

# Projet Master HPC-IA Mines ParisTech

2021 - 2022

Miguel Munoz Zuniga

November 30, 2021

## **Problem 1 - Comparison between the moments and maximum likelihood methods**

We consider a random variable  $X$  with probability density function given by :

$$f(x) = \frac{1}{\alpha} e^{-\frac{x-m}{\alpha}} e^{-e^{-\frac{x-m}{\alpha}}}$$

where  $m$  and  $\alpha > 0$  are two unknown parameters.

1. Check that  $f$  is a valid probability density function.
2. Given a random sample  $(X_i)_{i=1,\dots,n}$  of size  $n$ , design and propose a simple estimator for  $m$  and  $\alpha$  based on the method of moments.
3. Given a random sample  $(X_i)_{i=1,\dots,n}$  of size  $n$ , compute the log-likelihood as a function of  $m$  and  $\alpha$ .
4.  $X$  given above is said to follow a Gumbel distribution. Propose one or several numerical illustrations of the convergence of the proposed estimators with respect to the sample size  $n$  and the values of the unknown parameters. In particular, focus on the comparison of the efficiency of the method of moments estimator and the maximum likelihood ones.

## **Problem 2 - Calibration of the hyper-parameters of a gaussian process regressor by maximum likelihood**

Nous souhaitons approcher un modèle numérique, ou encore simulateur, représenté mathématiquement par une fonction  $f$ , dont le temps de calcul est important limitant de fait le nombre d'évaluations de  $f$  réalisables en un temps de calcul limité. La modélisation par processus Gaussien se prête particulièrement bien à ce contexte.

Pour illustrer cette stratégie nous allons approximer/modéliser une fonction simple  $x \rightarrow \sin(4\pi x)$  par un processus Gaussien.

Un processus aléatoire sur  $\mathbb{R}^d$  est une fonction de  $Z : \Omega \times \mathbb{R}^d \rightarrow \mathbb{R}$  telle que pour tout  $\omega \in \Omega$ ,  $Z(\omega, \cdot)$  est une réalisation (toute la fonction est une réalisation) du processus aléatoire  $Z$  et d'un autre côté  $Z(\cdot, x)$  est une variable aléatoire (scalaire). Dans la suite on omettra la référence à l'ensemble  $\Omega$  et noterons la variable aléatoire  $Z(\cdot, x)$  par  $Z(x)$ .

Un processus aléatoire Gaussien est un processus aléatoire tel que pour tout  $n$  et tout  $x_i \in \mathbb{R}^d$  ( $i = 1, \dots, n$ ), le vecteur  $(Z(x_1), \dots, Z(x_n))$  est un vecteur Gaussien.

Un processus Gaussien est complètement déterminé par la connaissance de sa fonction moyenne i.e.

$$\forall x \in \mathbb{R}^d \quad m(x) = \mathbb{E}[Z(x)]$$

et de sa fonction de covariance :

$$\forall x, x' \in \mathbb{R}^d \quad k(x, x') = \text{Cov}(Z(x), Z(x')).$$

On note

$$Z \sim \mathcal{PG}(m, k).$$

Dans la suite nous supposerons le processus centré i.e.  $\forall x \in \mathbb{R}^d, m(x) = 0$ .

Ainsi pour tout  $n$  et tout  $x_i \in \mathbb{R}^d$ ,  $i = 1, \dots, n$ , le vecteur  $(Z(x_1), \dots, Z(x_n))$  est un vecteur Gaussien de moyenne nulle et de matrice de covariance  $\Sigma$  tel que

$$\Sigma_{i,j} = k(x_i, x_j) = \text{Cov}(Z(x_i), Z(x_j))$$

et l'on note

$$\begin{pmatrix} Z(x_1) \\ \vdots \\ Z(x_n) \end{pmatrix} \sim \mathcal{N}(0, \Sigma)$$

Dorénavant  $d = 1$  et la fonction de covariance est considérée paramétré par  $\lambda$  et s'exprime :

$$k_\lambda(x, x') = \left(1 + \frac{x - x'}{\lambda} + \frac{(x - x')^2}{3\lambda^2}\right) \exp\left(-\frac{x - x'}{\lambda}\right). \quad (1)$$

La matrice de covariance  $\Sigma$  a donc pour termes les  $\Sigma_{i,j} = k_\lambda(x_i, x_j)$ .

**1. Simuler plusieurs réalisations d'un processus Gaussien de moyenne nulle et de fonction de covariance (1).**

- (a) Choisissez  $N$  (grand) valeurs equi-distribuées  $x_i \in [0, 1]$  tel que  $x_1 < \dots < x_N$ .  
Fixez  $\lambda = 1$  et calculez la matrice de covariance  $\Sigma$  correspondante. Soit  $L$  la décomposition de Cholesky de la matrice  $\Sigma$ . Soit  $G = (G_1, \dots, G_N)$  un vecteur dont les coordonnées sont des Gaussiennes centrées, réduites et indépendantes. En notant  $\mathbf{m} = (m(x_1), \dots, m(x_n))$  (que l'on considérera non nulle uniquement pour cette question), que vaut la moyenne et la matrice de covariance du vecteur aléatoire  $\mathbf{m} + L^t G$  ?
- (b) Simulez  $\mathbf{g} = (g_1, \dots, g_N)$  une réalisation de  $G$  et calculez  $(z_1, \dots, z_N) = L^t \mathbf{g}$
- (c) Tracez le graphe  $(x_i, z_i)_{i=1, \dots, N}$  qui représente une réalisation du processus Gaussien  $Z$  pour une valeur de paramètre  $\lambda = 1$
- (d) Recommencez (b) plusieurs fois et représentez sur un même graphe les différentes réalisations de  $Z$ .
- (e) Refaite (a) à (d) en changeant de valeur pour  $\lambda$ . Que constatez-vous ? Comment interpréter ce paramètre ?

**2. Estimation du paramètre  $\lambda$  par maximum de vraisemblance**

- (a) Divisez  $[0, 1]$  en une partition de  $n = 10$  intervalles et tirez un nombre aléatoire uniformément dans chacun d'entre eux. On notera  $x_1, \dots, x_n$  ces  $n$  valeurs que l'on

stockera.

(b) Soit  $\mathbf{z} = (z_1, \dots, z_n)$  le vecteur de taille  $n$  définit pour tout  $i$  par

$$z_i = \sin(4\pi x_i). \quad (2)$$

Ecrire, comme une fonction paramétrée par  $\lambda$ , la négative log-vraisemblance (i.e. moins le logarithme de la vraisemblance) associée au vecteur aléatoire Gaussien  $(Z(x_1), \dots, Z(x_n))$  dont  $\mathbf{z}$  est considéré comme une réalisation. A l'aide d'un optimiseur, minimisez cette negative log-vraisemblance par rapport à  $\lambda$ .

(c) Sur un même graphe représentez la fonction  $\sin(4\pi x)$  et quelques réalisations du processus Gaussien  $Z$  (en utilisant la technique proposée en 1.) en prenant pour  $\lambda$  la valeur optimale trouvée en 2.(b).

### 3. Simuler plusieurs réalisations d'un processus Gaussien conditionné à un ensemble de données.

Dans la suite le  $\lambda$  considéré sera celui obtenu dans 2. et nous reprenons également les mêmes  $n$  nombres  $(x_1, \dots, x_n)$  obtenus par simulation en 2.a. Soit  $N$  valeurs prises dans  $[0, 1]$  :  $x'_1, \dots, x'_N$ , distinctes des  $x_i$ . Nous connaissons la distribution Gaussienne des deux vecteurs aléatoires  $(Z(x_1), \dots, Z(x_n))$  et  $(Z(x'_1), \dots, Z(x'_N))$  qui sont de moyennes nulle et de matrices de covariance respectives  $\Sigma_n$  et  $\Sigma_N$  dont les termes sont calculés avec (1). Nous pouvons ensuite montrer que le vecteur aléatoire conditionné

$$Z(x'_1), \dots, Z(x'_N) | Z(x_1) = z_1, \dots, Z(x_n) = z_n$$

suit également une loi Gaussienne de vecteur moyenne

$$\boldsymbol{\mu} = \begin{pmatrix} \mathbf{k}_n^t(x'_1) \Sigma_n^{-1} \mathbf{z} \\ \vdots \\ \mathbf{k}_n^t(x'_N) \Sigma_n^{-1} \mathbf{z} \end{pmatrix} \quad (3)$$

où  $\mathbf{z}$  est le vecteur des données obtenu avec (2), et  $\mathbf{k}_n(x'_i) = (k_\lambda(x_1, x'_i), \dots, k_\lambda(x_n, x'_i))^t$ .

La matrice de covariance du vecteur conditionné est elle donnée par

$$\Sigma_{N|n} = \Sigma_N - \Sigma_{n,N}^t \Sigma_n^{-1} \Sigma_{n,N} \quad (4)$$

où  $\Sigma_N$  est la matrice de taille  $N \times N$  des covariances entre les  $N$  nouveaux points et  $\Sigma_{n,N}$  la matrice de taille  $n \times N$  dont la  $i$ -ième colonne correspond au vecteur de taille  $n$  des covariances du  $i$ -ième nouveau point avec les  $n$  points d'apprentissages.

- (a) Pour un  $x \in [0, 1]$  rappelez la moyenne et la variance de  $Z(x)$  et donnez celles de  $Z(x)$  sachant  $Z(x_1) = z_1, \dots, Z(x_n) = z_n$  que l'on notera respectivement  $\mu_n(x)$  et

$$\sigma_n^2(x) := \text{Var}[Z(x)|Z(x_1) = z_1, \dots, Z(x_n) = z_n].$$

On dit que le processus  $Z(x)$  est le processus Gaussien a priori et  $Z(x)$  sachant  $Z(x_1) = z_1, \dots, Z(x_n) = z_n$  le processus Gaussien a posteriori.

- (b) Tracez sur un même graphe la fonction à approximer :  $\sin(4\pi x)$ , la moyenne  $\mu_n(x)$  ainsi que  $\mu_n(x) + 2\sigma_n^2(x)$  et  $\mu_n(x) - 2\sigma_n^2(x)$ .
- (c) En utilisant les formules de la moyenne et de la covariance a posteriori du processus Gaussien conditionné donnés par (3) et (4) ainsi que la technique introduite en 1., avec  $N$  grand, tracez sur un même graphe un ensemble de réalisations du processus a posteriori ainsi que la fonction  $\sin(4\pi x)$ .

### Problem 3 - Simuler une loi de distribution a posteriori avec un algorithme MCMC

Considérez qu'une v.a.  $X$  suit une loi de Poisson  $\mathcal{P}(\theta)$  où  $\theta$  est un paramètre inconnu. D'autre part considérez que la loi de distribution a priori de  $\theta$  est une loi Gamma de paramètres fixés connus  $k$  et  $\lambda$  :  $\mathcal{G}(k, \lambda)$  (Voir distribution dans le cours). Le paramètre de translation  $\gamma$  est considéré nulle. Prenez  $x = 10$ . La loi a posteriori de  $\theta|X = x$  est connue, c'est une loi  $\mathcal{G}(k + x, \lambda + 1)$ . Nous à présent allons générer un échantillon de cette loi en utilisant le produit de la vraisemblance et du prior introduit ci-dessus avec un algorithme MCMC.

1. Afin de simuler une chaîne de Markov suivant la loi a posteriori Gamma objec-

tif, modifiez la vraisemblance et le prior de la fonction "posterior" de l'algorithme MCMC implémenté en python suivant : <https://moonbooks.org/Articles/Lalgorithme-de-Metropolis-Hastings-MCMC-avec-python/>. Générez une longue chaîne et prenez uniquement la seconde moitié de la chaîne comme échantillon représentatif.

2. Avec l'échantillon de 3.a faites une approximation à noyau de la densité associée et comparez cette densité estimée avec la vraie densité objectif de  $\mathcal{G}(k + x, \lambda + 1)$ .