Correction - AP 20

Réduction - Loi de Pareto

Exercice 1 (HEC 2012)

1. Soit $u = (x, y, z) \in \mathbb{R}^3$

$$u \in Ker\,(f) \Leftrightarrow M \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0 \Leftrightarrow \begin{cases} \frac{1}{m}y + \frac{1}{m^2}z = 0 \\ mx + \frac{1}{m}z = 0 \\ m^2x + my = 0 \end{cases} \underset{m\neq 0}{\Leftrightarrow} \begin{cases} z = -my \\ 2y = 0 \\ x = -\frac{1}{m}y \end{cases} \Leftrightarrow x = y = z = 0.$$

Ainsi, $Ker\,(f) = \{0\}$. Donc f est injective de $\mathbb{R}^3$ dans $\mathbb{R}^3$. Comme l'espace de départ et d'arrivée sont de même dimension, f bijective. Donc M est inversible et $Im\,(f) = \mathbb{R}^3$.

2. (a) On a :

$$M^2 = \begin{pmatrix} 0 & 1/m & 1/m^2 \\ m & 0 & 1/m \\ m^2 & m & 0 \end{pmatrix} \begin{pmatrix} 0 & 1/m & 1/m^2 \\ m & 0 & 1/m \\ m^2 & m & 0 \end{pmatrix} = \begin{pmatrix} 2 & \frac{1}{m} & \frac{1}{m^2} \\ m & 2 & \frac{1}{m} \\ m^2 & m & 2 \end{pmatrix} = M + 2I.$$

(b) Donc $M^2 - M - 2I = 0$ et le polynôme $P = X^2 - X - 2$ est annulateur de M

(c) Donc, si α est valeur propre de M alors $P\,(\alpha) = 0$ et donc $\alpha = -1$ et $\alpha = 2$.
Ainsi les seules valeurs propres **possibles** sont donc $\alpha = -1$ et 2. Reste à vérifier si elles le sont ou pas :

- Pour $\alpha = -1$: Soit $(x, y, z) \in \mathbb{R}^3$

$$(M + I) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0 \quad \Leftrightarrow \quad \begin{cases} x + \frac{1}{m}y + \frac{1}{m^2}z = 0 \\ mx + y + \frac{1}{m}z = 0 \\ m^2x + my + z = 0 \end{cases}$$
$$\Leftrightarrow \quad m^2x + my + z = 0 \Leftrightarrow z = -m^2x - my.$$

Les solutions sont donc $E_1(M) = \text{Vect}\left((1, 0, -m), (0, 1, -m)\right) \neq \{0\}$
Donc 1 est bien valeur propre et est associé à $\text{Vect}\left((1, 0, -m), (0, 1, -m)\right)$.
La famille $\left((1, 0, -m), (0, 1, -m)\right)$ est donc génératrice de $E_1(M)$ et elle est libre (deux vecteurs non colinéaires). C'est donc une base de $E_1(M)$ et $\dim\left(E_1(M)\right) = 2$.

- Pour $\alpha = 2$: Soit $(x, y, z) \in \mathbb{R}^3$.

$$(M - 2I) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0 \quad \Leftrightarrow \quad \begin{cases} -2x + \frac{1}{m}y + \frac{1}{m^2}z = 0 \\ mx - 2y + \frac{1}{m}z = 0 \\ m^2x + my - 2z = 0 \end{cases}$$
$$\Leftrightarrow \begin{cases} -2x + \frac{1}{m}y + \frac{1}{m^2}\left(\frac{m^2}{2}x + \frac{m}{2}y\right) = 0 \\ mx - 2y + \frac{1}{m}\left(\frac{m^2}{2}x + \frac{m}{2}y\right) = 0 \\ z = \frac{m^2}{2}x + \frac{m}{2}y \end{cases}$$
$$\Leftrightarrow \begin{cases} -\frac{3}{2}x + \frac{3}{2m}y = 0 \\ \frac{3}{2}mx - \frac{3}{2}y = 0 \\ z = \frac{m^2}{2}x + \frac{m}{2}y \end{cases}$$
$$\Leftrightarrow \begin{cases} 0 = 0 \\ y = mx \\ z = \frac{m^2}{2}x + \frac{m}{2}y \end{cases}$$
$$\Leftrightarrow \begin{cases} y = mx \\ z = m^2x \end{cases}$$

Les solutions sont donc $E_2(M) = \text{Vect}\left((1, m, m^2)\right)$.
Donc 2 est bien valeur propre et est associé à $\text{Vect}\left((1, m, m^2)\right)$ et $\left((1, m, m^2)\right)$ en est une famille libre (un seul vecteur non nul) et génératrice, donc une base. Ainsi, $\dim\left(E_2(M)\right) = 1$.

Les valeurs propres de M sont donc -1 et 2.
Par concaténation des bases des sous-espaces propres $E_{-1}(M)$ et $E_2(M)$, on obtient une famille libre de trois vecteurs propres de M. Comme son cardinal est égal à la dimension de $\mathscr{M}_{3,1}(\mathbb{R})$, c'est une base de $\mathscr{M}_{3,1}(\mathbb{R})$ et M est donc diagonalisable.

3. En posant $R = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & m \\ -m & -m & m^2 \end{pmatrix}$ inversible et $D = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$, on a $M = RDR^{-1}$ et $M^n = RD^nR^{-1}$ avec D diagonale, on a directement D^n.

   Il resterait donc à calculer explicitement R^{-1}.

4. (a) On passe par les matrices associées :
   - p a pour matrice :
   $$P = \frac{1}{3}(M + I)$$
   - q a pour matrice :
   $$Q = -\frac{1}{3}(M - 2I)$$

   On a :
   $$\begin{aligned} QP = PQ &= -\frac{1}{9}(M+I)(M-2I) = -\frac{1}{9}(M^2 - M - 2I) = 0 \\ P^2 &= \frac{1}{9}(M+I)(M+I) = \frac{1}{9}(M^2 + 2M + I) = \frac{1}{9}(M + 2I + 2M + I) = \frac{1}{3}(M+I) = P \\ Q^2 &= \frac{1}{9}(M-2I)(M-2I) = \frac{1}{9}(M^2 - 4M + 4I) = \frac{1}{9}(M + 2I - 4M + 4I) = -\frac{1}{3}(M - 2I) = \end{aligned}$$
   On démontre alors par récurrence que $P^n = P$ et $Q^n = Q$ pour $n \geq 1$. Finalement :
   $$p \circ q = q \circ p = 0, \quad p^n = p \text{ et } q^n = q \text{ pour } n \in \mathbb{N}^* \quad \text{et} \quad p^0 = q^0 = I \text{ pour } n = 0.$$

   (b) Toujours avec les matrices. On reconstitue M à partir de P et Q : $M = 2P - Q$ et comme $PQ = QP$ alors par le binôme (pour $n \geq 2$) :
   $$\begin{aligned} M^n &= (2P - Q)^n \\ &= \sum_{k=0}^{n} \binom{n}{k} 2^k (-1)^{n-k} P^k Q^{n-k} \\ &= \sum_{k=1}^{n-1} \binom{n}{k} 2^n (-1)^{n-k} PQ + (-1)^n Q + 2^n P \end{aligned}$$
   Comme $PQ = 0$, on a finalement :
   $$M^n = (-1)^n Q + 2^n P \quad \text{et donc} \quad f^n = (-1)^n q + 2^n p.$$
   Ce résultat est encore vrai pour $n = 0$ (car $p + q = Id$) et $n = 1$ (car $-q + 2p = f$). On a donc la formule pour tout $n \in \mathbb{N}$.

   (c) A partir de la formule obtenue précédemment,
   $$\begin{aligned} M^n &= (-1)^n Q + 2^n P \\ &= -\frac{(-1)^n}{3}(M - 2I) + \frac{2^n}{3}(M + I) \\ &= \frac{1}{3}\left(2^n - (-1)^n\right) M + \frac{1}{3}\left(2(-1)^n + 2^n\right) I. \end{aligned}$$
   On pose $a_n = \frac{1}{3}\left(2^n - (-1)^n\right)$ et $b_n = \frac{1}{3}\left(2(-1)^n + 2^n\right)$.

(d) Il faut vérifier si le produit par la candidate inverse vaut I :

$$\begin{aligned}
& \left[\frac{1}{3}\left(2^n-(-1)^n\right)M+\frac{1}{3}\left(2(-1)^n+2^n\right)I\right]\left[\frac{1}{3}\left(2^{-n}-(-1)^{-n}\right)M+\frac{1}{3}\left(2(-1)^{-n}+2^{-n}\right)I\right] \\
= \; & \frac{1}{9}\left(2^n-(-1)^n\right)\left(2^{-n}-(-1)^{-n}\right)M^2+\frac{1}{9}\left(2^n-(-1)^n\right)\left(2(-1)^{-n}+2^{-n}\right)M \\
& +\frac{1}{9}\left(2(-1)^n+2^n\right)\left(2^{-n}-(-1)^{-n}\right)M+\frac{1}{9}\left(2(-1)^n+2^n\right)\left(2(-1)^{-n}+2^{-n}\right)I \\
= \; & \frac{1}{9}\left(1-(-2)^n-\left(-\frac{1}{2}\right)^n+1\right)(M+2I)+\frac{1}{9}\left(2(-2)^n+1-2-\left(\frac{-1}{2}\right)^n\right)M \\
& +\frac{1}{9}\left(2\left(-\frac{1}{2}\right)^n-2+1-(-2)^n\right)M+\frac{1}{9}\left(4+2\left(\frac{-1}{2}\right)^n+2(-2)^n+1\right)I \\
= \; & I \qquad \text{(après simplification)}
\end{aligned}$$

Donc la formule est encore vrai pour n négatif et elle est donc vraie pour tout $n \in \mathbb{Z}$.

Exercice 2 (ESSEC I 2007)

1. (a) On reconnaît l'évènement contraire de la fonction de répartition :

$$P(X>x)=1-P(X\leq x)=1-F_X(x)=\begin{cases} 1 & \text{si } x<0 \\ e^{-\lambda x} & \text{si } x\geq 0 \end{cases}$$

(b) Pour x et y strictement positifs, comme $x+y\geq x$, $(X>x+y)\subset(X>x)$ donc :

$$\begin{aligned}
P_{(X>x)}(X>x+y) &= \frac{P((X>x)\cap(X>x+y))}{P(X>x)}=\frac{P(X>x+y)}{P(X>x)} \\
&= \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}}=e^{-\lambda y}=P(X>y).
\end{aligned}$$

Si X modélise la durée de vie d'un phénomène, le fait d'avoir durée au moins x (c'est-à-dire sachant $(X>x)$) ne change pas la loi de la durée de vie restante.
La durée de vie ne dépend pas du vécu. On peux donc la qualifier de "sans vieillissement".

2. (a) Par linéarité de l'espérance, S_n admet une espérance et :

$$E(S_n)=\sum_{k=1}^{n}E(X_k)=\frac{n}{\lambda}$$

et comme les (X_k) sont indépendantes,

$$V(S_n)=\sum_{k=1}^{n}V(X_k)=\frac{n}{\lambda^2}.$$

(b) On le prouve par récurrence sur n, en se servant de la propriété donnée par l'énoncé :

Ini. $S_1=X_1\hookrightarrow\mathcal{E}(\lambda)$ donc S_1 a pour densité :

$$f_1(t)=\begin{cases} 0 & \text{si } t<0 \\ \frac{\lambda^1}{(1-1)!}e^{-\lambda t}t^{1-1} & \text{si } t\geq 0 \end{cases}$$

Héré. Soit $n\in\mathbb{N}^*$. Supposons que S_n ait pour densité :

$$f_n(t)=\begin{cases} 0 & \text{si } t<0 \\ \frac{\lambda^n}{(n-1)!}e^{-\lambda t}t^{n-1} & \text{si } t\geq 0 \end{cases}$$

Remarquons que $S_{n+1} = S_n + X_{n+1}$, avec S_n et X_{n+1} indépendantes car S_n ne dépend que de $X_1, \ldots, X_n$ (lemme des coalitions). Donc S_{n+1} est à densité et une densité est donnée (d'après l'énoncé) par :

$$f_{n+1}(t) = \int_{-\infty}^{+\infty} f_n(x) f_1(t-x)\ dx$$

avec

$$f_1(t-x) = \begin{cases} 0 & \text{si } x > t \\ \lambda e^{-\lambda(t-x)} & \text{si } x \leq t \end{cases}$$

On en déduit que, si $t < 0$,

$$f_{n+1}(t) = \int_{-\infty}^{+\infty} 0\ dx = 0.$$

D'autre part si $t \geq 0$,

$$\begin{aligned} f_{n+1}(t) &= \int_{-\infty}^{0} 0\ dx + \int_0^t \frac{\lambda^{n+1} x^{n-1}}{(n-1)!} e^{-\lambda x} e^{-\lambda(t-x)}\ dx + \int_t^{+\infty} 0\ dx \\ &= \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda t} \int_0^t x^{n-1}\ dx = \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda t} \left[\frac{x^n}{n}\right]_0^t = \frac{\lambda^{n+1}}{n!} e^{-\lambda t} t^n. \end{aligned}$$

Ccl. Pour tout $n \in \mathbb{N}^*$, S_n a pour densité :

$$f_n(t) = \begin{cases} 0 & \text{si } t < 0 \\ \frac{\lambda^n}{(n-1)!} e^{-\lambda t} t^{n-1} & \text{si } t \geq 0 \end{cases}$$

3. Par relation de Chasles,

$$\int_{-\infty}^{+\infty} f(t)\ dt = \int_{-\infty}^{b} 0\ dt + \int_b^{+\infty} \frac{ab^a}{t^{a+1}}\ dt$$

La première intégrale converge et vaut 0.

La seconde est impropre en $+\infty$ et :

$$\int_b^M \frac{ab^a}{t^{a+1}}\ dt = \left[-\frac{b^a}{t^a}\right]_b^M = -\frac{b^a}{M^a} + \frac{b^a}{b^a} \xrightarrow[M \to +\infty]{} 1.$$

On en déduit que $\int_{-\infty}^{+\infty} f(t)\, dt$ converge et vaut 1.

(Les autres propriétés, évidentes, d'une densité de probabilité, sont admises par l'énoncé).

On considère ensuite l'intégrale :

$$\int_{-\infty}^{+\infty} t f(t)\ dt = \int_{-\infty}^{b} 0\ dt + \int_b^{+\infty} \frac{ab^a}{t^a}\ dt$$

La première intégrale converge absolument et vaut 0; la seconde converge absolument si et seulement si $a > 1$ (intégrale de Riemann), donc X admet une espérance si et seulement si $a > 1$ et :

$$E(X) = \int_b^{+\infty} \frac{ab^a}{t^a}\ dt$$

Or en passant par l'intégrale partielle on obtient :

$$\int_b^M \frac{ab^a}{t^a}\ dt = \left[\frac{ab^a}{(1-a)t^{a-1}}\right]_b^M = \frac{b^a}{(1-a)M^{a-1}} + \frac{ab^a}{(a-1)b^{a-1}} \xrightarrow[M \to +\infty]{} 0 + \frac{ab}{a-1}$$

donc :

$$E(X) = \frac{ab}{a-1}.$$

On considère enfin l'intégrale :

$$\int_{-\infty}^{+\infty} t^2 f(t)\ dt = \int_{-\infty}^{b} 0\ dt + \int_{b}^{+\infty} \frac{ab^a}{t^{a-1}}\ dt$$

La première intégrale converge absolument et vaut 0; la seconde converge absolument si et seulement si $a-1>1 \iff a>2$ (intégrale de Riemann), donc X admet un moment d'ordre deux et donc une variance si et seulement si $a>1$ et :

$$E(X^2) = \int_{b}^{+\infty} \frac{ab^a}{t^{a-1}}\ dt$$

Or en passant par l'intégrale partielle on obtient :

$$\int_{b}^{M} \frac{ab^a}{t^{a-1}}\ dt = \left[\frac{ab^a}{(2-a)t^{a-2}}\right]_b^M = \frac{b^a}{(2-a)M^{a-2}} + \frac{ab^a}{(a-2)b^{a-2}} \xrightarrow[M\to+\infty]{} 0 + \frac{ab^2}{a-2}$$

Enfin par formule de Koenig-Huygens,

$$\begin{aligned} V(X) &= \frac{ab^2}{a-2} - \frac{a^2b^2}{(a-1)^2} = ab^2\left(\frac{1}{a-2} - \frac{a}{(a-1)^2}\right) \\ &= ab^2 \times \frac{(a-1)^2 - a(a-2)}{(a-1)^2(a-2)} = \frac{ab^2}{(a-1)^2(a-2)}. \end{aligned}$$

4. La fonction de répartition de X est $F(x) = \displaystyle\int_{-\infty}^{x} f(t)\ dt$, nulle si $x<b$, et sinon :

$$F(x) = \int_b^x a\frac{b^a}{t^{a+1}}\ dt = \left[-\frac{b^a}{t^a}\right]_b^x = -\left(\frac{b}{x}\right)^a + 1.$$

La fonction de survie de X est donc donnée par :

$$G(x) = \begin{cases} 1 & \text{si } x<b \\ \left(\frac{b}{x}\right)^a & \text{si } x\geq b \end{cases}$$

5. Pour tout réel y positif ou nul et $x>a$ (pour que la probabilité conditionnelle soit définie),

$$\begin{aligned} P_{(X>x)}(X>x+y) &= \frac{P((X>x)\cap(X>x+y))}{P(X>x)} = \frac{P(X>x+y)}{P(X>x)} = \frac{\left(\frac{b}{x+y}\right)^a}{\left(\frac{b}{x}\right)^a} \\ &= \frac{x^a}{(x+y)^a} = \frac{1}{\left(1+\frac{y}{x}\right)^a} \xrightarrow[x\to+\infty]{} 1. \end{aligned}$$

Dans une telle modélisation de la durée de vie, plus on a vécu longtemps, plus la probabilité de vivre davantage augmente. C'est un phénomène dans lequel l'expérience fait gagner en endurance.

6. Pour tout $x\in\mathbb{R}$, par croissance de exp et comme $b>0$,

$$F_Y(x) = P(Y\leq x) = P\left(\ln\left(\frac{X}{b}\right)\leq x\right) = P\left(X\leq be^x\right) = F_X(be^x).$$

Or $be^x\geq b \Leftrightarrow e^x\geq 1 \Leftrightarrow x\geq 0$ (toujours par croissance de exp et comme $b>0$). Donc $F_Y(x)=0$ si $x<0$ et si $x\geq 0$,

$$F_Y(x) = 1-\left(\frac{b}{be^x}\right)^a = 1-e^{-ax}.$$

On reconnaît la fonction de répartition de la loi exponentielle de paramètre a. Donc $Y \hookrightarrow \mathcal{E}(a)$.

7. (a) $\beta \leq x_k$ alors $f_a(x_k) = a\frac{\beta^a}{x_k^{a+1}}$ pour tout k et

$$\mathcal{L}(a) = \frac{a^n \beta^{na}}{\left(\prod\limits_{k=1}^{n} x_k\right)^{a+1}}.$$

Tous les facteurs du produit étant strictement positifs,

$$\ln(\mathcal{L}(a)) = n\ln(a) + na\ln(\beta) - (a+1)\sum_{k=1}^{n}\ln(x_k).$$

(b) i. φ est dérivable sur $]0,+\infty[$ et

$$\varphi'(a) = \frac{n}{a} + n\ln(\beta) - \sum_{k=1}^{n}\ln(x_k) = \frac{n + \left(n\ln(\beta) - \sum\limits_{k=1}^{n}\ln(x_k)\right)a}{a}.$$

1er cas (oublié par l'énoncé) : $\sum\limits_{k=1}^{n}\ln(x_k) - n\ln(\beta) = 0$, alors φ est strictement croissante et n'admet pas de maximum.
On en déduit que l'énoncé aurait du imposer que les x_i ne sont pas tous égaux à β.
On en déduit alors qu'ils sont tous supérieurs ou égaux à β, et l'un d'entre eux au moins est strictement supérieur, donc :

$$\sum_{k=1}^{n}\ln(x_k) > \sum_{k=1}^{n}\ln(\beta) = n\ln(\beta)$$

et le numérateur de φ' est une fonction affine strictement décroissante qui s'annule pour

$$a = \frac{n}{\sum\limits_{k=1}^{n}\ln(x_k) - n\ln(\beta)} = w > 0$$

et comme le dénominateur est strictement positif, φ' est du signe de ce numérateur, donc φ est strictement croissante sur $]0;w]$ et strictement décroissante sur $[w;+\infty[$, donc admet un maximum strict atteint en w.

ii. On a trouvé à la question précédente :

$$w = \frac{n}{\sum\limits_{k=1}^{n}\ln(x_k) - n\ln(\beta)}.$$

iii. On a

$$\ln(\mathcal{L}(a)) = \varphi(a) \iff \mathcal{L}(a) = e^{\varphi(a)}$$

et par croissance de l'exponentielle $\mathcal{L}$ est maximale en w.

(c) i. On a vu (II 4) qu'avec X_k suivant une loi de Pareto de paramètres α et β, la variable $\ln\left(\frac{X_k}{\beta}\right)$ suivait alors une loi $\mathcal{E}(\alpha)$.
Au (I 2.b) on a vu qu'une somme de loi exponentielles indépendantes était à densité et avait pour densité f_n.
Les (X_k) étant indépendantes, $\sum\limits_{k=1}^{n}\ln\left(\frac{X_k}{\beta}\right)$ admet pour densité la fonction f_n avec $\lambda = \alpha$.

ii. Par théorème de transfert, W_n admet une espérance si et seulement si l'intégrale
$$\int_{-\infty}^{+\infty} \frac{n}{t} f_n(t)\ dt$$
est absolument convergente, ce qui est équivalent à sa convergence car la fonction intégrée est positive.
Or par relation de Chasles,
$$\int_{-\infty}^{+\infty} \frac{n}{t} f_n(t)\ dt = \int_{-\infty}^{0} 0\ dt + \int_{0}^{+\infty} \frac{n\alpha^n}{(n-1)!} e^{-\alpha t} t^{n-2} dt$$
La première intégrale converge et vaut 0, la seconde converge car on reconnaît, à une constante multiplicative près, f_{n-1} (et on sait que $n \geq 1$ donc $n-1 \geq 0$), donc W_n admet une espérance et :
$$E(W_n) = \frac{n\alpha}{n-1} \int_{0}^{+\infty} f_{n-1}(t)\ dt = \frac{n\alpha}{n-1}.$$
Enfin on en déduit par linéarité de l'espérance que la variable
$$W'_n = \frac{n-1}{n} W_n$$
admet une espérance, et c'est un estimateur de α car c'est une fonction ne dépendant pas de α et dépendant des variables aléatoires X_i identiques et indépendantes, et que :
$$E(W'_n) = \frac{n-1}{n} \times \frac{n\alpha}{n-1} = \alpha$$
donc W'_n est un estimateur sans biais de α.

(d) i. En admettant que le moment d'ordre 2 de W'_n est égal à
$$E(W'^2_n) = \frac{(n-1)\alpha^2}{n-2}$$
la formule de Koenig-Huygens donne :
$$V(W'_n) = E(W'^2_n) - (E(W'_n))^2 = \frac{(n-1)\alpha^2}{n-2} - \alpha^2 = \frac{\alpha^2}{n-2}.$$
Comme W'_n admet une variance, on peut appliquer l'inégalité de Bienaymé-Tchebychev, pour tout $\varepsilon > 0$,
$$\begin{aligned} P(W'_n - \varepsilon < \alpha < W'_n + \varepsilon) &= P(-\varepsilon < W'_n - \alpha < \varepsilon) = P(|W'_n - E(W'_n)| < \varepsilon) \\ &= 1 - P(|W'_n - E(W'_n)| \geq \varepsilon) \\ &\geq 1 - \frac{V(W'_n)}{\varepsilon^2} = 1 - \frac{\alpha^2}{\varepsilon^2(n-2)}. \end{aligned}$$

ii. On suppose dans cette question que α est strictement compris entre 1 et 2 et on prend $\varepsilon = 1/10$. On a alors :
$$1 - \frac{\alpha^2}{\varepsilon^2(n-2)} \geq 1 - \frac{400}{n-2} \geq 0,95 \iff 0,05 \geq \frac{400}{n-2} \iff n-2 \geq \frac{40000}{5} \iff n \geq 8002.$$
Finalement, si $n \geq 8002 = N$, on sait que :
$$P\left(W'_n - \frac{1}{10} \leq \alpha \leq W'_n + \frac{1}{10}\right) = P\left(\alpha \in \left[W'_n - \frac{1}{10}; W'_n + \frac{1}{10}\right]\right) \geq 0,95$$
On peut rajouter les égalités car W'_n est une variable à densité, la probabilité qu'elle soit égale à une valeur précise est nulle.
Ainsi, $\left[W'_n - \frac{1}{10}; W'_n + \frac{1}{10}\right]$ est un intervalle de confiance de α au niveau de confiance 95%.

8. (a) i. Pour tout k, X_k admet une espérance donc par linéarité de l'espérance, Y_n admet une espérance et :

$$E(Y_n) = c_n \sum_{k=1}^{n} \frac{\alpha\beta}{\alpha - 1} = \frac{nc_n\alpha\beta}{\alpha - 1}$$

Or on veut que Y_n estime β sans biais, ce qui est équivalent à :

$$E(Y_n) = \beta \iff c_n = \frac{\alpha - 1}{n\alpha}.$$

ii. Comme les X_k sont indépendantes et $\alpha > 2$ donc X_k admet une variance pour tout k,

$$V(Y_n) = c_n^2 \sum_{k=1}^{n} V(X_k) = n \times \frac{(\alpha-1)^2}{n^2\alpha^2} \times \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)} = \frac{\beta^2}{n\alpha(\alpha-2)} \xrightarrow[n\to+\infty]{} 0.$$

(b) i. Soit F la fonction de répartition commune aux X_k et G_n celle de Z_n. Pour tout x réel,

$$G_n(x) = P(Z_n \leq x) = 1 - P(\min(X_1, \ldots, X_n) > x) = 1 - P\left(\bigcap_{k=1}^{n}(X_k > x)\right).$$

Or les (X_k) sont indépendantes puis de même loi, on obtient :

$$G_n(x) = 1 - \prod_{k=1}^{n} P(X_k > x) = 1 - \prod_{k=1}^{n}(1 - F(x)) = 1 - (1 - F(x))^n.$$

Si $x < \beta$, $G_n(x) = 1 - (1-0)^n = 0$.
Si $x \geq \beta$,

$$G_n(x) = 1 - \left(1 - \left(1 - \left(\frac{\beta}{x}\right)^\alpha\right)\right)^n = 1 - \left(\left(\frac{\beta}{x}\right)^\alpha\right)^n = 1 - \left(\frac{\beta}{x}\right)^{n\alpha}$$

On reconnaît la fonction de répartition de la loi de Pareto de paramètres $n\alpha$ et β. Donc Z_n suit la loi de Pareto de paramètres $n\alpha$ et β.
Enfin on remarque que $\alpha > 2$ et $n \geq 1$ donc $n\alpha > 2$ et Z_n admet une espérance et une variance et :

$$E(Z_n) = \frac{n\alpha\beta}{n\alpha - 1} = \frac{\alpha\beta}{\alpha - \frac{1}{n}} \xrightarrow[n\to+\infty]{} \frac{\alpha\beta}{\alpha} = \beta.$$

ii. On veut que $E(d_nZ_n) = \beta$ (cette espérance existe bien par linéarité) ce qui donne :

$$d_n \times \frac{n\alpha\beta}{n\alpha - 1} = \beta \iff d_n = \frac{n\alpha - 1}{n\alpha}$$

qui ne dépend pas de β donc (d_nZ_n) est bien un estimateur de β, et il est sans biais.
On obtient de plus avec Z_n qui suit une loi de Pareto :

$$\begin{aligned} V(Z'_n) &= V(d_nZ_n) = d_n^2V(Z_n) = \frac{(n\alpha-1)^2}{n^2\alpha^2} \times \frac{n\alpha\beta^2}{(n\alpha-1)^2(n\alpha-2)} \\ &= \frac{\beta^2}{n\alpha(n\alpha-2)} \xrightarrow[n\to+\infty]{} 0. \end{aligned}$$

iii. On étudie le signe de $V(Z'_n) - V(Y_n)$ au voisinage de $+\infty$:

$$V(Z'_n) - V(Y_n) = \frac{\beta^2}{n\alpha(n\alpha-2)} - \frac{\beta^2}{n\alpha(\alpha-2)} = \frac{\beta^2}{n\alpha} \times \frac{\alpha - 2 - n\alpha + 2}{(\alpha-2)(n\alpha-2)} = \frac{\beta^2(1-n)}{n(\alpha-2)(n\alpha-2)}$$

Le dénominateur est toujours strictement positif car $\alpha > 2$ et $n\alpha > 2$, $\beta^2 > 0$ et $(1-n) < 0$ dès que $n \geq 2$ donc pour tout $n \geq 2$,

$$V(Z'_n) - V(Y_n) < 0 \Leftrightarrow V(Z'_n) < V(Y_n)$$

et Z'_n est donc un estimateur de β plus efficace que Y_n.