République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DÉPARTEMENT DE MATHÉMATIQUES

Mémoire présenté en vue de l'obtention du Diplôme :

MASTER en Mathématiques

Option : **Statistique**

Par

RAHIM Asma

Titre :

Statistique d'ordre : Théorie et applications

Membres du Comité d'Examen :

Pr.	**NECIR Abdelhakim**	UMKB	Président
Pr.	**SAYAH Abdallah**	UMKB	Encadreur
Dr.	**BERKANE Hassiba**	UMKB	Examinatrice

Juin 2022

Dédicace

Je dédie ce modeste travail :

A celui qui m'a indiqué la bonne voie, et qui a attendu avec patiences ce jour-là, à celui qui 0m'a initié à la vie, qui m'a appris la modestie :

Mon père.

A cette source de tendresse, qui a sacrifié sa vie pour parfaire mon éducation et qui me comble de bonheur :

Ma mère.

Mes chers sœurs et frères : **Siham, Razika, Amel, Meriem, Fatma Zohra, Reguia, Maroua et Khaoithar**. Et mes frères, **Hamza**, **Mohamed**, et surtout **Abdelmadjid** qui m'a beaucoup soutenu durant toutes mes études.

Je dédie ce travail à tous les membres de ma famille et à ceux qui m'ont aidé et m'ont encouragé.

Et tous ceux avec j'ai partagé ma vie universitaire dans le bonheur et le malheur.

Remerciements

Tout d'abord, je tiens à remercier mon Dieu le plus puissant « **Allah** » qui m'a donné le courage, la santé et la patience pour réaliser ce modeste travail.

Aussi, je tiens à exprimer mes sentiments de gratitude à mon encadreur **Mr. SAYAH Abdallah** de m'avoir encadré.

Ainsi, je la remercie pour ses conseils et ses orientations toute au long de la préparation de ce mémoire.

« Sans votre aide et votre encadrement, ce mémoire n'aurait jamais été complété ».

Je remercie **Mr. NECIR Abdelhakim** et **Mme. BERKANE Hassiba** pour leurs conseils pratiques et je remercie tous qui m'ont enseigné durant mon parcours universitaire.

Sans oublier de remercier tous ceux qui m'ont aidé de près ou du loin et ceux qui m'ont

encouragé à finir mon travail de fin d'étude.

Rahim Asma

Table des matières

Introduction

Le développement de la science au cours des dernières années est augmenté, et c'est ce que l'on observe dans le domaine la statistique, plus précisément dans les méthodes statistiques qui utilisent maintenant des observations ordonnée par ordre croissant. Ces dernières sont appelées statistiques d'ordre relative à un échantillon aléatoire. Celles-ci sont très peu étudiée, alors qu'elles sont un outil indispensable dans la théorie des valeurs extrêmes et la modélisation des risques par exemple.

La statistique d'ordre fait partie des outils fondamentaux de la statistique théorique. Les cas importants de la statistique d'ordre sont le minimum et le maximum, ainsi que les différents quantiles.

En raison de l'importance des statistiques d'ordre dans de nombreuses théories et son rôle efficace dans beaucoups d'applications, ce mémoire est une compilation de travaux de recherche liés aux statistiques d'ordre.

Le mémoire est partagé en 2 chapitre :

Chapitre1 : Dans ce chapitre, nous avons consacré à l'étude des distributions d'une statistique d'ordre, dont nous présentons les définitions et les propriétés des statistiques d'ordre associées à un échantillon. Après nous allons voir la fonction de répartition et la densité de probabilité de la $k-ième$ statistique d'ordre et plus précisément on s'intéresse aux valeurs extrêmes (la fonction de répartition du minimum et du maximum). De plus, nous présentons la densité jointe d'un couple de statistique d'ordre et ainsi, le calcul des moments : espérance, variance et covariance.

Chapitre2 : Dans ce chapitre, on va étudier les lois asymptotiques et les lois limites des valeurs extrêmes d'un échantillon ordonné.

Chapitre 1

Généralités

1.1 Statistique d'ordre

Soit $(X_1, \ldots, X_n)$ une suite de variables aléatoires indépendantes et indentiquement distribuées $(i.i.d)$, d'une densité connue f et d'une fonction de répartition F définie par : $(F(x) = P(X_n \leq x)\,, \; pour \; x \in \mathbb{R})\,.$

Soit S_n l'ensemble des permutation de $\{1, \ldots, n\}\,.$

1.1.1 Définition de la statistique d'ordre

La statistique d'ordre de l'échantillon $(X_1, \ldots, X_n)$ est le réarrangement croissant de $(X_1, \ldots, X_n)$. On le note par $(X_{1,n}, \ldots, X_{n,n})\,.$

On a $X_{1,n} \leq \ldots \leq X_{n,n}$, et il existe une permutation aléatoire $\sigma_n \in S_n$ telle que

$$(X_{1,n}, \ldots, X_{n,n}) = \left(X_{\sigma_n(1)}, \ldots, X_{\sigma_n(n)}\right)$$

Le vecteur $(X_{1,n}, \ldots, X_{n,n})$ est appelé l'échantillon ordonné associé à l'échantillon $(X_1, \ldots, X_n)$, et $X_{k,n}$ étant la $k^{ième}$ statistique d'ordre.

Remarque 1.1.1 *Si la loi de la variable X est absolument continue on peut conclure que :*

$$P\left[X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}\right] = 1$$

Définition 1.1.1 ***(Fonction de distribution emprique)*** *Soit $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d) définies sur le même espace de probabilité (Ω, F, P) d'une fonction de répartition commune F telle que,*

$$F(x) = P\left\{\omega \in \Omega \,/ X(\omega) \leq x\right\} = P\left\{X \leq x\right\}, \qquad x \in \mathbb{R}.$$

La fonction de répartition empirique : notée F_n est définie par :

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(X_i \leq x)}, \qquad x \in \mathbb{R}.$$

Définition 1.1.2 ***(Fonction des quantiles empiriques)*** *La fonction des quantiles ou l'inverse généralisé de la fonction de distribution F notée par Q : pour chaque entier $n \geq 1$*

$$Q(t) = F^{\leftarrow}(t) = \inf\left\{s : F(s) \geq t\right\}, \quad 0 < t < 1.$$

La fonction des quantiles empiriques : notée Q_n est définie par :

$$Q_n(t) = F^{\leftarrow}(t) = \inf\left\{s : F(s) \geq t\right\}, \quad 0 < t < 1,$$

où $F^{\leftarrow}$ est l'inverse généralisée de la fonction de distribution F.

Définition 1.1.3 ***(Distribution empirique et quantiles empirique)*** *Il existe une autre version de la définition de F_n en utilisant les statistiques d'ordre (s.o) comme suit :*

$$F_n(x) = \begin{cases} 0 & si \;\; x < X_{1,n} \\ \frac{i-1}{n} & si \;\; X_{i-1,n} \leq x < X_{i,n}, \;\; 2 \leq i \leq n \\ 1 & si \;\; x \geq X_{n,n.} \end{cases} .$$

De même on obtient une autre version de Q_n en utilisant les statistiques d'ordre (s.o) comme suit :

$$Q_n(t) = \begin{cases} X_{i,n} & si \;\; \frac{i-1}{n} < t \leq \frac{i}{n} \\ X_{n,n} & si \;\; 0 < t \leq 1 \end{cases}$$

1.1.2 Loi de la statistique d'ordre

Théorème 1.1.1 *Si la loi de X_1 possède une densité f, alors la statistique d'ordre $(X_{1,n}, \ldots, X_{n,n})$ possède la densité.*

$$f_{X_{1,n},\ldots,X_{n,n}}(x_1, x_2, \ldots, x_n) = n! \prod_{i=1}^{n} f(x_i), \quad -\infty < x_1 < \ldots < x_n < +\infty$$

Preuve : On rappelle que, puisque $f(x)$ est la dérivée de $F(x)$

$$f(x) = \lim_{\delta x \to 0} \frac{F(x + \delta x)}{\delta x}.$$

Soit :

$$f(x) = \lim_{\delta x \to 0} \left(\frac{P(x \leq X < x + \delta x)}{\delta x} \right)$$

$$P(x_1 \leq X_{1,1} < x_1 + \delta x_1, \ldots, x_n \leq X_{n,n} < x_n + \delta x_n)$$
$$= n! P(x_1 \leq X_{1,1} < x_1 + \delta x_1) \ldots P(x_n \leq X_{n,n} < x_n + \delta x_n),$$

d'où

$$\frac{P\left(x_1 \leq X_{1,1} < x_1 + \delta x_1, \ldots, x_n \leq X_{n,n} < x_n + \delta x_n\right)}{\delta x_1, \ldots, \delta x_n}$$
$$= n! \frac{P\left(x_1 \leq X_{1,1} < x_1 + \delta x_1\right)}{\delta x_1} \ldots \frac{P\left(x_n \leq X_{n,n} < x_n + \delta x_n\right)}{\delta x_n}.$$

Alors

$$f_{X_{1,n},\ldots,X_{n,n}}\left(x_1, x_2, \ldots, x_n\right) = n! \prod_{i=1}^{n} f\left(x_i\right), \quad -\infty < x_1 < \ldots < x_n < \infty.$$

Notation 1.1.1 *On note $f_{X_{k,n}}(x)$ par $f_k(x)$ et $F_{X_{k,n}}(x)$ par $F_k(x)$.*

1.1.3 Loi marginale d'une statistique d'ordre

La loi marginale de $\left(X_{(1)}, X_{(2)}, \ldots, X_{(k)}\right)$; $k \leq n$ admet la densit :

$$\frac{n!}{(n-k)!} 1_{C_k}\left(x_1, x_2, \ldots, x_n\right) \prod_{i=1}^{i=n} f\left(x_i\right) \left[\int_{x_k}^{+\infty} f(x)\, dx\right]^{n-k},$$

où :

$$C_k = \left[\left(x_1, x_{2,\ldots,} x_k\right) \in \mathbb{R}^n / x_1 < x_2 < \ldots < x_k\right]$$

1.1.4 Loi de la k-ième statistique d'ordre

Soit $X_{k,n}\left(1 \leq k \leq n\right)$ la $k^{ième}$ statistique d'ordre, l'événement

$$\left\{x < X_{k,n} \leq x + \delta x\right\}$$

peut être représenté comme suit :

$X_i \leq x$ pour $k-1$ des X_i, $x < X_i \leq x + \delta x$ pour exactement l'un des X_i et $X_i > x + \delta x$ pour $n-k$ restes des X_i. Il y a C_n^{k-1} manières de réaliser le premier événement et

C^1_{n-k+1} manières de réaliser le deuxième événement, et il reste une seule manière de réaliser le dernier événement.

Considérons δx assez petit, on peut écrire

$$\begin{aligned} P\left(x \leq X_{k,n} < x+\delta x\right) = \frac{n!}{(k-1)!1!\,(n-k)!}\left[P\left(X_i < x\right)\right]^{k-1} \\ P\left(x \leq X_{i,n} < x+\delta x\right)\left[P\left(X_i \geq x\right)\right]^{n-k}. \end{aligned}$$

Alors, la densité marginale $f_k\left(x\right)$ de la coordonnée $X_{k,n}$ $(k = 1, \ldots, n)$ de l'échantillon ordonné s'obtient comme suit :

$$\begin{aligned} f_k\left(x\right) &= \lim_{\delta x \rightarrow \infty}\left\{\frac{P\left(x \leq X_{k,n} < x+\delta x\right)}{\delta x}\right\} \\ &= \frac{n!}{(k-1)!\,(n-k)!}\left(F\left(x\right)^{k-1}\right)\left(1-F\left(x\right)\right)^{n-k} f\left(x\right), \quad x \in \mathbb{R}. \end{aligned}$$

Le calcul de la fonction de répartition $F_k\left(x\right)$ de $X_{k,n}$ est immédiat :

$$\begin{aligned} F_k\left(x\right) &= P\left(X_{k,n} \leq x\right) \\ &= P(\text{au moins } k \text{ des } X_i \text{ sont inférieurs à } x) \\ &= \sum_{i=k}^{n} P\left(\text{exactement } i \text{ de } X_1, \ldots, X_n \text{ sont inférieurs à } x\right) \\ F_k\left(x\right) &= \sum_{i=k}^{n} C_n^i\left[F\left(x\right)\right]^i\left[1-F\left(x\right)\right]^{n-i}, \quad x \in \mathbb{R}. \end{aligned}$$

1.1.5 Cas particulier des valeurs extrêmes $\mathbf{X}_{1,n}$ et $\mathbf{X}_{n,n}$

Dans un échantillon de taille n deux statistiques d'ordre sont particulièrement intéressantes pour l'étude des événements extrêmes ce sont :

$$X_{1,n} = \min\left(X_1, \ldots, X_n\right) \quad \text{et} \quad X_{n,n} = \max\left(X_1, \ldots, X_n\right)$$

On a :

$$\begin{aligned}
F_{X_{1,n}}(x) &= P\left(X_{1,n} \leq x\right) = 1 - P\left(X_{1,n} > x\right) \\
&= 1 - P\left(\min\left(X_1, \ldots, X_n\right) > x\right) \\
&= 1 - P\left\{\bigcap_{i=1}^{n} X_i > x\right\} \\
&= 1 - \prod_{i=1}^{n} P\left(X_i > x\right) \\
&= 1 - \prod_{i=1}^{n} \left[1 - P\left\{X_i > x\right\}\right] \\
&= 1 - \left[1 - F(x)\right]^n .
\end{aligned}$$

D'où on déduit :

$$\boxed{F_{X_{1,n}}(x) = 1 - \left[1 - F(x)\right]^n .}$$

et la densité :

$$\boxed{f_1(x) = f_{X_1}(x) = n\left[1 - F(x)\right]^{n-1} f(x) .}$$

De même pour $X_{n,n}$:

$$\begin{aligned}
F_{X_{n,n}}(x) &= P\left(X_{n,n} \leq x\right) \\
&= P\left(\max\left(X_1, \ldots, X_n\right) \leq x\right) \\
&= P\left\{\bigcap_{i=1}^{n} X_i \leq x\right\} \\
&= \prod_{i=1}^{n} P\left(X_i \leq x\right) \\
&= \prod_{i=1}^{n} F(x).
\end{aligned}$$

Alors :

$$\boxed{F_{X_{n,n}}(x) = [F(x)]^n .}$$

et la densité :

$$\boxed{f_n(x) = f_{X_n(x)} = n\,[F(x)]^{n-1} f(x) .}$$

Définition 1.1.4 *Une variable aléatoire X à valeurs dans $[0,1]$ est dite suivre la loi bêta de paramètres $r > 0$, $s > 0$ (ce que l'on note $\mathcal{B}(r,s)$) si elle est absolument continue et admet pour densité :*

$$f(x) = \begin{cases} \dfrac{1}{\mathcal{B}(r,s)} x^{r-1}(1-x)^{s-1} & si\ x \in [0,1] \\ 0 & sinon \end{cases}$$

La fonction bêta pour $r, s > 0$ est

$$\mathcal{B}(r,s) = \int_0^1 t^{r-1}(1-t)^{s-1}\, dt.$$

Lemme 1.1.1 *La variable aléatoire $Y_k = F(X_{k,n})$ suit une loi bêta de paramètres k et $n-k+1$.*

preuve : On a

$$\begin{aligned}
P(F(X_{k,n}) \leq u) &= P\left(X_{k,n} \leq F^{-1}(u)\right) \\
&= \frac{n!}{(k-1)!\,(n-k)!} \int_0^{F\left(F^{-1}(u)\right)} t^{k-1}(1-t)^{n-k}\,dt \\
&= \frac{1}{\mathcal{B}(k, n-k+1)} \int_0^u t^{k-1}(1-t)^{n-k+1-1}.
\end{aligned}$$

Si on pose $k-1 = r$ et $n-k+1 = s$ on remarque que cette distribution est la loi de bêta de paramètres k et $n-k+1$ Alors :

$$F(X_{k,n}) \sim \mathcal{B}(k, n-k+1)$$

Remarque 1.1.2 *Il est facile de démonter que*

$$\frac{n!}{(k-1)!\,(n-k)!} = \frac{1}{\mathcal{B}(k, n-k+1)}$$

Par l'utilisation de la relation suivante $\mathcal{B}(r,s) = \dfrac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$, où Γ est la fonction Gamma. On définit comme suit :

$$\Gamma(r) = \int_0^1 t^{r-1} e^{t-1} dt, r > 0.$$

Propriétés 1.1.1 ***(La fonction Gamma)***

La fonction gamma est entièrement caractérisée sur $\mathbb{R}_+^*$ par les trois propriétés suivantes :

1. $\Gamma(1) = 1$.

2. pour tout $x > 0$, on a $\Gamma(x+1) = x\Gamma(x)$.

3. pour tout $n > 0$, $\Gamma(n) = (n-1)!$

1.1.6 Loi jointe d'un couple $(X_{i,n}, X_{n,n})$

Soit un couple $(X_{i,n}, X_{n,n})$ avec $1 \leq i < j \leq n$, l'événement

$$\{x < X_{i,n} \leq x + \delta x, y < X_{j,n} \leq y + \delta y\}$$

peut être représenté comme suit :

$X_r \leq x$ pour $i-1$ des X_r, $x < X_r \leq x + \delta x$ pour l'un des X_r, $x + \delta x < X_r \leq y$ pour $j-i-1$ des X_r, $y < X_{j,n} \leq y + \delta y$ pour l'un des X_r et $X_r > y + \delta y$ pour $n-j$ restes des X_r.

Considérons δx et δy assez petits

$$\begin{aligned}
&P\left(x < X_{i,n} \leq x + \delta x, y < X_{j,n} \leq y + \delta y\right) \\
&= \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!} \left(F(x)\right)^{i-1} \left(F(x+\delta x) - F(x)\right) \\
&\left(F(y) - F(x+\delta x)\right)^{j-i-1} \left(F(y+\delta y) - F(y)\right) \left(1 - F(y+\delta y)\right)^{n-j}.
\end{aligned}$$

Alors on peut calculer la densité jointe $f_{i,,j}(x,y)$ de deux statistiques d'ordre $X_{i,n}$ et $X_{j,n}$ comme suit :

$$f_{i,,j}(x,y)=\lim_{\substack{\delta x\to 0\\ \delta y\to 0}}\left\{\frac{P\left(x<X_{i,n}\leq x+\delta x,y<X_{j,n}\leq y+\delta y\right)x}{\delta x\delta y}\right\}$$
$$=\frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!}\left[F(x)\right]^{i-1}\left[F(y)-F(x)\right]^{j-i-1}\left[1-F(y)\right]^{n-j}f(x)\,f(y)\,.$$

Pour $i=1$ et $j=n$, on obtient la densité jointe de $(X_{1,n},X_{n,n})$ comme suit :

$$f_{1,n}(x,y)=n\,(n-1)\,(F(y)-F(x))^{n-2}\,f(x)\,f(y)\,,-\infty<x<y<+\infty.$$

Conséquence :

A partir de la densité jointe d'un couple $(X_{i,n},X_{j,n})$, on trouve la densité jointe de n statistique d'ordre

$$f_{1,\ldots,n}(x_1,x_2,\ldots,x_n)=n!\prod_{i=1}^{n}f(x_i)\,,\quad -\infty<x_1<\ldots<x_n<+\infty.$$

La fonction de distribution jointe de $(X_{i,n},X_{j,n})$ est

$$F_{i,,j}(x,y)=P\left(\{X_{i,n}\leq x\}\cap\{X_{j,n}\leq y\}\right),\qquad x,y\in\mathbb{R}.$$

On a deux cas :

– $\mathbf{1}^{i\grave{e}re}$ **cas** : $x\geq y$

$$F_{i,,j}(x,y)=P\left(X_{i,n}\leq x,X_{j,n}\leq y\right)$$
$$=P\left(X_{j,n}\leq y\right)$$
$$F_{i,,j}(x,y)=F_{i,j}(y)$$

– $\mathbf{2}^{ième}$ **cas :** $x < y$

$$F_{i,,j}(x,y) = P\left(X_{i,n} \leq x, X_{j,n} \leq y\right)$$

$$= P(\text{au moins } i \text{ de } X_1, \ldots, X_n \text{ sont inférieurs à } x$$

$$\text{et au moins } j \text{ de } X_1, \ldots, X_n \text{ sont inférieurs à } y)$$

$$= \sum_{s=j}^{n} \sum_{r=i}^{s} P(\text{exactement } r \text{ de } X_1, \ldots, X_n \text{ sont inférieurs à } x$$

$$\text{et exactement } s \text{ de } X_1, \ldots, X_n \text{ sont inférieurs à } y)$$

$$F_{i,,j}(x,y) = \sum_{s=j}^{n} \sum_{r=i}^{s} \frac{n!}{r!\,(s-r)!\,(n-s)!} \left(F(x)\right)^{r} \left(F(y) - F(x)\right)^{s-r} \left(1 - F(y)\right)^{n-s}.$$

1.2 Les moments de la statistique d'ordre

Les principales relations sont données pour les moments de statistique d'ordre.

En prenant en compte l'égalité $F_{k:n}(x) = P\left\{X_{k,n} \leq x\right\} = \mathbf{1}_{F(x)}(k, n-k+1)$ nous obtenons la formule générale pour les moments de statistique d'ordre $X_{k,n}$ liés à une population avec un fonction de distribution F en fait :

$$\mu_{k,n}^{(r)} = E\left(X_{k,n}\right) = \int_{-\infty}^{+\infty} x^{r} dF_{k,n}(x)$$

$$= \frac{n!}{(k-1)!\,(n-k)!} \int_{-\infty}^{+\infty} x^{r} \left(F(x)\right)^{k-1} \left(1 - F(x)\right)^{n-k} dF(x) \qquad (1.1)$$

pour les fonctions de distribution continue F, cette égalité (1.1) peut s'exprimer comme suit :

$$\mu_{k,n}^{(r)} = \frac{n!}{(k-1)!\,(n-k)!} \int_{-\infty}^{+\infty} (G(u))^r u^{k-1} (1-u)^{n-k}\, du,$$

où G est l'inverse de F.

Il est intéressant de trouver les conditions correspondantes qui fournissent l'existence de différents moments des statistiques d'ordre.

Exemple 1.2.1 *Supposons que le moment* $\alpha_r = E\,|X|^r$ *, existe, (i, e)*

$$E\,|X|^r = \int_{-\infty}^{+\infty} |x|^r \, dF(x) < \infty. \tag{1.2}$$

Ensuite, en raison de(1.2), nous pouvons facilement déduire que

$$E\,|X_{k,n}|^r \leq \frac{n!}{(k-1)!\,(n-k)!} \int_{-\infty}^{+\infty} |x^r|\,(F(x))^{k-1} (1-F(x))^{n-k}\, dF(x)$$

$$\leq \frac{n!}{(k-1)!\,(n-k)!} \int_{-\infty}^{+\infty} |x|^r \, dF(x)$$

$$= \frac{n!}{(k-1)!\,(n-k)!} E\,|X|^r < \infty \tag{1.3}$$

Il résulte de (1.3) que l'existence du moment ar implique l'existence de tous les moments

$$E\,|X_{k,n}|^r\,, \quad 1 \leq k \leq n, \quad n = 1, 2, \ldots$$

1.2.1 Résultats généraux

Soit n fixè on note, sous réserve d'existence, $m = E(X)$ et $\sigma^2 = Var(X)$.

Alors :

1. $\sum_{k=1}^{n} E[X_{k,n}] = nE[X]$.

2. $\sum_{k=1}^{n} E\left[X_{k,n}\right] = nE\left[X^2\right].$

3. $\sum_{k=1}^{n} \sum_{s=1}^{n} E\left[X_{k,n}X_{s,n}\right] = nE\left[X^2\right] + n\left(n-1\right)m^2.$

4. $\sum_{k=1}^{n} \sum_{s=1}^{n} Cov\left[X_{k,n}, X_{s,n}\right] = n\sigma^2.$

De façon évidente, on a :

$$\forall j \in U \subseteq \mathbb{N}^* : \sum_{k=1}^{n} X_{k,n}^{j} = \sum_{k=1}^{n} X_{k}^{j}$$

1. si $j = 1$:
$$\sum_{k=1}^{n} E\left[X_{k,n}\right] = \sum_{k=1}^{n} E\left[X_k\right] = nm$$

2. si $j = 2$:
$$\sum_{k=1}^{n} E\left[X_{k,n}^2\right] = \sum_{k=1}^{n} E\left[X_k^2\right] = nE\left[X^2\right]$$

3. on a

$$\sum_{k=1}^{n} X_{k,n} = \sum_{k=1}^{n} X_{k\longrightarrow} \left(\sum_{k=1}^{n} X_{k,n}\right)^2 = \left(\sum_{k=1}^{n} X_k\right)^2$$

et en intégrant on trouve :

$$\begin{aligned}
&\Rightarrow E\left[\sum_{k=1}^{n} X_{k,n}\right]^2 = E\left[\sum_{k=1}^{n} X_k\right]^2 \\
&\Rightarrow \sum_{k=1}^{n} \sum_{s=1}^{n} E\left[X_{k,n}X_{s,n}\right] = \sum_{k=1}^{n} \sum_{s=1}^{n} E\left[X_k X_s\right] \\
&= nE\left[X^2\right] = \sum_{k=1}^{n} \sum_{s=1}^{n} E\left[X_k X_s\right] \\
&= nE\left[X^2\right] + n\left(n-1\right)m^2.
\end{aligned}$$

4.

$$\begin{aligned}\sum_{k=1}^{n}\sum_{s=1}^{n} Cov\left[X_{k,n}, X_{s,n}\right] &= \sum_{k=1}^{n}\sum_{s=1}^{n} E\left[X_{k,n}X_{s,n}\right] - \sum_{k=1}^{n}\sum_{s=1}^{n} E\left[X_{k}X_{s}\right] \\ &= nE\left[X^{2}\right] + n\left(n-1\right)m^{2} - n^{2}m^{2} \\ &= nE\left[X^{2}\right] + n^{2}m^{2} - n^{2}m^{2} - n^{2}m^{2} \\ &= n\left(E\left[X^{2}\right] - m^{2}\right) = n\sigma^{2}.\end{aligned}$$

Donc

$$\sum_{k=1}^{n}\sum_{s=1}^{n} Cov\left[X_{k,n}, X_{s,n}\right] = n\sigma^{2}.$$

Chapitre 2

Comportement asymptotique

Dans ce chapitre, nous étudions la loi asymptotique de $X_{k,n}$ et les lois limites des valeurs extrêmes d'une observation ordonnée.

Remarquons tout d'abord qu'un simple changement de variables permet de passer de la loi de $X_{n,n}$ à celle de $X_{1,n}$, en effet :

$$\min\left(X_1, X_2, \ldots\ldots, X_n\right) = -\max\left(-X_1, -X_2, \ldots\ldots, -X_n\right).$$

On obtiendra la loi de $X_{1,n}$ par le changement de variable $X \to -X$ dans la loi de $X_{n,n}$.

Ce qui nous aide à étudier les comportements des v.a aux rôles symétriques $X_{k,n}$, $X_{n-k+1,n}$.

Quand $n \to +\infty$: on a a deux types de comportements asymptotiques :

1. soit k est fixé, alors : $\lim_{n\to+\infty} \dfrac{k}{n} = 1$ ou 0, on est dans le cas des comportements des valeurs extrêmes.
2. soit k tend vers l'infini, alors : $\lim_{n\to+\infty} \dfrac{k}{n} = \alpha$, $0 < \alpha < 1$: cas d'un fractile empirique d'ordre α.

2.1 Convergence d'un fractile empirique

Théorème 2.1.1 *Si F est continue et strictement croissante et si $0 < \alpha < 1$, alors :*

$$X_{k,n} \underset{\longrightarrow}{P} x_\alpha \quad qand \quad n \to +\infty$$

où x_α est le fractile d'ordre α de X.

Preuve :

Le resultat $\frac{k}{n} \to \alpha$ alors, $k = [n\alpha]+1$, d'après la définition de la fonction de répartition empirique F_n :

$$F_n\left[X_{[n\alpha]+1,n}\right] = \frac{[n\alpha]}{n}.$$

$$\begin{aligned} F\left[X_{k,n}\right] - F\left[x_\alpha\right] &= F\left[X_{k,n}\right] - F_n\left[X_{k,n}\right] + F_n\left[X_{k,n}\right] - F\left[x_\alpha\right] \\ &= F\left[X_{k,n}\right] - F_n\left[X_{k,n}\right] + \frac{[n\alpha]}{n} - \alpha, \end{aligned}$$

on sait que :

$$d_k\left(F_n, F\right) = \sup\left|F(x) - F_n(x)\right| \underset{\longrightarrow}{P} \; 0,$$

donc,

$$F\left[X_{k,n}\right] - F\left[x_\alpha\right] \quad \underset{\longrightarrow}{P} \; 0$$

F est continue et strictement croissante, on conclut en composant par F^{-1} que:

$$X_{k,n} \;\; \underset{\longrightarrow}{P} \;\; x_\alpha \;\; \text{lorsque} \;\; n \to +\infty$$

Théorème 2.1.2 *Si F admet une densité f continue et strictement positive, alors on a :*

$$\sqrt{n}\left[X_{[n\alpha]+1,n}-x_{\alpha}\right] \quad \underrightarrow{loi} \quad N\left(0;\frac{\alpha\left(1-\alpha\right)}{f^{2}\left(x_{\alpha}\right)}\right)$$

Preuve. Posons $k=[n\alpha]+1$. La v.a :$\sqrt{n}\left[X_{k,n}-x_{\alpha}\right]$ suit une loi de densité h_{α} telle que :

$$h_{\alpha}\left(x\right)=\sqrt{n}\frac{n!}{\left(n-k\right)!\left(k-1\right)!}F^{k-1}\left(x_{\alpha}+n^{-1/2}x\right)\left[1-F\left(x_{\alpha}+n^{-1/2}x\right)\right]^{n-k}f^{k-1}\left(x_{\alpha}+n^{-1/2}x\right).$$

Prenons :

$$\underset{n\rightarrow+\infty}{[n\alpha]+1\rightarrow}n\alpha+1$$

En développant : $n!$, $\left(n\left(1-\alpha\right)-1\right)!$, $\left(n\alpha\right)!$ suivant la formule de Stirling :

$$m!=\left(m^{m}e^{-m}\right)\sqrt{2\pi m}\left(1+\varepsilon\left(m\right)\right),avec:\varepsilon\left(m\right)\rightarrow0\text{ quand }m\rightarrow\infty.$$

on trouve pour n assez grand :

$$h_{\alpha}\left(x\right)=\frac{1+\varepsilon\left(n\right)}{\sqrt{2\pi\alpha\left(1-\alpha\right)}}.\frac{f\left(x_{\alpha}+n^{-1/2}x\right)}{\left(1-\alpha\right)}\left[\frac{F\left(x_{\alpha}+n^{-1/2}x\right)}{\alpha}\right]^{n\alpha}\left[\frac{1-F\left(x_{\alpha}+n^{-1/2}x\right)}{\left(1-\alpha\right)}\right]^{n(1-\alpha)}$$

avec $\varepsilon\left(n\right)\rightarrow0$.

D'aprés la formule de Taylor on a :

$$\alpha^{-1}F\left(x_\alpha+n^{-1/2}x\right)=\alpha^{-1}F\left(x_\alpha\right)+n^{-1/2}xf\left(x_\alpha\right)+o\left(n\right)$$
$$=1+n^{-1/2}x\alpha^{-1}f\left(x_\alpha\right)+o\left(n\right),$$

avec

$$o\left(n\right)\rightarrow 0 \text{ et } F\left(x_\alpha\right)=\alpha.$$

De plus :

$$\frac{1-F\left(x_\alpha+n^{-1/2}x\right)}{\left(1-\alpha\right)}=\frac{1-F\left(x_\alpha+n^{-1/2}xf\left(x_\alpha\right)\right)}{\left(1-\alpha\right)}+\Psi\left(n\right)$$
$$=1-\frac{n^{-1/2}f\left(x\right)x}{\left(1-\alpha\right)}+\Psi\left(n\right),$$

avec

$$F\left(x_\alpha\right)=\alpha \quad \text{et} \quad \Psi\left(n\right)\rightarrow 0.$$

De plus

$$\left(1+n^{-1/2}x\alpha^{-1}f\left(x_\alpha\right)\right)=\exp\left[n\alpha\ln\left(1+n^{-1/2}x\alpha^{-1}f\left(x_\alpha\right)\right)\right]$$
$$=\exp\left[n^{-1/2}xf\left(x_\alpha\right)-x^2\frac{f^2\left(x\right)}{2\alpha}\right]+R_n$$

$$\left[1-\frac{n^{-1/2}xf\left(x_\alpha\right)}{(1-\alpha)}\right]\left[1+n^{-1/2}x\alpha^{-1}f\left(x_\alpha\right)\right]^{n\alpha}$$
$$=\exp\frac{1}{2}\left[\frac{-x^2f^2\left(x_\alpha\right)}{\alpha}-\frac{x^2f^2\left(x_\alpha\right)}{1-\alpha}\right].$$

Alors,

$$h_\alpha\left(x\right)=\frac{f\left(x\right)}{\sqrt{2\pi\alpha}}\exp\left[-\left(x^2f^2\left(x_\alpha\right)2\alpha\left(1-\alpha\right)\right)^{-1}\right]$$

elle n ést que la densité d'une loi normale de moyenne 0 et de la variance $\dfrac{\alpha\left(1-\alpha\right)}{f^2\left(x_\alpha\right)}$.

2.2 Convergence des valeurs extrêmes

Dans ce cas k : fixé, et $\alpha=0$ ou $\alpha=1$.

Soit $Y_k=F\left(X_{k,n}\right)$, et $G_k=nY_k$, on suppose que $n\to\infty$

Théorème 2.2.1 *On a*

$$G_k \underrightarrow{\ loi\ } \gamma\left(k,1\right),$$

ou $\gamma\left(k,1\right)$ est la loi gamma de paramètres k et 1.

Ce résultat provient du comportement asymptotique de la loi $n\mathcal{B}\left(k,n-k+1\right)$, qui converge en loi vers une loi gamma $\gamma\left(k,1\right)$. En effet, par définition d'une loi bêta sur $\left[0,1\right]$, G_k peut être mis sous la forme :

$$G_k=\frac{nA}{A+B}=\frac{A}{\frac{A}{n}+\frac{B}{n}},$$

où A et B sont des v.a. indépendantes de lois respectives $\gamma\left(k,1\right)$ et $\gamma\left(n-k+1,1\right)$.

On a :

$$E\left(\frac{A}{n}\right)=\frac{k}{n} \text{ et } Var\left(\frac{A}{n}\right)=\frac{k}{n^2}$$

donc $\frac{A}{n}$ converge en probabilité vers 0. De même :

$$E\left(\frac{B}{n}\right)=\frac{n-k+1}{n} \text{ et } Var\left(\frac{B}{n}\right)=\frac{n-k+1}{n^2}$$

et $\frac{B}{n}$ converge en probabilité vers 1.

2.3 Les lois des valeurs extrêmes

La distribution de $X_{n,n}$ devrait nous fournir des informations sur des événements extrêmes et comme la limite de cette distribution obtenue précédemment conduit à une loi dégénérée lorsque n tend vers l'infini, on recherche une loi non dégénérée pour le maximum de X. De façon analogue au théorème central limite, la théorie des valeurs extrêmes montre qu'il existe des suites $\{a_n\}$ et $\{b_n\}$ $n \in \mathbb{N}$, avec $a_n > 0$ et $b_n \in \mathbb{R}$, telles que

$$\lim_{n\to\infty} P\left(\frac{X_{n,n}-b_n}{a_n} \leq x\right)=\lim_{n\to\infty} F^n\left(a_n x+b_n\right)=H\left(x\right), \quad x\in\mathbb{R}$$

avec H est une distribution non dégénérée.

2.3.1 Les trois formes asymptotiques

Fisher et Tippett (1928), Gendenko (1943) et de Hann (1970) ont montré que les seules distributions limites non dégénérée H possibles sont les distributions de valeurs extrêmes. Le théorème suivant donne une condition nécessaire et suffisante pour l'existence d'une loi limite non dégénérée pour le maximum.

Théorème 2.3.1 ***(Fisher et Tippett(1928), Gnedenko (1943)) :*** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoire (i.i.d), s'il existe un réel et deux suites réelles* (a_n) *et* (b_n), $n \in \mathbb{N}$, *avec* $a_n > 0$ *et* $b_n \in \mathbb{R}$ *telles que :*

$$\lim_{n\to\infty} P\left[\frac{X_{n,n} - b_n}{a_n} \leq x\right] = H_\gamma(x),$$

pour tout x, où H est une fonction de distribution non dégénérée. Alors H est du même type que l'une des fonctions suivantes :

Gumbel (type 1) :

$$H_0(x) = \Lambda(x) = \exp(-\exp(-x)), \quad x \in \mathbb{R}.$$

Fréchet (type 2) :

$$H_\gamma(x) = \Phi_\gamma(x) = \begin{cases} 0, & x \leq 0, \\ \exp\left(-(x)^{-1/\gamma}\right), & x > 0, \end{cases} \quad \text{avec } \gamma > 0$$

Weibull (type 3) :

$$H_\gamma(x) = \Psi_\gamma(x) = \begin{cases} \exp\left(-(-x)^{-\gamma}\right), & x \leq 0, \\ 1, & x > 0, \end{cases} \quad \text{avec } \gamma < 0$$

- Ce théorème présente un intérêt important, car si l'ensemble des distributions est grand, l'ensemble des distributions des valeurs extrêmes est très petit. Ce théorème n'est valable que si les suites (a_n) et (b_n) existent et admettent des limites.
- Ce théorème est un résultat important car il n'est pas nécessaire de faire d'hypothèses paramétriques sur la loi des X_i La valeur de γ détermine le comportement de la queue de distribution.

Définition 2.3.1 ***(Distributions standard des valeurs extrêmes)*** *Les trois fonctions de distribution du théorème 2.3.1 sont les loi limites possibles pour le maximum s'appellent les distributions standard ou traditionnelle des valeurs extrêmes.*

Exemple 2.3.1 *(Coles* 2001*) Supposons que X suit une loi de probabilité loi exponentielle standard $\mathcal{E}(1)$ i.e :*

$$\forall x \in \mathbb{R}, \quad F(x) = P(X \leq x) = \begin{cases} 1 - e^{-x} & si \quad x > 0 \\ 0 & si \quad x \leq 0 \end{cases}$$

Si nous nous posons $a_n = 1$ et $b_n = \ln n$, nous aurons :

$$\forall x \in \mathbb{R}, \forall n \in \mathbb{N}^* \quad P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = P\{M_n \leq a_n x + b_n\} = F^n(x + \ln n).$$

Alors

$$P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = \begin{cases} \left[1 - e^{-(x+\ln n)}\right]^n & si \quad (x + \ln n) > 0 \\ 0 & si \quad (x + \ln n) \leq 0 \end{cases}$$
$$= \begin{cases} \left[1 + \frac{(-e^{-x})}{n}\right]^n & si \quad x > -\ln n \\ 0 & si \quad x \leq -\ln n \end{cases}$$

Donc

$$\forall x \in \mathbb{R}, \quad \lim_{n\to\infty} P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = \exp\left\{-e^{-x}\right\} = \Lambda(x).$$

On déduit que la loi exponentielle standard $\mathcal{E}(1)$ appartient au max domaine d'attraction de $\Lambda(x)$.

Exemple 2.3.2 *(Coles 2001) Supposons que X suit une loi de probabilité uniforme*

$\mathcal{U}[0,1]$ *i.e :*

$$\forall x \in \mathbb{R}, \quad F(x) = P(X \leq x) = \begin{cases} 0 & si \quad x < 0 \\ x & si \quad 0 \leq x < 1 \\ 1 & si \quad x \geq 1 \end{cases}$$

Si nous posons $a_n = n^{-1}$ et $b_n = 1$, nous aurons :

$$\forall x \in \mathbb{R}, \forall n \in \mathbb{N}^* \quad P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = P\left\{M_n \leq a_n x + b_n\right\} = F^n\left(1 + \frac{x}{n}\right)$$

$$P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = \begin{cases} 0 & si \quad 1 + \frac{x}{n} < 0 \\ \left(1 + \frac{x}{n}\right)^n & si \quad 0 \leq 1 + \frac{x}{n} < 1 \\ 1 & si \quad 1 + \frac{x}{n} \geq 1 \end{cases} = \begin{cases} 0 & si \quad x < -n \\ \left(1 + \frac{x}{n}\right)^n & si \quad -n \leq x < 0 \\ 1 & si \quad x \geq 0 \end{cases}$$

Car $\forall x \in \mathbb{R} \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$.

Donc :

$$\forall x \in \mathbb{R}, \quad \lim_{n\to\infty} P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = \begin{cases} e^x & si \quad x \leq 0 \\ 1 & si \quad x > 1 \end{cases} = \Psi_1(x).$$

On déduit que la loi uniforme $\mathcal{U}[0,1]$ appartient au max domaine d'attraction de $\Psi_1(x)$.

2.3.2 Distribution des valeurs extrêmes généralisée

Pour faciliter le travail avec les trois distribution limites, Jenkinson-Von Mises a donné une représentation qui a obtenu en introduisant les paramètres de localisation μ et de dispersion σ dans la paramétrisation des distributions extrêmes, on obtient la forme la plus générale de la distribution des valeurs extrêmes, notée GEVD (Generalized Extreme Value Distribution). Elle est simplement une reparamétrisation des distributions apparaissant dans le théorème 2.3.1

$$H_{\mu,\sigma,\gamma}(x) = \begin{cases} \exp\left\{-\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}}\right\} & pour \quad 1+\frac{\gamma}{\sigma}(x-\mu)>0, \\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) & pour \quad \gamma=0,\ x\in\mathbb{R}. \end{cases}$$

Avec $\mu \in \mathbb{R}$ et $\sigma > 0$.

En remplaçant $\left(\frac{x-\mu}{\sigma}\right)$ par x on obtient la forme standard de la GEVD :

$$H_{\gamma}(x) = \begin{cases} \exp\left\{-\left[1+\gamma x\right]^{-\frac{1}{\gamma}}\right\} & pour \quad \gamma\neq 0,\ 1+\gamma x>0, \\ \exp\left(-\exp\left(-x\right)\right) & pour \quad \gamma=0,\ x\in\mathbb{R}. \end{cases}$$

Où γ est le paramètres de forme.

Nous exprimons les trois distributions des valeur extrêmes Λ, Φ_γ et Ψ_γ et en termes de GEVD H_γ.

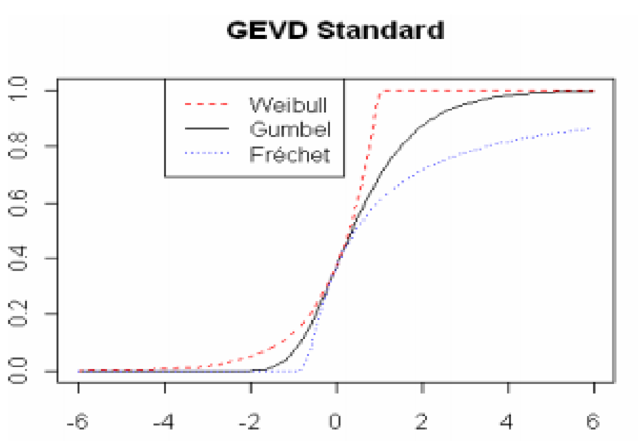

FIG. 2.1 – Distributions standard des valeurs extrêmes.

Remarque 2.3.1 *À partir de cette écriture, on peut distinguer 3 cas :*

– $\gamma = 0$ correspond à la loi de **Gumbel**.

– $\gamma > 0$ correspond à la loi de **Fréchet**.

– $\gamma < 0$ correspond à la loi de **Weibull**.

La densité de la loi GEV s'écrit pour $\gamma \neq 0$ comme suit :

$$h_{\mu,\sigma,\gamma}(x) = \frac{1}{\sigma}\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{(1+\gamma/\gamma)} \exp\left\{-\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}}\right\}.$$

La fonction de densité standard correspondante $h_{\mu,\sigma,\gamma}$ est :

$$h_{\mu,\sigma,\gamma}(x) = \begin{cases} H_{\gamma}(x)(1+\gamma x)^{-1/\gamma-1} & si \quad \gamma \neq 0,\ 1+\gamma x > 0, \\ \exp(-x-e^{-x}) & si \quad \gamma = 0,\ x \in \mathbb{R}. \end{cases}$$

La Figure 2.2 illustre les densities standard de GEV.

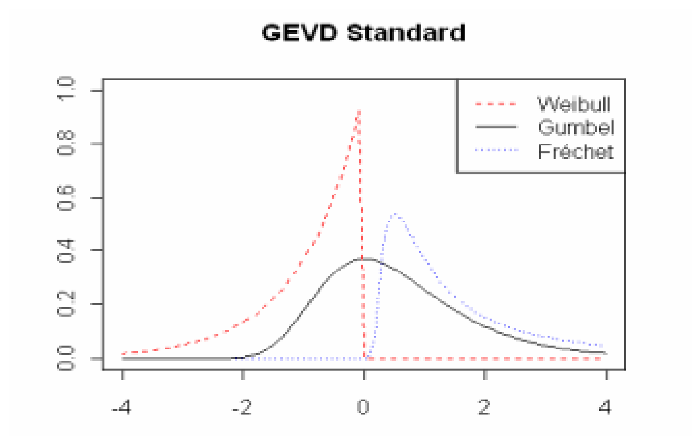

FIG. 2.2 – Densites standard des valeurs extrêmes.

Définition 2.3.2 ***(Domaines d'attraction)*** *Si F vérifie le **Théorème 2.3.1**, on dit alors que F apppartient au domaine d'attraction de H_{γ}, dénotée par $F \in \mathcal{D}(H_{\gamma})$.*

Avant de caractériser les domaines d'attractions, on définit les fonctions à variations.

2.3.3 Propriétés

Nous ne considéreronspar la suite que des variable standardisées :

- Type 1 : $F_1(y) = -\exp(-\exp(-y)), \; y \in \mathbb{R}$
- Type 2 : $F_2(y,\beta) = \begin{cases} 0, & y \leq 0, \\ \exp\left(-(y)^{-\beta}\right), & y > 0, \end{cases}$ avec $\beta > 0$
- Type 3 :$F_3(y,\beta) = \begin{cases} \exp\left(-(-y)^{\beta}\right), & y \leq 0, \\ 1, & y > 0, \end{cases}$ avec $\beta < 0$

Propriété 1 soit Y une v.a de type 1, la v.a $X = e^{-Y}$ suit une loi exponentielle $\gamma(1,1)$.

Propriété 2 Soient deux v.a X et Y, indépendantes, suivant une loi de type 1. La v.a $U = X - Y$ suit une loi logistique.

2.3.4 Relation entre les lois asymptotiques de $\mathbf{X}_{1,n}$ et $\mathbf{X}_{n,n}$

ll existe entre les lois asymptotiques de $X_{n,n}$ et de $X_{1,n}$ des relations aisément démontrables par changement de variables. On notera par F_i^* la f.r de la loi limite de $X_{1,n}\,(i = 1,2,3)$.

Ainsi, soit Y une v.a de type 1. On faisant la transformation $Z = -Y$:

$$\begin{aligned} F_z(Z) = P(Z < z) &= P(-Y < z) = 1 - F_Y(-z) \\ &= 1 - \exp\left(-\exp\left(\frac{x-(-\lambda)}{u}\right)\right) \\ &= F_1^*(x,-\lambda,u). \end{aligned}$$

Nous noterons :

$$F_1(., \lambda, u) \quad \xrightarrow{-Y} \quad F_1^*(., -\lambda, u) .$$

2.3.5 Eléments sur la démonstration des convergences en loi de $\mathbf{X}_{n,n}$

L'idée de base, due à Fréchet (1927) et à Fisher et Tippett (1928), consiste à découper l'échantillon de taille n en k sous-échantillons de taille m, $k = nm$, notons (X_i) , $i = 1$ à n, l'échantillon initial, et $\left(X_l^j\right)$, $j = 1$ à k et $l = 1$ à m, les éléments de chaque sous-échantillon. ll est évident que :

$$\sup_i X_i = \sup_j \sup_l X_l^j$$

Faisons tendre m vers l'infini pour k fixé, soit G la fonction de répartition de la loi limite $X_{n,n}$, $n \to \infty$. A une transformation affine de la variable près, de coefficients dépendant de k, on aura l'identité de G_k et de G :

$$G^k(x) = G(a_k x + b_k) \tag{2.1}$$

La relation(2.1) traduit ce que I'on appelle **le principe de stabilité**, et est l'équation fonctionnelle définissant toutes les lois limites possibles pour $X_{n,n}$. Les constantes a_k et b_k sont des constantes de normalisation, réelles et indépendantes de x

Définition 2.3.3 *Une fonction de répartition G est dite stable (sous-entendu."pour un maximum") s'il existe des suites $(a_n)_{n\in\mathbb{N}}$ et $a_n > 0$, et (b_n) telles que :*

$$G^n(x) = G(a_n\ x + b) , \quad x \in \mathbb{R}$$

Théorème 2.3.2 *Soit G une fonction de répartition stable. Alors G appartient nécessairement à l'un des types suivants :*

Type 1 :$F_1(y) = \exp(-\exp(-y))\,, \quad y \in \mathbb{R}.$

$$\text{Type 2 : } F_2(y,\beta) = \begin{cases} 0, & y \leq 0, \\ \exp\left(-(y)^{-\beta}\right), & y > 0, \end{cases}$$

$$\text{Type 3 :} F_3(y,\beta) = \begin{cases} \exp\left(-(-y)^{\beta}\right), & y \leq 0, \\ 1, & y > 0, \end{cases}$$

(β paramètre strictement positif.)

Si la mise en évidence des trois lois limites possibles pour $X_{n,n}$ a un intérêt propre, il est important d'essayer d'identifier. si possible, vers quel type de loi va converger $X_{n,n}$ selon sa loi initiale F.

Définition 2.3.4 *On appelle domaine d'attraction de $F(i = 1, 2, 3)$ l'ensemble $\mathcal{D}(F)$ des lois de probabilité F pour les quelles $(X_{1,\ldots,}X_n)$ étant un échantillon extrait de la loi F,*

$$X_{n,n} = \sup_{i=1 \text{à } n} X_i$$

va converger en loi vers la loi des extrêmes de type F_i.

Nous allons donner maintenant des conditions suffisantes de stabilité permettant de préciser le domaine d'attraction de chacun de ces types.

Définition 2.3.5 *Soit G une fonction de répartition sur $\mathbb{R}$, une fonction de répartition F sera dite appartenir au domaine d'attraction $\mathcal{D}(F)$ de G s'il existe deux suites $(a_n)_{n\in\mathbb{N}}$, $a_n > 0$. et (b_n) telles que :*

$$\lim_{n\to\infty} F^n(a_n\, x + b_n) = G(x)\,.$$

Pour tout x point de continuité de G.

Remarque 2.3.2 *Si G est stable, alors $G \in \mathcal{D}(G)$ et $\mathcal{D}(G)$ est non vide. En outre, on peut montrer qu'une fonction de répartition F ne peut appartenir qu'au domaine d'attraction d'une et une seule fonction G, représentant d'un type.*

Les caractéristiques des domaines d'attraction sont établies dans les théorèmes suivants.

Théorème 2.3.3 *Soit F une f.r de densité f positive admettant une dérivée f' négative sur (u,v), f étant nulle sur $[v,+\infty[$ (v fini ou non), si :*

$$\lim_{x \uparrow v} \frac{f'(x)(1-F(x))}{f^2(x)} = -1.$$

Alors : $F \in \mathcal{D}(F_1)$.

Théorème 2.3.4 *Soit F une f.r de densité positive f sur $[u,+\infty[$, si :*

$$\lim_{x \to +\infty} \frac{xf(x)}{1-F(x)} = \beta,\ \beta > 0.$$

Alors $F \in \mathcal{D}(F_2(.,\beta))$.

Théorème 2.3.5 *soit F une f.r de densité f positive sur un intervalle (u,v), et nulle sur $]v,+\infty[$, si :*

$$\lim_{x \uparrow v} \frac{(v-x)f(x)}{1-F(x)} = \beta,\ \beta > 0.$$

Alors : $F \in \mathcal{D}(F_3(.,\beta))$.

Exemple 2.3.3 ***a)*** *Soit* $X \rightsquigarrow \gamma(1,1)$, *loi exponenrielle de densité* $f(x) = e^{-x}\mathbf{1}_{\mathbb{R}_+}(x)$.
La fonction de répartition F_n *de* $X_{n,n}$ *est :*

$$F_n(x) = \left(1 - e^{-x}\right)^n \mathbf{1}_{\mathbb{R}_+}(x)$$

Soit :

$$Z_n = X_{(n)} - \ln n$$
$$F_{Z_n}(x) = F_n(z + \ln n) = \left(1 - \frac{1}{n}e^{-z}\right)^n - \exp\left(-e^{-z}\right) \text{ } quand \text{ } n \to \infty.$$

L'appartenance de la loi $\gamma(1,1)$ au domaine d'attraction de F_1 peut être retrouvée à I'aide de la condition du théorème12, en effet, pour $x > 0$:

$$\frac{f'(x)(1 - F(x))}{f^2(x)} = -\frac{e^{-x}e^{-x}}{(e^{-x})^2} = -1.$$

b) Soit X de loi logistique, de $f.r.$ F telle que :

$$F(x) = \frac{1}{1 + e^{-x}}$$

La $f.r$ de $X_{n,n}$ est $F_n(x) = (1 + e^{-x})^{-n}$. soit $Z_n = X_{n,n} - \ln n$,

$$F_{Z_n}(x) = \left\{1 + e^{-(x + \ln n)}\right\}^{-n} = \left(1 - \frac{1}{n}e^{-z}\right)^n \to \exp\left(-e^{-z}\right) \text{ } quand \text{ } n \to \infty.$$

De même que pour l'exemple **a)** :

$$\frac{f'(x)(1 - F(x))}{f^2(x)} = e^{-x} - 1 \to -1 \text{ } quand \text{ } n \to \infty.$$

c) Soit X suivant une loi de Cauchy de densité :

$$\begin{aligned} f(x) &= \frac{1}{\pi(1+x^2)}. \\ \frac{xf(x)}{1-F(x)} &= \frac{x}{(1+x^2)\left(\frac{\pi}{2}-Arctgx\right)} = \frac{x}{(1+x^2)\,Arctg\frac{1}{x}} pour\ x>0. \\ \lim_{x\to+\infty}\frac{xf(x)}{1-F(x)} &= \lim_{x\to+\infty}\frac{x^2}{1+x^2}=1. \end{aligned}$$

La loide Cauchy appartient à $\mathcal{D}(F_2(.,1))$.

d) On établit que la loi $\gamma(p,1)$ et la loi $N(0,1)$ appartiennent à $\mathcal{D}(F_1)$, et la loi uniforme $U_{[0,1]}$ à $\mathcal{D}(F_3(.,1))$.

Remarque 2.3.3 *On peut vérifier l'appartenance de chaque loi à son propre domaine d'attraction. Ainsi, pour $F_1(y)$:*

$$\begin{aligned} \frac{f'(y)(1-F(y))}{f^2(y)} &= e^{y+e^{-y}}\left(1-e^{-e^{-y}}\right)\left(e^{-y}-1\right) \\ &= e^{y}\left(e^{e^{-y}}-1\right)\left(e^{-y}-1\right). \end{aligned}$$

Or :

$$e^{e^{-y}}-1 \sim e^{-y} \quad quand\ y\to+\infty$$

et donc

$$\frac{f'(y)(1-F(y))}{f^2(y)} \to -1\ quand\ y\to+\infty$$

2.4 La loi de l'etendue

2.4.1 Résultat général

Un cas particulier de fonction des extrêmes très utilisé en statistique descriptive est l'étendue $W = X_{n,n} - X_{1,n}$, ou plus généralement les quasi-étendues de la forme $X_{(l)} - X_{(k)}\,(l > k)$.

Nous considérons la v.a $Z_{l,k} = X_{(l)} - X_{(k)}$. A partir de la loi du couple $\left(X_{(l)}, X_{(k)}\right)$ établie au paragraphe 2, on cherche la loi marginale de $Z_{l,k}$ après avoir fait le changement de variables

$$\begin{pmatrix} X_{(k)} \\ X_{(l)} \end{pmatrix} - \begin{pmatrix} U \\ Z_{l,k} \end{pmatrix} = \begin{pmatrix} X_{(k)} \\ X_{(l)} - X_{(k)} \end{pmatrix}.$$

La loi jointe du couple $(U, Z_{l,k})$ s'obtient simplement, en remarquant que la valeur absolue du jacobien du changement de variables est 1 :

$$\varphi(u,z) = \frac{1}{\mathcal{B}(k, l-k)\,\mathcal{B}(l, n-l+1)} f(u)\, f(u+z)$$
$$F^{k-1}(u)\,[1 - F(u+z)]^{n-1}\,[F(u+z) - F(u)]^{l-k-1}.$$

En appliquant l'intégrale par rapport à u, on obtient la densité $\varphi_{l,k}(z)$ de $Z_{l,k}$:

$$\varphi_{l,k}(z) = \frac{1}{\mathcal{B}(k, l-k)\,\mathcal{B}(l, n-l+1)} \int_{\mathbf{R}} F^{k-1}(u)\,[1 - F(u+z)]^{n-1}$$
$$[F(u+z) - F(u)]^{l-k-1} f(u)\, f(u+z)\, du.$$

A titre de cas particulier, considérons la longueur des mailles, c'est-à-dire l'écart entre deux coordonnées successive $Z_{k+1,k}\,(k = 1 \text{ à } n-1)$:

$$\varphi_{k+1,k}(z)=\frac{1}{\mathcal{B}(k,l-k)\,\mathcal{B}(l,n-l+1)}\int_{\mathbf{R}}F^{k-1}(u)\left[1-F(u+z)\right]^{n-k-1}f(u)\,f(u+z)\,du.$$

En calculant $E(Z_{k+1,k})$, on obtient après changement de variables :

$$E(Z_{k+1,k})=C_n^k\int_{\mathbf{R}}\left[1-F(v)\right]^{n-k}F^k(v)\,dv.$$

Exemple 2.4.1 *Soit X v.a de loi exponentielle $y(1)$.*

$$\begin{aligned}\varphi_{k+1,k}(z)&=\int_0^{+\infty}\frac{n!}{(k-1)!\,(n-k-1)!}\left(1-e^{-u}\right)^{k-1}e^{-(n-k-1)(z+u)}e^{-u}e^{-(z+u)}du\\&=\frac{n!e^{-(n-k)z}}{(k-1)!\,(n-k-1)!}\int_0^1(1-t)^{k-1}t^{n-k}dt.\end{aligned}$$

2.4.2 L'étendue W

a/ Loi de W

Pour obtenir la loi de l'étendue $W=X_{n,n}-X_{1,n}$ il suffit de faire $l=n$ et $k=1$ dans $\varphi_{l,k}(z)$. La densité $\varphi_{n,1}$ de W est donnée par :

$$\varphi_{n,1}(w)=n(n-1)\int_{\mathbf{R}}\left[F(w+u)-F(u)\right]^{n-2}f(u)\,f(w+u)\,du.$$

Sa fonction de répartition $F_{n,1}(w)$ est :

$$F_{n,1}(w)=\int_0^w\varphi_{n,1}(z)\,dz.$$

$$= \int_{-\infty}^{+\infty} n(n-1)\left(\int_0^w [F(z+u) - F(uvv)]^{n-2} f(u) f(z+u)\, dz\right) du$$
$$= \int_{-\infty}^{+\infty} n(n-1) f(u)\, du \left(\int_u^{w+u} [F(v) - F(u)]^{n-2} f(v)\, dv\right)$$
$$= \int_{-\infty}^{+\infty} n [F(u+w) - F(u)]^{n-1} dF(u).$$

b/ Espérance de l'étendue

On a

$$E(W) = E(X_{n,n}) - E(X_{1,n})$$

Avec les notations précédentes :

$$E(W) = \sum_{k=1}^{n-1} E(Z_{k+1,k}) = \sum_{k=1}^{n-1} \int_{\mathbf{R}} C_n^k (1 - F(v))^{n-k} F^k(v)\, dv$$
$$E(W) = \int_{\mathbf{R}} \sum_{k=1}^{n-1} C_n^k (1 - F(v))^{n-k} F^k(v)\, dv,$$

ainsi, on peut écrire :

$$E(W) = \int_{\mathbf{R}} [1 - [1 - F(v)]^n - F^n(v)]\, dv,$$

en utilisant :

$$\sum_{m=0}^{n} C_n^m [1-F]^{n-m} [F]^m = 1.$$

Exemple 2.4.2 *Soit* X *une variable aléatoire de loi uniforme sur* $[0, 1]$. *La densite de* $W = X_{n,n} - X_{1,n}$ *est :*

$$
\begin{aligned}
\varphi_{n,1}(w) &= n(n-1)\int_0^{1-w}[(u+w)-u]^{n-2}\,du \\
&= n(n-1)(1-w)\,w^{n-1}\mathbf{1}_{[0,1]}(w) \\
E(w) &= \int_0^1 [1-(1-w)^n - w^n]\,dw = \frac{(n-1)}{(n+1)}.
\end{aligned}
$$

Conclusion

L'objectif de ce mémoire, a été de traiter les statistiques d'ordre et d'obtenir la loi de chaque statistique correspondante. De plus, en s'intéresse à trouver les valeures extrêmes, ainsi que les caractéristiques qui nous donne plus des informations sur cette distribution.

En fait, les statistiques d'ordre constituent le point de départ en valeur extrême. Elles sont considérées comme un outil essentiel pour la modélisation des risques. Ceux-ci sont utilisés dans plusieurs applications. Les statistique d'ordre ont beaucoup d'importance dans le développement de la science, elles jouent un rôle très important dans la pratique, elles sont utilisées dans plusieurs domaines d'applications comme l'astronomie, l'acoustique et dans les sciences techniques et la médecine.

Bibliographie

[1] Arnold, B.C, Balakrishnan, N. et Nagaraja, H.N. (1992). A First course in order statistics.John Wiley & Sons, inc.

[2] Abdelmalek, B.B. (2017). mémoire Observation ordonnées et application. Biskra.

[3] Berkane. H. (2021) . cours mastre2. Biskra.

[4] Mohammad, A. et Valery, B.N. et Mohammad, S. (2013) . An introduction to order statistics. Atlantis Studies in Probability and Statistics, 3. Atlantis Press, Paris.

[5] Samah, B. (2010). mémoire determination du nombre de statistique díordre extremes. Biskra.

[6] Statistiques díordre https ://www.math.u-bordeaux.fr/~mchabano/Agreg/ProbaAgreg1213-COURS1-Statordre.pdf

[7] Tassi, P, (1985). Méthodes statistique. Economica, Paris.

Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous :

Abréviation ou notations	:	Explications
$\underset{\rightarrow}{P}$	:	convergence en probabilité
$E[X]$	:	espérance mathématique ou moyenne du v.a X.
$Var(X)$	:	variance
$v.a$	:	variable aléatoire.
$i.i.d$	:	identiquement et identiquement distribuée.
$f.r$	:	fonction de répartition.
H_γ	:	distribution des valeurs extrêmes (EVD).
F	:	fonction de répartition.
$\mathcal{D}(F)$	:	domaine d'attraction de F.
$\mathbf{1}_{C_n}$	:	fonction indicatrice de l'événement C_n.
$\Gamma(.)$	:	fonction de gamma.
a_k et b_k	:	des constantes de normalisation.
$cov(X_{k,n}, X_{s,n})$	:	covariance du $X_{k,n}$ et $X_{s,n}$.

$X_{1,n}$	:	minimum de $(X_1, X_{2,} \ldots, X_n)$
$X_{n,n}$	:	maximum de $(X_1, X_{2,} \ldots, X_n)$
$(X_{1,n}, X_{2,n}, \ldots, X_{n,n})$	:	statistique d'ordre associées $(X_1, \ldots, X_n)$
$(X_1, \ldots, X_n)$	:	échantillon de taille n va définie sur (Ω, F, P) de X
GEVD	:	distribution des valeurs extrêmes généralisée
(Ω, F, P)	:	espace de probabilité
Q	:	fonction inverse généralisée de F
F_n	:	fonction de répartition empirique
Q_n	:	fonction des quantiles empiriques
$s.o$	:	statistique d'ordre
GEV	:	valeurs extrêmes généralisé

Résumé

La statistique d'ordre est l'un des principaux outils de la statistique inférentielle. Les cas particuliers des statistiques d'ordre sont les valeurs minimale et maximale d'échantillon, d'où cette importance. Dans ce mémoire, nous avons présentés les différentes définitions, propriétés et des résultats généraux associées aux statistiques d'ordre relative à un échantillon et les comportements asymptotiques.

Mots-clés: statistique d'ordre, comportements asymptotique.

Abstract

The order statistics is one of the main tools of inferential statistics. The special cases of order statistics are the minimum and maximum sample values. In this memory, we have presented the various definitions, properties and general results associated with the order statistics relative to a sample and the asymptotic behavior.

Key words: order statistics, asymptotic behavior.

ملخص

إحصاء الترتيب تعد من بين الأدوات الأساسية في الإحصائيات و الاستدلالات. تتمثل الحالات الخاصة المهمة لإحصائيات الترتيب في الحد الأدنى و الحد الأقصى لقيمة العينة, و من هنا تأتي أهميتها. في هذه المذكرة نقوم بتقديم تعريفات مختلفة, الخصائص و النتائج العامة المرتبطة بالإحصائيات المرتبة للعينة و السلوك المقارب لها.

الكلمات المفتاحية: إحصاء الترتيب, السلوك المقارب.