



Processus gaussien a posteriori pour données fonctionnelles

Résumé

Dans ce chapitre, nous démontrons l'ensemble des résultats que nous utiliserons dans les chapitres suivants. En particulier, nous souhaitons calculer une loi a posteriori dans un modèle où loi a priori et vraisemblance sont des processus gaussiens. Notre méthodologie requiert le calcul de densités de processus gaussiens. Ces densités sont calculées à l'aide de la théorie des espaces de Hilbert à noyau reproduisant. Dans un premier temps, nous commençons par définir les données fonctionnelles et les processus gaussiens, avant d'introduire les concepts dont nous aurons besoin. Dans un second temps, nous démontrons le résultat principal dans le cas d'une seule observation, avant de le généraliser au cas d'observations multiples.

2.1 Les données fonctionnelles

2.1.1 Généralités

L'analyse de données fonctionnelles [102] consiste à mettre en place des méthodes statistiques dans lesquelles les observations sont des fonctions, c'est-à-dire des courbes. Beaucoup de domaines d'application font appel à des courbes et des signaux, comme la spectrométrie ou lors de l'étude de courbes de croissance. Avec le développement actuel du phénotypage où les données sont recueillies en temps continu, de plus en plus d'utilisateurs ont besoin d'outils capables de classer des courbes.

Une variable aléatoire est dite fonctionnelle si ses valeurs sont dans un espace de dimension infinie. Une observation d'une variable fonctionnelle est appelée donnée fonctionnelle. Le plus souvent, une donnée fonctionnelle est définie comme la trajectoire d'un processus stochastique $Y = (Y_t)_{t \in [0, T]}$. Sauf mention explicite, dans toute la suite de cette thèse, nous considérerons uniquement des processus dont les trajectoires appartiennent à l'espace $L^2([0, T])$, espace des fonctions de carré intégrable sur $[0, T]$. Cet espace étant polonais¹, il nous garantit l'existence des probabilités conditionnelles d'après Dudley [35].

En pratique, une donnée fonctionnelle n'est jamais observée continûment, mais en un nombre

1. Un espace polonais est un espace métrique, complet et séparable

fini de temps d'observation. Il est donc possible de résumer chaque courbe comme un vecteur et ainsi de transformer le problème en dimension finie, mais cette approche néglige l'aspect fonctionnel. L'utilisation de modèles fonctionnels présente l'avantage de pouvoir prendre en compte la corrélation temporelle des données. Une des spécificités des données fonctionnelles est également la possibilité d'utiliser l'information contenue dans les dérivées. Certains auteurs [19] ont montré qu'elles pouvaient révéler des caractéristiques importantes des jeux de données.

2.1.2 Les processus gaussiens

Les processus gaussiens [104, 114] jouent un rôle crucial dans la théorie des processus stochastiques car :

1. beaucoup de processus stochastiques peuvent être approchés par des processus gaussiens,
2. beaucoup de calculs sont facilités dans le cadre des processus gaussiens.

Rappelons que les processus gaussiens sont la généralisation des lois normales multivariées aux espaces de dimension infinie et qu'un processus est gaussien si, et seulement si, toutes ses lois fini-dimensionnelles sont des lois normales multivariées. Un processus gaussien de loi $P_{m,K}$ est défini par le biais de deux fonctions qui sont sa fonction moyenne m et sa fonction de covariance K , qui est symétrique et définie positive (voir annexe A.3). Rappelons qu'une fonction de deux variables K définie sur $[0, T] \times [0, T]$ est dite :

- (i) définie positive si pour tout entier non nul $n \in \mathbb{N} \setminus \{0\}$, tous $t_1, \dots, t_n \in [0, T]$ et $a_1, \dots, a_n \in \mathbb{R}$,

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j K(t_i, t_j) \geq 0, \quad (2.1.1)$$

- (ii) symétrique si pour tous $s, t \in [0, T]$,

$$K(s, t) = K(t, s), \quad (2.1.2)$$

Dans la littérature, les processus gaussiens sont souvent notés $GP(m, K)$, mais pour des raisons de commodité, nous les noterons $P_{m,K}$ dans toute la suite. La fonction de covariance influe sur la régularité des trajectoires du processus. Le lecteur pourra se référer par exemple à Cramér & Leadbetter [22] ou encore Shi & Choi [114] concernant le choix de cette dernière.

2.2 Densité d'un processus gaussien

2.2.1 Objectif

Notons $Y = (Y_t)_{t \in [0, T]}$ un processus gaussien quelconque et P_Y sa mesure de probabilité associée (voir annexe A.3). Notre objectif est de trouver une mesure de référence P pour laquelle on puisse exprimer la dérivée de Radon-Nikodym $\frac{dP_Y}{dP}$, c'est-à-dire une expression de la densité de probabilité de P_Y par rapport à P .

Un problème récurrent en traitement du signal, et qui rejoint notre but initial, est de pouvoir extraire un signal depuis une observation bruitée. Formellement, cela revient à considérer deux processus stochastiques $X = (X_t)_{t \in [0, T]}$ et $\epsilon = (\epsilon_t)_{t \in [0, T]}$, le premier étant appelé "signal" et le second "bruit". En notant $Y = X + \epsilon$ le processus observé, l'un des objectifs en traitement du signal est de déterminer $\frac{dP_{X+\epsilon}}{dP_\epsilon}$.

Ce problème est en particulier équivalent au problème de test d'hypothèse suivant :

(H0) $Y = \epsilon$,

(H1) $Y = X + \epsilon$,

dans lequel le rapport de vraisemblance est égal à $\frac{dP_{X+\epsilon}}{dP_\epsilon}(Y)$, que nous noterons $L(Y)$ dans toute la suite et que nous appellerons processus de vraisemblance².

Pour la bonne compréhension de la suite, le lecteur peut se référer à l'annexe A.4 pour une brève définition d'une intégrale d'un processus.

2.2.2 Travaux précurseurs pour un bruit blanc gaussien

Parmi les travaux précurseurs, Price [100, 101] est le pionnier dans le cas où ϵ est un bruit blanc gaussien $P_{0,R}$, pour lequel la fonction de covariance R est donnée par :

$$R(s, t) = \delta_s(t).$$

Supposons de plus $X \sim P_{0,K}$ avec X et ϵ indépendants. Historiquement, à condition que :

1. la fonction $K(\bullet, \bullet)$ soit continue sur $[0, T] \times [0, T]$, (2.2.1)

2. $\int_0^T K(t, t) dt < \infty$, (2.2.2)

Price a montré que l'on pouvait écrire :

$$L(Y) = \frac{1}{\sqrt{B(Y)}} e^{\frac{1}{2} \int_0^T \int_0^T H(s, t) Y_s Y_t ds dt},$$

où $B(Y)$ est un terme déterministe de biais et $H(\bullet, \bullet)$ est une fonction appelée résolvant de Fredholm de K , solution de l'équation intégrale :

$$H(s, t) + \int_0^T H(\tau, t) K(s, \tau) d\tau = K(s, t).$$

Dans le courant des années 1960, Stratanovich & Sosulin [119–121] et Schweppe [111] ont voulu généraliser cette formule en introduisant un processus stochastique noté \widehat{X}_1 , qui est une fonction du passé de Y , c'est-à-dire que $\widehat{X}_1(t)$ dépend uniquement des valeurs $\{Y_s, s < t\}$. Cependant, le calcul de \widehat{X}_1 n'est la plupart du temps pas réalisable hormis par approximations, et ces approches souffrent d'un problème pratique.

Il est possible de s'affranchir de la condition d'indépendance entre le signal et le bruit. Dans ce cas, on écrira $Cov(Y_s, Y_t) = \delta_s(t) + K(s, t)$, où :

$$K(s, t) = \mathbb{E}(X_s X_t) + \mathbb{E}(X_s \epsilon_t) + \mathbb{E}(\epsilon_s X_t).$$

Dans ce cas, la fonction K reste symétrique mais n'est plus forcément définie positive au sens de l'équation (2.1.1). Pour pallier ce problème, il nous faut supposer que :

1. la fonction $(s, t) \mapsto Cov(Y_s, Y_t)$ est définie positive sur $[0, T] \times [0, T]$, (2.2.3)

2. $\int_0^T \int_0^T K^2(s, t) ds dt < \infty$. (2.2.4)

2. Il s'agit bien d'un processus, en tant que dérivée de Radon-Nikodym de deux processus.

Remarquons que les conditions (2.2.1) et (2.2.2) ci-dessus impliquent les conditions (2.2.3) et (2.2.4). Sous ces conditions, une nouvelle expression du rapport de vraisemblance a été donnée par Shepp [113] en 1966, sous la forme :

$$L(Y) = \frac{1}{\sqrt{C(Y)}} e^{J(Y)} e^{\int_0^T \int_0^T H(s,t) K(s,t) ds dt},$$

où $C(Y)$ est une fonction déterministe et $J(Y)$ fait intervenir l'intégrale double de Wiener centrée. Dans un rapport de 1969, Kailath [56] a montré l'expression suivante :

$$L(Y) = e^{\int_0^T \widehat{X}_1(t) Y_t dt - \frac{1}{2} \int_0^T \widehat{X}_1(t)^2 dt},$$

où \int désigne l'intégrale d'Itô et $\widehat{X}_1(t) = E(X_t | \{Y_s, s < t\})$. Des détails sur l'intégrale d'Itô se trouvent par exemple dans le livre de Doob [31].

Toutes ces formules ne sont pas facilement explicites et restent donc très peu utilisées en pratique. Remarquons simplement que dans cette dernière formule, si X est déterministe et égal à la fonction m , alors $\widehat{X}_1 = m$ et on retrouve la formule du rapport de vraisemblance pour le problème de test suivant :

$$(H0) Y = \epsilon,$$

$$(H1) Y = m + \epsilon.$$

2.2.3 Travaux précurseurs pour un bruit gaussien quelconque

Supposons dans cette sous-section que ϵ est le processus gaussien quelconque $P_{0,R}$ et que X est déterministe et égal à la fonction continue m . Dans la littérature, une première approche due à Grenander [46] a été de considérer les développements de Karhunen-Loève. A condition que la fonction R soit continue sur $[0, T] \times [0, T]$ et que $\int_0^T \int_0^T R^2(s, t) ds dt < \infty$, on sait que l'on peut trouver des valeurs propres $(\lambda_k)_{k \geq 1}$ et des fonctions propres $(\psi_k)_{k \geq 1}$ vérifiant pour tout entier $k \geq 1$:

$$\int_0^T R(s, t) \psi_k(s) ds = \lambda_k \psi_k(t),$$

et on a alors la décomposition suivante :

$$\epsilon_t = \sum_{k=1}^{\infty} \epsilon_k \psi_k(t),$$

$$\epsilon_k = \int_0^T \epsilon_t \psi_k(t) dt.$$

La convergence ci-dessus est une convergence en moyenne quadratique pour chaque $t \in [0, T]$. Les coefficients ϵ_k sont des variables aléatoires non corrélées, de moyenne nulle et de variance λ_k . Il est alors possible de déduire :

$$L(Y) = e^{\sum_{k=1}^{\infty} \frac{m_k Y_k}{\lambda_k} - \frac{1}{2} \sum_{k=1}^{\infty} \frac{m_k^2}{\lambda_k}},$$

où $m_k = \int_0^T m(t) \psi_k(t) dt$ et $Y_k = \int_0^T Y_t \psi_k(t) dt$. En posant de plus $a(t) = \sum_{k=1}^{\infty} \frac{m_k}{\lambda_k} \psi_k(t)$, l'expression peut également s'écrire :

$$L(Y) = e^{\int_0^T a(t) Y_t dt - \frac{1}{2} \int_0^T a(t) m(t) dt}.$$

En revanche, même si la fonction a est solution d'une équation difficile, celle-ci est souvent peu explicitable. La théorie des espaces de Hilbert à noyau reproduisant (RKHS) permet de contourner de nombreux problèmes. Nous allons voir en quoi cette théorie offre un cadre de travail plus simple pour expliciter des densités de processus gaussiens.

2.2.4 Espaces de Hilbert à noyau reproduisant

Commençons par donner quelques définitions et propriétés générales sur les espaces RKHS, abréviation de *Reproducing Kernel Hilbert Space* en anglais. Rappelons avant tout qu'un espace vectoriel H , muni d'un produit scalaire $(\bullet, \bullet)_H$, est un espace de Hilbert si c'est de plus un espace complet pour la norme induite par le produit scalaire, c'est-à-dire la norme définie par $\|u\|_H = \sqrt{(u, u)_H}$.

Un théorème énoncé par Moore [85] en 1935 et démontré, entre autres, par Aronszajn [5] en 1950 justifie l'existence des espaces RKHS et peut se résumer de la façon suivante :

Théorème 2.1 (Moore, 1935)

Soit K la fonction de covariance d'un processus gaussien sur $[0, T]$. Alors il existe un unique espace de Hilbert, que l'on note $H(K)$ et que l'on appelle espace de Hilbert à noyau reproduisant et de noyau K , défini comme l'espace des fonctions réelles f sur $[0, T]$ vérifiant :

- (i) $\forall s \in [0, T], K(\bullet, s) \in H(K)$, où $K(\bullet, s)$ est la fonction $s' \mapsto K(s', s)$,
- (ii) $\forall t \in [0, T], \forall f \in H(K), f(t) = (f, K(\bullet, t))_K$.

$(\bullet, \bullet)_K$ désigne le produit scalaire dans l'espace $H(K)$.

Remarque 2.1

La propriété (ii) est dite propriété reproduisante et justifie le nom de la théorie RKHS.

Une question naturelle est alors la suivante : comment expliciter le produit scalaire ? Pour trouver un produit scalaire, le plus simple reste de le construire. A chaque fonction de covariance étant associé un unique espace RKHS, on peut dans certains cas proposer un produit scalaire et montrer qu'il vérifie les propriétés (i) et (ii) précédentes ; l'espace étant unique, il s'agira alors du produit scalaire de $H(K)$. Kailath, Geesey & Weinert [57] ou encore Weinert [135] proposent des écritures de produits scalaires pour différentes fonctions de covariance. Les auteurs y décrivent également de nombreux espaces RKHS. Précisons qu'il est aussi possible d'approcher numériquement un produit scalaire. Nous le verrons à la fin de cette sous-section.

Revenons à présent à la théorie de traitement du signal. Parzen [96] a démontré le résultat suivant :

Théorème 2.2 (Parzen, 1963)

$\frac{dP_{m,K}}{dP_{0,K}}$ existe si, et seulement si, $m \in H(K)$. Dans ce cas, il est possible de montrer au travers l'utilisation des lois fini-dimensionnelles que l'on aboutit à l'expression suivante :

$$\frac{dP_{m,K}}{dP_{0,K}}(Y) = e^{(Y,m)_K - \frac{1}{2}(m,m)_K},$$

en supposant que K soit une fonction faiblement continue sur $[0, T] \times [0, T]$. Une fonction K est dite faiblement continue sur $[0, T] \times [0, T]$ si :

$$\forall t \in [0, T], K(\bullet, t) \text{ est continue sur } [0, T], \quad (2.2.5a)$$

$$\forall t \in [0, T], \text{ il existe une boule ouverte } S(t) \text{ contenant } t \quad (2.2.5b)$$

$$\text{et une constante } M(t) \text{ telle que pour tout } t' \in S(t), K(t', t') \leq M(t).$$

On remarque dans cette expression que $(Y, m)_K$ définit un produit scalaire entre le processus stochastique Y et une fonction connue de l'espace $H(K)$. La notation $(Y, m)_K$ ne définit donc pas réellement un produit scalaire et cache l'expression d'une intégrale stochastique. Pour définir une telle notation, il est nécessaire d'établir une correspondance entre les éléments de $H(K)$ et ceux d'un autre espace de Hilbert que nous allons définir.

Définition 2.1

Soit $Y = (Y_t)_{t \in [0, T]}$ un processus stochastique. On définit $L_2(Y)$ comme l'espace des variables aléatoires qui sont des combinaisons linéaires finies des variables aléatoires Y_t ou bien des limites en moyenne quadratique de telles combinaisons linéaires.

Intuitivement, on peut interpréter $L_2(Y)$ comme l'espace de toutes les fonctions linéaires du processus Y .

Définition 2.2

Soit $f = (f_t)_{t \in [0, T]}$ une famille de vecteurs d'un espace de Hilbert H . On définit $L_2(f)$ comme l'espace des vecteurs qui sont des combinaisons linéaires finies des vecteurs f_t ou bien des limites de telles combinaisons linéaires.

Une propriété importante reprise par Parzen [94] est la suivante :

Théorème 2.3 (Parzen, 1961)

Soient $Y = (Y_t)_{t \in [0, T]}$ un processus stochastique de fonction de covariance K et $f = (f_t)_{t \in [0, T]}$ une famille de vecteurs d'un espace de Hilbert H . On dit que la famille f est une représentation du processus stochastique Y si pour tous $s, t \in [0, T]$, on a :

$$(f_s, f_t)_H = K(s, t).$$

Il existe alors une congruence ψ de $L_2(f)$ sur $L_2(Y)$ vérifiant pour tout $t \in [0, T]$, $\psi(f_t) = Y_t$, et toute variable aléatoire $U \in L_2(Y)$ peut s'écrire sous la forme $U = \psi(g)$, pour un unique vecteur $g \in L_2(f)$. Cette congruence est un isomorphisme et préserve les produits scalaires :

$$(f_s, f_t)_{L_2(f)} = (\psi(f_s), \psi(f_t))_{L_2(Y)}.$$

En particulier, nous remarquons que $K(s, t) = (K(\bullet, s), K(\bullet, t))_K$ d'après la propriété reproductrice (ii) du théorème 2.1. Ainsi, la famille de fonctions $(K(\bullet, t))_{t \in [0, T]}$ est une représentation du processus stochastique Y . D'après le théorème 2.3, il existe une congruence ψ de $L_2((K(\bullet, t))_{t \in [0, T]})$ sur $L_2(Y)$. De plus, il est connu que $L_2((K(\bullet, t))_{t \in [0, T]}) = H(K)$ d'après Parzen [94]. Ainsi, nous déduisons que la congruence ψ est à valeurs de $H(K)$ dans $L_2(Y)$. Si $g \in H(K)$, alors $(Y, g)_K$, produit scalaire entre un élément $g \in H(K)$ et le processus stochastique Y , est l'image $U \in L_2(Y)$ correspondant à la congruence définie précédemment, c'est-à-dire $U = \psi(g)$. Il s'agit donc d'une variable aléatoire de l'espace $L_2(Y)$. On obtient également les deux propriétés suivantes pour tous réel $t \in [0, T]$ et fonctions $f, g \in H(K)$:

- (i) $(Y, K(\bullet, t))_K = Y_t$,
- (ii) $(f, g)_K = \mathbb{E}((Y, f)_K(Y, g)_K)$,

Un autre théorème intéressant est le théorème de représentation intégrale.

Définition 2.3 (Famille aléatoire orthogonale)

Soit (Q, \mathcal{B}, μ) un espace mesuré. On dit qu'une famille de variables aléatoires $(Z(B))_{B \in \mathcal{B}}$ de Q est une famille aléatoire orthogonale de noyau de covariance μ si pour tous $B_1, B_2 \in \mathcal{B}$, on a :

$$\mathbb{E}(Z(B_1)Z(B_2)) = \mu(B_1 \cap B_2).$$

Définition 2.4

Soit (Q, \mathcal{B}, μ) un espace mesuré. On définit $L_2(Q)$ comme l'espace de Hilbert de toutes les fonctions f \mathcal{B} -mesurables définies sur Q , et vérifiant de plus :

$$\int_Q f^2 d\mu < \infty.$$

Le produit scalaire sur l'espace $L_2(Q)$ est défini par :

$$(f, g)_{L_2(Q)} = \int_Q f g d\mu.$$

Théorème 2.4 (Parzen, 1961)

Soit $Y = (Y_t)_{t \in [0, T]}$ un processus stochastique de fonction de covariance K . S'il existe un espace mesuré (Q, \mathcal{B}, μ) et une famille de fonctions $f = (f_t)_{t \in [0, T]}$ de Q , vérifiant de plus :

$$K(s, t) = \int_Q f_s f_t d\mu,$$

alors la famille f est une représentation du processus stochastique Y .

Si de plus, la famille f engendre l'espace $L_2(Q)$, alors il existe une famille aléatoire orthogonale $(Z(B))_{B \in \mathcal{B}}$ de Q et de noyau de covariance μ , telle que :

$$Y_t = \int_Q f_t dZ,$$

et toute variable aléatoire $U \in L_2(Y)$ peut s'écrire sous la forme :

$$U = \int_Q g dZ,$$

pour un unique vecteur $g \in L_2(Q)$.

D'après ce théorème, nous voyons bien que l'expression $(Y, g)_K$ est une variable aléatoire pouvant s'écrire comme une intégrale stochastique. Terminons cette sous-section en citant deux résultats importants, correspondant aux théorèmes 6E et 9B de Parzen [93] :

Définition 2.5 (Fonction non singulière)

Une fonction K définie sur $[0, T] \times [0, T]$ est dite non singulière sur tout sous-ensemble de $[0, T] \times [0, T]$ si quels que soient l'entier naturel non nul $n \in \mathbb{N}$ et les nombres réels $t_1, \dots, t_n \in [0, T]$, la matrice $[K(t_i, t_j)]_{1 \leq i, j \leq n}$ est inversible.

Théorème 2.5 (Parzen, 1959)

Soit K une fonction de covariance d'un processus stochastique. Supposons que K est une fonction faiblement continue sur $[0, T] \times [0, T]$ et qu'elle est non singulière sur tout sous-ensemble de

$[0, T] \times [0, T]$. Nous avons alors pour toutes fonctions $f, g \in H(K)$ et toute suite $(t_i)_{i \geq 1}$ de $[0, T]$ dense dans $[0, T]$:

$$\lim_{n \rightarrow \infty} f_n^\top K_n^{-1} g_n = (f, g)_K,$$

où $f_n = (f(t_1), \dots, f(t_n))$, $K_n = [K(t_i, t_j)]_{1 \leq i, j \leq n}$ et $g_n = (g(t_1), \dots, g(t_n))$.

Théorème 2.6 (Parzen, 1959)

Soit $Y = (Y_t)_{t \in [0, T]}$ un processus stochastique de fonction moyenne nulle et de fonction de covariance K . Supposons que K est une fonction faiblement continue sur $[0, T] \times [0, T]$ et qu'elle est non singulière sur tout sous-ensemble de $[0, T] \times [0, T]$. Nous avons alors pour toute fonction $g \in H(K)$ et toute suite $(t_i)_{i \geq 1}$ de $[0, T]$ dense dans $[0, T]$:

$$\lim_{n \rightarrow \infty} Y_n^\top K_n^{-1} g_n = (Y, g)_K = \psi(g),$$

où $Y_n = (Y(t_1), \dots, Y(t_n))$, $K_n = [K(t_i, t_j)]_{1 \leq i, j \leq n}$ et $g_n = (g(t_1), \dots, g(t_n))$ et ψ est la congruence définie précédemment. La convergence existe en moyenne quadratique mais aussi presque sûrement.

Ces théorèmes sont très importants en pratique car ils fournissent une méthode d'approximation des produits scalaires. A défaut de pouvoir expliciter le produit scalaire, la méthode d'approximation la plus courante consiste à approcher $(f, g)_K$ par $f_n^\top K_n^{-1} g_n$, où f_n , K_n et g_n sont des versions discrétisées des fonctions f , K et g en des temps d'observations donnés. En 1961, Parzen [94] propose une méthode itérative pour estimer les produits scalaires, lorsque la fonction de covariance est connue analytiquement ou seulement numériquement. Sa méthode consiste en une première estimation H_0 , puis à construire une suite de fonctions H_n par récurrence. Cette suite H_n est alors telle que :

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left(\left| (Y, g)_K - \int_0^T H_n(t) X_t dt \right|^2 \right) = 0,$$

ce qui permet d'obtenir une estimation de $(Y, g)_K$. En 1965, Weiner [134] développe une autre méthode itérative, encore utilisée à l'heure actuelle. Cependant, la complexité de la méthode est importante. Finalement, c'est plus récemment en 2009 que Oya et al. [92] proposent une nouvelle approche pour évaluer numériquement un produit scalaire. Dans un premier temps, leur méthode requiert de résoudre un problème aux valeurs propres "généralisé", puis à faire usage des valeurs et vecteurs propres trouvés afin d'obtenir l'estimation souhaitée.

En comparaison aux développements de Karhunen-Loève et aux travaux précurseurs, la théorie RKHS offre donc un cadre de travail plus souple tout en faisant intervenir des quantités facilement mesurables numériquement.

2.3 Processus gaussien a posteriori

2.3.1 Cas d'une seule observation

Dans toute cette section, nous adoptons les notations suivantes :

$t^n = (t_1, \dots, t_n)$ lorsque (t_1, \dots, t_n) est un vecteur,

$X(t^n) = (X(t_1), \dots, X(t_n))$ lorsque X est une fonction d'une seule variable,

$X(t^n) = (X_{t_1}, \dots, X_{t_n})$ lorsque X est un processus stochastique.

donc

$$\text{rg}(X + Y) \geq \text{rg}Y = r.$$

Finalement, considérons le cas où Y est une matrice symétrique et positive quelconque. Comme elle est positive, il existe une matrice orthogonale P telle que $P^\top Y P$ est une matrice diagonale. Ainsi, comme P^\top et P sont inversibles, on obtient :

$$\text{rg}(X + Y) = \text{rg}(P^\top (X + Y) P) = \text{rg}(P^\top X P + P^\top Y P).$$

Comme la matrice $P^\top X P$ est à nouveau positive et que la matrice $P^\top Y P$ est diagonale, on peut faire usage de ce qui précède.

En conclusion, $\text{rg}[(W + C)(t_i, t_j)]_{1 \leq i, j \leq n} \geq n$ donc $[(W + C)(t_i, t_j)]_{1 \leq i, j \leq n}$ est inversible. ■

Théorème 2.7

Soient $X \sim P_{m, W}$ et $\epsilon \sim P_{0, C}$. Supposons que X et ϵ sont indépendants et notons $Y = X + \epsilon$. Enfin, supposons que les trajectoires de Y sont continues p.s.. Alors, la loi a posteriori $X|Y$ est également un processus gaussien P_{m^*, W^*} , donné par :

$$m^*(t) = m(t) + (W(\bullet, t), Y - m)_{W+C}, \quad (2.3.1a)$$

$$W^*(t, t') = W(t, t') - (W(\bullet, t), W(\bullet, t'))_{W+C}, \quad (2.3.1b)$$

où $(\bullet, \bullet)_{W+C}$ désigne le produit scalaire dans l'espace $H(W + C)$.

Remarque 2.2

Supposer Y à avoir des trajectoires continues presque sûrement n'est pas une hypothèse restrictive en soi. En effet, c'est le cas des processus Brownien, d'Ornstein-Uhlenbeck et de nombreux autres processus intéressants en pratique.

Remarque 2.3

Lorsque nous avons débuté ce travail de thèse, nous ne connaissions pas l'existence de résultats similaires [21, 34, 128]. Cependant, la preuve ne nous ayant pas paru triviale, nous avons pris la peine de démontrer notre théorème ex nihilo. Notons toutefois que, par chance, notre théorème diffère des résultats existants, en ce sens que :

- Cox [21] se place directement dans un espace de Banach séparable et Van der Vaart [128] dans un espace de Hilbert, alors que nous ne faisons aucune hypothèse a priori sur l'espace des trajectoires de nos processus,
- Driscoll [34] ne considère pas de répétitions, se plaçant uniquement dans une optique de traitement du signal.

PREUVE. Fixons $t^n = (t_1, \dots, t_n) \in \mathcal{T}^n$. Pour commencer, cherchons la loi conditionnelle de :

$$X(t^n)|Y. \quad (2.3.2)$$

La loi de (2.3.2) est entièrement caractérisée par sa fonction caractéristique, c'est-à-dire la fonction $\psi : u \mapsto \mathbb{E}(\exp(iu^\top X(t^n))|Y)$, où $u = (u_1, \dots, u_n) \in \mathbb{R}^n$. On écrit $\mathbb{E}(\bullet|Y) = \mathbb{E}(\bullet|\mathcal{B}_Y)$ avec $\mathcal{B}_Y = \sigma(Y_t, t \in \mathcal{T}) = \sigma(Y)$.

Soit $\tau^m = (\tau_1, \dots, \tau_m) \in \mathcal{T}^m$, où $(\tau_m)_{m \geq 1}$ est une suite quelconque de \mathcal{T} qui est dense dans \mathcal{T} . D'après le résultat de O'Hagan [91], il est connu que :

$$X(t^n)|Y(\tau^m) \sim \mathcal{N}(m_m^*(t^n), W_m^*(t^n, t^n)), \quad (2.3.3)$$

avec :

$$m_m^*(t) = m(t) + W_m(t)^\top (W_m + C_m)^{-1} (Y(\tau^m) - m(\tau^m)), \quad (2.3.4a)$$

$$W_m^*(t, t') = W(t, t') - W_m(t)^\top (W_m + C_m)^{-1} W_m(t'), \quad (2.3.4b)$$

où nous écrivons $W_m(t) = (W(\tau_1, t), \dots, W(\tau_m, t))$, $W_m = [W(\tau_i, \tau_j)]_{1 \leq i, j \leq m}$ et de même $C_m = [C(\tau_i, \tau_j)]_{1 \leq i, j \leq m}$. En conséquence, comme la loi de (2.3.3) est une normale multivariée, nous pouvons calculer sa fonction caractéristique comme suit :

$$\psi_m(u) = \mathbb{E}(\exp(iu^\top X(t^n)) | Y(\tau^m)) = \exp\left(iu^\top m_m^*(t^n) - \frac{1}{2}u^\top W_m^*(t^n, t^n)u\right). \quad (2.3.5)$$

A partir de maintenant, étudions la limite $\lim_{m \rightarrow \infty} \psi_m(u)$ dans l'équation (2.3.5) et montrons qu'elle est égale à $\psi(u)$ p.s..

D'une part, d'après les résultats de la théorie RKHS, nous savons que $W_m(t)^\top (W_m + C_m)^{-1} W_m(t')$ converge lorsque $m \rightarrow \infty$ si, et seulement si, $W(\bullet, t) \in H_{W+C}$ et $W(\bullet, t') \in H_{W+C}$. D'après le théorème 12 de Berlinet et al. [6], nous avons toujours $H_W \subset H_{W+C}$, car $(W + C) - W = C$ est une fonction de covariance. Ainsi, comme pour chaque $t \in \mathcal{T}$ on a $W(\bullet, t) \in H_W$, on obtient $W(\bullet, t) \in H_{W+C}$. D'après le lemme 2.1 et le théorème 2.5,

$$W_m(t)^\top (W_m + C_m)^{-1} W_m(t') \xrightarrow{m \rightarrow \infty} (W(\bullet, t), W(\bullet, t'))_{W+C}.$$

Donc $W_m^*(t^n, t^n)$ converge vers $W^*(t^n, t^n)$ lorsque $m \rightarrow \infty$, et qui est donné par l'équation (2.3.1).

D'autre part, étudions à présent la limite $\lim_{m \rightarrow \infty} W_m(t)^\top (W_m + C_m)^{-1} (Y(\tau^m) - m(\tau^m))$. En premier lieu, $Y - m \sim P_{0, W+C}$. Il est facile de le vérifier car nous pouvons écrire pour chaque entier non nul n et tout $t^n = (t_1, \dots, t_n) \in \mathcal{T}^n$, $((Y - m)(t_1), \dots, (Y - m)(t_n)) = (X(t^n) + \epsilon(t^n) - m(t^n))$, qui est une normale multivariée $\mathcal{N}(0, [(W + C)(t_i, t_j)]_{1 \leq i, j \leq n})$ en raison de l'indépendance entre X et ϵ .

En faisant usage à nouveau du lemme 2.1 et du théorème 2.6, nous savons que $\phi_m = W_m(t)^\top (W_m + C_m)^{-1} (Y(\tau^m) - m(\tau^m))$ converge vers $\phi(W(\bullet, t))$, où ϕ désigne la congruence de H_{W+C} sur $L_2((Y_t - m_t)_{t \in \mathcal{T}})$ telle que $\phi((W + C)(\bullet, t)) = Y_t - m_t$. Cette convergence existe en tant que limite en moyenne quadratique mais aussi presque sûrement, ce qui signifie que pour presque toute réalisation de Y , on a $\phi_m \xrightarrow{m \rightarrow \infty} \phi(W(\bullet, t))$.

Ecrivons maintenant $(W(\bullet, t), Y - m)_{W+C}$ au lieu de $\phi(W(\bullet, t))$. Alors $m_m^*(t^n) \xrightarrow{m \rightarrow \infty} m^*(t^n)$ p. s. et

$$\mathbb{E}(\exp(iu^\top X(t^n)) | Y(\tau^m)) \xrightarrow{m \rightarrow \infty} \exp\left(iu^\top m^*(t^n) - \frac{1}{2}u^\top W^*(t^n, t^n)u\right) \text{ p.s..} \quad (2.3.6)$$

Cela conclut la première partie de la preuve. Si nous montrons maintenant que :

$$\mathbb{E}(\exp(iu^\top X(t^n)) | Y(\tau^m)) \xrightarrow{m \rightarrow \infty} \mathbb{E}(\exp(iu^\top X(t^n)) | Y) \text{ p.s.,} \quad (2.3.7)$$

nous déduisons que la loi de $X(t^n) | Y$ est une normale multivariée, de moyenne $m^*(t^n)$ et de covariance $W^*(t^n, t^n)$ et ainsi la loi de $X | Y$ sera le processus gaussien P_{m^*, W^*} . En effet, nous savons d'après le théorème d'extension de Kolmogorov que le processus stochastique est entièrement défini par la famille de ses lois fini-dimensionnelles.

Soient $Z_m = \mathbb{E}(\exp(iu^\top X(t^n))|Y(\tau^m))$ et $Z = \mathbb{E}(\exp(iu^\top X(t^n))|Y)$. Z est intégrable et, si nous écrivons $\mathcal{B}_m = \sigma(Y(\tau_1), \dots, Y(\tau_m))$, Z_m est une martingale par rapport à \mathcal{B}_m d'après la suite d'égalités suivantes :

$$\begin{aligned} \forall k \geq m, \quad \mathbb{E}(Z_k|\mathcal{B}_m) &= \mathbb{E}(\mathbb{E}(\exp(iu^\top X(t^n))|\mathcal{B}_k)|\mathcal{B}_m), \\ &= \mathbb{E}(\exp(iu^\top X(t^n))|\mathcal{B}_m), \\ &= Z_m. \end{aligned} \tag{2.3.8}$$

De plus, Z est une martingale fermée dans le sens où :

$$\begin{aligned} \mathbb{E}(Z|\mathcal{B}_m) &= \mathbb{E}(\mathbb{E}(\exp(iu^\top X(t^n))|Y)|\mathcal{B}_m), \\ &= \mathbb{E}(\exp(iu^\top X(t^n))|\mathcal{B}_m), \\ &= Z_m, \end{aligned} \tag{2.3.9}$$

car $\mathcal{B}_m \subset \mathcal{B}_Y$. Ainsi, d'après le théorème 10.5.1 de Dudley [35], on obtient :

$$Z_n \xrightarrow[n \rightarrow \infty]{} \mathbb{E}(Z|\mathcal{B}_\infty) \text{ p.s.,}$$

où \mathcal{B}_∞ désigne la plus petite tribu contenant tous les \mathcal{B}_m , c'est-à-dire $\sigma(Y_{\tau_i}, i \in \mathbb{N} \setminus \{0\})$.

Montrons à présent que $\mathcal{B}_\infty = \sigma(Y)$. D'une part, il est clair que $\mathcal{B}_\infty \subset \sigma(Y)$. D'autre part, il nous faut montrer que $\sigma(Y) \subset \mathcal{B}_\infty$. Comme $\sigma(Y)$ est générée par la famille d'ensembles $\{Y_t \in B\}$ pour tout $t \in \mathcal{T}$ et tout $B \in \mathcal{B}_\mathbb{R}$, il suffit de montrer que chaque $\{Y_t \in B\} \in \mathcal{B}_\infty$.

Soit $(\tau_i^*)_{i \geq 1}$ une sous-suite de $(\tau_i)_{i \geq 1}$ telle que :

$$\tau_i^* \xrightarrow[i \rightarrow \infty]{} t.$$

D'après la définition d'une tribu, $Y_{\tau_i^*}$ est mesurable pour la tribu \mathcal{B}_∞ et il en est de même pour $\lim_{i \rightarrow \infty} Y_{\tau_i^*}$. Cependant, Y ayant par hypothèse des trajectoires continues p.s., nous avons $\lim_{i \rightarrow \infty} Y_{\tau_i^*} = Y_t$ donc $Y_t \in \mathcal{B}_\infty$.

En conclusion, les expressions (2.3.6) et (2.3.7) sont identiques. Toutes les convergences ci-dessus étant presque sûres, nous concluons la preuve. ■

2.3.2 Cas de multiples observations

La seconde partie de cette section est consacrée à la généralisation du théorème 2.7 pour des processus gaussiens i.i.d. $\epsilon_1, \dots, \epsilon_n$ centrés et de fonction de covariance C .

Lemme 2.2

Soient X, Y, Z des processus stochastiques, à valeurs d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ dans un espace polonais muni de sa tribu borélienne $(\mathcal{X}, \mathcal{B})$. Supposons Z indépendant de X et de Y (signifiant que la tribu $\sigma(Z)$ est indépendante de $\sigma(X, Y)$). Alors, il existe $N \in \mathcal{A}$ tel que $\mathbb{P}(N) = 0$ et

$$\forall w \notin N, \mathbb{P}_X(\bullet|Y, Z)(w) = \mathbb{P}_X(\bullet|Y)(w),$$

où $\mathbb{P}_X(\bullet|Y, Z)$ et $\mathbb{P}_X(\bullet|Y)$ désignent respectivement la loi conditionnelle de X sachant (Y, Z) et de X sachant Y .

PREUVE. Le fait que l'espace d'arrivée des processus soit polonais assure de l'existence de $\mathbb{P}_X(\bullet|Y, Z)$ et $\mathbb{P}_X(\bullet|Y)$. Comme \mathcal{X} est séparable, nous appliquons la proposition 2.1.4 de Dudley [35], ce qui permet d'affirmer qu'il existe une base d'ouverts $\theta = (\theta_i)_{i \in \mathbb{N}}$ qui génère la tribu : $\sigma(\theta) = \mathcal{B}$.

Ecrivons alors $\tau = \{\theta_{i_1} \cap \dots \cap \theta_{i_k}, k \in \mathbb{N}, \theta_{i_j} \in \theta\}$, qui est un ensemble dénombrable. Puis, écrivons :

$$N_{i_1, \dots, i_k} = \{w \in \Omega : \mathbb{P}_X(\theta_{i_1} \cap \dots \cap \theta_{i_k} | Y, Z)(w) \neq \mathbb{P}_X(\theta_{i_1} \cap \dots \cap \theta_{i_k} | Y)(w)\}.$$

D'après la propriété 9.7(k) de Williams [136], comme $\mathbb{P}_X(A|Y, Z)$ est une version de $\mathbb{E}(\mathbb{1}_X(A)|Y, Z)$, $\mathbb{P}(N_{i_1, \dots, i_k}) = 0$. Ainsi $\mathbb{P}(\cup_{k \in \mathbb{N}} \cup_{i_1, \dots, i_k} N_{i_1, \dots, i_k}) = 0$.

En conséquence, en écrivant $N = \cup_{k \in \mathbb{N}} \cup_{i_1, \dots, i_k} N_{i_1, \dots, i_k}$, on a pour tout $w \notin N$ l'égalité $\mathbb{P}_X(D|Y, Z)(w) = \mathbb{P}_X(D|Y)(w)$ pour chaque $D \in \tau$. Or, deux mesures finies sur un π -système (qui est ici τ) étant égales sur $\sigma(\tau) = \mathcal{B}$, on obtient finalement :

$$\mathbb{P}_X(\bullet|Y, Z)(w) = \mathbb{P}_X(\bullet|Y)(w).$$

■

Ainsi, en notant respectivement $\mathcal{L}(X|Y, Z)$ et $\mathcal{L}(X|Y)$ une version de la loi conditionnelle de X sachant (Y, Z) et de X sachant Y , nous avons l'égalité :

$$\mathcal{L}(X|Y, Z) = \mathcal{L}(X|Y),$$

lorsque $Z \perp (X, Y)$.

Lemme 2.3

Soient $\epsilon_1, \dots, \epsilon_n$ des processus gaussiens indépendants $P_{0, C}$. Alors $(\epsilon_1 - \bar{\epsilon}, \dots, \epsilon_n - \bar{\epsilon})$ et $\bar{\epsilon}$ sont indépendants.

PREUVE. Prouvons dans un premier temps que ce résultat est vrai pour des normales multivariées indépendantes $\epsilon_1, \dots, \epsilon_n$ suivant $\mathcal{N}(0, \tilde{C})$. Le vecteur $(\epsilon_1 - \bar{\epsilon}, \dots, \epsilon_n - \bar{\epsilon}, \bar{\epsilon})$ est une combinaison linéaire des composantes du vecteur $(\epsilon_1, \dots, \epsilon_n)$. Par indépendance, $(\epsilon_1, \dots, \epsilon_n)$ a une loi normale multivariée, et il en est de même pour toute combinaison linéaire des composantes du vecteur, donc $(\epsilon_1 - \bar{\epsilon}, \dots, \epsilon_n - \bar{\epsilon}, \bar{\epsilon})$ a une loi normale multivariée.

En conclusion, une condition nécessaire et suffisante pour prouver que $(\epsilon_1 - \bar{\epsilon}, \dots, \epsilon_n - \bar{\epsilon})$ est indépendant de $\bar{\epsilon}$ est que pour chaque entier $i \in \{1, \dots, n\}$, $\epsilon_i - \bar{\epsilon}$ et $\bar{\epsilon}$ sont non corrélés.

Clairement, $\mathbb{E}(\epsilon_i - \bar{\epsilon}) = 0$ et $\mathbb{E}(\bar{\epsilon}) = 0$. On en déduit ainsi les covariances suivantes :

$$\begin{aligned} \text{Cov}(\epsilon_i - \bar{\epsilon}, \bar{\epsilon}) &= \mathbb{E}((\epsilon_i - \bar{\epsilon})\bar{\epsilon}^\top) \\ &= \mathbb{E}((\epsilon_i - \bar{\epsilon})\bar{\epsilon}^\top) \\ &= \frac{1}{n} \sum_{j=1}^n \mathbb{E}(\epsilon_i \epsilon_j^\top) - \frac{1}{n^2} \sum_{j, k=1}^n \mathbb{E}(\epsilon_j \epsilon_k^\top) \\ &= \frac{1}{n} \tilde{C} - \frac{1}{n^2} \sum_{i=1}^n \tilde{C} \\ &= 0. \end{aligned}$$

Les variables étant non corrélées, nous venons de prouver le lemme 2.3 dans le cas de lois fini-dimensionnelles. A présent, il nous faut généraliser ce résultat à la tribu entière. En réalité, il nous faut montrer que $\sigma(\epsilon_1 - \bar{\epsilon}, \dots, \epsilon_n - \bar{\epsilon})$ et $\sigma(\bar{\epsilon})$ sont des tribus indépendantes. Or, d'après ce qui précède, les lois fini-dimensionnelles de $(\epsilon_1 - \bar{\epsilon}, \dots, \epsilon_n - \bar{\epsilon})$ et $\bar{\epsilon}$ sont indépendantes. Le résultat s'ensuit donc. ■

Corollaire 2.1

Soient X et $\epsilon_1, \dots, \epsilon_n$ des processus gaussiens avec X indépendant de chaque ϵ_i . Pour chaque $i \in \{1, \dots, n\}$, supposons que $\epsilon_i \sim P_{0,C}$. Écrivons de même $Y_i = X + \epsilon_i$, $Y = (Y_1, \dots, Y_n)$ et $\epsilon = (\epsilon_1, \dots, \epsilon_n)$. Alors on a $\mathcal{L}(X|Y) = \mathcal{L}(X|\bar{Y})$.

PREUVE. Il est clair que $\sigma(Y) = \sigma(X + \bar{\epsilon}, \epsilon - \bar{\epsilon})$, où $\epsilon - \bar{\epsilon} = (\epsilon_1 - \bar{\epsilon}, \dots, \epsilon_n - \bar{\epsilon})$.

Comme X est indépendant de ϵ , X est également indépendant de $\epsilon - \bar{\epsilon}$. D'après le lemme 2.3, $\bar{\epsilon}$ est indépendant de $\epsilon - \bar{\epsilon}$. Ainsi, la somme $X + \bar{\epsilon}$ est indépendante de $\epsilon - \bar{\epsilon}$.

Nous déduisons donc notre résultat de la suite d'égalités suivantes et du lemme 2.2 :

$$\begin{aligned} \mathcal{L}(X|Y) &= \mathcal{L}(X|X + \bar{\epsilon}, \epsilon - \bar{\epsilon}) \\ &= \mathcal{L}(X|X + \bar{\epsilon}) \\ &= \mathcal{L}(X|\bar{Y}). \end{aligned}$$

■

Nous sommes maintenant capables de généraliser les résultats du théorème 2.7 au cas de multiples observations.

Théorème 2.8

Soient $X \sim P_{m,W}$ et $\epsilon_1, \dots, \epsilon_n$ des processus gaussiens $P_{0,C}$. Supposons que X et chaque ϵ_i sont indépendants et notons $Y_i = X + \epsilon_i$ pour tout $i \in \{1, \dots, n\}$. Enfin, supposons que les trajectoires de Y sont continues p.s.. Alors, la loi a posteriori $X|Y_1, \dots, Y_n$ est également un processus gaussien P_{m^*,W^*} , donné par :

$$m^*(t) = m(t) + (W(\bullet, t), \bar{Y} - m)_{W + \frac{C}{n}}, \quad (2.3.10a)$$

$$W^*(t, t') = W(t, t') - (W(\bullet, t), W(\bullet, t'))_{W + \frac{C}{n}}, \quad (2.3.10b)$$

où $(\bullet, \bullet)_{W + \frac{C}{n}}$ désigne le produit scalaire dans l'espace $H(W + \frac{C}{n})$.

PREUVE. Observons avant tout que $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = X + \bar{\epsilon}$. Comme les processus ϵ_i sont des processus gaussiens indépendants $P_{0,C}$, nous savons que $\bar{\epsilon} \sim P_{0, \frac{C}{n}}$.

De plus, d'après le corollaire 2.1, on a $\mathcal{L}(X|Y) = \mathcal{L}(X|\bar{Y})$. Ainsi, en appliquant le théorème 2.7 et en remplaçant Y par \bar{Y} et ϵ par $\bar{\epsilon}$, nous obtenons le résultat escompté. ■