Karhunen-Loève decomposition of Gaussian measures on Banach spaces

Jean-Charles Croix

GT APSSE - April 2017, the 13th

joint work with Xavier Bay.
Sommaire

1 Preliminaries on Gaussian processes

2 Gaussian measures on Banach spaces
   - Definitions
   - Karhunen-loève decomposition in the Hilbert case
   - Karhunen-loève decomposition on Banach spaces

3 Examples on continuous functions

4 Conclusions
Consider:

- $\mathcal{T}$ a non-empty set,
- $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space.

A square integrable stochastic process is a collection $X = (X_t)_{t \in \mathcal{T}}$ of square integrable real random variables:

$$\forall t \in \mathcal{T}, \ X_t : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \in L^2(\Omega)$$
Preliminaries on Gaussian processes

Square integrable stochastic process

Consider:
- $\mathcal{T}$ a non-empty set,
- $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space.

A square integrable stochastic process is a collection $X = (X_t)_{t \in \mathcal{T}}$ of square integrable real random variables:

$$\forall t \in \mathcal{T}, \ X_t : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \in L^2(\Omega)$$

Mean function and covariance kernel

Let $X$ be a square integrable process, then
- $\forall t \in \mathcal{T}, \ m(t) = \mathbb{E}[X_t]$,
- $\forall (t, s) \in \mathcal{T}^2, \ K(s, t) = \mathbb{E}[(X_t - m(t))(X_s - m(s))]$.

From now on, all processes will be centered: $m \equiv 0$. 
The covariance kernel $K$ is a symmetric kernel:

$$\forall (s, t) \in \mathcal{T}^2, \ K(s, t) = K(t, s),$$

and of positive type:

$$\forall n \in \mathbb{N}, \ \forall (t_1, \ldots, t_n) \in \mathcal{T}, \ \forall (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n, \ \sum_{i=1}^{n} \alpha_i \alpha_j K(t_i, t_j) \geq 0.$$ 

In short, $K$ is a positive definite kernel on $\mathcal{T}$.

**Gaussian process**

A square integrable stochastic process $X$ is Gaussian if and only if:

$$\forall n \in \mathbb{N}, \ \forall (t_1, \ldots, t_n) \in \mathcal{T}^n, \ (X_{t_1}, \ldots, X_{t_n}) \text{ is a Gaussian vector.}$$

Equivalently,

$$\forall \mu = \sum_{i=1}^{n} \alpha_i \delta_{t_i}, \ \langle X, \mu \rangle = \sum_{i=1}^{n} \alpha_i X_{t_i} \text{ is a real Gaussian variable.}$$
**Preliminaries on Gaussian processes**

### Gaussian space

Let $X$ be a Gaussian process $X$ on $\mathcal{T}$, the associated Gaussian space is

$$H_X = \{X_t, t \in \mathcal{T}\}_{L^2(\Omega, \mathcal{F}, \mathbb{P})}$$

The mapping

$$U : Z \in H_X \rightarrow (t \rightarrow \langle Z, X_t \rangle_{L^2}) \in \mathbb{R}^\mathcal{T}$$

is a linear and injective.

### Reproducing Kernel Hilbert space (RKHS)

Let $X$ be a Gaussian process on $\mathcal{T}$, the associated RKHS is

$$\mathcal{H}_X = U(H_X).$$

It is a Hilbert space with $\langle f, g \rangle_{\mathcal{H}_X} = \langle U^{-1}(f), U^{-1}(g) \rangle_{H_X}$ as inner product.
The previous mapping $U$ is an isometric isomorphism, thus

$$H_K \equiv \mathcal{H}_K$$

as Hilbert spaces.

**Moore-Aronszajn theorem**

Let $T$ a set and $K$ a positive-definite kernel on $T$, then there exists a unique Hilbert subspace $\mathcal{H}_K \subset \mathbb{R}^T$ such that:

1. $\mathcal{H}_K = \{K(t,.) : t \in T\}$,
2. $\forall t \in T, \forall h \in \mathcal{H}_K$, $h(t) = \langle h, K_t \rangle_{\mathcal{H}_K}$.

Conversely, let $T$ be a set and $K$ a positive definite symmetric kernel, there exists a Gaussian process on $T$ with $K$ as covariance kernel (Kolmogorov extension theorem).
Let $\mathcal{T}$ be a set and $\mathcal{C}_\mathcal{T}$ the cylindrical $\sigma$-algebra on $\mathbb{R}^\mathcal{T}$. Any stochastic process $X$ on $\mathcal{T}$ is a $(\mathcal{C}_\mathcal{T}, \mathcal{F})$-measurable mapping

$$X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}^\mathcal{T}, \mathcal{C}_\mathcal{T}).$$

The process $X$ defines a probability measure on $(\mathbb{R}^\mathcal{T}, \mathcal{C}_\mathcal{T})$

$$\mu_X = \mathbb{P} \circ X^{-1}.$$  

The dual space $(\mathbb{R}^\mathcal{T})^* = \text{span}(\delta_t, t \in \mathcal{T})$ and

$$\forall \mu \in (\mathbb{R}^\mathcal{T}), \mu_X \circ \mu^{-1}$$

is a Gaussian measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

This property will be fundamental in the rest of this talk.
From now on:
- \((E, \| \cdot \|_E)\) will always be a real (separable) Banach space,
- \(\mathcal{B}(E)\) its Borel \(\sigma\)-algebra.

A probability measure \(\gamma\) on \((E, \mathcal{B}(E))\) is (centered) Gaussian if and only if \(\gamma \circ f^{-1}\) is a (centered) Gaussian measure on \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\) for all \(f \in E^*\).

We will always consider centered Gaussian measure. The characteristic functional is well defined and

\[
\forall f \in E^*, \; \hat{\gamma}(f) := \int_E e^{i \langle x, f \rangle_{E, E^*}} \gamma(dx) = e^{-\frac{C_{\gamma}(f, f)}{2}} \in \mathbb{R}
\]

with \(C_{\gamma}(f, g) = \int_E \langle x, f \rangle_{E, E^*} \langle x, g \rangle_{E, E^*} \gamma(dx)\) (covariance).
The covariance is bilinear, symmetric, positive and continuous.

**Covariance operator**

Let $\gamma$ be a Gaussian measure on $(E, \mathcal{B}(E))$ then

$$R_\gamma : f \in E^* \rightarrow \int_E \langle x, f \rangle_{E, E^*} x_{\gamma}(dx) \in E$$

is the covariance operator associated to $\gamma$.

It is characterized by the following relation

$$\forall (f, g) \in E^* \times E^*, \quad \langle R_\gamma f, g \rangle_{E, E^*} = C_\gamma(f, g).$$

Moreover, it is a positive symmetric kernel: $\forall (f, g) \in E^* \times E^*

$$
\begin{align*}
\langle R_\gamma f, g \rangle_{E, E^*} &= \langle R_\gamma g, f \rangle_{E, E^*}, \\
\langle R_\gamma f, f \rangle_{E, E^*} &\geq 0
\end{align*}$$

and $R_\gamma$ is a compact (nuclear) operator.
Similarly to Gaussian processes, define on $R_{\gamma}(E^*)$

$$\langle h, g \rangle_\gamma = \langle R_{\gamma} \hat{h}, \hat{g} \rangle_{E,E^*}$$

with $R_{\gamma} \hat{h} = h$ and $R_{\gamma} \hat{g} = g$. The Cameron-Martin space is

$$H(\gamma) := \{ R_{\gamma} \overset{\sim}{f}, \ f \in E^* \} \subset E$$

and we have the \textit{reproducing property}

$$\forall h \in H(\gamma), \ \forall f \in E^*, \ \langle h, f \rangle_{E,E^*} = \langle h, R_{\gamma} f \rangle_\gamma.$$

- $(H(\gamma), \langle ., . \rangle_\gamma)$ is a Hilbertian subspace of $E$,
- $R_{\gamma}(X^*)$ is dense in $H(\gamma)$,
- $R_{\gamma}$ may not be injective.
As we just saw, $R_\gamma : E^* \to E$ need not be injective.

However, $R_\gamma f = 0$ implies $f = 0$, $\gamma$-almost everywhere.

**Gaussian space**

Let $\gamma$ be a Gaussian measure on $(E, \mathcal{B}(E))$ and

$$i : f \in E^* \to f \in L^2(\gamma),$$

the Gaussian space is

$$E_{\gamma}^* = \overline{i(E^*)}^{L^2(\gamma)}.$$

For all Gaussian measures, Gaussian and Cameron-Martin spaces are isometric

$$E_{\gamma}^* \equiv H(\gamma).$$
Consider now:

- \((H, \langle ., . \rangle_H)\) a (separable) Hilbert space,
- \(B_{H^*}\) is weakly compact,
- \(H = H^*\) (Riesz representation theorem),
- \(\gamma\) Gaussian measure on \((H, B(H))\).

**Continuity**

\(h \in H \rightarrow \langle R_\gamma h, h \rangle_H\) is weakly sequentially continuous.

As consequence, consider \(h_0 \in B_H\) such that:

\[
\sup_{\|h\|_H \leq 1} \langle R_\gamma h, h \rangle_H = \langle R_\gamma h_0, h_0 \rangle_H = \lambda_0,
\]

then \(h_0\) is an eigenvector of \(R_\gamma\) associated to the eigenvalue \(\lambda_0\).
The previous $h_0$ may be used to split $H$ in two orthogonal subspaces

$$H = \mathbb{R}h_0 \oplus \perp H_1.$$  

From this, define:

- $P_0 : h \in H \rightarrow \langle h, h_0 \rangle_H h_0$,
- $\gamma_{\lambda_0} = \gamma \circ P_0^{-1}$,
- $\gamma_1 = \gamma \circ (I - P_0)^{-1}$.

**Orthogonal split of the Gaussian measure**

The two measures $\gamma_{\lambda_0}$ and $\gamma_1$ on $(H, B(H))$ are Gaussian and

$$\gamma = \gamma_{\lambda_0} * \gamma_1,$$

$$R_{\gamma_{\lambda_0}} = P_0 R_{\gamma} P_0 = \lambda_0 \langle h_0, \cdot \rangle_H h_0,$$

$$R_{\gamma_1} = (I - P_0) R_{\gamma} (I - P_0) = R_{\gamma} - R_{\gamma_{\lambda_0}}$$

with $\mathbb{R}h_0 = H(\gamma_{\lambda_0})$, $H_1 = (\mathbb{R}h_0)^\perp$ both equipped with $\langle \cdot, \cdot \rangle_{\gamma}$. 

Karhunen-loève decomposition on Hilbert spaces
Next, apply the splitting method to the (conditional) Gaussian measure $\gamma_1$ and so on and so forth...

**Properties**

- $(h_n)$ is orthonormal in $H$,
- $(h_n)$ is dense in $H(\gamma)$,
- $R_\gamma = \sum_{n \geq 0} \lambda_n \langle \cdot, h_n \rangle_H h_n$ with $\sum_{n \geq 0} \lambda_n < \infty$.

In particular, $R_\gamma$ is compact (nuclear) and

$$h = \sum_{n \geq 0} \langle h, h_n \rangle_H h_n, \ \gamma \ a.e.$$ 

The standard theory uses the compacity of $R_\gamma$ as an input and the spectral theorem. Moreover, the class of all Gaussian covariances is equal to symmetric, positive and nuclear operators.
The following result is already known (Theorem 3.5.1 p.112 in XX) and is called "Karhunen-loève" decomposition in some texts:

**Decomposition theorem**

Let $\gamma$ be a Gaussian measure on a locally convex space $E$, $H(\gamma)$ the Cameron-Martin space, $(e_n)$ an orthonormal basis, $(\xi_n)_n$ a sequence of independent standard Gaussian random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ then

$$\sum_{n\geq 0} \xi_n(\omega)e_n$$

converges almost surely in $E$, and has distribution $\gamma$.

We will now construct a "particular" basis in $R_\gamma(E^*)$ which corresponds to decomposition of the covariance operator.
Consider now the general case:

- $(E, \|\cdot\|_E)$ is a Banach space,
- $B_{E^*}$ is weakly compact (Banach Alaoglu theorem),
- $\gamma$ Gaussian measure on $(E, \mathcal{B}(E))$.

**Variance continuity**

$f \in E^* \rightarrow \langle R_\gamma f, f \rangle_{E,E^*}$ is weakly sequentially continuous.

$$\exists f_0 \in B_{E^*}, \quad \sup_{f \in B_{X^*}} \langle R_\gamma f, f \rangle_{E,E^*} = \langle R_\gamma f_0, f_0 \rangle_{E,E^*} = \lambda_0.$$  

Now, write $R_\gamma f_0 = \lambda_0 x_0 \in E$.

Spectral theory is not defined in this context.
Karhunen-loève decomposition on Banach spaces

\((f_0, x_0)\) will be useful to split "orthogonally" the Banach space

\[ E = \mathbb{R}x_0 + X_1, \]

and similarly, define

- \( P_0 : x \in E \rightarrow \langle x, f_0 \rangle_{E, E^*} x_0, \)
- \( \gamma_{\lambda_0} = \gamma \circ P_0^{-1}, \)
- \( \gamma_1 = \gamma \circ (I - P_0)^{-1}. \)

**Orthogonal split of the Gaussian measure**

The two measures \( \gamma_{\lambda_0} \) and \( \gamma_1 \) on \((E, \mathcal{B}(E))\) are Gaussian and

\[
\begin{align*}
\gamma &= \gamma_{\lambda_0} \ast \gamma_1, \\
R_{\gamma_{\lambda_0}} &= P_0 R_\gamma P_0^* = \lambda_0 \langle h_0, . \rangle_{H} h_0, \\
R_{\gamma_1} &= (I - P_0) R_\gamma (I - P_0)^* = R_\gamma - R_{\gamma_{\lambda_0}}
\end{align*}
\]

with \( \mathbb{R} h_0 = H(\gamma_{\lambda_0}), \ H_1 = (\mathbb{R} h_0)^\perp \) both equipped with \( \langle . , . \rangle_\gamma \).
Again, repeat the splitting method on $\gamma_1$.

### Properties

- $\|x_n\|_E = \|f_n\|_{E^*} = 1$,
- $(x_n)$ is dense in $H(\gamma)$,
- $R_\gamma = \sum_{n \geq 0} \lambda_n \langle \cdot, f_n \rangle_{E, E^*} x_n$.

Again, we have:

$$x = \sum_{n \geq 0} \langle x, f_n^* \rangle_{E, E^*} x_n, \gamma \text{ a.e.}$$

and

$$R_\gamma = \sum_{n \geq 0} \lambda_n \langle x_n, \cdot \rangle_{E, E^*} x_n.$$ 

The class of Gaussian covariance operators is still an open problem.
Consequences

So far, we have the following remarks:

**Approximation of covariance operator**

In the previous decomposition, we have:

\[
\left\| R_{\gamma} - \sum_{k=0}^{n} \lambda_k \langle x_k, \cdot \rangle_{E, E^*} x_k \right\| = \lambda_{n+1}
\]

where the norm is in \( \mathcal{L}(X^*, X) \).

**Maximum variance**

\[
\sup_{f \in B_{E^*}} \langle R_{\gamma n} f, f \rangle_{E, E^*} = \sup_{f \in B_{E^*}} \text{Var}(\langle P_n X, f \rangle_{E, E^*}) = \lambda_n
\]

- \( \sum_{n \geq 0} \lambda_n \) may not be finite,
- The decomposition may not be unique.
Specify a Gaussian measure in practice

Given a function space $E$, to specify a Gaussian measure on a function space:

- choose a covariance kernel $K$ such that:
  \[
  \mathbb{P}[(t \to X_t) \in E] = 1
  \]

- choose directly a Covariance operator (ex: inverse of elliptic differential operator)

Sample path from stochastic processes have properties linked to the Kernel:

- Continuity (Kolmogorov continuity theorem),
- Integrability,
- Differentiability…
Example on the Wiener measure

Now, consider the standard Brownian motion with kernel $K(s,t) = s \wedge t$.

- On $H = L^2([0,1], \mathbb{R})$, the decomposition is:

  $h_n : t \rightarrow \sqrt{2} \sin \left( \left( n + \frac{\pi}{2} \right) t \right)$,

  $\lambda_n = \frac{1}{(n + \frac{\pi}{2})^2}$

- On $E = C([0,1], \mathbb{R})$, it becomes:

  $h'_n : t \rightarrow \sqrt{2^p} \left( 1 \left\{ \frac{2k}{2^p+1}, \frac{2k+1}{2^p+1} \right\} - 1 \left\{ \frac{2k+1}{2^p+1}, \frac{2k+2}{2^p+1} \right\} \right)$,

  $x_n : t \rightarrow \sqrt{2^{p+2}} h'_n(t)$,

  $\lambda_n = \frac{1}{2^{p+2}}$, $n = 2^p + k$, $k = 0, \ldots, 2^p - 1$, $p \geq 0$

which represent "hat functions" on dyadic intervals.
Example with different kernels

Figure: Wiener measure basis functions.
Example with different kernels

**Figure:** Wiener measure basis variances.
Example with different kernels

Figure: Gaussian kernel basis functions.
Example with different kernels

Figure: Gaussian kernel basis variances.
Example with different kernels

Figure: Matern 3/2 kernel basis functions.
Example with different kernels

Figure: Matern 3/2 kernel basis variances.
Figure: Simulation of brownian motion using different bases.
Conclusion

A new method to decompose Gaussian measures on Banach spaces.

- The method can be applied to non-Gaussian square integrable measures,
- Gaussian covariance characterization is still an open problem,
- Links to optimal design in Kriging,
- Links with Bayesian inverse problems,
- Precise quantification of uncertainty.