Université de Montpellier Année 2024-2025
M2 SSD - HAX918X

Examen de contrôle continu - vendredi 18 octobre 2024
Durée 1h - Seul le formulaire de probabilités est autorisé

Exercise 1 (8 pts) La distribution de von Mises est une distribution de probabilité qui s'applique principalement aux données circulaires, c'est-à-dire des variables aléatoires dont les valeurs sont des angles mesurés sur un cercle, comme des directions ou des orientations. La distribution de von Mises est souvent vue comme une analogie circulaire de la distribution normale. La fonction de densité de probabilité de la distribution de von Mises pour un angle x (en radians) est donnée par :

$$f(x|\mu,\kappa) \propto \exp\left(\kappa\cos(x-\mu)\right)\mathbf{1}_{]-\pi,\pi]}(x)$$

— $\mu \in]-\pi,\pi]$ est le paramètre de localisation (l'angle moyen) exprimé en radian,
— $\kappa > 0$ est le paramètre de concentration, analogue à l'inverse de la variance dans une distribution normale (plus κ est grand, plus les observations sont concentrées autour de μ).

Dans la suite, nous supposons que $\mu = 0$ et $\kappa = 2$.

1) (4 pts) Proposer une méthode de simulation suivant $f(x|0,2)$.

2) (4 pts) Donner le script du code R permettant de mettre en oeuvre la stratégie présentée à la question précédente.

Correction 1) La densité de la loi cible est

$$f(x|0,2) \propto \exp(2\cos(x))\mathbf{1}_{]-\pi,\pi]}(x)\,.$$

Comme $\forall x \in]-\pi,\pi]$, nous avons

$$\exp(2\cos(x)) \leq \exp(2)\,,$$

nous pouvons utiliser l'algorithme acceptation-rejet avec pour loi instrumentale une loi uniforme sur $]-\pi,\pi]$. Pour simuler suivant $f(\cdot|0,2)$ il suffit donc de simuler une réalisation y suivant une loi uniforme sur $]-\pi,\pi]$ et d'accepter cette simulation comme une simulation suivant $f(\cdot|0,2)$ si $u \leq \frac{\exp(2\cos(y))}{\exp(2)}$ où u est une simulation suivant une loi uniforme sur $[0,1]$.

2) Voici le code R associé à la stratégie décrite à la question précédente :

```
x <- rep(0,N)
for (i in 1:N)
{
test <- TRUE
while (test)
{
y <- runif(1,-pi,pi)
if (runif(1)<=exp(2*cos(y))/exp(2)) { x[i] <- y ; test <- FALSE }
}
}
```

Exercice 2 (4 pts) Le code R ci-dessous permet de simuler un n-échantillon suivant une loi de probabilité de densité $f(y)$

```
n <- 10000
x_m <- 1
alpha <- 1/2
u <- runif(n)
y <- x_m / (1 - u)^(1 / alpha)
```

Expliciter cette densité, être vigilant quant au support de la loi.

Correction La méthode de simulation utilisée est celle de l'inversion de la fonction de répartition. Pour retrouver la densité, nous allons calculer la fonction de répartition et la dériver. Soit $u \in]0,1[$,

$$y = \frac{1}{(1-u)^2}$$

on remarque que $y \geq 1$ et

$$\Longleftrightarrow u = 1 - 1/\sqrt{y}\,.$$

La densité associé est alors

$$f(y) = \frac{1}{2y^{3/2}} \mathbf{1}_{y \geq 1}\,.$$

Il s'agit d'une loi de Pareto de paramètres $(1, 1/2)$.

Exercice 3 (4 pts) Nous considérons un n-échantillon $(x_1, \ldots, x_n)$ suivant une loi Exponentielle de paramètre $\lambda > 0$. Nous nous plaçons dans le paradigme bayésien et supposons que λ est distribué suivant une loi a priori Gamma de paramètres $\alpha = 2$ et $\beta = 1$.

1) (2 pts) Expliciter la loi a posteriori de λ.

2) (2 pts) Donner un intervalle de crédibilité de niveau 95% pour λ.

Correction 1)

$$\pi(\lambda | x_1, \ldots, x_n) \propto \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right) \lambda \exp(-\lambda) \mathbf{1}_{\lambda > 0}$$

$$\lambda | x_1, \ldots, x_n \sim \mathcal{G}a\left(n+2, \sum_{i=1}^n x_i + 1\right)\,.$$

2) Un intervalle de crédibilité de niveau 95% pour λ est

$$\left[F^{-1}_{\mathcal{G}a\left(n+2, \sum_{i=1}^n x_i + 1\right)}(0.025), F^{-1}_{\mathcal{G}a\left(n+2, \sum_{i=1}^n x_i + 1\right)}(0.975)\right]$$

Exercice 4 (4 pts) Nous considérons un n-échantillon $(x_1, \ldots, x_n)$ suivant une loi uniforme sur $[0, \theta]$. Nous nous plaçons dans le paradigme bayésien et souhaitons estimer θ. Nous supposons que θ est distribué suivant une loi a priori Exponentielle de paramètre $\lambda = 1$.

1) (2 pts) Expliciter la loi a posteriori de θ.

2) (2 pts) Donner l'intervalle de crédibilité HPD de niveau 95% pour θ.

Correction 1)

$$\pi(\theta | x_1, \ldots, x_n) \propto \theta^{-n} \exp(-\theta) \prod_{i=1}^n \mathbf{1}_{\theta > x_i}$$

$$\pi(\theta | x_1, \ldots, x_n) \propto \theta^{-n} \exp(-\theta) \mathbf{1}_{\theta > \max(x_1, \ldots, x_n)}$$

Cela ne correspond pas à une loi connue.

$$\pi(\theta | x_1, \ldots, x_n) = \frac{\theta^{-n} \exp(-\theta)}{\int_{\max(x_1,\ldots,x_n)}^{\infty} \theta^{-n} \exp(-\theta) \mathrm{d}\theta} \mathbf{1}_{\theta > \max(x_1, \ldots, x_n)}\,.$$

2) On remarque que la densité a posteriori est une fonction décroissante en θ. Ainsi, l'intervalle de crédibilité HPD de niveau 95% pour θ est

$$[\max(x_1, \ldots, x_n), b]$$

où $b > \max(x_1, \ldots, x_n)$ est tel que $\dfrac{\int_{\max(x_1,\ldots,x_n)}^{b} \theta^{-n} \exp(-\theta) \mathrm{d}\theta}{\int_{\max(x_1,\ldots,x_n)}^{\infty} \theta^{-n} \exp(-\theta) \mathrm{d}\theta} = 0.95.$