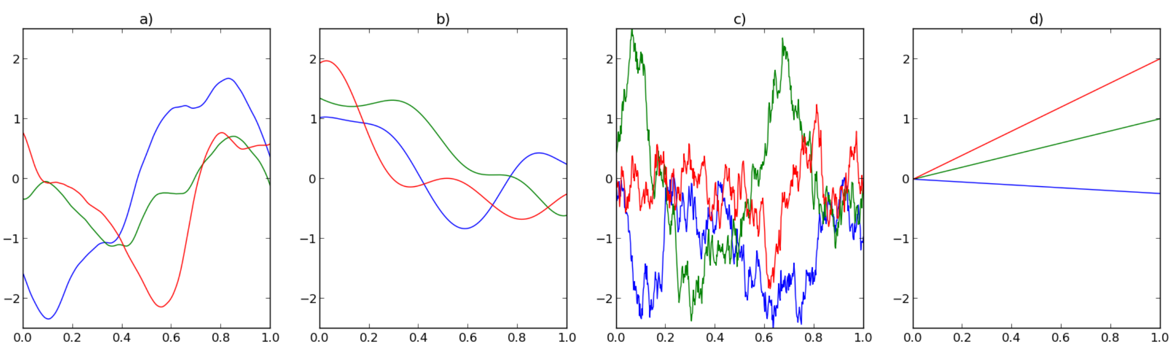
# Examen de modélisation par processus Gaussiens

ENSM-SE, MAFQ – 28 janvier 2014

Les notes de cours sont autorisées pour cet examen.

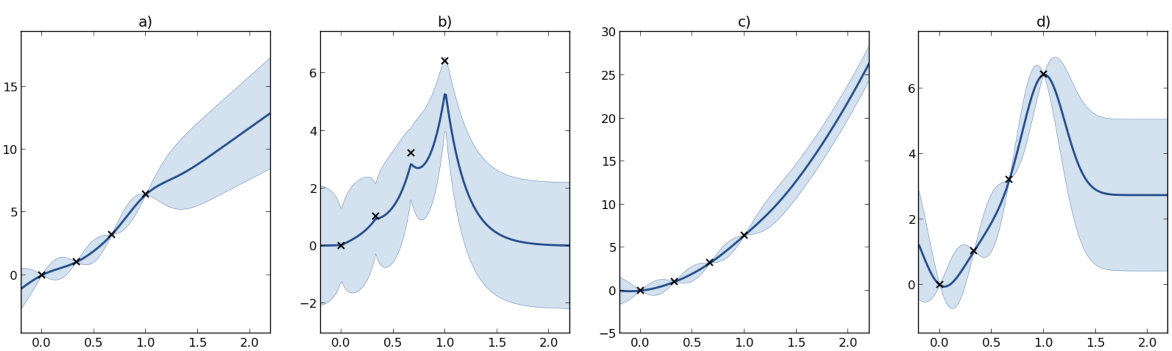
## Exercice 1 :

Les graphiques ci-dessous représentent des trajectoires de processus Gaussiens centrés. Quels noyaux intuitez-vous ? Justifiez brièvement vos réponses.




## Exercice 2 :

Pour les quatre modèles représentés ci-dessous, indiquez en justifiant brièvement :
- le type de krigeage utilisé (simple, ordinaire, universel) ;
- si la tendance est connue ou estimée ;
- le noyau utilisé ;
- la présence de pépite ou de bruit d'observation.




Proposez un modèle qui vous semble mieux adapté aux données.

## Exercice 3 :

L'objectif de cet exercice est de construire un processus périodique et de calculer sa covariance. Pour cela, nous allons considérer la restriction sur un cercle d'un processus Gaussien $Z$ défini sur $\mathbb{R}^2$. Par la suite, nous ferons l'hypothèse que $Z$ est centré et de noyau gaussien isotrope.

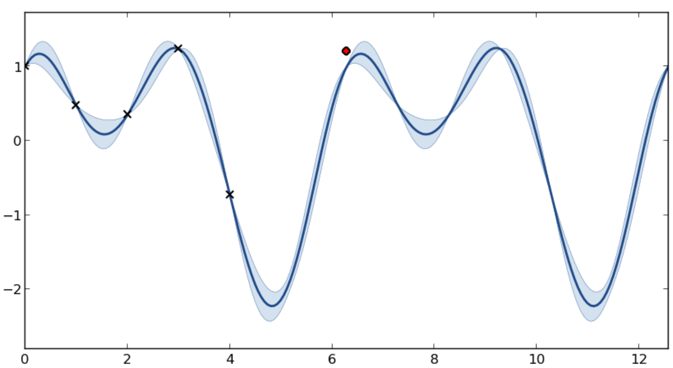
**1.** Donnez l'expression de la covariance de $Z$ en faisant apparaître les paramètres de variance et de portée $\sigma^2$ et $\theta$.

Soit $\mathcal{C}$ un cercle centré sur $(0,0)$ et de rayon 1. On paramétrise $\mathcal{C}$ par $(\cos(t), \sin(t)),\ t \in \mathbb{R}$. Soit $Z_\mathcal{C}$ la restriction de $Z$ à $\mathcal{C}$ : $\forall t \in [0, 2\pi]$, $Z_\mathcal{C}(t) = Z(\cos(t), \sin(t))$.

**2.** $Z_\mathcal{C}$ est il un processus Gaussien ? Est-il centré ? Calculez sa moyenne et son noyau de covariance $k_\mathcal{C}$. Le processus obtenu est il stationnaire ?

**3.** Soit $Y$ un processus Gaussien centré indexé par $\mathbb{R}$ de covariance $k_\mathcal{C}$. Pour une valeur $t$ donnée, montrez que $Y(t) = Y(t + 2k\pi)$ avec probabilité 1 (pour cela, il suffira de montrer que la variable aléatoire $Y(t) - Y(t + 2k\pi)$ est d'espérance et de variance nulle).

**4.** La figure ci-dessous montre un modèle obtenu avec 5 points d'observations $X = (0, 1, 2, 3, 4)$ pour les observations suivantes $Y = (1.00, 0.47, 0.35, 1.24, -0.73)$. Que remarquez-vous de particulier et comment l'expliquez vous ?




**5.** On suppose maintenant que l'on dispose d'une observations supplémentaire $y = 1.2$ située en $x = 2\pi$ (point rouge sur le graphique). Quelle est alors la vraisemblance des observations (la valeur numérique veut être obtenue sans calcul) ?

**6.** L'observation supplémentaire rend la matrice de covariance non-inversible. Quelle stratégie préconisez-vous pour construire le modèle.

**7.** Que peut on faire si on souhaite rendre compte que la fonction modélisée n'est pas purement périodique mais qu'il n'y a pas de bruit d'observation ou de pépite ? Cela améliore-t-il le conditionnement de la matrice de covariance ?

**Questions bonus :** Proposez une autre méthode permettant de construire des processus périodiques. Le processus obtenu est il continu ?

## Exercice 4 :

Dans cet exercice, nous étudierons comment prendre en compte dans les modèles des observations de la dérivée de la fonction à approximer.

Soit $Z$ un processus gaussien centré indexé par $\mathbb{R}$ de noyau gaussien $k_{Z,Z} = \exp(-(x-y)^2)$. On notera $Z'(x) = \frac{\mathrm{d}Z(x)}{\mathrm{d}x}$. Nous admettrons que $Z'(x)$ est bien défini et que le calcul de la dérivée commute avec la covariance : $\mathrm{cov}\left(\frac{\mathrm{d}Z(x)}{\mathrm{d}x}, Y\right) = \frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{cov}(Z(x), Y))$.

**1.** En utilisant la propriété précédente, donnez l'expression de la cross-covariance $k_{Z',Z}(x,y) = \mathrm{cov}(Z'(x), Z(y))$. Le graphe de la fonction obtenue est représenté sur la Figure 1 (figure du milieu).

**2.** De manière similaire, quelle est la covariance de $k_{Z',Z'}(x,y) = \mathrm{cov}(Z'(x), Z'(y))$. Cette fonction est représentée sur le graphique de droite de la Figure 1.

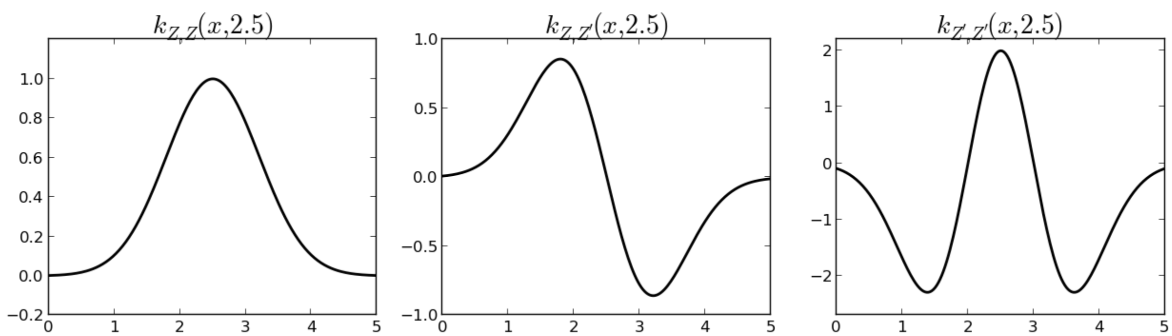



FIGURE 1 – Noyau gaussien et ses dérivées.

Nous allons pour le moment nous intéresser à la prédiction de $Z$ à partir d'une seule observation de $Z'$ en X=7.

**3.** Exprimez le meilleur prédicteur ainsi que la variance de prédiction comme des espérances et variances conditionnelles puis détaillez leur expression en fonction de $k_{Z,Z}$et $k_{Z',Z'}$ et/ou $k_{Z',Z}$.

**4.** Représentez graphiquement l'allure du meilleur prédicteur ainsi que des intervalles de confiance. Vous pourrez vous aider de la variance de prédiction représentée sur le graphe figure 2.a.

On suppose maintenant que l'on dispose de certaines observations de $Z$ pour des points $(X_1, \dots, X_m)$ ainsi que des observations de la dérivée en $(X_{m+1}, \dots, X_n)$.

**5.** Donnez l'expression du meilleur comme une fonction de $k_{Z,Z}$, $k_{Z',Z'}$ et $k_{Z',Z}$.

**6.** La figure 3 représente un modèle avec deux observations de $Z$ et de deux observations de la dérivée. Retrouvez les coordonnées (abscisses) de ces quatre points d'observations (justifiez votre réponse).

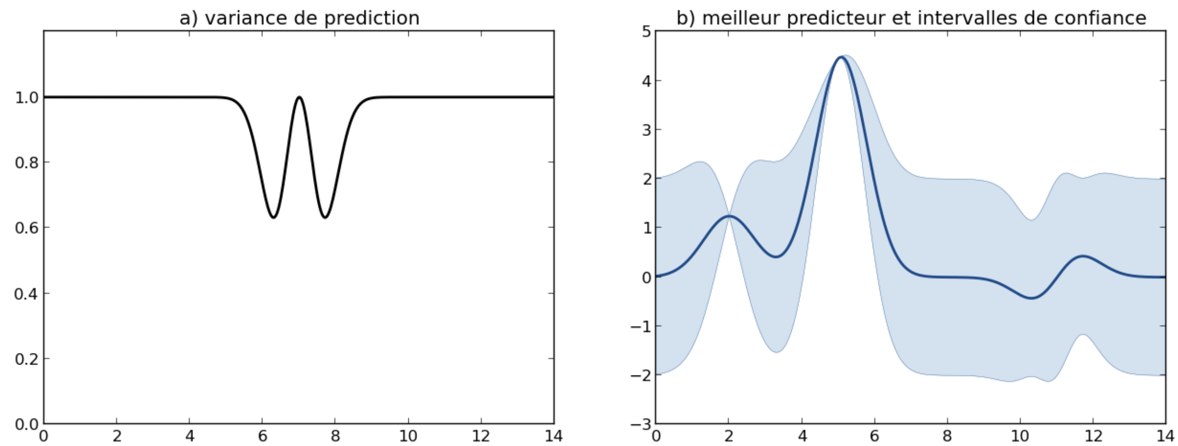



FIGURE 2 – a) Variance de prédiction pour le modèle basé sur une observation de la dérivée en 7. b) Meilleur prédicteur et intervalles de confiance pour deux observations de $Z$ et deux observations de $Z'$.

## Exercice 5 :

Le but de cet exercice est de modéliser les retombées radioactives sur la suisse suite à la catastrophe de Tchernobyl. Pour cela, on dispose de mesures de pluviométrie pour 100 stations disséminées en Suisse pour la date du 8 mai 1986 (soit deux semaines après l'accident) ainsi que d'un modèle permettant de relier la pluviométrie à la radioactivité déposée en surface.

**1.** La première étape de la modélisation consiste à effectuer une transformation de type Box-Cox sur les mesures $Z$ : $\tilde{Z} = \frac{Z^\lambda - 1}{\lambda}$. Pourquoi effectuer cette transformation ?

**2.** Détaillez l'ensemble des étapes que vous mettriez en œuvre pour obtenir en tout point une estimation de la pluviométrie.

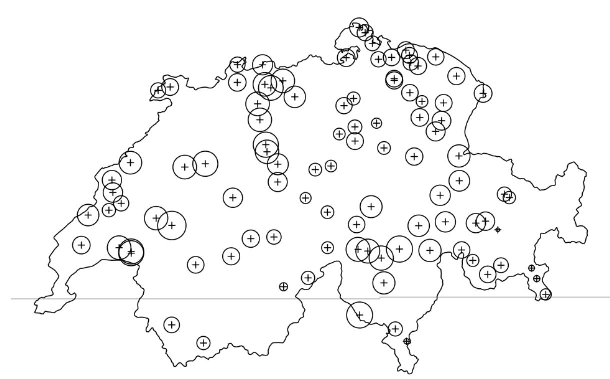


FIGURE 3 – Localisation des stations météo. La taille des cercles est proportionnelle à la hauteur des précipitations.