Université Paris-Dauphine Année 2017–2018
L3 - Statistique (1) : Estimation

**Examen Partiel Octobre 2018 - Durée 2h00 - Sans documents**

**Important : Suivant les réglements 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 R (6 pts)

Soit $(x_1, \dots, x_n)_{n\geq 1}$ un échantillon de variables aléatoires indépendantes identiquement distribuées de loi exponentielle de paramètre $\lambda = 20$. On rappelle que la densité d'une loi exponentielle de paramètre $\lambda > 0$ est définie pour tout réel $x$ par

$$f_\lambda(x) = \lambda e^{-\lambda x}\mathbf{1}_{x\geq 0}, \quad \text{et que} \quad \mathbb{E}[X] = \frac{1}{\lambda}, \mathbb{V}\mathrm{ar}[X] = \frac{1}{\lambda^2}$$

**1. (1 pt)** Décrivez brièvement chacune des fonctions suivantes : `rexp`, `dexp`, `pexp` et qexp.

**2. Estimateur de $1/\lambda$ (2 pts)** On tape les commandes suivantes dans R

```
> n <- 500
> K <- 500
> lambda <- 20
> y <- numeric(K)
> for (i in 1:K){
    y[i] <- mean(rexp(n, lambda))
  }
> y <- sqrt(n)*(lambda*y-1)
> qqnorm(y)
> qqline(y, col = "firebrick")
```

Parmi les diagrammes Quantiles-Quantiles de la Figure 1, lequel correspond au code ci-dessus ? Justifier.
**Indication :** on pourra déterminer la loi de y.

**Estimateur de $\lambda$ :** on s'intéresse aux propriétés de l'estimateur $\hat{\lambda}$ du paramètre $\lambda$ défini par $\hat{\lambda} = 1/\overline{X}_n$, où $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, .

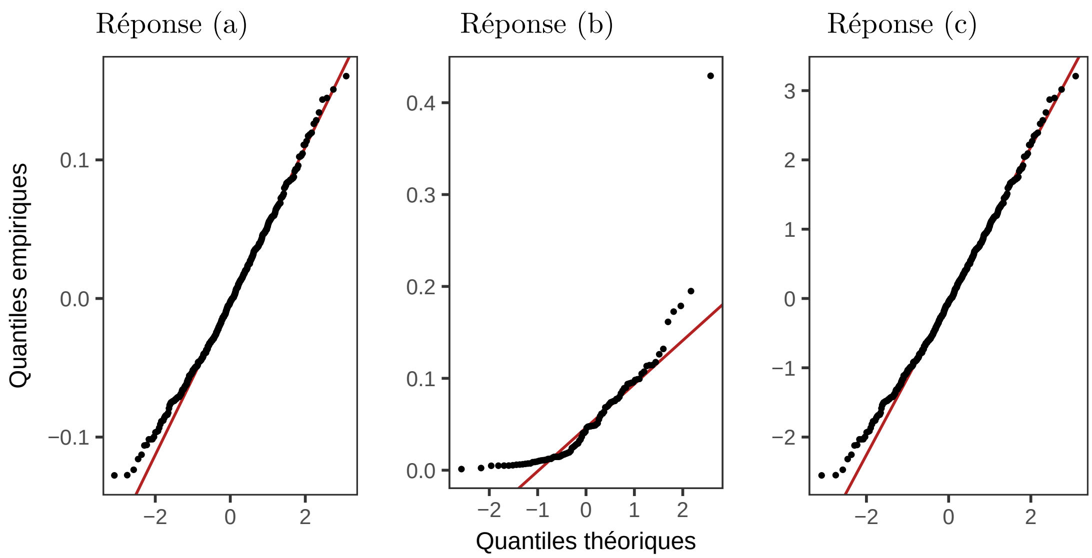



FIGURE 1 – Diagramme Quantiles-Quantiles de la question 2.

**3 (0.5 pt)** On note $\hat{F}_n$ la fonction de répartion empirique. Parmi les fonctions suivantes, laquelle permet de simuler un échantillon de taille $n$ suivant $\hat{F}_n$ ? Justifier.

Réponse (a)

```
> simu <- function(n, x){
    return(sample(x, n, TRUE))
  }
```

Réponse (b)

```
> simu <- function(n, x){
    return(sum(x<=n))
  }
```

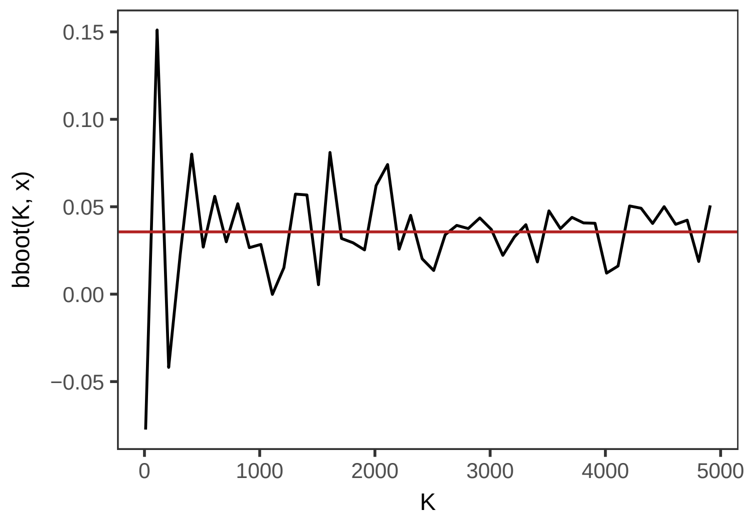
Réponse (c)

```
> simu <- function(n, x){
    return(sample(x, n, FALSE))
  }
```

**5 (1pt)** Expliquer ce que fait la fonction bboot.

```
> bboot <- function(K, x){
    ans <- numeric(K)
    for(i in 1:K){
      ans[i] <- 1/mean(simu(n,x))
    }
    return(mean(ans) - 1/mean(x))
  }
```

**6 (0.5pt)** On trace l'évolution de la fonction bboot en fonction de K. La ligne continue représente la valeur moyenne de bboot. Que pouvez-vous en conclure sur l'estimateur $\hat{\lambda}$ ?




**6 (1pt)** Proposer une fonction `vboot` qui renvoit un estimateur bootstrap de la variance de l'estimateur $\hat{\lambda}$.

## Exercice 2 (6 pts)

On considère un rectangle de longueur $U$ et de largeur $V$. On suppose que $U$ et $V$ sont deux variables aléatoires indépendantes de loi uniforme sur $]0,1[$. On s'intéresse aux variables aléatoires $X = UV$ et $Y = \frac{U}{V}$, modélisant respectivement l'aire et l'allongement du rectangle.

**1 (1 pt)** Calculer $\mathbb{E}(X)$ et $\mathbb{E}(Y)$.

**2 (1 pt)** Montrer que la fonction $h : (u, v) \mapsto \left(uv, \frac{u}{v}\right)$ est une bijection de $]0,1[^2$ dans $D = \{(x, y) \in \mathbb{R}_+^2 : \ 0 < x < 1, \ x < y < \frac{1}{x}\}$.

**3 (1.5 pts)** En déduire la densité du couple $(X, Y)$.

**4 (1.5 pts)** Calculer les densités marginales de $X$ et de $Y$. Ces variables sont-elles indépendantes ?

**5 (1 pt)** Si on ne veut plus privilégier une direction par rapport à l'autre, on peut caractériser l'allongement par $Z = \min\left(\frac{U}{V}, \frac{V}{U}\right)$. Calculer la densité de $Z$.

## Exercice 3 (5 pts)

On dit qu'une variable aléatoire $Y$ suit une loi de Gumbel si elle admet pour densité $f_Y(x) = e^{-x-e^{-x}}$.

**1 (0.5 pt)** Vérifier que $f_Y$ est bien une densité.

**2 (1 pt)** Calculer la fonction de répartition de Y.

**3 (1 pt)** Soit $(X_n)$ une suite de variables aléatoires indépendantes et identiquement distribuées de loi exponentielle de paramètre 1, calculer la fonction de répartition de $M_n = \max(X_1, \ldots, X_n)$ et de $M_n - \ln(n)$.

**4 (1 pt)** Rappeler la définition de la convergence en loi et montrer que $M_n - \ln(n)$ converge en loi vers $Y$.

**5 (1.5 pts)** Rappeler la définition de la convergence en probabilités. Montrer que $\frac{M_n}{\ln(n)}$ converge vers 1 en probabilités.

## Exercice 4 (6 pts)

On rappelle qu'une variable aléatoire $Y$ suit la loi de Cauchy standard si sa densité est $f_Y(y) = \pi^{-1}\{1+y^2\}^{-1}$ et que la fonction caractéristique de cette loi est $\Psi(t) = \mathbb{E}_Y[\exp\{itY\}] = \exp\{-|t|\}$. Une loi de Cauchy de paramètres $\mu$ et $\sigma$ est définie comme la loi de la transformée $\sigma Y + \mu$ d'une variable aléatoire de loi de Cauchy standard.

**1 (0.5 pt)** Donner la densité de la loi de Cauchy de paramètres $\mu$ et $\sigma$ et sa fonction caractéristique.

**2 (0.5 pt)** Montrer que la fonction de répartition de la loi de Cauchy standard est

$$F(x) = \mathbb{P}(X \leq x) = \frac{1}{\pi}\text{arctg}(x) + \frac{1}{2}\,.$$

**3 (1 pt)** Vérifier que la moyenne de la loi de Cauchy de paramètres $\mu$ et $\sigma$ n'est pas définie. Démontrer que la médiane de cette loi vaut $\mu$.

**4 (1 pt)** On considère un échantillon iid $X_1, \ldots, X_n$ de loi de Cauchy de paramètres $\mu$ et $\sigma$ inconnus. Montrer que la loi de la moyenne de l'échantillon $\bar{X}_n = \{X_1 + \ldots + X_n\}/n$ est la même loi de Cauchy. En déduire que la loi des grands nombres ne s'applique pas à cette statistique. Proposer un estimateur convergent de la médiane $\mu$ de la loi de Cauchy.

**5 (1 pt)** Vérifier que la variance de la loi de Cauchy de paramètres $\mu$ et $\sigma$ n'est pas définie. Démontrer que la déviation médiane à la médiane, définie comme la médiane de $|X - \mu|$ vaut $\sigma$. Proposer un estimateur convergent de ce paramètre $\sigma$.

**6 (1 pt)** En considérant la distribution jointe de l'échantillon iid $X_1, \ldots, X_n$ de loi de Cauchy de paramètres $\mu$ et $\sigma$, démontrer que la loi de Cauchy ne peut constituer une famille exponentielle.