Statistiques

Deux méthodes classiques d'estimation ponctuelle

Julian Tugaut

Télécom Saint-Étienne

Sommaire

Plan

Introduction - 1

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.

Introduction - 2

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.

Plan

Méthode des moments - Modèle

Soit un modèle statistique indexé par un ensemble Θ. Supposons donné un n-échantillon $(X_1, \cdots, X_n)$ d'une loi intégrable sous $\mathbb{P}_\theta$ pour chaque $\theta \in \Theta$. On cherche un estimateur $\widehat{\theta_n}$ de θ^*.

Méthode des moments - Modèle

Soit un modèle statistique indexé par un ensemble Θ. Supposons donné un n-échantillon $(X_1, \cdots, X_n)$ d'une loi intégrable sous $\mathbb{P}_\theta$ pour chaque $\theta \in \Theta$. On cherche un estimateur $\widehat{\theta_n}$ de θ^*. Supposons dans un premier temps que Θ est une partie de $\mathbb{R}$.

Méthode des moments - Modèle

Soit un modèle statistique indexé par un ensemble Θ. Supposons donné un n-échantillon $(X_1, \cdots, X_n)$ d'une loi intégrable sous $\mathbb{P}_\theta$ pour chaque $\theta \in \Theta$. On cherche un estimateur $\widehat{\theta_n}$ 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 $\mathbb{E}_{\widehat{\theta_n}} =: m\left(\widehat{\theta_n}\right)$ coïncide avec la moyenne empirique observée.

Exemple introductif - 1

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 $\mathbb{R}$.

Exemple introductif - 1

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

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i = m\left(\widehat{\theta_n}\right) .$$

Exemple introductif - 2

Si m réalise une bijection de Θ dans $m(\Theta)$, alors $\widehat{\theta_n} = m^{-1}\left(\overline{X_n}\right)$.

Exemple introductif - 2

Si m réalise une bijection de Θ dans $m(\Theta)$, alors $\widehat{\theta_n} = m^{-1}\left(\overline{X_n}\right)$.
Pour la loi de Poisson de paramètre θ^*, la moyenne est égale à θ^*.
De fait, on prend $\widehat{\theta_n} := \overline{X_n}$.

Exemple introductif - 2

Si m réalise une bijection de Θ dans $m(\Theta)$, alors $\widehat{\theta_n} = m^{-1}\left(\overline{X_n}\right)$.
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 $\frac{1}{n}\sum_{i=1}^{n} x[i]$.

Méthode générale - 1

D'abord, on suppose que Θ est une partie de $\mathbb{R}^d$ avec $d \geq 1$.

Méthode générale - 1

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.

Méthode générale - 1

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 $\left(\mathbb{E}_\theta(X^{k_1}), \cdots, \mathbb{E}_\theta(X^{k_d})\right)$ est injective ; où $1 \leq k_1 < \cdots < k_d \leq p$.

Méthode générale - 2

Exemple

Si l'on cherche le paramètre $\theta \subset \mathbb{R}_+^*$ 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.

Méthode générale - 2

Exemple

Si l'on cherche le paramètre $\theta \subset \mathbb{R}_+^*$ 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

$$\frac{1}{n}\sum_{i=1}^{n} X_i^{k_l} = m_{k_l}\left(\widehat{\theta_n}\right) ,$$

pour tout $l \in [\![1; d]\!]$ où $m_k(\widehat{\theta_n}) := \mathbb{E}_{\widehat{\theta_n}}(X^k)$ pour tout $k \geq 1$.

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.

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 $\widehat{p_n} := \frac{1}{\overline{X_n}} = \frac{n}{\sum_{k=1}^n X_k}$.

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 $\mathrm{Var}[X] = \frac{(b-a)^2}{12}$. Ainsi, on peut écrire

$$a = \mathbb{E}[X] - \sqrt{3}\sqrt{\mathrm{Var}[X]} \quad \text{et} \quad b = \mathbb{E}[X] - \sqrt{3}\sqrt{\mathrm{Var}[X]}.$$

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 $\mathrm{Var}[X] = \frac{(b-a)^2}{12}$. Ainsi, on peut écrire

$$a = \mathbb{E}[X] - \sqrt{3}\sqrt{\mathrm{Var}[X]} \quad \text{et} \quad b = \mathbb{E}[X] - \sqrt{3}\sqrt{\mathrm{Var}[X]}.$$

Par conséquent, on prendra les estimateurs de la méthode des moments sont $\widehat{a_n} := \overline{X_n} - \sqrt{3}S_n$ et $\widehat{b_n} := \overline{X_n} + \sqrt{3}S_n$.

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 $\widehat{\lambda_n} := \frac{1}{\overline{X_n}} = \frac{n}{\sum_{k=1}^{n} X_k}$.

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 $\widehat{\sigma_n^2} := S_n^2$.

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 $\widehat{\lambda_n} := \frac{r}{\overline{X_n}} = \frac{nr}{\sum_{k=1}^n X_k}$.

Loi de Pareto - 1

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

$$\int_{x_0}^{\infty} \alpha \left(\frac{x_0}{x}\right)^{\alpha} \mathrm{d}x = \frac{\alpha}{\alpha - 1} x_0 \,.$$

Alors, on prendra comme estimateur $\widehat{\alpha_n} := \frac{\overline{X_n}}{\overline{X_n} - x_0}$.

Loi de Pareto - 2

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

$$\mathrm{Var}[X] = \frac{\alpha x_0^2}{(\alpha - 1)^2 (\alpha - 2)} \,.$$

Loi de Pareto - 2

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

$$\mathrm{Var}[X] = \frac{\alpha x_0^2}{(\alpha - 1)^2(\alpha - 2)} \,.$$

On est ensuite amené à résoudre

$$\frac{\alpha}{\alpha - 1} x_0 = \overline{X_n} \quad \text{et} \quad \frac{\alpha x_0^2}{(\alpha - 1)^2(\alpha - 2)} = S_n^2 \,.$$

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.

Limites - 1

Limites - 1

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.

Limites - 2

Limites - 2

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.

Limites - 2

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.

Plan

Philosophie de la méthode - 1

Philosophie de la méthode - 1

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 : $\theta_1, \cdots, \theta_r$.

Philosophie de la méthode - 1

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 : $\theta_1, \cdots, \theta_r$.
On dispose d'un n-échantillon $(X_1, \cdots, X_n)$ dont les valeurs que l'on a observées sont $(x[1], \cdots, x[n])$. On suppose enfin que la variable aléatoire X est discrète quelle que soit la valeur de θ.

Philosophie de la méthode - 2

On admet également que l'on ne dispose d'aucun *a priori* (aussi appelé *prior* en statistiques bayésiennes) sur le choix entre $\theta_1, \cdots, \theta_k$. En d'autres termes, $\mathbb{P}(\theta = \theta_1) = \cdots = \mathbb{P}(\theta = \theta_r) = \frac{1}{r}$.

Philosophie de la méthode - 2

On admet également que l'on ne dispose d'aucun *a priori* (aussi appelé *prior* en statistiques bayésiennes) sur le choix entre $\theta_1, \cdots, \theta_k$. En d'autres termes, $\mathbb{P}(\theta = \theta_1) = \cdots = \mathbb{P}(\theta = \theta_r) = \frac{1}{r}$. Alors, d'après la formule de Bayes, on a :

$$\mathbb{P}(\theta = \theta_l \mid X_1 = x[1], \cdots, X_n = x[n]) = \frac{\prod_{i=1}^n \mathbb{P}(X = x[i] \mid \theta = \theta_l)}{\sum_{k=1}^r \prod_{i=1}^n \mathbb{P}(X = x[i] \mid \theta = \theta_k)} .$$

Philosophie de la méthode - 3

Il est alors raisonnable de choisir le mode, c'est-à-dire dans le cas présent la valeur I telle que la probabilité conditionnelle $\mathbb{P}(\theta = \theta_I \mid X_1 = x[1], \cdots, X_n = x[n])$ soit la plus élevée. Pour ce faire, il suffit de maximiser la quantité $\prod_{i=1}^n \mathbb{P}(X = x[i] \mid \theta = \theta_I)$.

Philosophie de la méthode - 3

Il est alors raisonnable de choisir le mode, c'est-à-dire dans le cas présent la valeur I telle que la probabilité conditionnelle $\mathbb{P}(\theta = \theta_I \mid X_1 = x[1], \cdots, X_n = x[n])$ soit la plus élevée. Pour ce faire, il suffit de maximiser la quantité $\prod_{i=1}^n \mathbb{P}(X = x[i] \mid \theta = \theta_I)$. Cette quantité sera appelée vraisemblance. On cherche ainsi à maximiser la vraisemblance.

Exemple introductif - 1

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.

Exemple introductif - 1

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.

Exemple introductif - 2

La vraisemblance est ici :

$$L_{(x[1],\cdots,x[n])}(\theta) := \prod_{i=1}^{n} \frac{\theta^{x[i]}}{x[i]!} e^{-\theta} = \frac{\exp\left[-\left(n\theta - \sum_{i=1}^{n} x[i] \log(\theta)\right)\right]}{\prod_{i=1}^{n} x[i]!}\,.$$

Exemple introductif - 3

Pour maximiser cette fonction, on minimise $\theta \mapsto -\log(L_{(x[1],\cdots,x[n])}(\theta))$. Pour ce faire, on dérive et l'on a un unique point critique qui est atteint en

$$\widehat{\theta_n} := \frac{1}{n}\sum_{i=1}^{n} x[i] \,.$$

Exemple introductif - 3

Pour maximiser cette fonction, on minimise $\theta \mapsto -\log(L_{(x[1],\cdots,x[n])}(\theta))$. Pour ce faire, on dérive et l'on a un unique point critique qui est atteint en

$$\widehat{\theta_n} := \frac{1}{n}\sum_{i=1}^{n} x[i]\,.$$

On peut ensuite vérifier que ce minimum critique correspond bien à un maximum global.

Exemple introductif - 3

Pour maximiser cette fonction, on minimise $\theta \mapsto -\log(L_{(x[1],\cdots,x[n])}(\theta))$. Pour ce faire, on dérive et l'on a un unique point critique qui est atteint en

$$\widehat{\theta_n} := \frac{1}{n}\sum_{i=1}^{n} x[i]\,.$$

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.

Méthode générale - 1

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

$$f(x,\theta) := \mathbb{P}_\theta\left(X = x\right)$$

si X est une variable aléatoire discrète.

Méthode générale - 2

Définition

On appelle fonction de vraisemblance de θ pour une réalisation $(x[1], \cdots, x[n])$ d'un échantillon, la fonction de θ :

$$L_{(x[1],\cdots,x[n])}(\theta) := \prod_{i=1}^{n} f(x[i], \theta) \,.$$

Méthode générale - 2

Définition

On appelle fonction de vraisemblance de θ pour une réalisation $(x[1], \cdots, x[n])$ d'un échantillon, la fonction de θ :

$$L_{(x[1],\cdots,x[n])}(\theta) := \prod_{i=1}^{n} f(x[i], \theta) \,.$$

La méthode qui consiste à estimer θ par la valeur qui maximise $L_{(x[1],\cdots,x[n])}$ (la vraisemblance) s'appelle "méthode du maximum de vraisemblance".

Méthode générale - 3

On appelle $\hat{\theta}$ l'estimateur associé :

$$\hat{\theta} := \left\{ \theta \;:\; L_{(x[1],\cdots,x[n])}(\hat{\theta}) = \sup_{\theta \in \Theta} L_{(x[1],\cdots,x[n])}(\theta) \right\} .$$

Méthode générale - 3

On appelle $\hat{\theta}$ l'estimateur associé :

$$\hat{\theta} := \left\{ \theta \ : \ L_{(x[1],\cdots,x[n])}(\hat{\theta}) = \sup_{\theta \in \Theta} L_{(x[1],\cdots,x[n])}(\theta) \right\} .$$

Ceci est un problème d'optimisation. On utilise généralement le fait que si $L_{(x[1],\cdots,x[n])}$ est dérivable et si $L_{(x[1],\cdots,x[n])}$ admet un maximum global en une valeur, alors la dérivée première de $L_{(x[1],\cdots,x[n])}$ s'y annule et sa dérivée seconde y est négative.

Méthode générale - 4

Réciproquement, si la dérivée première de $L_{(x[1],\cdots,x[n])}$ s'annule en $\theta_0 := \theta^*$, et si sa dérivée seconde est négative strictement en θ^*, alors θ^* est un point où $L_{(x[1],\cdots,x[n])}$ admet un maximum local (et non global). Il est alors nécessaire de vérifier que le maximum est global.

Méthode générale - 4

Réciproquement, si la dérivée première de $L_{(x[1],\cdots,x[n])}$ s'annule en $\theta_0 := \theta^*$, et si sa dérivée seconde est négative strictement en θ^*, alors θ^* est un point où $L_{(x[1],\cdots,x[n])}$ 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.

Loi de Bernoulli - 1

On suppose que X suit la loi de Bernoulli de paramètre $p \in]0;1[$. On observe n valeurs $x[1], \cdots, x[n]$. 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

$$L_{(x[1],\cdots,x[n])}(p) := (1-p)^{n(1-q)} p^{nq} \, .$$

Loi de Bernoulli - 2

On prend le logarithme népérien et on dérive par rapport à p pour trouver l'unique point critique qui est

$$\widehat{p_n} = q\,.$$

Loi de Bernoulli - 2

On prend le logarithme népérien et on dérive par rapport à p pour trouver l'unique point critique qui est

$$\widehat{p_n} = q\,.$$

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.

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

$$L_{(x[1],\cdots,x[n])}(p) = \prod_{i=1}^{n} p(1-p)^{x[i]-1} = \left(\frac{p}{1-p}\right)^n (1-p)^{\sum_{i=1}^{n} x[i]} .$$

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

$$L_{(x[1],\cdots,x[n])}(p) = \prod_{i=1}^{n} p(1-p)^{x[i]-1} = \left(\frac{p}{1-p}\right)^n (1-p)^{\sum_{i=1}^{n} x[i]}.$$

On prend le logarithme népérien, on dérive et l'on aboutit à l'unique point critique suivant :

$$p = \frac{n}{\sum_{i=1}^{n} x[i]}.$$

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

$$L_{(x[1],\cdots,x[n])}(p) = \prod_{i=1}^{n} p(1-p)^{x[i]-1} = \left(\frac{p}{1-p}\right)^n (1-p)^{\sum_{i=1}^{n} x[i]}.$$

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.

Loi uniforme - 1

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

$$L_{(x[1],\cdots,x[n])}(a,b) = \frac{1}{(b-a)^n}\prod_{i=1}^{n}\mathbb{1}_{[a;b]}(x[i])\,.$$

Loi uniforme - 1

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

$$L_{(x[1],\cdots,x[n])}(a, b) = \frac{1}{(b-a)^n}\prod_{i=1}^{n}\mathbb{1}_{[a;b]}(x[i])\,.$$

Pour maximiser, il faut d'abord s'assurer que $x[i] \in [a; b]$ pour tout $i \in [\![1; n]\!]$. On a donc $\widehat{a_n} \leq \min(X_1, \cdots, X_n)$ et $\max(X_1, \cdots, X_n) \leq \widehat{b_n}$. Puis, il faut minimiser $b - a$ en respectant cette contrainte.

Loi uniforme - 2

On aboutit donc à

$$\widehat{a_n} := \min(X_1, \cdots, X_n) \quad \text{et} \quad \widehat{b_n} = \max(X_1, \cdots, X_n)\,.$$

Loi uniforme - 2

On aboutit donc à

$$\widehat{a_n} := \min(X_1, \cdots, X_n) \quad \text{et} \quad \widehat{b_n} = \max(X_1, \cdots, X_n)\,.$$

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.

Loi exponentielle

On suppose que X suit la loi exponentielle de paramètre $\lambda > 0$. Alors, la vraisemblance est

$$L_{(x[1],\cdots,x[n])}(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n x[i]} \prod_{i=1}^n H(x[i])\,.$$

Loi exponentielle

On suppose que X suit la loi exponentielle de paramètre $\lambda > 0$. Alors, la vraisemblance est

$$L_{(x[1],\cdots,x[n])}(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n x[i]} \prod_{i=1}^n H(x[i])\,.$$

On prend le logarithme népérien, on dérive et l'on trouve un unique point critique :

$$\lambda_0 = \frac{n}{\sum_{i=1}^n x[i]}\,.$$

Loi exponentielle

On suppose que X suit la loi exponentielle de paramètre $\lambda > 0$. Alors, la vraisemblance est

$$L_{(x[1],\cdots,x[n])}(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n x[i]} \prod_{i=1}^n H(x[i])\,.$$

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.

Loi normale - 1

On suppose que X suit la loi normale de paramètres m et $\sigma^2 > 0$. Alors, la vraisemblance est :

$$L_{(x[1],\cdots,x[n])}(m,\sigma^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x[i]-m)^2\right\}.$$

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.

Loi normale - 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.

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

$$L_{(x[1],\cdots,x[n])} = \lambda^{nr} \frac{\prod_{i=1}^{n} x[i]^{r-1}}{((r-1)!)^n} e^{-\lambda \sum_{i=1}^{n} x[i]} \prod_{i=1}^{n} H(x[i]) \,.$$

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

$$L_{(x[1],\cdots,x[n])} = \lambda^{nr} \frac{\prod_{i=1}^n x[i]^{r-1}}{((r-1)!)^n} e^{-\lambda \sum_{i=1}^n x[i]} \prod_{i=1}^n H(x[i]) \,.$$

Le calcul fournit immédiatement $\widehat{\lambda_n} = \frac{r}{\overline{X_n}}$, ce qui correspond à la méthode des moments.

Loi de Pareto - 1

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 :

$$L_{(x[1],\cdots,x[n])}(\alpha) = \alpha^n x_0^{n\alpha} \left(\prod_{i=1}^{n} x[i] \right)^{-\alpha-1} \prod_{i=1}^{n} \mathbb{1}_{[x_0;+\infty[}(x[i]) \,.$$

Loi de Pareto - 1

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 :

$$L_{(x[1],\cdots,x[n])}(\alpha) = \alpha^n x_0^{n\alpha} \left(\prod_{i=1}^{n} x[i] \right)^{-\alpha-1} \prod_{i=1}^{n} \mathbb{1}_{[x_0;+\infty[}(x[i])\,.$$

On prend le logarithme népérien et l'on dérive par rapport à α et l'on obtient :

$$\frac{\partial}{\partial \alpha} \log\left(L_{(x[1],\cdots,x[n])}(\alpha) \right) = \frac{n}{\alpha} + n \log(x_0) - \sum_{i=1}^{n} \log(x[i])\,.$$

Loi de Pareto - 2

Ainsi, l'estimateur associé est

$$\widehat{\alpha_n} := \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{x[i]}{x_0}\right)} .$$

Loi de Pareto - 2

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

$$\widehat{x_0(n)} := \min\left(X_1, \cdots, X_n\right)$$

et de fait

$$\widehat{\alpha_n} := \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{x[i]}{\min(x[1], \cdots, x[n])}\right)} \,.$$

Loi de Pareto - 3

Ainsi, on pouvait définir l'estimateur du maximum de vraisemblance sans hypothèse supplémentaire ; contrairement à la méthode des moments.

Loi de Cauchy - 1

On trouve ici une vraisemblance égale à

$$L_{(x[1],\cdots,x[n])}(a,b) = \frac{1}{\pi^n} \frac{b^n}{\prod_{i=1}^{n} \left(b^2 + (x[i]-a)^2\right)} \,.$$

Dérivons la log-vraisemblance par rapport à b :

$$\frac{\partial}{\partial b} \log\left(L_{(x[1],\cdots,x[n])}(a,b)\right) = \frac{n}{b} - 2b \sum_{i=1}^{n} \frac{1}{b^2 + (x[i]-a)^2} \,.$$

Loi de Cauchy - 2

On dérive maintenant par rapport à a :

$$\frac{\partial}{\partial a} \log \left(L_{(x[1],\cdots,x[n])}(a,b) \right) = 2 \sum_{i=1}^{n} \frac{x[i]-a}{b^2 + (x[i]-a)^2} \, .$$

Le calcul explicite nécessite d'utiliser les outils numériques pour trouver a_0 et b_0 satisfaisant $\frac{\partial}{\partial a} \log \left(L_{(x[1],\cdots,x[n])}(a_0,b_0) \right) = 0$ et $\frac{\partial}{\partial b} \log \left(L_{(x[1],\cdots,x[n])}(a_0,b_0) \right) = 0$.