2 Partie théorique : étude mathématique

2.1/ Etude probabiliste

1. Soient α et θ deux réels strictements positifs. Déterminer le réel C pour que la fonction : $x \to \begin{cases} \frac{C}{x^{\alpha+1}} & si\ x \geq \theta \\ 0 & sinon. \end{cases}$ soit une densité de probabilité.

On sait que f est une densité si :

$$\begin{aligned} \int_{-\infty}^{+\infty} f(t)dt = 1 \Leftrightarrow & \int_{\theta}^{+\infty} \frac{C}{x^{\alpha+1}} dx = 1 \\ & \Leftrightarrow C \int_{\theta}^{+\infty} \frac{1}{x^{\alpha+1}} dx = 1 \\ & \Leftrightarrow C[\frac{-1}{\alpha x^{\alpha}}]_{\theta}^{+\infty} = 1 \\ & \Leftrightarrow \frac{C}{\alpha \theta^{\alpha}} = 1 \\ & \Leftrightarrow C = \alpha \theta^{\alpha} \end{aligned}$$

On obtient que :$f(x) = \begin{cases} \frac{\alpha}{\theta} (\frac{\theta}{x})^{\alpha+1} & si\ x \geq \theta \\ 0 & sinon \end{cases}$

Dans toute la suite, on notera cette densité f.

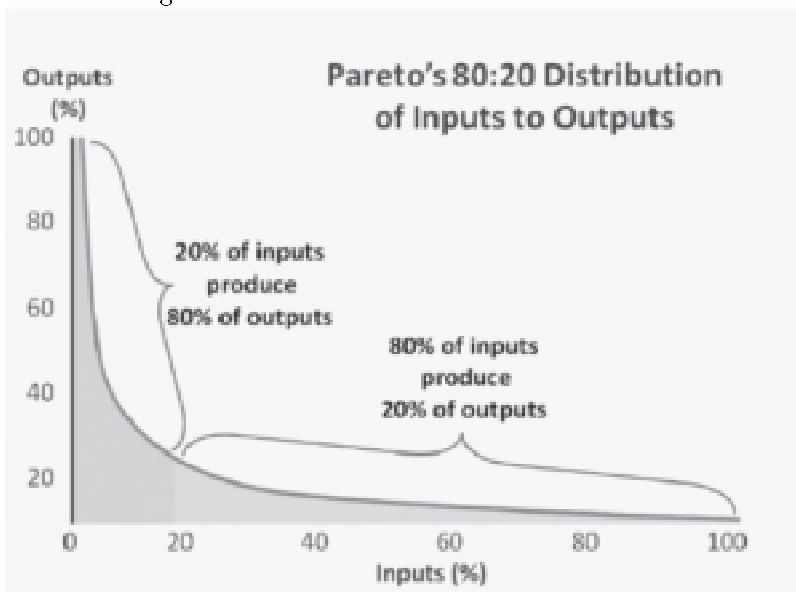
Définition : On dit que la variable aléatoire X suit une loi de Pareto de paramètres α et θ si et seulement si sa densité est f. On notera alors $X \sim P(\alpha, \theta)$.

Dans ce qui suit, X désignera une variable aléatoire de loi $P(\alpha, \theta)$

2. Dans quelle contexte cette loi est-elle utilisée? Donner des exemples de modélisation avec une telle loi.

Après quelques recherches, on apprend que cette loi est utilisé en économie principalement. Elle est néanmoins très discutée car beaucoup la trouvent trop imprécise. A l'aide de la loi de Pareto on peut notemment modeliser de la gestion de stock, de ventes, de projet ou de production. On appelle aussi cette loi la loi des 80-20 car elle permet d'observer que 80% des effets sont le produit de seulement 20% des causes. En effet, si on modélise un problème comme étant causé par les valeurs que peuvent prendre X, on voit qu'après θ, approximativement 20% des valeurs de cause constituent 80% de l'intégrale de la densité de probabilité du problème (voir figure 1). On peut généraliser ce raisonnement pour différentes valeurs de α (on a alors des proportions différentes de 80-20).

Figure 1: Densité de loi de Pareto schématisée



3. A quelle condition la variable X posède-t-elle une espérance ? une variance ? On se placera sous cette dernière condition dans la suite.

- X admet une espérance si :

$$\int_{-\infty}^{+\infty} |x| f(x)dx \ converge$$
$$\Leftrightarrow \int_{\theta}^{+\infty} |x| \frac{\alpha}{\theta}(\frac{\theta}{x})^{\alpha+1}dx = \theta^{\alpha}\alpha \int_{\theta}^{+\infty} \frac{1}{x^{\alpha}}dx \quad ,(\theta > 0 \Leftrightarrow |x| = x) \ converge$$
$$\Leftrightarrow \alpha > 1 \ (Intégrale \ de \ Riemann)$$

X admet une espérance si $\alpha > 1$ et cette espérance vaut :

$$E(X) = \int_{-\infty}^{+\infty} x f(x)dx$$

- X admet une variance si X admet une espérance et si :

$$\int_{-\infty}^{+\infty} (x - E(X))^2 f(x)dx \ converge$$
$$\Leftrightarrow \int_{\theta}^{+\infty} x^2 f(x)dx + E(X)^2 \int_{\theta}^{+\infty} f(x)dx - 2E(X) \underbrace{\int_{\theta}^{+\infty} xf(x)dx}_{E(X)} = \int_{\theta}^{+\infty} x^2 f(x)dx - E(X)^2 \ converge$$
$$\Leftrightarrow \int_{\theta}^{+\infty} x^2 f(x)dx \ converge$$
$$\Leftrightarrow \theta^\alpha \alpha \int_{\theta}^{+\infty} \frac{1}{x^{\alpha-1}} \ converge$$
$$\Leftrightarrow \alpha - 1 > 1 \ (Intégrale \ de \ Riemman)$$
$$\Leftrightarrow \alpha > 2$$

X admet une variance si $\alpha > 2$ et cette variance vaut :

$$V(X) = \int_{-\infty}^{+\infty} (x - E(x))^2 f(x)dx = E(X^2) - E(X)^2$$

4. Calculer l'espérance et la variance de X.

- Calcul de l'espérance :

$$\begin{aligned} E(X) &= \theta^\alpha \alpha \int_{\theta}^{+\infty} \frac{1}{x^\alpha} dx \\ &= \theta^\alpha \alpha [\frac{-1}{(\alpha-1)x^{\alpha-1}}]_{\theta}^{+\infty} \\ &= \theta^\alpha \alpha \frac{1}{(\alpha-1)\theta^{\alpha-1}} \\ &= \frac{\theta\alpha}{(\alpha-1)} \end{aligned}$$

- Calcul de la variance :

$$\begin{aligned} V(x) &= E(X^2) - E(X)^2 \\ E(X^2) &= \frac{\theta^2\alpha}{(\alpha-2)} \\ V(X) &= \frac{\theta^2\alpha}{(\alpha-2)} - \frac{\theta^2\alpha^2}{(\alpha-1)^2} \\ V(X) &= \frac{\theta^2\alpha(\alpha-1)^2 - \theta^2\alpha^2(\alpha-2)}{(\alpha-1)^2(\alpha-2)} \\ V(X) &= \frac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)} \end{aligned}$$

5. Calculer la fonction de répartition F de X.

La fonction de répartition, notée $F_X(x)$ se calcule suivant la formule suivante :

$$\begin{aligned}
F_X(x) =& P(X \leq x) = \int_{-\infty}^{x} f(t)dt \\
=& \begin{cases} 0 & si\ x < \theta \\ \int_\theta^x \frac{\alpha}{\theta}(\frac{\theta}{t})^{\alpha+1}dt & si\ x \geq \theta \end{cases} \\
=& \begin{cases} 0 & si\ x < \theta \\ \alpha\theta^\alpha[\frac{-1}{\alpha t^\alpha}]_\theta^x = \frac{\cancel{\alpha\theta^\alpha}}{\cancel{\alpha\theta^\alpha}} - \frac{\cancel{\alpha}\theta^\alpha}{\cancel{\alpha}x^\alpha} & si\ x \geq \theta \end{cases} \\
=& \begin{cases} 0 & si\ x < \theta \\ 1 - (\frac{\theta}{x})^\alpha & si\ x \geq \theta \end{cases}
\end{aligned}$$

6. Déterminer la fonction réciproque de F.

$$\begin{aligned}
y =& 1 - (\frac{\theta}{x})^\alpha \\
(\frac{\theta}{x})^\alpha =& 1 - y \\
x^\alpha =& \frac{\theta^\alpha}{1-y} \\
x =& \frac{\theta}{(1-y)^{1/\alpha}}
\end{aligned}$$

On a donc :

$$F^{-1}(x) = \frac{\theta}{(1-x)^{1/\alpha}}$$

2.2/ Simulation d'une valeur d'une variable de loi de Pareto

L'objet de ce paragraphe est de proposer une méthode de construction d'une valeur d'une variable aléatoire de loi P(α,θ).

7. Expliquer de manière formelle comment simuler une valeur d'une variable aléatoire de loi P(α,θ).

On va appliquer la méthode d'inversion.

Soit X une variable réelle de loi continue et strictement croissante. Alors si U est de loi uniforme sur [0,1], la variable $F^{-1}(U)$ a la même loi que X.

Dans le cadre de la loi de Pareto :

$$\begin{cases} F(\theta^-) = 0 \\ F(\theta) = 1 - (\frac{\theta}{\theta})^\alpha = 0 \end{cases}$$

Donc F est continue en θ, de plus F est continue de manière triviale sur $\mathbb{R}\backslash\{\theta\}$,donc F est continue sur $\mathbb{R}$.

F est également strictement croissante sur $[\theta, +\infty[$ de manière évidente.

On peut donc appliquer la méthode d'inversion à la loi de Pareto pour en simuler une valeur x en simulant une réalisation u de U et en posant $x = F^{-1}(u) = \frac{\theta}{(1-u)^{1/\alpha}}$.

2.3/ Etude statistique

On considère n données réelles $x_1, x_2, ..., x_n$. On fait l'hypothèse que ces données sont les réalisations de n variables aléatoires $X_1, X_2, ..., X_n$,indépendantes et de même loi $P(\alpha, \theta = 1)$. On suppose que le paramètre α est inconu et on s'intéresse au problème de son estimation par la méthode du maximum de vraisemblance.

8. Ecrire la log-vraisemblance de l'échantillon et trouver l'estimation du maximum de vraisemblance de α. En déduire alors l'estimateur du maximum de vraisemblance du paramètre α. On le notera $\widehat{\alpha}_n$.

$$\begin{aligned}
L(x,\alpha) &= \prod_{j=1}^{n} \alpha(\frac{1}{x_j})^{\alpha+1} \\
log(L(x,\alpha)) &= nln(\alpha) - \sum_{j=1}^{n}(\alpha+1)ln(x_j) \\
&= nln(\alpha) - (\alpha+1)s_n \ avec \ s_n = \sum_{j=1}^{n} ln(x_j) \\
t_n &= arg \max_{z\in[1,+\infty]}(LL(x,z)) \\
g(z) &= nln(z) - (z+1)s_n \\
g'(z) &= \frac{n}{z} - s_n \\
g''(z) &= \frac{-n}{z^2}
\end{aligned}$$

Donc $g'(z) = 0 \Leftrightarrow z = \frac{n}{s_n}$ et $g''(z) < 0$. L'unique valeur t_n qui maximise la vraisemblance sachant $x_1, x_2, ..., x_n$ est donc égale à $t_n = \frac{n}{s_n}$

On peut alors construire un estimateur du maximum de vraisemblance de α noté $\widehat{\alpha}_n$: $\widehat{\alpha}_n = \frac{n}{\sum_{j=1}^{n} log(X_j)}$

9. Calculer $E[log(X_1)]$ et $Var(log(X_1))$.

$$
\begin{aligned}
E(log(X_1)) =& \int_1^{+\infty} \alpha ln(x)(\frac{1}{x})^{\alpha+1}dx \\
=& \alpha \int_1^{+\infty} ln(x)(\frac{1}{x})^{\alpha+1}dx
\end{aligned}
$$

On va procéder par intégration par parties :$u(x) = ln(x), u'(x) = \frac{1}{x}, v'(x) = \frac{1}{x^{\alpha+1}}, v(x) = \frac{-1}{\alpha x^{\alpha}}$

$$
\begin{aligned}
\int_1^{+\infty} ln(x)(\frac{1}{x})^{\alpha+1}dx =& [\frac{-ln(x)}{\alpha x^{\alpha}}]_1^{+\infty} + \int_1^{+\infty} \frac{1}{\alpha x^{\alpha+1}}dx \\
=& 0 + [\frac{-1}{\alpha^2 x^{\alpha}}]_1^{+\infty} = \frac{1}{\alpha^2} \\
\text{donc } E(log(X_1)) =& \frac{1}{\alpha}
\end{aligned}
$$

Calcul de $Var(log(X_1))$:

$$
\begin{aligned}
Var(log(X_1)) =& E((log(X_1))^2) - E(log(X_1))^2 \\
E((log(X_1))^2) =& \alpha \int_1^{+\infty} ln^2(x)(\frac{1}{x})^{\alpha+1}dx \\
=& \alpha(\underbrace{[\frac{-ln^2(x)}{\alpha x^{\alpha}}]_1^{+\infty}}_{0} + \int_1^{+\infty} \frac{2ln(x)}{\alpha x^{\alpha+1}}dx) \\
=& \frac{2}{\alpha^2} \\
Var(log(X_1)) =& \frac{1}{\alpha^2}
\end{aligned}
$$

10. Proposer un estimateur sans biais et convergent du paramètre $\beta = \frac{1}{\alpha}$. On notera cet estimateur $B_n$et $b_n$sa réalisation.

$$
B_n = \frac{1}{n} \sum_{i=1}^{+\infty} log(X_i)
$$

B_n est un estimateur du paramètre $\beta = \frac{1}{\alpha}$ sans biais et convergent. En effet, $E(log(X_1)) = \frac{1}{\alpha}$ donc par linéarité de l'espérance, on a

$$
E(B_n) = \frac{\cancel{n}}{\cancel{n}}\beta = \beta
$$

De même, puisque les variables X_i indépendantes, on a la linéarité de la variance de leur somme, donc

$$Var(B_n) = \frac{n}{n^2}\beta^2 \to 0$$

11. Enoncer le TLC que satisfait cet estimateur.

$E(log(X_1)^2) < \infty$ car la variance de $log(X_i)$ existe. Cette variable admet donc un momet d'ordre 2. $B_n$vérifie donc le Théorème limite centrale suivant :

$$\sqrt{n}(\underbrace{\frac{1}{n}\sum_{j=1}^{n}}_{B_n} log(X_i) - E(log(X_i))) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} N(0, Var(log(X_i)))$$

$$\sqrt{n}(B_n - \beta) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} N(0, \beta^2)$$

12. Proposer un estimateur du paramètre α. On le notera A_n.

$$A_n = B_n^{-1} = \frac{n}{\sum_{i=1}^{n} log(X_i)} = \widehat{\alpha_n}$$

13. En déduire la convergence de $A_n$vers α et énoncer le TLC que satisfait A_n.

D'après la loi forte des grands nombres, on a :

$$B_n \underset{n\to\infty}{\overset{p.s}{\to}} \beta$$

$$\text{donc } \frac{1}{B_n} = A_n \underset{n\to\infty}{\overset{p.s}{\to}} \frac{1}{\beta} = \alpha$$

$A_n$converge donc bien vers α.

On a :

$$\sqrt{n}(B_n - \beta) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} N(0, \beta^2)$$

On utilise la delta-méthode en utilisant $g(x) = \frac{1}{x}$, dérivable en α de manière évidente ($\alpha > 0$)

$$\sqrt{n}(\frac{1}{B_n} - \frac{1}{\beta}) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} N(0, (-\frac{1}{\beta^2})^2\beta^2)$$

On obtient donc finalement :

$$\sqrt{n}(A_n - \alpha) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} N(0, \alpha^2)$$

2.4 Récapitulatif de la partie mathématique

Fonction de densité	$f(x)$	$\begin{cases} \frac{\alpha}{\theta}(\frac{\theta}{x})^{\alpha+1} & si\ x \geq \theta \\ 0 & sinon \end{cases}$
Fonction de répartition	$F(x)$	$\begin{cases} 0 & si\ x < \theta \\ 1-(\frac{\theta}{x})^{\alpha} & si\ x \geq \theta \end{cases}$
Fonction inverse de la fonction de répartition	$F^{-1}(x)$	$F^{-1}(x) = \frac{\theta}{(1-x)^{1/\alpha}}$
Esperance	$E(X)$	$\frac{\theta\alpha}{(\alpha-1)}\ (\alpha > 1)$
Variance	$V(x)$	$\frac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)}(\alpha > 2)$
Estimateur du paramètre α	A_n	$\frac{n}{\sum_{i=1}^{n} log(X_i)}$