Metamodeling and data assimilation

I/ Gaussian process and uncertainty quantification

exemple 2 : Bayesian optimization

metamodelisation : on peut creer nos données

(On dispose d'un simulateur ou d'un algo de ML qu'on veut regler)

↓

metamodeling – computer experiments – surrogate modeling

exemple 1

we want to create a new vehicule

on veut faire des crashs tests

on a des simulateurs

simulator input → simulator → output

on n'a pas beaucoup de ressources pour faire tous les tests

⇒ on cree un metamodel

si on a 100 experiences

on fait 20 experiences avec le vrai modele physique

on ~~applique~~ construit les metamodel

on fait de nouvelles experiences

on enrichit le metamodel

processus aléatoires

famille de variables aléatoires $\{Z(x)\}_{x \in X}$

$(\Omega, F, \mathbb{P})$ espace probabiliste

Y: variable aléatoire réelle

$Y : \Omega \to \mathbb{R}$
$\quad w \to Y(w)$

$(Z(x))_{x \in X}$ processus aléatoire

$Z(x) : \Omega \to \mathbb{R}$
$\quad w \to Z(x)(w)$

trajectoire

w fixé

$X \to \mathbb{R}$
$x \to Z(x)\, w$

Exemple 1:

$X = \mathbb{R}$

$(Z(x))_{x \in X}$ bruit blanc

ie $\forall x \neq x'$ $Z(x) \perp\!\!\!\perp Z(x')$

$Z(x)$ a m̂ loi que $Z(x')$

exemple 2 : mouvement brownien en dim 1

$X = \mathbb{R}_+$

$\text{cov}(\cancel{Z(x)}, \cancel{Z(x')}) = \min(t, t')$

~~exemple 3~~

exemple 3

processus gaussien

noyau : squared exponential

$X = [0, 1]$

$k(x, x') = \text{cov}(Z(x), Z(x')) = e^{-\frac{(x-x')^2}{l^2}}$

second order RP

Y is a second order RP when all the random variables $Y(x) \in L^2(\mathbb{R})$

$\Rightarrow$ they all have variance

covariance function or kernel

$k(x, x') = \text{cov}(Y(x), Y(x'))$

stationnarité

1/ Y is strongly stationnary if for all locations $x_1, \dots, x_n \in X$

the law of $(Y(x_1 + h), \dots, Y(x_n + h))$ does not depend on h

2/ Y is weakly stationnary if for all locations $x_1, \dots, x_n \in X$ the first two moments of

$(Y(x_1 + h)$

processus gaussien

un processus $(Z(x))$ est gaussien ssi

$\forall (x_1, \dots, x_n)$ la loi du vecteur $(Z(x_1), \dots, Z(x_n))$

est gaussienne.

La loi d'un PG dépend de

$m : x \longrightarrow \mathbb{E}(Z(x))$ moyenne (fct quelconque)

$$k : \begin{array}{l} \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R} \\ (x, x') \longmapsto \operatorname{cov}(Z(x), Z(x')) \end{array}$$

k doit etre semi définie positive

$\forall x_1, \dots, x_n \in \mathcal{X}$
$\forall n \geq 1$
$\operatorname{Cov}\begin{pmatrix} Z(x_1) \\ \vdots \\ Z(x_n) \end{pmatrix}$ est semi définie positive

$$\operatorname{cov}\begin{pmatrix} Z(x_1) \\ \vdots \\ Z(x_n) \end{pmatrix} = \begin{pmatrix} & \vdots & \\ \cdots & \operatorname{cov}(Z(x_i), Z(x_j)) & \cdots \end{pmatrix}_{1 \leq i,j \leq n}$$

$$= \begin{pmatrix} k(x_i, x_j) \end{pmatrix}_{1 \leq i,j \leq n}$$

$= k(X, X)$

avec $X = \{x_1, \dots, x_n\}$

Donc $\forall X$
$\forall \alpha \in \mathbb{R}^n$

$\alpha^T k(X, X) \alpha \geq 0$

$\Leftrightarrow \sum_{1 \leq i,j \leq n} \alpha_i \alpha_j k(x_i, x_j) \geq 0$

RKHS et processus,

$Y_0 \sim GP(0, k_0)$

↑ par ex : Matern $\frac{5}{2}$

à partir de Y_0 : on peut avoir plusieurs noyaux

$Y \sim GP(0, k)$

dont les trajectoires sont impaires

$Y(x) = Y_0(x) - Y_0(-x)$

$L : f \to f(x) - f(-x)$ est linéaire

$Y(-x) = Y_0(-x) - Y_0(x) = -Y(x)$

Exercice 1

Montrer que Y est un GP

soit $(x_1, \dots, x_n) \in X$

$$\begin{pmatrix} Y(x_1) \\ \vdots \\ Y(x_n) \end{pmatrix} = \begin{pmatrix} Y_0(x_1) - Y_0(-x_1) \\ \vdots \\ Y_0(x_n) - Y_0(-x_n) \end{pmatrix} = \underbrace{\begin{pmatrix} Y_0(x_1) \\ \vdots \\ Y_0(x_n) \end{pmatrix}}_{\text{vecteur Gaussien}} - \underbrace{\begin{pmatrix} Y_0(-x_1) \\ \vdots \\ Y_0(-x_n) \end{pmatrix}}_{\text{vecteur gaussien}}$$

car $\begin{pmatrix} -x_1 \\ \vdots \\ -x_n \end{pmatrix} \in X$ et $Y_0 \sim GP$

~~Ainsi par comb linéaire :~~ $\cancel{\begin{pmatrix} Y(x_1) \\ \vdots \\ Y(x_n) \end{pmatrix}}$ ~~vecteur gaussien~~ → ce n'est pas vrai

~~$\forall (x_1, \dots, x_n) \in X$~~

$\Rightarrow Y \sim GP$ $\quad L : f \to f(x) - f(x_0)$ linéaire

$L(Y_0) = Y$

Y_0 est un GP

2/

$$\cancel{\mathbb{E}\left[\begin{pmatrix} Y(x_1) \\ \vdots \\ Y(x_n) \end{pmatrix}\right] = \mathbb{E}\left[\begin{pmatrix} Y_0(x_1) \\ \vdots \\ Y_0(x_n) \end{pmatrix} - \begin{pmatrix} Y_0(-x_1) \\ \vdots \\ Y_0(-x_n) \end{pmatrix}\right] = \mathbb{E}\left[\begin{pmatrix} Y_0(x_1) \\ \vdots \\ Y_0(x_n) \end{pmatrix}\right] - \mathbb{E}\left[\begin{pmatrix} Y_0(-x_1) \\ \vdots \\ Y_0(-x_n) \end{pmatrix}\right]}$$

car $\mathbb{E}[Y(x)] = \mathbb{E}[Y_0(x) - Y_0(-x)]$

$= \mathbb{E}[Y_0(x)] - \mathbb{E}[Y_0(-x)]$

$= 0 - 0$

$= 0$ car Y_0 PG $(0, k_0)$

$\Rightarrow Y$ est centré

$k(x, x') = cov(Y(x), Y(x'))$

$= cov(Y_0(x) - Y_0(-x), Y_0(x') - Y_0(-x'))$

$= cov(Y_0(x), Y_0(x')) - cov(Y_0(x), Y_0(-x'))$

$- cov(Y_0(-x), Y_0(x')) + cov(Y_0(-x), Y_0(-x'))$

$= k_0(x, x') - k_0(x, -x') - k_0(-x, x') + k_0(-x, -x')$

Correction

1/ considerons une comb linéaire

$\alpha_1 Y(x_1) + \dots + \alpha_n Y(x_n) = U$

mq U est de loi normale

$U = \alpha_1 \cdot Y_0(x_1) - \alpha_1 Y_0(-x_1) + \dots + \alpha_n Y_0(x_n) - \alpha_n Y_0(-x_n)$

Y_0 est un processus Gaussien

U est une comb linéaire extrait de Y_0 aux points $x_1, \dots, x_n, -x_1, \dots, -x_n$

$\Rightarrow U$ est de loi normale

3ème méthode

$$\begin{bmatrix} Y_0(x_1) \\ \vdots \\ Y_0(x_n) \\ Y_0(-x_1) \\ \vdots \\ Y_0(-x_n) \end{bmatrix}$$ est un vecteur Gaussien car Y est un GP

$$\underbrace{\begin{bmatrix} Y(x_1) \\ \vdots \\ Y(x_n) \end{bmatrix}}_{\to \text{ vecteur Gaussien}} = \begin{bmatrix} 1 & \cdots & 0 & 1 & \cdots & 0 \\ 0 & 1 \, 0 \cdots & 0 & 0 & 1 \, 0 \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} Y_0(x_1) \\ \vdots \\ Y_0(x_n) \\ Y_0(-x_1) \\ \vdots \\ Y_0(-x_n) \end{bmatrix}$$

Exercice 2 : Paving the way for GP regression with derivatives

$$Y \sim GP(0, k)$$

$$Z = \begin{pmatrix} Y(x) \\ Y(x_1) \\ Y'(x_1) \end{pmatrix}$$

GP Regression : $Y(x) \mid Y(x_1), \dots, Y(x_n)$

si on a les dérivées en $x_1, \dots, x_n$

On aimerait calculer

$$Y(x) \mid Y(x_1), \dots, Y(x_n), Y'(x_1), \dots, Y'(x_n)$$

un premier pas est de calculer la loi de :

$$\begin{pmatrix} Y(x) \\ Y(x_1) \\ \vdots \\ Y(x_n) \\ Y'(x_1) \\ \vdots \\ Y'(x_n) \end{pmatrix} \quad \text{et} \quad \text{avec } n=1 \quad \begin{pmatrix} Y(x) \\ Y(x_1) \\ Y'(x_1) \end{pmatrix}$$

$$Y \sim GP(0, k) \Rightarrow Y' \sim GP\left(0, \frac{\partial^2 k}{\partial s\, \partial t}\right)$$

$$\mathbb{E}[Z] = \begin{pmatrix} \mathbb{E}[Y(x)] \\ \mathbb{E}[Y(x_1)] \\ \mathbb{E}[Y'(x_1)] \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Mq Z est un processus Gaussien

L'application : $Y \longrightarrow \begin{pmatrix} Y(x) \\ Y(x_1) \\ Y'(x_1) \end{pmatrix}$ est linéaire

comme Y est GP alors $Z = \begin{pmatrix} Y(x) \\ Y(x_1) \\ Y'(x_1) \end{pmatrix}$

est un vecteur Gaussien

$$\mathrm{Cov}(Z) = \begin{pmatrix} \mathrm{Var}(Y(x)) & & \\ \mathrm{cov}(Y(x), Y(x_1)) & \mathrm{Var}(Y(x_1)) & \\ \mathrm{cov}(Y(x), Y'(x_1)) & \mathrm{cov}(Y(x_1), Y'(x_1)) & \mathrm{Var}(Y'(x_1)) \end{pmatrix}$$

$$\mathrm{Var}(Y(x)) = k_Y(x, x)$$

$$\underbrace{\mathrm{cov}(Y(x), Y'(x_1))}_{} = \left.\frac{\partial k_Y(s,t)}{\partial t}\right|_{(s,t)=(x,x_1)}$$

on a une dérivée à la 2ème composante $\Rightarrow$ il faut dériver par rapport à la 2ème

$$\mathrm{Var}(Y'(x_1)) = \left.\frac{\partial^2 k(s,t)}{\partial s\, \partial t}\right|_{(s,t)=(x_1,x_1)}$$

$$\Gamma_X = \begin{pmatrix} \mathrm{Var}(X_1), & \mathrm{cov}(X_1, X_2) & \mathrm{cov}(X_1, X_3) \\ \mathrm{cov}(X_2, X_1) & \mathrm{cov}(X_2, X_2) & \mathrm{cov}(X_2, X_3) \\ \mathrm{cov}(X_3, X_1) & \mathrm{cov}(X_3, X_2) & \mathrm{cov}(X_3, X_3) \end{pmatrix}$$

Exercice 3

$$\frac{\partial u}{\partial t}(x,t) - \alpha \cdot \frac{\partial^2 u}{\partial x^2}(x,t) = 0$$

$u(x,0) = \phi(x)$

$\downarrow$ transformée de Fourier

$$u(x,t) = \int_{\mathbb{R}} S(x-y,t)\,\phi(y)\,dy$$

$$S(x,t) = (4\pi\alpha t)^{-1/2} \exp\left(-\frac{x^2}{4\alpha t}\right)$$

on peut considérer $(\phi(x)) \sim GP(0, k_0)$

Mq u est un processus gaussien centré

+ calculer son noyau

$L\phi : \phi \to \int S(x-y,t)\,\phi(y)\,dy$

L est linéaire et $u = L(\phi)$

$\Rightarrow$ u est un processus gaussien

$\mathbb{E}(u(x,t)) = \mathbb{E}\left[\int_{\mathbb{R}} S(x-y,t)\cdot\phi(y)\cdot dy\right]$

$= \int_{\mathbb{R}} S(x-y,t) \cdot \underbrace{\mathbb{E}[\phi(y)]}_{=0} \cdot dy$

Fubini Tonelli

$= 0$

$\Rightarrow$ u est centré

$k_u(x,x') = \mathrm{cov}(u(x,t), u(x',t'))$

$= \mathrm{cov}\left(\int S(x-y,t)\cdot\phi(y)\cdot dy, \int S(x'-y,t')\,\phi(y)\,dy\right)$

$= \iint_{\mathbb{R}^2} S(x-y,t) \cdot S(x'-y',t')\,\mathrm{cov}(\phi(y),\phi(y'))\cdot dy\,dy'$

$= \iint_{\mathbb{R}^2} S(x-y,t)\; S(x'-y',t')\; k_0(y,y')\; dy\cdot dy'$

Exercice 4

RKHS et processus

si $k(x, x') = \text{cov}(Y(x), Y(x'))$
alors k est psd

def : noyau : $(k(x_i, x_j))$ est a psd (positive semi definite)

soient $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}$; $x_1, \dots, x_n \in \mathcal{X}$

$$\sum_{i,j} \alpha_i \alpha_j k(x_i, x_j) = \sum_{i,j} \alpha_i \alpha_j \,\text{cov}(Y(x_i), Y(x_j))$$

$$= \text{Cov}\Big(\underbrace{\sum_i \alpha_i Y(x_i)}_{U}, \sum_j \alpha_j Y(x_j)\Big)$$

bilinearité de la covariance

$$= \text{Cov}(U, U)$$

$$= \text{Var}\Big(\sum_i \alpha_i Y(x_i)\Big) \geq 0$$

Remarque

Si Y est centré : $k(x, x') = \mathbb{E}[Y(x) Y(x')]$

$$= \langle Y(x), Y(x') \rangle_{L^2(\mathbb{P})}$$

$$= \langle \phi(x), \phi(x') \rangle_H$$

avec $\phi : x \mapsto Y(x)$ et $H : L^2(\mathbb{P})$: espace de Hilbert

propriétés

k noyau $\Rightarrow c^2 k$ noyau $\forall c \in \mathbb{R}$

$$k_1 \oplus k_2 \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} \right) = k_1(x_1, x_1') + k_2(x_2, x_2')$$

$$k_1 \otimes k_2 \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} \right) = k_1(x_1, x_1') \times k_2(x_2, x_2')$$

f

$f : \mathcal{X} \to U$

$$k_f(x, x') = k(f(x), f(x'))$$

Exercice 5 : transformations et processus gaussiens

Y GP

$k \leftrightarrow Y \quad k(x, x') = \text{cov}(Y(x), Y(x'))$

$c^2 k \longleftrightarrow Z = cY$

$$c^2 k(x, x') = \text{cov}(cY(x), cY(x'))$$

Z est un GP
car $g \mapsto cg$ est linéaire
et Y est un GP

4/ $f : U \to \mathcal{X}$ $\quad k$ noyau sur $\mathcal{X}$

$$k_f(x, x') = k(f(x), f(x'))$$

$$k(f(x), f(x')) = \text{cov}(Y(f(x)), Y(f(x')))$$

$$Z(x) = Y(f(x))$$

$$\text{cov}(Z(x), Z(x')) = \text{cov}(Y(f(x)), Y(f(x')))$$

$$= k(u, u')$$

avec $u = f(x) \quad u' = f(x')$

$$= k(f(x), f(x'))$$

$$= k_f(x, x')$$

$Y \to x \mapsto Y(f(x))$ est linéaire par rapport à Y

$\Rightarrow Z$ est un GP

autre preuve

Soit une CP extraite de Z en $x_1, \dots, x_n$

$\alpha_1 Z(x_1) + \dots + \alpha_n Z(x_n)$

$= \alpha_1 Y(f(x_1)) + \dots + \alpha_n Y(f(x_n))$

c'est une CP extraite de Y aux points

$u_1, \dots, u_n$

or Y est un GP

$\Rightarrow$ Z est un GP

$$R_1 + R_2 \xleftrightarrow{Y_1 \perp\!\!\!\perp Y_2} Y_1 + Y_2 = Z$$

Z est un GP

$$R_1 R_2 \quad \xrightarrow{Y_1 \perp\!\!\!\perp Y_2} \quad Z = Y_1 Y_2$$

Indication 1/ vérifier que Z est centré

2/ calculer $R_Z(x,x') = \mathbb{E}[Z(x) \times Z(x')]$

$\mathbb{E}[Z] = \mathbb{E}[Y_1 Y_2] = \mathbb{E}[Y_1]\mathbb{E}[Y_2]$
$= 0$

$R_Z(x,x') = \operatorname{cov}(Z(x), Z(x'))$
$= \mathbb{E}(Z(x) \times Z(x'))$
$= \mathbb{E}[Y_1(x) Y_1(x') Y_2(x), Y_2(x')]$
$= \mathbb{E}[Y_1(x) Y_1(x')] . \mathbb{E}[Y_2(x) Y_2(x')]$
$= R_1(x,x') . R_2(x,x')$

⚠ Z n'est pas un GP

si le PG est stationnaire en 18 alors

$$k(x,x') = c(x-x') = c(k)$$

Bochner's Theorem :

The kernel of a real valued stationnary GP on $\mathbb{R}^d$ is the fourrier transform of a probability distribution

$$k(x,x') = \int_{\mathbb{R}^d} \cos(2\pi \langle x-x', r\rangle) dp(r)$$

Exercice 6 : proof of the direct sense

1/ mq $(x,x') \longmapsto \cos(2\pi\langle x-x', r\rangle)$ est semi définie positive

soit $(\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d$

$$\sum_{k=1}^{d}\sum_{j=1}^{d} \alpha_k \alpha_j\, k(x_k, x_j)$$

$$= \sum_{1\leq k,j\leq d} \alpha_k \alpha_j \cos(2\pi\langle x_k - x_j', r\rangle)$$

or $\cos(2\pi\langle x_k - x_j', r\rangle) = Re\left(e^{i2\pi\langle x_k - x_j', r\rangle}\right)$

$$= \sum_{k,j} \alpha_k \alpha_j Re\left(e^{i2\pi\langle x_k, r\rangle} \cdot e^{-i2\pi\langle x_j, r\rangle}\right)$$

$$= Re\Big(\underbrace{\sum_k \alpha_k e^{i2\pi\langle x_k, r\rangle}}_{a} \cdot \underbrace{\sum_j \alpha_j e^{-i2\pi\langle x_j, r\rangle}}_{=\bar{a}}\Big)$$

$$= Re(a \times \bar{a}) = Re(|a|^2) = |a|^2 \geq 0$$

2/ En déduire que $k : (x,x') \longmapsto \int \cos(2\pi\langle x-x', r\rangle) \underbrace{dp(r)}_{\text{mesure spectrale}}$ est un noyau

Soient $\alpha_1, \dots, \alpha_d \in \mathbb{R}^d$

$$\sum_{1\leq i,j\leq d} \alpha_i \alpha_j \int \cos(2\pi\langle x_i - x_j, r\rangle) \cdot dp(r)$$

$$= \int \underbrace{\sum_{1\leq i,j\leq d} \alpha_i \alpha_j \cos(2\pi\langle x_i - x_j, r\rangle)}_{\geq 0} \cdot dp(r)$$

Exercice 7 : the square exp kernel on a Hilbert space

$(H, \langle\cdot,\cdot\rangle)$ Hilbert $\quad k(x,x') = e^{-\|x-x'\|^2}$

1/ $k(x,x') = e^{-\langle x-x', x-x'\rangle}$

$$= e^{-\|x\|^2 - \|x'\|^2 + 2\langle x,x'\rangle}$$

$$= e^{-\|x\|^2} \cdot e^{-\|x'\|^2} \cdot e^{2\langle x,x'\rangle}$$

$$= g(x)\, g(x')\, e^{2\langle x,x'\rangle}$$

2/ $k_2(x,x') = g(x)\,g(x')$

Mq k_2 est un noyau

$$\sum_{i,j} \alpha_i \alpha_j k_2(x,x') = \sum_{i,j} \alpha_i \alpha_j g(x_i) \cdot g(x_j)$$

$$= \sum_i \alpha_i g(x_i) \cdot \sum_j \alpha_j g(x_j) \geq 0$$

$\Rightarrow k_2$ noyau

3/ si k_0 un noyau mq e^{k_0} noyau

$$\sum_{i,j} \alpha_i \alpha_j e^{k_0(x_i, x_j)}$$

$$\sum_{i,j} \alpha_i \alpha_j \sum_{n=0}^{+\infty} \frac{(k_0(x_i,x_j))^n}{n!}$$

$$\sum_{n=0}^{+\infty} \sum_{i,j} \frac{\alpha_i \alpha_j (k_0(x_i,x_j))^n}{n!} \longrightarrow \geq 0$$

k_0 noyau $\Rightarrow k_0^n(x_i, x_j)$

limit of kernels

Rappels vecteurs gaussiens

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Gamma\right)$$

$$\Gamma = \left(\operatorname{cov}(X_i, X_j)\right)_{1 \leq i,j \leq 2}$$

$X_2 \mid X_1 = x_1$ suit une loi normale

$x_1 \to \mathbb{E}[X_2 \mid X_1 = x_1]$ → est linéaire

$\operatorname{Var}(X_2 \mid X_1 = x_1)$ → constant

Démonstration dans le cas X_1 et X_2 centrés

X_1 et X_2 var

$\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$

calculons $\mathbb{E}_L(X_2 \mid X_1) = \alpha + \beta x_1$

↓

meilleure approximation de X_2 comme fonction affine de X_1 au sens L_2

Le produit scalaire dans L^2 est

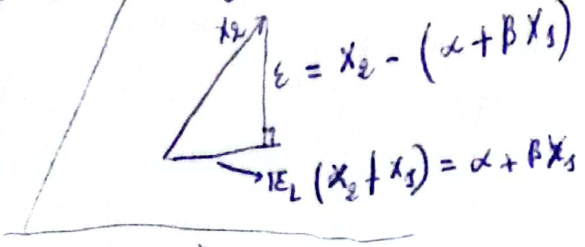
$$\langle U, V \rangle = \mathbb{E}[U \times V]$$

Mq $\mathbb{E}_L(X_2 \mid X_1) = \dfrac{\operatorname{cov}(X_1, X_2)}{\operatorname{Var}(X_2)} \cdot X_1$

$\operatorname{cov}(U,V) = \mathbb{E}\left[(U - \mathbb{E}(U))(V - \mathbb{E}(V))\right]$

$\operatorname{Var}(U) = \operatorname{cov}(U, U)$

$\operatorname{cov}(U, V) = \mathbb{E}[X_1 X_2]$



$\operatorname{Vect}(1, X_1)$

↓

la variable aléatoire constante

$$\varepsilon \perp \operatorname{Vect}(1, X_1) \Leftrightarrow \begin{cases} \varepsilon \perp 1 \\ \varepsilon \perp X_1 \end{cases}$$

$\Leftrightarrow \langle \varepsilon, X_1 \rangle = 0$ et $\langle \varepsilon, 1 \rangle = 0$

$\varepsilon = X_2 - (\alpha + \beta X_1)$

$\varepsilon \perp 1$

$\langle \varepsilon, 1 \rangle = 0 \Leftrightarrow \mathbb{E}\left[(X_2 - (\alpha + \beta X_1)) \cdot 1\right] = 0$

$\Leftrightarrow \mathbb{E}[X_2] - \alpha - \mathbb{E}[X_1] = 0$

$\Leftrightarrow \alpha = 0$

$\varepsilon \perp X_1$

$\langle \varepsilon, X_1 \rangle \Leftrightarrow \mathbb{E}\left[(X_2 - \beta X_1) \cdot X_1\right] = 0$

$\Leftrightarrow \mathbb{E}[X_1 X_2] - \beta\, \mathbb{E}[X_1^2] = 0$

$$\beta = \frac{\mathbb{E}[X_1 X_2]}{\mathbb{E}[X_1^2]} = \frac{\operatorname{cov}(X_1, X_2)}{\operatorname{Var}(X_1)}$$

En fait on a ici que

$$\mathbb{E}_L(X_2 \mid X_1) = \underbrace{\mathbb{E}(X_2 \mid X_1)}$$

projec $\perp$ de X_2 sur toutes les fonctions de X_1 de carré intégrable

i/ $\varepsilon \perp X_1$

ii/ $\begin{pmatrix} \varepsilon \\ X_1 \end{pmatrix}$ est un vecteur gaussien

$$\begin{pmatrix} \varepsilon \\ X_1 \end{pmatrix} = \begin{pmatrix} X_2 - \beta X_1 \\ X_1 \end{pmatrix} = \begin{pmatrix} -\beta & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

de i et ii

$$\operatorname{Cov}(\varepsilon, X_1) = \begin{pmatrix} \operatorname{Var}(\varepsilon) & 0 \\ 0 & \operatorname{Var}(X_1) \end{pmatrix}$$

↓

orthogonalité

$\Rightarrow \varepsilon$ et X_1 sont indépendants ?

$$f_{\binom{\varepsilon}{X_1}}\binom{\varepsilon}{x_1} = \frac{1}{2\pi |\Gamma|} e^{-\frac{1}{2}(\varepsilon \;\; x_1)\Gamma^{-1}\binom{\varepsilon}{x_1}}$$

$$= e^{-\frac{1}{2}\frac{\varepsilon^2}{\operatorname{Var}(\varepsilon)}} \times e^{-\frac{1}{2}\frac{x_1^2}{\operatorname{Var}(x_1)}}$$

~~comme~~

ainsi $f_{\binom{\varepsilon}{X_1}}\binom{\varepsilon}{x_1} = f_\varepsilon(\varepsilon) \cdot f_{X_1}(x_1)$

$\Rightarrow \varepsilon \perp\!\!\!\perp X_1$

$\Rightarrow \varepsilon \perp\!\!\!\perp g(X_1)$ pour toute fonction g

$\Rightarrow \mathbb{E}_L(X_2|X_1) = $ proj $\perp$ de X_2 sur $L^2(X_1)$

$\Rightarrow \mathbb{E}_L(X_2|X_1) = \mathbb{E}[X_2|X_1]$

b/ Mq que $X_2|X_1$ est une loi normale

$X_2 = \varepsilon + \beta X_1$

sachant que $X_1 = x_1$ et $\varepsilon \perp\!\!\!\perp X_1$

$X_2 \overset{\text{loi}}{=} \varepsilon + \beta x_1$

de loi ~~linéaire~~ normale car $\varepsilon = X_2 - (\alpha + \beta X_1)$

comb linéaire de X_1 et X_2

$\Rightarrow X_2 | X_1 = x_1$ est de loi normale

comme $\varepsilon \sim N(\mu_\varepsilon, \sigma_\varepsilon^2)$

$X_2 | X_1 = x_1 \sim N(\mu_\varepsilon + \beta x_1, \sigma_\varepsilon^2)$

$\mu_\varepsilon = \mathbb{E}[\varepsilon] = \mathbb{E}[X_2 - \beta X_1 - \alpha]$

$= \underbrace{\mathbb{E}[X_2]}_{=0} - \beta \underbrace{\mathbb{E}[X_1]}_{=0} - \underbrace{\alpha}_{=0}$

$= 0$

$\mathbb{E}[X_2 | X_1 = x_1] = \beta x_1 = \dfrac{\operatorname{cov}(X_1, X_2)}{\operatorname{Var}(X_1)}$

$\sigma_\varepsilon^2 = \operatorname{Var}(\varepsilon) = \mathbb{E}[\varepsilon^2]$

$\mathbb{E}[\varepsilon] = 0$

$= \|\varepsilon\|^2$

$= \|X_2\|^2 - \|\mathbb{E}_L(X_2|X_1)\|^2$

pythagore

$= \operatorname{Var}(X_2) - \beta^2 \operatorname{Var}(X_1)$

$= \operatorname{Var}(X_2) - \left(\dfrac{\operatorname{cov}(X_1, X_2)}{\operatorname{Var}(X_1)}\right)^2 \operatorname{Var}(X_1)$

$= \operatorname{Var}(X_2) - \dfrac{\operatorname{cov}(X_1, X_2)^2}{\operatorname{Var}(X_1)}$

$$\operatorname{Var}(X_2 | X_1 = x_1) = \operatorname{Var}(X_2) - \frac{\operatorname{cov}(X_1, X_2)^2}{\operatorname{Var}(X_1)}$$

En général

$$\begin{pmatrix} V \\ W \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_V \\ \mu_W \end{pmatrix}, \begin{pmatrix} \Gamma_V & \Gamma_{V,W} \\ \Gamma_{W,V} & \Gamma_W \end{pmatrix}\right)$$

$\Gamma_{W,V} = \Gamma_{V,W}^T$

$V | W = w$ est un vecteur gaussien

$\mathbb{E}(V | W = w) = \mu_V + \Gamma_{V,W} \cdot \Gamma_W^{-1}(w - \mu_W)$

$\operatorname{cov}(V | W = w) = \Gamma_V - \Gamma_{V,W} \cdot \Gamma_W^{-1} \cdot \Gamma_{W,V}$

en particulier

$V = X_2 \sim N(0, \operatorname{Var}(X_2))$

$W = X_1 \sim N(0, \operatorname{Var}(X_1))$

$$\begin{pmatrix} V \\ W \end{pmatrix} = \begin{pmatrix} X_2 \\ X_1 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \operatorname{Var}(X_2) & \operatorname{cov}(X_2, X_1) \\ \operatorname{cov}(X_1, X_2) & \operatorname{Var}(X_1) \end{pmatrix}\right)$$

$$\frac{\ell_{\mu e}}{xe} \times \frac{xe}{\check{h}e} \times \frac{\check{h}e}{\ell e} =$$

$$\frac{\ell_{\mu e}}{\check{h}e} \times \frac{\check{h}e}{\ell e} = \frac{\ell_{\mu e}}{\ell e}$$

Exercice 11

But : montrer que si $Y \sim GP(m, k)$ (m : moyenne, k : noyau)

alors $Y \mid Y(x_1) = y_1, \dots Y(x_n) = y_n$

est un GP (m_c, k_c) avec

$$m_c(x) = m(x) + \underbrace{k(x,X)}_{1\times n} \underbrace{k(X,X)^{-1}}_{n\times n} \underbrace{(y - m(X))}_{n\times 1}$$

$k(x, X) = (k(x, x_1), \dots, k(x, x_n))$

vecteur $1 \times n$

1/ $V = Y(x)$ c'est un point

$W = \begin{pmatrix} Y(x_1) \\ \vdots \\ Y(x_n) \end{pmatrix}$

Mq $U = \begin{pmatrix} V \\ W \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_V \\ \mu_W \end{pmatrix}, \begin{pmatrix} \Gamma_V & \Gamma_{V,W} \\ \Gamma_{V,W} & \Gamma_W \end{pmatrix}\right)$?

W est un vecteur gaussien car Y est un PG

V est aussi un vecteur gaussien

$V \sim N(m(x), \sigma_x) \sim N(m(x), k(x,x))$

$W \sim N\left(\begin{pmatrix} m(x_1) \\ \vdots \\ m(x_n) \end{pmatrix}, \begin{pmatrix} Var(X_1) & cov(X_1, X_2) & \\ & \ddots & \\ cov(X_n, X_1) & & Var(X_n) \end{pmatrix}\right)$

$\sim N\left(\begin{pmatrix} m(x_1) \\ \vdots \\ m(x_n) \end{pmatrix}, k(X,X)\right)$

$U \sim N\left(\begin{pmatrix} m(x) \\ m(x_1) \\ \vdots \\ m(x_n) \end{pmatrix}, \begin{pmatrix} k(x,x) & k(x,X) \\ k(X,x) & k(X,X) \end{pmatrix}\right)$

Mq V est un vecteur gaussien ? On ne peut pas dire que si V et Y Gaussiens alors $\binom{U}{V}$ est gaussien (ceci est vrai si V et V sont indépendants)

car $\Rightarrow (x, x_1, \dots, x_n)$ est un ensemble de $(n+1)$ points + Y est un GP $\Rightarrow U = \begin{pmatrix} Y(x) \\ \vdots \\ Y(x_n) \end{pmatrix}$ est gaussien

$V \mid W = w$ est un vecteur gaussien

$\mathbb{E}[V \mid W = w] = m(x) + k(x,X) \cdot k(X,X)^{-1} \cdot (y - m(X))$

$Var(V \mid W = w) = k(x,x) - k(x,X)\, k(X,X)^{-1} k(X,x)$

2/ $m_c(x_i) = \mathbb{E}[Y(x_i) \mid Y(x_1) = y_1, \dots Y(x_n) = y_n]$

$= \mathbb{E}[y_i \mid \dots\dots$ (constante)

$= y_i$

$k_c(x_i, x) = cov(Y(x_i), Y(x) \mid Y(x_1) = y_1, \dots, Y(x_n) = y_n)$

$= cov(y_i, Y(x) \mid \dots\dots)$ (déterministe)

$= 0$

<u>ième méthode</u>

$$m_c(x_i) = m(x_i) + \underbrace{k(x_i, X)\, k(X,X)^{-1}}_{v} \begin{pmatrix} y_1 - m(x_1) \\ \vdots \\ y_n - m(x_n) \end{pmatrix}$$

$$k(x_i, X)\, k(X,X)^{-1} = v$$

$$k(x_i, X) = v \times k(X,X)$$

$$= v \times \begin{pmatrix} k(x_1, X) \\ \vdots \\ k(x_1, X) \end{pmatrix}$$

$$v = (0, 0, \dots, 1, 0, \dots 0) = e$$

ième cordonnée

$$m_c(x_i) = m(x_i) + y_i - m(x_i)$$

$$= y_i$$

$$k_c(x_i, x) = k(x_i, x) - k(x_i, X) . k(X,X)^{-1} . k(X, x)$$

$$k(x_i, X) . k(X,X)^{-1} = v$$

$$k(x_i, X) = v . k(X,X)$$

$$v = (0, \dots, 1, 0, \dots, 0)$$

ième position

$$k_c(x_i, x) = 0$$

3/ Intervalle de prévision pour une va

$$Z \sim \mathcal{N}(\mu_Z, \sigma_Z^2)$$

$$\mu_Z \pm 2\sigma_Z \quad \text{à } 95\%$$

$$\mathbb{P}\left(Z \in [\mu_Z - 2\sigma_Z, \mu_Z + 2\sigma_Z]\right) \simeq 95\%$$

$$q_{\mathcal{N}(0,1)}^{(2,5\%)} \simeq q_{\mathcal{N}(0,1)}^{(97,5\%)}$$

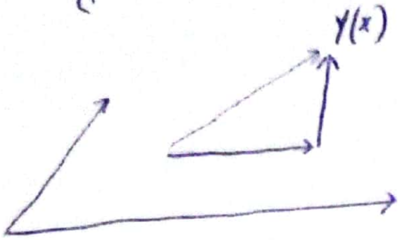
ici appliquée à la loi de $Y(x) | Y(x_i) = y_i$ i de 1 à n

$$m_c(x) + 2\sqrt{k_c(x,x)}$$

$\hat{Y}$ unbiased

$$\hat{Y}(x) = w_0(x) + w_1(x)\, Y(x_1) + \dots + w_n(x)\, Y(x_n)$$

min $\mathbb{E}\left[\left(Y(x) - \hat{Y}(x)\right)^2\right]$ under $\mathbb{E}\left(\hat{Y}(x)\right) = \mathbb{E}\left[Y(x)\right]$



Vect$(1, Y(x_1), \dots, Y(x_n))$

$$\hat{Y}(x) = \mathbb{E}_L\left(Y(x) \mid Y(x_1), \dots, Y(x_n)\right)$$

si Y est un PG alors $\mathbb{E}_L\left(Y(x) \mid Y(x_1), \dots, Y(x_n)\right)$

$$= \mathbb{E}\left(Y(x) \mid Y(x_1), \dots, Y(x_n)\right)$$

$$Y(x) = m(x) + Z(x)$$

$$m(x) = \beta_1 f_1(x) + \dots + \beta_p (x)$$

Z is a centered GP with kernel $k(x, y; \theta)$

Observations bruitées et filtrage

$$y_i = Y(x_i) + \varepsilon_i \quad i = 1 \text{ à } n$$

Y : $GP(m, R)$

ε_i : bruit, $\varepsilon_1, \dots, \varepsilon_n$ indep, $\varepsilon_i \sim \mathcal{N}(0, \tau_i^2)$

On suppose de plus que $\varepsilon_1, \dots, \varepsilon_n$ indep de Y

On veut aussi calculer la loi de

$$\underbrace{Y(x)}_{\text{signal original non bruité}} \mid \underbrace{Y(x_i) + \varepsilon_i = y_i}_{\text{observations non bruitées}} \quad i = 1 \text{ à } n$$

$$U = \begin{bmatrix} Y(x) \\ Y(x_1) + \varepsilon_1 \\ \vdots \\ Y(x_n) + \varepsilon_n \end{bmatrix} \begin{matrix} \} V \\ \\ \} W \\ \\ \end{matrix}$$

Exercice 18

Montrer que U est un vecteur gaussien

Calculer $\mathbb{E}[U]$ et $cov(U)$

Considerons une combinaison linéaire

$$\alpha_0 Y(x) + \alpha_1 Y(x_1) + \alpha_1 \varepsilon_1 + \dots + \alpha_n Y(x_n) + \alpha_n \varepsilon_n$$

~~$Y(x)$, $Y(x_1)$, …, $Y(x_n)$, ε_1, …, ε_n suivent des~~

~~Pareil~~ $C_1 = \alpha_1 Y(x) + \dots + \alpha_n Y(x_n)$ de loi normale car Y est un GP

$\Rightarrow$ ~~par comb linéaire~~ + $\varepsilon_1, \dots, \varepsilon_n$ de loi normale + indep

~~cette comb linéaire suit une loi normale~~

$\Rightarrow C_2 = \alpha_1 \varepsilon_1 + \alpha_2 \varepsilon_2 + \dots + \alpha_n \varepsilon_n$

$\Rightarrow$ ~~U est un vecteur Gaussien~~ de loi normale

C_1 et C_2 indep

$\Rightarrow C_1 + C_2$ de loi normale

$\Rightarrow U$ est Gaussien

$$\mathbb{E}[U] = \begin{bmatrix} \mathbb{E}[Y(x)] \\ \mathbb{E}[Y(x_1) + \varepsilon_1] \\ \vdots \\ \mathbb{E}[Y(x_n) + \varepsilon_n] \end{bmatrix} = \begin{bmatrix} m(x) \\ m(x_1) \\ \vdots \\ m(x_n) \end{bmatrix} = \begin{bmatrix} m(x) \\ m(X) \end{bmatrix}$$

$$cov(V) = Cov(x, x) = R(x, x)$$

$$cov(W) = \begin{bmatrix} Var(Y(x_1) + \varepsilon_1) & & \\ cov(Y(x_2) + \varepsilon_2, Y(x_1) + \varepsilon_1) & \ddots & \\ & & Var(Y(x_n) + \varepsilon_n) \end{bmatrix}$$

$$Var(Y(x_i) + \varepsilon_i) = Var(Y(x_i)) + Var(\varepsilon_i)$$

car $\varepsilon_i \perp\!\!\!\perp Y$ $= R(x_i, x_i) + \tau_i^2$

$$Cov(Y(x_i) + \varepsilon_i, Y(x_j) + \varepsilon_j)$$

$$= cov(Y(x_i), Y(x_j)) + cov(\varepsilon_i, \varepsilon_j) + cov(Y(x_i), \varepsilon_j) + cov(\varepsilon_i, Y(x_j))$$

$$= R(x_i, x_j)$$

$$cov(W) = \begin{bmatrix} R(x_1, x_1) + \tau_1^2 & R(x_1, x_2) & \cdots & R(x_1, x_n) \\ R(x_2, x_1) & \ddots & & \\ & & \ddots & \\ R(x_n, x_1) & & & R(x_n, x_n) + \tau_n^2 \end{bmatrix}$$

$$cov(U) = \left[\begin{array}{c|ccc} R(x, x) & R(x_1, x) & \cdots & R(x_n, x) \\ \hline R(x_1, x) & & & \\ \vdots & & cov(W) = R(X, X) + \Delta & \\ R(x_n, x) & & & \end{array}\right]$$

avec le bloc $R(x, X)$ en haut et $R(X, x)$ à gauche.

$$\Delta = \begin{pmatrix} \tau_1^2 & & (0) \\ & \ddots & \\ (0) & & \tau_n^2 \end{pmatrix}$$

Exercice 13

$$Y \sim GP(0, k)$$

$$k(s,t) \longrightarrow \operatorname{cov}\big(Y(x), Y'(x_s)\big) = \frac{\partial}{\partial t} \operatorname{cov}\big(Y(s), Y(t)\big) = \frac{\partial k(s,t)}{\partial t}\Big|_{(s,t)=(x,x_s)}$$

$$U = \begin{bmatrix} Y(x) \\ Y'(x_s) \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k_Y(x,x) & \dfrac{\partial k(s,t)}{\partial t}\Big|_{(s,t)=(x,x_s)} \\ \dfrac{\partial k(s,t)}{\partial s}\Big|_{(s,t)=(x,x_s)} & \dfrac{\partial^2 k}{\partial s\, \partial t}\Big|_{(s,t)=(x_s,x_s)} \end{bmatrix} \right)$$

$\operatorname{cov}(Y'(x_s), Y(x))$

$$\mathbb{E}\big[Y(x) \mid Y'(x_s) = d_s\big] = 0 + \frac{\partial k(s,t)}{\partial t}(x,x_s) \cdot \frac{\partial^2 k}{\partial s\, \partial t}(x_s,x_s) \cdot \underbrace{\big(Y'(x_s) - 0\big)}_{= d_s}$$

$$\operatorname{Var}\big[Y(x) \mid Y'(x_s) = d_s\big] = k(x,x) - \frac{\partial k(s,t)}{\partial t}(x,x_s) \cdot \frac{\partial^2 k}{\partial s\, \partial t}(x_s,x_s) \cdot \frac{\partial k}{\partial s}(x,x_s)$$

$$= k(x,x) - \left(\frac{\partial k(s,t)}{\partial t}(x,x_s)\right)^2 \cdot \frac{\partial^2 k}{\partial s\, \partial t}(x_s,x_s)$$

Objectif : chercher le min d'une fonction f

$f : \mathbb{R}^p \to \mathbb{R}$

Contrainte : l'évaluation de f est couteux

Soit $X = (x_1, \dots, x_n) \quad x_i \in \mathbb{R}^p$

$\tilde{Y} = (y_1 = f(x_1), \dots, y_n = f(x_n))$

$\{Y(x) ; x \in X \subset \mathbb{R}^p\} \sim GP(\mu, k)$

$\mathbb{E}[Y(x)] = \mu$

$\mathrm{cov}(Y(x), Y(x')) = k(x, x')$

On définit ~~Soit $X = \{x_1, \dots, x_n\}$~~ $x_i \in \mathbb{R}^p$

$Y_c(x) = \{Y(x) \mid Y(x) = \tilde{Y} , x \in X\} \sim GP(\mu_c(x), k_c(x, x'))$

$\mu_c(x) = \mu(x) - k(x, X)\, k(X, X^{-1})\, Y$

$k_c(x, x')$

$$\boxed{f_{min} = \min(y_1, \dots, y_n)} = \min(\tilde{Y})$$

$\{(f_{min} - Y_c(x))_+ , x \in X\}$

$\varepsilon_i : X \to \mathbb{R}$

$x \mapsto \mathbb{E}(f_{min} - Y_c(x))_+$

$$x_{new} = \underset{x \in X}{\mathrm{argmax}}\ \varepsilon_i(x)$$

2eme étape

$X_2 = \{x_1, \dots, x_n, x_{new}\}$

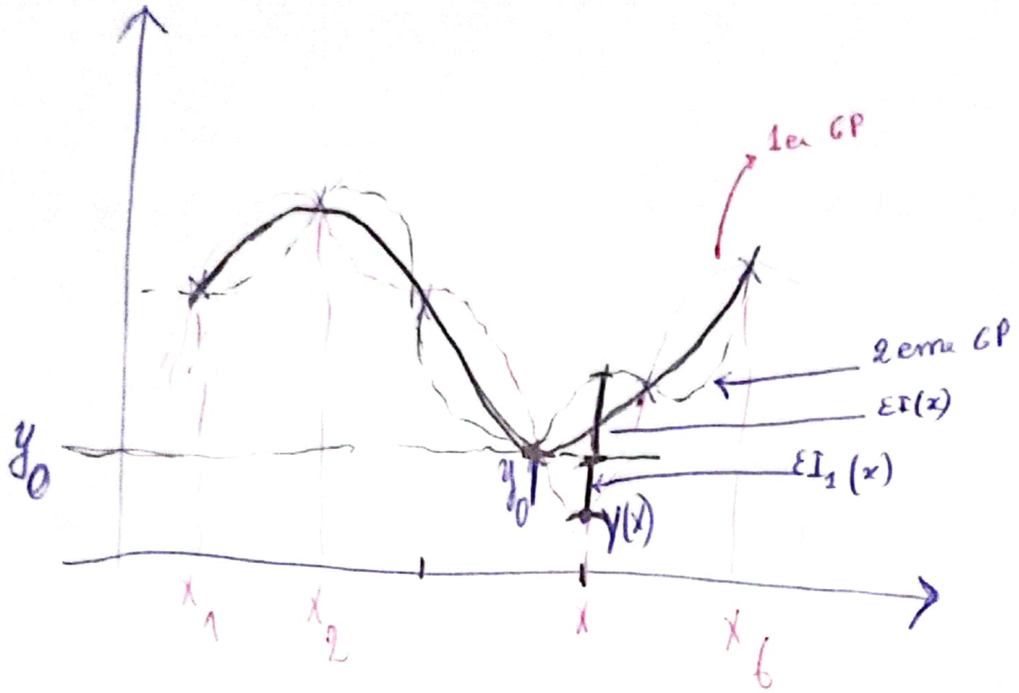
$Y_2 = \{y_1, \dots, y_n, Y_c(x_{new})\}$

$f_{min2} = \min(\tilde{Y}_2)$

$Y_{c_2}(x) = \{Y(x) \mid Y(x) = \tilde{Y}_2 , x \in X\}$

$\{(f_{min} - Y_{c_2}(x)$

E



on calcule EI pour chaque

on prend x qui maximise le critère d'amélioration pour l'ajouter ensuite a l'ensemble

$$x = \arg\max EI(x)$$

Exercice 17 : calcul du critère EI

$$EI(x) = \mathbb{E}\left(\left(y_0 - Y(x)\right)_+ \mid Y(x_i) = y_i \quad i \in [\![1,n]\!]\right)$$

$y_0 = \min(y_1, \dots, y_n)$

avec $Y(x) \sim GP(0, R)$

Mq $EI(x) = s_c(x)\left(z_0\, \Phi(z_0) + \phi(z_0)\right)$

avec ϕ : densité loi $N(0,1)$

Φ cdf de la loi $N(0,1)$

$$z_0 = \frac{y_0 - m_c(x)}{s_c(x)}$$

La loi de $Y(x) \mid Y(x_i) = y_i \quad i \in [\![1,n]\!]$ est une loi

$N\left(m_c(x), s_c^2(x)\right)$

$$EI(x) = \int_{\mathbb{R}} \underbrace{(y_0 - y)_+}_{\text{valeur de } Y(x)} \cdot f_{Y(x) \mid Y(x_i) = y_i}(y)\, dy$$

$$= \int_{\mathbb{R}} (y_0 - y)_+ \frac{1}{s_c^2(x)\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(y - m_c(x))^2}{s_c(x)^2}}\, dy$$

$$= \int_{\frac{y_0 - m_c(x)}{s_c(x)}}^{+\infty}$$

si $y > y_0$

$EI(x) = 0$

si $y < y_0$ $(y_0 - y)_+ = y_0 - y$

$$EI(x) = \int_{-\infty}^{y_0} (y_0 - y) \cdot \frac{1}{\sqrt{2\pi}\, s_c(x)} e^{-\frac{1}{2}\left(\frac{y - m_c(x)}{s_c(x)}\right)^2} dy$$

$$u = \frac{y - m_c(x)}{s_c(x)}$$

$$EI(x) = \int_{-\infty}^{\frac{y_0 - m_c(x)}{s_c(x)}} \left(y_0 - u \cdot s_c(x) - m_c(x)\right) \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du$$

$$z_0 = \frac{y_0 - m_c(x)}{s_c(x)}$$

$$EI(x) = \int_{-\infty}^{z_0} \left(z_0 \cdot s_c(x) - u \cdot s_c(x)\right) \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

$$= s_c(x) \cdot \left[\int_{-\infty}^{z_0} \frac{z_0}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du - \int_{-\infty}^{z_0} \frac{u}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du\right]$$

$$= s_c(x)\left[z_0\, \Phi(z_0) + \int_{-\infty}^{z_0} \frac{(-u)}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}} \cdot du\right]$$

$$= s_c(x)\left[z_0\, \Phi(z_0) + \left[\frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}\right]_{-\infty}^{z_0}\right]$$

$$= s_c(x) = \left[z_0\, \Phi(z_0) + \frac{e^{-\frac{z_0^2}{2}}}{\sqrt{2\pi}}\right]$$

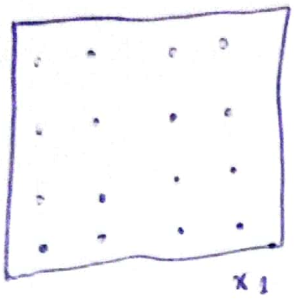
$$= s_c(x)\left[z_0\, \Phi(z_0) + \phi(z_0)\right]$$

Exercice 16

1/ $f_{sim} : [0,1]^2 \to \mathbb{R}$ depends on a single variable x_1

$\Rightarrow$ the other one is inactive

x_2



x_1

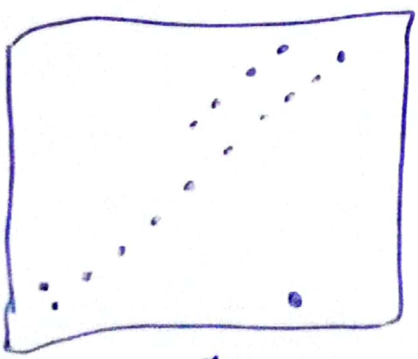
$f_{sim}(1,0) = f_{sim}(1,1)$
$= f_{sim}(1,\frac{1}{3}) = f_{sim}(1,\frac{2}{3})$

comme x_2 est inactive
et x_1 prend seulement 4 valeurs différentes
parmi les 16 qui sont disponibles
$\Rightarrow$ 75% de l'info est perdue

Pour resoudre ce problème, on utilise un hypercube
latin : toutes les valeurs de x_1 (et x_2 ...) seront
différentes
$\Rightarrow$ pas de perte d'information

2/

x_8



x_7

...

Le plan n'est pas adapté en particulier si
$f_{sim}(x_1, x_2, \dots, x_8) = g(x_7, x_8)$

car dans le plan (x_7, x_8) on n'explore pas
bien l'espace.

Si $f(x_1, x_2, \dots, x_8) = g(x_7 - x_8)$

il n'y a que 3 valeurs de $x_7 - x_8$

$\Rightarrow$ on perd $\frac{13}{16}$ d'info

Analyse de Sensibilité

Exercice 1.9 : ANOVA en dimension 2

X_1, X_2 va indépendantes

X_i de loi ν_i $\quad i = 1$ à 2

$f \in L^2(\nu)$ $\quad \nu = \nu_1 \otimes \nu_2$

we want to prove the Sobol Hoeffding

$$Y = f(X_1,X_2) = f_0 + f_1(X_1) + f_2(X_2) + f_{1,2}(X_1,X_2)$$

avec $\mathbb{E}(f_I(X_I) \mid X_J) = 0$ si $J \subsetneq I$

$\Rightarrow$ si $I = \{1\}$ $\quad \mathbb{E}(f_I(X_I) \mid X_J) = \mathbb{E}(f_1(X_1)) = 0$

$J = \phi$

si $I = \{2\}$ $\quad \mathbb{E}[f_2(X_2)] = 0$

$J = \phi$

si $I = \{1,2\}$

$J \to \phi$ $\quad \mathbb{E}[f_{1,2}(X_1,X_2)] = 0$

$J \to \{1\}$ $\quad \mathbb{E}[f_{1,2}(X_1,X_2) \mid X_1] = 0$

$J \to \{2\}$ $\quad \mathbb{E}[f_{1,2}(X_1,X_2) \mid X_2] = 0$

1| Si une telle decomposition existe alors elle est données par

$f_0 = \mathbb{E}(f(X))$

$i = 1,2$ $\quad f_i(X_i) = \mathbb{E}(f(X) \mid X_i) - f_0$

$f_{1,2}(X_1,X_2) = \mathbb{E}(f(X) \mid X_1,X_2) - f_1(X_1) - f_2(X_2) - f_0$

$$f(X_1,X_2) = f_0 + f_1(X_1) + f_2(X_2) + f_{1,2}(X_1,X_2)$$

$$\mathbb{E}[f(X_1,X_2)] = f_0 + \mathbb{E}[f_1(X_1)] + \mathbb{E}[f_2(X_2)] + \mathbb{E}[f_{1,2}(X_1,X_2)]$$

$X = (X_1, X_2)$

d'apres les hypotheses

$\mathbb{E}[f_1(X_1)] = 0 \quad \mathbb{E}[f_2(X_2)] = 0$

$\mathbb{E}[f_{1,2}(X_1,X_2)] = 0$

ainsi $\boxed{\mathbb{E}[f(X)] = f_0}$

$$\mathbb{E}(f(X_1,X_2) \mid X_1) = f_0 + \mathbb{E}[f_1(X_1) \mid X_1] + \mathbb{E}[f_2(X_2) \mid X_1] + \underbrace{\mathbb{E}[f_{1,2}(X_1,X_2) \mid X_1)]}_{=0 \text{ par hypothèse}}$$

on connait deja X_1

$\Rightarrow \mathbb{E}[f_1(X_1) \mid X_1] = f_1(X_1)$

$\mathbb{E}[f_2(X_2) \mid X_1] = \mathbb{E}[f_2(X_2)] = 0$

car $X_1 \perp\!\!\!\perp X_2$

Ainsi

$\mathbb{E}[f(X_1,X_2) \mid X_1] = f_0 + f_1(X_1)$

$$\boxed{f_1(X_1) = \mathbb{E}[f(X_1,X_2) \mid X_1) - f_0}$$

$$\mathbb{E}[f(X_1,X_2)|X_1,X_2] = \mathbb{E}[f_0|X_1,X_2] + \mathbb{E}[f_1(X_1)|X_1,X_2] + \mathbb{E}[f_2(X_2)|X_1,X_2] + \mathbb{E}[f_{1,2}(X_1,X_2)|X_1,X_2]$$

$$= f_0 + f_1(X_1) + f_2(X_2) + f_{1,2}(X_1,X_2)$$

$$f_{1,2}(X_1,X_2) = \mathbb{E}[f(X_1,X_2)|X_1,X_2] - f_0 - f_1(X_1) - f_2(X_2)$$

~~2) si les hypothèses sont vérifiées~~

$$\cancel{f_{1,2}(X_1,X_2) = \mathbb{E}[f(X)|X_1,X_2] - f_1(X_1) - f_2(X_2) - f_0}$$

$$\cancel{= \mathbb{E}[f(X)|X_1,X_2] - \mathbb{E}[f(X)|X_1] + f_0 - \mathbb{E}[f(X)|X_2] + f_0 - f_0}$$

$$\cancel{= \mathbb{E}[f(X)|X_1,X_2] - \mathbb{E}[f(X)|X_1] - \mathbb{E}[f(X)|X_2] + \mathbb{E}[f(X)]}$$

si les hypothèses sont vérifiées

$$f_1(X_1) = \mathbb{E}[f(X)|X_1] - f_0$$

$$\mathbb{E}[f_1(X_1)] = \mathbb{E}[\mathbb{E}(f(X)|X_1)] - \mathbb{E}[f_0]$$

$$= \mathbb{E}[f(X)] - f_0$$

car $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$

$$= f_0 - f_0$$

$$= 0$$

de même

$$\mathbb{E}[f_2(X_2)] = 0$$

$$\mathbb{E}[f_{1,2}(X_1,X_2)|X_1] = \mathbb{E}[\overbrace{\mathbb{E}[f(X)|X_1,X_2]}^{=f(X)}|X_1]$$

$$- \mathbb{E}[f_1(X_1)|X_1] - \mathbb{E}[f_2(X_2)|X_1] - \mathbb{E}[f_0|X_1]$$

$$= \mathbb{E}[f(X)|X_1] - \mathbb{E}[f_1(X_1)|X_1] - \underbrace{\mathbb{E}[f_2(X_2)|X_1]}_{=0 \text{ car } X_2 \perp\!\!\!\perp X_1} - f_0$$

$$= 0$$

2/ $f_{1,2}(X_1,X_2) = \mathbb{E}[f(X)|X_1,X_2] - \cancel{\mathbb{E}[f(X)|X_1]} - \cancel{\mathbb{E}[f(X)|X_2]} + \cancel{\mathbb{E}[f(X)]}$

$f_1(X_1)$ $f_2(X_2)$ f_0

$= \mathbb{E}[f(X)|X_1,X_2] - \mathbb{E}[f(X)|X_1] + f_0 - \mathbb{E}[f(X)|X_2] + f_0 - f_0$

$= \mathbb{E}[f(X)|X_1,X_2] - \mathbb{E}[f(X)|X_1] - \mathbb{E}[f(X)|X_2] + \mathbb{E}[f(X)]$

3/ $I = \{1\}$ $I' = \emptyset$

$\mathbb{E}[f_I(X_I)\ f_{I'}(X_{I'})] = 0$

$I = \{1\}$ $I' = \{2\}$

$\mathbb{E}[f_1(X_1)\ f_2(X_2)] = \mathbb{E}[f_1(X_1)] \cdot \mathbb{E}[f_2(X_2)]$

$=$

$I = \{1\}$ $I' = \{1,2\}$

$\mathbb{E}[f_1(X_1)\ f_{1,2}(X_1,X_2)]$

$= \mathbb{E}[\mathbb{E}[\underbrace{f_1(X_1)}_{\text{devient déterministe}}\ f_{1,2}(X_1,X_2) \mid X_1]]$

$= \mathbb{E}[f_1(X_1) \cdot \underbrace{\mathbb{E}[f_{1,2}(X_1,X_2)|X_1]}_{0 \text{ par hypothèse}}]$

$= 0$

4/ $f(X_1,X_2) = X_1$ X_1 is centered

$f(X_1,X_2) = 0 + X_1 + 0 + 0$

$= \underbrace{0}_{f_0} + \underbrace{0}_{f_1(X_1)} + \underbrace{0}_{f_2(X_2)} + \underbrace{X_1}_{f_{1,2}(X_1,X_2)}$

si la 2eme était vraie

$\mathbb{E}[f_{1,2}(X_1,X_2)|X_1) = \mathbb{E}[X_1|X_1] = X_1 \neq 0$ absurde

Décomposition de Sobol - Hoeffding

$X_1, \dots, X_d$ va indépendantes

$$f(X_1, \dots, X_d) = f_0 + f_1(X_1) + \dots + f_d(X_d) + f_{1,2}(X_1, X_2) + \dots + f_{1,2,\dots,d}(X_1, X_2, \dots, X_d)$$

On a existence et unicité d'une décomposition

f : $X = (X_1, \dots, X_d)$

$$f(X) = f_0 + \sum_{i=1}^{d} f_i(X_i) + \sum_{i<j} f(X_i, X_j) + \dots + f_{1,2,\dots,d}(X_1, X_2, \dots, X_d) \qquad I \subseteq \{1, \dots, d\}$$

avec $\mathbb{E}(f_I(X_I) \mid X_J) = 0$ si $J \subsetneq I$

On a

$$f_0 = \mathbb{E}[f(X)]$$

$$f_i(X_i) = \mathbb{E}[f(X) \mid X_i] - f_0 \text{ : effet principal}$$

$$f_{i,j}(X_i, X_j) = \mathbb{E}(f(X) \mid X_i, X_j] - f_i(X_i) - f_j(X_j) - f_0 \text{ : interaction d'ordre 2}$$

$$= \mathbb{E}[f(X) \mid X_i, X_j] - \mathbb{E}[f(X) \mid X_i] - \mathbb{E}[f(X) \mid X_j] + f_0$$

$$f_I(X_I) = \mathbb{E}[f(X) \mid X_I] - \sum_{J \subset I} f_J(X_J) \longrightarrow \text{Recursion Formula}$$

$$= \sum_{J \subseteq I} (-1)^{|I| - |J|} \mathbb{E}[f(X) \mid X_J] \longrightarrow \text{inclusion - Exclusion}$$

En outre : tous les termes sont centrés et orthogonaux

Si $I \neq J$ $\quad \mathbb{E}(f_I(X_I)\, f_J(X_J)) = 0 \iff \text{cov}(f_I(X_I), f_J(X_J)) = 0$

Décomposition de la Variance (ANOVA) $\rightarrow$ analysis of Variance

$$\mathrm{Var}(f(X)) = \sum_{i=1}^{d} \mathrm{Var}\, f_i(X_i) + \sum_{i<j} \mathrm{Var}(f_{i,j}(X_i, X_j)) + \dots + \mathrm{Var}(f_{1,2,\dots,d}(X_1, \dots, X_d))$$

On mesure l'influence de X_i sur $Y = f(X)$ par

$$\boxed{S_i = \frac{\mathrm{Var}(f_i(X_i))}{\mathrm{Var}(f(X))}}$$

% de variance expliquée par X_i

$\rightarrow$ Indice de Sobol $\in [0,1]$

$$S_i^{tot} = \frac{\mathrm{Var}\left(\sum_{\substack{I \subseteq \{1,\dots,d\} \\ i \in I}} f_I(X_I) \right)}{\mathrm{Var}(f(X))}$$

$$= \sum_{i \in I} \frac{\mathrm{Var}\, f_I(X_I)}{\mathrm{Var}(f(X))}$$

: indice de Sobol total

% de variance expliqué par X_i et toutes ses interactions

$$S_I = \frac{\mathrm{Var}(f_I(X_I))}{D}$$

% de variance expliquée par le terme $f_I(X_I)$

Rq : $\sum_{I \subseteq \{1,\dots,d\}} S_I = 1$

Exercice 21 : Ishigami function

$$\Delta = [-\pi, \pi]^3 \qquad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$f(x) = \sin(x_1) + A \sin^2(x_2) + B\, x_3^4 \sin(x_1) \qquad \text{avec } A = 7 \qquad B = 0.1$$

① Calculer la décomposition de Sobol Hoeffding de $f(X_1, X_2, X_3)$

avec X_1, X_2, X_3 uniformes sur $[-\pi, \pi]$ et indépendantes

Calculer les indices de Sobol

On donne $a = \int_{-\pi}^{\pi} \sin^2(u) \cdot \frac{du}{2\pi} = \frac{1}{2}$

$$b = \int_{-\pi}^{\pi} u^4 \cdot \frac{du}{2\pi} = \frac{\pi^4}{5}$$

$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = \mathbb{E}[X_3] = 0$$

$$f_0 = \mathbb{E}[f(X)] = \mathbb{E}[\sin(X_1)] + A\, \mathbb{E}[\sin^2(X_2)] + B \cdot \mathbb{E}[X_3^4 \sin(X_1)]$$

$$= \int_{-\pi}^{\pi} \sin(x_1) \cdot \frac{1}{2\pi} \cdot dx_1$$

$$\mathbb{E}[\sin(X_1)] = 0$$

$$\mathbb{E}[\sin^2(X_2)] = \int_{-\pi}^{\pi} \sin^2(x_2) \cdot \frac{1}{2\pi} \cdot dx_2 = \frac{1}{2}$$

$$\mathbb{E}[X_3^4 \sin(X_1)] = \mathbb{E}[X_3^4]\, \mathbb{E}[\sin(X_1)] \quad \text{par indep}$$

$$= \int_{-\pi}^{\pi} x_3^4 \cdot \frac{1}{2\pi} \cdot dx_3 \cdot \int_{-\pi}^{\pi} \sin(x_1) \cdot \frac{1}{2\pi} \cdot dx_1 = 0$$

$$f_0 = \frac{A}{2}$$

$f_1(X_1) = \mathbb{E}[f(X)|X_1]$

$= \mathbb{E}[\sin(X_1)|X_1) + A\,\mathbb{E}[\sin^2(X_2)|X_1] + B\,\mathbb{E}[X_3^4 \sin(X_3)|X_1] - \mathbb{E}[f]$

$= \mathbb{E}[\sin(X_1)|X_1) + A\,\mathbb{E}[\sin^2(X_2) + B \sin(X_3) \cdot \mathbb{E}[X_3^4] - f_0$

$= \sin X_1 + A\,\mathbb{E}[\sin^2(X_2) + B \sin(X_3) \cdot \mathbb{E}[X_3^4] - f_0$

car $X_1 \perp X_2$

$= \sin X_1 + \frac{A}{2} + \frac{B.\pi^4}{5} \sin(X_1) - \frac{A}{2}$

$= \sin(X_1)\left(1 + B.\frac{\pi^4}{5}\right)$

$f_2(X_2) = \mathbb{E}[f(X)|X_2] = \mathbb{E}[\sin(X_1)] + A\sin^2(X_2) + B\,\mathbb{E}[X_3^4 \sin X_3] - f_0$

$= A\sin^2(X_2) - \frac{A}{2}$

$f_3(X_3) = \mathbb{E}[f(X)|X_3] = \mathbb{E}[\sin(X_1)] + A.\mathbb{E}[\sin^2(X_2)] + B X_3^4 . \mathbb{E}[\sin(X_1)] - f_0$

$= 0 + \frac{A}{2} + 0 - \frac{A}{2}$

$= 0$

$f_{1,3}(X_1, X_3) = \mathbb{E}[f(X)|X_1, X_3] = \sin(X_1) + A.\mathbb{E}[\sin^2(X_2)] + B X_3^4 \sin(X_1) - f_1(X_1) - f_3(X_3) - f_0$

$= \frac{A}{2} + \sin(X_1)(1 + B X_3^4) - \sin(X_1)\left(1 + B\frac{\pi^4}{5}\right) - 0 - \frac{A}{2}$

~~$f_{1,2}(X_1, X_2) = \mathbb{E}[f(X)|X_1, X_2] = \sin(X_1) + A\sin(X_2) + B \sin(X_1)\,\mathbb{E}[X_3^4]$~~

~~$= \sin(X_1) + A\sin(X_2)$~~

$= \sin(X_1)\left(B X_3^4 - B.\frac{\pi^4}{5}\right)$

$f_1(X_1) + f_2(X_2) + f_3(X_3) + f_{1,3}(X_1, X_3) = f(X)$

$\Rightarrow$ On peut arrêter le développement et déduire que

$f_{1,2}(X_1, X_2) = 0 \quad f_{2,3}(X_2, X_3) = 0$ et $f_{1,2,3}(X_1, X_2, X_3) = 0$

Calcul des indices de Sobol

$$S_i = \frac{D_i}{D} \qquad D_i = Var\left(f_i(X_i)\right)$$

$$S_I = \frac{D_I}{D} \qquad D_I = Var\left(f_I(X_I)\right)$$

Astuce 1 : $\mathbb{E}\left[f_I(X_I)\right] = 0 \Rightarrow D_I = \mathbb{E}\left[f_I^2(X_I)\right]$

Astuce 2 : $D = \sum_I D_I$

$$D_1 = \mathbb{E}\left[f_1^2(X_1)\right] = \mathbb{E}\left[\sin^2(X_1)\left(1 + B\frac{\pi^4}{5}\right)^2\right]$$

$$= \left(1 + B\frac{\pi^4}{5}\right)^2 \cdot \int_{-\pi}^{\pi} \sin^2(x_1) \cdot \frac{1}{2\pi} \cdot dx_1 \cancel{= \frac{1}{2}}$$

$$= \frac{1}{2}\left(1 + B\frac{\pi^4}{5}\right)^2$$

$$D_0 = \mathbb{E}\left[f_0^2\right] = \frac{A^2}{4}$$

$$D_2 = \mathbb{E}\left[f_2^2(X_2)\right] = \mathbb{E}\left[A^2 \sin^4(X_2)\right] - \mathbb{E}\left[A^2 \sin^2(X_2)\right] + \mathbb{E}\left[\frac{A^2}{2}\right]$$

$$= \frac{1}{2\pi} A^2 \cdot \frac{3\pi}{4} - \frac{A^2}{2} + \frac{A^2}{2}$$

$$= \frac{3A^2}{8}$$

$$D_3 = \mathbb{E}\left[f_3^2(X_3)\right] = 0$$

$$D_{1,2} = 0$$

$$D_{2,3} = 0$$

$$D_{1,3} = \mathbb{E}\left[\sin^2(X_1)\right] \cdot \mathbb{E}\left[B^2\left(X_3^4 - \frac{\pi^4}{5}\right)^2\right]$$

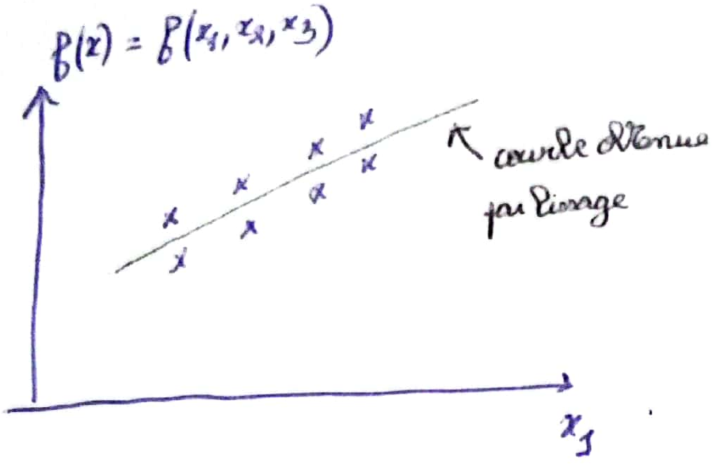
$$= B^2\left(\mathbb{E}\left[X_3^8\right] - \frac{2\pi^4}{5}\mathbb{E}\left[X_3^4\right] + \frac{\pi^8}{5}\right) \cdot \frac{A}{2}$$

$$= B^2 \cdot \left(\frac{\pi^8}{9} - \frac{\pi^4}{5}\right)\frac{A}{2}$$

Vérification numérique des résultats : Estimation de $x_1 \longrightarrow \mathbb{E}\{f(X) \mid X_1 = x_1\}$

idée ① : simuler $x^{(1)}, \dots, x^{(N)}$ selon la loi de X

② représenter $f(x^{(i)})$ contre $x_1^{(i)}$ et estimer la courbe moyenne de



~~Estimat~~

Exercice : courbe 1 : On remarque que la tendance donne $f_1(X_1)$ est sous forme de sinus (période = π)

courbe 2 : On remarque la tendance donne $f_2(X_2)$ est sous forme de $\sin^2$ (période = π)

courbe 3 : On remarque que $f_3(X_3) = 0$

Random Process: family of RV $(Y(x)) = Y$

Trajectory: $x \mapsto Y(x)(w)$

Second order RP: when $Y(x) \in L^2(\Omega)$

mean: $x \to \mathbb{E}[Y(x)]$; Kernel $k(x,x') = cov(Y(x),Y(x'))$

$Var(Y(x)) = k(x,x)$

Stationarity: the law $(Y(x_1+h),\dots,Y(x_n+h))$ does not depend on h

weak stationarity: $\mathbb{E}[Y(x)] = m$ (constant)

$k(x,x') = c(x-x')$: depends only on $x-x'$

Gaussian Process: if $(Y(x_1),\dots, Y(x_n))$ is a Gaussian vector $\Rightarrow Y \sim GP(m,k)$

(it is stationnary $\Leftrightarrow$ weakly stationnary)

independance of $Y(x)$ and $Y(x')$ $\Leftrightarrow$ $k(x,x') = 0$

Stability of GP

1/ Y a GP and L linear operator $\Rightarrow LY$ a GP

a/ $Y \sim GP(0,k) \Rightarrow LY \sim GP(0,k_L)$

$k_L(s,t) = L_s L_t \cdot k(s,t)$

b/ $Y \sim GP(0,k) \Rightarrow Y' \sim GP(0,k_2)$

$k_2(s,t) = \frac{\partial^2 k_0}{\partial s\, \partial t}(s,t)$

matrice de covariance

$$cov\begin{pmatrix} Z(x_1)\\ Z(x_2)\\ Z(x_3)\end{pmatrix} = \begin{pmatrix} Var(Z(x_1)) & cov(Z(x_1),Z(x_2)) \\ cov(Z(x_2),Z(x_1)) & Var(Z(x_2),Z(x_2)) \\ cov(Z(x_3),Z(x_1)) & cov(Z(x_3),Z(x_2)) \end{pmatrix}$$

$$= k(X,X)$$

Mq Y est un GP: 1/ prendre une comb linéaire et mq c'est suit une loi normale

example of kernels of 1-D GP

1/ cosine: $k(x,x') = \cos(2\pi(x-x'))$

2/ sinc: $k(x,x') = \frac{\sin(\pi(x-x'))}{\pi(x-x')}$

3/ squared exp: $k(x,x') = \exp\left(-\frac{1}{2}\frac{(x-x')^2}{\theta^2}\right)$

4/ exp: $k(x,x') = \exp\left(-\frac{|x-x'|}{\theta}\right)$

5/ matern 3/2: $\left(1+\sqrt{3}\frac{|x-x'|}{\theta}\right)\exp\left(-\sqrt{3}\frac{|x-x'|}{\theta}\right)$

6/ matern 5/2: $\left(1+\sqrt{5}\frac{|x-x'|}{\theta} + \frac{5}{3}\frac{(x-x')^2}{\theta^2}\right)\exp\left(-\sqrt{5}\frac{|x-x'|}{\theta}\right)$

operations on kernels

k kernel $\Rightarrow ck$ kernel

k_1, k_2 kernels $\Rightarrow$ k_1+k_2 kernel

$k_1 \oplus k_2$ kernel (on $(x_1, x_2) \in X^2$)

$k_1 k_2$ kernel

$k_1 \otimes k_2$ kernel

k kernel on X^2 $\Rightarrow$ $k_f(x,x') = k(f(x),f(x'))$ kernel on U^2

$f: U \to X$

How to obtain kernels on d-dimensional space

The tensor product is widely used to define kernels on d-dimensional spaces from 1-dimensional space

$k(x,x') = e^{-\|x-x'\|^2}$

Gaussian Process approach

$$U = \begin{pmatrix} V \\ W \end{pmatrix}$$

$$V = (y(x)) \qquad W = \begin{pmatrix} y(x_1)\\ \vdots \\ y(x_n)\end{pmatrix}$$

U est un vecteur gaussien

$$U \sim N\left(\begin{pmatrix} m(x)\\ m(x_1)\\ \vdots\\ m(x_n)\end{pmatrix}, \begin{pmatrix} k(x,x) & k(x,X)\\ k(X,x) & k(X,X)\end{pmatrix}\right)$$

$m_c(x) = \mathbb{E}[V \mid W=w]$

$= m(x) + k(x,X)\,k(X,X)^{-1}(y-m(x))$

$Var(V \mid W=w) = k(x,x) - k(x,X)\,k(X,X)^{-1}\,k(X,x)$

intervalle de confiance

$Z \sim N(\mu_Z, \sigma_Z^2)$

$\mu_Z \pm 2\sigma_Z$ à 95%

$\mathbb{P}[Z \in [\underbrace{\mu_Z - 2\sigma_Z}_{q_{2,5}}, \underbrace{\mu_Z + 2\sigma_Z}_{q_{97,5}}]] = 95\%$

Observations bruitées

$$V = (y(x)) \qquad W = \begin{pmatrix} y(x_1)+\varepsilon_1\\ \vdots \\ y(x_n)+\varepsilon_n\end{pmatrix}$$

$$\mathbb{E}[U] = \begin{pmatrix} m(x)\\ m(x_1)\\ \vdots\\ m(x_n)\end{pmatrix} \qquad \varepsilon_i \sim N(0,\sigma_i^2)$$

$cov(W) = k(X,X) + \Delta$

$$\Delta = \begin{bmatrix} \sigma_1^2 & & (0) \\ & \ddots & \\ (0) & & \sigma_n^2 \end{bmatrix}$$

$$cov(U) = \begin{bmatrix} k(x,x) & k(x,X) \\ k(X,x) & k(X,X)+\Delta \end{bmatrix}$$

Bayesian optimization

is a powerful technique for optimizing time consuming functions.
The main principle behind it is to build a probabilistic model on the function and using it to guide the search for the best solution:

- it involves
  1/ initial sampling of the function (initial evaluations)
  2/ create a surrogate model to estimate the function's behavior typically using gaussian process (thanks to their flexibility in modeling complex functions and uncertainty: capture mean and variance of the function)
  3/ using an acquisition function to determine the next point to evaluate. It balances exploration and exploitation (trying unexplored regions and sampling in regions with high predicted values) EX: Expected improvement — EGO algorithm
  4/ evaluate the expensive objective function at the selected point
  5/ updating the surrogate model with the new data
  6/ repeat 3, 4, 5 iteratively
  7/ stop the algorithm based on a stopping criterion
  8/ return the best solution

pseudo code: initial DOE $\{x_1, \dots, x_n\}$ and $y = \{y_1, \dots, y_n\}$

for i in range nb_iterations
 build surrogate model: GP model with X and y
 compute acquisition function: $EI(x)$
 find maximum of the criterion: $\max EI(x)$
 evaluate the objective function to this max point
 ~~add it to the dataset~~ → update the DOE
 find the best solution

tune the parameters of a model ⇒ apply BO to a cross validation error criterion

if one parameter is discrete and the squared exponential kernel?

the usual squared exponential kernel is based on distance
if the variable is ordinal (number) it can make sense (distance between these levels)
if it is not ordinal, it is much less making sense.
there are specific kernels for this case.
in the ordinal case, such kernels account for non stationnarities: the distance between levels 1 & 2 and 3 is not necessarily the same between levels 2 and 3

pour tuner les hyperparamètres: On utilise un hypercube latin design (LHD) using LHS: Latin hypercube sampling
tous les valeurs de x_1 et x_2 seront différentes: diminuer le risque d'oversampling ou d'undersampling.
~~⇒ diminuer le nombre d'expériences à effe~~
On explore bien l'espace

~~calcul de~~

LHD construction

is based on the idea of stratified sampling
the projection of points that belong to some interval are forced to equal the theorical one

EGO: Efficient global optimization

criterion with noise: Expected quantile improvement criterion

analyse de sensibilité

$f_0 = \mathbb{E}[f(X)]$

$f_i(X_i) = \mathbb{E}[f(X) \mid X_i] - f_0$

$f_{i,j}(X_i, X_j) = \mathbb{E}[f(X) \mid X_i, X_j] - f_i - f_j - f_0$

$$f_I(X_I) = \mathbb{E}[f(X) \mid X_I] - \sum_{J \subset I} f_J(X_J)$$

$$= \sum_{J \subset I} (-1)^{|I|-|J|} \mathbb{E}[f(X) \mid X_J]$$

les termes sont centrés et orthogonaux

$$S_i = \frac{Var(f_i(X_i))}{Var(f(X))} \text{ ; indice de Sobol}$$

$$S_i^T = \sum_{i \in I} \frac{Var(f_I(X_I))}{Var(f(X))}$$

$$S_I = \frac{Var(f_I(X_I))}{Var(f(X))}$$

Rq $\sum S_I = 1$

Ici $\mathbb{E}[f_I(X_I)] = 0 \Rightarrow D_I = \mathbb{E}[f_I^2(X_I)]$

$$D = \sum_I D_I$$