Bayesian nonparametrics - An introduction

J. Rousseau

Oxford University & Dauphine

Applibugs

Outline

Outline

Introduction : notions générales en bayésien

- **Modèle** $X \sim P_\theta$, $\theta \in \Theta$, $X \in \mathcal{X}$
- **Proba a priori** : Proba sur Θ : $d\pi(\theta)$
- **Loi jointe**

$$d\lambda^\pi(X,\theta) = dP(X|\theta)d\pi(\theta): \quad \lambda^\pi(A\times B) = \int_B P(A|\theta)d\pi(\theta), \quad A \subset \mathcal{X}, B$$

- **Loi marginale de X**

$$dm_\pi(X) = \int dP(X|\theta)d\pi(\theta), \quad m_\pi(A) = \int P(A|\theta)d\pi(\theta), \quad A \subset \mathcal{X}$$

- **Loi a posteriori** version $d\pi(\theta|X)$ satisfaisant : pour tout $A \in \mathcal{X}$, $B \in \Theta$ mesurables

$$\int_B P(A|\theta)d\pi(\theta) = \int_A d\pi(B|X)dm_\pi(X)$$

Quelques exemples de problèmes non paramétriques

- **Distribution** $X \sim P$ avec $X \in \mathbb{R}$ et P inconnue. modèle non dominé

$$\Theta = \{P : \text{proba sur} \mathbb{R}\}$$

- **Densité** $X \sim P$ $P << \mu$ $f = dP/d\mu$ modèle dominé par μ

$$\Theta = \{f : \mathbb{R} \to \mathbb{R}^+, \int f d\mu = 1\}$$

- **Régression** $Y = f(X) + \epsilon$, f inconnue $X \in \mathcal{X}$ (ex $[0,1]^d$, $\mathbb{R}^d$)
- **Survie** données censurées $X \in \mathbb{R}^+$ $C \in \mathbb{R}^+$, $X \sim F$, $C \sim G$. On observe (Y, δ) avec $Y = X \wedge C$ $\delta = \mathbb{1}_{C \leqslant X}$
  Paramètre d'intéret : Fonction de Hasard cumulée

$$dH(x) = P(X \in (x, x + dx) | X > x)$$

Quelques exemples de problèmes non paramétriques

- **Distribution** $X \sim P$ avec $X \in \mathbb{R}$ et P inconnue. modèle non dominé

$$\Theta = \{P : \text{proba sur} \mathbb{R}\}$$

- **Densité** $X \sim P$ $P << \mu$ $f = dP/d\mu$ modèle dominé par μ

$$\Theta = \{f : \mathbb{R} \to \mathbb{R}^+, \int f d\mu = 1\}$$

- **Régression** $Y = f(X) + \epsilon$, f inconnue $X \in \mathcal{X}$ (ex $[0,1]^d$, $\mathbb{R}^d$)
- **Survie** données censurées $X \in \mathbb{R}^+$ $C \in \mathbb{R}^+$, $X \sim F$, $C \sim G$. On observe (Y, δ) avec $Y = X \wedge C$ $\delta = \mathbb{1}_{C \leqslant X}$
Paramètre d'intéret : Fonction de Hasard cumulée

$$dH(x) = P(X \in (x, x + dx) | X > x)$$

Quelques exemples de problèmes non paramétriques

- **Distribution** $X \sim P$ avec $X \in \mathbb{R}$ et P inconnue. modèle non dominé
$$\Theta = \{P : \text{proba sur} \mathbb{R}\}$$
- **Densité** $X \sim P$ $P << \mu$ $f = dP/d\mu$ modèle dominé par μ
$$\Theta = \{f : \mathbb{R} \to \mathbb{R}^+, \int f d\mu = 1\}$$
- **Régression** $Y = f(X) + \epsilon$, f inconnue $X \in \mathcal{X}$ (ex $[0,1]^d$, $\mathbb{R}^d$)
- **Survie** données censurées $X \in \mathbb{R}^+$ $C \in \mathbb{R}^+$, $X \sim F$, $C \sim G$. On observe (Y, δ) avec $Y = X \wedge C$ $\delta = \mathbb{1}_{C \leqslant X}$
Paramètre d'intéret : Fonction de Hasard cumulée
$$dH(x) = P(X \in (x, x + dx) | X > x)$$

Quelques exemples de problèmes non paramétriques

- **Distribution** $X \sim P$ avec $X \in \mathbb{R}$ et P inconnue. modèle non dominé
$$\Theta = \{P : \text{proba sur} \mathbb{R}\}$$
- **Densité** $X \sim P$ $P << \mu$ $f = dP/d\mu$ modèle dominé par μ
$$\Theta = \{f : \mathbb{R} \to \mathbb{R}^+, \int f d\mu = 1\}$$
- **Régression** $Y = f(X) + \epsilon$, f inconnue $X \in \mathcal{X}$ (ex $[0,1]^d$, $\mathbb{R}^d$)
- **Survie** données censurées $X \in \mathbb{R}^+$ $C \in \mathbb{R}^+$, $X \sim F$, $C \sim G$. On observe (Y, δ) avec $Y = X \wedge C$ $\delta = \mathbb{1}_{C \leqslant X}$
Paramètre d'intéret : Fonction de Hasard cumulée
$$dH(x) = P(X \in (x, x + dx) | X > x)$$

- Modèles à données dépendantes
  - **D**ensité spectrale dans les modèles stationnaires

$Y = (Y_1, ..., Y_n)$

$$Cov(Y_j, Y_{j+h}) = \gamma(h) = \int_{-\pi}^{\pi} f(\lambda) e^{i\lambda h} d\lambda$$

$f : [-\pi, \pi] \to \mathbb{R}^+$ paire.

  - **D**iffusions ergodiques

$$dX(t) = b(X(t)) + \sigma(X(t)) dB(t)$$

  - **Chaines de Markov**

$$\mathcal{L}(X_n | X_j, j \leqslant n-1) = \mathcal{L}(X_n | X_{n-1})$$

$K(x|y) =$ densité du noyau de transition. A estimer.

Types de problèmes

- **Construction de lois de probabilités sur** Θ Ensembles de dimension infinie....contraintes de formes ou autres....
- **Etude de certaines propriétés probabilistes de ces lois**
- **Semi paramétrique**
- **Algorithmes d'implémentation**
- **Propriétés fréquentistes (asymptotiques) des procédures Bayésiennes**

Familles de lois a priori

- A priori sur des probabilités : Dirichlet Process et extensions
- A priori sur des densités : Mélanges nonparamétriques, Arbre de Polyas, bases ou GP + transformations
- A priori sur des fonctions : GP , mélanges généralisés, utilisations de dictionnaires
- A priori sur des variétés ou autres objets : à partir de dictionnaires

Familles de lois a priori

- A priori sur des probabilités : Dirichlet Process et extensions
- A priori sur des densités : Mélanges nonparamétriques, Arbre de Polyas, bases ou GP + transformations
- A priori sur des fonctions : GP , mélanges généralisés, utilisations de dictionnaires
- A priori sur des variétés ou autres objets : à partir de dictionnaires

Familles de lois a priori

- A priori sur des probabilités : Dirichlet Process et extensions
- A priori sur des densités : Mélanges nonparamétriques, Arbre de Polyas, bases ou GP + transformations
- A priori sur des fonctions : GP , mélanges généralisés, utilisations de dictionnaires
- A priori sur des variétés ou autres objets : à partir de dictionnaires

Familles de lois a priori

- A priori sur des probabilités : Dirichlet Process et extensions
- A priori sur des densités : Mélanges nonparamétriques, Arbre de Polyas, bases ou GP + transformations
- A priori sur des fonctions : GP , mélanges généralisés, utilisations de dictionnaires
- A priori sur des variétés ou autres objets : à partir de dictionnaires

Outline

Espace $\mathcal{M}(\mathcal{X})$

$(\mathcal{X}, \mathcal{B})$: espace des observations .

$$\mathcal{M}(\mathcal{X}) = \{ \text{ Probabilités sur } \mathcal{X}\}.$$

- **Mesurabilité** Typiquement par rapport à la tribu borélienne associée à la convergence faible.
- **Construction de Probabilités sur** $\mathcal{X}$: Exemple type processus de Dirichlet .

Definition (Processus de Dirichlet)

Soit α une mesure de probabilité sur $\mathcal{X}$ et $M > 0$. On dit que $P \in \mathcal{M}(\mathcal{X})$ est distribué selon le processus de Dirichlet $DP(\alpha, M)$ si pour tout $k \in \mathbb{N}^*$ et toute partition $(B_j)_{j \leqslant k}$ de $\mathbb{R}$

$$(P(B_1), ..., P(B_k)) \sim \mathcal{D}(M\alpha(B_1), ..., M\alpha(B_k)).$$

Remarque : Pour toute mesure finie α et $M > 0$ il existe probabilité sur l'ensemble $\mathcal{M}(\mathcal{X})$ des probas sur $\mathcal{X}$

Propriétés du processus de Dirichlet et représentations

► **Représentation de Sethuraman** Si $P \sim DP(\alpha, M)$ alors presque surement

$$P(dx) = \sum_{j=1}^{\infty} p_j \delta_{(\theta_j)}, \quad \theta_j \overset{iid}{\sim} \alpha$$

et $(p_j)_j$: stick breaking

$$p_j = V_j \prod_{i<j} (1 - V_i), \quad V_i \overset{iid}{\sim} \text{Beta}(1, M)$$

• Si $P \sim DP(\alpha, M)$ alors P est discrête (p.s.)

$\rightarrow$ Généralisations possibles : Autres lois pour V_j ou pour (V_j, θ_j)

e.g. Pitman Yorr process

Marginalisation : Polya urn représentation et le processus du restaurant chinois

- **Model**

$$X_i|P \overset{iid}{\sim} P,\ i \leqslant n \quad P \sim DP(\alpha, M)$$

- **Loi marginal de** $(X_1, \cdots, X_n)$

$$X_1 \sim \bar{\alpha},$$
$$X_2|X_1 \sim \frac{M\alpha + \delta_{(X_1)}}{M+1}$$
$$X_n|X_1, \cdots, X_{n-1} \sim \frac{M\alpha + \sum_{i=1}^{n-1} \delta_{(X_i)}}{M+n-1}$$

- **Partition model on the observations** (EPPF formula) Si α est une loi sans atomes

$$\mathbb{P}((C_1, \cdots, C_k)) = \frac{M^k \Gamma(M) \prod_{j=1}^{k} \Gamma(n_j)}{\Gamma(M+n)}$$

Loi a posteriori et centrage

- **LE processus de Dirichlet est centrée sur α**

$$\forall A, \quad E[P(A)] = \alpha(A), \quad E\left(\int f(x)dP(x)\right) = \int f(x)\alpha(dx)$$

- **loi a posteriori**

Theorem

Si $X_i|P \overset{iid}{\sim} P$ et $P \sim DP(\alpha, M)$ alors la loi a posteriori de $P|(X_1, \cdots, X_n)$ est un $DP((M\alpha + n\mathbb{P}_n)/(M+n), M+n)$

$$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{(x_i)}$$

Le Dirichlet Process est une famille conjuguée

- **Nombre de Clusters**

$$K_n \sim M \log n, \quad \text{as} \quad n \rightarrow +\infty$$

Loi a posteriori et centrage

- **LE processus de Dirichlet est centrée sur α**

$$\forall A, \quad E[P(A)] = \alpha(A), \quad E\left(\int f(x)dP(x)\right) = \int f(x)\alpha(dx)$$

- **loi a posteriori**

Theorem

Si $X_i|P \overset{iid}{\sim} P$ et $P \sim DP(\alpha, M)$ alors la loi a posteriori de $P|(X_1, \cdots, X_n)$ est un $DP((M\alpha + n\mathbb{P}_n)/(M+n), M+n)$

$$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{(x_i)}$$

Le Dirichlet Process est une famille conjuguée

- **Nombre de Clusters**

$$K_n \sim M \log n, \quad \text{as} \quad n \to +\infty$$

Loi a posteriori et centrage

- **LE processus de Dirichlet est centrée sur α**

$$\forall A, \quad E[P(A)] = \alpha(A), \quad E\left(\int f(x)dP(x)\right) = \int f(x)\alpha(dx)$$

- **loi a posteriori**

Theorem

Si $X_i|P \overset{iid}{\sim} P$ et $P \sim DP(\alpha, M)$ alors la loi a posteriori de $P|(X_1, \cdots, X_n)$ est un $DP((M\alpha + n\mathbb{P}_n)/(M+n), M+n)$

$$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{(x_i)}$$

Le Dirichlet Process est une famille conjuguée

- **Nombre de Clusters**

$$K_n \sim M \log n, \quad \text{as} \quad n \to +\infty$$

Outline

Estimation de la Probabilité

- **Convergence**
- **Fonction de répartition**

$$F(x) = P(X \leqslant x), \quad \alpha((-\alpha, x]) = A(x), \quad F_n(x) = \mathbb{P}_n((-\infty, x])$$

Si $X_1, ..., X_n | P \overset{i.i.d}{\sim} P$ et si $P \sim \mathcal{D}(M, \alpha)$ alors $\forall x \in \mathbb{R}$

$$[F(x)|X_1, ..., X_n] \sim \text{beta}(MA(x)+nF_n(x), M(1-A(x))+n(1-F_n(x))$$

Donc

$$E^{\pi}[F(x)|X^n] = \frac{M}{M+n}A(x) + \frac{n}{M+n}F_n(x) = F_n(x) + O_P(n^{-1}),$$

$$V^{\pi}[F(x)|X^n] \leqslant \frac{F_n(x)(1-F_n(x))}{n} + O(n^{-2})$$

Clustering ou estimation de densité

- **Principe**
  But trouver à quelle population appartient chacun des X_i.
  Modèle standard : $[X_i|Z_i = j] \sim g_{\theta_j}$, Z_i : variable d'allocation. Si $j = 1, ..., k$ k connu : mélange clasique.
  si k inconnu $\Rightarrow$ Dirichlet Proc mixture
- **Dirichlet Mixture : représentation hierarchique**

$$\begin{array}{rcl} [X_i|\theta_i] & \sim & g_{\theta_i}, \quad \text{ex :}\mathcal{N}(\mu_j, \sigma_j^2) \\ \theta_i|P & \overset{i.i.d}{\sim} & P \\ P & \sim & DP(M, \alpha) \end{array}$$

- **Représentation Marginale**

$$X_i \overset{iid}{\sim} f_P(x) = \sum_{j=1}^{\infty} p_j g_{\theta_j}(x)$$

- **Estimation de densité**

DP mixtures : posterior distribution

Il existe bcp d'algos pour simuler dans la loi a posteriori $[P|X^n]$ en particulier si $P \sim DP(\alpha, M)$

- Algorithmes marginaux : On récupère la loi a posteriori de $[(\theta_1, \cdots, , \theta_n)|X^n]$

$$L_n(X_1, \cdots, X_n, d\theta_1, \cdots, d\theta_n) = \prod_{i=1}^{n} g_{\theta_i}(X_i) \underbrace{m_n(d\theta_1, \cdots, d\theta_n)}_{\text{Urne de Polya}}$$

On en déduit

$$\pi_i(d\theta_i|\theta_{-i}, X^n) = \sum_{j \neq i} q_{i,j}(X)\delta_{(\theta_j)}(d\theta_i) + q_{i,0}G_{b,i}(d\theta_i)$$

avec

$$q_{i,j}(X) \propto \begin{cases} g_{\theta_j}(X_i) & \text{si } j \neq i, j \geqslant 1 \\ \int_\Theta g_\theta(X_i)\alpha(\theta)d\theta & \text{si } j = 0 \end{cases}$$

$$G_{b,i}(\theta) \propto g_\theta(X_i)\alpha(\theta)$$

DP mixtures : posterior distribution II

Algorithmes conditionnels : basés sur la représentation Stick-Breaking

$$L_n(X_1, \cdots, X_n, P) = \prod_{i=1}^{n} \left(\sum_j p_j g_{\theta_j}(X_i) \right)$$

- Slice sampling : Walker , Kalli et al. Augmented approach

$$L_n(X_1, \cdots, X_n, U_1, \cdots, U_n, P) = \prod_{i=1}^{n} \left(\sum_j \mathbb{I}_{U_i \leqslant p_j} g_{\theta_j}(X_i) \right) \quad \Rightarrow$$

$$L_n(X_1, \cdots, X_n, \underbrace{U_1, \cdots, U_n}_{\text{augmented}}, \underbrace{s_1, \cdots, s_n}_{\text{allocations}}, P) = \prod_{i=1}^{n} \mathbb{I}_{U_i \leqslant p_{s_i}} g_{\theta_{s_i}}(X_i)$$

- Others : Neal, retrospective sampling (Roberts& Papaspiliopoulos)

DP mixtures : posterior distribution II

Algorithmes conditionnels : basés sur la représentation Stick-Breaking

$$L_n(X_1, \cdots, X_n, P) = \prod_{i=1}^{n} \left(\sum_j p_j g_{\theta_j}(X_i) \right)$$

- Slice sampling : Walker , Kalli et al. Augmented approach

$$L_n(X_1, \cdots, X_n, U_1, \cdots, U_n, P) = \prod_{i=1}^{n} \left(\sum_j \mathbb{I}_{U_i \leqslant p_j} g_{\theta_j}(X_i) \right) \quad \Rightarrow$$

$$L_n(X_1, \cdots, X_n, \underbrace{U_1, \cdots, U_n}_{\text{augmented}}, \underbrace{s_1, \cdots, s_n}_{\text{allocations}}, P) = \prod_{i=1}^{n} \mathbb{I}_{U_i \leqslant p_{s_i}} g_{\theta_{s_i}}(X_i)$$

- Others : Neal, retrospective sampling (Roberts& Papaspiliopoulos)

Variants and extensions

- **There are lots of variants to the DP**
  - Extensions around the stick breaking (e.g. Pitman Yor and co)
  - Extensions around the Polya Urn (e.g. Gibbs priors)
  - Extensions around Normalized Gamma process : e.g. Normalized completely random measures
  - DDP : dépendance
- **Variations around Feature allocation models** chaque indiv peut appartenir à plusieurs groupes : e.g. Indian Buffet process

Variants and extensions

- **There are lots of variants to the DP**
  - Extensions around the stick breaking (e.g. Pitman Yor and co)
  - Extensions around the Polya Urn (e.g. Gibbs priors)
  - Extensions around Normalized Gamma process : e.g. Normalized completely random measures
  - DDP : dépendance
- **Variations around Feature allocation models** chaque indiv peut appartenir à plusieurs groupes : e.g. Indian Buffet process

Variants and extensions

- **There are lots of variants to the DP**
  - Extensions around the stick breaking (e.g. Pitman Yor and co)
  - Extensions around the Polya Urn (e.g. Gibbs priors)
  - Extensions around Normalized Gamma process : e.g. Normalized completely random measures
  - DDP : dépendance
- **Variations around Feature allocation models** chaque indiv peut appartenir à plusieurs groupes : e.g. Indian Buffet process

Variants and extensions

- **There are lots of variants to the DP**
  - Extensions around the stick breaking (e.g. Pitman Yor and co)
  - Extensions around the Polya Urn (e.g. Gibbs priors)
  - Extensions around Normalized Gamma process : e.g. Normalized completely random measures
  - DDP : dépendance
- **Variations around Feature allocation models** chaque indiv peut appartenir à plusieurs groupes : e.g. Indian Buffet process

Gaussian Process

Une autre famille de loi a priori : Processus gaussiens

Definition (Processus Gaussien)

Soit $\mathcal{X}$ et $W = (W(x), x \in \mathcal{X})$ est un processus gaussien centré ssi il existe $K(.,.)$ tq pour tout $(x_1, \cdots, x_n)$

$$(W(x_1), \cdots, W(x_n)) \sim \mathcal{N}(0, (K(x_i, x_j), i, j \leqslant n))$$

et $K(.,.)$ est symétrique et semi définie positive.

- **Interprétation comme a priori** Modèle NP

$$y_i = f(x_i) + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2), \quad x_i \in [0, 1]$$

- Paramètre : (f, σ^2), $f : [0, 1] \to \mathbb{R}$
- Prior sur (f, σ^2) : $\sigma^2 \sim \pi_\sigma$ et

$$f \sim GP(\mu, K), \quad \text{e.g.} \quad \mu = 0$$

Gaussian Process

Une autre famille de loi a priori : Processus gaussiens

Definition (Processus Gaussien)

Soit $\mathcal{X}$ et $W = (W(x), x \in \mathcal{X})$ est un processus gaussien centré ssi il existe $K(.,.)$ tq pour tout $(x_1, \cdots, x_n)$

$$(W(x_1), \cdots, W(x_n)) \sim \mathcal{N}(0, (K(x_i, x_j), i, j \leqslant n))$$

et $K(.,.)$ est symétrique et semi définie positive.

- **Interprétation comme a priori** Modèle NP

$$y_i = f(x_i) + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2), \quad x_i \in [0, 1]$$

  - Paramètre : (f, σ^2), $f : [0, 1] \to \mathbb{R}$
  - Prior sur (f, σ^2) : $\sigma^2 \sim \pi_\sigma$ et

$$f \sim GP(\mu, K), \quad \text{e.g.} \quad \mu = 0$$

Gaussian Process

Une autre famille de loi a priori : Processus gaussiens

Definition (Processus Gaussien)

Soit $\mathcal{X}$ et $W = (W(x), x \in \mathcal{X})$ est un processus gaussien centré ssi il existe $K(.,.)$ tq pour tout $(x_1, \cdots, x_n)$

$$(W(x_1), \cdots, W(x_n)) \sim \mathcal{N}(0, (K(x_i, x_j), i, j \leqslant n))$$

et $K(.,.)$ est symétrique et semi définie positive.

- **Interprétation comme a priori** Modèle NP

$$y_i = f(x_i) + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2), \quad x_i \in [0, 1]$$

  - Paramètre : (f, σ^2), $f : [0, 1] \to \mathbb{R}$
  - Prior sur (f, σ^2) : $\sigma^2 \sim \pi_\sigma$ et

$$f \sim GP(\mu, K), \quad \text{e.g.} \quad \mu = 0$$

Choix du noyau ?

- K : Mvt Brownien $K(s,t) = s \wedge t$
- K : exponentiel $K(s,t) = e^{-\lambda(s-t)^2}$

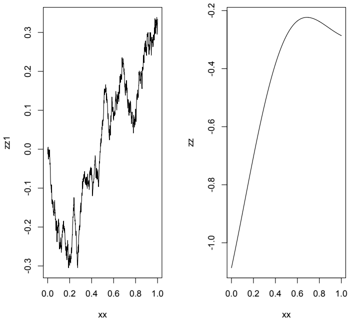

FIGURE: Gaussian processes : left : Brownian motion, right : exponential

Choix du noyau ?

- K : Mvt Brownien $K(s,t) = s \wedge t$
- K : exponentiel $K(s,t) = e^{-\lambda(s-t)^2}$

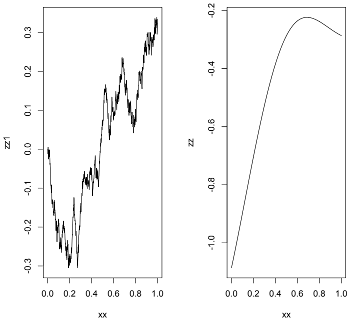

FIGURE: Gaussian processes : left : Brownian motion, right : exponential

Quelques exemples connus

- Brownien ou Rieman Liouville integrated B $\alpha > 0$

$$R^{\alpha}(t) = \int_0^t \frac{(t-s)^{\alpha-1/2}}{\Gamma(\alpha+1/2)} dB_s, \quad I_{0+}^k f(x) = \int_0^x \frac{(x-t)^{k-1}}{\Gamma(k)} f(t)dt$$

- Serie representation [Karhunen Loeve expansion] : Hilbert Space

$$f = \sum_{i=1}^{\infty} \theta_i \phi_i, \quad (\phi_i)_i = \text{BON } \mathbb{H} \quad \theta_i \overset{ind}{\sim} \mathcal{N}(0, \tau_i^2), \quad \tau_i \downarrow 0$$

- Krigging/interpolation process .
  $W = (W_x, x \in \mathcal{X}) = GP(0, K)$. Fixe $x_1, \cdots, x_N$

$$W^*(x) = E[W_x | W_{x_1}, \cdots, W_{x_n}] = \sum_{i=1}^{N} a_i(x) W_{x_i}, \quad a(x) = \Sigma^{-1}\sigma(x)$$

et

$$\Sigma = (K(x_i, x_j), i, j \leqslant N), \quad \sigma(x) = (K(x; x_i), i \leqslant N)^T$$

Quelques exemples connus

- Brownien ou Rieman Liouville integrated B $\alpha > 0$

$$R^{\alpha}(t) = \int_0^t \frac{(t-s)^{\alpha-1/2}}{\Gamma(\alpha+1/2)} dB_s, \quad I_{0+}^k f(x) = \int_0^x \frac{(x-t)^{k-1}}{\Gamma(k)} f(t)dt$$

- Serie representation [Karhunen Loeve expansion] : Hilbert Space

$$f = \sum_{i=1}^{\infty} \theta_i \phi_i, \quad (\phi_i)_i = \text{BON } \mathbb{H} \quad \theta_i \overset{ind}{\sim} \mathcal{N}(0, \tau_i^2), \quad \tau_i \downarrow 0$$

- Krigging/interpolation process .
$W = (W_x, x \in \mathcal{X}) = GP(0, K)$. Fixe $x_1, \cdots, x_N$

$$W^*(x) = E[W_x | W_{x_1}, \cdots, W_{x_n}] = \sum_{i=1}^{N} a_i(x) W_{x_i}, \quad a(x) = \Sigma^{-1}\sigma(x)$$

et

$$\Sigma = (K(x_i, x_j), i, j \leqslant N), \quad \sigma(x) = (K(x; x_i), i \leqslant N)^T$$

Quelques exemples connus

- Brownien ou Rieman Liouville integrated B $\alpha > 0$

$$R^\alpha(t) = \int_0^t \frac{(t-s)^{\alpha-1/2}}{\Gamma(\alpha+1/2)} dB_s, \quad I_{0+}^k f(x) = \int_0^x \frac{(x-t)^{k-1}}{\Gamma(k)} f(t)dt$$

- Serie representation [Karhunen Loeve expansion] : Hilbert Space

$$f = \sum_{i=1}^\infty \theta_i \phi_i, \quad (\phi_i)_i = \text{BON } \mathbb{H} \quad \theta_i \overset{ind}{\sim} \mathcal{N}(0, \tau_i^2), \quad \tau_i \downarrow 0$$

- Krigging/interpolation process .
$W = (W_x, x \in \mathcal{X}) = GP(0, K)$. Fixe $x_1, \cdots, x_N$

$$W^*(x) = E[W_x | W_{x_1}, \cdots, W_{x_n}] = \sum_{i=1}^N a_i(x) W_{x_i}, \quad a(x) = \Sigma^{-1}\sigma(x)$$

et

$$\Sigma = (K(x_i, x_j), i, j \leqslant N), \quad \sigma(x) = (K(x; x_i), i \leqslant N)^T$$

Comprendre W

- **Sample paths** $x \rightarrow W(x)$. Typiquement restreindre $W \in (\mathbb{B}, \|.\|)$ banach séparable. e.g. prendre la version continue du Brownien sur $[0,1]$.
- **Support** RKHS de W est relié

$$\mathbb{H} = \{h_H : \mathcal{X} \rightarrow \mathbb{R}; h_H(x) = E[W(x)H], \quad H \in \bar{\text{lin}}(W)$$

et

$$< h_{H_1}, h_{H_2} >_{\mathbb{H}} = E[H_1 H_2]$$

alors

$$h(x) =< h, K(x,.) >_{\mathbb{H}}, \quad h \in \mathbb{H}$$

Remarque : $(\mathbb{H}, < .,. >_{\mathbb{H}})$ *est un espace de Hilbert*
Support de $W \in (\mathbb{B}, \|.\|)$ est $\bar{\mathbb{H}}$ pour la norme $\|.\|$

Utilisations dans différents modèles

- Régression sur $[0,1]^d$

$$f \sim GP(K)$$

- Classification : transformation dans $[0,1]$: η

$$f(x) = \eta(W(x)), \quad x \in [0,1]^d$$

- Densité

$$f(x) = \frac{e^{W(x)}}{\int_0^1 e^{W(y)}dy}, \quad \text{ou} \quad f(x) = \frac{g(x)F(h(x)W(x))}{\int_{\mathbb{R}} g(x)F(h(x)W(x))dx}$$

- Gaussian processes on manifolds (e.g. Ghosal on the sphere, Castillo)
- Hierarchical Gaussian process to adapt over smoothness

$$W \sim GP(K), \quad f(x) = W(\tau x)$$

  - $\tau > 1$ détend le temps : augmentation de la régularité
  - $\tau < 1$ rétrécit le temps : dérégularise

Utilisations dans différents modèles

- Régression sur $[0,1]^d$

$$f \sim GP(K)$$

- Classification : transformation dans $[0,1]$: η

$$f(x) = \eta(W(x)), \quad x \in [0,1]^d$$

- Densité

$$f(x) = \frac{e^{W(x)}}{\int_0^1 e^{W(y)} dy}, \quad \text{ou} \quad f(x) = \frac{g(x)F(h(x)W(x))}{\int_{\mathbb{R}} g(x)F(h(x)W(x))dx}$$

- Gaussian processes on manifolds (e.g. Ghosal on the sphere, Castillo)
- Hierarchical Gaussian process to adapt over smoothness

$$W \sim GP(K), \quad f(x) = W(\tau x)$$

  - $\tau > 1$ détend le temps : augmentation de la régularité
  - $\tau < 1$ rétrécit le temps : dérégularise

Utilisations dans différents modèles

- Régression sur $[0,1]^d$

$$f \sim GP(K)$$

- Classification : transformation dans $[0,1]$: η

$$f(x) = \eta(W(x)), \quad x \in [0,1]^d$$

- Densité

$$f(x) = \frac{e^{W(x)}}{\int_0^1 e^{W(y)} dy}, \quad \text{ou} \quad f(x) = \frac{g(x)F(h(x)W(x))}{\int_{\mathbb{R}} g(x)F(h(x)W(x))dx}$$

- Gaussian processes on manifolds (e.g. Ghosal on the sphere, Castillo)
- Hierarchical Gaussian process to adapt over smoothness

$$W \sim GP(K), \quad f(x) = W(\tau x)$$

  - $\tau > 1$ détend le temps : augmentation de la régularité
  - $\tau < 1$ rétrécit le temps : dérégularise

Utilisations dans différents modèles

- Régression sur $[0,1]^d$

$$f \sim GP(K)$$

- Classification : transformation dans $[0,1]$: η

$$f(x) = \eta(W(x)), \quad x \in [0,1]^d$$

- Densité

$$f(x) = \frac{e^{W(x)}}{\int_0^1 e^{W(y)} dy}, \quad \text{ou} \quad f(x) = \frac{g(x)F(h(x)W(x))}{\int_{\mathbb{R}} g(x)F(h(x)W(x))dx}$$

- Gaussian processes on manifolds (e.g. Ghosal on the sphere, Castillo)
- Hierarchical Gaussian process to adapt over smoothness

$$W \sim GP(K), \quad f(x) = W(\tau x)$$

  - $\tau > 1$ détend le temps : augmentation de la régularité
  - $\tau < 1$ rétrécit le temps : dérégularise

Utilisations dans différents modèles

- Régression sur $[0,1]^d$

$$f \sim GP(K)$$

- Classification : transformation dans $[0,1]$: η

$$f(x) = \eta(W(x)), \quad x \in [0,1]^d$$

- Densité

$$f(x) = \frac{e^{W(x)}}{\int_0^1 e^{W(y)} dy}, \quad \text{ou} \quad f(x) = \frac{g(x)F(h(x)W(x))}{\int_{\mathbb{R}} g(x)F(h(x)W(x))dx}$$

- Gaussian processes on manifolds (e.g. Ghosal on the sphere, Castillo)
- Hierarchical Gaussian process to adapt over smoothness

$$W \sim GP(K), \quad f(x) = W(\tau x)$$

- $\tau > 1$ détend le temps : augmentation de la régularité
- $\tau < 1$ rétrécit le temps : dérégularise

Utilisations dans différents modèles

- Régression sur $[0,1]^d$

$$f \sim GP(K)$$

- Classification : transformation dans $[0,1]$: η

$$f(x) = \eta(W(x)), \quad x \in [0,1]^d$$

- Densité

$$f(x) = \frac{e^{W(x)}}{\int_0^1 e^{W(y)} dy}, \quad \text{ou} \quad f(x) = \frac{g(x)F(h(x)W(x))}{\int_{\mathbb{R}} g(x)F(h(x)W(x))dx}$$

- Gaussian processes on manifolds (e.g. Ghosal on the sphere, Castillo)
- Hierarchical Gaussian process to adapt over smoothness

$$W \sim GP(K), \quad f(x) = W(\tau x)$$

  - $\tau > 1$ détend le temps : augmentation de la régularité
  - $\tau < 1$ rétrécit le temps : dérégularise

Utilisations dans différents modèles

- Régression sur $[0,1]^d$

$$f \sim GP(K)$$

- Classification : transformation dans $[0,1]$: η

$$f(x) = \eta(W(x)), \quad x \in [0,1]^d$$

- Densité

$$f(x) = \frac{e^{W(x)}}{\int_0^1 e^{W(y)} dy}, \quad \text{ou} \quad f(x) = \frac{g(x) F(h(x) W(x))}{\int_{\mathbb{R}} g(x) F(h(x) W(x)) dx}$$

- Gaussian processes on manifolds (e.g. Ghosal on the sphere, Castillo)
- Hierarchical Gaussian process to adapt over smoothness

$$W \sim GP(K), \quad f(x) = W(\tau x)$$

  - $\tau > 1$ détend le temps : augmentation de la régularité
  - $\tau < 1$ rétrécit le temps : dérégularise