, il s'ensuit la borne suivante :

$$\mathbb{P}(|\overline{X}_n - \mu| > \epsilon) \leq 2 \exp \{-n\epsilon^2\gamma\},$$

où $\gamma := \frac{1}{\beta} + \frac{1}{1-\alpha}$. De fait, on prendra ici $\epsilon := \sqrt{-\frac{1}{n\gamma} \log \left(\frac{\alpha}{2}\right)}$; ce qui ne dépend pas de p .

De manière générale, la fonction Λ^* est convexe et l'argument de son minimum global est atteint en $\mu = \mathbb{E}[X]$. De fait, on aura **toujours** un résultat de la forme $O\left(\sqrt{-\frac{\log(\alpha)}{n}}\right)$.

Généralités sur les tests statistiques

20.1 Introduction

Dans le chapitre suivant (voir page 307), nous étudierons différents tests statistiques.

Avant cela, nous allons introduire le lecteur à des principes généraux sur les tests d'hypothèse. Nous avons opté pour l'approche suivante : on commence par donner des exemples pour motiver l'utilisation de ces tests et justifier de leur importance pratique (et théorique), ensuite on présente les différents types de tests. Suite à cela, on donne l'algorithme général de mise en place d'un test statistique. Enfin, on parlera de l'erreur de première espèce ainsi que de l'erreur de seconde espèce.

Les pré-requis sont une compréhension fine de la probabilité conditionnelle.

L'objectif majeur de ce chapitre est de comprendre les différentes étapes qui régissent l'algorithme d'un test statistique. Notamment, on attend du lecteur qu'à l'issue du chapitre celui-ci soit en mesure de reproduire les grandes étapes d'un test d'hypothèse.

20.1.1 Motivations

En médecine générale comme dans diverses spécialités médicales, quand le patient est malade, l'une des premières choses que fait le médecin est de prescrire un test sanguin. Après la déplaisante piqûre effectuée par l'infirmière, l'échantillon de sang va dans un laboratoire médical où certaines quantités sont dénombrées.

Pour savoir si le patient a des soucis au niveau de la thyroïde par exemple, la quantité d'hormone TSH est comptée. Si elle est trop élevée, ce n'est pas bon. Si elle est trop basse, ce n'est pas bon non plus. Avec les reins, c'est la créatinine qui doit être testée.

Nanmoins, comment décider de ce qui est trop élevé et de ce qui ne l'est pas ? On pourrait faire une moyenne et décider que tout ce qui n'est pas pile dans la moyenne est mauvais. Mais, ce serait négliger les fluctuations d'échantillonnage

et ce serait partir du principe que les lois de l'arithmétique et du nombre exact président à la destinée des êtres vivants. Cette hypothèse est peu convaincante. Pourtant, il y a bien une région critique dans laquelle on considère que le patient souffre d'hypothyroïdie...

Dans les chapitres précédents, on utilise parfois le théorème central limite sans faire tendre la taille de l'échantillon vers l'infini. Plus précisément, on considère qu'au-delà de trente, on est dans le cadre gaussien. Pourtant, on peut très facilement imaginer une loi de probabilité construite afin qu'il faille au moins cinq cents variables aléatoires avant que la moyenne d'échantillon ne vérifie le modèle gaussien avec une bonne précision. Et même si l'on a suffisamment d'échantillons, l'expression "avec une bonne précision" est mathématiquement floue. Il convient alors de décider si oui ou non on est bien dans le cadre gaussien.

Il est naturel de dire que le poids et la taille ne sont pas des caractères indépendants dans l'ensemble de la population humaine. En effet, plus l'individu est grand, plus il a de chance d'être lourd, *a priori* en tout cas. Mais comment le prouver sérieusement ? On pourrait appliquer le Théorème 16.6.13 sur un échantillon mais alors rien ne dit que les fluctuations propres à chaque individu ne sont pas responsables de la non-validité de la formule. Il est donc important de faire un choix en se fondant sur une statistique précise.

20.1.2 Définition

Un test d'hypothèse est un algorithme qui permet de décider entre deux hypothèses. Pour ce faire, le test d'hypothèse analyse une statistique élaborée à partir d'un échantillon aléatoire, tout en quantifiant le risque d'erreur. Ou plutôt : les risques d'erreurs, voir Section 20.3 et Section 20.4.

20.1.3 Deux familles de tests

Il y a deux grandes familles de tests statistiques : les tests paramétriques et les tests non paramétriques.

20.1.3.1 Tests paramétriques

Les tests paramétriques présupposent que l'on connaît le genre de loi que suivent les variables aléatoires sous-jacentes : Bernoulli $\mathcal{B}(p)$, Poisson $\mathcal{P}(\lambda)$, normale $\mathcal{N}(m, \sigma^2)$...

Ces hypothèses sont généralement difficiles à vérifier (et nécessitent donc un test non paramétrique pour commencer). Il convient de noter que les tests paramétriques sont puissants. Ainsi, dans le cas des grands échantillons ou si l'on est exactement dans le modèle gaussien, il est facile de décider si oui ou non les paramètres proposés sont les bons.

20.1.3.2 Tests non paramétriques

Les tests non paramétriques ne présupposent rien. Il n'y a en particulier aucune hypothèse de normalité à satisfaire au préalable. Toutefois, leur puissance est plus modérée.

De fait, on les emploie quand on ne peut faire d'autres tests. Leur avantage est de pouvoir être utilisés même avec des échantillons de faible taille.

20.1.4 Quels tests peut-on effectuer ?

Dans ce paragraphe, on donne quatre types de tests, tous classiques.

20.1.4.1 Test de conformité

Pour commencer, on veut savoir si une valeur pré-établie du paramètre de la loi sous-jacente est bonne ou non. On confronte alors la valeur théorique de la statistique liée à cette valeur pré-établie à la valeur pratique de la statistique prise sur un échantillon.

Notamment, on peut effectuer des tests sur la moyenne, la variance, la proportion...

20.1.4.2 Test d'ajustement

Ici, on vérifie si la distribution empirique dans l'échantillon est représentée de façon crédible ou non par une distribution pré-établie. En particulier, on peut effectuer un test de normalité pour savoir si la distribution respecte bien l'hypothèse de normalité. Suite à cela, un test paramétrique peut être appliqué.

Il convient de noter que la loi sous-jacente peut être continue ou discrète.

On parle aussi de test d'adéquation.

20.1.4.3 Test d'homogénéité

Dans un test d'homogénéité, on essaie de vérifier si oui ou non divers échantillons sont issus d'une même population et donc si l'hypothèse d'homogénéité est vérifiée. Dit autrement, on cherche à vérifier si la loi sous-jacente des différentes variables aléatoires est commune aux différents échantillons.

On parle aussi de test de comparaison.

20.1.4.4 Test d'indépendance

Le dernier type de test que nous présentons consiste à décider si deux variables aléatoires sous-jacentes sont indépendantes dans l'échantillon considéré.

Dit autrement, les deux caractères sont-ils indépendants (la non-validité du Théorème 16.6.13 vient en fait des fluctuations d'échantillonnage) ou non ?

On parle aussi de test d'association.

20.2 Principe général

L'algorithme régissant la mise en place d'un test d'hypothèse est une succession d'étapes simples, précises et chacune de ces étapes est d'une grande importance. La liste des étapes est la suivante :

1. Formuler l'hypothèse nulle (H_0).
2. Définir l'hypothèse alternative (H_1).
3. Contrôler avec une statistique liée à (H_0) (et à (H_1) de fait).
4. Établir la distribution de ladite statistique.
5. Choisir le niveau de signification du test, α .
6. Déterminer la région critique associée.
7. Calculer la valeur empirique de la statistique (dans l'échantillon).
8. Décider en fonction des étapes **6.** et **7.**.

20.2.1 Formuler (H_0)

La première étape consiste à définir l'hypothèse nulle, que l'on note (H_0). Il s'agit de l'hypothèse que l'on cherche à contrôler. Cela peut être une conformité à la loi normale par exemple. L'hypothèse nulle consiste à dire qu'il n'y a pas suffisamment de différences entre les valeurs théoriques et les valeurs empiriques pour affirmer que l'hypothèse que l'on a formulée est fausse.

Il est important de comprendre que l'objectif du test d'hypothèse n'est pas d'accepter (H_0) mais plutôt de ne pas la rejeter ; ou le cas contraire de la rejeter.

20.2.2 Définir (H_1)

Une fois formulée l'hypothèse nulle (H_0), une manière classique de procéder est de prendre comme hypothèse alternative (H_1) le complémentaire de (H_0). Cependant, cette dernière n'est pas obligatoirement le complémentaire de (H_0).

Par exemple, si l'on fait un test sur une proportion et que les deux seules proportions plausibles sont $p := \frac{2}{7}$ et $p := \frac{4}{7}$, alors si (H_0) correspond à $p = \frac{2}{7}$, il s'ensuit que (H_1) n'est pas $p \neq \frac{2}{7}$ mais $p = \frac{4}{7}$.

Par suite, si l'on rejette (H_0), on accepte (H_1).

20.2.3 Contrôler avec une statistique

Pour que les tests statistiques aient un sens, il convient que ces tests soient fondés sur le calcul des probabilités et sur les théorèmes des statistiques. C'est pourquoi, après que (H_0) et (H_1) ont été formulées, on va devoir contrôler ces deux hypothèses au moyen d'une statistique S à partir de l'échantillon considéré.

Par exemple, pour un test du χ^2 d'ajustement, voir la page 319, on considère

$$\chi^2 := \sum_{i=1}^r \frac{(N_i - np_i)^2}{np_i}.$$

La plupart du temps, dans les cas concrets, on se ramène à des tests connus et cette étape consiste à identifier le bon test (et donc la bonne statistique à utiliser).

20.2.4 Établir la distribution de la statistique de contrôle

On dispose de deux hypothèses principales : (H_0) et (H_1) ; ces dernières étant éventuellement complémentaires. Selon ces hypothèses, la loi de la statistique va différer.

Dans cette étape, on doit établir la loi de variabilité de la statistique de contrôle conditionnellement à (H_0) . Et, si possible on fait de même conditionnellement à l'hypothèse alternative (H_1) ; bien que dans le cas d'un test sur la moyenne, la loi forte des grands nombres nous permet de nous affranchir de cette possibilité.

20.2.5 Choisir α

Un test d'hypothèse amène nécessairement à des risques d'erreur. L'on aimerait que les deux erreurs qui peuvent survenir (celle de première espèce et celle de seconde espèce) soient aussi petites que possibles.

L'erreur qui nous intéresse le plus est la première à savoir la probabilité de rejeter (H_0) conditionnellement au fait que (H_0) soit vrai. Cette erreur, que l'on quantifie par α est celle que l'on va contrôler.

On se fixe donc $\alpha > 0$ en fonction de l'étude que l'on fait avec le test statistique en question. Généralement, on prend $\alpha = 0.1\%$ ou 1% ou 2% voire 5% .

20.2.6 Déterminer la région critique associée

Par définition de la région critique associée, on doit trouver le seuil critique $q_c(\alpha)$ tel que

$$\mathbb{P}_{(H_0)}(S > q_c(\alpha)) = \alpha,$$

si l'hypothèse (H_0) est de la forme $p = p_0$ et (H_1) de la forme $p > p_0$. Si jamais (H_1) est de la forme $p < p_0$, alors $q_c(\alpha)$ satisfait

$$\mathbb{P}_{(H_0)}(S < q_c(\alpha)) = \alpha.$$

Enfin, si (H_1) est de la forme $p \neq p_0$, $q_c(\alpha)$ satisfait

$$\mathbb{P}_{(H_0)}(|S| > q_c(\alpha)) = \alpha.$$

La région critique avec laquelle on rejette l'hypothèse (H_0) sera donc de la forme $[q_c(\alpha); +\infty[$, $] - \infty; q_c(\alpha)]$ ou $] - q_c(\alpha); q_c(\alpha)[^c$.

20.2.7 Calculer la valeur empirique de la statistique

On calcule la valeur de la statistique $S(X_1, \dots, X_n)$ dans l'échantillon. On obtient une valeur $s(x[1], \dots, x[n])$ où $x[1], \dots, x[n]$ sont les réalisations dans l'échantillon.

20.2.8 Décider

Si $s(x[1], \dots, x[n])$ est dans la région critique, on rejette (H_0) donc on accepte l'hypothèse alternative (H_1) . Si la valeur empirique n'est pas dans la région critique, on ne rejette pas (H_0) .

Il est essentiel de comprendre que l'on n'accepte pas (H_0) . On ne la rejette simplement pas.

20.3 Erreur de première espèce

L'erreur de première espèce consiste à rejeter (H_0) alors que (H_0) est vraie. On contrôle cette erreur par α . Il est essentiel de bien comprendre que le paramètre α n'est pas un paramètre de précision. Ainsi, au premier abord, on pourrait penser que plus α est petit, moins on se trompe. Ceci n'est pas exact. Plus α est petit et moins l'on rejettera (H_0) .

20.4 Erreur de seconde espèce

L'erreur de seconde espèce consiste à ne pas rejeter (H_0) alors que (H_1) est vraie. Cette erreur est notée β . On n'a pas de prise sur β . On peut éventuellement calculer cette valeur si l'on dispose de statistiques simples.

La quantité $1 - \beta$ est appelée la puissance du test. Ainsi, les tests paramétriques ont une erreur de seconde espèce plus faible que celle des tests non paramétriques.

Il est crucial de bien comprendre qu'*a priori*, α et β n'ont rien à voir. En effet,

$$\alpha := \mathbb{P}_{(H_0)}(\text{rejeter } (H_0)) = \frac{\mathbb{P}((H_0), \text{rejeter } (H_0))}{\mathbb{P}((H_0))},$$

tandis que

$$\beta := \mathbb{P}_{(H_0)^c}(\text{ne pas rejeter } (H_0)) = \frac{\mathbb{P}((H_0)^c, \text{ ne pas rejeter } (H_0))}{\mathbb{P}((H_0)^c)}$$

dans le cas où $(H_1) = (H_0)^c$.

Tests d'hypothèse classiques

21.1 Introduction

Dans ce chapitre, on présente quelques tests d'hypothèse. En particulier, nous verrons des tests sur la moyenne, sur la variance, sur la fréquence d'apparition d'un caractère. Également, nous présenterons des tests de comparaison (entre moyennes, entre variances, entre proportions) ainsi que le très classique test du χ^2 d'ajustement dans différents cas. Nous donnerons également les clés pour mener un test d'indépendance et un test d'homogénéité. Enfin, nous aborderons brièvement le test de Kolmogorov-Smirnov.

Dans les Sections 2 à 7, nous supposons que les lois sous-jacentes sont des gaussiennes. Et, si les lois sous-jacentes ne sont pas gaussiennes, on appliquera le théorème central de la limite pour s'y ramener (cas des grands échantillons).

Tous les chapitres précédents sont à étudier avant d'attaquer celui-ci. En particulier, on invite le lecteur à connaître par cœur les différentes étapes qui régissent tout test statistique avant de s'attaquer au présent chapitre.

Les premières sections du chapitre sont dans un style "catalogue". De fait, plutôt que de retenir par cœur les résultats, nous insistons pour que le lecteur soit en mesure de les retrouver à partir des calculs ou tout du moins de savoir où ils se situent dans le livre dans un premier temps. Il est également essentiel qu'à l'issue de ce chapitre, le lecteur connaisse les hypothèses permettant d'appliquer un test du khi-deux d'ajustement pour une loi discrète. Également, il faudra connaître le nombre de degrés de liberté, au moins dans le cas où on n'utilise pas d'estimateur. Il est aussi crucial que le lecteur sache que l'utilisation d'un estimateur abaisse le nombre de degrés de liberté mais aussi qu'il soit au courant que la loi du Khi-deux peut être utilisée pour des tests d'indépendance ou d'homogénéité. Enfin, un attendu extrêmement important est de savoir qu'il existe un très grand nombre d'autres tests que ceux donnés ici. Outre les tests du Khi-deux ou de Kolmogorov-Smirnov, on peut citer sans être exhaustif : le test Anova, le test de Shapiro-Wilks, le test de Lilliefors, le test d'Anderson-Darling, le test de D'Agostino, le test de Jarque-Bera, le test de Ansari - Bradley, le test de Klotz, le test de Mood, le test de Siegel-Tukey, le test Rho de Spearman et bien d'autres...

21.2 Tests sur la moyenne μ

On suppose ici que la loi sous-jacente est $\mathcal{N}(\mu, \sigma^2)$. On suit le principe général énoncé à la Section 20.2.

Étape 1. L'hypothèse (H_0) est la suivante : $\mu = \mu_0$ où μ_0 est un réel dont on pense qu'il est la moyenne.

Étape 2. L'hypothèse (H_1) peut correspondre à un test bilatéral à savoir $\mu \neq \mu_0$ comme il peut s'agir d'un test unilatéral : $\mu > \mu_0$ (ou au contraire $\mu < \mu_0$). On voit ainsi que l'on dispose de trois hypothèses alternatives. Notons-les respectivement (H_1^b), (H_1^+) et (H_1^-). Les trois cas seront traités par la suite.

Étape 3. La statistique naturelle pour approximer la moyenne est la moyenne d'échantillon \overline{X}_n puisqu'il s'agit d'un estimateur sans biais et convergent de μ . On note toutefois que la loi de \overline{X}_n dépend de σ^2 . En effet, la moyenne d'échantillon suit la loi normale de paramètres μ et $\frac{\sigma^2}{n}$. Ainsi, deux cas peuvent se produire : on connaît σ^2 ou on ne connaît pas σ^2 .

21.2.1 En supposant la variance σ^2 connue

Étape 4. On suppose ici que la variance σ^2 est connue. Alors, sous l'hypothèse (H_0), la variable aléatoire $Z_n := \frac{\overline{X}_n - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$ suit la loi normale centrée réduite. On note que Z_n ne dépend d'aucun paramètre inconnu.

Étape 5. On se donne ensuite un niveau de signification (erreur de première espèce) α .

21.2.1.1 Test bilatéral : (H_1^b)

Étape 6. La région critique consiste ici à résoudre

$$\mathbb{P}(|Z_n| > q_c(\alpha)) = \alpha.$$

On trouve $q_c(\alpha) := q_{1-\frac{\alpha}{2}}$, le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^b :=]-\infty; -q_{1-\frac{\alpha}{2}}] \cup [q_{1-\frac{\alpha}{2}}; +\infty[.$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^b$, on rejette l'hypothèse (H_0) et on accepte (H_1^b); sinon on ne la rejette pas.

21.2.1.2 Test unilatéral : (H_1^+)

Étape 6. La région critique consiste ici à résoudre

$$\mathbb{P}(Z_n > q_c(\alpha)) = \alpha.$$

On trouve $q_c(\alpha) := q_{1-\alpha}$, le quantile d'ordre $1 - \alpha$ de la loi normale centrée réduite. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^+ := [q_{1-\alpha}; +\infty[.$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^+$, on rejette l'hypothèse (H_0) et on accepte (H_1^+) ; sinon on ne la rejette pas.

21.2.1.3 Test unilatéral : (H_1^-)

Étape 6. La région critique consiste ici à résoudre

$$\mathbb{P}(Z_n < q_c(\alpha)) = \alpha.$$

On trouve $q_c(\alpha) := q_\alpha$, le quantile d'ordre α de la loi normale centrée réduite. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^- :=]-\infty; q_\alpha].$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^-$, on rejette l'hypothèse (H_0) et on accepte (H_1^-) ; sinon on ne la rejette pas.

21.2.2 En supposant la variance σ^2 inconnue

Étape 4. On suppose ici que la variance σ^2 est inconnue. On va alors l'estimer par $\widehat{S}_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Comme on l'a déjà vu, la variable $\frac{\bar{X}_n - \mu}{\sqrt{\frac{\widehat{S}_n^2}{n}}}$ suit la loi de Student à $n - 1$ degrés de liberté. Alors, sous l'hypothèse (H_0) , la variable aléatoire $Z_n := \frac{\bar{X}_n - \mu_0}{\sqrt{\frac{\widehat{S}_n^2}{n}}}$ suit la loi de Student à $n - 1$ degrés de liberté. On note que Z_n ne dépend d'aucun paramètre inconnu.

Étape 5. On se donne ensuite un niveau de signification (erreur de première espèce) α .

21.2.2.1 Test bilatéral : (H_1^b)

Étape 6. La région critique consiste ici à résoudre

$$\mathbb{P}(|Z_n| > q_c(\alpha)) = \alpha.$$

On trouve $q_c(\alpha) := t_{1-\frac{\alpha}{2}}$, le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de Student à $n - 1$ degrés de liberté. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^b :=]-\infty; -t_{1-\frac{\alpha}{2}}] \cup [t_{1-\frac{\alpha}{2}}; +\infty[.$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^b$, on rejette l'hypothèse (H_0) et on accepte (H_1^b) ; sinon on ne la rejette pas.

21.2.2.2 Test unilatéral : (H_1^+)

Étape 6. La région critique consiste ici à résoudre

$$\mathbb{P}(Z_n > q_c(\alpha)) = \alpha.$$

On trouve $q_c(\alpha) := t_{1-\alpha}$, le quantile d'ordre $1 - \alpha$ de la loi de Student à $n - 1$ degrés de liberté. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^+ := [t_{1-\alpha}; +\infty[.$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^+$, on rejette l'hypothèse (H_0) et on accepte (H_1^+) ; sinon on ne la rejette pas.

21.2.2.3 Test unilatéral : (H_1^-)

Étape 6. La région critique consiste ici à résoudre

$$\mathbb{P}(Z_n < q_c(\alpha)) = \alpha.$$

On trouve $q_c(\alpha) := t_\alpha$, le quantile d'ordre α de la loi de Student à $n - 1$ degrés de liberté. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^- :=]-\infty; t_\alpha].$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^-$, on rejette l'hypothèse (H_0) et on accepte (H_1^-) ; sinon on ne la rejette pas.

21.3 Tests sur la variance σ^2

On suppose ici que la loi sous-jacente est $\mathcal{N}(\mu, \sigma^2)$. On suit à nouveau le principe général énoncé à la Section 20.2.

Étape 1. L'hypothèse (H_0) est la suivante : $\sigma^2 = \sigma_0^2$ où σ_0^2 est un réel strictement positif dont on pense qu'il est la variance.

Étape 2. L'hypothèse (H_1) peut correspondre à un test bilatéral à savoir $\sigma^2 \neq \sigma_0^2$ comme il peut s'agir d'un test unilatéral : $\sigma^2 > \sigma_0^2$ (ou au contraire $\sigma^2 < \sigma_0^2$). On voit ainsi que l'on dispose de trois hypothèses alternatives. Notons-les respectivement (H_1^b), (H_1^+) et (H_1^-). Les trois cas seront traités par la suite.

Étape 3. La statistique naturelle pour approximer la variance est la variance d'échantillon $S_n^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ puisqu'il s'agit d'un estimateur asymptotiquement sans biais et convergent de σ^2 . On note toutefois que la loi de S_n^2 dépend de μ . Ainsi, deux cas peuvent se produire : on connaît μ ou on ne connaît pas μ .

21.3.1 En supposant la moyenne μ connue

Étape 4. On suppose ici que la moyenne μ est connue. Alors, sous l'hypothèse (H_0), la variable aléatoire $Z_n := \frac{nS_n^2}{\sigma_0^2}$ suit la loi du χ^2 à n degrés de liberté. On note que Z_n ne dépend d'aucun paramètre inconnu.

Étape 5. On se donne ensuite un niveau de signification (erreur de première espèce) α .

21.3.1.1 Test bilatéral : (H_1^b)

Étape 6. La région critique consiste ici à trouver deux valeurs $q_c^1(\alpha)$ et $q_c^2(\alpha)$ telles que

$$\mathbb{P}(q_c^1(\alpha) \leq Z_n \leq q_c^2(\alpha)) = 1 - \alpha.$$

On trouve par exemple $q_c^1(\alpha) := \chi_{\frac{\alpha}{2}}^2(n)$, le quantile d'ordre $\frac{\alpha}{2}$ de la loi du χ^2 à n degrés de liberté. Et alors $q_c^2(\alpha) := \chi_{1-\frac{\alpha}{2}}^2(n)$, le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi du χ^2 à n degrés de liberté. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^b :=]0; \chi_{\frac{\alpha}{2}}^2(n)] \cup [\chi_{1-\frac{\alpha}{2}}^2(n); +\infty[.$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^b$, on rejette l'hypothèse (H_0) et on accepte (H_1^b); sinon on ne la rejette pas.

21.3.1.2 Test unilatéral : (H_1^+)

Étape 6. La région critique consiste ici à trouver une valeur $q_c(\alpha)$ telle que

$$\mathbb{P}(Z_n \geq q_c(\alpha)) = \alpha.$$

On trouve ainsi $q_c(\alpha) := \chi_{1-\alpha}^2(n)$, le quantile d'ordre $1 - \alpha$ de la loi du χ^2 à n degrés de liberté. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^+ := [\chi_{1-\alpha}^2(n); +\infty[.$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^+$, on rejette l'hypothèse (H_0) et on accepte (H_1^+); sinon on ne la rejette pas.

21.3.1.3 Test unilatéral : (H_1^-)

Étape 6. La région critique consiste ici à trouver une valeur $q_c(\alpha)$ telle que

$$\mathbb{P}(Z_n \leq q_c(\alpha)) = \alpha.$$

On trouve ainsi $q_c(\alpha) := \chi_\alpha^2(n)$, le quantile d'ordre α de la loi du χ^2 à n degrés de liberté. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^- := [0; \chi_\alpha^2(n)].$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^-$, on rejette l'hypothèse (H_0) et on accepte (H_1^-); sinon on ne la rejette pas.

21.3.2 En supposant la moyenne μ inconnue

Étape 4. On suppose ici que la moyenne μ est inconnue. On l'estime alors par \overline{X}_n . De fait, on considère la statistique $\widetilde{S}_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ et l'on sait que $(n-1) \frac{\widetilde{S}_n^2}{\sigma^2}$ suit la loi du χ^2 à $n - 1$ degrés de liberté. Alors, sous l'hypothèse (H_0), la variable aléatoire $Z_n := (n-1) \frac{\widetilde{S}_n^2}{\sigma_0^2}$ suit la loi du χ^2 à $n - 1$ degrés de liberté. On note que Z_n ne dépend d'aucun paramètre inconnu.

Étape 5. On se donne ensuite un niveau de signification (erreur de première espèce) α .

21.3.2.1 Test bilatéral : (H_1^b)

Étape 6. La région critique consiste ici à trouver deux valeurs $q_c^1(\alpha)$ et $q_c^2(\alpha)$ telles que

$$\mathbb{P}(q_c^1(\alpha) \leq Z_n \leq q_c^2(\alpha)) = 1 - \alpha.$$

On trouve par exemple $q_c^1(\alpha) := \chi_{\frac{\alpha}{2}}^2(n-1)$, le quantile d'ordre $\frac{\alpha}{2}$ de la loi du χ^2 à $n-1$ degrés de liberté. Et alors $q_c^2(\alpha) := \chi_{1-\frac{\alpha}{2}}^2(n-1)$, le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi du χ^2 à $n-1$ degrés de liberté. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^b := \left[0; \chi_{\frac{\alpha}{2}}^2(n-1)\right] \cup \left[\chi_{1-\frac{\alpha}{2}}^2(n-1); +\infty\right].$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^b$, on rejette l'hypothèse (H_0) et on accepte (H_1^b) ; sinon on ne la rejette pas.

21.3.2.2 Test unilatéral : (H_1^+)

Étape 6. La région critique consiste ici à trouver une valeur $q_c(\alpha)$ telle que

$$\mathbb{P}(Z_n \geq q_c(\alpha)) = \alpha.$$

On trouve ainsi $q_c(\alpha) := \chi_{1-\alpha}^2(n-1)$, le quantile d'ordre $1 - \alpha$ de la loi du χ^2 à $n-1$ degrés de liberté. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^+ := \left[\chi_{1-\alpha}^2(n-1); +\infty\right].$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^+$, on rejette l'hypothèse (H_0) et on accepte (H_1^+) ; sinon on ne la rejette pas.

21.3.2.3 Test unilatéral : (H_1^-)

Étape 6. La région critique consiste ici à trouver une valeur $q_c(\alpha)$ telle que

$$\mathbb{P}(Z_n \leq q_c(\alpha)) = \alpha.$$

On trouve ainsi $q_c(\alpha) := \chi_\alpha^2(n-1)$, le quantile d'ordre α de la loi du χ^2 à $n-1$ degrés de liberté. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^- := \left[0; \chi_\alpha^2(n-1)\right].$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^-$, on rejette l'hypothèse (H_0) et on accepte (H_1^-) ; sinon on ne la rejette pas.

21.4 Test sur une proportion f

On se donne une grande population et l'on s'intéresse à la fréquence d'apparition d'un certain caractère chez les individus de ladite population.

On tire n individus au hasard et indépendamment. On suppose que la population est suffisamment grande pour que $(X_i)_{i \in \llbracket 1; n \rrbracket}$ soit bien une suite de variables aléatoires indépendantes et identiquement distribuées où $X_i = 1$ si l'individu i présente le caractère et $X_i = 0$ sinon.

On pose f la proportion d'individus dans l'échantillon qui présentent le caractère.

Étape 1. L'hypothèse (H_0) est la suivante : $f = f_0$ où $f_0 \in [0; 1]$ et où l'on suspecte que f_0 soit la proportion f .

Étape 2. L'hypothèse (H_1) peut correspondre à un test bilatéral à savoir $f \neq f_0$ comme il peut s'agir d'un test unilatéral : $f > f_0$ (ou au contraire $f < f_0$). On voit ainsi que l'on dispose de trois hypothèses alternatives. Notons-les respectivement (H_1^b) , (H_1^+) et (H_1^-) . Les trois cas seront traités par la suite.

Étape 3. La statistique naturelle pour approximer la fréquence f est la proportion d'échantillon $F_n := \frac{1}{n} \sum_{i=1}^n X_i$ puisqu'il s'agit d'un estimateur sans biais et convergent de f en tant que moyenne d'échantillon. En supposant que n est assez grand, le théorème central de la limite s'applique si bien que l'on peut considérer que F_n suit une loi normale de paramètres p et $\frac{p(1-p)}{n}$.

Étape 4. Ainsi, sous l'hypothèse (H_0) , la variable aléatoire F_n suit la loi normale de paramètres p_0 et $\frac{p_0(1-p_0)}{n}$.

Étape 5. On se donne ensuite un niveau de signification (erreur de première espèce) α .

21.4.0.1 Test bilatéral : (H_1^b)

Étape 6. La région critique consiste ici à trouver une valeur $q_c(\alpha)$ telle que

$$\mathbb{P}(|F_n - p_0| \leq q_c(\alpha)) = 1 - \alpha.$$

On trouve $q_c(\alpha) := q_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}$, où $q_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^b := \left] -\infty; p_0 - q_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}} \right] \cup \left[p_0 + q_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}; +\infty \right[.$$

On note que l'ensemble \mathcal{J}_α^b n'est pas inclus dans $[0; 1]$.

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^b$, on rejette l'hypothèse (H_0) et on accepte (H_1^b) ; sinon on ne la rejette pas.

21.4.0.2 Test unilatéral : (H_1^+)

Étape 6. La région critique consiste ici à trouver une valeur $q_c(\alpha)$ telle que

$$\mathbb{P}(F_n - p_0 \geq q_c(\alpha)) = \alpha.$$

On trouve $q_c(\alpha) := q_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}}$, où $q_{1-\alpha}$ est le quantile d'ordre $1 - \alpha$ de la loi normale centrée réduite. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^+ := \left[p_0 + q_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}}; +\infty \right[.$$

On note que l'intervalle \mathcal{J}_α^+ n'est pas inclus dans $[0; 1]$.

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^+$, on rejette l'hypothèse (H_0) et on accepte (H_1^+) ; sinon on ne la rejette pas.

21.4.0.3 Test unilatéral : (H_1^-)

Étape 6. La région critique consiste ici à trouver une valeur $q_c(\alpha)$ telle que

$$\mathbb{P}(F_n - p_0 \leq q_c(\alpha)) = \alpha.$$

On trouve $q_c(\alpha) := q_\alpha \sqrt{\frac{p_0(1-p_0)}{n}}$, où q_α est le quantile d'ordre α de la loi normale centrée réduite. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha^- := \left] -\infty; p_0 - q_\alpha \sqrt{\frac{p_0(1-p_0)}{n}} \right].$$

On note que l'intervalle \mathcal{J}_α^- n'est pas inclus dans $[0; 1]$.

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha^-$, on rejette l'hypothèse (H_0) et on accepte (H_1^-) ; sinon on ne la rejette pas.

21.5 Test de comparaison entre deux moyennes

Dans cette section et dans les suivantes, le test sera bilatéral.

On suppose ici que l'on a deux suites de variables aléatoires indépendantes et identiquement distribuées $(X_i)_{i \in \llbracket 1; n_1 \rrbracket}$ et $(Y_i)_{i \in \llbracket 1; n_2 \rrbracket}$. On suppose de plus que les deux suites sont mutuellement indépendantes.

L'objectif du test est de vérifier si les moyennes sous-jacentes des deux suites sont les mêmes.

Si $n_1 = n_2$, en posant $Z_i := X_i - Y_i$ pour tout $i \in \llbracket 1; n_1 \rrbracket$, on peut se ramener à la Section 21.2 où l'on prend $\mu_0 := 0$. Dorénavant, on suppose $n_1 \neq n_2$. De plus, X_i suit une loi normale de paramètres μ_X et σ_X^2 tandis que Y suit une loi normale de paramètres μ_Y et σ_Y^2 .

Étape 1. L'hypothèse (H_0) est la suivante : $\mu_X = \mu_Y$ où μ_X est la moyenne de X et μ_Y est celle de Y .

Étape 2. L'hypothèse (H_1) est ici $\mu_X \neq \mu_Y$.

Étape 3. On pose $\overline{X}_{n_1} := \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$ et $\overline{Y}_{n_2} := \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$. La statistique naturelle pour comparer les moyennes est alors $\overline{X}_{n_1} - \overline{Y}_{n_2}$. En effet, celle-ci suit la loi normale de paramètres $\mu_X - \mu_Y$ et $\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}$.

21.5.1 En supposant les variances σ_X^2 et σ_Y^2 connues

Étape 4. On suppose ici que les variances sont connues. Alors, sous l'hypothèse (H_0) , la variable aléatoire $Z_{n_1, n_2} := \frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}}$ suit la loi normale centrée réduite. On note que Z_{n_1, n_2} ne dépend d'aucun paramètre inconnu.

Étape 5. On se donne ensuite un niveau de signification (erreur de première espèce) α .

Étape 6. La région critique consiste ici à résoudre

$$\mathbb{P}(|Z_{n_1, n_2}| > q_c(\alpha)) = \alpha.$$

On trouve $q_c(\alpha) := q_{1-\frac{\alpha}{2}}$, le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha :=]-\infty; -q_{1-\frac{\alpha}{2}}] \cup [q_{1-\frac{\alpha}{2}}; +\infty[.$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha$, on rejette l'hypothèse (H_0) et on accepte (H_1) ; sinon on ne la rejette pas.

21.5.2 En supposant les variances σ_X^2 et σ_Y^2 inconnues

Étape 4. On suppose ici que les variances sont inconnues. Néanmoins, on les suppose égales à σ^2 à savoir que l'on dispose de l'hypothèse d'homoscédasticité.

On peut donc l'estimer par $\widetilde{S}_X^2 := \frac{1}{n_1-1} \sum_{i=1}^{n_1} (X_i - \overline{X}_{n_1})^2$ ainsi que par $\widetilde{S}_Y^2 := \frac{1}{n_2-1} \sum_{i=1}^{n_2} (Y_i - \overline{Y}_{n_2})^2$. Et, on estime donc σ^2 par $\theta_{n_1, n_2} := \frac{(n_1-1)\widetilde{S}_X^2 + (n_2-1)\widetilde{S}_Y^2}{n_1+n_2-2}$. On en déduit que $(n_1+n_2-2)\frac{\theta_{n_1, n_2}^2}{\sigma^2}$ suit une loi du χ^2 à n_1+n_2-2 degrés de liberté. Alors, sous l'hypothèse (H_0) , la variable aléatoire $Z_{n_1, n_2} := \frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\theta_{n_1, n_2}^2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ suit la loi de Student à n_1+n_2-2 degrés de liberté. On note que Z_{n_1, n_2} ne dépend d'aucun paramètre inconnu.

Étape 5. On se donne ensuite un niveau de signification (erreur de première espèce) α .

Étape 6. La région critique consiste ici à résoudre

$$\mathbb{P}(|Z_{n_1, n_2}| > q_c(\alpha)) = \alpha.$$

On trouve $q_c(\alpha) := t_{1-\frac{\alpha}{2}}$, le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de Student à n_1+n_2-2 degrés de liberté. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha :=]-\infty; -t_{1-\frac{\alpha}{2}}] \cup [t_{1-\frac{\alpha}{2}}; +\infty[.$$

Étape 7. On calcule ensuite la valeur de Z_{n_1, n_2} dans l'échantillon.

Étape 8. Ainsi, si $Z_{n_1, n_2} \in \mathcal{J}_\alpha$, on rejette l'hypothèse (H_0) et on accepte (H_1) ; sinon on ne la rejette pas.

Remarque 21.5.1. Si l'on veut s'économiser l'hypothèse d'homoscédasticité, on doit alors supposer que n_1 et n_2 sont suffisamment grands; typiquement supérieurs ou égaux à 30. On peut alors appliquer le théorème central de la limite. De fait, la statistique à utiliser devient

$$Z_{n_1, n_2} := \frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{\widetilde{S}_X^2}{n_1} + \frac{\widetilde{S}_Y^2}{n_2}}}.$$

Et, Z_{n_1, n_2} suit approximativement la loi normale centrée réduite. On dispose alors d'un test asymptotique.

21.6 Test de comparaison entre deux variances

On suppose ici que l'on a deux suites de variables aléatoires indépendantes et identiquement distribuées $(X_i)_{i \in [1; n_1]}$ et $(Y_i)_{i \in [1; n_2]}$. On suppose de plus que les deux suites sont mutuellement indépendantes et qu'il s'agit de lois normales de moyennes respectives μ_X et μ_Y et de variances respectives σ_X^2 et σ_Y^2 .

L'objectif du test est de vérifier si les variances sous-jacentes des deux suites sont les mêmes; c'est-à-dire si on a homoscédasticité.

Étape 1. L'hypothèse (H_0) est la suivante : $\sigma_X^2 = \sigma_Y^2$.

Étape 2. L'hypothèse (H_1) est ici $\sigma_X^2 \neq \sigma_Y^2$.

Étape 3. On pose $\widetilde{S}_X^2 := \frac{1}{n_1-1} \sum_{i=1}^{n_1} (X_i - \overline{X}_{n_1})^2$ et $\widetilde{S}_Y^2 := \frac{1}{n_2-1} \sum_{i=1}^{n_2} (Y_i - \overline{Y}_{n_2})^2$.

Étape 4. On sait que $\frac{(n_1-1)\widetilde{S}_X^2}{\sigma_X^2}$ suit la loi du χ^2 à $n_1 - 1$ degrés de liberté. De même, $\frac{(n_2-1)\widetilde{S}_Y^2}{\sigma_Y^2}$ suit la loi du χ^2 à $n_2 - 1$ degrés de liberté. Comme les variables sont indépendantes, il s'ensuit que

$$\frac{\widetilde{S}_X^2/\sigma_X^2}{\widetilde{S}_Y^2/\sigma_Y^2}$$

suit la loi de Fisher-Snedecor à $n_1 - 1$ et $n_2 - 1$ degrés de liberté, voir Section 10.11. Puis, sous l'hypothèse (H_0), on peut dire que

$$Z_{n_1, n_2} := \frac{\widetilde{S}_X^2}{\widetilde{S}_Y^2}$$

suit la loi de Fisher-Snedecor à $n_1 - 1$ et $n_2 - 1$ degrés de liberté.

Étape 5. On se donne ensuite un niveau de signification (erreur de première espèce) α .

Étape 6. La région critique consiste ici à trouver deux valeurs $q_c^1(\alpha)$ et $q_c^2(\alpha)$ telles que

$$\mathbb{P}(q_c^1(\alpha) \leq Z_{n_1, n_2} \leq q_c^2(\alpha)) = 1 - \alpha.$$

On prend par exemple $q_c^1(\alpha) := f_{\frac{\alpha}{2}}$, le quantile d'ordre $\frac{\alpha}{2}$ de la loi de Fisher-Snedecor à $n_1 - 1$ et $n_2 - 1$ degrés de liberté. Et, l'on prend $q_c^2(\alpha) := f_{1-\frac{\alpha}{2}}$, le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de Fisher-Snedecor à $n_1 - 1$ et $n_2 - 1$ degrés de liberté. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha := [0; f_{\frac{\alpha}{2}}] \cup [f_{1-\frac{\alpha}{2}}; +\infty[.$$

Étape 7. On calcule ensuite la valeur de Z_{n_1, n_2} dans l'échantillon.

Étape 8. Ainsi, si $Z_{n_1, n_2} \in \mathcal{J}_\alpha$, on rejette l'hypothèse (H_0) et on accepte (H_1); sinon on ne la rejette pas.

21.7 Test de comparaison entre deux proportions

On suppose ici que l'on a deux caractères de fréquences d'apparition respectives f_1 et f_2 . On note $X_i := 1$ si l'individu i présente le premier caractère et $X_i := 0$ sinon. De même, on note $Y_i := 1$ si l'individu i présente le second caractère et $Y_i := 0$ sinon.

L'objectif du test est de vérifier si les proportions sous-jacentes sont les mêmes.

Étape 1. L'hypothèse (H_0) est la suivante : $f_1 = f_2$.

Étape 2. L'hypothèse (H_1) est ici $f_1 \neq f_2$.

Étape 3. On pose $\overline{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$ et $\overline{Y}_n := \frac{1}{n} \sum_{i=1}^n Y_i$. La statistique naturelle pour comparer les proportions est alors $\overline{X}_n - \overline{Y}_n$. En effet, celle-ci suit approximativement la loi normale de paramètres $f_1 - f_2$ et de variance $\frac{f_1(1-f_1)}{n} + \frac{f_2(1-f_2)}{n}$; si n est grand. Mais, on ne connaît ni f_1 ni f_2 .

Étape 4. Sous l'hypothèse (H_0), on peut dire que

$$F_n := \frac{1}{2} (\overline{X}_n + \overline{Y}_n)$$

est un estimateur sans biais et convergent presque sûrement vers $f := f_1 = f_2$.
Puis :

$$Z_n := \frac{\overline{X}_n - \overline{Y}_n}{\sqrt{\frac{2F_n(1-F_n)}{n}}}$$

suit approximativement la loi normale centrée réduite si n est grand ; en utilisant le lemme de Slutsky (voir Théorème 14.5.5).

Étape 5. On se donne ensuite un niveau de signification (erreur de première espèce) α .

Étape 6. La région critique consiste ici à résoudre

$$\mathbb{P}(|Z_n| > q_c(\alpha)) = \alpha.$$

On trouve ainsi $q_c(\alpha) := q_{1-\frac{\alpha}{2}}$, le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha :=]-\infty; -q_{1-\frac{\alpha}{2}}] \cup [q_{1-\frac{\alpha}{2}}; +\infty[.$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha$, on rejette l'hypothèse (H_0) et on accepte (H_1) ; sinon on ne la rejette pas.

21.8 Test du χ^2 d'ajustement

Dans cette section, on présente un grand classique : le test du χ^2 d'ajustement.

On va commencer par montrer les justifications théoriques dudit test. Puis, on l'appliquera dans le cas d'une loi discrète, dans le cas d'une loi continue et enfin dans le cas où la loi dépend d'un paramètre θ à estimer ponctuellement.

21.8.1 Justification théorique du test

Donnons-nous des variables aléatoires $(X_i)_{i \in \llbracket 1; n \rrbracket}$ indépendantes et identiquement distribuées suivant la loi discrète suivante :

$$\mathbb{P}(X_i = k) = p_k \text{ pour } k \in \llbracket 1; r \rrbracket, \text{ avec } \sum_{k=1}^r p_k = 1.$$

Ainsi, il y a r classes différentes. Maintenant, pour toute classe $k \in \llbracket 1; r \rrbracket$, on pose :

$$N_k := \sum_{i=1}^n \mathbb{1}_{\{X_i=k\}}.$$

N_k correspond au nombre d'occurrences de la valeur k dans l'échantillon considéré. C'est donc une variable aléatoire. On sait que $\frac{1}{n}N_k$ converge presque sûrement vers p_k quand n tend vers l'infini d'après la loi forte des grands nombres. Et, d'après le théorème central de la limite, $\frac{\frac{1}{n}N_k - p_k}{\sqrt{\frac{p_k(1-p_k)}{n}}}$ tend en loi vers la loi normale centrée réduite.

Il est ainsi naturel **d'imaginer** que $\frac{(N_k - np_k)^2}{np_k(1-p_k)}$ puisse tendre vers une loi du χ^2 à 1 degré de liberté puis l'on peut **envisager** que la somme pour k allant de 1 à r de cette quantité tend en loi vers une loi du χ^2 quand n tend vers l'infini. Le cas échéant, on pourrait créer un test qui discrimine si oui ou non le r -uplet (p_1^0, \dots, p_r^0) correspond à la distribution réelle.

Plus rigoureusement, on introduit

$$\chi_n^2 := \sum_{k=1}^r \frac{(N_k - np_k)^2}{np_k} = n \sum_{k=1}^r \frac{\left(\frac{N_k}{n} - p_k\right)^2}{p_k}. \quad (21.1)$$

Et, l'on dispose du théorème suivant :

Théorème 21.8.1 (Théorème de Pearson). *Quand n tend vers l'infini, la variable aléatoire χ_n^2 définie par (21.1) tend en loi vers une loi du Khi-deux à $r - 1$ degrés de liberté.*

La preuve du théorème est omise.

Remarque 21.8.2. *En pratique, on se contente d'avoir $np_k \geq 5$ pour tout $k \in \llbracket 1; r \rrbracket$. Mais, il est préférable que l'on ait $np_k \geq 10$.*

Remarque 21.8.3. *Lorsque l'effectif attendu np_k d'une classe est plus petit que 5, il est recommandé de regrouper cette classe avec une autre qui lui est adjacente avant de procéder au test du χ^2 . Le test d'ajustement porte alors sur la distribution dans les classes obtenues après le regroupement.*

21.8.2 Pour une loi discrète

Soit $p_k^0 \in [0; 1]$ pour tout $k \in \llbracket 1; r \rrbracket$, avec $\sum_{k=1}^r p_k^0 = 1$.

Étape 1. On se donne l'hypothèse (H_0) suivante : $p = p^0$ c'est-à-dire que l'on a $p_k = p_k^0$ pour tout $k \in \llbracket 1; r \rrbracket$.

Étape 2. L'hypothèse alternative (H_1) est $p \neq p^0$ c'est-à-dire qu'il existe $k_0 \in \llbracket 1; r \rrbracket$ tel que $p_{k_0} \neq p_{k_0}^0$.

Étape 3. On introduit la statistique $Z_n := \sum_{k=1}^r \frac{(N_k - np_k^0)^2}{np_k^0} = n \sum_{k=1}^r \frac{\left(\frac{N_k}{n} - p_k^0\right)^2}{p_k^0}$ où N_k est le nombre d'occurrences de la classe k dans l'échantillon considéré.

Étape 4. On sait alors d'après le Théorème 21.8.1 que sous (H_0), Z_n converge en loi vers une loi du χ^2 à $r - 1$ degrés de liberté. Et, par la loi forte des grands nombres (Théorème 15.3.2), il est assez facile de montrer que Z_n tend presque sûrement vers l'infini sous (H_1).

Étape 5. On se donne ensuite un niveau de signification (erreur de première espèce) α .

Étape 6. La région critique consiste ici à résoudre

$$\mathbb{P}(Z_n > q_c(\alpha)) = \alpha.$$

On trouve ainsi $q_c(\alpha) := q_{1-\alpha}$, le quantile d'ordre $1 - \alpha$ de la loi du χ^2 à $r - 1$ degrés de liberté. Ceci nous donne la région critique suivante :

$$\mathcal{J}_\alpha := [q_{1-\alpha}; +\infty[.$$

Étape 7. On calcule ensuite la valeur de Z_n dans l'échantillon.

Étape 8. Ainsi, si $Z_n \in \mathcal{J}_\alpha$, on rejette l'hypothèse (H_0) et on accepte (H_1); sinon on ne la rejette pas.

Exemple 21.8.4. *On simule cent lancers d'un dé à six faces par un programme. On obtient la répartition suivante :*

TABLE 21.1 – Dé électronique

Valeurs isolées x_k	1	2	3	4	5	6
Effectifs N_k	17	22	18	14	13	16

Peut-on considérer, au risque de 5%, que la simulation est bien celle d'un dé non pipé ?

Considérer le dé comme équilibré revient à supposer l'hypothèse (H_0) : "l'échantillon est celui d'une loi équirépartie sur un ensemble à $r = 6$ éléments". On a ici $p_1^0 = p_2^0 = \dots = p_6^0 = \frac{1}{6}$.

On organise les calculs de la manière suivante :

TABLE 21.2 – Test du χ^2 sur le dé

x_k	1	2	3	4	5	6	Somme
N_k	17	22	18	14	13	16	100
np_k^0	16.67	16.67	16.67	16.67	16.67	16.67	100
$\frac{(N_k - np_k^0)^2}{np_k^0}$	0.0067	1.7067	0.1067	0.4267	0.8067	0.0267	$\chi_{obs}^2 = 3.08$

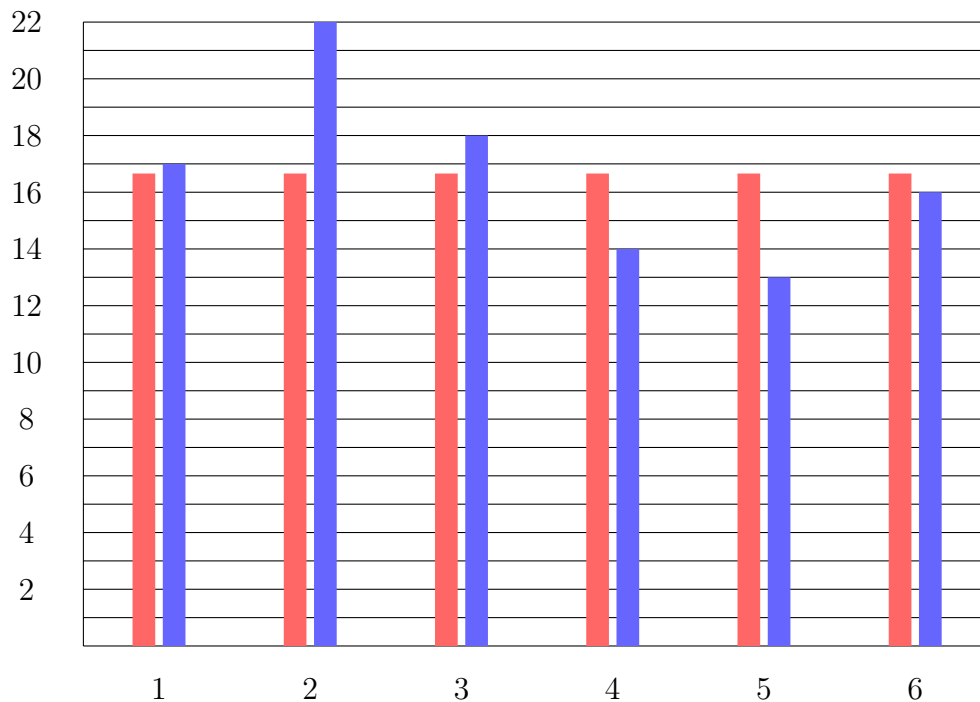
Pour la loi du Khi-deux à $r - 1 = 5$ degrés de liberté, une lecture de la table donne le quantile d'ordre 0.95 égal à $\chi_{0.95}^2(5) \approx 11.1$.

Comme $\chi_{obs}^2 < 11.1$, on ne rejette pas (H_0).

Il s'agit, au risque d'erreur de 5%, d'une simulation d'un dé équilibré.

Regardons par ailleurs les deux distributions (effectifs théoriques et effectifs réels) sur un diagramme en barres :

FIGURE 21.1 – Comparaison des effectifs réels (en bleu) et théoriques (en rouge)



21.8.3 Pour une loi continue

Si l'on suppose maintenant que la loi suivie par X est continue et qu'on veut la confronter à une densité f , on se ramène alors à une discrétisation de ladite loi. Pour ce faire, on se donne $r - 1$ réels deux à deux distincts a_1, \dots, a_{r-1} . On suppose par ailleurs que l'on a $a_1 < a_2 < \dots < a_k < a_{k+1} < \dots < a_{r-1}$. On pose ensuite

$$p_k^0 := \int_{a_{k-1}}^{a_k} f(x) dx,$$

pour tout $k \in \llbracket 1; r \rrbracket$ où $a_0 := -\infty$ et $a_r := +\infty$.

On peut ainsi se ramener au cas du paragraphe précédent en posant

$$N_k := \sum_{i=1}^n \mathbb{1}_{\{X_i \in]a_{k-1}; a_k\}}.$$

21.8.4 Pour des lois qui dépendent d'un paramètre θ

Nous souhaitons maintenant mettre en place un test permettant de décider si la loi de l'échantillon appartient ou non à une famille de lois $(p^{(\theta)})_{\theta \in \Theta}$ indexée par un paramètre θ à valeurs dans $\Theta \subset \mathbb{R}^d$. On suppose que l'on dispose de $\hat{\theta}_n$, l'estimateur du maximum de vraisemblance, voir la page 279.

Théorème 21.8.5. *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de même loi $p^{(\theta^*)}$ (où $\theta^* \in \Theta \subset \mathbb{R}^d$ est inconnu). Notons $\hat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ^* et $p_k(\hat{\theta}_n) := p_k^{(\hat{\theta}_n)}$ pour $k \in \llbracket 1; n \rrbracket$. Alors,*

$$\widetilde{\chi}_n^2 := \sum_{k=1}^r \frac{\left(N_k - np_k(\hat{\theta}_n)\right)^2}{np_k(\hat{\theta}_n)}$$

converge en loi vers une loi du χ^2 à $q - d - 1$ degrés de liberté.

Remarque 21.8.6. *On perd ainsi un degré de liberté par paramètre réel estimé.*

Remarque 21.8.7. *Pour certaines familles de lois classiques comme la loi de Poisson, la loi normale ou la loi binomiale, l'estimateur du maximum de vraisemblance coïncide avec l'estimateur de la méthode des moments, comme on l'a déjà vu.*

21.9 Test du χ^2 d'indépendance

Supposons qu'on ait une population subdivisée en r classes selon une variable X et en s classes selon une variable Y . On désigne par $p_{k,l}$ la proportion des

individus dans la population appartenant à la k -ème classe selon X et à la l -ème classe selon Y .

Avec cette notation, la proportion des individus de la population appartenant à la classe k selon la variable X est

$$p_{k,\bullet} := \sum_{l=1}^s p_{k,l}.$$

De même, la proportion des individus de la population appartenant à la classe l selon la variable Y est

$$p_{\bullet,l} := \sum_{k=1}^r p_{k,l}.$$

Il y a indépendance entre les variables aléatoires X et Y avec la classification utilisée (voir Théorème 16.6.13) si et seulement si

$$p_{k,l} = p_{k,\bullet} p_{\bullet,l}$$

pour tout $k \in \llbracket 1; r \rrbracket$ et pour tout $l \in \llbracket 1; s \rrbracket$.

Le problème qu'on se propose de résoudre est celui du test de cette hypothèse d'indépendance à l'aide des résultats d'un échantillon de taille n extrait de la population.

Supposons qu'on observe une proportion $p_{k,l}(n)$ d'individus appartenant à la cellule (k, l) , définie comme la classe k pour la variable X et la classe l pour la variable Y . On définit de même $p_{k,\bullet}(n)$ et $p_{\bullet,l}(n)$.

On introduit la statistique :

$$Z_n := n \sum_{k=1}^r \sum_{l=1}^s \frac{(p_{k,l}(n) - p_{k,\bullet}(n)p_{\bullet,l}(n))^2}{p_{k,\bullet}(n)p_{\bullet,l}(n)}. \quad (21.2)$$

On peut montrer que sous l'hypothèse d'indépendance, Z_n converge en loi vers une loi du Khi-deux à $(r-1)(s-1)$ degrés de liberté. La preuve de ce résultat est omise.

Remarque 21.9.1. *Pour que le résultat décrit ci-dessus soit assez précis, l'effectif attendu sous l'hypothèse (H_0) d'indépendance dans chaque cellule doit être supérieur ou égal à 5. Cependant, lorsque le nombre de cellules est grand, on peut tolérer un certain nombre de cellules avec des effectifs attendus plus petits que 5 sans trop affecter la justesse du test.*

Proposition 21.9.2 (Règle de décision). *Soit $\alpha \in]0; 1[$. On teste l'hypothèse (H_0) contre l'hypothèse (H_1) au risque d'erreur α .*

Soit $\chi_{1-\alpha}^2 [(r-1)(s-1)]$ le quantile d'ordre $1-\alpha$ de la loi du Khi-deux à $(r-1)(s-1)$ degrés de liberté. La région critique pour laquelle on rejette (H_0) est alors :

$$\mathcal{J}_\alpha := [\chi_{1-\alpha}^2 [(r-1)(s-1)]; +\infty[.$$

En d'autres termes, on rejette (H_0) si la valeur observée de Z_n dans l'échantillon est dans \mathcal{J}_α .

21.10 Test du χ^2 d'homogénéité

Considérons r populations, chacune subdivisée en s classes distinctes selon une même variable aléatoire X . On dit que les populations sont homogènes si la distribution est la même dans les r populations.

Pour tester cette hypothèse notée (H_0) , on extrait un échantillon dans chacune des populations et on observe la variable X . Ici, $N_{k,l}$ désigne le nombre d'observations dans la k -ème population appartenant à la l -ème classe de X . De même, on appelle n la taille commune des échantillons et $N_{\bullet,l}$ le nombre total d'observations appartenant à la l -ème classe.

On introduit la statistique

$$Z_n := \sum_{k=1}^r \sum_{l=1}^s \frac{\left(N_{k,l} - \frac{N_{\bullet,l}}{r}\right)^2}{\frac{N_{\bullet,l}}{r}}. \quad (21.3)$$

On peut montrer que sous l'hypothèse d'homogénéité, Z_n converge en loi vers une loi du Khi-deux à $(r-1)(s-1)$ degrés de liberté. La preuve de ce résultat est omise.

Remarque 21.10.1. *Pour que le résultat décrit ci-dessus soit assez précis, l'effectif attendu sous (H_0) dans chaque cellule doit être supérieur ou égal à 5. Cependant, lorsque le nombre de cellules est grand, on peut tolérer un certain nombre de cellules avec des effectifs attendus plus petits que 5 sans trop affecter la justesse du test.*

Proposition 21.10.2 (Règle de décision). *Soit $\alpha \in]0; 1[$. On teste l'hypothèse (H_0) contre l'hypothèse (H_1) au risque d'erreur α .*

Soit $\chi_{1-\alpha}^2 [(r-1)(s-1)]$ le quantile d'ordre $1-\alpha$ de la loi du Khi-deux à $(r-1)(s-1)$ degrés de liberté. La région critique pour laquelle on rejette (H_0) est alors :

$$\mathcal{J}_\alpha := [\chi_{1-\alpha}^2 [(r-1)(s-1)]; +\infty[.$$

En d'autres termes, on rejette (H_0) si la valeur observée de Z_n dans l'échantillon est dans \mathcal{J}_α .

21.11 Tests de Kolmogorov-Smirnov

Le test du χ^2 n'est pas bien adapté à la comparaison d'une statistique à une loi continue. Dans ce cas, on peut se tourner vers le test de Kolmogorov-Smirnov.

21.11.1 Test d'ajustement

On va commencer par énoncer un lemme très pratique.

Lemme 21.11.1. *Soit (X_1, \dots, X_n) un n -échantillon de la loi μ , de fonction de répartition F et F_n la fonction de répartition empirique associée :*

$$F_n(x) := \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{]-\infty; x]}(X_k) .$$

On suppose que F est continue. Alors la loi de $D_n := \|F_n - F\|_\infty$ ne dépend pas de la loi μ . On dit que D_n est une statistique libre.

Remarque 21.11.2. *Le calcul numérique de D_n est très simple car il ne fait intervenir qu'un nombre fini de points.*

Prenant pour hypothèse (H_0) : "l'échantillon suit la loi μ ", si $\mathbb{P}(D_n \leq M) = \alpha$, on rejette avec un risque α l'hypothèse (H_0) si l'observation de la variable aléatoire D_n dépasse M .

Il est également théoriquement possible de calculer explicitement la fonction de répartition de D_n (c'est une fonction polynomiale par morceaux), mais on n'obtient pas des formules fermées. À partir de cela, on peut construire des tables de la loi de D_n pour les petites valeurs de n . Pour les grandes valeurs de n , le calcul de la fonction de répartition de D_n est numériquement plus compliqué. Heureusement, nous disposons du théorème suivant, dit de Kolmogorov-Smirnov, qui permet de bonnes approximations.

Théorème 21.11.3. *Soit (X_1, \dots, X_n) un n -échantillon de la loi μ et F_n la fonction de répartition empirique associée :*

$$F_n(x) := \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{]-\infty; x]}(X_k) .$$

On suppose de plus que F (la fonction de répartition de μ) est continue. Alors, $\sqrt{n} \|F_n - F\|_\infty$ converge en loi vers la loi de Kolmogorov-Smirnov μ_{KS} , qui est caractérisée par sa fonction de répartition :

$$\mu_{KS} (]-\infty; t]) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k \exp \{-2k^2 t^2\} .$$

Ainsi, comme la série $\sum_{k=1}^{\infty} (-1)^k \exp \{-2k^2 t^2\}$ converge très rapidement, il suffit de calculer les trois premiers termes pour obtenir quatre chiffres significatifs dès que $t > 0.56$.

21.11.2 Test d'homogénéité

Le test de Kolmogorov-Smirnov pour tester l'homogénéité de deux échantillons est très semblable au précédent. Supposons que l'on ait un premier échantillon (X_1, \dots, X_n) de loi μ ayant la fonction de répartition continue F . Supposons qu'on ait un deuxième échantillon (Y_1, \dots, Y_n) de loi ν ayant la fonction de répartition continue G . On veut tester l'hypothèse $(H_0) : \mu = \nu$ contre l'hypothèse alternative $(H_1) : \mu \neq \nu$.

La statistique

$$D_{m,n} := \sup_{x \in \mathbb{R}} |F_n(x) - G_m(x)|$$

est encore une statistique libre et, lorsque m et n tendent vers l'infini, la quantité $\sqrt{\frac{mn}{m+n}} D_{m,n}$ converge en loi vers la loi de Kolmogorov-Smirnov ; précédemment définie.

Régression linéaire

22.1 Introduction

Dans ce chapitre, on étudiera la régression linéaire simple et uniquement la régression linéaire simple. Il convient néanmoins de savoir que l'on peut être amené à effectuer de la régression linéaire multiple.

On suppose que l'on a un n -échantillon de valeurs $(x[i], y[i])$ où $x[i]$ est la valeur du caractère quantitatif X pour l'individu i tandis que $y[i]$ est celle du caractère Y .

Tout ce qui précède dans le livre est un pré-requis. Il en est de même pour les équations de droite.

Le principal objectif de ce chapitre est de savoir calculer les coefficients de la droite de régression et aussi de savoir à quoi elle sert (s'ajuster au plus près, dans le cadre du modèle gaussien, du nuage de points). Notamment, il est attendu du lecteur qu'il comprenne bien que la philosophie sous-jacente à la droite de régression est ici intimement liée au cadre gaussien.

22.2 Équation de la droite de régression

On souhaite obtenir une droite qui s'ajuste "au mieux" au nuage de points. Il s'avère que les mots "au mieux" peuvent varier selon le modèle que l'on utilise.

On suppose ici que l'on a

$$y[i] = \alpha x[i] + \beta + \varepsilon[i],$$

où ε est un Bruit Blanc de variance σ^2 . En d'autres termes, pour tout i , on a $\mathbb{E}[\varepsilon[i]] = 0$, $\text{Var}[\varepsilon[i]] = \sigma^2$ et de plus, $\mathbb{E}[\varepsilon[i]\varepsilon[j]] = 0$ si $i \neq j$.

De fait, si $\mathbb{E}[\varepsilon[i]^4] < +\infty$, on dispose de la limite

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n \varepsilon[i]^2 = \sigma^2,$$

d'après la loi forte des grands nombres.

Ainsi, minimiser le bruit revient à minimiser la moyenne des distances au carré :

$$D(\alpha, \beta) := \frac{1}{n} \sum_{i=1}^n |y[i] - \alpha x[i] - \beta|^2 .$$

On dispose ici d'une fonction de \mathbb{R}^2 dans \mathbb{R}_+ . De plus, la fonction D est de la forme :

$$D(\alpha, \beta) = \beta^2 + \alpha^2 \frac{1}{n} \sum_{i=1}^n x[i]^2 + 2\alpha\beta \frac{1}{n} \sum_{i=1}^n x[i] + L(\alpha, \beta) ,$$

où L est une fonction affine. On regarde donc la forme quadratique $Q(\alpha, \beta) := \beta^2 + \alpha^2 \frac{1}{n} \sum_{i=1}^n x[i]^2 + 2\alpha\beta \frac{1}{n} \sum_{i=1}^n x[i]$. La trace et le déterminant de la matrice associée à cette forme quadratique sont positifs. De fait, $D(\alpha, \beta)$ tend vers l'infini dès que $\alpha^2 + \beta^2$ tend vers l'infini.

On en déduit que le minimiseur de D est atteint en un point critique. On regarde maintenant les dérivées partielles de D par rapport à α et à β :

$$\frac{\partial}{\partial \alpha} D(\alpha, \beta) = -\frac{2}{n} \sum_{i=1}^n x[i] (y[i] - \alpha x[i] - \beta)$$

ainsi que

$$\frac{\partial}{\partial \beta} D(\alpha, \beta) = -\frac{2}{n} \sum_{i=1}^n (y[i] - \alpha x[i] - \beta) .$$

Ainsi, on trouve $\beta_0 = \frac{1}{n} \sum_{i=1}^n y[i] - \alpha_0 \frac{1}{n} \sum_{i=1}^n x[i]$. Cette égalité implique $\beta_0 = \bar{y} - \alpha_0 \bar{x}$, ce qui traduit l'absence de direction privilégiée dans le Bruit Blanc.

Également, on a

$$\overline{x^2} \alpha_0 = \overline{xy} - \beta_0 \bar{x} = \overline{xy} - \bar{y} \bar{x} + \alpha_0 \bar{x}^2 .$$

Il s'ensuit

$$\alpha_0 = \frac{s_{xy}}{s_x^2} \text{ et } \beta_0 = \bar{y} - \bar{x} \frac{s_{xy}}{s_x^2} ,$$

où s_{xy} est la covariance des séries marginales tandis que s_x^2 est la variance empirique de la série x .

22.3 Lien avec $r_{x,y}$

On peut maintenant se demander la valeur du bruit moyen σ^2 . On estime celle-ci par

$$D(\alpha_0, \beta_0) = \frac{1}{n} \sum_{i=1}^n |y[i] - \alpha_0 x[i] - \beta_0|^2 .$$

Or, $\beta_0 = \bar{y} - \alpha_0 \bar{x}$ donc on en déduit :

$$\begin{aligned} D(\alpha_0, \beta_0) &= \frac{1}{n} \sum_{i=1}^n |(y[i] - \bar{y}) - \alpha_0 (x[i] - \bar{x})|^2 \\ &= \frac{1}{n} \sum_{i=1}^n |y[i] - \bar{y}|^2 + \alpha_0^2 \frac{1}{n} \sum_{i=1}^n |x[i] - \bar{x}|^2 - 2\alpha_0 \frac{1}{n} \sum_{i=1}^n (y[i] - \bar{y})(x[i] - \bar{x}) \\ &= s_y^2 + \left(\frac{s_{xy}}{s_x^2} \right)^2 s_x^2 - 2 \frac{s_{xy}}{s_x^2} s_{xy} \\ &= s_y^2 \left(1 + \frac{s_{xy}^2}{s_x^2 s_y^2} - 2 \frac{s_{xy}^2}{s_x^2 s_y^2} \right) \\ &= s_y^2 (1 - r_{xy}^2) , \end{aligned}$$

où r_{xy} est le coefficient de corrélation linéaire.

On utilise aussi la valeur R^2 (sur excel par exemple) qui correspond au carré du coefficient de corrélation linéaire entre y et \hat{y} avec $\hat{y}[i] := \alpha_0 x[i] + \beta_0 = \frac{s_{xy}}{s_x^2} x[i] + \bar{y} - \bar{x} \frac{s_{xy}}{s_x^2}$. Ce coefficient de corrélation est donc égal à

$$\begin{aligned} r_{y\hat{y}} &= \frac{1}{n} \sum_{i=1}^n y[i] \left(\frac{s_{xy}}{s_x^2} x[i] + \bar{y} - \bar{x} \frac{s_{xy}}{s_x^2} \right) - \bar{y}^2 \\ &= \frac{s_{xy} \bar{x} \bar{y}}{s_x^2} + \bar{y}^2 - \bar{x} \bar{y} \frac{s_{xy}}{s_x^2} - \bar{y}^2 \\ &= \frac{s_{xy}^2}{s_x^2} . \end{aligned}$$

Ce coefficient correspond à la valeur résiduelle de la variance de y : $s_y^2 - \sigma^2$. Il s'agit en d'autres termes de la variance expliquée.

Troisième partie

Annexes

Rappels et compléments

23.1 Introduction

Il n'y a aucun pré-requis post-bac pour ce chapitre.

Celui-ci a pour objectif principal de revoir les fondements mathématiques nécessaires à la compréhension du chapitre suivant sur les éléments de théorie de la mesure. Notamment, du vocabulaire sur les ensembles, des rappels sur les suites et sur les fonctions doivent être maîtrisés. Il en est de même de l'intégration au sens de Riemann. Enfin, la topologie générale est brièvement rappelée.

23.2 Ensembles

Dans cette section, nous donnons quelques rappels sur les ensembles, que l'on appelle parfois "classes", "familles" ou "collections".

Il arrive qu'un ensemble puisse être entièrement décrit. Ainsi, $\{x_0\}$ est l'ensemble qui ne contient qu'un élément : x_0 . On dit de lui qu'il s'agit d'un singleton. On peut aussi décrire un ensemble fini comme suit : $E = \{x_1, \dots, x_n\}$ où $\#E = n$.

Lorsque l'ensemble n'est pas de cardinal fini, on le décrit au travers d'une propriété satisfaite par chacun de ses éléments et uniquement par ses éléments. Par exemple, l'ensemble $[0; 1]$ est l'ensemble des nombres réels x tels que $0 \leq x \leq 1$.

Lorsque l'on aborde la théorie de la mesure, il n'est pas surprenant de voir arriver des ensembles d'ensembles. On parle ici plutôt de classes d'ensembles. Supposons ainsi que l'on dispose d'une classe d'ensembles $(A_i)_{i \in I}$ décrite en fonction d'un indice $i \in I$.

Alors, on peut regarder l'intersection de toutes ces familles de même que leur réunion :

$$\bigcap_{i \in I} A_i := \{E : E \in A_i \text{ pour tout } i \in I\},$$

et

$$\bigcup_{i \in I} A_i := \{E : E \in A_{i_0} \text{ pour au moins un } i_0 \in I\}.$$

Le produit cartésien de n ensembles A_1, \dots, A_n est l'ensemble des n -uplets (a_1, \dots, a_n) tels que $a_i \in A_i$.

La borne supérieure d'un ensemble (de réels) majoré correspond au plus petit de ses majorants tandis que la borne inférieure d'un ensemble (de réels) minoré correspond au plus grand de ses minorants.

23.3 Suites

On introduit ici la notion de limite supérieure et celle de limite inférieure.

On considère une suite de réels $(x_n)_{n \in \mathbb{N}}$. Alors, on pose pour tout $p \in \mathbb{N}$:

$$y_p := \sup_{k \geq p} x_k.$$

On peut aussi écrire

$$y_p := \sup E_p,$$

où $E_p := \{x_k : k \geq p\}$. On remarque $E_{p+1} \subset E_p$ pour tout $p \geq 0$. Par conséquent,

$$y_{p+1} \leq y_p.$$

La suite $(y_p)_{p \in \mathbb{N}}$ est ainsi décroissante et on peut donc définir son infimum (qui peut éventuellement être égal à $-\infty$).

On pose alors :

$$\limsup_{n \rightarrow +\infty} x_n = \lim_{p \rightarrow +\infty} y_p = \inf_{p \in \mathbb{N}} y_p.$$

Cette quantité porte le nom de limite supérieure de la suite.

On peut aussi définir la limite inférieure. Pour ce faire, on pose :

$$z_p := \inf_{k \geq p} x_k.$$

On peut aussi écrire

$$z_p := \inf E_p,$$

où $E_p := \{x_k : k \geq p\}$. On remarque cette fois $E_{p+1} \subset E_p$ pour tout $p \geq 0$. Par conséquent,

$$z_{p+1} \geq z_p.$$

La suite $(z_p)_{p \in \mathbb{N}}$ est ainsi croissante et on peut donc définir son supremum (qui peut éventuellement être égal à $+\infty$).

On pose alors :

$$\liminf_{n \rightarrow +\infty} x_n = \lim_{p \rightarrow +\infty} z_p = \sup_{p \in \mathbb{N}} z_p.$$

Cette quantité porte le nom de limite inférieure de la suite.

On remarque

$$\liminf_{n \rightarrow +\infty} x_n \leq \limsup_{n \rightarrow +\infty} x_n.$$

Par ailleurs, la suite est convergente si et seulement si $\liminf_{n \rightarrow +\infty} x_n = \limsup_{n \rightarrow +\infty} x_n$.

23.4 Fonctions

Une fonction f d'un ensemble E vers un ensemble F est une application (ou une transformation) qui, à chaque élément de E fait correspondre un élément de F .

Si $A \subset E$, on pose

$$f(A) := \{y \in F : \exists x \in A \text{ tel que } y = f(x)\} \subset F.$$

L'ensemble E est appelé le domaine de f tandis que l'ensemble $f(E)$ est son image.

Si $B \subset F$, on pose

$$f^{-1}(B) := \{x \in E : f(x) \in B\} \subset E.$$

Si f est une fonction d'un ensemble E vers \mathbb{R} , on introduit les fonctions f^+ et f^- par

$$f^+(x) := \max\{f(x); 0\} \quad \text{et} \quad f^-(x) := \max\{-f(x); 0\}.$$

On remarque alors que ces deux fonctions sont à valeurs positives ou nulles. De plus :

$$|f(x)| = f^+(x) + f^-(x) \quad \text{et} \quad f(x) = f^+(x) - f^-(x).$$

La fonction f^+ est appelée partie positive de f tandis que f^- est sa partie négative.

23.5 Topologie

La topologie générale est plus subtile que la topologie des espaces métriques. Néanmoins, recourir à la généralité permet de faire le pont avec la théorie de la mesure. En effet, les tribus et les topologies satisfont des propriétés similaires.

On se donne un ensemble Ω . Alors, une topologie sur Ω est une collection \mathcal{T} de sous-ensembles de Ω telle que :

- \mathcal{T} contient les sous-ensembles triviaux : $\Omega \in \mathcal{T}$ et $\emptyset \in \mathcal{T}$.
- \mathcal{T} est stable par intersection finie : si $\mathcal{O}_i \in \mathcal{T}$ pour tout $i \in \llbracket 1; n \rrbracket$, alors $(\mathcal{O}_1 \cap \dots \cap \mathcal{O}_n) \in \mathcal{T}$.
- \mathcal{T} est stable par réunion quelconque : si $\mathcal{O}_i \in \mathcal{T}$ pour tout $i \in I$ où I est un ensemble *quelconque* d'indices, alors $(\bigcup_{i \in I} \mathcal{O}_i) \in \mathcal{T}$.

Il est ici essentiel de bien comprendre que l'on ne suppose pas $\emptyset \subset \mathcal{T}$ (ce qui est une trivialité) mais bien $\emptyset \in \mathcal{T}$. De même, si \mathcal{U} est une collection de sous-ensembles de Ω , on n'a pas nécessairement la stabilité par intersection finie.

Prenons par exemple $\Omega := \{1; 2; 3; 4\}$. Alors, si $\mathcal{U} = \{\emptyset; \{1; 2; 3\}; \{2; 3; 4\}; \Omega\}$, il est immédiat que \mathcal{U} ne contient pas $\{2; 3\}$ alors même que $\{2; 3\} = \{1; 2; 3\} \cap \{2; 3; 4\}$.

Si \mathcal{T} est une topologie sur Ω , on dit que (Ω, \mathcal{T}) est un espace topologique. La plupart du temps, on fera l'abus de langage consistant à omettre la topologie : ainsi Ω est un espace topologique. Néanmoins, il convient de garder à l'esprit qu'un même ensemble peut admettre différentes topologies.

Si Ω est un espace topologique (topologie : \mathcal{T}), on dit que les éléments de \mathcal{T} , dont on rappelle qu'il s'agit d'ensembles, sont appelés des ouverts.

De la topologie, on en déduit la notion de convergence. Si Ω_1 et Ω_2 sont des espaces topologiques, on dit que f est une application continue de Ω_1 dans Ω_2 si $f^{-1}(\mathcal{G})$ est un ouvert de Ω_1 pour tout ouvert \mathcal{G} de Ω_2 .

Cette dernière définition de la continuité est consistante et cohérente avec celle donnée dans le cas des espaces métriques (le fameux "pour tout $\epsilon > 0$, il existe").

23.6 Intégration de Riemann

L'intégration de Riemann permet de définir la notion d'intégrale pour des fonctions continues par morceaux sur un intervalle fermé et borné $[a; b]$.

L'idée est de définir l'intégrale des fonctions en escaliers. La fonction f est une fonction en escalier s'il existe x_1, \dots, x_n satisfaisant :

$$a < x_1 < x_2 < \dots < x_n < b,$$

et en posant $x_0 := a$ ainsi que $x_{n+1} := b$, alors pour tout $i \in \llbracket 0; n \rrbracket$, la fonction f est constante sur l'intervalle $]x_i; x_{i+1}[$ et vaut $\alpha_i \in \mathbb{R}$.

L'intégrale de f sur $[a; b]$ est alors

$$\int_{[a;b]} f := \sum_{i=0}^n \alpha_i (x_{i+1} - x_i).$$

Or, toute fonction continue par morceaux peut être approchée par des fonctions en escalier. Ainsi, on peut construire leur intégrale.

Puis, en utilisant l'algèbre des fonctions en escalier comme élément de base, on en déduit la linéarité de l'intégrale au sens de Riemann.

Éléments de théorie de la mesure

24.1 Introduction

Le chapitre précédent est un pré-requis à l'étude de celui-ci. Également, il peut être pertinent d'avoir étudié tout le livre avant de le reprendre avec le point de vue "théorie de la mesure" ; lequel est plus rigoureux.

L'objectif de ce chapitre est avant tout de clarifier la plupart des passages du livre dans lesquels nous avons triché. Notamment, nous ajoutons de la rigueur à la définition d'une probabilité, des variables aléatoires... La théorie de la mesure peut être ardue voire inutile en première lecture. Néanmoins, elle devient nécessaire dès que l'on va plus loin que les bases des probabilités. Ainsi, il n'est pas envisageable de comprendre les processus stochastiques sans avoir fait au préalable la théorie de la mesure. La notion de tribu est abordée très succinctement en classes préparatoires et pas du tout en DUT.

Le principal intérêt des tribus est que certaines mesures ne peuvent pas être définies proprement sur 2^Ω . Il est ainsi facile de construire des ensembles non boréliens et même des ensembles non Lebesguiens pour lesquels la mesure de Lebesgue ne peut pas être définie.

Il est donc crucial de se restreindre à une classe d'ensembles telle que l'on puisse y définir la mesure en question : c'est la tribu.

Les attendus à la fin du chapitre sont de pouvoir identifier dans le reste du livre tous les endroits où nous avons triché.

24.2 Tribus

La théorie de la mesure est fondée sur la notion de tribus. En effet, dans le Chapitre 2, nous avons modélisé les événements comme étant les sous-ensembles de l'univers Ω . Et, nous avons muni Ω d'une probabilité c'est-à-dire d'une application de 2^Ω dans $[0; 1]$ satisfaisant de bonnes propriétés. Néanmoins, l'existence d'une telle application peut présupposer qu'on ne la définit que sur un sous-ensemble de 2^Ω . Ce sous-ensemble est justement une tribu.

Définition 24.2.1. *On dit que $\mathcal{A} \subset 2^\Omega$ est une algèbre de Boole si elle vérifie les trois propriétés suivantes :*

- $\Omega \in \mathcal{A}$.
- Pour tout $A \in \mathcal{A}$, $A^c \in \mathcal{A}$.
- Pour tout $A, B \in \mathcal{A}$, $A \cup B \in \mathcal{A}$.

Remarque 24.2.2. *Il est assez facile de prouver que si \mathcal{A} est une algèbre de Boole, alors $\emptyset \in \mathcal{A}$.*

Remarque 24.2.3. *Il est tout aussi facile de prouver que pour tout $n \in \mathbb{N}$, pour tout $A_1, \dots, A_n \in \mathcal{A}$, on a $\bigcup_{i=1}^n A_i \in \mathcal{A}$ et $\bigcap_{i=1}^n A_i \in \mathcal{A}$.*

Exercice 24.2.4. *Démontrer les deux remarques.*

La notion d'algèbre de Boole a le mauvais goût de ne pas exploiter la dénombrabilité. En effet, nous avons considéré des suites de variables aléatoires. Et donc, pour pouvoir mesurer les événements relatifs à cette suite de variables aléatoires, il nous faut la stabilité par réunion dénombrable.

On aboutit donc à la notion de tribu.

Définition 24.2.5. *On dit que $\mathcal{F} \subset 2^\Omega$ est une tribu si elle vérifie les trois propriétés suivantes :*

- $\Omega \in \mathcal{F}$.
- Pour tout $A \in \mathcal{F}$, $A^c \in \mathcal{F}$.
- Pour tout $(A_i)_{i \in \mathcal{I}}$ où $A_i \in \mathcal{F}$ et \mathcal{I} est un ensemble fini ou infini dénombrable d'indices, $\bigcup_{i \in \mathcal{I}} A_i \in \mathcal{F}$.

Remarque 24.2.6. *Une tribu est une algèbre de Boole stable par réunion dénombrable.*

Remarque 24.2.7. *Une tribu est aussi stable par intersection dénombrable. En effet, $\bigcap_{i \in \mathcal{I}} A_i = \left(\bigcup_{i \in \mathcal{I}} A_i^c\right)^c$ d'après les lois de Morgan, voir le Théorème 1.3.33 à la page 17. Puis, A_i^c est un élément de \mathcal{F} donc sa réunion dénombrable aussi. On passe encore à la complémentation et la preuve est achevée.*

Définition 24.2.8. *On appelle "événement" tout élément de la tribu.*

Exemple 24.2.9. *L'ensemble 2^Ω est une tribu sur Ω .*

On peut aussi remarquer qu'une tribu est un π -système et un λ -système :

Proposition 24.2.10. *Soit $\mathcal{F} \subset 2^\Omega$. C'est une tribu si elle vérifie les deux propriétés suivantes :*

- \mathcal{F} est un π -système c'est-à-dire : pour tout $A, B \in \mathcal{F}$, $A \cap B \in \mathcal{F}$.
- \mathcal{F} est un λ -système (ou une classe monotone) c'est-à-dire :
 - $\Omega \in \mathcal{F}$.
 - Pour tout $A, B \in \mathcal{F}$, alors $A \subset B$ implique $B \cap A^c \in \mathcal{F}$.
 - Pour toute suite d'événements $(A_n)_n$ avec $A_n \subset A_{n+1}$, on a $\bigcup_{n=1}^\infty A_n \in \mathcal{F}$.

Exemple 24.2.11. Si $\Omega := \mathbb{R}$, alors l'ensemble des demi-droites de la forme $] -\infty; a]$ avec $a \in \mathbb{R}$ est un π -système.

Remarque 24.2.12. Le précédent π -système engendre la tribu des boréliens et c'est pourquoi la fonction de répartition F_X caractérise la loi de probabilité de la variable aléatoire X .

Donnons maintenant la preuve de la Proposition 24.2.10.

Démonstration. Soit \mathcal{F} qui est un π -système et un λ -système. D'abord, le premier axiome est immédiatement satisfait puisque $\Omega \in \mathcal{F}$. Ensuite, soit $A \in \mathcal{F}$. Alors, $A \subset \Omega \in \mathcal{F}$ d'où $A^c = \Omega \cap A^c \in \mathcal{F}$. Puis, comme $A \cap B \in \mathcal{F}$ pour tout $A, B \in \mathcal{F}$, il s'ensuit que \mathcal{F} est stable par réunion finie. Soit maintenant $(A_n)_n$ une suite d'évènements dans \mathcal{F} . Alors, on pose $B_n := \bigcup_{i=1}^n A_i$. Cette suite est croissante c'est-à-dire que l'on a $B_n \subset B_{n+1}$ pour tout $n \geq 1$. Donc $\bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$. Or, $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} \bigcup_{i=1}^n A_i = \bigcup_{i=1}^{\infty} A_i$. Par conséquent, une classe d'évènements \mathcal{F} qui est à la fois un π -système et un λ -système est une tribu.

Prouvons maintenant la réciproque. Soit donc \mathcal{F} une tribu. Alors, elle est stable par intersection finie ce qui en fait un π -système. Puis, $\Omega \in \mathcal{F}$. Ensuite, si $A, B \in \mathcal{F}$ avec $A \subset B$, alors $B \cap A^c \in \mathcal{F}$ vu que $A^c \in \mathcal{F}$. Enfin, si on considère une suite croissante d'évènements dans \mathcal{F} , sa réunion appartient à \mathcal{F} vu que \mathcal{F} est stable par réunion dénombrable. La preuve est ainsi achevée. \square

On peut être amené à considérer des sous-tribus.

Définition 24.2.13. Si \mathcal{F} est une tribu sur Ω et si $\mathcal{F}' \subset \mathcal{F}$ est une tribu, on dit que \mathcal{F}' est une sous-tribu de \mathcal{F} .

Remarque 24.2.14. Il est crucial de bien comprendre qu'une sous-tribu de \mathcal{F} n'est pas nécessairement constituée de sous-ensembles d'éléments de \mathcal{F} . Ainsi, si $\mathcal{F}' \subset \mathcal{F}$ et si \mathcal{F} comme \mathcal{F}' sont deux tribus de Ω , alors $B \subset A$ avec $A \in \mathcal{F}$ n'implique ni $B \in \mathcal{F}'$ ni $B \in \mathcal{F}$.

On va maintenant effectuer des opérations sur les tribus. On va d'abord voir les intersections de tribus.

Proposition 24.2.15. Soit Ω un ensemble non-vide. Soit \mathcal{T} un ensemble de tribus sur Ω . En d'autres termes, chaque élément de \mathcal{T} est une tribu sur Ω c'est-à-dire un ensemble de sous-ensembles de Ω . On suppose que $\mathcal{T} \neq \emptyset$. Alors, $\mathcal{A} := \bigcap_{\mathcal{F} \in \mathcal{T}} \mathcal{F}$ est une tribu.

On note que l'on ne suppose ni que \mathcal{T} est fini ni qu'il est infini dénombrable.

Il convient de noter que \mathcal{T} est un ensemble d'ensembles d'ensembles en tant qu'ensemble d'ensembles de sous-ensembles de Ω .

Démonstration. Vérifions les trois axiomes de base d'une tribu. D'abord, $\Omega \in \mathcal{F}$ pour tout $\mathcal{F} \in \mathcal{T}$ donc $\Omega \in \mathcal{A}$. Ensuite, soit $A \in \mathcal{A}$ alors $A \in \mathcal{F}$ pour tout $\mathcal{F} \in \mathcal{T}$. D'où $A^c \in \mathcal{F}$ pour tout $\mathcal{F} \in \mathcal{T}$ si bien que $A^c \in \mathcal{A}$. Soit maintenant une famille dénombrable $(A_i)_{i \in \mathcal{I}}$ d'éléments de \mathcal{A} . Alors, $A_i \in \mathcal{F}$ pour tout $\mathcal{F} \in \mathcal{T}$ d'où pour tout \mathcal{F} , on a $\bigcup_{i \in \mathcal{I}} A_i \in \mathcal{F}$ et par conséquent, $\bigcup_{i \in \mathcal{I}} A_i \in \mathcal{A}$. \square

Voyons maintenant ce qu'est la tribu engendrée par une famille de tribus.

Définition 24.2.16. Soit $(\mathcal{F}_i)_{i \in \mathcal{I}}$ une famille de tribus sur Ω . On note

$$\sigma(\mathcal{F}_i; i \in \mathcal{I})$$

la plus petite tribu (au sens de l'inclusion) qui contient toutes les tribus \mathcal{F}_i .

Proposition 24.2.17. Cette tribu $\sigma(\mathcal{F}_i; i \in \mathcal{I})$ existe.

Démonstration. L'ensemble \mathcal{T} des tribus qui contiennent toutes les tribus \mathcal{F}_i est non vide puisqu'il contient 2^Ω . Ainsi, on peut faire l'intersection de toutes les tribus \mathcal{F} dans \mathcal{T} . \square

Remarque 24.2.18. Si \mathcal{F} est une tribu, alors $\sigma(\mathcal{F}) = \mathcal{F}$.

On peut aller plus loin en considérant la tribu engendrée par une famille d'ensembles.

Proposition 24.2.19. Soit $(A_i)_{i \in \mathcal{I}}$ une famille de parties de Ω (c'est-à-dire que l'on a $A_i \subset \Omega$). Pour tout $i \in \mathcal{I}$, la plus petite tribu contenant A_i est la tribu $\mathcal{F}_i := \{\emptyset, A_i, A_i^c, \Omega\}$. Et, la plus petite tribu contenant tous les ensembles A_i est définie par $\sigma(\mathcal{F}_i; i \in \mathcal{I})$. Cette tribu sera notée $\sigma(A_i; i \in \mathcal{I})$. On remarque au passage $\sigma(\mathcal{F}_i; i \in \mathcal{I}) = \sigma(\bigcup_{i \in \mathcal{I}} \mathcal{F}_i)$.

24.3 Fonctions mesurables

Définition 24.3.1. Soit Ω un ensemble muni d'une tribu \mathcal{F} . On dit que (Ω, \mathcal{F}) est un espace mesurable.

Définition 24.3.2. Soient $(\Omega_1, \mathcal{F}_1)$ et $(\Omega_2, \mathcal{F}_2)$ deux espaces mesurables. On dit que l'application f de Ω_1 dans Ω_2 est mesurable si quel que soit $A \in \mathcal{F}_2$, alors $f^{-1}(A) := \{x \in \Omega_1 : f(x) \in A\} \in \mathcal{F}_1$.

On remarquera la similitude entre cette définition et celle de la continuité dans la topologie générale.

Remarque 24.3.3. Si f est une application mesurable de $(\Omega_1, \mathcal{F}_1)$ dans $(\Omega_2, \mathcal{F}_2)$ et si g est une application mesurable de $(\Omega_2, \mathcal{F}_2)$ dans $(\Omega_3, \mathcal{F}_3)$, alors $g \circ f$ est une application mesurable de $(\Omega_1, \mathcal{F}_1)$ dans $(\Omega_3, \mathcal{F}_3)$.

Théorème 24.3.4 (Théorème fondamental de la mesurabilité). Soit f une application quelconque de Ω_1 dans Ω_2 . Alors, pour toute tribu \mathcal{F}_2 sur Ω_2 , l'ensemble $f^{-1}(\mathcal{F}_2) := \{f^{-1}(A) : A \in \mathcal{F}_2\}$ est une tribu sur Ω_1 . Et, pour tout $\mathcal{A} \subset 2^{\Omega_2}$, on a $\sigma(f^{-1}(\mathcal{A})) = f^{-1}(\sigma(\mathcal{A}))$.

Le théorème est admis.

Définition 24.3.5. La tribu $\sigma(f^{-1}(\mathcal{A})) = f^{-1}(\sigma(\mathcal{A}))$ est appelée la tribu image de \mathcal{A} par f .

Remarque 24.3.6. S'il n'y a pas d'ambiguïté sur la tribu \mathcal{F}_2 de l'espace d'arrivée Ω_2 , on note $\sigma(f) := f^{-1}(\mathcal{F}_2)$. C'est aussi la plus petite tribu sur Ω_1 qui rend l'application f mesurable. On dira que $\sigma(f)$ est la tribu engendrée par f .

Corollaire 24.3.7. Soient $(\Omega_1, \mathcal{F}_1)$ et $(\Omega_2, \mathcal{F}_2)$ deux espaces mesurables. Soit $\mathcal{C} \subset \mathcal{F}_2$ tel que $\mathcal{F}_2 = \sigma(\mathcal{C})$. Alors, f est une application mesurable de $(\Omega_1, \mathcal{F}_1)$ dans $(\Omega_2, \mathcal{F}_2)$ si et seulement si pour tout $C \in \mathcal{C}$, $f^{-1}(C) \in \mathcal{F}_1$.

24.4 Tribu borélienne

Quand une tribu rencontre une topologie, on obtient un Borélien. Pour être plus rigoureux :

Définition 24.4.1. Si Ω est un ensemble muni d'une topologie, la tribu borélienne est la tribu engendrée par les ouverts de cette topologie.

Notation 24.4.2. La tribu borélienne est notée $\mathcal{B}(\Omega)$.

Remarque 24.4.3. Si l'espace d'arrivée est $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on parle généralement d'application mesurable en omettant de préciser l'espace d'arrivée.

Lemme 24.4.4. Soient $(\Omega_1, \mathcal{T}_1)$ et $(\Omega_2, \mathcal{T}_2)$ deux espaces topologiques. Alors toute application continue de $(\Omega_1, \mathcal{T}_1)$ dans $(\Omega_2, \mathcal{T}_2)$ est mesurable de $(\Omega_1, \mathcal{B}(\Omega_1))$ dans $(\Omega_2, \mathcal{B}(\Omega_2))$.

Démonstration. En effet, par définition de la continuité, pour tout ouvert \mathcal{O}_2 de Ω_2 , $f^{-1}(\mathcal{O}_2)$ est un ouvert de Ω_1 donc un élément de la tribu $\mathcal{B}(\Omega_1)$. On applique ensuite le Corollaire 24.3.7. \square

Théorème 24.4.5. Soit f une application de Ω (muni de la tribu \mathcal{F}) dans \mathbb{R} muni de sa tribu borélienne. Alors, f est mesurable si et seulement si pour tout $a \in \mathbb{R}$, l'ensemble $f^{-1}(] - \infty; a])$ est dans \mathcal{F} .

Démonstration. Il suffit pour s'en convaincre de prouver que l'ensemble des intervalles de la forme $] - \infty; a]$ engendrent la tribu borélienne. Soit ainsi \mathcal{M} la tribu engendrée par ces intervalles.

Étape 1. En passant au complémentaire, il est immédiat que tout intervalle de la forme $]a; +\infty[$ est dans cette tribu \mathcal{M} .

Étape 2. Soient $a < b$. Alors, $]a; b] =] - \infty; b] \cap]a; +\infty[$. Comme une tribu est stable par intersection finie en tant qu'algèbre de Boole, l'intervalle $]a; b]$ est bien inclus dans \mathcal{M} .

Étape 3. Soient $a < b$. Alors, on remarque :

$$]a; b[= \bigcap_{n \in \mathbb{N}}]a; b - \frac{1}{2^n}[.$$

Une tribu étant stable par réunion dénombrable, il s'ensuit que tout intervalle ouvert borné est inclus dans la tribu \mathcal{M} .

Étape 4. Soit maintenant un ouvert borné quelconque \mathcal{O} . On note $\mathcal{R}_{\mathcal{O}}$ l'ensemble des intervalles ouverts bornés $]q; r[$ dont les bornes q et r sont des rationnels et tels que $]q; r[\subset \mathcal{O}$. L'ensemble $\mathcal{R}_{\mathcal{O}}$ est clairement fini ou infini dénombrable puisqu'en bijection avec $\{(q, r) \in \mathbb{Q}^2 : q < r \text{ et }]q; r[\subset \mathcal{O}\} \subset \mathbb{Q}^2$.

On pose

$$\mathcal{G} := \bigcup_{\mathcal{I} \in \mathcal{R}_{\mathcal{O}}} \mathcal{I}.$$

Alors, $\mathcal{G} \subset \mathcal{O}$. Montrons maintenant $\mathcal{O} \subset \mathcal{G}$. Soit $x \in \mathcal{O}$. Comme \mathcal{O} est ouvert, il existe $\epsilon > 0$ tel que $]x - \epsilon; x + \epsilon[\subset \mathcal{O}$. L'ensemble des rationnels étant dense dans celui des réels, on en déduit qu'il existe $q, r \in \mathbb{Q}$ tels que

$$x - \epsilon < q < x < r < x + \epsilon.$$

Ainsi, $\mathcal{I}_0 :=]q; r[\subset]x - \epsilon; x + \epsilon[\subset \mathcal{O}$. L'intervalle \mathcal{I}_0 est donc un élément de $\mathcal{R}_{\mathcal{O}}$. Conséquemment, $\mathcal{I}_0 \subset \mathcal{G}$. Puis l'on en déduit $x \in \mathcal{G}$. Ceci implique : $\mathcal{O} \subset \mathcal{G}$ d'où $\mathcal{O} = \mathcal{G}$.

Toutefois, l'ouvert $\mathcal{G} = \mathcal{O}$ est une réunion finie ou infinie dénombrable d'éléments de \mathcal{M} . Il s'ensuit que tout ouvert borné est un élément de la tribu \mathcal{M} .

Étape 5. Soit maintenant un ouvert quelconque \mathcal{O} , non nécessairement borné. Alors :

$$\mathcal{O} = \bigcup_{n \in \mathbb{N}} \mathcal{O}_n,$$

où $\mathcal{O}_n := \mathcal{O} \cap]-n; n[$. La tribu \mathcal{M} étant stable par réunion dénombrable, l'ouvert \mathcal{O} est bien dans \mathcal{M} . □

24.5 Mesures

Dans cette section, on s'intéresse aux mesures et plus précisément aux mesures positives; ce qu'est toute probabilité.

Définition 24.5.1. Une mesure positive μ est une fonction définie sur une tribu \mathcal{F} à valeurs dans $[0; +\infty]$ qui possède la propriété d'additivité dénombrable.

En d'autres termes, pour toute famille $(A_i)_{i \in I}$ d'éléments disjoints de \mathcal{F} où I est un ensemble fini ou infini dénombrable d'indices, on a

$$\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i),$$

où la somme précédente est éventuellement infinie.

Remarque 24.5.2. Dans la définition précédente, on demande l'additivité dénombrable au sens où l'additivité est vraie pour une réunion finie ou une réunion infinie dénombrable. Il convient de noter que l'on pourrait se contenter de supposer l'additivité quand la réunion est infinie dénombrable.

Remarque 24.5.3. Par la suite, il arrivera que l'on omette de préciser que la mesure est finie. En effet, on ne s'intéresse pas aux mesures signées dans cet ouvrage.

Définition 24.5.4. Le triplet $(\Omega, \mathcal{F}, \mu)$ est un espace mesuré tandis que (Ω, \mathcal{F}) est un espace mesurable.

Définition 24.5.5. Si $\mu(\Omega) = 1$, alors μ est une mesure de probabilité et le triplet $(\Omega, \mathcal{F}, \mu)$ est un espace probabilisé.

Théorème 24.5.6. Soit μ une mesure positive sur une tribu \mathcal{F} . Alors :

1. $\mu(\emptyset) = 0$.
2. Si $A, B \in \mathcal{F}$ avec $A \subset B$, alors $\mu(A) \leq \mu(B)$.
3. Si $(A_n)_{n \in \mathbb{N}}$ est une suite croissante d'évènements alors $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A_\infty)$ où $A_\infty := \bigcup_{n \in \mathbb{N}} A_n$.
4. Si $(A_n)_{n \in \mathbb{N}}$ est une suite décroissante d'évènements et si de plus $\mu(A_0) < \infty$ alors $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A_\infty)$ où $A_\infty := \bigcap_{n \in \mathbb{N}} A_n$.

Démonstration. Pour le premier point, il suffit de prendre $A_1 = A_2 = \emptyset$. Il vient immédiatement

$$\mu(\emptyset) = \mu(\emptyset \cup \emptyset) = \mu(\emptyset) + \mu(\emptyset),$$

d'où $\mu(\emptyset) = 0$.

Concernant le deuxième point, il suffit d'utiliser $B = (B \cap A) \cup (B \cap A^c)$ et de remarquer que $\mu(B \cap A^c) \geq 0$.

Le troisième point est immédiat après que l'on a défini $B_n := A_n \cap A_{n-1}^c$ pour $n \geq 1$ et $B_0 := A_0$. On a alors $B_n \in \mathcal{F}$ et il s'agit d'une suite d'évènements deux à deux disjoints dont la réunion est telle que

$$A_\infty = \bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} B_n,$$

d'où

$$\mu(A_\infty) = \lim_{n \rightarrow \infty} \sum_{k=1}^n \mu(B_k).$$

Or, $\mu(B_k) + \mu(A_{k-1}) = \mu(A_k)$. Ainsi, avec la convention $A_{-1} = \emptyset$, on aboutit à

$$\sum_{k=1}^n \mu(B_k) = \mu(A_n),$$

et la limite annoncée s'ensuit.

En ce qui concerne le dernier point, on pose $C_n := A_0 \cap A_n^c$. On a alors une suite croissante d'évènements et de plus $A_0 \cap A_\infty^c = \bigcup_{n \in \mathbb{N}} C_n$ si bien que l'application du résultat du point précédent nous donne

$$\mu(A_0) - \mu(A_\infty) = \mu(A_0 \cap A_\infty^c) = \lim_{n \rightarrow \infty} \mu(C_n) = \mu(A_0) - \lim_{n \rightarrow \infty} \mu(A_n).$$

Comme $\mu(A_0) < \infty$, le résultat est immédiat. □

Nous allons maintenant nous intéresser aux ensembles négligeables. Il serait en effet bienvenue que si $B \in \mathcal{F}$ est tel que $\mu(B) = 0$ alors tout sous-ensemble de B est lui aussi de mesure nulle.

Néanmoins, ceci n'a aucune raison d'être vrai. En effet, la mesure est définie sur la tribu et rien ne nous assure que les sous-ensembles de $B \in \mathcal{F}$ tel que $\mu(B) = 0$ soient eux-mêmes des évènements (c'est-à-dire des éléments de la tribu).

C'est pourquoi on enrichit généralement la tribu considérée par les ensembles négligeables pour la mesure μ .

Définition 24.5.7. Soit un espace mesuré $(\Omega, \mathcal{F}, \mu)$. On dit que $A \subset \mathcal{F}$ est un ensemble négligeable s'il existe $B \in \mathcal{F}$ tel que $\mu(B) = 0$ et $A \subset B$.

Notation 24.5.8. L'ensemble des négligeables se note \mathcal{N} .

Définition 24.5.9. On appelle tribu complétée la plus petite tribu qui contient \mathcal{F} et \mathcal{N} . On la note généralement \mathcal{F}^* .

Comme on peut toujours se permettre de compléter la tribu considérée, il est par la suite peu restrictif de supposer que tout ensemble négligeable appartient à ladite tribu.

Gardons toutefois à l'esprit qu'un certain nombre de considérations techniques ont été passées sous silence. Notamment, si $E \in \mathcal{F}^*$, alors il existe $A, B \in \mathcal{F}$ tels que $A \subset E \subset B$ et $\mu(B \cap A^c) = 0$. Par ailleurs, on pose $\mu(E) = \mu(A) = \mu(B)$. Il est *crucial* que la mesure de E ne dépende pas du couple (A, B) ainsi défini.

On achève cette section par un théorème fondamental ; par lequel on prouve que la fonction de répartition caractérise entièrement la loi de probabilité.

Théorème 24.5.10. *Soit un ensemble Ω muni d'un π -système \mathcal{P} et soit \mathcal{F} la tribu engendrée. Supposons que l'on ait deux mesures μ_1 et μ_2 sur \mathcal{F} telles que $\mu_1(\Omega) = \mu_2(\Omega) < \infty$. Alors $\mu_1 = \mu_2$.*

La preuve utilise le théorème de Dynkin et donc on l'admet.

24.6 Intégration de Lebesgue

L'idée de l'intégrale au sens de Lebesgue est "simplement" de construire l'intégration à partir d'un espace plus fin que celui des fonctions en escalier.

Il s'agit de l'espace des fonctions étagées. Une telle fonction est définie sur un espace mesurable (Ω, \mathcal{F}) et son image (c'est-à-dire $f(\Omega)$) est un sous-ensemble fini de \mathbb{R} .

Notamment, on construit l'intégrale au sens de Lebesgue, en premier lieu du moins, pour les fonctions étagées positives c'est-à-dire les fonctions étagées à valeurs dans $[0; +\infty]$.

Soit s une fonction étagée positive. On note $\alpha_1, \dots, \alpha_n$ les différentes valeurs prises par la fonction s et $A_i := \{\omega \in \Omega : s(\omega) = \alpha_i\}$. Alors, on peut écrire

$$s(x) = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(x).$$

Cette fonction s est mesurable pour peu que chacun des ensembles A_i le soit aussi.

Théorème 24.6.1. *Soit une fonction f mesurable et à valeurs dans $[0; +\infty]$. Il existe une suite $(s_n)_{n \in \mathbb{N}}$ de fonctions étagées positives et mesurables telle que pour tout $\omega \in \Omega$, $s_i(\omega) \leq s_{i+1}(\omega)$ et de plus $\lim_{n \rightarrow \infty} s_n(\omega) = f(\omega)$.*

La preuve est omise.

Avec la convention $0 \times \infty = 0$, si s est une fonction étagée positive telle que

$$s = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i},$$

l'intégrale de Lebesgue de s par rapport à la mesure μ sur un élément $E \in \mathcal{F}$ est définie ainsi :

$$\int_E s d\mu := \sum_{i=1}^n \alpha_i \mu(A_i \cap E).$$

En particulier,

$$\int_{\Omega} s d\mu = \sum_{i=1}^n \alpha_i \mu(A_i).$$

La convention $0 \times \infty = 0$ prend ici tout son sens : si la fonction s est nulle et si E est un ensemble de mesure infinie, l'intégrale de s par rapport à μ sur cet ensemble E est nulle.

Définition 24.6.2. Si f est une fonction mesurable positive, son intégrale par rapport à μ sur un ensemble $E \in \mathcal{F}$ est

$$\int_E f d\mu := \sup \int_E s d\mu,$$

où le supremum est pris sur l'ensemble de toutes les fonctions étagées positives et mesurables s telles que $s \leq f$.

On dispose des premières propriétés suivantes :

- Si $0 \leq f \leq g$ alors $\int_E f d\mu \leq \int_E g d\mu$.
- Si $f \geq 0$ et $A \subset B$, alors $\int_A f d\mu \leq \int_B f d\mu$.
- Si $f \geq 0$ et si $c \in \mathbb{R}_+$, on a $\int_E c \times f d\mu = c \times \int_E f d\mu$.
- Si $f(x) = 0$ pour tout $x \in E$, alors $\int_E f d\mu = 0$ même si la mesure de E est infinie.
- Si $\mu(E) = 0$, alors $\int_E f d\mu = 0$ même si $f(x) = +\infty$ pour tout $x \in E$.
- Si $f(\omega) \geq 0$ pour tout $\omega \in \Omega$, alors $\int_E f d\mu = \int_\Omega \mathbf{1}_E f d\mu$.

On dispose aussi de la linéarité suivante :

$$\int_\Omega (s + t) d\mu = \int_\Omega s d\mu + \int_\Omega t d\mu.$$

Donnons maintenant les grands théorèmes de l'intégration.

Théorème 24.6.3 (Théorème de convergence monotone de Lebesgue). Soit une suite croissante $(f_n)_{n \in \mathbb{N}}$ de fonctions mesurables positives telles que $f_n(\omega) \rightarrow f(\omega)$ pour tout $\omega \in \Omega$ quand n tend vers l'infini. Alors la fonction f est elle aussi mesurable et de plus

$$\int_\Omega f_n d\mu \rightarrow \int_\Omega f d\mu,$$

dans la limite $n \rightarrow +\infty$.

Théorème 24.6.4. Si f_n est une fonction mesurable de Ω dans $[0; +\infty]$ pour tout $n \in \mathbb{N}$ alors la fonction f définie par

$$f(x) := \sum_{n=0}^{\infty} f_n(x)$$

est mesurable et de plus

$$\int_\Omega f d\mu = \sum_{n=0}^{\infty} \int_\Omega f_n d\mu.$$

Théorème 24.6.5 (Lemme de Fatou). *Si f_n est une fonction mesurable de Ω dans $[0; +\infty]$ pour tout $n \in \mathbb{N}$ alors $\liminf_{n \rightarrow \infty} f_n$ est mesurable et de plus*

$$\int_{\Omega} \left(\liminf_{n \rightarrow \infty} f_n \right) d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} f_n d\mu.$$

Voici une version abstraite de la formule de transfert :

Théorème 24.6.6. *Soit une fonction mesurable et positive f . On pose pour tout $E \in \mathcal{F}$:*

$$\Psi(E) := \int_E f d\mu.$$

Alors, Ψ est une mesure sur \mathcal{F} et

$$\int_{\Omega} g d\Psi = \int_{\Omega} g f d\mu,$$

pour toute fonction mesurable et positive g sur Ω .

On peut maintenant définir l'intégrale de Lebesgue par rapport à la mesure μ pour toute fonction mesurable pour peu que celle-ci soit intégrable.

Définition 24.6.7. *Par $L^1(\mu)$, notons l'ensemble de toutes les fonctions réelles mesurables f sur Ω telles que*

$$\int_{\Omega} |f| d\mu < +\infty.$$

Cette condition nous assure que

$$\int_{\Omega} f^+ d\mu < +\infty \quad \text{et} \quad \int_{\Omega} f^- d\mu < +\infty.$$

On peut alors définir l'intégrale de Lebesgue de f par

$$\int_{\Omega} f d\mu := \int_{\Omega} f^+ d\mu - \int_{\Omega} f^- d\mu.$$

Par construction, il vient :

Théorème 24.6.8 (Linéarité de l'intégrale). *Pour toutes les fonctions $f, g \in L^1(\mu)$ et pour toutes les constantes réelles α et β , la fonction $\alpha f + \beta g$ est un élément de $L^1(\mu)$. De plus :*

$$\int_{\Omega} (\alpha f + \beta g) d\mu = \alpha \int_{\Omega} f d\mu + \beta \int_{\Omega} g d\mu.$$

Un résultat crucial est l'inégalité triangulaire :

$$\left| \int_{\Omega} f d\mu \right| \leq \int_{\Omega} |f| d\mu,$$

pour toute fonction f dans $L^1(\mu)$.

Exercice 24.6.9. *Démontrer l'inégalité triangulaire.*

Théorème 24.6.10 (Théorème de convergence dominée de Lebesgue). *Soit une suite $(f_n)_{n \in \mathbb{N}}$ de fonctions réelles mesurables telles que $f(\omega) = \lim_{n \rightarrow +\infty} f_n(\omega)$ pour tout $\omega \in \Omega$. S'il existe une fonction $g \in L^1(\mu)$ telle que pour tout $n \in \mathbb{N}$ et pour tout $\omega \in \Omega$, on ait*

$$|f_n(\omega)| \leq g(\omega).$$

Alors f est une fonction intégrable et de plus

$$\lim_{n \rightarrow +\infty} \int_{\Omega} |f_n - f| d\mu = 0,$$

si bien que

$$\lim_{n \rightarrow +\infty} \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu$$

24.7 Retour sur les chapitres précédents

De fait, les premiers chapitres de cet ouvrage doivent se lire comme suit.

Une probabilité sur un ensemble Ω n'est pas une application de 2^{Ω} dans $[0; 1]$. Il s'agit d'une application de la tribu \mathcal{F} pour laquelle cette mesure de probabilité est bien définie. Le côté bien défini ou non est entièrement passé sous silence de par sa difficulté.

Une variable aléatoire X n'est pas seulement une application de Ω dans \mathbb{R} . En effet, il faut de plus que l'image réciproque de tout borélien de \mathbb{R} par X soit un élément de la tribu \mathcal{F} dont est muni l'ensemble Ω .

Le théorème d'unicité des mesures pour peu qu'elles coïncident sur un π -système plus le résultat stipulant que le π -système des intervalle de la forme $] -\infty; a]$ engendre la tribu borélienne achèvent de prouver que si deux mesures de probabilité (donc de même masse totale égale à 1, par axiome) coïncident sur ces intervalles (ce qui définit la fonction de répartition), alors ces deux mesures sont les mêmes sur toute la tribu.

Notamment, il s'ensuit que la donnée d'une fonction de répartition d'une variable aléatoire X est équivalente à la donnée de la loi de probabilité de cette même variable aléatoire.

Enfin, la linéarité de l'espérance est un simple corollaire de la linéarité de l'intégrale par rapport à la mesure de probabilité \mathbb{P} .

Puis, on peut tout étendre au cas multi-dimensionnel et faire le pont entre les variables aléatoires discrètes et les variables aléatoires à densité. À vrai dire, on peut même traiter tous les types possibles et imaginables de variables aléatoires réelles et de vecteurs aléatoires.

Tables des lois de probabilité usuelles

Dans ce chapitre, nous donnons les tables des principales lois de probabilité.

Tables des lois discrètes

Pour commencer, on donne les tables de deux lois discrètes : la loi binomiale et la loi de Poisson.

Les tables de la loi binomiale donnent les probabilités que la variable aléatoire X (suivant ladite loi binomiale) soit égale à k : p_k . La fonction de répartition est également tabulée : $\sum_{i=0}^k p_i$.

Les différents paramètres des lois binomiales sont pour $n = 10, n = 20, n = 30, n = 40$ ainsi que pour les valeurs suivantes de p : 0.05, 0.1, 0.2, 0.3, 0.4 et 0.5.

La précision est de quatre chiffres significatifs. Il convient donc de noter que les valeurs 0.0000 ou 1.0000 sont des approximations.

Concernant la loi de Poisson, celle-ci n'a qu'un paramètre $\lambda > 0$ et les tables sont données pour $\lambda = 0.1, \lambda = 0.2, \lambda = 0.3, \lambda = 0.4, \lambda = 0.5, \lambda = 0.6, \lambda = 0.7, \lambda = 0.8, \lambda = 0.9, \lambda = 1, \lambda = 2, \lambda = 3, \lambda = 4, \lambda = 5, \lambda = 6, \lambda = 7, \lambda = 8, \lambda = 9, \lambda = 10, \lambda = 11, \lambda = 12, \lambda = 13, \lambda = 14, \lambda = 15, \lambda = 16$.

TABLE 25.1 – Table de la loi binomiale avec $p = 0.05$

k	$n = 10$		$n = 20$		$n = 30$		$n = 40$		$n = 50$	
	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$
0	0.5987	0.5987	0.3585	0.3585	0.2146	0.2146	0.1285	0.1285	0.0769	0.0769
1	0.3151	0.9138	0.3773	0.7358	0.3389	0.5535	0.2706	0.3991	0.2025	0.2794
2	0.0746	0.9884	0.1887	0.9245	0.2587	0.8122	0.2777	0.6768	0.2611	0.5405
3	0.0105	0.9989	0.0596	0.9841	0.1270	0.9392	0.1851	0.8619	0.2199	0.7604
4	0.0010	0.9999	0.0133	0.9974	0.0451	0.9843	0.0901	0.9520	0.1360	0.8964
5	0.0001	1.0000	0.0023	0.9997	0.0124	0.9967	0.0341	0.9861	0.0658	0.9622
6	0.0000	1.0000	0.0003	1.0000	0.0027	0.9994	0.0105	0.9966	0.0260	0.9882
7	0.0000	1.0000	0.0000	1.0000	0.0005	0.9999	0.0027	0.9993	0.0086	0.9968
8	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000	0.0006	0.9999	0.0024	0.9992
9	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000	0.0007	0.9999
10	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000

TABLE 25.2 – Table de la loi binomiale avec $p = 0.10$

k	$n = 10$		$n = 20$		$n = 30$		$n = 40$		$n = 50$	
	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$
0	0.3487	0.3487	0.1216	0.1216	0.0424	0.0424	0.0148	0.0148	0.0052	0.0052
1	0.3874	0.7361	0.2701	0.3917	0.1413	0.1837	0.0657	0.0805	0.0286	0.0338
2	0.1937	0.9298	0.2852	0.6769	0.2277	0.4114	0.1423	0.2228	0.0779	0.1117
3	0.0574	0.9872	0.1901	0.8670	0.2360	0.6474	0.2003	0.4231	0.1386	0.2503
4	0.0112	0.9984	0.0898	0.9568	0.1771	0.8245	0.2059	0.6290	0.1809	0.4312
5	0.0015	0.9999	0.0319	0.9887	0.1023	0.9268	0.1647	0.7937	0.1849	0.6161
6	0.0001	1.0000	0.0089	0.9976	0.0474	0.9742	0.1068	0.9005	0.1541	0.7702
7	0.0000	1.0000	0.0020	0.9996	0.0180	0.9922	0.0576	0.9581	0.1077	0.8779
8	0.0000	1.0000	0.0003	0.9999	0.0058	0.9980	0.0264	0.9845	0.0642	0.9421
9	0.0000	1.0000	0.0001	1.0000	0.0015	0.9995	0.0104	0.9949	0.0334	0.9755
10	0.0000	1.0000	0.0000	1.0000	0.0004	0.9999	0.0036	0.9985	0.0151	0.9906
11	-	-	0.0000	1.0000	0.0001	1.0000	0.0011	0.9996	0.0062	0.9968
12	-	-	0.0000	1.0000	0.0000	1.0000	0.0003	0.9999	0.0022	0.9990
13	-	-	0.0000	1.0000	0.0000	1.0000	0.0001	0.0000	0.0007	0.9997
14	-	-	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0002	0.9999
15	-	-	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000

TABLE 25.3 – Table de la loi binomiale avec $p = 0.20$

k	$n = 10$		$n = 20$		$n = 30$		$n = 40$		$n = 50$	
	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$
0	0.1074	0.1074	0.0115	0.0115	0.0012	0.0012	0.0001	0.0001	0.0000	0.0000
1	0.2684	0.3758	0.0577	0.0692	0.0093	0.0105	0.0014	0.0015	0.0002	0.0002
2	0.3020	0.6778	0.1369	0.2061	0.0337	0.0442	0.0064	0.0079	0.0011	0.0013
3	0.2013	0.8791	0.2053	0.4114	0.0785	0.1227	0.0206	0.0285	0.0044	0.0057
4	0.0881	0.9672	0.2182	0.6296	0.1325	0.2552	0.0474	0.0759	0.0128	0.0185
5	0.0264	0.9936	0.1746	0.8042	0.1723	0.4275	0.0854	0.1613	0.0295	0.0480
6	0.0055	0.9991	0.1091	0.9133	0.1795	0.6070	0.1246	0.2859	0.0554	0.1034
7	0.0008	0.9999	0.0546	0.9679	0.1538	0.7608	0.1512	0.4371	0.0870	0.1904
8	0.0001	1.0000	0.0221	0.9900	0.1105	0.8713	0.1560	0.5931	0.1169	0.3073
9	0.0000	1.0000	0.0074	0.9974	0.0676	0.9389	0.1387	0.7318	0.1364	0.4437
10	0.0000	1.0000	0.0020	0.9994	0.0355	0.9744	0.1074	0.8392	0.1399	0.5836
11	-	-	0.0005	0.9999	0.0161	0.9905	0.0733	0.9125	0.1271	0.7107
12	-	-	0.0001	1.0000	0.0064	0.9969	0.0443	0.9568	0.1032	0.8139
13	-	-	0.0000	1.0000	0.0022	0.9991	0.0238	0.9806	0.0755	0.8894
14	-	-	0.0000	1.0000	0.0007	0.9998	0.0115	0.9921	0.0499	0.9393
15	-	-	0.0000	1.0000	0.0002	1.0000	0.0050	0.9971	0.0299	0.9692
16	-	-	0.0000	1.0000	0.0000	1.0000	0.0019	0.9990	0.0164	0.9856
17	-	-	0.0000	1.0000	0.0000	1.0000	0.0007	0.9997	0.0081	0.9937
18	-	-	0.0000	1.0000	0.0000	1.0000	0.0002	0.9999	0.0038	0.9975
19	-	-	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000	0.0016	0.9991
20	-	-	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0006	0.9997
21	-	-	-	-	0.0000	1.0000	0.0000	1.0000	0.0002	0.9999
22	-	-	-	-	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000

TABLE 25.4 – Table de la loi binomiale avec $p = 0.30$

k	$n = 10$		$n = 20$		$n = 30$		$n = 40$		$n = 50$	
	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$
0	0.0282	0.0282	0.0008	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.1211	0.1493	0.0068	0.0076	0.0003	0.0003	0.0000	0.0000	0.0000	0.0000
2	0.2335	0.3828	0.0279	0.0355	0.0018	0.0021	0.0001	0.0001	0.0000	0.0000
3	0.2668	0.6496	0.0716	0.1071	0.0072	0.0093	0.0005	0.0006	0.0000	0.0000
4	0.2001	0.8497	0.1304	0.2375	0.0209	0.0302	0.0020	0.0026	0.0002	0.0002
5	0.1030	0.9527	0.1789	0.4164	0.0464	0.0766	0.0060	0.0086	0.0005	0.0007
6	0.0367	0.9894	0.1916	0.6080	0.0829	0.1595	0.0152	0.0238	0.0018	0.0025
7	0.0090	0.9984	0.1643	0.7723	0.1219	0.2814	0.0315	0.0553	0.0048	0.0073
8	0.0015	0.9999	0.1144	0.8867	0.1501	0.4315	0.0557	0.1110	0.0110	0.0183
9	0.0001	1.0000	0.0653	0.9520	0.1573	0.5888	0.0849	0.1959	0.0219	0.0402
10	0.0000	1.0000	0.0309	0.9829	0.1416	0.7304	0.1128	0.3087	0.0387	0.789
11	-	-	0.0120	0.9949	0.1103	0.8407	0.1319	0.4406	0.0601	0.1390
12	-	-	0.0038	0.9987	0.0748	0.9155	0.1366	0.5772	0.0839	0.2229
13	-	-	0.0010	0.9997	0.0444	0.9599	0.1260	0.7032	0.1050	0.3279
14	-	-	0.0003	1.0000	0.0232	0.9831	0.1042	0.8074	0.1189	0.4468
15	-	-	0.0000	1.0000	0.0105	0.9936	0.0775	0.8849	0.1224	0.5692
16	-	-	0.0000	1.0000	0.0043	0.9979	0.0518	0.9367	0.1147	0.6839
17	-	-	0.0000	1.0000	0.0015	0.9994	0.0313	0.9680	0.0983	0.7822
18	-	-	0.0000	1.0000	0.0004	0.9998	0.0172	0.9852	0.0772	0.8594
19	-	-	0.0000	1.0000	0.0002	1.0000	0.0085	0.9937	0.0558	0.9152
20	-	-	0.0000	1.0000	0.0000	1.0000	0.0039	0.9976	0.0370	0.9522
21	-	-	-	-	0.0000	1.0000	0.0015	0.9991	0.0227	0.9749
22	-	-	-	-	0.0000	1.0000	0.0006	0.9997	0.0128	0.9877
23	-	-	-	-	0.0000	1.0000	0.0002	0.9999	0.0067	0.9944
24	-	-	-	-	0.0000	1.0000	0.0001	1.0000	0.0032	0.9976
25	-	-	-	-	0.0000	1.0000	0.0000	1.0000	0.0015	0.9991
26	-	-	-	-	0.0000	1.0000	0.0000	1.0000	0.0006	0.9997
27	-	-	-	-	0.0000	1.0000	0.0000	1.0000	0.0002	0.9999
28	-	-	-	-	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000

TABLE 25.5 – Table de la loi binomiale avec $p = 0.40$

k	$n = 10$		$n = 20$		$n = 30$		$n = 40$		$n = 50$	
	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$
0	0.0060	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0404	0.0464	0.0005	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.1209	0.1673	0.0031	0.0036	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.2150	0.3823	0.0124	0.0160	0.0003	0.0003	0.0000	0.0000	0.0000	0.0000
4	0.2508	0.6331	0.0350	0.0510	0.0012	0.0015	0.0000	0.0000	0.0000	0.0000
5	0.2007	0.8338	0.0746	0.1256	0.0042	0.0057	0.0001	0.0001	0.0000	0.0000
6	0.1114	0.9452	0.1244	0.2500	0.0115	0.0172	0.0005	0.0006	0.0000	0.0000
7	0.0425	0.9877	0.1659	0.4159	0.0263	0.0435	0.0015	0.0021	0.0001	0.0001
8	0.0106	0.9983	0.1797	0.5956	0.0505	0.0940	0.0040	0.0061	0.0001	0.0002
9	0.0016	0.9999	0.1597	0.7553	0.0823	0.1763	0.0095	0.0156	0.0006	0.0008
10	0.0001	1.0000	0.1172	0.8725	0.1152	0.2915	0.0196	0.0352	0.0014	0.0022
11	-	-	0.0710	0.9435	0.1396	0.4311	0.0357	0.0709	0.0035	0.0057
12	-	-	0.0355	0.9790	0.1474	0.5785	0.0576	0.1285	0.0076	0.0133
13	-	-	0.0145	0.9935	0.1360	0.7145	0.0827	0.2112	0.0147	0.0280
14	-	-	0.0049	0.9984	0.1101	0.8246	0.1062	0.3174	0.0260	0.0540
15	-	-	0.0013	0.9997	0.0783	0.9029	0.1228	0.4402	0.0415	0.0955
16	-	-	0.0003	1.0000	0.0490	0.9519	0.1279	0.5681	0.0605	0.1561
17	-	-	0.0000	1.0000	0.0279	0.9798	0.1204	0.6885	0.0808	0.2369
18	-	-	0.0000	1.0000	0.0119	0.9917	0.1026	0.7911	0.0987	0.3356
19	-	-	0.0000	1.0000	0.0054	0.9971	0.0791	0.8702	0.1109	0.4465
20	-	-	0.0000	1.0000	0.0020	0.9991	0.0554	0.9256	0.1145	0.5610
21	-	-	-	-	0.0007	0.9998	0.0352	0.9608	0.1091	0.6701
22	-	-	-	-	0.0002	1.0000	0.0203	0.9811	0.0959	0.7660
23	-	-	-	-	0.0000	1.0000	0.0106	0.9917	0.0778	0.8438
24	-	-	-	-	0.0000	1.0000	0.0049	0.9966	0.0584	0.9022
25	-	-	-	-	0.0000	1.0000	0.0022	0.9988	0.0405	0.9427
26	-	-	-	-	0.0000	1.0000	0.0008	0.9996	0.0259	0.9686
27	-	-	-	-	0.0000	1.0000	0.0003	0.9999	0.0154	0.9840
28	-	-	-	-	0.0000	1.0000	0.0001	1.0000	0.0084	0.9924
29	-	-	-	-	0.0000	1.0000	0.0000	1.0000	0.0042	0.9966
30	-	-	-	-	0.0000	1.0000	0.0000	1.0000	0.0020	0.9986
31	-	-	-	-	-	-	0.0000	1.0000	0.0009	0.9995
32	-	-	-	-	-	-	0.0000	1.0000	0.0003	0.9998
33	-	-	-	-	-	-	0.0000	1.0000	0.0001	0.9999
34	-	-	-	-	-	-	0.0000	1.0000	0.0001	1.0000

TABLE 25.6 – Table de la loi binomiale avec $p = 0.50$

k	$n = 10$		$n = 20$		$n = 30$		$n = 40$		$n = 50$	
	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$
0	0.0010	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0097	0.0107	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0440	0.0547	0.0002	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.1172	0.1719	0.0011	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.2051	0.3770	0.0046	0.0059	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.2460	0.6230	0.0148	0.0207	0.0002	0.0002	0.0000	0.0000	0.0000	0.0000
6	0.2051	0.8281	0.0370	0.0577	0.0005	0.0007	0.0000	0.0000	0.0000	0.0000
7	0.1172	0.9453	0.0739	0.1316	0.0019	0.0026	0.0000	0.0000	0.0000	0.0000
8	0.0440	0.9893	0.1201	0.2517	0.0055	0.0081	0.0001	0.0001	0.0000	0.0000
9	0.0097	0.9990	0.1602	0.4119	0.0133	0.0214	0.0002	0.0003	0.0000	0.0000
10	0.0010	1.0000	0.1762	0.5881	0.0280	0.0494	0.0008	0.0011	0.0000	0.0000
11	-	-	0.1602	0.7483	0.0508	0.1002	0.0021	0.0032	0.0000	0.0000
12	-	-	0.1201	0.8684	0.0806	0.1808	0.0051	0.0083	0.0002	0.0002
13	-	-	0.0739	0.9423	0.1115	0.2923	0.0109	0.0192	0.0003	0.0005
14	-	-	0.0370	0.9793	0.1355	0.4278	0.0211	0.0403	0.0008	0.0013
15	-	-	0.0148	0.9941	0.1444	0.5722	0.0366	0.0769	0.0020	0.0033
16	-	-	0.0046	0.9987	0.1355	0.7077	0.0572	0.1341	0.0044	0.0077
17	-	-	0.0011	0.9998	0.1115	0.8192	0.0807	0.2148	0.0087	0.0164
18	-	-	0.0002	1.0000	0.0806	0.8998	0.1031	0.3179	0.0161	0.0325
19	-	-	0.0000	1.0000	0.0508	0.9506	0.1194	0.4373	0.0270	0.0595
20	-	-	0.0000	1.0000	0.0280	0.9786	0.1254	0.5627	0.0418	0.1013
21	-	-	-	-	0.0133	0.9919	0.1194	0.6821	0.0598	0.1611
22	-	-	-	-	0.0055	0.9974	0.1031	0.7852	0.0788	0.2399
23	-	-	-	-	0.0019	0.9993	0.0807	0.8659	0.0960	0.3359
24	-	-	-	-	0.0005	0.9998	0.0572	0.9231	0.1080	0.4439
25	-	-	-	-	0.0002	1.0000	0.0366	0.9597	0.1122	0.5561
26	-	-	-	-	0.0000	1.0000	0.0211	0.9808	0.1080	0.6641
27	-	-	-	-	0.0000	1.0000	0.0109	0.9917	0.0960	0.7601
28	-	-	-	-	0.0000	1.0000	0.0051	0.9968	0.0788	0.8389
29	-	-	-	-	0.0000	1.0000	0.0021	0.9989	0.0598	0.8987
30	-	-	-	-	0.0000	1.0000	0.0008	0.9997	0.0418	0.9405
31	-	-	-	-	-	-	0.0002	0.9999	0.0270	0.9675
32	-	-	-	-	-	-	0.0001	1.0000	0.0161	0.9836
33	-	-	-	-	-	-	0.0000	1.0000	0.0087	0.9923
34	-	-	-	-	-	-	0.0000	1.0000	0.0044	0.9967
35	-	-	-	-	-	-	0.0000	1.0000	0.0020	0.9987
36	-	-	-	-	-	-	0.0000	1.0000	0.0008	0.9995
37	-	-	-	-	-	-	0.0000	1.0000	0.0003	0.9998
38	-	-	-	-	-	-	0.0000	1.0000	0.0002	1.0000

TABLE 25.7 – Table de la loi de Poisson - 1

k	$\lambda = 0.1$		$\lambda = 0.2$		$\lambda = 0.3$		$\lambda = 0.4$		$\lambda = 0.5$	
	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$
0	0.9048	0.9048	0.8187	0.8187	0.7408	0.7408	0.6703	0.6703	0.6065	0.6065
1	0.0905	0.9953	0.1638	0.9825	0.2222	0.9630	0.2681	0.9384	0.3033	0.9098
2	0.0045	0.9998	0.0164	0.9989	0.0333	0.9963	0.0536	0.9920	0.0758	0.9856
3	0.0002	1.0000	0.0011	1.0000	0.0033	0.9996	0.0072	0.9992	0.0126	0.9982
4	0.0000	1.0000	0.0000	1.0000	0.0003	0.9999	0.0007	0.9999	0.0016	0.9998
5	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000	0.0001	1.0000	0.0002	1.0000
k	$\lambda = 0.6$		$\lambda = 0.7$		$\lambda = 0.8$		$\lambda = 0.9$		$\lambda = 1.0$	
	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$
0	0.5488	0.5488	0.4966	0.4966	0.4493	0.4493	0.4066	0.4066	0.3679	0.3679
1	0.3293	0.8781	0.3476	0.8442	0.3595	0.8088	0.3659	0.7725	0.3679	0.7358
2	0.0988	0.9769	0.1217	0.9659	0.1438	0.9526	0.1647	0.9372	0.1839	0.9197
3	0.0198	0.9967	0.0284	0.9943	0.0383	0.9909	0.0494	0.9866	0.0613	0.9810
4	0.0030	0.9997	0.0050	0.9993	0.0077	0.9986	0.0111	0.9977	0.0153	0.9963
5	0.0003	1.0000	0.0007	1.0000	0.0012	0.9998	0.0020	0.9997	0.0031	0.9994
6	0.0000	1.0000	0.0000	1.0000	0.0002	1.0000	0.0003	1.0000	0.0005	0.9999
7	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000
k	$\lambda = 2.0$		$\lambda = 3.0$		$\lambda = 4.0$		$\lambda = 5.0$		$\lambda = 6.0$	
	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$
0	0.1353	0.1353	0.0498	0.0498	0.0183	0.0183	0.0067	0.0067	0.0025	0.0025
1	0.2707	0.4060	0.1494	0.1992	0.0733	0.0916	0.0337	0.0404	0.0149	0.0174
2	0.2707	0.6767	0.2240	0.4232	0.1465	0.2381	0.0842	0.1246	0.0446	0.0620
3	0.1805	0.8572	0.2240	0.6472	0.1954	0.4335	0.1404	0.2650	0.0892	0.1512
4	0.0902	0.9474	0.1680	0.8152	0.1954	0.6289	0.1755	0.4405	0.1339	0.2851
5	0.0361	0.9835	0.1008	0.9160	0.1563	0.7852	0.1755	0.6160	0.1606	0.4457
6	0.0120	0.9955	0.0504	0.9664	0.1042	0.8894	0.1462	0.7622	0.1606	0.6063
7	0.0034	0.9989	0.0216	0.9880	0.0595	0.9489	0.1044	0.8666	0.1377	0.7440
8	0.0009	0.9998	0.0081	0.9961	0.0298	0.9787	0.0653	0.9319	0.1033	0.8473
9	0.0002	1.0000	0.0027	0.9988	0.0132	0.9919	0.0363	0.9682	0.0688	0.9161
10	0.0000	1.0000	0.0008	0.9996	0.0053	0.9972	0.0181	0.9863	0.0413	0.9574
11	0.0000	1.0000	0.0003	0.9999	0.0019	0.9991	0.0082	0.9945	0.0225	0.9799
12	0.0000	1.0000	0.0001	1.0000	0.0006	0.9997	0.0034	0.9979	0.0113	0.9912
13	0.0000	1.0000	0.0000	1.0000	0.0002	0.9999	0.0013	0.9992	0.0052	0.9964
14	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000	0.0005	0.9997	0.0022	0.9986
15	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0002	0.9999	0.0009	0.9995
16	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000	0.0004	0.9999
17	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000

TABLE 25.8 – Table de la loi de Poisson - 2

k	$\lambda = 7.0$		$\lambda = 8.0$		$\lambda = 9.0$		$\lambda = 10.0$		$\lambda = 11.0$	
	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$
0	0.0009	0.0009	0.0003	0.0003	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000
1	0.0064	0.0073	0.0027	0.0030	0.0011	0.0012	0.0005	0.0005	0.0002	0.0002
2	0.0223	0.0296	0.0107	0.0137	0.0050	0.0062	0.0023	0.0028	0.0010	0.0012
3	0.0521	0.0817	0.0286	0.0423	0.0150	0.0212	0.0076	0.0104	0.0037	0.0049
4	0.0912	0.1729	0.0573	0.0996	0.0337	0.0549	0.0189	0.0293	0.0102	0.0151
5	0.1277	0.3006	0.0916	0.1912	0.0607	0.1156	0.0378	0.0671	0.0224	0.0375
6	0.1490	0.4496	0.1221	0.3133	0.0911	0.2067	0.0631	0.1302	0.0411	0.0786
7	0.1490	0.5986	0.1396	0.4529	0.1171	0.3238	0.0901	0.2203	0.0646	0.1432
8	0.1304	0.7290	0.1396	0.5925	0.1318	0.4556	0.1126	0.3329	0.0888	0.2320
9	0.1014	0.8304	0.1241	0.7166	0.1318	0.5874	0.1251	0.4580	0.1085	0.3405
10	0.0710	0.9014	0.0993	0.8159	0.1186	0.7060	0.1251	0.5831	0.1194	0.4599
11	0.0452	0.9466	0.0722	0.8881	0.0970	0.8030	0.1137	0.6968	0.1194	0.5793
12	0.0264	0.9730	0.0481	0.9362	0.0728	0.8758	0.0948	0.7916	0.1094	0.6887
13	0.0142	0.9872	0.0296	0.9658	0.0504	0.9262	0.0729	0.8645	0.0926	0.7813
14	0.0071	0.9943	0.0169	0.9827	0.0324	0.9586	0.0521	0.9166	0.0728	0.8541
15	0.0033	0.9976	0.0090	0.9917	0.0194	0.9780	0.0347	0.9513	0.0534	0.9075
16	0.0015	0.9991	0.0045	0.9962	0.0109	0.9889	0.0217	0.9730	0.0367	0.9442
17	0.0006	0.9997	0.0021	0.9983	0.0058	0.9947	0.0128	0.9858	0.0237	0.9679
18	0.0002	0.9999	0.0009	0.9992	0.0029	0.9976	0.0071	0.9929	0.0145	0.9824
19	0.0001	1.0000	0.0004	0.9996	0.0014	0.9990	0.0037	0.9966	0.0084	0.9908
20	0.0000	1.0000	0.0003	0.9999	0.0006	0.9996	0.0019	0.9985	0.0046	0.9954
21	0.0000	1.0000	0.0001	1.0000	0.0003	0.9999	0.0009	0.9994	0.0024	0.9978
22	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000	0.0004	0.9998	0.0012	0.9990
23	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0002	1.0000	0.0006	0.9996
24	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0003	0.9999
25	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000

TABLE 25.9 – Table de la loi de Poisson - 3

k	$\lambda = 12.0$		$\lambda = 13.0$		$\lambda = 14.0$		$\lambda = 15.0$		$\lambda = 16.0$	
	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$	p_k	$\sum_{i=0}^k p_i$
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0004	0.0005	0.0002	0.0002	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000
3	0.0018	0.0023	0.0008	0.0010	0.0004	0.0005	0.0002	0.0002	0.0001	0.0001
4	0.0053	0.0076	0.0027	0.0037	0.0013	0.0018	0.0007	0.0009	0.0003	0.0004
5	0.0127	0.0203	0.0070	0.0107	0.0037	0.0055	0.0019	0.0028	0.0010	0.0014
6	0.0255	0.0458	0.0152	0.0259	0.0087	0.0142	0.0048	0.0076	0.0026	0.0040
7	0.0437	0.0895	0.0281	0.0540	0.0174	0.0316	0.0104	0.0180	0.0060	0.0100
8	0.0655	0.1550	0.0457	0.0997	0.0304	0.0620	0.0194	0.0374	0.0120	0.0220
9	0.0874	0.2424	0.0661	0.1658	0.0473	0.1093	0.0324	0.0698	0.0213	0.0433
10	0.1048	0.3472	0.0859	0.2517	0.0663	0.1756	0.0486	0.1184	0.0341	0.0774
11	0.1144	0.4616	0.1015	0.3532	0.0844	0.2600	0.0663	0.1847	0.0496	0.1270
12	0.1144	0.5760	0.1099	0.4631	0.0984	0.3584	0.0829	0.2676	0.0661	0.1931
13	0.1056	0.6816	0.1099	0.5730	0.1060	0.4644	0.0956	0.3632	0.0814	0.2745
14	0.0905	0.7721	0.1021	0.6751	0.1060	0.5704	0.1024	0.4656	0.0930	0.3675
15	0.0724	0.8445	0.0885	0.7636	0.0989	0.6693	0.1024	0.5680	0.0992	0.4667
16	0.0543	0.8988	0.0719	0.8355	0.0866	0.7559	0.0960	0.6640	0.0992	0.5659
17	0.0383	0.9371	0.0550	0.8905	0.0713	0.8272	0.0847	0.7487	0.0934	0.6593
18	0.0255	0.9626	0.0397	0.9302	0.0554	0.8826	0.0706	0.8193	0.0830	0.7423
19	0.0161	0.9787	0.0272	0.9574	0.0409	0.9235	0.0558	0.8751	0.0699	0.8122
20	0.0097	0.9884	0.0177	0.9751	0.0286	0.9521	0.0418	0.9169	0.0559	0.8681
21	0.0055	0.9939	0.0109	0.9860	0.0191	0.9712	0.0299	0.9468	0.0426	0.9107
22	0.0030	0.9969	0.0065	0.9925	0.0121	0.9833	0.0204	0.9672	0.0310	0.9417
23	0.0016	0.9985	0.0037	0.9962	0.0074	0.9907	0.0133	0.9805	0.0216	0.9633
24	0.0008	0.9993	0.0020	0.9982	0.0043	0.9950	0.0083	0.9888	0.0144	0.9777
25	0.0004	0.9997	0.0010	0.9992	0.0024	0.9974	0.0050	0.9938	0.0092	0.9869
26	0.0002	0.9999	0.0005	0.9997	0.0013	0.9987	0.0029	0.9967	0.0057	0.9926
27	0.0001	1.0000	0.0002	0.9999	0.0007	0.9994	0.0016	0.9983	0.0034	0.9960
28	0.0000	1.0000	0.0001	1.0000	0.0003	0.9997	0.0009	0.9992	0.0019	0.9979
29	0.0000	1.0000	0.0000	1.0000	0.0002	0.9999	0.0004	0.9996	0.0010	0.9989
30	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000	0.0002	0.9998	0.0006	0.9995
31	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0001	0.9999	0.0003	0.9998
32	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000	0.0001	0.9999
33	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000

Tables des lois continues

On donne maintenant les tables des lois continues, en commençant par la loi normale centrée réduite.

La lecture de la table est ainsi faite : à l'intersection de la ligne $a.b$ et de la colonne $0.0c$, on donne $\Phi(a.bc)$. Par exemple, $\Phi(2.72) = \Phi(2.7 + 0.02) = 0.9967$.

La précision est de quatre chiffres significatifs.

TABLE 25.10 – Table de la fonction de répartition de la loi normale

t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7290	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9779	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

On donne également $\Phi(t)$ pour des “grandes” valeurs de t .

TABLE 25.11 – Cas des grandes valeurs de $t - 1$

t	3.0	3.1	3.2	3.3	3.4
$\Phi(t)$	0.99865	0.99904	0.99931	0.99952	0.99966

TABLE 25.12 – Cas des grandes valeurs de $t - 2$

t	3.5	3.6	3.8	4.0	4.5
$\Phi(t)$	0.99976	0.999841	0.999928	0.999968	0.999997

La table suivante donne le quantile d'ordre $1 - \frac{\gamma}{2}$ de la loi normale centrée réduite. En d'autres termes, à l'intersection de la ligne $0.a$ et de la colonne $0.0b$, on donne la solution x à l'équation $\Phi(x) = 1 - \frac{\gamma}{2}$.

Ainsi, si l'on cherche le quantile d'ordre 0.975, on prend $\gamma = 0.05$ d'où ce quantile vaut 1.9600. La précision est de quatre chiffres significatifs.

TABLE 25.13 – Quantile d'ordre $1 - \frac{\gamma}{2}$ de la loi normale centrée réduite

γ	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	$+\infty$	2.5758	2.3263	2.1701	2.0537	1.9600	1.8808	1.8119	1.7507	1.6954
0.1	1.6449	1.5982	1.5548	1.5141	1.4758	1.4395	1.4051	1.3722	1.3408	1.3106
0.2	1.2816	1.2536	1.2265	1.2004	1.1750	1.1503	1.1264	1.1031	1.0803	1.0581
0.3	1.0364	1.0152	0.9945	0.9741	0.9542	0.9346	0.9154	0.8965	0.8779	0.8596
0.4	0.8416	0.8239	0.8064	0.7892	0.7722	0.7554	0.7388	0.7225	0.7063	0.6903
0.5	0.6745	0.6588	0.6433	0.6280	0.6128	0.5978	0.5828	0.5681	0.5534	0.5388
0.6	0.5244	0.5101	0.4959	0.4817	0.4677	0.4538	0.4399	0.4261	0.4125	0.3989
0.7	0.3853	0.3719	0.3585	0.3451	0.3319	0.3186	0.3055	0.2924	0.2793	0.2663
0.8	0.2533	0.2404	0.2275	0.2147	0.2019	0.1891	0.1764	0.1637	0.1510	0.1383
0.9	0.1257	0.1130	0.1004	0.0878	0.0753	0.0627	0.0502	0.0376	0.0251	0.0125

On donne ici la table des quantiles de la loi de Student avec trois chiffres significatifs.

Ainsi, à l'intersection de la ligne $n = n_0$ et de la colonne $\gamma = \gamma_0$, on trouve la solution de l'équation $\mathbb{P}(X_0 \leq x) = \gamma_0$ où X_0 est une variable aléatoire suivant la loi de Student à n_0 degrés de liberté.

On remarquera que la loi de Student à un nombre infini de degrés de liberté est la loi normale centrée réduite.

TABLE 25.14 – Quantiles d'ordre γ de la loi de Student à n degrés de liberté

$n \backslash \gamma$	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995
1	0.158	0.325	0.510	0.727	1.00	1.38	1.96	3.08	6.31	12.7	31.8	63.7
2	0.142	0.289	0.445	0.617	0.816	1.06	1.39	1.89	2.92	4.30	6.96	9.92
3	0.137	0.277	0.424	0.584	0.765	0.978	1.25	1.64	2.35	3.18	4.54	5.84
4	0.134	0.271	0.414	0.569	0.741	0.941	1.19	1.53	2.13	2.78	3.75	4.60
5	0.132	0.267	0.408	0.559	0.727	0.920	1.16	1.48	2.01	2.57	3.36	4.03
6	0.131	0.265	0.404	0.553	0.718	0.906	1.13	1.44	1.94	2.45	3.14	3.71
7	0.130	0.263	0.402	0.549	0.711	0.896	1.12	1.42	1.90	2.36	3.00	3.50
8	0.130	0.262	0.399	0.546	0.706	0.889	1.11	1.40	1.86	2.31	2.90	3.36
9	0.129	0.261	0.398	0.543	0.703	0.883	1.10	1.38	1.83	2.26	2.82	3.25
10	0.129	0.260	0.397	0.542	0.700	0.879	1.09	1.37	1.81	2.23	2.76	3.17
11	0.129	0.260	0.396	0.540	0.697	0.876	1.09	1.36	1.80	2.20	2.72	3.11
12	0.128	0.259	0.395	0.539	0.695	0.873	1.08	1.36	1.78	2.18	2.68	3.06
13	0.128	0.259	0.394	0.538	0.694	0.870	1.08	1.35	1.77	2.16	2.65	3.01
14	0.128	0.258	0.393	0.537	0.692	0.868	1.08	1.34	1.76	2.14	2.62	2.98
15	0.128	0.258	0.393	0.536	0.691	0.866	1.07	1.34	1.75	2.13	2.60	2.95
16	0.128	0.258	0.392	0.535	0.690	0.865	1.07	1.34	1.75	2.12	2.58	2.92
17	0.128	0.257	0.392	0.534	0.689	0.863	1.07	1.33	1.74	2.11	2.57	2.90
18	0.127	0.257	0.392	0.534	0.688	0.862	1.07	1.33	1.73	2.10	2.55	2.88
19	0.127	0.257	0.391	0.533	0.688	0.861	1.07	1.33	1.73	2.09	2.54	2.86
20	0.127	0.257	0.391	0.533	0.687	0.860	1.06	1.32	1.72	2.09	2.53	2.84
21	0.127	0.257	0.391	0.532	0.686	0.859	1.06	1.32	1.72	2.08	2.52	2.83
22	0.127	0.256	0.390	0.532	0.686	0.858	1.06	1.32	1.72	2.07	2.51	2.82
23	0.127	0.256	0.390	0.532	0.685	0.858	1.06	1.32	1.71	2.07	2.50	2.81
24	0.127	0.256	0.390	0.531	0.685	0.857	1.06	1.32	1.71	2.06	2.49	2.80
25	0.127	0.256	0.390	0.531	0.684	0.856	1.06	1.32	1.71	2.06	2.48	2.79
26	0.127	0.256	0.390	0.531	0.684	0.856	1.06	1.32	1.70	2.06	2.48	2.78
27	0.127	0.256	0.389	0.531	0.684	0.855	1.06	1.31	1.70	2.05	2.47	2.77
28	0.127	0.256	0.389	0.530	0.683	0.855	1.06	1.31	1.70	2.05	2.47	2.76
29	0.127	0.256	0.389	0.530	0.683	0.854	1.05	1.31	1.70	2.04	2.46	2.76
30	0.127	0.256	0.389	0.530	0.683	0.854	1.05	1.31	1.70	2.04	2.46	2.75
∞	0.126	0.253	0.385	0.524	0.674	0.842	1.04	1.28	1.64	1.96	2.33	2.58

On donne ici la table des quantiles de la loi du Khi-deux avec trois chiffres significatifs.

Ainsi, à l'intersection de la ligne $n = n_0$ et de la colonne $\gamma = \gamma_0$, on trouve la solution de l'équation $\mathbb{P}(X_0 \leq x) = \gamma_0$ où X_0 est une variable aléatoire suivant la loi du Khi-deux à n_0 degrés de liberté.

TABLE 25.15 – Quantiles d'ordre γ de la loi du χ^2 à n degrés de liberté

$n \backslash \gamma$	0.005	0.01	0.025	0.05	0.10	0.20	0.50	0.80	0.90	0.95	0.99	0.995
1	0.00	0.00	0.00	0.00	0.02	0.06	0.46	1.64	2.71	3.84	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.45	1.39	3.22	4.61	5.99	9.21	10.6
3	0.07	0.12	0.22	0.35	0.58	1.00	2.37	4.64	6.25	7.81	11.3	12.8
4	0.21	0.30	0.48	0.71	1.06	1.65	3.36	5.99	7.78	9.49	13.3	14.9
5	0.41	0.55	0.83	1.15	1.61	2.34	4.35	7.29	9.24	11.1	15.1	16.7
6	0.68	0.87	1.24	1.64	2.20	3.07	5.35	8.56	10.6	12.6	16.8	18.5
7	0.99	1.24	1.69	2.17	2.83	3.82	6.35	9.80	12.0	14.1	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	4.59	7.34	11.0	13.4	15.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.38	8.34	12.2	14.7	16.9	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.18	9.34	13.4	16.0	18.3	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	6.99	10.3	14.6	17.3	19.7	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	7.81	11.3	15.8	18.5	21.0	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	8.63	12.3	17.0	19.8	22.4	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	9.47	13.3	18.2	21.1	23.7	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	10.3	14.3	19.3	22.3	25.0	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.2	15.3	20.5	23.5	26.3	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.0	16.3	21.6	24.8	27.6	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	12.9	17.3	22.8	26.0	28.9	34.8	37.2
19	6.83	7.63	8.91	10.1	11.7	13.7	18.3	23.9	27.2	30.1	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	14.6	19.3	25.0	28.4	31.4	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	15.4	20.3	26.2	29.6	32.7	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	16.3	21.3	27.3	30.8	33.9	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	17.2	22.3	28.4	32.0	35.2	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	18.1	23.3	29.6	33.2	36.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	18.9	24.3	30.7	34.4	37.7	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	19.8	25.3	31.8	35.6	38.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	20.7	26.3	32.9	36.7	40.1	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	21.6	27.3	34.0	37.9	41.3	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	22.5	28.3	35.1	39.1	42.6	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	23.4	29.3	36.2	40.3	43.8	50.9	53.7
40	20.7	22.1	24.4	26.5	29.0	32.3	39.3	47.3	51.8	55.8	63.7	66.8
50	28.0	29.7	32.3	34.8	37.7	41.3	49.3	58.2	63.2	67.5	76.2	79.5
60	35.5	37.5	40.5	43.2	46.5	50.6	59.3	69.0	74.4	79.1	88.4	92.0

On donne ici la table du quantile d'ordre 0.975 de la loi de Fisher avec trois chiffres significatifs.

Ainsi, à l'intersection de la ligne $d_2 = 17$ et de la colonne $d_1 = 15$, on trouve la solution de l'équation $\mathbb{P}(X_0 \leq x) = 0.975$ où X_0 est une variable aléatoire suivant la loi de Fisher à $d_1 = 15$ et $d_2 = 17$ degrés de liberté.

TABLE 25.16 – Quantiles d'ordre 0.975 de la loi $\mathcal{F}(d_1, d_2)$

$d_2 \backslash d_1$	1	2	3	4	5	6	8	10	15	20	30	$+\infty$
1	648	800	864	900	922	937	957	969	985	993	1001	1018
2	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.5	39.5
3	17.4	16.0	15.4	15.1	14.9	14.7	14.5	14.4	14.3	14.2	14.1	13.9
4	12.2	10.6	9.98	9.60	9.36	9.20	8.98	8.84	8.66	8.56	8.46	8.26
5	10.0	8.43	7.76	7.39	7.15	6.98	6.76	6.62	6.43	6.33	6.23	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.60	5.46	5.27	5.17	5.07	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.90	4.76	4.57	4.47	4.36	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.43	4.30	4.10	4.00	3.89	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.10	3.96	3.77	3.67	3.56	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.85	3.72	3.52	3.42	3.31	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.66	3.53	3.33	3.23	3.12	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.51	3.37	3.18	3.07	2.96	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.39	3.25	3.05	2.95	2.84	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.29	3.15	2.95	2.84	2.73	2.49
15	6.20	4.76	4.15	3.80	3.58	3.41	3.20	3.06	2.86	2.76	2.64	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.12	2.99	2.79	2.68	2.57	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.06	2.92	2.72	2.62	2.50	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.01	2.87	2.67	2.56	2.44	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	2.96	2.82	2.62	2.51	2.39	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	2.91	2.77	2.57	2.46	2.35	2.09
22	5.79	4.38	3.78	3.44	3.22	3.05	2.84	2.70	2.50	2.39	2.27	2.00
24	5.72	4.32	3.72	3.38	3.15	2.99	2.78	2.64	2.44	2.33	2.21	1.94
26	5.66	4.27	3.67	3.33	3.10	2.94	2.73	2.59	2.39	2.28	2.16	1.88
28	5.61	4.22	3.63	3.29	3.06	2.90	2.69	2.55	2.34	2.23	2.11	1.83
30	5.57	4.18	3.59	3.25	3.03	2.87	2.65	2.51	2.31	2.20	2.07	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.53	2.39	2.18	2.07	1.94	1.64
50	5.34	3.98	3.39	3.06	2.83	2.67	2.46	2.32	2.11	1.99	1.87	1.55
60	5.29	3.93	3.34	3.01	2.79	2.63	2.41	2.27	2.06	1.94	1.82	1.48
80	5.22	3.86	3.28	2.95	2.75	2.57	2.36	2.21	2.00	1.88	1.75	1.40
100	5.18	3.83	3.25	2.92	2.70	2.54	2.32	2.18	1.97	1.85	1.71	1.31
$+\infty$	5.02	3.69	3.12	2.79	2.57	2.41	2.19	2.05	1.83	1.71	1.57	1.00

On donne ici la table du quantile d'ordre 0.95 de la loi de Fisher avec trois chiffres significatifs.

Ainsi, à l'intersection de la ligne $d_2 = 17$ et de la colonne $d_1 = 15$, on trouve la solution de l'équation $\mathbb{P}(X_0 \leq x) = 0.95$ où X_0 est une variable aléatoire suivant la loi de Fisher à $d_1 = 15$ et $d_2 = 17$ degrés de liberté.

TABLE 25.17 – Quantiles d'ordre 0.95 de la loi $\mathcal{F}(d_1, d_2)$

$d_2 \backslash d_1$	1	2	3	4	5	6	8	10	15	20	30	$+\infty$
1	161	200	216	225	230	234	239	242	246	248	250	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.85	8.79	8.70	8.66	8.62	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.96	5.86	5.80	5.75	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.74	4.62	4.56	4.50	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.06	3.94	3.87	3.81	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.64	3.51	3.44	3.38	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.35	3.22	3.15	3.08	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.14	3.01	2.94	2.86	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.98	2.85	2.77	2.70	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.85	2.72	2.65	2.57	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.85	2.75	2.62	2.54	2.47	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.77	2.67	2.53	2.46	2.38	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.60	2.46	2.39	2.31	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.54	2.40	2.33	2.25	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.49	2.35	2.28	2.19	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.45	2.31	2.23	2.15	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.41	2.27	2.19	2.11	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.38	2.23	2.16	2.07	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.35	2.20	2.12	2.04	1.84
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.30	2.15	2.07	1.98	1.78
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.25	2.11	2.03	1.94	1.73
26	4.23	3.37	2.98	2.74	2.59	2.47	2.32	2.22	2.07	1.99	1.90	1.69
28	4.20	3.34	2.95	2.71	2.56	2.45	2.29	2.19	2.04	1.96	1.87	1.65
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.16	2.01	1.93	1.84	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.08	1.92	1.84	1.74	1.51
50	4.03	3.18	2.79	2.56	2.40	2.29	2.13	2.03	1.87	1.78	1.69	1.44
60	4.00	3.15	2.76	2.53	2.37	2.25	2.10	1.99	1.84	1.75	1.65	1.39
80	3.96	3.11	2.72	2.49	2.33	2.21	2.06	1.95	1.79	1.70	1.60	1.32
100	3.94	3.09	2.70	2.46	2.31	2.19	2.03	1.93	1.77	1.68	1.57	1.28
$+\infty$	3.84	3.00	2.60	2.37	2.21	2.10	1.94	1.83	1.67	1.57	1.46	1.00

Bibliographie

- [1] Mohamed Boucetta. **Éléments d'analyse réelle**. Ellipses.
- [2] Nicolas Bouleau. **Probabilités de l'ingénieur**. Hermann.
- [3] Pierre Brémaud **Introduction aux Probabilités**. Springer.
- [4] Marie-Line Chabanol et Jean-Jacques Ruch. **Probabilités et statistiques pour l'épreuve de modélisation à l'agrégation de mathématiques**. Ellipses.
- [5] Jacques Dauxois et Claudie Hassenforder. **Toutes les probabilités et les statistiques**. Ellipses.
- [6] Olivier Garet et Aline Kurtzmann **De l'intégration aux probabilités**. Ellipses.
- [7] Claude Gasquet et Patrick Witomski **Analyse de Fourier et Applications**. Dunod.
- [8] Jean Jacod et Philip Protter **L'essentiel en théorie des probabilités**. Cassini.
- [9] Benjamin Jourdain. **Probabilités et statistique**. Ellipses.
- [10] Jean-Claude Laleuf. **Processus et intégrales stochastiques**. Ellipses.
- [11] Sabin Lessard et Monga. **Statistique concepts et méthodes**. Masson.

- [12] Olivier Marchal. **Fondements des probabilités**. Ellipses.
- [13] Olivier Marchal. **Statistiques appliquées**. Ellipses.
- [14] Olivier Marchal. **Cours et exercices corrigés de statistiques inférentielles**. Ellipses.
- [15] Étienne Marceau. **Modélisation et évaluation quantitative des risques en actuariat**. Springer.
- [16] Raymond A. Marie. **Introduction aux probabilités**. Ellipses.
- [17] Abderrahmane Ouagga. **Analyse. Fonctions d'une à plusieurs variables réelles**. Ellipses.
- [18] Jean-Yves Ouvrard. **Probabilités 1, capes et agrégation**. Cassini.
- [19] Jean-Yves Ouvrard. **Probabilités 2, master et agrégation**. Cassini.
- [20] Jean Pellaumail. **Probabilités, statistiques, files d'attente**. Dunod.
- [21] Patrick Royis. **Éléments de théorie des probabilités**. Ellipses.
- [22] Walter Rudin. **Analyse réelle et complexe**. Dunod.

Index

A			
Algèbre de Boole	341	Convergence p.s.	202, 207, 271, 294
Algorithme	302	Corrélation linéaire	74, 255
Axiome des probabilités totales	29,	Courbe des fréquences cumulées	233
	81, 82, 199	Covariance	73, 128, 194, 255, 256,
			330
B			
Biais	267, 270, 288	D	
Bijection	20	Dénombrabilité	20
Borélien	345, 352	Densité conditionnelle	191
Box-plot	240	Densité de probabilité	118, 119, 122,
			131, 136, 138, 148, 174, 180,
C			189, 197, 211, 236
Càdlàg	55	Diagramme à moustaches	240
Caractère	226, 259, 307	Diagramme circulaire	228
Caractère continu	227	Diagramme de Venn	9
Caractère discret	227	Diagramme en barres	229
Caractère qualitatif	226, 227	Diagramme en bâtons	60, 87, 91,
Caractère qualitatif nominal	226		96–98, 101–105, 231
Caractère qualitatif ordinal	226	Dirac	136
Caractère quantitatif	226, 230, 255	Dispersion	69, 71, 126
Cardinal d'un ensemble	20	Distribution	50, 53, 60, 77, 87, 93,
Cas gaussien	155, 162, 217, 288, 289		118, 135, 148, 172, 187, 250,
Centrer	68, 126		271
Classes	232, 335	Distribution conditionnelle	250
Classes modales	234	Distribution régulière	173
Coefficient de corrélation linéaire	331	Données brutes	227
Coefficient de dispersion	246	Données groupées	227
Collections	335	Droite de régression	329
Composante	184, 188	E	
Convergence	199	Écart-type	69, 71, 72, 126, 227, 242
Convergence dans L^p	204, 207, 271	Échantillon	226, 258, 303
Convergence en loi	90, 94, 97, 109,	Effectif marginal	248
	199, 209, 218, 294, 325, 326	Effectifs corrigés	233
Convergence en proba	204, 206, 207,	Effectifs cumulés	231
	261, 268		

- Ensemble des réalisations possibles 46
 Ensembles 335
 Ensembles d'ensembles 335, 341
 Ensembles disjoints 13, 25, 39
 Ensembles négligeables 348
 Équiprobabilité 33, 96
 Erreur de première espèce 303, 304, 308
 Erreur de seconde espèce 303, 304
 Erreur quadratique moyenne 270, 272, 273
 Espace fondamental 22
 Espace mesurable 347
 Espace mesuré 347
 Espace probabilisé 347
 Espérance 62, 63, 97, 122, 125, 135, 137, 144, 167, 236, 238, 260
 Estimateur 266, 276, 281
 Estimateur convergent 268, 308
 Estimateur sans biais 264, 267, 308
 Estimation ponctuelle 243, 257, 287
 Étendue 246
 Évènements 24, 27, 81, 342, 348
 Expérience aléatoire 21
 Extinction des grands noms 115
- F**
- Familles 335
 Fonction caractéristique 109, 171, 180, 186, 189, 196, 211
 Fonction de répartition 51, 53, 90, 94, 117, 119, 133, 137, 140, 145, 149, 185, 189, 210, 218, 231, 235, 270, 272, 326, 353
 Fonction de survie 100
 Fonction en escalier 338, 349
 Fonction étagée 349
 Fonction Gamma d'Euler 153, 155, 162
 Fonction génératrice 109
 Fonction mesurable 344
 Forme quadratique 196
 Formule de transfert 65, 110, 122, 134, 351
 Fréquence conditionnelle 249
 Fréquence marginale 249
 Fréquence statistique 265
- Fréquences 38, 230
 Fréquences cumulées 231
- H**
- Histogramme 233
 Homoscédasticité 293, 316, 317
 Hypothèse alternative 302
 Hypothèse nulle 302
- I**
- I.I.D. 57, 201, 214, 216, 217, 219, 292, 294, 314, 316, 320
 Implication 25
 Indépendance 38, 39, 55, 67, 73, 79, 82, 87, 92, 112, 114, 125, 128, 142, 183, 189, 194, 214, 250, 252
 Inégalité de Berry-Esseen 218
 Inégalité de Bienaymé-Tchebychev 70, 127, 214, 262, 269, 289
 Inégalité de Cauchy-Schwarz 194
 Inégalité de Jensen 205
 Inégalité de Markov 67, 70, 125
 Intégrale de Lebesgue 349, 351
 Intégrale de Riemann 338
 Interquartile 247
 Intervalle interquartile 247
 Intervalles de confiance 162, 198, 217, 243, 270, 287, 293
- K**
- Kurtosis 129, 264
- L**
- λ -système 342, 343
 Lemme de Fatou 351
 Lemme de Slutsky 210, 294, 319
 Lemmes de Borel-Cantelli 41–43
 LGN de Kolmogorov 216
 Limite inférieure 336
 Limite supérieure 336
 Linéarité de l'espérance 65, 68, 85, 93, 124, 353
 Linéarité de l'intégrale 351, 353
 Loi binomiale 79, 111, 175, 219, 265, 294
 Loi conditionnelle 190

- Loi conjointe 185, 190
 Loi d'Erlang 138, 151, 153, 155, 179, 278, 283, 295
 Loi de Benford 104, 112
 Loi de Bernoulli 78, 84, 86, 87, 95, 110, 175, 218, 265, 277, 281, 293, 294, 297
 Loi de Cauchy 122, 167, 179, 279, 285
 Loi de Fisher 165, 318
 Loi de Kolmogorov-Smirnov 326
 Loi de Laplace 126, 179
 Loi de Pareto 166, 278, 284
 Loi de Poisson 90, 111, 176, 219, 258, 276, 293, 296
 Loi de probabilité 47, 49–51, 68, 95, 98, 171, 184, 189, 209, 249, 258, 353
 Loi de Rademacher 95, 111, 176, 190, 195, 200, 209
 Loi de Rayleigh 158
 Loi de Student 162, 167, 183, 291, 309
 Loi de Weibull 152, 159, 278, 284
 Loi de Zipf 103, 112
 Loi du Khi-deux 155, 162, 264, 291, 312, 320
 Loi du tout ou rien 42
 Loi \mathcal{E} 136, 151, 152, 155, 177, 277, 282, 295
 Loi faible des grands nombres 214, 271
 Loi forte des grands nombres 202, 211, 215, 294, 329
 Loi Gamma 153, 155, 275, 295
 Loi géométrique 99, 112, 136, 177, 200, 201, 277, 282
 Loi hypergéométrique 102, 112
 Loi log-normale 160
 Loi logarithmique 104, 112
 Loi logistique 149
 Loi normale 90, 94, 138, 156, 158, 160, 162, 163, 178, 193, 262, 270, 277, 283, 289, 294, 308
 Loi normale centrée réduite 142, 145, 177, 190, 217, 218, 255, 291, 294, 308, 309, 317
 Loi parente 257, 266
 Loi sans mémoire 101, 136
 Loi triangulaire 147, 178
 Loi triangulaire discrète 97, 111, 176
 Loi \mathcal{U} 97, 118, 119, 122, 123, 132, 177, 239, 254, 266, 271, 273, 276, 277, 282
 Loi uniforme discrète 96, 111, 176
 Loi zêta 105, 112
 Lois de Morgan 17, 41
 Lois des grands nombres 71, 213, 279
- ### M
- Marginale 185, 190, 247, 330
 Matrice de covariance 139, 195
 Maximum de vraisemblance 271, 274, 275, 279, 281, 323
 Médiane 227
 Mesure 348, 349
 Mesure de probabilité 347
 Méthode des moments 274, 275, 277, 280, 281, 284
 Méthodes de Monte Carlo 214
 Modalités 226, 247
 Mode 98, 231, 234
 Modèle binomial 79, 83
 Moment 78, 279
 Moyenne 227, 236, 238
 Moyenne conditionnelle 256
 Moyenne d'échantillon 260, 267, 269–271, 273, 283, 308
- ### N
- Normalité asymptotique 270, 271
 Nuage de points 253, 329
- ### O
- Ouvert 338
- ### P
- Partition de l'univers 18, 26, 114, 238
 π -système 342, 343, 352
 Population mère 258
 Principe d'Heisenberg 178
 Probabilité 27, 62, 199, 352
 Probabilité conditionnelle 35, 36, 279
 Probabilité d'un intervalle 121, 140

- Processus de Galton-Watson 115
 Produit de convolution 77, 87, 93,
 131, 135, 147, 151, 154, 180,
 271
 Proportion d'échantillon 265, 270,
 293
- Q**
- Qualité des estimateurs 266
 Quantile 146, 165, 166, 233, 239,
 289, 291, 294, 308
 Queue de distribution 71
- R**
- Recensement 226
 Réduire 74, 129
 Région critique 302, 308
 Régression linéaire 253, 329
 Répartition uniforme 233, 235, 238
- S**
- Série à valeurs classées 232, 241,
 244, 255
 Série à valeurs isolées 230, 241, 255
 Série entière 110
 Séries statistiques 227
 Séries statistiques doubles 247
 Séries statistiques simples 230
 Skewness 129
 Sous-tribu 343
 Statistiques 155, 162, 198, 223, 259,
 266, 267, 300, 301
 Statistiques descriptives 225
 Statistiques inférentielles 243, 258
 Statistiques libres 326
- T**
- T.D.S.S.S. 164, 174
 Tableau de contingence 247
 Test d'adéquation 301
 Test d'ajustement 301
 Test d'association 302
 Test d'homogénéité 301
 Test d'indépendance 252, 301
 Test de comparaison 301
 Test de conformité 301
 Test de Kolmogorov-Smirnov 325
- Tests du Khi-deux 198, 210, 303
 Tests non paramétriques 300
 Tests paramétriques 300
 Tests statistiques 165, 299, 307
 Théorème central de la limite 138,
 213, 217, 219, 270, 271, 289,
 294, 300, 307, 317
 Théorème de Bayes 37, 279
 Théorème de Cochran 198, 264
 Théorème de convergence dominée
 de Lebesgue 110, 173, 210,
 352
 Théorème de convergence monotone
 350
 Théorème de Pearson 320
 Théorie de la mesure 172, 174, 179,
 206, 335, 338, 341
 Théorie probabiliste des nombres 106
 Topologie 338
 Transformée de Fourier 171, 172,
 178, 180
 Transformée de Laplace 171
 Tribu 24, 27, 36, 49, 51, 119, 259,
 338, 341, 342, 348, 352
 Tribu borélienne 345
 Tribu complétée 348
 Tribu engendrée 344, 345
 Tribu image 345
- U**
- Univers 22, 38, 46, 258
- V**
- Valeur principale de Cauchy 179
 Valeur R^2 331
 Variable aléatoire 45, 48, 171, 226,
 258, 352
 Variable aléatoire à densité 117, 353
 Variable aléatoire continue 47
 Variable aléatoire discrète 47, 59,
 117, 121, 353
 Variance 68, 69, 72, 126, 135, 137,
 144, 167, 242, 260, 270, 330
 Variance corrigée 264, 268
 Variance d'échantillon 262, 268, 270,
 273, 283
 Variance de la somme 73

<i>INDEX</i>		377
Vecteur propre	178	
Vecteurs aléatoires	119, 164, 183, 353	
Vecteurs aléatoires à densité	188	
Vecteurs aléatoires discrets		187
Vecteurs aléatoires gaussiens		193, 264
Vraisemblance		280