

Université Paul Sabatier
Faculté des Sciences et de l'ingénierie

PROCESSUS GAUSSIENS

L3 FLEX : MACHINE LEARNING

2023-2024

Responsable UE : Fabrice Gamboa

11 avril 2024

0.1 Introduction

Dans ces notes de cours, nous nous intéressons à une fonction

$$f : [0, 1]^d \rightarrow \mathbb{R}$$

Cette fonction est inconnue (nous ne savons pas comment $f(x)$ est calculée pour $x \in [0, 1]^d$). Nous ne pouvons observer que quelques valeurs de f de la forme

$$\begin{pmatrix} (x_1, f(x_1)) \\ \vdots \\ (x_n, f(x_n)) \end{pmatrix}.$$

Notre objectif est, essentiellement, d'utiliser ces valeurs (nos données statistiques) pour inférer la fonction f , comme illustré dans la Figure 1.

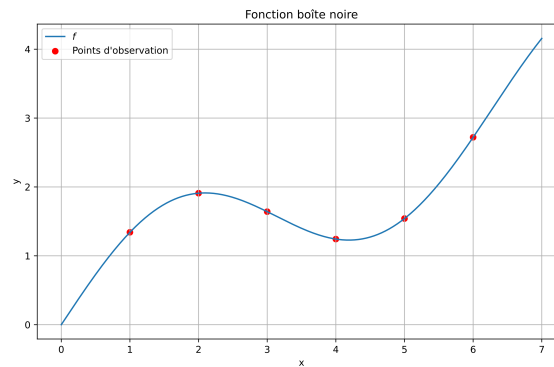


FIGURE 1 – Fonction inconnue et points d'observation.

Domaines d'application : Lorsque f représente un modèle informatique ou une expérience informatique, choisir $x \in [0, 1]^d$ correspond à définir les valeurs des d paramètres de simulation. Après une (longue et coûteuse) computation, le modèle informatique fournit le résultat de la simulation $f(x)$. Nous ne savons pas comment f fonctionne, car f implique, par exemple, des modèles physiques complexes et des schémas numériques. Ainsi, il y a un nombre limité de résultats

$$\begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix}.$$

Exemples : Simulation de composants aéronautiques, industrie automobile, ingénierie nucléaire. Les composants de x peuvent correspondre à des paramètres géométriques, des caractéristiques matérielles, des concentrations physiques,...

0.2 Rappels sur les vecteurs gaussiens

Proposition 1 : Soit V un vecteur aléatoire dans \mathbb{R}^n . Les trois assertions suivantes sont équivalentes :

1. Toute combinaison linéaire de V suit une distribution gaussienne. Autrement dit, pour tout vecteur a fixé de dimension $n \times 1$, il existe $\mu \in \mathbb{R}$ et $\sigma^2 \geq 0$ tels que

$$a^\top V = \sum_{i=1}^n a_i V_i \sim \mathcal{N}(0, \sigma^2)$$

2. Il existe $m \in \mathbb{R}^n$ et une matrice Σ symétrique non-négative définie (SNND) de dimension $n \times n$ telle que pour tout vecteur u de dimension $n \times 1$,

$$\phi_V(u) = \exp\left(1u^\top m - \frac{1}{2}u^\top \Sigma u\right)$$

où ϕ_V est la fonction caractéristique de V , avec $\phi_V(u) = \mathbb{E}\left(e^{iu^\top V}\right)$.

3. Il existe un vecteur m' de dimension $n \times 1$, une matrice K de dimension $n \times r$, avec $r \leq n$, et un vecteur w de dimension $r \times 1$ avec des composantes indépendantes suivant une distribution $\mathcal{N}(0, 1)$, tels que

$$V = m' + Kw.$$

De plus, avec m, Σ comme dans 2 et m', K comme dans 3, nous avons

$$m = m' = \mathbb{E}(V) \quad (\text{vecteur moyen})$$

et

$$\Sigma = KK^\top = \text{cov}(V) \quad (\text{matrice de covariance})$$

Nous disons que V est un vecteur gaussien lorsqu'il satisfait aux trois assertions précédentes (il suffit de vérifier qu'une des trois est satisfaite pour montrer que V est un vecteur gaussien, puisqu'elles sont équivalentes).

0.3 Processus gaussiens

Définition 2 : Soit (Ω, \mathcal{A}, P) un espace probabilisé et

$$Z : (\Omega, \mathcal{A}, P) \times [0, 1]^d \rightarrow \mathbb{R} \\ (\omega, x) \rightarrow Z(\omega, x).$$

On dit que Z est un processus gaussien sur $[0, 1]^d$ lorsque, pour tout $n \in \mathbb{N}$ et pour tout $x_1, \dots, x_n \in [0, 1]^d$, la fonction $\omega \mapsto (Z(\omega, x_1), \dots, Z(\omega, x_n))$ est un vecteur gaussien (aléatoire).

Pour tout $\omega, x \mapsto Z(\omega, x)$ est une fonction de $[0, 1]^d$ dans \mathbb{R} . Ainsi, un processus gaussien est une fonction aléatoire. Nous appelons $x \mapsto Z(\omega, x)$ une trajectoire, ou un chemin d'échantillonnage, de Z .

Exemple :

Soient X_1, X_2 deux variables aléatoires indépendantes avec une distribution $\mathcal{N}(0, 1)$. Soit

$$Z(x) = X_1 + \cos(x)X_2$$

pour $x \in [0, 1]$. Alors Z est un processus gaussien. En effet, pour tout $n \in \mathbb{N}$ et $x_1, \dots, x_n \in [0, 1]$,

$$\begin{pmatrix} Z(x_1) \\ \vdots \\ Z(x_n) \end{pmatrix} = \begin{pmatrix} X_1 + \cos(x_1)X_2 \\ \vdots \\ X_1 + \cos(x_n)X_2 \end{pmatrix} = \begin{pmatrix} 1 & \cos(x_1) \\ \vdots & \vdots \\ 1 & \cos(x_n) \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

est un vecteur gaussien d'après le point 3 de la Proposition 1.

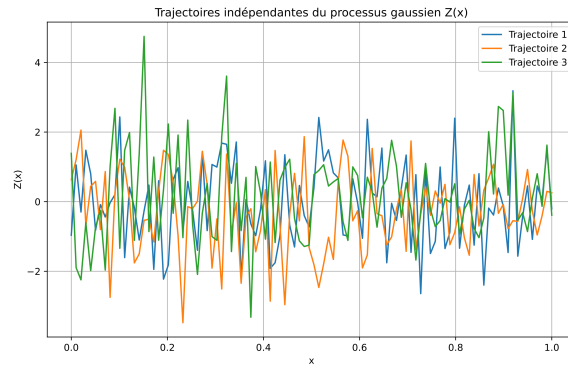


FIGURE 2 – Exemples de chemins de processus gaussiens

0.3.1 Fonction moyenne

Définition 3 : Soit Z un processus gaussien sur $[0, 1]^d$. La fonction

$$\begin{aligned} m : [0, 1]^d &\rightarrow \mathbb{R} \\ x &\rightarrow \mathbb{E}(Z(x)) \end{aligned}$$

est appelée la fonction moyenne de Z .

0.3.2 Fonction de covariance

Définition 4 : Soit Z un processus gaussien sur $[0, 1]^d$. La fonction

$$\begin{aligned} K : [0, 1]^d \times [0, 1]^d &\rightarrow \mathbb{R} \\ (x, y) &\rightarrow \text{cov}(Z(x), Z(y)) \end{aligned}$$

est appelée la fonction de covariance de Z .

Si la fonction K dépend uniquement de $x - y$, c'est-à-dire $x - y = x' - y' \implies K(x, y) = K(x', y')$, alors nous disons que K est stationnaire. Dans ce cas, par abus de notation, nous écrivons $K(x, y) = K(x - y)$.

Proposition 5 : Soit Z un processus gaussien sur $[0, 1]^d$, avec une fonction moyenne constante et une fonction de covariance stationnaire. Alors Z est un processus stationnaire. C'est-à-dire, pour tout $n \in \mathbb{N}$, pour tout $x_1, \dots, x_n \in [0, 1]^d$, pour tout $\delta \in \mathbb{R}^d$ tel que $x_1 + \delta, \dots, x_n + \delta \in [0, 1]^d$, la distribution de

$$\begin{pmatrix} Z(x_1) \\ \vdots \\ Z(x_n) \end{pmatrix}$$

est la même que celle de

$$\begin{pmatrix} Z(x_1 + \delta) \\ \vdots \\ Z(x_n + \delta) \end{pmatrix}$$

La stationnarité signifie que la distribution est invariante par translation.

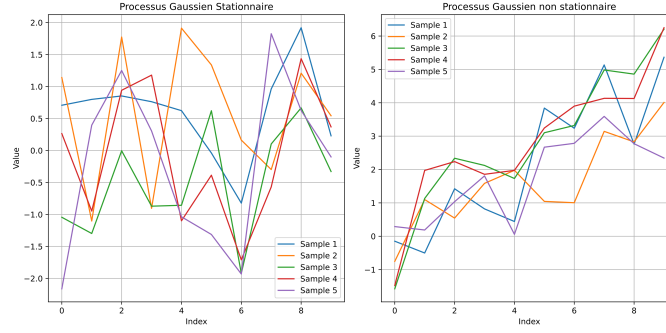


FIGURE 3 – Exemples de chemins de processus gaussiens

0.4 Prédiction par processus gaussien

Théorème : (Conditionnement gaussien) Nous considérons un vecteur gaussien de taille $(n_1 + n_2) \times 1$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

où Y_1 est de taille $n_1 \times 1$ et Y_2 est de taille $n_2 \times 1$. Nous écrivons le vecteur moyen de $(Y_1^\top, Y_2^\top)^\top$ comme

$$\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$$

où m_1 est de taille $n_1 \times 1$ et m_2 est de taille $n_2 \times 1$. Enfin, nous écrivons la matrice de covariance de $(Y_1^\top, Y_2^\top)^\top$ comme

$$\begin{pmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2}^\top & \Sigma_2 \end{pmatrix}$$

où Σ_1 est de taille $n_1 \times n_1$, $\Sigma_{1,2}$ est de taille $n_1 \times n_2$ et Σ_2 est de taille $n_2 \times n_2$. Nous supposons que Σ_1 est inversible. Alors, pour tout $y \in \mathbb{R}^{n_1}$, conditionnellement à $Y_1 = y_1$, le vecteur aléatoire Y_2 est un vecteur gaussien avec un vecteur moyen

$$\mathbb{E}(Y_2 \mid Y_1 = y_1) = m_2 + \Sigma_{1,2}^\top \Sigma_1^{-1} (y_1 - m_1)$$

et une matrice de covariance

$$\text{Cov}(Y_2 \mid Y_1 = y_1) = \Sigma_2 - \Sigma_{1,2}^\top \Sigma_1^{-1} \Sigma_{1,2}.$$

Interprétation du théorème de conditionnement gaussien :

- Nous nous trouvons dans un cadre favorable avec des vecteurs gaussiens car la distribution conditionnelle reste gaussienne. Habituellement, lorsque nous avons deux vecteurs aléatoires dont nous connaissons uniquement la densité jointe, la distribution conditionnelle n'est pas aussi simplement disponible : nous devons effectuer une intégration numérique pour calculer les moments et effectuer des simulations Monte Carlo pour obtenir des réalisations conditionnelles.
- En plus de savoir que la distribution conditionnelle est gaussienne, nous avons des formules explicites pour le vecteur moyen conditionnel et la matrice de covariance.
- Le vecteur moyen conditionnel est égal au vecteur moyen inconditionnel m_2 plus une correction/ajustement $\Sigma_{1,2}^\top \Sigma_1^{-1} (y_1 - m_1)$ avec
 - $\Sigma_{1,2}$: les covariances entre Y_1 et Y_2 .
 - $y_1 - m_1$: l'écart de Y_1 par rapport à sa moyenne.

- Si $Y_1 = m_1$, alors nous ne mettons pas à jour la moyenne de Y_2 . Si Y_1 et Y_2 sont indépendants, $\Sigma_{1,2} = 0$ et donc la moyenne conditionnelle est égale à la moyenne inconditionnelle. C'est normal, car sous l'indépendance, la distribution conditionnelle est égale à la distribution inconditionnelle.
- La matrice de covariance conditionnelle est égale à la matrice de covariance inconditionnelle Σ_2 , moins une correction/ajustement $\Sigma_{1,2}^\top \Sigma_1^{-1} \Sigma_{1,2}$ qui est une matrice semi-positive définie.
 - Nous diminuons la matrice de covariance en conditionnant.
 - Nous diminuons les incertitudes sur Y_2 en observant Y_1 .
- Encore une fois, si Y_1 et Y_2 sont indépendants, l'ajustement est nul.

Exemple (fait en cours) On considère trois variables aléatoires indépendantes et identiquement distribuées A, B, C de loi $\mathcal{N}(0, 1)$. On pose, pour tout $(t, s) \in [0, 1]^2$, $Y(t, s) = tsA + (s - 2t)B + C$.

1. Montrer soigneusement que Y est un processus gaussien sur $[0, 1]^2$.
2. Quelle est la fonction moyenne de Y ?
3. Quelle est la fonction de covariance de Y ?
4. Quelle est la loi de $Y(1, 1)$?
5. Quelle est la loi du vecteur $(Y(1, 1), Y(0, 1))$ conditionnellement à $Y(1, 0) = 2$.

1. Soient $(t_1, n_1), \dots, (t_n, n_n) \in (0, 1)^2$ et $\alpha_1, \dots, \alpha_n \in \mathbb{R}$

$$\begin{aligned} & \alpha_1 Y(t_1, s_1) + \alpha_2 Y(t_2, s_2) + \dots + \alpha_n Y(t_n, s_n) \\ &= \alpha_1 (t_1 s_1 A + (s_1 - 2t_1)B + C) + \dots + \alpha_n (t_n s_n A + (s_n - 2t_n)B + C) \\ &= (\alpha_1 t_1 s_1 + \dots + \alpha_n t_n s_n)A + (\alpha_1 (s_1 - 2t_1) + \dots + \alpha_n (s_n - 2t_n))B + (\alpha_1 + \dots + \alpha_n)C \end{aligned}$$

est une combinaison linéaire de $\begin{pmatrix} A \\ B \\ C \end{pmatrix}$ qui est un vecteur gaussien car composantes gaussiennes iid.

Donc Y est un processus gaussien.

2. $\mathbb{E}[Y(t, s)] = ts\mathbb{E}[A] + (s - 2t)\mathbb{E}[B] + \mathbb{E}[C] = 0 \times ts + 0 \times (s - 2t) + 0 = 0$

- 3.

$$\begin{aligned} \text{cov}(Y(t_1, s_1), Y(t_2, s_2)) &= \text{cov}(t_1 s_1 A + (s_1 - 2t_1)B + C, t_2 s_2 A + (s_2 - 2t_2)B + C) \\ &= t_1 s_1 t_2 s_2 + (s_1 - 2t_1)(s_2 - 2t_2) + 1 \end{aligned}$$

La fonction de covariance est :

$$K((t_1, s_1), (t_2, s_2)) = t_1 s_1 t_2 s_2 + (s_1 - 2t_1)(s_2 - 2t_2) + 1$$

4. $Y(1, 1) \sim \mathcal{N}(m(1, 1), K((1, 1), (1, 1))) = \mathcal{N}(0, 1 + (1 - 2)(1 - 2) + 1) = \mathcal{N}(0, 3)$

- 5.

$$\begin{pmatrix} Y(1, 0) \\ Y(1, 1) \\ Y(0, 1) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 & 3 & -1 \\ 3 & 3 & 0 \\ -1 & 0 & 2 \end{pmatrix}\right)$$

Par exemple $K((1, 0), (1, 0)) = 5$, $K((1, 0), (1, 1)) = 3$, $K((1, 0), (0, 1)) = -1$, ...

Donc

$$E\left(\begin{pmatrix} Y(1, 1) \\ Y(0, 1) \end{pmatrix} \middle| y(1, 0) = 2\right) = \begin{pmatrix} 3 \\ -1 \end{pmatrix} \times 5^{-1} \times 2 = \begin{pmatrix} 6/5 \\ -2/5 \end{pmatrix}$$

 et

$$\begin{aligned} \text{Cov} \left(\begin{pmatrix} Y(1,1) \\ Y(0,1) \end{pmatrix} \middle| y(1,0) = 2 \right) &= \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} - \begin{pmatrix} 3/5 \\ -1/5 \end{pmatrix} \begin{pmatrix} 3 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 6/5 & -3/5 \\ -3/5 & 9/5 \end{pmatrix} \end{aligned}$$

$$\text{D'où } ((Y(1,1), Y(0,1)) | Y(1,0) = 2) \sim \mathcal{N} \left(\begin{pmatrix} 6/5 \\ -2/5 \end{pmatrix}, \begin{pmatrix} 6/5 & -3/5 \\ -3/5 & 9/5 \end{pmatrix} \right)$$