

Sauf mention contraire, la mesure de référence est la mesure de Lebesgue.

Formulaire

Loi	Notation	Densité
Exponentielle	$\mathcal{E}(\lambda)$	$f(x \lambda) = \lambda e^{-\lambda x} \mathbb{1}_{x>0}$
Gamma	$\mathcal{G}a(a, b)$	$f(x a, b) = \frac{1}{\Gamma(a)} x^{a-1} e^{-bx} b^a \mathbb{1}_{x>0}$
Normale	$\mathcal{N}(\mu, \sigma^2)$	$f(x \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} \mathbb{1}_{x \in \mathbb{R}}$

French – English Lexicon

- échantillon : *sample*
- famille exponentielle : *exponential family*
- espace naturel des paramètres : *natural parameter space*
- fonction génératrice des moments : *moment-generating function*
- i.i.d.* : *independent and identically distributed*

Exercice 1

..... / 5.5

Dans cet exercice il vous est demandé de donner la ou les bonnes réponses. Seules les réponses justifiées seront validées. Il n'y a pas de points négatifs, mais toute réponse fausse conduit à une note nulle.

1. On considère f définie, pour $x \in \mathbb{R}$, par $f(x) = (x + 0.5) \mathbb{1}_{\{x \in]-1, 1[\}}$. Pour la mesure de Lebesgue

- (a) $x \mapsto 2f(x)$ est une densité (c) $x \mapsto f(x)/2$ est une densité
(b) f est une densité (d) Aucune de ces propositions

(d) f n'est pas une fonction positive et ne peut donc pas définir une densité.

2. Soit $n, p \in \mathbb{N}^*$. On considère la variable aléatoire $X : \omega \mapsto 2^n \mathbb{1}_{\omega \in [0, 1/n]}$, pour $\omega \in [0, 1]$. Alors,

- (a) $\mathbb{E}[X^p]$ est non définie (b) $\mathbb{E}[X^p] = 2^{np}/n$ (c) $\mathbb{E}[X^p] = 2^{np}n$ (d) $\mathbb{E}[X^p] = 2^{np-1}/n^2$

(b) On a

$$\mathbb{E}[X^p] = \int_0^1 X(w)^p \mathbb{P}(dw) = \int_0^{\frac{1}{n}} 2^{np} \mathbb{P}(dw) = \frac{2^{np}}{n}.$$

3. Soit un jeu de données $\mathbf{x}$ qui ne contient que les valeurs possibles suivantes : $\mathbf{aa}$, $\mathbf{aA}$, $\mathbf{AA}$. Pour représenter la distribution de $\mathbf{x}$, on utilise la commande

- (a) `barplot(table(x)/length(x), names.arg = c(aa, aA, AA))`
(b) `hist(x, xlab = c(aa, aA, AA))`
(c) `hist(table(x)/length(x), xlab = c(aa, aA, AA))`
(d) `barplot(x, names.arg = c(aa, aA, AA))`

(a) La commande `barplot` permet de tracer la distribution d'une variable aléatoire discrète en spécifiant la hauteur (correspondant à l'effectif ou la fréquence) de chacune des modalités. La fonction `table` permet

d'obtenir le nombre de fois où chacune des modalités apparaît dans le jeu de données.

4. Soit X une variable aléatoire réelle de densité définie pour $x \in [0, 1]$ par $f(x) = (\pi\sqrt{x(1-x)})^{-1}$.

- (a) $\mathbb{E}[X] = 2$ (b) $\mathbb{E}[X]$ est non définie (c) $\mathbb{E}[X] = 1/2$ (d) $\mathbb{E}[X] = 0$

(c) $\mathbb{E}[X]$ est bien définie car $x \mapsto xf(x) = \sqrt{x/(1-x)}$ est intégrable en 1 (critère de Riemann). X et $1 - X$ ont même loi (invariance de la densité). On en déduit que $\mathbb{E}[X] = \mathbb{E}[1 - X]$. D'où le résultat.

5. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires telles que pour tout $\varepsilon > 0$

$$\sum_{n \in \mathbb{N}} \mathbb{P}[|X_n - X| > \varepsilon] < +\infty.$$

- (a) X_n converge en loi vers X . (c) X_n converge en probabilité vers X .
(b) X_n converge dans L^p vers X . (d) X_n converge presque sûrement vers X .

(a, c, d) Le lemme de Borel Cantelli donne que X_n converge presque sûrement vers X . On utilise ensuite que la convergence presque-sûre implique les convergences en probabilité et en loi.

6. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires *i.i.d.* telle que $\mathbb{E}[X_1^2] < +\infty$. On pose

$$Y_n = \frac{1}{2n} \sum_{k=1}^n (X_{2k-1} - X_{2k})^2.$$

- (a) $\mathbb{E}[Y_n] = \mathbb{V}\text{ar}[X_1]$ (d) $\sqrt{n}(Y_n - \mathbb{E}[X_1^2])$ converge en loi vers $\mathcal{N}(0, \mathbb{V}\text{ar}[X_1])$
(b) Y_n converge presque sûrement vers $\mathbb{V}\text{ar}[X_1]$
(c) Y_n converge presque sûrement vers $\mathbb{E}[X_1^2]$

(a, b) $((X_{2k-1} - X_{2k})^2)_{k \in \mathbb{N}^*}$ est une suite de variables aléatoires *i.i.d.* (comme transformation mesurable des variables *i.i.d.* (X_{2k-1}, X_{2k})). De plus comme X_1 et X_2 sont identiquement distribuées, on a

$$\mathbb{E}[(X_1 - X_2)^2] = \mathbb{E}[X_1^2] + \mathbb{E}[X_2^2] - 2\mathbb{E}[X_1 X_2] = 2\mathbb{E}[X_1^2] - 2\mathbb{E}[X_1]^2 = 2\mathbb{V}\text{ar}[X_1].$$

Par hypothèse $\mathbb{E}[X_1^2] < +\infty$, d'où $\mathbb{E}[X_1] < +\infty$ (inégalité de Jensen) et donc $\mathbb{V}\text{ar}[X_1] < +\infty$. La loi forte des grands nombres permet alors de conclure que Y_n converge presque sûrement vers $\mathbb{V}\text{ar}[X_1]$.

7. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables *i.i.d.* de loi uniforme sur $[0, \theta]$, avec $\theta > 0$. Parmi les variables suivantes, donner celle(s) qui sui(ven)t asymptotiquement la loi $\mathcal{N}(0, 1)$.

- (a) $\sqrt{3n} \left(1 - \theta/(2\bar{X}_n)\right)$ (c) $\sqrt{n} \left(1 - (2\bar{X}_n)/\theta\right)$ (e) Aucune
(b) $2\sqrt{3n} \left(1 - \theta/\bar{X}_n\right)$ (d) $\frac{\sqrt{n}}{3} \left(1 - (2\bar{X}_n)/\theta\right)$

(a) $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires *i.i.d.* de variance finie : $\mathbb{V}\text{ar}[X_1] = \theta^2/12$. Le théorème Central Limite permet donc d'écrire

$$\sqrt{n} \left(\bar{X}_n - \frac{\theta}{2}\right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}\left(0, \frac{\theta^2}{12}\right).$$

En appliquant la méthode delta pour la fonction $g : x \mapsto 1/x$ qui est dérivable sur $\mathbb{R}$ avec $g'(\theta/2) = -4/\theta^2 \neq 0$, on obtient

$$\sqrt{n} \left(\frac{1}{\bar{X}_n} - \frac{2}{\theta} \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N} \left(0, \frac{16}{\theta^4} \frac{\theta^2}{12} \right) \equiv \mathcal{N} \left(0, \frac{4}{3\theta^2} \right). \quad \Leftrightarrow \quad \sqrt{3n} \left(\frac{\theta}{2\bar{X}_n} - 1 \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N} (0, 1).$$

(c, d) suivent asymptotiquement une loi normale mais dont les variances ne sont pas égales à 1.

8. Soit une famille exponentielle définie par $f(x | \theta) = c(\theta)h(x) \exp[\eta(\theta)T(x)]$, pour $x \in \mathbb{R}$ et $\theta \in \mathbb{R}$. On suppose que η est une bijection de $\mathbb{R}$ dans $\mathbb{R}$, de bijection réciproque $\tau \mapsto \eta^{-1}(\tau)$.

- | | |
|---|---|
| (a) $\mathbb{E}[T(X)] = -\frac{d}{d\theta} \log c(\theta)$ | (d) $\mathbb{E}[T(X)] = -\frac{d}{d\theta} c(\theta)$ |
| (b) $\mathbb{E}[X] = -\frac{d}{d\theta} \log c(\theta)$ | (e) $\mathbb{E}[X] = -\frac{d}{d\theta} c(\theta)$ |
| (c) $\mathbb{E}[T(X)] = -\frac{d}{d\tau} [\log(c \circ \eta^{-1})](\eta(\theta))$ | (f) $\mathbb{E}[T(X)] = -\frac{d}{d\tau} [c \circ \eta^{-1}](\eta(\theta))$ |

(c) Pour une famille exponentielle sous forme canonique, de paramètre τ , de fonction de partition $a(\cdot)$ et de statistique $T(\cdot)$, on a le résultat suivant

$$\mathbb{E}[T(X)] = -\frac{d}{d\tau} \log a(\tau).$$

La forme canonique de la famille exponentielle est ici obtenue en faisant le changement de paramétrisation $\tau = \eta(\theta)$. La forme canonique est donc $f(x | \tau) = c[\eta^{-1}(\tau)]h(x) \exp[\tau T(x)]$. On a donc $a(\tau) = c \circ \eta^{-1}(\tau)$ et

$$\mathbb{E}[T(X)] = -\frac{d}{d\tau} [\log(c \circ \eta^{-1})](\eta(\theta)).$$

(a) est faux car $c(\cdot)$ n'est pas la fonction de partition d'une forme canonique

9. Soit $k \in \mathbb{N}^*$. On considère $f(\cdot | v)$ la densité paramétrée par $v \in \mathbb{R}$ et définie sur $\mathbb{R}^n$ par

$$f(x | v) = C(v) \exp \left[-\sum_{i=1}^n (x_i - v)^k \right].$$

- (a) $f(\cdot | v)$ ne définit pas une famille exponentielle.
- (b) $f(\cdot | v)$ forme une famille exponentielle de dimension k .
- (c) $f(\cdot | v)$ forme une famille exponentielle de dimension n .
- (d) $f(\cdot | v)$ forme une famille exponentielle de dimension $k - 2$.
- (e) $f(\cdot | v)$ forme une famille exponentielle de dimension $k - 1$.

(e) En utilisant le binôme de Newton, on peut écrire la densité sous la forme

$$f(x | v) = C(v) \exp \left((-1)^{k+1} n v \right) \exp \left(-\sum_{i=1}^n x_i^k \right) \exp \left(-\sum_{i=1}^n \sum_{j=1}^{k-1} (-v)^{k-j} x_i^j \right).$$

On en déduit que $f(\cdot | v)$ forme une famille exponentielle de dimension $k - 1$.

10. On enregistre dans un vecteur x le temps passé par n consommateurs sur un site de vente en ligne. On considère que le modèle statistique est une loi gamma de paramètre $(1, 2)$. Si dans le contexte de cette expérience je souhaite obtenir la valeur de la fonction de répartition empirique au point t , j'utilise la commande

- | | |
|----------------------------------|--|
| (a) <code>mean(x) <= t</code> | (d) <code>mean(x <= t)</code> |
| (b) <code>pgamma(t, 1, 2)</code> | (e) <code>sum(x <= t)</code> |
| (c) <code>qgamma(t, 1, 2)</code> | (f) <code>mean(pgamma(x, 1, 2) <= t)</code> |

(d) La fonction de répartition empirique correspond à la moyenne empirique des variables $\mathbb{1}_{\{X_1 \leq t\}}, \dots, \mathbb{1}_{\{X_n \leq t\}}$.

Exercice 2

..... /7

Soit X une variable aléatoire de loi de Kumaraswamy $K(a, b)$ de paramètre $(a, b) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$, dont la densité est donnée pour $x \in \mathbb{R}$ par

$$f(x | a, b) = abx^{a-1}(1 - x^a)^{b-1}\mathbb{1}_{[0,1]}(x).$$

1. Calculer la fonction de répartition de X , notée F .

Pour tout $t \in [0, 1]$,

$$F(t) = ab \int_0^t x^{a-1}(1 - x^a)^{b-1} dx = \left[-(1 - x^a)^b \right]_0^t = 1 - (1 - t^a)^b.$$

2.a Montrer que si U est une variable aléatoire de loi uniforme sur $[0, 1]$ alors $Y = (1 - U^{1/b})^{1/a}$ est une variable aléatoire qui suit la loi $K(a, b)$.

Pour tout $t \in [0, 1]$,

$$\mathbb{P}[Y \leq t] = \mathbb{P}\left[(1 - U^{1/b})^{1/a} \leq t\right] = \mathbb{P}\left[U \geq (1 - t^a)^b\right] = 1 - \mathbb{P}\left[U \leq (1 - t^a)^b\right] = 1 - (1 - t^a)^b = F(t).$$

2.b Écrire le code R d'une fonction `rkum(n, a, b)` permettant de simuler n réalisations indépendantes de la loi $K(a, b)$. Le code n'utilisera ni boucle `for`, ni boucle `while`.

```
rkum <- function(n, a, b) {
  return((1 - runif(n)^(1/b))^(1/a))
}
```

3. Dans quel(s) cas la densité de f définit une famille exponentielle? Lorsque cela a un sens, écrire la forme canonique et l'espace naturel du ou des paramètres.

- (a) Lorsque a et b sont inconnus.
- (b) Lorsque a est supposé connu et b inconnu.
- (c) Lorsque a est supposé inconnu et b connu.
- (d) Jamais.

La densité peut s'écrire

$$f(x | a, b) = ab \frac{x}{1-x^a} \mathbb{1}_{[0,1]}(x) \exp[a \ln(x) + b \ln(1-x^a)].$$

Dès lors que a est supposé inconnu, la loi $K(a, b)$ ne définit pas une famille exponentielle car on ne peut pas trouver de changement de mesure h indépendant de a .

Si a est supposé connu, on reconnaît une famille exponentielle sous forme canonique avec

$$c(b) = ab, \quad h(x) = \frac{x^{a-1}}{1-x^a} \mathbb{1}_{[0,1]}(x), \quad \eta(b) = b \quad \text{et} \quad T(x) = \ln(1-x^a).$$

L'espace naturel du paramètre b lorsque $a > 0$ est connu est

$$\Theta = \left\{ b \in \mathbb{R} \mid \int_{\mathbb{R}} h(x) \exp(bT(x)) dx = \int_0^1 x^{a-1} (1-x^a)^{b-1} dx < +\infty \right\} =]0, +\infty[,$$

car l'intégrande est intégrable au voisinage de 1 pour $b > 0$. Alternativement, on peut utiliser la caractérisation de l'espace naturel via la fonction de partition de la forme canonique

$$\Theta = \{b \in \mathbb{R} \mid c(b) = ab > 0\} =]0, +\infty[.$$

4. Donner la fonction de répartition et la densité de la variable aléatoire $Y = -\ln(X)$.

La fonction de répartition s'écrit pour $t \in \mathbb{R}_+$

$$F_Y(t) = \mathbb{P}[Y \leq t] = \mathbb{P}[-\ln(X) \leq t] = \mathbb{P}(X \geq e^{-t}) = 1 - F(e^{-t}).$$

D'après 1., on a donc pour tout $t > 0$,

$$F_Y(t) = \left(1 - e^{-at}\right)^b.$$

F_Y est dérivable sur $]0, +\infty[$. On en déduit donc pour tout $t > 0$,

$$f_Y(t) = F'_Y(t) = abe^{-at} \left(1 - e^{-at}\right)^{b-1}.$$

5.a Soient $a > 0$ et X_n de densité $f(\cdot | a, 1+n^{-1})$, $n \in \mathbb{N}^*$. Montrer que $Y_n = -\ln(X_n)$ converge en loi, lorsque n tend vers $+\infty$, vers une loi exponentielle dont on donnera la valeur du paramètre.

Pour montrer la convergence en loi de Y_n , on montre que sa fonction de répartition converge vers la fonction de répartition d'une loi exponentielle. La fonction de répartition de Y_n est donnée pour $t > 0$ par

$$F_{Y_n}(t) = \left(1 - e^{-at}\right)^{1+\frac{1}{n}} = \exp \left[\left(1 + \frac{1}{n}\right) \ln \left(1 - e^{-at}\right) \right].$$

On a donc tout $t > 0$

$$F_{Y_n}(t) \xrightarrow{n \rightarrow \infty} 1 - e^{-at}.$$

La limite correspond à la fonction de répartition de la loi $\mathcal{E}(a)$.

5.b Écrire un code R pour illustrer numériquement la convergence en loi de Y_n .

Pour illustrer numériquement la convergence, on peut tracer le diagramme quantile-quantile (*qq-plot*) des quantiles empiriques d'un échantillon de m réalisations *i.i.d.* de Y_n en fonction des quantiles de la loi exponentielle de paramètre a . Voici le code pour $m = 1000$ et $n = 10000$.

```
m <- 1000
n <- 10000
# --- Simulation de Y_n
y_n <- -log(rkum(m, a + 1/n))
# --- Diagramme quantile-quantile
plot(qexp(ppoints(100), a), quantile(y_n, ppoints(100)))
abline(0, 1, col = "firebrick")
```

Alternativement, on peut comparer la fonction de répartition de Y_n avec la fonction de répartition de la loi exponentielle $\varepsilon(a)$.

```
plot(x <- seq(0, 20, .01), pexp(x, a), type = "l")
lines(x, pexp(x, a)^(1+1/n), col = "mediumseagreen")
```

6. Soit $b > 0$, donner la fonction de répartition de la variable $Z = -\ln(1 - X)$. En déduire une suite de variables aléatoires Z_n , $n \in \mathbb{N}^*$, qui sont une transformation de variables X_n de densité f dont on précisera les paramètres et qui converge en loi, lorsque n tend vers $+\infty$, vers la loi exponentielle de paramètre b .

La fonction de répartition de Z est donnée pour tout $t > 0$

$$F_Z(t) = \mathbb{P}[-\ln(1 - X) \leq t] = \mathbb{P}[1 - X \geq e^{-t}] = \mathbb{P}[X \leq 1 - e^{-t}] = 1 - \left[1 - (1 - e^{-t})^a\right]^{b-1}.$$

On pose X_n de densité $f(\cdot | 1 + \frac{1}{n}, b)$ et $Z_n = -\ln(1 - X_n)$. La fonction de répartition de Z_n est, pour $t > 0$,

$$F_{Z_n}(t) = 1 - \left[1 - (1 - e^{-t})^{1+\frac{1}{n}}\right]^b.$$

Pour tout $t > 0$, $1 - e^{-t} \in]0, 1[$ et $\lim_{n \rightarrow +\infty} (1 - e^{-t})^{1+\frac{1}{n}} = 1 - e^{-t}$. Par continuité de la fonction $x \mapsto 1 - (1 - x)^b$ sur $]0, 1[$, F_{Z_n} converge bien vers la fonction de répartition d'une loi exponentielle de paramètre $b > 0$.

Exercice 3

..... /8

Lorsque Y suit une loi gamma $\mathcal{G}(\alpha, \beta)$ (c.f. Formulaire), $(\alpha, \beta) \in \Theta \subset \mathbb{R}^2$, on dit que $X = 1/Y$ suit une loi inverse gamma de paramètres α, β , notée $\text{Inv} - \mathcal{G}(\alpha, \beta)$. Dans cet exercice, on suppose que **les deux paramètres α and β sont inconnus**.

1. Montrer que la densité de X est

$$f_X(x | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(\frac{-\beta}{x}\right) \mathbb{1}_{\{x>0\}}.$$

Le changement de variable $g : x \mapsto 1/x$ réalise un $\mathcal{C}^1$ -difféomorphisme de $\mathbb{R}_+^*$ dans $\mathbb{R}_+^*$. On obtient donc que la densité de $X = g(Y)$ est

$$f_X(x) = f_Y(g^{-1}(x)) \left| \frac{d}{dx} g^{-1}(x) \right| = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{x}\right)^{\alpha-1} \exp\left(\frac{-\beta}{x}\right) \frac{1}{x^2} \mathbb{1}_{\{1/x>0\}} = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(\frac{-\beta}{x}\right) \mathbb{1}_{\{x>0\}}.$$

2. Montrer que la famille ainsi définie est une famille exponentielle et calculer l'espace naturel des paramètres. La famille ainsi définie est-elle sous forme canonique? Minimale? Régulière?

On peut écrire la densité

$$f_X(x | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{x} \exp\left[-\alpha \log(x) - \frac{\beta}{x}\right] \mathbb{1}_{\{x>0\}} = c(\alpha, \beta) h(x) \exp[\langle \eta(\alpha, \beta), T(x) \rangle]$$

avec

$$c(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}, \quad h(x) = \frac{1}{x} \mathbb{1}_{\{x>0\}}, \quad \eta(\alpha, \beta) = (\alpha, \beta)^\top, \quad T(x) = (-\log(x), -1/x)^\top.$$

Il s'agit donc d'une famille exponentielle sous forme canonique. Son espace naturel des paramètres est

$$\Theta = \left\{ \tau = (\tau_1, \tau_2) \in \mathbb{R}^2 \mid \int_{\mathbb{R}} h(x) \exp(\langle \tau, T(x) \rangle) dx = \int_0^{+\infty} \frac{1}{x^{1+\tau_1}} \exp\left(\frac{-\tau_2}{x}\right) dx < +\infty \right\}$$

La fonction $x \mapsto x^{-(1+\tau_1)} \exp(-\tau_2/x)$ est intégrable en 0 si $\tau_2 > 0$ et est intégrable en $+\infty$ pour $1 + \tau_1 > 1$. On en déduit donc que $\Theta =]0, +\infty[\times]0, +\infty[$. Θ étant un ouvert de $\mathbb{R}^2$, la famille est régulière.

Soit $(a, b, c) \in \mathbb{R}^* \times \mathbb{R}^* \times \mathbb{R}$. L'ensemble $\{x \in \mathbb{R}, a \log(x) + b/x = c\}$ contient au plus 2 points. Il est donc de mesure nulle pour la mesure de Lebesgue, i.e., $\mathbb{P}[a \log(X) + b/X = c] = 0$. La famille est donc minimale.

3.a Calculer $\mathbb{E}[1/X]$ en fonction de α, β .

La famille étant sous forme canonique, on peut directement utiliser les dérivées partielles de la log fonction de partition $\log c(\alpha, \beta)$ pour trouver $\mathbb{E}[T(X)]$. En prenant la dérivée partielle par rapport à β , on obtient

$$\mathbb{E}\left[-\frac{1}{X}\right] = -\frac{\partial}{\partial \beta} \log c(\alpha, \beta), \quad \text{donc} \quad \mathbb{E}\left[\frac{1}{X}\right] = \frac{\partial}{\partial \beta} [\alpha \log \beta - \log \Gamma(\alpha)] = \frac{\alpha}{\beta}.$$

3.b Calculer $\mathbb{E}[\log(X)]$ en fonction de α, β et de la fonction digamma définie pour $x > 0$ par

$$\Psi(x) = \Gamma'(x)/\Gamma(x).$$

En prenant la dérivée partielle par rapport à α , on obtient

$$\mathbb{E}[-\log(X)] = -\frac{\partial}{\partial \alpha} \log c(\alpha, \beta) \quad \text{donc} \quad \mathbb{E}[\log(X)] = \frac{\partial}{\partial \alpha} [\alpha \log \beta - \log \Gamma(\alpha)] = \log(\beta) - \Psi(\alpha).$$

4.a Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables *i.i.d.* suivant la loi $\text{Inv} - \mathcal{Ga}(\alpha, 1)$. Montrer qu'il existe une suite $(a_n)_{n \in \mathbb{N}^*}$ telle que

$$a_n \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{et} \quad a_n \log \left(\prod_{i=1}^n X_i \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} -\Psi(\alpha).$$

Les $(\log(X_n))_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires *i.i.d.* comme transformation mesurable des $(X_n)_{n \in \mathbb{N}^*}$ qui sont *i.i.d.*. De plus,

$$\mathbb{E}[\log(X_1)] = \log(1) - \Psi(\alpha) = -\Psi(\alpha) < \infty$$

La loi faible des grands nombres donne

$$\frac{1}{n} \sum_{i=1}^n \log(X_i) = \frac{1}{n} \log \left(\prod_{i=1}^n X_i \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} -\Psi(\alpha).$$

En prenant pour $n \in \mathbb{N}^*$, $a_n = 1/n$, on obtient le résultat voulu.

4.b On admet que $\text{Var}(\log(X_1)) = \sigma^2 < \infty$. Trouver une suite $(c_n)_{n \in \mathbb{N}^*}$ dont l'expression ne dépend ni de σ^2 ni de $\Psi(\alpha)$ telle que

$$c_n \left(\frac{1}{n} \sum_{i=1}^n \log(X_i) + \Psi(\alpha) \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1).$$

On a déjà établi que $(\log(X_n))_{n \in \mathbb{N}^*}$ sont *i.i.d.*. De plus, l'hypothèse de l'énoncé donne $\text{Var}(\log(X_1)) = \sigma^2 < \infty$. Le théorème central limite donne

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \log(X_i) + \Psi(\alpha) \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \sigma^2).$$

Posons

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n \log(X_i)^2 - \left(\frac{1}{n} \sum_{i=1}^n \log(X_i) \right)^2.$$

Les $(\log(X_n)^2)_{n \in \mathbb{N}^*}$ sont *i.i.d.* comme transformation mesurable des $(X_n)_{n \in \mathbb{N}^*}$ qui sont *i.i.d.*. Par ailleurs, $\mathbb{E}[\log(X_1)^2] = \sigma^2 + \Psi(\alpha)^2 < \infty$. Ainsi on obtient avec la loi faible des grands nombres,

$$\frac{1}{n} \sum_{i=1}^n \log(X_i)^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \sigma^2 + \Psi(\alpha)^2.$$

De plus, $g : x \mapsto x^2$ est continue sur $\mathbb{R}$ et par le théorème d'application continue et la question 4.a,

$$\left(\frac{1}{n} \sum_{i=1}^n \log(X_i) \right)^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \Psi(\alpha)^2.$$

Ainsi,

$$\hat{\sigma}^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \sigma^2.$$

Finalement, $h : x \mapsto 1/\sqrt{x}$ étant continue pour tout $x > 0$, le théorème d'application continue donne

$$\frac{1}{\sqrt{\hat{\sigma}^2}} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \frac{1}{\sigma} \in \mathbb{R}.$$

En appliquant le lemme de Slutsky, on obtient alors

$$\sqrt{\frac{n}{\hat{\sigma}^2}} \left(\frac{1}{n} \sum_{i=1}^n \log(X_i) + \Psi(\alpha) \right) \xrightarrow[n \rightarrow \infty]{d} \frac{1}{\sigma} \mathcal{N}(0, \sigma^2) \equiv \mathcal{N}(0, 1).$$

On obtient donc le résultat avec

$$c_n = \sqrt{\frac{n}{\hat{\sigma}^2}}.$$

4.c Pour $n \in \mathbb{N}^*$ et $\varepsilon \in]0, 1[$, en déduire la valeur de t_n en fonction de c_n et ε telle que

$$\mathbb{P} \left[-t_n - \frac{1}{n} \sum_{i=1}^n \log(X_i) \leq \Psi(\alpha) \leq t_n - \frac{1}{n} \sum_{i=1}^n \log(X_i) \right] \xrightarrow[n \rightarrow \infty]{} \varepsilon.$$

D'après la question précédente, on sait que pour tout $t \in \mathbb{R}$, on a

$$\mathbb{P} \left[-t \leq c_n \left(\frac{1}{n} \sum_{i=1}^n \log(X_i) + \Psi(\alpha) \right) \leq t \right] \xrightarrow[n \rightarrow \infty]{} \mathbb{P}[-t \leq Z \leq t], \quad \text{avec } Z \sim \mathcal{N}(0, 1).$$

Autrement dit,

$$\mathbb{P} \left[-\frac{t}{c_n} - \frac{1}{n} \sum_{i=1}^n \log(X_i) \leq \Psi(\alpha) \leq \frac{t}{c_n} - \frac{1}{n} \sum_{i=1}^n \log(X_i) \right] \xrightarrow[n \rightarrow \infty]{} \mathbb{P}[-t \leq Z \leq t].$$

Il suffit donc de choisir t tel que $\mathbb{P}[-t \leq Z \leq t] = \varepsilon$ et de prendre $t_n = t/c_n$. Si l'on note Φ la fonction de répartition de la loi $\mathcal{N}(0, 1)$, on cherche t qui est solution de

$$\mathbb{P}[-t \leq Z \leq t] = 2\Phi(t) - 1 = \varepsilon \quad \Leftrightarrow \quad t = \Phi^{-1} \left(\frac{\varepsilon + 1}{2} \right) := q_{\frac{1+\varepsilon}{2}}.$$

Il faut donc prendre pour t le quantile d'ordre $(1 + \varepsilon)/2$ de la loi $\mathcal{N}(0, 1)$. D'où le résultat.

- 5.a** On souhaite trouver une approximation de $\Psi(\alpha)$ pour $\alpha = 63$. Donner un code R qui,
1. pour $n = 1000$, permet de simuler un échantillon $\mathbf{x} = (x_1, \dots, x_n)$ suivant la loi $\text{Inv} - \mathcal{G}\mathbf{a}(63, 1)$ à partir du générateur de la loi gamma ;
 2. pour cet échantillon $\mathbf{x}$ renvoie un intervalle de confiance asymptotique pour $\Psi(63)$ au niveau $\varepsilon = 90\%$.
- Le code n'utilisera ni boucle `for`, ni boucle `while`.

```
n <- 1000
alpha <- 63
# --- Simulation de x
x <- 1/rgamma(n, shape = alpha, rate = 1)
# --- Intervalle de confiance asymptotique
epsilon <- 0.9
# ----- Estimation de la variance
y <- log(x)
sigma2_hat <- mean(y^2) - mean(y)^2
# Remarque : asymptotiquement sigma2_hat est équivalent à var(y)
# ----- Calcul de t_n
t_n <- qnorm(.5 * (1 + epsilon)) * sqrt(sigma2_hat/n)
# ----- Calcul de l'intervalle
c(-1, 1) * t_n - mean(y)

[1] 4.126367 4.139655
```

- 5.b** (Bonus) Donner un code R d'une fonction donnant l'ensemble des valeurs de la suite $(a_k \log(\prod_{i=1}^k x_i))_{k \in \mathbb{N}^*}$ pour un vecteur $\mathbf{x}$ donné. Le code n'utilisera ni boucle `for`, ni boucle `while`.

```
# --- Option n°1
cumsum(log(x))/(1:length(x))
# --- Option n°2
log(cumprod(x))/(1:length(x))
```