

Examen Partiel – Octobre 2019

Durée 2h00 – Sans documents

Important. Suivant les règlements en vigueur,

1. les enseignants présents lors de l'épreuve ne peuvent communiquer que sur les fautes d'énoncé potentielles. Toute autre question durant la composition ne sera pas acceptée.
2. les étudiants sont tenus de se lever au moment de l'annonce de fin de la composition. En cas de refus, le responsable de l'UE sera fondé à ne pas prendre en compte la copie incriminée.
3. l'identification des copies et intercalaires doit se faire au moment de la remise de chaque copie par les enseignants et surveillants. Il ne sera pas accordé de délai pour cette raison en fin d'épreuve.

Exercice 1 (7 pts – **Les questions peuvent être résolues indépendamment**). Pour représenter les cours de divers instruments financiers comme les actions ou les taux d'intérêt, on décide d'utiliser le modèle statistique $\mathcal{P}$ associé à la loi log-normale, notée $\log - \mathcal{N}(\mu, \sigma^2)$, dont la densité est donnée par

$$f(y \mid \mu, \sigma^2) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(\ln(y) - \mu)^2}{2\sigma^2} \right], \quad \mu \in \mathbb{R} \quad \text{et} \quad \sigma^2 \in \mathbb{R}_+.$$

1. (1 pt) Justifier que la loi log-normale appartient à une famille exponentielle. Cette famille est-elle minimale ?
2. (1 pt) Dans R, la loi log-normale est désignée par `lnorm` et a pour paramètre `meanlog` et `sdlog`. Donner les codes R qui, pour `meanlog = 1` et `sdlog = 0.5`, permettent de simuler un échantillon y de 100 réalisations de la loi log-normale, d'évaluer la densité f au point 0.7, de calculer le quantile d'ordre 90% de la loi log-normale et le quantile empirique d'ordre 90% de l'échantillon y .

On observe différents cours $y_1, \dots, y_n$ supposés être des réalisations *i.i.d* du modèle $\mathcal{P}$. On souhaite estimer les paramètres μ et σ^2 à l'aide des estimateurs suivants

$$\hat{\mu}_n = \frac{1}{n} \sum_{k=1}^n \ln(Y_k). \quad \text{et} \quad \hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n [\ln(Y_k) - \hat{\mu}_n]^2.$$

3. (a) (0.5 pt) Montrer que si Y suit la loi $\log - \mathcal{N}(\mu, \sigma^2)$, alors $X = \ln(Y)$ suit la loi normale $\mathcal{N}(\mu, \sigma^2)$.
(b) (1.5 pts) En déduire que les estimateurs $\hat{\mu}_n$ et $\hat{\sigma}_n^2$ convergent respectivement vers μ et σ^2 et que $\hat{\sigma}_n^2$ est biaisé, c'est-à-dire $\mathbb{E}[\hat{\sigma}_n^2] \neq \sigma^2$.
4. (1 pts) La médiane de la loi $\log - \mathcal{N}(\mu, \sigma^2)$ est $m = \exp(\mu)$. Donner un intervalle de confiance pour m au niveau de confiance asymptotique $1 - \alpha$ en fonction de $\hat{\mu}_n$.

Dans la suite, on considère que l'on a chargé dans R un vecteur d'observations y correspondant aux différents cours financiers que l'on étudie.

5. (1 pt) Écrire le code R donnant une approximation bootstrap paramétrique d'un intervalle de confiance pour σ^2 au niveau 95%.
6. (1 pt) Écrire le code R permettant d'évaluer l'erreur quadratique moyenne $\mathbb{E}[(\hat{\sigma}_n^2 - \sigma^2)^2]$ à l'aide d'une procédure bootstrap non paramétrique.

Solution Exercice 1.

1. La densité peut s'écrire sous la forme $f(y \mid \mu, \sigma^2) = c(\mu, \sigma^2)h(y) \exp(\eta(\mu, \sigma^2)^T T(y))$ avec

$$c(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \quad (0.2pt), \quad h(y) = \frac{1}{y} \quad (0.2pt),$$

$$\eta(\mu, \sigma^2) = \begin{pmatrix} \mu/\sigma^2 \\ -1/(2\sigma^2) \end{pmatrix} \quad (0.2pt) \quad \text{et} \quad T(y) = \begin{pmatrix} \ln(y) \\ \ln(y)^2 \end{pmatrix} \quad (0.2pt).$$

Les composantes de T sont linéairement indépendantes. La famille est donc minimale (0.2 pt).

2. (0.25 pt par réponse)

```
y <- rlnorm(100, meanlog = 1, sdlog = 0.5)
dlnorm(0.7, meanlog = 1, sdlog = 0.5)
qlnorm(0.9, meanlog = 1, sdlog = 0.5)
quantile(y, probs = 0.9)
```

3. (a) Soit h une fonction borélienne positive. En considérant le changement de variable $x \mapsto \ln x$ qui réalise un $\mathcal{C}^1$ -difféomorphisme de $\mathbb{R}_+^*$ dans $\mathbb{R}$, on obtient

$$\begin{aligned} \mathbb{E}[h(X)] &= \int_0^{+\infty} h(\ln y) \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln(y) - \mu)^2}{2\sigma^2}\right] dy \\ &= \int_{-\infty}^{+\infty} h(x) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] dx. \end{aligned}$$

Donc X suit la loi $\mathcal{N}(\mu, \sigma^2)$.

- (b) La suite $\ln Y_1, \dots, \ln Y_n$ est une suite de variables *i.i.d.* de loi $\mathcal{N}(\mu, \sigma^2)$ et sont donc de carré intégrable (0.2 pt). La loi des grands nombres nous permet donc de dire que

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n \ln Y_k &= \hat{\mu}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}[\ln Y_1] = \mu \quad (0.4pt) \\ \frac{1}{n} \sum_{k=1}^n (\ln Y_k)^2 &\xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}[(\ln Y_1)^2] = \text{Var}[\ln Y_1] + \mathbb{E}[\ln Y_1]^2 = \sigma^2 + \mu^2. \quad (0.2pt) \end{aligned}$$

Par ailleurs, on a

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n (\ln Y_k)^2 - (\hat{\mu}_n)^2.$$

En utilisant le théorème de continuité, on obtient que

$$\hat{\sigma}_n^2 \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \sigma^2 + \mu^2 - \mu^2 = \sigma^2. \quad (0.2pt)$$

On peut également en déduire que

$$\mathbb{E} [\hat{\sigma}_n^2] = \mathbb{E}[(\ln Y_1)^2] - \mathbb{E}[(\hat{\mu}_n)^2] = \sigma^2 + \mu^2 - \mathbb{E}[(\hat{\mu}_n)^2].$$

Or les variables étant *i.i.d*, on a

$$\begin{aligned} \mathbb{E}[(\hat{\mu}_n)^2] &= \text{Var}[\hat{\mu}_n] + \mathbb{E}[\hat{\mu}_n]^2 \\ &= \frac{1}{n} \text{Var}[\ln Y_1] + \mu^2 = \frac{1}{n} \sigma^2 + \mu^2. \end{aligned}$$

On en déduit que

$$\mathbb{E}[\hat{\sigma}_n^2] = \frac{n-1}{n} \sigma^2. \quad (0.5pt)$$

4. Les variables $\ln Y_1, \dots, \ln Y_n$ étant indépendantes et de carré intégrable, le théorème central limite donne

$$\sqrt{n} \left(\frac{1}{n} \sum_{k=1}^n \ln Y_k - \mathbb{E}[\ln Y_1] \right) = \sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \text{Var}[\ln Y_1]) = \mathcal{N}(0, \sigma^2). \quad (0.2pt)$$

La fonction $g : x \mapsto \exp(x)$ ne s'annule pas en μ (0.1 pt), donc la delta méthode nous permet d'écrire

$$\sqrt{n}(g(\hat{\mu}_n) - g(\mu)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, [g'(\mu)]^2 \sigma^2) = \mathcal{N}(0, \sigma^2 \exp(2\mu)). \quad (0.2pt)$$

La fonction $(x, y) \mapsto (xy)^{-1/2}$ étant continue sur $\mathbb{R}_+^* \times \mathbb{R}_+^*$, on en déduit que

$$\frac{1}{\sqrt{\hat{\sigma}_n^2 \exp(2\hat{\mu}_n)}} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \frac{1}{\sqrt{\sigma^2 \exp(2\mu)}}. \quad (0.2pt)$$

Le théorème de Slutsky permet alors de conclure

$$\sqrt{\frac{n}{\hat{\sigma}_n^2 \exp(2\hat{\mu}_n)}} (g(\hat{\mu}_n) - m) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1). \quad (0.1pt)$$

L'intervalle de confiance asymptotique au niveau $1 - \alpha$ pour m est

$$\left[g(\hat{\mu}_n) - q_{1-\alpha/2} \sqrt{\frac{\hat{\sigma}_n^2 \exp(2\hat{\mu}_n)}{n}}, g(\hat{\mu}_n) + q_{1-\alpha/2} \sqrt{\frac{\hat{\sigma}_n^2 \exp(2\hat{\mu}_n)}{n}} \right]$$

avec $q_{1-\alpha/2}$ le quantile d'ordre $1 - (\alpha/2)$ de la loi $\mathcal{N}(0, 1)$ (0.2 pt).

5. On se donne un échantillon y .

```

# Estimation sur l'observation
mu_hat <- mean(log(y))
sigma2_hat <- mean((log(y) - mu_hat)^2)
# Procédure bootstrap paramétrique
n <- length(y)
K <- 1000
boot_sigma <- numeric(n)
for (i in 1:K) {
  y_boot <- rlnorm(n, mu_hat, sqrt(sigma2_hat))
  boot_sigma[i] <- mean((log(y_boot) - mu_hat)^2)
}
# Intervalle de confiance
quantile(boot_sigma, prob = c(0.025, 0.975))

```

6. On se donne un échantillon y .

```

# Estimation sur l'observation
mu_hat <- mean(log(y))
sigma2_hat <- mean((log(y) - mu_hat)^2)
# Procédure bootstrap non paramétrique
n <- length(y)
K <- 1000
boot_sigma <- numeric(n)
for (i in 1:K) {
  y_boot <- sample(y, n, replace = TRUE)
  boot_sigma[i] <- mean((log(y_boot) - mu_hat)^2)
}
# Estimation de l'erreur quadratique moyenne
mean((boot_sigma - sigma2_hat)^2)

```

Exercice 2 (5 pts). On définit la loi de Weibull par la densité suivante :

$$f(x | \lambda) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} \exp \left[- \left(\frac{x}{\lambda}\right)^k \right] \mathbb{I}_{x>0},$$

où $k > 1$ est connu et λ est inconnu. On note $x_1, \dots, x_n$ un échantillon *i.i.d.* de cette loi.

- (1.5 pts) Montrer que cette loi appartient à une famille exponentielle et en donner la forme canonique, avec comme statistique $T(x) = x^k$. Cette famille est-elle minimale ? régulière ?
- (2 pts) Calculer la fonction génératrice des moments associée à la statistique naturelle. En déduire $\mathbb{E}[X^k]$, $\mathbb{E}[X^{2k}]$ et $\text{Var}[X^k]$ pour X suivant une loi de Weibull.
- (0.5 pt) Montrer que $\hat{\lambda}_n^k = \frac{1}{n} \sum_{i=1}^n x_i^k$ est un estimateur non biaisé de λ^k , c'est-à-dire $\mathbb{E}[\hat{\lambda}_n^k] = \lambda^k$. En déduire que $\hat{\lambda}_n = \sqrt[k]{\frac{1}{n} \sum_{i=1}^n x_i^k}$ est un estimateur biaisé de λ .
- (1pt) Trouver une suite (a_n) telle que

$$\frac{\sqrt{n}}{a_n} (\hat{\lambda}_n - \lambda) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

En déduire un intervalle de confiance asymptotique de niveau α pour λ .

5. (1 pt) Quelle est la densité de X^k ?

Solution Exercice 2.

1. La densité f peut s'écrire $f(x | \lambda) = c(\lambda)h(x) \exp(\eta(\lambda) \cdot T(x))$ avec

$$c(\lambda) = \frac{k}{\lambda^k} \text{ (0.25pt)}, \quad h(x) = x^{k-1} \mathbf{1}_{x>0} \text{ (0.25pt)}, \quad \eta(\lambda) = -\lambda^{-k} \text{ (0.25pt)} \quad \text{et} \quad T(x) = x^k.$$

La famille est de dimension 1 donc elle est minimale (0.25 pt). L'espace naturel des paramètres est :

$$\Theta = \left\{ \theta \mid \int_0^{+\infty} x^{k-1} \exp(\theta x^k) dx < \infty \right\}.$$

La fonction $x \mapsto x^{k-1} \exp(\theta x^k)$ est intégrable en $+\infty$ lorsque $\theta < 0$, autrement dit $\Theta = \mathbb{R}_-^*$ (0.5 pt). La forme canonique est obtenue pour la paramétrisation $\theta = \eta(\lambda)$, soit $\lambda = (-\theta)^{-1/k}$, et on peut en déduire que $\lambda \in \mathbb{R}_+^*$.

2. Le cumulante est $\psi(\theta) = -\log[c \circ \eta^{-1}(\theta)] = -\log(-k\theta)$. On déduit que la fonction des génératrice des moments de $T(X)$ est

$$M_{T(X)}(t) = \exp[\psi(\theta + t) - \psi(\theta)] = \frac{c \circ \eta^{-1}(\theta)}{c \circ \eta^{-1}(\theta + t)} = \frac{\theta}{\theta + t}. \quad \text{(0.5pt)}$$

On utilise que $\mathbb{E}[X^k] = \mathbb{E}[T(X)] = M'_{T(X)}(0)$ et $\mathbb{E}[X^{2k}] = \mathbb{E}[T(X)^2] = M''_{T(X)}(0)$. Pour tout t réel, on a

$$M'_{T(X)}(t) = -\frac{\theta}{(\theta + t)^2} \quad \text{et} \quad M''_{T(X)}(t) = \frac{2\theta}{(\theta + t)^3}.$$

On en déduit

$$\mathbb{E}[X^k] = -\frac{1}{\theta} = \lambda^k \quad \text{(0.5pt)} \quad \text{et} \quad \mathbb{E}[X^{2k}] = \frac{2}{\theta^2} = 2\lambda^{2k}. \quad \text{(0.5pt)}$$

Il en résulte que $\text{Var}[X^k] = \theta^{-2} = \lambda^{2k}$ (0.5 pt).

3. L'espérance étant linéaire, on a $\mathbb{E}[\hat{\lambda}_n^k] = \mathbb{E}[X^k] = \lambda^k$. Donc c'est un estimateur non biaisé de λ^k (0.25 pt). La fonction $g : y \mapsto \sqrt[k]{y}$ est strictement convexe donc l'inégalité de Jensen donne

$$\mathbb{E}[\hat{\lambda}] = \mathbb{E}[g(\hat{\lambda}_n^k)] > g(\mathbb{E}[\hat{\lambda}_n^k]) = \lambda.$$

On obtient donc un estimateur biaisé de λ (0.25 pt).

4. $X_1^k, \dots, X_n^k$ est une suite de variables *i.i.d* de carré intégrable ($\text{Var}[X^k] < +\infty$) (0.1 pt). Le théorème central limite donne

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i^k - \mathbb{E}[X_1^k] \right) = \sqrt{n}(\hat{\lambda}_n^k - \lambda^k) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \lambda^{2k}).$$

La fonction $g : y \mapsto \sqrt[k]{y}$ est telle que $g(\lambda^k) = \lambda \neq 0$ (0.1 pt). On peut donc appliquer la delta méthode avec la fonction g et on obtient

$$\sqrt{n}[g(\hat{\lambda}_n^k) - g(\lambda^k)] = \sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \lambda^{2k}(g'(\lambda^k))^2) = \mathcal{N}(0, \lambda^2/k^2). \quad \text{(0.2pt)}$$

Autrement dit, on a

$$\frac{k}{\lambda} \sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Or en utilisant la loi forte des grands nombre et la continuité de la fonction $x \mapsto k/\sqrt[k]{x}$, on obtient que $k/\hat{\lambda}_n$ converge en probabilité vers k/λ (0.2 pt). Il suffit donc de prendre $a_n = \hat{\lambda}_n/k$ (0.2 pt) et le théorème de Slutsky permet de conclure

$$\frac{k}{\hat{\lambda}_n} \sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

On en déduit l'intervalle de confiance asymptotique au niveau α

$$\left[\hat{\lambda}_n - q_{(1+\alpha)/2} \frac{a_n}{\sqrt{n}}, \hat{\lambda}_n + q_{(1+\alpha)/2} \frac{a_n}{\sqrt{n}} \right],$$

avec $q_{(1+\alpha)/2}$ le quantile d'ordre $(1+\alpha)/2$ de la loi $\mathcal{N}(0, 1)$ (0.2 pt).

5. Soit h une fonction borélienne positive, en considérant le changement de variable $t \mapsto t^k$ qui réalise un $\mathcal{C}^1$ difféomorphisme de $\mathbb{R}_+$ dans $\mathbb{R}_+$, on a

$$\begin{aligned} \mathbb{E}[h(X^k)] &= \int_0^{+\infty} h(x^k) \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{x}{\lambda}\right)^k\right] dx \\ &= \int_0^{+\infty} h(t) \frac{1}{\lambda^k} \exp\left(-\frac{t}{\lambda^k}\right) dt. \end{aligned}$$

On en déduit que X^k suit une loi exponentielle de paramètre λ^{-k} (1 pt).

Exercice 3 (3 pts). On considère un portefeuille d'assurance où le nombre de sinistres X_i de l'assuré i dans une année suit une loi de Poisson de paramètre θ_i . Les θ_i sont distincts, inconnus et aléatoires de loi π . Autrement dit, on connaît uniquement la loi de X_i conditionnellement à θ_i

$$\mathbb{P}[X_i = n \mid \theta_i] = \frac{e^{-\theta_i} \theta_i^n}{n!} \quad \forall n \in \mathbb{N}.$$

On note X le nombre de sinistres dans une année d'un assuré pris au hasard. Dans la suite, on cherche à étudier la loi de X marginale, c'est-à-dire la loi du nombre de sinistres dans une année sans connaître la valeur de θ .

Les réponses aux questions 2. et 3. ne dépendent pas de la question 1.

1. (1 pts) On suppose que π est la loi Gamma de paramètre (r, λ) de densité par rapport à la mesure de Lebesgue

$$f_{r,\lambda}(t) = \frac{1}{\Gamma(r)} \lambda^r t^{r-1} \exp(-\lambda t) \mathbb{I}_{t \geq 0},$$

où $r > 1$ et $\lambda > 0$. En admettant que $\mathbb{P}[X = n] = \mathbb{E}_\pi[\mathbb{P}[X = n \mid \theta]]$, montrer que la loi de X marginale suit une loi binomiale négative de paramètre (a, p) (avec $a > 0$ et $0 < p < 1$) que l'on précisera, c'est-à-dire qu'elle a pour fonction de masse,

$$\forall n \in \mathbb{N}, \quad \mathbb{P}[X = n] = \frac{\Gamma(a+n)}{n! \Gamma(a)} p^a (1-p)^n.$$

2. (1 pt) Montrer que si le paramètre a est fixé, la loi binomiale négative de paramètre (a, p) appartient à une famille exponentielle. Est-elle sous forme canonique ?
3. (1 pt) En utilisant la forme canonique de cette famille exponentielle, déduire l'espérance et la variance de X .

Solution Exercice 3.

1. On a

$$\begin{aligned}
 \mathbb{P}[X = n] &= \mathbb{E}_\pi [\mathbb{P}[X = n \mid \theta]] \\
 &= \frac{\lambda^r}{\Gamma(r)} \frac{1}{n!} \int \theta^{r-1} \exp(-\lambda\theta) \theta^n \exp(-\theta) \mathbb{1}_{\theta \geq 0} d\theta \\
 &= \frac{\lambda^r}{\Gamma(r)} \frac{1}{n!} \int \theta^{n+r-1} \exp(-(\lambda+1)\theta) \mathbb{1}_{\theta \geq 0} d\theta \\
 &= \frac{\lambda^r \Gamma(n+r)}{\Gamma(r)(\lambda+1)^{n+r}} \frac{1}{n!} \int f_{n+r, \lambda+1}(\theta) d\theta \\
 &= \frac{\lambda^r \Gamma(n+r)}{\Gamma(r)(\lambda+1)^{n+r} n!}.
 \end{aligned}$$

On reconnaît la fonction de masse d'une binomiale négative avec $a = r > 0$ et $p = \lambda/(\lambda+1) \in]0, 1[$ (1 pt).

2. Si a est fixé, on a

$$\begin{aligned}
 \mathbb{P}(Y = n) &= \frac{\Gamma(a+n)}{n! \Gamma(a)} p^a (1-p)^n = \frac{\Gamma(a+n)}{n! \Gamma(a)} p^a \exp(n \log(1-p)) \\
 &= c(p) h(n) \exp(\eta(p) T(n))
 \end{aligned}$$

avec

$$\begin{aligned}
 c(p) &= p^a \quad (0.2pt), \quad h(n) = \frac{\Gamma(a+n)}{n! \Gamma(a)} \quad (0.2pt), \\
 \eta(p) &= \log(1-p) \quad (0.2pt) \quad \text{et} \quad T(n) = n \quad (0.2pt).
 \end{aligned}$$

Elle n'est pas sous forme canonique (0.2 pt).

3. La forme canonique s'obtient en considérant la paramétrisation $\theta = \eta(p) = \log(1-p)$. La constante de normalisation de la forme canonique est alors $\tilde{c}(\theta) = c \circ \eta^{-1}(\theta) = (1 - \exp(\theta))^a$. On a alors

$$\begin{aligned}
 \mathbb{E}[X] &= \mathbb{E}[T(X)] = -\frac{\partial}{\partial \theta} \log \tilde{c}(\theta) = \frac{a \exp(\theta)}{1 - \exp(\theta)} = \frac{a(1-p)}{p} \quad (0.5pt) \\
 \text{Var}[X] &= \text{Var}[T(X)] = -\frac{\partial^2}{\partial \theta^2} \log(c(\theta)) = \frac{a \exp(\theta)}{(1 - \exp(\theta))^2} = \frac{a(1-p)}{p^2} \quad (0.5pt)
 \end{aligned}$$

En remplaçant par les valeurs de a et de p obtenues à la question précédente, on a

$$\mathbb{E}[X] = \frac{\lambda}{r} \quad \text{et} \quad \text{Var}[X] = \frac{r}{\lambda} + \frac{r}{\lambda^2}.$$

Exercice 4 (6 pts – Les questions peuvent être résolues indépendamment). Soit f une densité de probabilité sur $\mathbb{R}$. On définit l'ensemble

$$\mathfrak{A} = \left\{ (u; v) \in \mathbb{R}^+ \times \mathbb{R}; 0 \leq u^2 \leq f(v/u) \right\},$$

1. (2 pts)

- (a) Définissant $\mathcal{U}(\mathfrak{A})$ comme la loi uniforme sur l'ensemble $\mathfrak{A}$, montrez que, si $(U, V) \sim \mathcal{U}(\mathfrak{A})$, la densité de la transformée $(W, X) = (U^2, V/U)$ est

$$\tilde{f}(w, x) \propto \mathbb{I}_A(w^{1/2}, w^{1/2}x) \times w^{1/2} \times \frac{1}{2w^{1/2}} \propto \mathbb{I}_{0 \leq w \leq f(x)}$$

où $\propto$ dénote le signe de proportionnalité, $f \propto g$ signifiant qu'il existe une constante $c > 0$ telle que $f(x) = cg(x)$ pour tout x dans $\mathbb{R}$. En déduire que la loi de $X = V/U$ a pour densité f .

- (b) Montrer que le résultat s'étend au cas où f est seulement proportionnelle à une densité de probabilité.
2. (1 pt) Si l'ensemble $\mathfrak{A}$ est borné, proposer une méthode de simulation de f fondée sur des simulations uniformes sur un certain carré de $\mathbb{R}^+ \times \mathbb{R}$.
3. (1 pt) Montrer que l'ensemble $\mathfrak{A}$ est borné par la courbe

$$u(x) = \sqrt{f(x)} \quad \text{et} \quad v(x) = x\sqrt{f(x)}, \quad \text{avec} \quad x \in \mathbb{R}.$$

4. (2 pts)

- (a) Déterminer cette courbe dans le cas où f est la densité de la loi Normale $\mathcal{N}(0, 1)$ et montrer que l'ensemble $\mathfrak{A}$ associé est inclus dans un carré dont on déterminera les bornes

$$0 \leq \sqrt{f(x)} \leq b \quad c \leq x\sqrt{f(x)} \leq d$$

- (b) Proposer un code R mettant en œuvre cette méthode pour la simulation de la loi Normale $\mathcal{N}(0, 1)$.
- (c) Sachant que le rapport de la surface de $\mathfrak{A}$ à celle du carré l'entourant vaut 0.73 (admis), discuter de l'efficacité de cette méthode.

Solution Exercice 4.

1. (a) Si $(U, V) \sim \mathcal{U}(\mathfrak{A})$, sa densité est l'indicatrice de $\mathfrak{A}$. Par la formule du changement de variable, la densité de (W, X) est donnée par la densité de (U, V) en $(w^{1/2}, w^{1/2}x)$ multipliée par le Jacobien

$$\frac{d(u, v)}{d(w, x)} = \det \begin{vmatrix} \frac{dw^{1/2}}{du} & \frac{dw^{1/2}x}{dv} \\ \frac{dw^{1/2}}{dv} & \frac{dw^{1/2}x}{du} \end{vmatrix} = \det \begin{vmatrix} 1/2w^{-1/2} & 1/2w^{-1/2}x \\ 0 & w^{1/2} \end{vmatrix} = 1/2$$

soit donc

$$\tilde{f}(w, x) \propto \mathbb{I}_A(w^{1/2}, w^{1/2}x) \times w^{1/2} \times \frac{1}{2w^{1/2}} \propto \mathbb{I}_{0 \leq w \leq f(x)}.$$

Intégrant cette densité en w donne $f(x)$, densité marginale de X . Et le $1/2$ représente la surface de $\mathfrak{A}$.

- (b) Si f est seulement proportionnelle à une densité de probabilité, $f(x) \propto \tilde{f}$, avec $\int f(x)dx = c$, le volume de $\mathfrak{A}$ est multiplié par c , ce qui redonne $\tilde{f}$ comme densité marginale de X . (0.5 pt)
2. Si l'ensemble $\mathfrak{A}$ est borné en u , $0 \leq u \leq a$ et en v , $c \leq v \leq d$, il est inclus dans le carré $(0, a) \times (b, c) \subset \mathbb{R}^+ \times \mathbb{R}$. Donc simuler des points uniformément dans le carré $(0, a) \times (b, c)$ jusqu'à ce qu'ils appartiennent à $\mathfrak{A}$ produit des simulations suivant $\mathcal{U}(\mathfrak{A})$.
3. La frontière de l'ensemble $\mathfrak{A}$ correspond à l'existence d'un couple (u, v) tel que $u^2 = f(v/u)$. Dénnotant $x = v/u$, il existe donc un $x \in \mathbb{R}$ tel que $u = \sqrt{f(x)}$ et $v = x\sqrt{f(x)}$. Inversement, pour tout $x \in \mathbb{R}$, $(u(x), v(x))$ appartient à la frontière de l'ensemble $\mathfrak{A}$.
4. (a) Si $f(x) = \exp\{x^2/2\}$ (en omettant la constante puisqu'elle ne sert pas), la frontière de l'ensemble $\mathfrak{A}$ est donnée par

$$\{(\exp\{x^2/4\}, x \exp\{x^2/4\}); x \in \mathbb{R}\}$$

représenté sur la figure ??.

On a alors $\exp\{x^2/4\} \leq 1$ et le maximum de $b(x) = x \exp\{x^2/4\}$ est atteint pour

$$b'(x) = (1 - 2x^2/4) \exp\{x^2/4\} = 0 \quad \text{soit} \quad x = \sqrt{2}$$

- (b) Par conséquent, simuler des variables uniformes sur le carré $(0, 1) \times (-\sqrt{2}, \sqrt{2})$ produit des simulations normales :

```
K <- 1e+06
U <- runif(K)
V <- runif(K, -sqrt(2), sqrt(2))
X <- (V/U) [U^2 < exp(-V^2/U^2/2)]
```

- (c) Le rapport de la surface de $\mathfrak{A}$ à celle de $(0, 1) \times (-\sqrt{2}, \sqrt{2})$ étant de

$$\int f(x) dx / 2b(d - c) = 0.73$$

la méthode gaspille une simulation sur 4, donc utilise en moyenne 8/3 uniformes pour simuler une normale, alors qu'il existe des algorithmes utilisant en moyenne une seule uniforme pour simuler une variable normale.