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 pour obtenir le diplôme de

Master en "**Mathematiques Appliquées**"

Option : **Statistique**

Par

BEN SEGHIER Faten

Titre :

Statistiques d'Ordre et Théorie des Espacements

Devant le Jury :

NECIR Abdelhakim	**Pr**	UMKB	Président
BENAMEUR Sana	**M.C.B**	UMKB	Rapporteur
BERKANE HASSIBA	**M.A.A**	UMKB	Examinatreur

Soutenu Publiquement le 26/06/2022

Dédicace

It is with genuine gratitude and warm regard that i dedicate this work :

To my lovely parents, who have accompanied me in this process.

To my darling sisters and brothers, for their support and for helping me all the time.

To all my family members.

To my beloved friends, for their support and for the great time that we had inside and outside the college.

To all those who are dear to my heart.

Remerciements

*Firstly, i express my sincere gratitude, thanks and appreciation to my supervisor **Dr. BENAMEUR Sana**, for her great work, valuable and useful advices and her patience with me during this time.*

*I'm also deeply grateful to **Pr. RIFFI Mohamed I**, for his big help and valuable information.*

*I would like to extend a special thanks to the members of examination committee : **Pr. NECIR Abdelhakim** and **Dr. BERKANE Hassiba**, it's a big pleasure to me, to be the members of the examination committee for my memory.*

Thank's

Table des matières

Table des figures

Introduction

Les valeurs ordonnées dans l'ordre croissant d'un échantillon de variables aléatoire sont appelées les statistiques d'ordre. Ils ont un large éventail d'applications dans de nombreux domaines de statistique et de la probabilité, notamment l'estimation des paramètres de distributions, la théorie des valeurs extrêmes et les statistiques non paramétrique, etc.

Les espacements, les écarts entre les statistiques d'ordre successives, et leurs distributions sont également d'un grand intérêt dans la caractérisation de certaines distributions, en particulier la distribution uniforme et la distribution exponentielle. Ils sont très importants dans l'inférence statistique, par exemple sont utilisés dans les tests de qualité d'ajustement, l'estimation d'entropie non paramétrique et dans les tests d'uniformité, etc.

Dans la littérature, les statistiques d'ordre et leurs espacements ont reçu beaucoup d'attention de la part de nombreux chercheurs, citons le livre de David (1970 [4] *et* 1981 [5])*, David et Nagaraja* (2003) *[6], Arnold et al* (1992) [2]*, Ahsanullah et al* (2013) [1]*, Shahbaz et al* (2016) [13] *et le livre de Devroye* (1986) [7]*, ainsi que l'article de Pyke*(1965) [9]*, Riffi* (2006)[12]*.*

En raison de l'importance des statistiques d'ordre et leurs espacements, ce mémoire a pour objectif de présenter les différentes propriétés et distributions de ce type de données.

En plus d'une introduction et une conclusion, ce mémoire se compose de trois chapitres.

Chapitre 1 (Introduction aux statistiques d'ordre). *Dans ce chapitre, nous* *intéressons aux différentes notions et propriétés des statistiques d'ordre et leurs distributions dans le cas continu et discret.*

Chapitre 2 (Espacements uniformes et exponentiels). *Ce chapitre est principalement consacré à l'étude des espacements des statistiques d'ordre. On commence par explorer les espacements uniformes. Ensuite, nous illustrons les propriétés et les distributions des espacements exponentiels. Enfin, on introduit la relation entre ces deux types d'espacements.*

Chapitre 3 (Simulation sous R). *Nous appliquons dans ce chapitre les résultats les plus importants des deux chapitres précédents pour faire des simulations sous logiciel R. La première section conçue pour les différentes méthodes de génération des statistiques d'ordre. Alors que la deuxième section est dédiée à la simulation des espacements.*

Chapitre 1

Introduction aux statistiques d'ordre

Dans ce chapitre, on regroupe d'abord quelques définitions et propriétés des statistiques d'ordre, puis on procède progressivement à ses distributions dans le cas continu et le cas discret. Pour plus de détaille vous pouvez consulter les ouvrages : David et Nagaraja (2003) [6], Arnold et al (1992) [2], Ahsanullah et al (2013) [1], Shahbaz et al (2016) [13].

1.1 Statistiques d'ordre continues

Soit $X_1, X_2, \ldots, X_n$ une suite de variables aléatoires (v.a.'s) continues indépendantes et identiquement distribuées (i.i.d.), d'une fonction de densité f et d'une fonction de répartition F, avec $F(x) = \mathbb{P}(X \leq x),\ x \in \mathbb{R}$.

1.1.1 Définitions et propriétés

Définition 1.1.1 (Statistiques d'ordre)

Les statistiques d'ordre des variables aléatoires i.i.d. $X_1, X_2, \ldots, X_n$ est son réarrangement croissant, notée par $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$, avec

$$X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}.$$

Définition 1.1.2 (Statistiques d'ordre extrêmes)

Les variables définissent les statistiques d'ordre extrêmes sont

- *Statistique d'ordre minimum : est la plus petite statistique d'ordre, qui a pour définition*
$$X_{1:n} = \min\{X_1, X_2, \ldots, X_n\}.$$
- *Statistique d'ordre maximum : est la plus grande statistique d'ordre, qui a pour forme*
$$X_{n:n} = \max\{X_1, X_2, \ldots, X_n\}.$$

Remarque 1.1.1 *Pour $1 \leq k \leq n$, $X_{k:n}$ est appelée la $k^{ième}$ statistique d'ordre.*

Propriété 1.1.1

- *Les statistiques d'ordre sont dépendantes et non identiquement distribuées.*
- *La relation entre la statistique d'ordre minimum et la statistique d'ordre maximum est la suivante*
$$X_{1:n} = -\max\{-X_1, -X_2, \ldots, -X_n\}.$$
- *Si la loi est symétrique ($f(x) = f(-x)$, $F(x) = F(-x)$), alors*
$$X_{k:n} \stackrel{d}{=} -X_{n-k+1:n},$$
où $\stackrel{d}{=}$ désigne l'égalité en distribution.
- *Si on pose $U_{1:n}, U_{2:n}, \ldots, U_{n:n}$ les statistiques d'ordre associées à n v.a.'s indépendantes et uniformément distribuées sur $[0,1]$, alors*
$$\{F(X_{i:n}),\ 1 \leq i \leq n\} \stackrel{d}{=} \{U_{i:n},\ 1 \leq i \leq n\}.$$
$$\{F^{\leftarrow}(U_{i:n}),\ 1 \leq i \leq n\} \stackrel{d}{=} \{X_{i:n},\ 1 \leq i \leq n\},$$
où $F^{\leftarrow}$ est la fonction inverse généralisée de F.

- *La distance entre les statistiques d'ordre extrêmes est appelée l'étendue de l'échantillon, notée R, donnée par*

$$R = X_{n:n} - X_{1:n}.$$

1.1.2 Distributions d'une statistique d'ordre

Soient $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ les statistiques d'ordre associées à l'échantillon $(X_1, X_2, \ldots, X_n)$ de v.a.'s i.i.d., d'une fonction de répartition F et d'une densité f.

Distributions des statistiques d'ordre extrêmes

Lemme 1.1.1 *La fonction de répartition et la fonction de densité de la statistique d'ordre minimum $X_{1:n}$ sont données respectivement par*

$$F_{1:n}(x) = 1 - [1 - F(x)]^n, \ x \in \mathbb{R}, \tag{1.1}$$

et

$$f_{1:n}(x) = n[1 - F(x)]^{n-1} f(x), \ x \in \mathbb{R}.$$

Preuve.

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

$$f_{1:n}(x) = \frac{dF_{1:n}(x)}{dx} = n[1 - F(x)]^{n-1} f(x), \ x \in \mathbb{R}.$$

■

Lemme 1.1.2 *La fonction de répartition et la fonction de densité de la statistique d'ordre maximum $X_{n:n}$ sont données respectivement par*

$$F_{n:n}(x) = [F(x)]^n, \ x \in \mathbb{R},$$

et

$$f_{n:n}(x) = n[F(x)]^{n-1} f\left(x\right), \ x \in \mathbb{R}.$$

Preuve.

$$\begin{aligned} F_{n:n}(x) &= \mathbb{P}(X_{n:n} \leq x) \\ &= \mathbb{P}\left(\bigcap_{i=1}^{n} X_i \leq x\right) \\ &= \prod_{i=1}^{n} \mathbb{P}\left(X_i \leq x\right) \\ &= [F(x)]^n, \ x \in \mathbb{R}. \end{aligned}$$

$$f_{n:n}(x) = \frac{dF_{n:n}(x)}{dx} = n[F(x)]^{n-1} f\left(x\right), \ x \in \mathbb{R}.$$

■

Distribution de la $k^{ième}$ statistique d'ordre

La fonction de répartition de la $k^{ième}$ statistique d'ordre $X_{k:n}$, pour $1 \leq k \leq n$ est donnée par

$$F_{k:n}(x) = \sum_{r=k}^{n} C_n^r [F(x)]^r [1 - F(x)]^{n-r}, \ x \in \mathbb{R}, \tag{1.2}$$

et la fonction de densité de $X_{k:n}$ est la suivante

$$f_{k:n}(x) = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x), \ x \in \mathbb{R},$$

où $C_n^r = \dfrac{n!}{r!(n-r)!}$ est le nombre de combinaisons de r éléments parmi n éléments (sans remise).

Preuve.

$$
\begin{aligned}
F_{k:n}(x) &= \mathbb{P}(X_{k:n} \leq x) \\
&= \mathbb{P}(\text{Il y a un nombre supérieur ou égal à } k \text{ de } X_i \text{ inférieurs ou égal à } x) \\
&= \mathbb{P}\left(\sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq x\}} \geq k\right) \\
&= \sum_{r=k}^{n} C_n^r [F(x)]^r [1-F(x)]^{n-r},\ x \in \mathbb{R},
\end{aligned}
$$

où $\mathbb{I}_A$ est la fonction indicatrice de A telle que

$$
\mathbb{I}_A(x) = \begin{cases} 1 & si \quad x \in A, \\ 0 & si \quad x \notin A. \end{cases}
$$

Pour $1 \leq i \leq n$, les v.a.'s $\mathbb{I}_{\{X_i \leq x\}}$ étant i.i.d. de loi de Bernoulli de paramètre $F(x)$ ce qui implique que la loi de $\sum\limits_{i=1}^{n} \mathbb{I}_{\{X_i \leq x\}}$ est une Binomiale de paramètres n et $F(x)$ i.e.

$$
\sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq x\}} \sim Bin(n, F(x)).
$$

Pour la preuve de $f_{k:n}$, voir Shahbaz et al (2016) [13], page 10. ■

Lemme 1.1.3 *En utilisant la relation entre la somme des termes d'une loi binomiale et la fonction bêta incomplète on trouve une autre forme de $F_{k:n}$, donnée par*

$$
F_{k:n}(x) = \frac{n!}{(k-1)!(n-k)!} \int_0^{F(x)} t^{k-1}(1-t)^{n-k} dt,\ x \in \mathbb{R}. \tag{1.3}
$$

Preuve. *Pour $a > 0,\ b > 0$ et $0 \leq p \leq 1$, la fonction bêta incomplète donnée par*

$$
\begin{aligned}
I_p(a,b) &= \frac{\int_0^p t^{a-1}(1-t)^{b-1} dt}{B(a,b)} \\
&= \int_0^x \frac{(a+b-1)!}{(a-1)!(b-1)!} t^{a-1}(1-t)^{b-1} dt,\ 0 < x < 1,
\end{aligned}
$$

avec

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \frac{(x-1)!(y-1)!}{(x+y-1)!},$$

et

$$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt = (x-1)!,$$

sont la fonction bêta d'Euler et la fonction gamma d'Euler (respectivement).

En utilisant la relation

$$\sum_{r=i}^{n} C_n^r p^r (1-p)^{n-r} = \int_0^p \frac{n!}{(i-1)!(n-i)!} t^{i-1}(1-t)^{n-i}dt, \ 0 < p < 1,$$

on trouve

$$F_{k:n}(x) = \int_0^{F(x)} \frac{n!}{(k-1)!(n-k)!} t^{k-1}(1-t)^{n-k}dt = I_{F(x)}(k, n-k+1), \ x \in \mathbb{R}.$$

■

Moments de statistiques d'ordre

Pour $1 \leq k \leq n$ et $m \geq 1$, le moment d'ordre m de la $k^{ième}$ statistique d'ordre $X_{k:n}$ est donné par

$$u_{k:n}^{(m)} = \int_{-\infty}^{+\infty} x^m f_{k:n}(x)dx = \frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^{+\infty} x^m [F(x)]^{k-1}[1-F(x)]^{n-k} f(x)dx.$$

Il y a une autre forme du moment qui peut être obtenue en utilisant la propriété (1.1.1), pour $U_{1:n}, U_{2:n}, \ldots, U_{n:n}$ les statistiques d'ordre associées à n v.a.'s indépendantes et uniformément distribuées sur $[0,1]$, on trouve

$$u_{k:n}^{(m)} = \frac{n!}{(k-1)!(n-k)!} \int_0^1 [F^{\leftarrow}(u)]^m u^{k-1}[1-u]^{n-k}du. \tag{1.4}$$

1.1.3 Distribution jointe de statistiques d'ordre

Distribution jointe d'un couple de statistiques d'ordre

Lemme 1.1.4 *La fonction de répartition d'un couple de statistiques d'ordre* $(X_{k:n}, X_{j:n})$ *avec* $X_{k:n} \leq X_{j:n},\ 1 \leq k < j \leq n$*, est donnée par :*

Pour $x_k < x_j$, on a

$$F_{k,j:n}(x_k, x_j) = \sum_{r=j}^{n} \sum_{i=k}^{r} \frac{n!}{i!(r-i)!(n-r)!} [F(x_k)]^i [F(x_j) - F(x_k)]^{r-i} [1 - F(x_j)]^{n-r}. \tag{1.5}$$

Pour $x_k \geq x_j$, on a

$$F_{k,j:n}(x_k, x_j) = F_{j:n}(x_j).$$

La densité jointe du couple $(X_{k:n}, X_{j:n})$ est définie comme suit

$$f_{k,j:n}(x_k, x_j) = \frac{n!}{(k-1)!(j-k-1)!(n-j)!} \left[F\left(x_k\right)\right]^{k-1} f(x_k) \left[F(x_j) - F(x_k)\right]^{j-k-1} \tag{1.6}$$

$$\times f(x_j)[1 - F(x_j)]^{n-j},\ -\infty < x_k < x_j < +\infty.$$

Preuve. Voir Arnold et al (1992) [2], page $16 - 18$. ■

Corollaire 1.1.1 *La densité jointe des statistiques d'ordre extrêmes* $(X_{1:n}, X_{n:n})$ *est*

$$f_{1,n:n}(x_1, x_n) = n(n-1) \left[F(x_n) - F(x_1)\right]^{n-2} f(x_1) f(x_2),\ -\infty < x_1 < x_n < +\infty.$$

Lemme 1.1.5 *En utilisant la relation entre la somme des termes d'une loi binomiale et la fonction bêta incomplète on trouve une autre forme de* $F_{k,j:n}$*, pour* $x_k < x_j,\ 1 \leq k < j \leq n$

$$F_{k,j:n}(x_k, x_j) = \frac{n!}{(k-1)!(j-k-1)!(n-j)!} \int_0^{F(x_k)} \int_{t_1}^{F(x_j)} t_1^{k-1} (t_2 - t_1)^{j-k-1} (1 - t_2)^{n-j} dt_1 dt_2.$$

Preuve. Voir (Arnold et al, 1992) [2], page $18 - 19$. ■

Distribution jointe du n-uplet de statistiques d'ordre

Lemme 1.1.6 *Soient $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ les statistiques d'ordre associées à l'échantillon $(X_1, X_2, \ldots, X_n)$ de variables aléatoires i.i.d., pour $x_1, x_2, \ldots, x_n$ des observations avec $x_1 < x_2 < \ldots < x_n,\ x_k \in \mathbb{R},\ 1 \leq k \leq n$, la densité jointe du n-uplet de statistique d'ordre est donnée par*

$$f_{1,2,\ldots,n:n}(x_1, x_2, \ldots, x_n) = n! \prod_{k=1}^{n} f(x_k),\ -\infty < x_1 < x_2 < \ldots < x_n < +\infty. \quad (1.7)$$

Preuve.

$$\begin{aligned}
f_{1,2,\ldots,n:n}(x_1, x_2, \ldots, x_n) &= \lim_{\delta x_1 \to 0, \ldots, \delta x_n \to 0} \frac{\mathbb{P}\left(x_1 < X_{1:n} \leq x_1 + \delta x_1, \ldots, x_n < X_{n:n} \leq x_n + \delta x_n\right)}{\delta x_1 \ldots \delta x_n} \\
&= \lim_{\delta x_1 \to 0, \ldots, \delta x_n \to 0} \frac{n!\mathbb{P}\left(x_1 < X_1 \leq x_1 + \delta x_1, \ldots, x_n < X_n \leq x_n + \delta x_n\right)}{\delta x_1 \ldots \delta x_n} \\
&= \lim_{\delta x_1 \to 0, \ldots, \delta x_n \to 0} \frac{n!\mathbb{P}\left(x_1 < X_1 \leq x_1 + \delta x_1\right) \ldots \mathbb{P}(x_n < X_n \leq x_n + \delta x_n)}{\delta x_1 \ldots \delta x_n} \\
&= n! \prod_{k=1}^{n} f(x_k).
\end{aligned}$$

■

1.1.4 Fonction de répartition et quantile empirique

Soient $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ les statistiques d'ordre associées à l'échantillon $(X_1, X_2, \ldots, X_n)$ de variables aléatoires i.i.d., d'une fonction de répartition commune F.

Définition 1.1.3 (Fonction de répartition empirique)

La fonction de répartition empirique de l'échantillon $(X_1, X_2, \ldots, X_n)$ est donnée par

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq x\}} = \begin{cases} 0 & si \quad x < X_{1:n}, \\ \dfrac{i}{n} & si \quad X_{i:n} \leq x < X_{i+1:n},\ 1 \leq i \leq n-1, \\ 1 & si \quad x \geq X_{n:n}. \end{cases}$$

Définition 1.1.4 (Quantile)

La fonction quantile associée à F est l'inverse généralisée de F, définie par

$$Q(\alpha) = F^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}, \ 0 < \alpha < 1.$$

Définition 1.1.5 (Quantile empirique)

La fonction quantile empirique d'un échantillon $(X_1, X_2, \ldots, X_n)$ est définie par

$$Q_n(\alpha) = F_n^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : F_n(x) \geq \alpha\}, \ 0 < \alpha < 1.$$

Remarque 1.1.2 *La fonction quantile entraîne que*

$$Q_n(\alpha) = X_{i:n}, \ \frac{i-1}{n} < \alpha \leq \frac{i}{n}, \ 1 \leq i \leq n.$$

1.2 Statistiques d'ordre discrètes

Soit $(X_1, X_2, \ldots, X_n)$ un échantillon de variables aléatoires i.i.d. discrètes et $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ les statistiques d'ordre associées à cet échantillon, on suppose que les v.a.'s ont des valeurs dans $\mathbb{N}$.

1.2.1 Fonction de masse d'une statistique d'ordre

Pour trouver la fonction de masse (noté pmf : probability mass function) de la $k^{ième}$ statistique d'ordre dans le cas discret on utilise la fonction de répartition $F_{k:n}$ définie dans la section (1.1.2), comme suit

$$\begin{aligned} f_{k:n}(x) &= \mathbb{P}(X_{k:n} = x) \\ &= \mathbb{P}(X_{k:n} \leq x) - \mathbb{P}(X_{k:n} < x) \\ &= F_{k:n}(x) - F_{k:n}(x-), \end{aligned}$$

où $F(x-) = P(X < x)$.

En utilisant la première forme de $F_{k:n}$ (somme binomiale, 1.2), on trouve

$$\begin{aligned} f_{k:n}(x) &= \sum_{r=k}^{n} C_n^r [F(x)]^r [1-F(x)]^{n-r} - \sum_{r=k}^{n} C_n^r [F(x-)]^r [1-F(x-)]^{n-r} \\ &= \sum_{r=k}^{n} C_n^r [(F(x))^r (1-F(x))^{n-r} - (F(x-))^r (1-F(x-))^{n-r}]. \end{aligned}$$

En utilisant la deuxième forme de $F_{k:n}$ (bêta incomplète, 1.3), on obtient une autre représentation pour $f_{k:n}(x)$, telle que

$$\begin{aligned} f_{k:n}(x) &= \frac{n!}{(k-1)!(n-k)!} \int_0^{F(x)} u^{k-1}(1-u)^{n-k} du - \frac{n!}{(k-1)!(n-k)!} \int_0^{F(x-)} u^{k-1}(1-u)^{n-k} du \\ &= \frac{n!}{(k-1)!(n-k)!} \left[\int_0^{F(x)} u^{k-1}(1-u)^{n-k} du - \int_0^{F(x-)} u^{k-1}(1-u)^{n-k} du \right] \\ &= \frac{n!}{(k-1)!(n-k)!} \int_{F(x-)}^{F(x)} u^{k-1}(1-u)^{n-k} du. \end{aligned}$$

Remarque 1.2.1 *La fonction de masse des statistiques d'ordre extrêmes $X_{1:n}$ et $X_{1:n}$ (repectivement)*

$$\begin{aligned} f_{1:n}(x) &= \mathbb{P}(X_{1:n} = x) \\ &= \mathbb{P}(X_{1:n} \leq x) - \mathbb{P}(X_{1:n} < x) \\ &= F_{1:n}(x) - F_{1:n}(x-) \\ &= [1-F(x-)]^n - [1-F(x)]^n . \end{aligned}$$

$$\begin{aligned} f_{n:n}(x) &= \mathbb{P}(X_{n:n} = x) \\ &= \mathbb{P}(X_{n:n} \leq x) - \mathbb{P}(X_{n:n} < x) \\ &= F_{n:n}(x) - F_{n:n}(x-) \\ &= [F(x)]^n - [F(x-)]^n . \end{aligned}$$

Moments de statistiques d'ordre

Les moments dans le cas discret ont la même forme que dans le cas continu, donné dans l'équation (1.4). Par exemple, pour $1 \leq k \leq n$, l'espérance de la $k^{ième}$ statistique d'ordre $X_{k:n}$ est donné par

$$u_{k:n} = \frac{n!}{(k-1)!(n-k)!} \int_0^1 F^{\leftarrow}(u) u^{k-1} (1-u)^{n-k} \, du,$$

puisque $F^{\leftarrow}(u)$ n'a pas une bonne forme pour la plupart des distributions discrètes cette approche est souvent peu pratique.

Le support de plusieurs distributions discrètes standard est un sous-ensemble d'entiers non négatifs, il est donc possible d'utiliser la fonction de répartition directement pour obtenir les moments de $X_{k:n}$. Par exemple pour $1 \leq k \leq n$, l'espérance de la $k^{ième}$ statistique d'ordre est donné par

$$u_{k:n} = \sum_{x=0}^{\infty} [1 - F_{k:n}(x)],$$

et le moment d'ordre 2 de $X_{k:n}$ est donné par

$$u_{k:n}^{(2)} = 2 \sum_{x=0}^{\infty} x [1 - F_{k:n}(x)] + u_{k:n}.$$

Preuve. Remarquons tout d'abord que si $u_{k:n}$ existe, $r\mathbb{P}(X_{k:n} > r) \to 0$ quand $r \to \infty$. Pour l'espérance on a

$$\begin{aligned} \sum_{x=0}^{r} x f_{k:n}(x) &= \sum_{x=0}^{r} x \left[\mathbb{P}(X_{k:n} > x-1) - \mathbb{P}(X_{k:n} > x)\right] \\ &= \sum_{x=0}^{r-1} [(x+1) - x] \mathbb{P}(X_{k:n} > x) - r\mathbb{P}(X_{k:n} > r), \\ &= \sum_{x=0}^{r-1} \mathbb{P}(X_{k:n} > x) - r\mathbb{P}(X_{k:n} > r), \end{aligned}$$

alors

$$u_{k:n} = \sum_{x=0}^{\infty} x f_{k:n}(x) = \lim_{r\to\infty} \sum_{x=0}^{r} x f_{k:n}(x) = \sum_{x=0}^{\infty} \mathbb{P}(X_{k:n} > x) = \sum_{x=0}^{\infty} [1 - F_{k:n}(x)].$$

Pour le moment d'ordre 2 on a

$$\begin{aligned}
u_{k:n}^{(2)} &= \sum_{x=0}^{\infty} x^2 f_{k:n}(x) = \lim_{r\to\infty} \sum_{x=0}^{r} x^2 f_{k:n}(x) \\
&= \lim_{r\to\infty} \sum_{x=0}^{r} x^2 \left[\mathbb{P}(X_{k:n} > x-1) - \mathbb{P}(X_{k:n} > x)\right] \\
&= \lim_{r\to\infty} \sum_{x=0}^{r-1} [(x+1)^2 - x^2] \mathbb{P}(X_{k:n} > x) - \lim_{r\to\infty} r^2 \mathbb{P}(X_{k:n} > r) \\
&= \lim_{r\to\infty} 2 \sum_{x=0}^{r-1} x \mathbb{P}(X_{k:n} > x) + \lim_{r\to\infty} \sum_{x=0}^{r-1} \mathbb{P}(X_{k:n} > x) - \lim_{r\to\infty} r^2 \mathbb{P}(X_{k:n} > r) \\
&= 2 \sum_{x=0}^{\infty} x \mathbb{P}(X_{k:n} > x) + \sum_{x=0}^{\infty} \mathbb{P}(X_{k:n} > x) \\
&= 2 \sum_{x=0}^{\infty} x [1 - F_{k:n}(x)] + u_{k:n}.
\end{aligned}$$

■

1.2.2 Fonction de masse jointe de statistiques d'ordre

Soient $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ les statistiques d'ordre associées à un échantillon de variables aléatoires i.i.d. discrètes.

Théorème 1.2.1 (Arnold et al (1992)**)**

Soient $x_1 \leq x_2 \leq \ldots \leq x_n$ des observations tel que $x_1 = \ldots = x_{r_1} < x_{r_1+1} = \ldots = x_{r_2} < \ldots = x_{r_{m-1}} < x_{r_{m-1}+1} = \ldots = x_{r_m} = x_n$, $1 \leq r_1 < r_2 < \ldots < r_m = n$.

Pour $1 \leq m \leq n$ et $r_0 = 0$ la fonction de masse jointe du n-uplet de statistiques d'ordre est donnée par

$$f_{1,2,\ldots,n:n}(x_1, x_2, \ldots, x_n) = n! \prod_{s=1}^{m} \frac{[f(x_s)]^{r_s - r_{s-1}}}{(r_s - r_{s-1})!}.$$

Théorème 1.2.2 (Arnold et al (1992)**)**

Pour $i_0 = 0,\ 1 \leq i_1 < i_2 < \ldots < i_k \leq n,\ k < n$ *et* $x_{i_1} \leq x_{i_2} \leq \ldots \leq x_{i_k}$, *la fonction de masse jointe de* $X_{i_1:n}, X_{i_2:n}, \ldots, X_{i_k:n}$ *est donnée par*

$$f_{i_1,i_2,\ldots,i_k:n}(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = \frac{n!}{(n-i_k)!\prod\limits_{r=1}^{k}(i_r - i_{r-1} - 1)!}$$
$$\times \int\limits_B \left[\prod_{r=1}^{k}\left(u_{i_r} - u_{i_{r-1}}\right)^{i_r - i_{r-1} - 1}\right](1 - u_{i_k})^{n-i_k}\, du_{i_1}\ldots du_{i_k},$$

où $u_0 = 0$ *et* $B = \{(u_{i_1}, \ldots, u_{i_k}) : u_{i_1} \leq u_{i_2} \leq \ldots \leq u_{i_k},\ F(x_r-) \leq u_r \leq F(x_r),\ r = i_1, \ldots, i_k\}$.

Corollaire 1.2.1 *Pour* $1 \leq i_1 < i_2 \leq n$ *et* $x_{i_1} \leq x_{i_2}$, *la fonction de masse jointe du couple* $(X_{i_1:n}, X_{i_2:n})$ *est donnée par*

$$f_{i_1,i_2:n}(x_{i_1}, x_{i_2}) = \frac{n!}{(i_1-1)!(i_2-i_1-1)!(n-i_2)!}\int\limits_B u_{i_1}^{i_1-1}(u_{i_2} - u_{i_1})^{i_2-i_1-1}(1-u_{i_2})^{n-i_2}\, du_{i_1}du_{i_2},$$

où $B = \{(u_{i_1}, u_{i_2}) : u_{i_1} \leq u_{i_2},\ F(x_r-) \leq u_r \leq F(x_r),\ r = i_1, i_2\}$.

Lemme 1.2.1 *Soit* $(X_{k:n}, X_{j:n})$ *un couple de statistiques d'ordre avec* $X_{k:n} \leq X_{j:n}$ $x_k \leq x_j,\ 1 \leq k < j \leq n$ *la fonction de masse jointe du couple* $(X_{k:n}, X_{j:n})$ *est donnée par*

$$f_{k,j:n}(x_k, x_j) = \frac{n!}{(k-1)!(j-k-1)!(n-j)!}\int\limits_{F(x_k-)}^{F(x_k)}\int\limits_{F(x_j-)}^{F(x_j)} u_k^{k-1}(u_j - u_k)^{j-k-1}(1-u_j)^{n-j}du_k du_j.$$

Preuve. *En utilisant le corollaire* (1.2.1), *pour* $x_k \leq x_j,\ 1 \leq k < j \leq n$, *la fonction de masse jointe du couple* $(X_{k:n}, X_{j:n})$ *est donnée par*

$$f_{k,j:n}(x_k, x_j) = \frac{n!}{(k-1)!(j-k-1)!(n-j)!}\int\limits_B u_k^{k-1}(u_j - u_k)^{j-k-1}(1-u_j)^{n-j}du_k du_j,$$

où $B = \{(u_k, u_j) : u_k \leq u_j,\ F(x_r-) \leq u_r \leq F(x_r),\ r = k, j\}$. ■

Chapitre 2

Espacements uniformes et exponentiels

Un autre ensemble de statistiques utiles pour comprendre la distribution de probabilité sont les espacements, qui sont les écarts entre les statistiques d'ordre successifs. Dans ce chapitre, on s'intéresse aux distributions et à certaines propriétés des espacements des statistiques d'ordre uniforme et exponentiel. Les références de base utilisées sont : Pyke (1965) [10], Devroye (1986) [7] et Riffi (2006) [12].

2.1 Espacements uniformes

2.1.1 *Définition*

Soit $(X_1, X_2, \ldots, X_n)$ un échantillon de v.a.'s i.i.d. d'une loi uniforme sur $[0, 1]$, telle que

$$f(x) = \begin{cases} 1 & si\ 0 \leq x \leq 1, \\ 0 & \text{sinon}, \end{cases} \tag{2.1}$$

et

$$F(x) = \begin{cases} 0 & si \quad x < 0, \\ x & si \quad 0 \leq x \leq 1, \\ 1 & si \quad x > 1. \end{cases} \tag{2.2}$$

Soit $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ les statistiques d'ordre associées à cet échantillon.

D'après les équations (1.7) et (2.1) la densité jointe du n-uplet de statistique d'ordre $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ est

$$f_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n) = \begin{cases} n! & si \quad 0 \leq x_1 < x_2 < \ldots < x_n \leq 1, \\ 0 & \text{sinon.} \end{cases}$$

Définition 2.1.1 (Espacements uniformes)

Les espacements uniformes de l'échantillon $(X_1, X_2, \ldots, X_n)$ sont définis par les statistiques

$$D_i = X_{i:n} - X_{i-1:n}, \; 1 \leq i \leq n+1,$$

avec par convention $X_{0:n} = 0$ et $X_{n+1:n} = 1$.

Remarque 2.1.1

1. $D_i \geq 0, \; 1 \leq i \leq n+1.$
2. $\sum_{i=1}^{n+1} D_i = 1.$

2.1.2 *Distributions des* espacements uniformes

Théorème 2.1.1 (Pyke (1965) [10]**)**

Le vecteur des espacements uniformes $D = (D_1, D_2, \ldots, D_{n+1})$ est uniformément distribué sur le simplex

$$A_n = \left\{ (d_1, d_2, \ldots, d_n); d_i \geq 0, \; \sum_{i=1}^{n} d_i \leq 1 \right\},$$

avec la densité de probabilité jointe de D est

$$f_D(d_1, d_2, ..., d_{n+1}) = \begin{cases} n! & si \; d_i \geq 0, \; d_1 + d_2 + \ldots + d_{n+1} = 1, \\ 0 & sinon. \end{cases} \tag{2.3}$$

Preuve. La preuve se poursuit en transformant l'ensemble des espacements à partir de l'ensemble des statistiques d'ordre et en appliquant le théorème de densité jacobien. On note la transformation inverse

$$\begin{pmatrix} d_1 = x_1 \\ d_2 = x_2 - x_1 \\ \vdots \\ d_i = x_i - x_{i-1} \\ \vdots \\ d_n = x_n - x_{n-1} \end{pmatrix} \overset{\Psi^{-1}}{\longrightarrow} \begin{pmatrix} x_1 = d_1 \\ x_2 = d_1 + d_2 \\ \vdots \\ x_i = d_1 + d_2 + \ldots + d_i \\ \vdots \\ x_n = d_1 + d_2 + \ldots + d_n \end{pmatrix}.$$

Le jacobien de cette transformation, i.e. le déterminant de la matrice formée par $\dfrac{\partial x_i}{\partial d_i}$ $(i = 1, \ldots, n)$ est

$$\begin{vmatrix} \frac{\partial x_1}{\partial d_1} & \frac{\partial x_1}{\partial d_2} & \cdots & \frac{\partial x_1}{\partial d_n} \\ \frac{\partial x_2}{\partial d_1} & \frac{\partial x_2}{\partial d_2} & \cdots & \frac{\partial x_2}{\partial d_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial d_1} & \frac{\partial x_n}{\partial d_2} & \cdots & \frac{\partial x_n}{\partial d_n} \end{vmatrix} = \begin{vmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 1 & & \cdots & 1 \end{vmatrix} = 1.$$

Alors la densité jointe du vecteur D est

$$\begin{aligned} f_D(d_1, d_2, \ldots, d_n) &= f_{1,2,\ldots,n}(\Psi^{-1}(d_1, d_2, \ldots, d_n)) \left| J_{\Psi^{-1}}(d_1, d_2, \ldots, d_n) \right| \\ &= \begin{cases} n!, & si\ d_i \geq 0,\ d_1 + d_2 + \ldots + d_n \leq 1, \\ 0 & \text{sinon.} \end{cases} \end{aligned}$$

Utilisant l'équivalence entre le simplex $\{d_i \geq 0\ (1 \leq i \leq n),\ d_1 + d_2 + \ldots + d_n \leq 1\}$ et le simplex $\{d_i \geq 0\ (1 \leq i \leq n+1),\ d_1 + d_2 + \ldots + d_{n+1} = 1\}$, la densité jointe du vecteur D est

$$f_D(d_1, d_2, \ldots, d_{n+1}) = \begin{cases} n! & si\ d_i \geq 0,\ d_1 + d_2 + \ldots + d_{n+1} = 1, \\ 0 & \text{sinon.} \end{cases}$$

■

De l'équation (2.3), on remarque que la densité jointe du vecteur D est constante (indépendante des espacements uniformes), cela signifie que les espacements uniformes sont des v.a.'s interchangeables. Cela implique que la distribution de D_i $(1 \leq i \leq n+1)$ est à celle de D_1 et la distribution jointe du couple (D_i, D_j) avec $(i \neq j)$ est la même que celle de (D_1, D_2).

Théorème 2.1.2 *Soient $D_1, D_2, \ldots, D_{n+1}$ les espacements uniformes de l'échantillon $(X_1, X_2, \ldots, X_n)$, la fonction de répartition et la densité de D_i $(1 \leq i \leq n+1)$ sont données respectivement par*

$$F_{D_i}(x) = 1 - (1-x)^n, \ x \in [0 \ 1],$$

et

$$f_{D_i}(x) = n(1-x)^{n-1}, \ x \in [0 \ 1]. \tag{2.4}$$

En effet, en utilisant les résultats du lemme (1.1.1) et les équations (2.1) et (2.2) on trouve

$$F_{D_i}(x) = F_{D_1}(x) = F_{X_{1:n}}(x) = 1 - (1-x)^n, \ x \in [0 \ 1].$$

$$f_{D_i}(x) = n(1-x)^{n-1}, \ x \in [0 \ 1].$$

Théorème 2.1.3 *Pour $x, y \geq 0$, $x + y \leq 1$ et $(i \neq j)$ la densité et la fonction répartition jointe du couple (D_i, D_j) sont données respectivement par*

$$f_{(D_i,D_j)}(d_1, d_2) = n(n-1)(1-d_1-d_2)^{n-2}, \tag{2.5}$$

et

$$F_{(D_i,D_j)}(d_1, d_2) = 1 - (1-d_1)^n - (1-d_2)^n + (1-d_1-d_2)^n.$$

Preuve. On a $f_{(D_i,D_j)} = f_{(D_1,D_2)}$, pour trouver la densité jointe du couple (D_1, D_2) on utilise le théorème de densité jacobien.

On note la transformation inverse

$$\begin{pmatrix} d_1 = x_1 \\ d_2 = x_2 - x_1 \end{pmatrix} \overset{\Psi^{-1}}{\longrightarrow} \begin{pmatrix} x_1 = d_1 \\ x_2 = d_1 + d_2 \end{pmatrix}.$$

Le jacobien de cette transformation inverse vaut 1. En utilisant les équations (1.6) (2.1) et (2.2), on trouve la densité jointe du couple $(X_{1:n}, X_{2:n})$ est donnée par

$$f_{1,2:n}(x_1, x_2) = n(n-1)(1-x_2)^{n-2}.$$

Pour $x, y \geq 0$ et $x + y \leq 1$, la formule du changement de variable donne

$$\begin{aligned} f_{(D_1,D_2)}(d_1, d_2) &= f_{1,2:n}(\Psi^{-1}(d_1, d_2))\,|J_{\Psi^{-1}}(d_1, d_2)| \\ &= n(n-1)(1-d_1-d_2)^{n-2}, \end{aligned}$$

alors

$$f_{(D_i,D_j)}(d_1, d_2) = n(n-1)(1-d_1-d_2)^{n-2}.$$

Pour trouver la fonction de répartition du couple (D_i, D_j) on va intégrer la densité jointe du couple (D_i, D_j) deux fois comme suit

$$\begin{aligned} F_{(D_i,D_j)}(d_1, d_2) &= \int_0^{d_1}\int_0^{d_2} f_{(D_i,D_j)}(t, z)dzdt \\ &= \int_0^{d_1}\int_0^{d_2} n(n-1)(1-t-z)^{n-2}dzdt \\ &= \int_0^{d_1}\int_0^{d_2} n(n-1)(1-t)^{n-2}\left(1-\frac{z}{1-t}\right)^{n-2} dzdt \\ &= \int_0^{d_1} n(n-1)(1-t)^{n-2}\int_0^{d_2}\left(1-\frac{z}{1-t}\right)^{n-2} dzdt, \end{aligned}$$

on pose $q = 1 - \dfrac{z}{1-t} \Rightarrow -(1-t)dq = dz$, alors

$$
\begin{aligned}
F_{(D_i,D_j)}(d_1,d_2) &= \int_0^{d_1} n(n-1)(1-t)^{n-2} \int_1^{1-\frac{d_2}{1-t}} -q^{n-2}(1-t)dqdt \\
&= \int_0^{d_1} n(n-1)(1-t)^{n-1}\left[1-\left(1-\frac{d_2}{1-t}\right)^{n-1}\right]\frac{1}{(n-1)}dt \\
&= \int_0^{d_1} n(1-t)^{n-1}dt - \int_0^{d_1} n(1-t)^{n-1}\left(1-\frac{d_2}{1-t}\right)^{n-1} dt \\
&= \int_0^{d_1} n(1-t)^{n-1}dt - \int_0^{d_1} n(1-t-d_2)^{n-1}dt,
\end{aligned}
$$

pour la première intégrale on pose $v = 1-t \Rightarrow -dv = dt$, alors

$$
I_1 = \int_1^{1-d_1} -n(v)^{n-1}dv = 1-(1-d_1)^n,
$$

et pour la deuxième intégrale on pose $v = 1-t-d_2 \Rightarrow -dv = dt$, alors

$$
I_2 = \int_{1-y}^{1-d_1-d_2} n(v)^{n-1}dv = (1-d_1-d_2)^n - (1-d_2)^n,
$$

on déduit que

$$
F_{(D_i,D_j)}(d_1,d_2) = 1-(1-d_1)^n-(1-d_2)^n+(1-d_1-d_2)^n.
$$

■

2.1.3 Propriétés des espacements uniformes

Soient $D_1, D_2, \ldots, D_{n+1}$ les espacements uniformes de l'échantillon $(X_1, X_2, \ldots, X_n)$ pour $i = 1, 2, \ldots, n$ et $j = 1, 2, \ldots, n$ avec $i \neq j$, on a

- Le moment d'ordre m : $E(D_i^m) = \dfrac{m!n!}{(n+m)!}$.

- La variance : $Var(D_i) = \dfrac{n}{(n+1)^2(n+2)}$.
- La covariance : $Cov(D_i, D_j) = -\dfrac{1}{(n+1)^2(n+2)}$, tel que $E(D_i, D_j) = \dfrac{1}{(n+1)(n+2)}$.

En effet, en utilisant l'équation (2.4), on trouve

$$\begin{aligned}
E(D_i^m) &= \int_0^1 x^m f_{D_i}(x)dx = \int_0^1 nx^m(1-x)^{n-1}dx \\
&= nB(m+1, n) = n\frac{m!(n-1)!}{(n+m)!} = \frac{m!n!}{(n+m)!}.
\end{aligned}$$

$$Var(D_i) = E(D_i^2) - E(D_i)^2 = \frac{2}{(n+1)(n+2)} - \frac{1}{(n+1)^2} = \frac{n}{(n+1)^2(n+2)}.$$

En utilisant l'équation (2.5), on trouve

$$\begin{aligned}
E(D_i, D_j) &= \int_0^1 \int_0^{1-x} xyf_{(D_i,D_j)}(x,y)dydx \\
&= \int_0^1 \int_0^{1-x} n(n-1)xy(1-x-y)^{n-2}dydx \\
&= \int_0^1 \int_0^{1-x} n(n-1)x\frac{y}{1-x}(1-x)^{n-1}\left(1-\frac{y}{1-x}\right)^{n-2} dydx \\
&= \int_0^1 n(n-1)x(1-x)^{n-1} \int_0^{1-x} \frac{y}{1-x}\left(1-\frac{y}{1-x}\right)^{n-2} dydx,
\end{aligned}$$

on pose $r = \dfrac{y}{1-x}$, alors

$$\begin{aligned}
E(D_i, D_j) &= \int_0^1 n(n-1)x(1-x)^n dx \int_0^1 r(1-r)^{n-2}dr \\
&= n(n-1)B(2, n+1)B(2, n-1) \\
&= n(n-1)\frac{(n)!}{(n+2)!}\frac{(n-2)!}{(n)!} \\
&= \frac{1}{(n+1)(n+2)}.
\end{aligned}$$

Par conséquent

$$\begin{aligned}
Cov(D_i, D_j) &= E(D_i, D_j) - E(D_i)E(D_j) \\
&= \frac{1}{(n+1)(n+2)} - \frac{1}{(n+1)^2} \\
&= -\frac{1}{(n+1)^2(n+2)}.
\end{aligned}$$

Théorème 2.1.4 *Soient $E_1, E_2, \ldots, E_{n+1}$ une suite de v.a.'s i.i.d. de loi exponentielle standard. Les espacements uniformes $D_1, D_2, \ldots, D_{n+1}$ ont la même distribution que*

$$\frac{E_1}{S}, \frac{E_2}{S}, \ldots, \frac{E_{n+1}}{S},$$

pour $S = \sum_{i=1}^{n+1} E_i$.

Preuve. *Premièrement, nous devons savoir que la densité jointe de $(E_1, E_2, \ldots, E_n, S)$ est donné par*

$$\begin{aligned}
f_{(E_1,E_2,\ldots,E_n,E_{n+1})}(e_1, e_2, \ldots, e_n, e_{n+1}) &= \prod_{i=1}^{n} e^{-e_i} e^{-e_{n+1}} \\
&= \prod_{i=1}^{n} e^{-e_i} e^{-(s-e_1-e_2-\ldots-e_n)} \\
&= e^{-s},\ \ s > 0,\ \ e_i \geq 0,\ \ i = 1, 2, \ldots, n.
\end{aligned}$$

On note la transformation inverse suivant

$$\left(\begin{array}{l}
y_1 = \dfrac{e_1}{s} \\
y_2 = \frac{e_2}{s} \\
\vdots \\
y_n = \dfrac{e_n}{s} \\
y_{n+1} = s \\
\qquad = \sum_{i=q}^{n+1} e_i = e_1 + e_2 + \ldots + e_n + e_{n+1}
\end{array}\right) \overset{\Psi^{-1}}{\longrightarrow} \left(\begin{array}{l}
e_1 = sy_1 \\
e_2 = sy_2 \\
\vdots \\
e_n = sy_n \\
e_{n+1} = y_{n+1} - e_1 - e_2 - \ldots e_n \\
\qquad = y_{n+1} - s(y_1 + y_2 + \ldots y_n)
\end{array}\right).$$

Le jacobien de cette transformation inverse est

$$\begin{vmatrix} \frac{\partial e_1}{\partial y_1} & \frac{\partial e_1}{\partial y_2} & \dots & \frac{\partial e_1}{\partial y_n} & \frac{\partial e_1}{\partial y_{n+1}} \\ \frac{\partial e_2}{\partial y_1} & \frac{\partial e_2}{\partial y_2} & \dots & \frac{\partial e_2}{\partial y_n} & \frac{\partial e_2}{\partial y_{n+1}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial e_n}{\partial y_1} & \frac{\partial e_n}{\partial y_2} & \dots & \frac{\partial e_n}{\partial y_n} & \frac{\partial e_n}{\partial y_{n+1}} \\ \frac{\partial e_{n+1}}{\partial y_1} & \frac{\partial e_{n+1}}{\partial y_2} & \dots & \frac{\partial e_{n+1}}{\partial y_n} & \frac{\partial e_{n+1}}{\partial y_{n+1}} \end{vmatrix} = \begin{vmatrix} s & 0 & \dots & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & \ddots & \ddots & 0 & \vdots \\ \vdots & & & \ddots & s & 0 \\ -s & -s & \dots & \dots & -s & 1 \end{vmatrix} = s^n.$$

Alors, la densité jointe du vecteur $(Y_1, Y_2, \dots, Y_n, Y_{n+1})$ *est*

$$f_{(Y_1,Y_2,\dots,Y_n,Y_{n+1})}(y_1, y_2, \dots, y_n, s) = f_{(E_1,E_2,\dots,E_n,E_{n+1})}(\Psi^{-1}(y_1, y_2, \dots, y_n, s)) \tag{2.6}$$

$$\times |J_{\Psi^{-1}}(y_1, y_2, \dots, y_n, s)| \tag{2.7}$$

$$= e^{-s}s^n, \;\; s > 0, \;\; y_i \geq 0 \; (1 \leq i \leq n), \; \sum_{i=1}^{n} y_i \leq 1.$$

Finalement, la densité marginale de $(Y_1, Y_2, \dots, Y_n)$ *est obtenue en intégraut la densité jointe du vecteur* $(Y_1, Y_2, \dots, Y_n, Y_{n+1})$ (2.6) *par rapport à* s, *ce qui nous donne*

$$f_{(Y_1,Y_2,\dots,Y_n)}(y_1, y_2, \dots, y_n) = \int_0^{+\infty} e^{-s}s^n ds$$

$$= n!, \;\; y_i \geq 0 \; (1 \leq i \leq n), \; \sum_{i=1}^{n} y_i \leq 1,$$

alors, la densité jointe du vecteur $(Y_1, Y_2, \dots, Y_{n+1}) = \left(\frac{E_1}{S}, \frac{E_2}{S}, \dots, \frac{E_n}{S}, \frac{E_{n+1}}{S}\right)$ *est*

$$f_{(Y_1,Y_2,\dots,Y_{n+1})}(y_1, y_2, \dots, y_{n+1}) = n!, \;\; y_i \geq 0 \; (1 \leq i \leq n+1), \; \sum_{i=1}^{n+1} y_i \leq 1.$$

$\blacksquare$

2.2 Espacements exponentiels

2.2.1 *Définition*

Soit $(X_1, X_2, \ldots, X_n)$ un échantillon de v.a.'s i.i.d. de loi exponentielle de paramètre $\lambda > 0$, telle que

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & si \quad x \geq 0, \\ 0 & si \quad x < 0, \end{cases} \tag{2.8}$$

et

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & si \quad x \geq 0, \\ 0 & si \quad x < 0. \end{cases} \tag{2.9}$$

Soit $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ les statistiques d'ordre associées à cet échantillon.

D'après les équations (1.7) et (2.8), la densité jointe du n-uplet de $(X_{1:n}, X_{2:n}, \ldots, X_{n:n})$ est

$$f_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n) = n!\lambda \exp\left(-\lambda \sum_{i=1}^{n} x_i\right), \; 0 \leq x_1 < x_2 < \ldots < x_n < +\infty. \tag{2.10}$$

Définition 2.2.1 (Espacements exponentiels)

Les espacements exponentiels de l'échantillon $(X_1, X_2, \ldots, X_n)$ *sont définis par les statistiques*

$$\tilde{D}_i = X_{i:n} - X_{i-1:n}, \; 1 \leq i \leq n,$$

avec par convention $X_{0:n} = 0$.

Remarque 2.2.1

1. $\tilde{D}_i \geq 0, \; 1 \leq i \leq n.$
2. $\sum_{i=1}^{n} \tilde{D}_i = X_{n:n}.$

2.2.2 *Distributions des* espacements exponentiels

Théorème 2.2.1 (Pyke (1965) [10]**, Riffi** (2006) [12]**)**

La densité jointe du vecteur des espacements exponentiels $\tilde{D} = \left(\tilde{D}_1, \tilde{D}_2, \ldots, \tilde{D}_n\right)$ *est donnée par*

$$f_{\tilde{D}}(\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_n) = \begin{cases} \prod_{i=i}^{n} \lambda(n-i+1)e^{-\lambda(n-i+1)\tilde{d}_i} & si\ \tilde{d}_i > 0, 1 \leq i \leq n, \\ 0 & sinon. \end{cases} \tag{2.11}$$

Preuve. On note la transformation inverse suivante

$$\begin{pmatrix} \tilde{d}_1 = x_1 \\ \tilde{d}_2 = x_2 - x_1 \\ \vdots \\ \tilde{d}_i = x_i - x_{i-1} \\ \vdots \\ \tilde{d}_n = x_{n:n} - x_{n-1} \end{pmatrix} \overset{\Psi^{-1}}{\longrightarrow} \begin{pmatrix} x_{1:n} = \tilde{d}_1 \\ x_2 = \tilde{d}_1 + \tilde{d}_2 \\ \vdots \\ x_i = \tilde{d}_1 + \tilde{d}_2 + \ldots + \tilde{d}_i \\ \vdots \\ x_n = \tilde{d}_1 + \tilde{d}_2 \ldots + \tilde{d}_n \end{pmatrix}.$$

Le jacobien de cette transformation inverse vaut 1. Alors, la densité jointe du vecteur $\tilde{D}$ est

$$\begin{aligned} f_{\tilde{D}}(\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_n) &= f_{1,2,\ldots,n}(\Psi^{-1}(\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_n)) \left| J_{\Psi^{-1}}(\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_n) \right| \\ &= n!\lambda e^{-\lambda \sum_{i=1}^{n} (\tilde{d}_1 + \tilde{d}_2 + \ldots + \tilde{d}_i)} \\ &= n!\lambda e^{-\lambda \sum_{i=1}^{n} (n-i+1)\tilde{d}_i} \\ &= \begin{cases} \prod_{i=i}^{n} \lambda(n-i+1)e^{-\lambda(n-i+1)\tilde{d}_i} & si\ \tilde{d}_i > 0,\ 1 \leq i \leq n, \\ 0 & \text{sinon.} \end{cases} \end{aligned}$$

■

Théorème 2.2.2 (Riffi (2006) [**12**]**)**

Soient $X_{1:n} < X_{2:n} < \ldots < X_{n:n}$ *les statistiques d'ordre associées à des v.a.'s i.i.d. de loi exponentielle de paramètre* $\lambda > 0$*, pour* $1 \leq i < j \leq n$ *an a* $X_{j:n} - X_{i:n} \stackrel{d}{=} X_{j-i:n-i}$ *et la densité de* $X_{j:n} - X_{i:n}$ *est donnée par*

$$f_{X_{j:n}-X_{i:n}}(x) = \frac{(n-i)!}{(j-i-1)!(n-j)!}\lambda e^{-\lambda x(n-j+1)}(1-e^{-\lambda x})^{j-i-1}, \ x > 0.$$

Preuve. *Voir Riffi* (2006) *[12], le théorème* 3.1 *et le théorème* 3.2*, page* $4-5$*.* ■

Corollaire 2.2.1 *Pour* $1 \leq i < k+i \leq n$ *et* $1 \leq k < n$*, on trouve*

$$X_{k+i:n} - X_{i:n} \stackrel{d}{=} X_{k:n-i}$$

et la densité de $X_{k+i:n} - X_{i:n}$ *est donnée par*

$$f_{X_{k+i:n}-X_{i:n}}(x) = \frac{(n-i)!}{(k-1)!(n-k-i)!}\lambda e^{-\lambda x(n-k-i+1)}(1-e^{-\lambda x})^{k-1}, \ x > 0.$$

Théorème 2.2.3 *Les espacements exponentiels* $\tilde{D}_i$*,* $1 \leq i \leq n$*, sont des variables aléatoires i.i.d. de loi exponentielle de paramètre* $\tilde{\lambda} = \lambda(n-i+1)$ *respectivement. La densité de* $\tilde{D}_i$ *est donnée par*

$$f_{\tilde{D}_i}(\tilde{d}) = \lambda(n-i+1)e^{-\lambda(n-i+1)\tilde{d}}, \ \tilde{d} > 0, \ 1 \leq i \leq n, \tag{2.12}$$

et la fonction de répartition est donnée par

$$F_{\tilde{D}i}(\tilde{d}) = 1 - e^{-\lambda(n-i+1)\tilde{d}}, \ \tilde{d} > 0, \ 1 \leq i \leq n. \tag{2.13}$$

Preuve. Pour $X_{0:n} = 0, \ 1 \leq i \leq n$, on a $\tilde{D}_i = X_{i:n} - X_{i-1:n}$, en utilisant le corollaire (2.2.1), on trouve $\tilde{D}_i \stackrel{d}{=} X_{1:n-i+1}$ et la densité de $\tilde{D}_i$ est

$$\begin{aligned} f_{\tilde{D}_i}(\tilde{d}) &= \frac{(n-i+1)!}{(n-i)!}\lambda e^{-\lambda\tilde{d}(n-i+1)} \\ &= \lambda(n-i+1)e^{-\lambda(n-i+1)\tilde{d}}, \ \tilde{d} > 0, \ 1 \leq i \leq n, \end{aligned}$$

et la fonction de répartition de $\tilde{D}_i$ est

$$F_{\tilde{D}_i}(\tilde{d}) = \int_0^{\tilde{d}} \lambda(n-i+1)e^{-\lambda(n-i+1)x}dx$$
$$= 1 - e^{-\lambda(n-i+1)\tilde{d}}, \ \tilde{d} > 0, \ 1 \leq i \leq n.$$

■

Remarque 2.2.2 *D'après les équations (2.11) et (2.12), on remarque que*

$$f_{\tilde{D}}(d_1, d_2, \ldots, d_n) = \prod_{i=1}^{n} f_{\tilde{D}_i}(d),$$

alors, les espacements exponentiels $\tilde{D}_1, \tilde{D}_2, \ldots, \tilde{D}_n$ sont indépendants et non identiquement distribués.

Lemme 2.2.1 *Pour $1 \leq i_1 < i_2 < \ldots < i_k \leq n, 1 \leq k \leq n$, la densité jointe du vecteur $\left(\tilde{D}_{i_1}, \tilde{D}_{i_2}, \ldots, \tilde{D}_{i_k}\right)$ est*

$$f_{\left(\tilde{D}_{i_1}, \tilde{D}_{i_2}, \ldots, \tilde{D}_{i_k}\right)}(\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_k) = \prod_{r=1}^{k} f_{\tilde{D}_{i_r}}(\tilde{d}_r) = \prod_{r=1}^{k} \lambda(n-i_r+1)e^{-\lambda(n-i_r+1)\tilde{d}_r}.$$

Corollaire 2.2.2 *Pour$1 \leq i < j \leq n$, la densité jointe du couple $(\tilde{D}_i, \tilde{D}_j)$ est*

$$f_{\left(\tilde{D}_i, \tilde{D}_j\right)}(\tilde{d}_1, \tilde{d}_2) = \lambda^2(n-i+1)(n-j+1)e^{-\lambda\left[(n-i+1)\tilde{d}_1 + (n-j+1)\tilde{d}_2\right]}.$$

2.2.3 Propriétés des espacements exponentiels

Soient $\tilde{D}_1, \tilde{D}_2, \ldots, \tilde{D}_n$ les espacements exponentiels associés à des variables aléatoires i.i.d. de loi exponentielle de paramètre $\lambda > 0$. Pour $i = 1, 2, \ldots n$ et $j = 1, 2, \ldots, n$ avec $i \neq j$, on a

- L'espérance : $E(\tilde{D}_i) = \dfrac{1}{\lambda(n-i+1)}$.
- Le moment d'ordre 2 : $E(\tilde{D}_i^2) = \dfrac{2}{\lambda^2(n-i+1)^2}$.

- La variance : $Var(\tilde{D}_i) = \dfrac{1}{\lambda^2(n-i+1)^2}$.
- La covariance : $Cov(\tilde{D}_i, \tilde{D}_j) = 0$.

Preuve. Par intégration par parties, on trouve

$$E(\tilde{D}_i) = \int_0^{+\infty} x f_{\tilde{D}_i}(x)dx = \int_0^{+\infty} x\lambda(n-i+1)e^{-\lambda(n-i+1)x}dx = \int_0^{+\infty} e^{-\lambda(n-i+1)x}dx = \frac{1}{\lambda(n+i-1)}.$$

$$\begin{aligned}
E(\tilde{D}_i^2) &= \int_0^{+\infty} x^2 f_{\tilde{D}_i}(x)dx \\
&= \int_0^{+\infty} x^2\lambda(n-i+1)e^{-\lambda(n-i+1)x}dx \\
&= \int_0^{+\infty} 2xe^{-\lambda(n-i+1)x}dx \\
&= \int_0^{+\infty} 2\frac{e^{-\lambda(n-i+1)x}}{\lambda(n-i+1)}dx \\
&= \frac{2}{\lambda^2(n-i+1)^2}.
\end{aligned}$$

$$Var(\tilde{D}_i) = E(\tilde{D}_i^2) - E(\tilde{D}_i)^2 = \frac{2}{\lambda^2(n-i+1)^2} - \frac{1}{\lambda^2(n+i-1)^2} = \frac{1}{\lambda^2(n-i+1)^2}.$$

$$Cov(\tilde{D}_i, \tilde{D}_j) = E(\tilde{D}_i, \tilde{D}_j) - E(\tilde{D}_i)E(\tilde{D}_j) = E(\tilde{D}_i)E(\tilde{D}_j) - E(\tilde{D}_i)E(\tilde{D}_j) = 0.$$

■

Théorème 2.2.4 (Sukhatme (1937) [**14**]**)**

Soient $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ *les statistiques d'ordre associées à l'échantillon* $(X_1, X_2, \ldots, X_n)$ *de v.a.'s i.i.d. de loi exponentielle de paramètre* $\lambda > 0$*. Pour* $X_{0:n} = 0$*, on a*

$$\hat{D}_i = (n-i+1)(X_{i:n} - X_{i-1:n}),\ 1 \leq i \leq n,$$

sont des v.a.'s i.i.d. de loi exponentielle de paramètre $\lambda > 0$*.*

De plus

$$\frac{X_1}{n}, \frac{X_1}{n}+\frac{X_2}{n-1}, \dots, \frac{X_1}{n}+\frac{X_2}{n-1}+\dots+\frac{X_n}{1},$$

ont la même distribution que $X_{1:n}, X_{2:n}, \dots, X_{n:n}$.

Preuve.

Première partie : *On a*

$$\hat{D}_i = (n-i+1)(X_{i:n} - X_{i-1:n}) = (n-i+1)\tilde{D}_i.$$

D'apres le Théorème (2.2.2) et l'équation (2.13) on trouve la fonction de répartition de $\hat{D}_i$

$$\begin{aligned} F_{\hat{D}_i}(\tilde{d}) &= \mathbb{P}(\hat{D}_i \leq \tilde{d}) = \mathbb{P}((n-i+1)\tilde{D}_i \leq \tilde{d}) = \mathbb{P}\left(\tilde{D}_i \leq \frac{\tilde{d}}{n-i+1}\right) \\ &= F_{\tilde{D}_i}\left(\frac{\tilde{d}}{n-i+1}\right) \\ &= 1 - e^{-\lambda \tilde{d}},\ \tilde{d} > 0,\ 1 \leq i \leq n, \end{aligned}$$

et la densité de $\hat{D}_i$

$$f_{\hat{D}_i}(\tilde{d}) = \frac{dF_{\hat{D}_i}(\tilde{d})}{d(\tilde{d})} = \lambda e^{-\lambda \tilde{d}},\ \tilde{d} > 0,\ 1 \leq i \leq n.$$

Deuxième partie : *On note la transformation inverse suivante*

$$\begin{pmatrix} y_1 = \dfrac{x_1}{n} \\ y_2 = \dfrac{x_1}{n} + \dfrac{x_2}{n-1} \\ \vdots \\ y_i = \dfrac{x_1}{n} + \dfrac{x_2}{n-1} + \dots + \dfrac{x_i}{n-i+1} \\ \vdots \\ y_n = \dfrac{x_1}{n} + \dfrac{x_2}{n-1} + \dots + \dfrac{x_n}{1} \end{pmatrix} \overset{\Psi^{-1}}{\longrightarrow} \begin{pmatrix} x_1 = ny_1 \\ x_2 = (y_2 - y_1)(n-1) \\ \vdots \\ x_i = (y_i - y_{i-1})(n-i+1) \\ \vdots \\ x_n = (y_n - y_{n-1}) \end{pmatrix}.$$

Le jacobien de cette transformation inverse est

$$\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \frac{\partial x_1}{\partial y_3} & \dots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \frac{\partial x_2}{\partial y_3} & \dots & \frac{\partial x_2}{\partial y_n} \\ \frac{\partial x_3}{\partial y_1} & \frac{\partial x_3}{\partial y_2} & \frac{\partial x_3}{\partial y_3} & & \frac{\partial x_3}{\partial y_n} \\ \vdots & & & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \frac{\partial x_n}{\partial y_3} & \dots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} = \begin{vmatrix} n & 0 & \dots & \dots & \dots & 0 \\ -(n-1) & (n-1) & 0 & \dots & \dots & 0 \\ 0 & -(n-2) & (n-2) & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & -1 & 1 \end{vmatrix} = n!.$$

En utilisant l'équation (2.8), la densité jointe du n-uplet de $(X_1, X_2, \dots, X_n)$ *est*

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda e^{-\lambda \sum_{i=1}^{n} x_i}, \ x \geq 0.$$

Alors, la densité jointe du vecteur $(Y_1, Y_2, \dots, Y_n)$ *est*

$$\begin{aligned} f_{(Y_1,Y_2,\dots,Y_n)}(y_1, y_2, \dots, y_n) &= f(\Psi^{-1}(y_1, y_2, \dots, y_n)) \left| J_{\Psi^{-1}}(y_1, y_2, \dots, y_n) \right| \\ &= n! \lambda e^{-\lambda \sum_{i=1}^{n} (y_i - y_{i-1})(n-i+1)} \\ &= n! \lambda e^{-\lambda \sum_{i=1}^{n} y_i}, \ 0 \leq y_1 < y_2 < \dots < y_n < +\infty. \end{aligned}$$

À partir l'équation (2.10), on remarque que

$$f_{(Y_1,Y_2,\dots,Y_n)}(y_1, y_2, \dots, y_n) = f_{1,2,\dots,n}(x_1, x_2, \dots, x_n).$$

■

Corollaire 2.2.3 (Malmquist (1950) [9]**)**

Soient $0 \leq U_{1:n} \leq U_{2:n} \leq \dots \leq U_{n:n} \leq 1$ *les statistiques d'ordre associées à* $U_1, U_2, \dots, U_n$ *; une suite de v.a.'s i.i.d. uniformément distribuées sur* $[0, 1]$.

Pour $U_{n+1:n} = 1$, *on a*

$$\left\{ \left(\frac{U_{i:n}}{U_{i+1:n}} \right)^i, \ 1 \leq i \leq n \right\} \stackrel{d}{=} U_1, U_2, \ldots, U_n.$$

Preuve. Premièrement, nous devons savoir que

$$U_i \stackrel{d}{=} 1 - e^{-E_i}, \ 1 \leq i \leq n,$$

où $(E_1, E_2, \ldots, E_n)$ un échantillon de v.a.'s i.i.d. de loi exponentielle standard.

De plus, d'aprés la propriété (4), on a

$$U_{i:n} \stackrel{d}{=} 1 - e^{-E_{i:n}}, \ 1 \leq i \leq n,$$

où $E_{1:n}, E_{2:n}, \ldots, E_{n:n}$ les statistiques d'ordre associées à l'échantillon $(E_1, E_2, \ldots, E_n)$.

On sait que

$$1 - U_{n-i+1:n} \stackrel{d}{=} U_{i:n}, \ 1 \leq i \leq n.$$

Alors,

$$\left(\frac{U_{i:n}}{U_{i+1:n}} \right)^i \stackrel{d}{=} \left(\frac{1 - U_{n-i+1:n}}{1 - U_{n-i:n}} \right)^i \stackrel{d}{=} \left(\frac{e^{-E_{n-i+1:n}}}{e^{-E_{n-i:n}}} \right)^i \stackrel{d}{=} e^{i(-E_{n-i+1:n} + E_{n-i:n})}.$$

Donc,

$$\log \left(\frac{U_{i:n}}{U_{i+1:n}} \right)^i \stackrel{d}{=} i \left(-E_{n-i+1:n} + E_{n-i:n} \right), \ 1 \leq i \leq n.$$

D'après le théorème (2.2.4), on a $i\left(-E_{n-i+1:n} + E_{n-i:n}\right)$ sont des v.a.'s i.i.d. de loi exponenielle, implique que $\log \left(\frac{U_{i:n}}{U_{i+1:n}} \right)^i$ $(1 \leq i \leq n)$, sont des v.a.'s i.i.d. de loi exponenielle. On peut conclure que $\left(\frac{U_{i:n}}{U_{i+1:n}} \right)^i$ sont des v.a.'s i.i.d. uniformément distribuées sur $[0, 1]$. ■

2.3 Espacements uniformes en fonction des espacements exponentiels

Soit $(X_1, X_2, \ldots, X_n)$ un échantillon de v.a.'s i.i.d. de loi exponentielle standard telle que

$$F(x) = 1 - e^{-x}, \; x \geq 0.$$

Pour $i = 1, 2, \ldots, n$, on pose

$$Y_i = \sum_{j=1}^{i} X_j (n - j + 1)^{-1}.$$

D'après le théorème (2.2.4), les v.a.'s $Y_1, Y_2, \ldots, Y_n$ ont la même distribution que les statistiques d'ordre associées à l'échantillon $(X_1, X_2, \ldots, X_n)$.

De plus, on peut observer que les v.a.'s $X_i(n - i + 1)^{-1}$ ont la même distribution des espacements exponentiels.

En effet, pour $X_{0:n} = 0$

$$\begin{aligned}
\tilde{D}_i &= X_{i:n} - X_{i-1} \\
&\stackrel{d}{=} Y_i - Y_{i-1} \\
&\stackrel{d}{=} \frac{X_1}{n} + \frac{X_2}{n-1} + \ldots + \frac{X_i}{n-i+1} - \left(\frac{X_1}{n} + \frac{X_2}{n-1} + \ldots + \frac{X_{i-1}}{n-i+2} \right) \\
&\stackrel{d}{=} \frac{X_i}{n-i+1}.
\end{aligned}$$

Soient $U_{1:n}, U_{2:n}, \ldots, U_{n:n}$ les statistiques d'ordre associées à $(U_1, U_2, \ldots, U_n)$; un échantillon de v.a.'s i.i.d. uniformément distribuées sur $[0, 1]$. D'aprés la propriété (4), on a

$$U_{i:n} \stackrel{d}{=} F(X_{i:n}) \stackrel{d}{=} F(Y_i) = 1 - e^{-Y_i}, \; 1 \leq i \leq n.$$

Pour $U_{0:n} = 0$, les espacements uniformes peuvent s'écrire en fonction des espacements exponentiels, comme suit

$$\begin{aligned}
D_i &= U_{i:n} - U_{i-1:n} \\
&\overset{d}{=} 1 - e^{-Y_i} - \left(1 - e^{-Y_{i-1}}\right) \\
&\overset{d}{=} e^{-Y_{i-1}} - e^{-Y_i} \\
&\overset{d}{=} e^{-Y_{i-1}} \left(1 - e^{-Y_i} e^{Y_{i-1}}\right) \\
&\overset{d}{=} (1 - U_{i-1:n}) \left(1 - e^{-\sum\limits_{j=1}^{i} X_j (n-j+1)^{-1} + \sum\limits_{j=1}^{i-1} X_j (n-j+1)^{-1}}\right) \\
&\overset{d}{=} (1 - U_{i-1:n}) \left(1 - e^{-X_i (n-i+1)^{-1}}\right),
\end{aligned}$$

et pour $U_{n+1:n} = 1$ on a

$$D_{n+1} = 1 - U_{n:n} \overset{d}{=} 1 - \left(1 - e^{-Y_n}\right) \overset{d}{=} e^{-Y_n} = e^{-\sum\limits_{j=1}^{n} X_j (n-j+1)^{-1}}.$$

Chapitre 3

Simulation sous R

Dans ce chapitre, on va illustrer certains des résultats importants, que l'on a obtenus dans les deux premiers chapitres, sur des données simulées à l'aide du logiciel R. Tout d'abord, on s'intéresse aux différents algorithmes pour générer un échantillon de statistiques d'ordre, à savoir la simulation par : la méthode directe la représentation de Réney, la méthode d'inversion et la fonction de répartition empirique. Par la suite, on va simuler des espacements uniformes et exponentiels.

3.1 Simulation des statistiques d'ordre

3.1.1 Méthode directe

Après avoir générer un échantillon de variables aléatoire `X`, on peut obtenir les statistiques d'ordre lui associées directement, en utilisant la commande `sort(X)` du logiciel R. En effet, pour générer un échantillon de variables aléatoires, on utilise la command `"rfunc"`, où le suffix `"func"` se change selon la loi de probabilité. De plus, on peut remplacer la lettre `r` par la lettre `d`, `p` ou `q` pour obtenir la densité, la fonction de répartition, ou le quantile (respectivement). Le tableau (3.1) résume le code de quelques lois usuelles.

Loi	**Commande**
Gauss(normale) $\mathcal{N}(\mu, \sigma^2)$	`rnorm(n,mean,sd)`
Exponentielle $\exp(\lambda)$	`rexp(n,rate)`
Gamma $\Gamma(a, \lambda)$	`rgamma(n,shape,scale)`
Poisson $\mathcal{P}(\lambda)$	`rpois(n,lambda)`
Beta $B(\alpha, \beta)$	`rbeta(n,shape1,shape2)`
Binomiale $Bin(n, p)$	`rbinom(n,size,prob)`
Géométrique $\mathcal{G}(p)$	`rgeom(n,prob)`
Uniforme $\mathcal{U}[a, b]$	`runif(n,min,max)`

TAB. 3.1 – Lois de probabilités usuelles sous logiciel R

Exemple 3.1.1 *Soit* $X = (X_1, X_2, \ldots, X_n)$ *un échantillon de v.a.'s i.i.d de loi exponentielle standard* $(\lambda = 1)$*, de taille* $n = 100$*. Pour générer les statistiques d'ordre associées à cet échantillon, on utilise le code suivant*

```
n=100
X=numeric(n)
for(i in 1 :n){X[i]=rexp(1)}
XO=sort(X)
```

3.1.2 Représentation de Réney

Soit $(E_1, E_2, \ldots, E_{n+1})$ un échantillon de v.a.'s de loi exponentielle standard. Soient $U_{1:n}, U_{2:n}, \ldots, U_{n:n}$ les statistiques d'ordre associées aux v.a.'s indépendantes uniformément distribuées sur $[0, 1]$. Pour $S_i = \sum_{j=1}^{i} E_j,\ 1 \leq i \leq n+1$ on a

$$(U_{1:n}, U_{2:n}, \ldots, U_{n:n}) \stackrel{d}{=} \left(\frac{S_1}{S_{n+1}}, \frac{S_2}{S_{n+1}}, \ldots, \frac{S_n}{S_{n+1}} \right).$$

Le code suivant permet de générer des statistiques d'ordre associées à n v.a.'s de loi uniforme standard selon la représentation de Réney et de comparer sa distribution avec la distribution des statistiques d'ordre obtenues par la méthode directe.

```
n=250
U=numeric(n)
Y=numeric(n)
S=numeric(n+1)
E=rexp(n+1)
for(i in 1 :n){U[i]=runif(1)}
UO=sort(U)
S[n+1]=sum(E)
for (i in 1 :n) {S[i]=sum(E[1 :i])
Y[i]=S[i]/S[n+1]}
qqplot(UO,Y)
ks.test(UO,Y)
```

La figure (3.1) représente les quantiles de l'échantillon obtenu par la représentation de Réney $Y = \left(\frac{S_1}{S_{n+1}}, \frac{S_2}{S_{n+1}}, \ldots, \frac{S_n}{S_{n+1}}\right)$, en fonction des quantiles de l'échantillon des statistiques d'ordre obtenu par la méthode directe $UO = (U_{1:n}, U_{2:n}, \ldots, U_{n:n})$. On remarque que ces deux quantiles sont alignés, ce qui prouve graphiquement l'égalité en distribution des deux échantillons

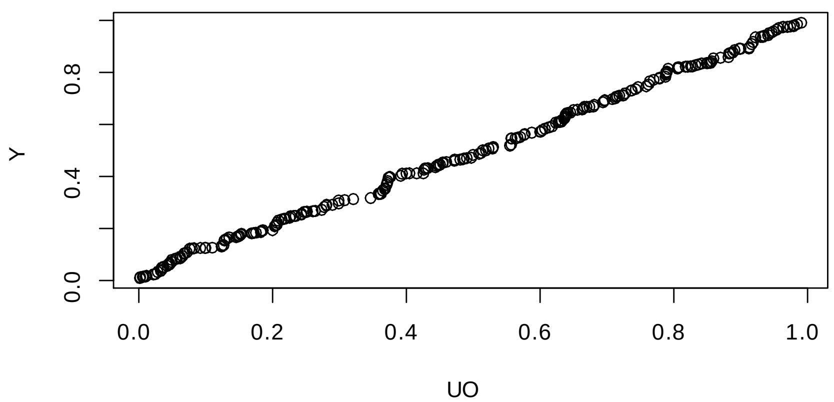

FIG. 3.1 – Comparaison des distributions de statistiques d'ordre obtenues par la représentation de Réney et la méthode directe

Au seuil de signification 5%, on peut confirmer ce résultat à l'aide du test non-paramétrique de Kolmogorov-Smirnov. La sortie du commande `ks.test` nous donne la "$p - value = 0.341$" qui est supérieure à 0.05, ce qui nous conduit à accepter l'hypothèse d'égalité des distributions de ces deux échantillons.

3.1.3 Méthode d'inversion

Soit $X = (X_1, X_2, \ldots, X_n)$ un échantillon de v.a.'s i.i.d. d'une fonction de répartition commune

$$F(x) = \mathbb{P}(X \leq x),\ x \in \mathbb{R}.$$

Soient $U_{1:n}, U_{2:n}, \ldots, U_{n:n}$ les statistiques d'ordre associées à un échantillon $(U_1, U_2, \ldots, U_n)$ de v.a.'s indépendantes issue d'une loi uniforme standard. On a

$$\{F^{\leftarrow}(U_{i:n}), 1 \leq i \leq n, n \geq 1\} \stackrel{d}{=} \{X_{i:n}, 1 \leq i \leq n, n \geq 1\}, \tag{3.1}$$

où $(X_{i:n}, 1 \leq i \leq n, n \geq 1)$ les statistiques d'ordre relatives à l'échantillon $(X_1, X_2, \ldots, X_n)$.

Pour générer un échantillon ordonné de de v.a.'s de fonction de répartition F, il suffit de générer un échantillon ordonné de v.a.'s uniforme sur $[0, 1]$ et de lui appliquer la transformation des quantiles.

Exemple 3.1.2 *Soit $\lambda > 0$, si $X \sim \exp(\lambda)$, la fonction de répartition est donnée par*

$$F(x) = 1 - \exp(-\lambda x).$$

Ce qui implique que

$$F^{-1}(u) \stackrel{d}{=} -\frac{1}{\lambda} \log(1 - u),\ 0 < u < 1$$

Par conséquent, on obtient

$$\left\{F^{-1}(U_{i:n}), 1 \leq i \leq n, n \geq 1\right\} \stackrel{d}{=} \left\{-\frac{1}{\lambda} \log(1 - U_{i:n}), 1 \leq i \leq n, n \geq 1\right\}$$

Dans le code suivant, on génère un échantillon ordonné associé à un échantillon de v.a.'s de loi exponentielle de paramètre $\lambda = 2$, de taille $n = 200$ à travers la transformation des quantiles. Ensuite, on compare la distributions de cet échantillon avec celle des statistiques d'ordre obtenues par la méthode directe.

```
n=200
U=numeric(n)
X=numeric(n)
lambda=2
for(i in 1 :n){U[i]=runif(1)}
UO=sort(U)
Y=(-1/lambda)*(log(1-UO))
for(i in 1 :n){X[i]=rexp(1,lambda)}
XO=sort(X)
qqplot(Y,XO)
ks.test(Y,XO)
```

L'alignement des quantiles de l'échantillon $Y = (X_{1:n}, X_{2:n}, \ldots .X_{n:n})$ et les quantiles de l'échantillon $XO = (F^{\leftarrow}(U_{1:n}), F^{\leftarrow}(U_{2:n}), \ldots, F^{\leftarrow}(U_{n:n}))$, le long de la bissectrice dans la figure (3.2), prouve l'égalité en distribution de ces deux échantillons.

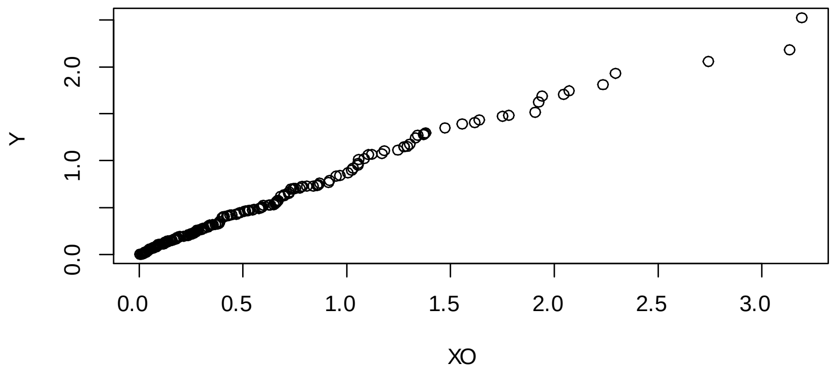

FIG. 3.2 – Comparaison des distributions de statistiques d'ordre obtenues par la méthode directe et la méthode d'inversion relatives à un échantillon de loi exponentielle

Pour confirmer ce résultat, on utilise le test de Kolmogorov-Smirnov, ici la $p - value = 0.9228$ est supérieur à 0.05 donc les deux échantillons ont la même distribution.

3.1.4 Méthode par la fonction de répartition empirique

Soit $(X_1, X_2, \ldots, X_n)$ un échantillon de v.a.'s i.i.d et $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ les statistiques d'ordre associées à cet échantillon. Dans la section (1.1.4), nous avons constatés que l'expression de la fonction de répartition empirique est

$$F_n(x) = \begin{cases} 0 & si \quad x < X_{1:n}, \\ \dfrac{i}{n} & si \quad X_{i:n} \leq x < X_{i+1:n}, \; 1 \leq i \leq n-1, \\ 1 & si \quad x \geq X_{n:n}, \end{cases}$$

par conséquent, les statistiques d'ordre $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ sont la solution de l'équation suivante

$$F_n(x) - \frac{i}{n} = 0.$$

Pour générer des statistiques d'ordre à travers cette méthode, on propose le code suivant.

```
n=400
U=numeric(n) ;X0=numeric(n)
for (i in 1 :n){U[i]=runif(1)}
U0=sort(U)
Fn=function(x,U){Y=numeric(n)
for(i in 1 :n){if (U[i]<=x) Y[i]=1 else Y[i]=0}
mean(Y)}
for(i in 1 :n){G=function(x){Fn(x,U)-(i/n)}
X=uniroot(G,lower=0,upper=1)
X0[i]=X$root}
qqplot(U0,X0)
ks.test(U0,X0)
```

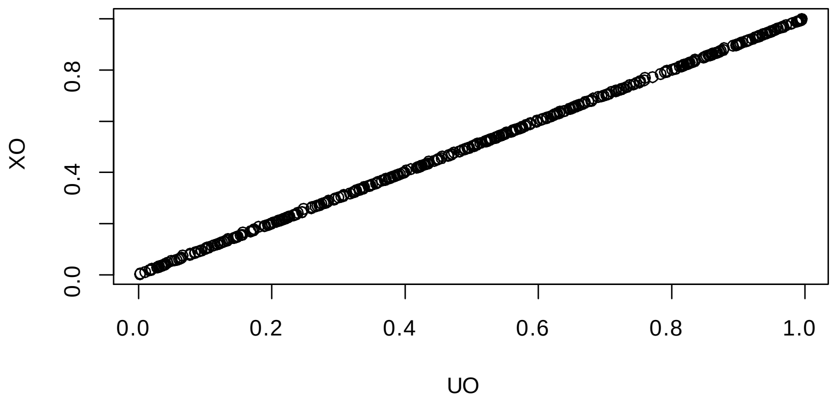

Fig. 3.3 – Comparaison des distributions de statistiques d'ordre obtenues par la méthode directe et la méthode de la fonction de réprtition empirique

La figure (3.3) présente une superposition parfaite du quantile des statistiques d'ordre XO obtenues à travers la fonction de répartition empirique et le quantile de l'échantillon ordonné UO par la méthode directe. De plus, la $p - value = 1$ du test de Kolmogorov-Smirnov est supérieur au seuil de signification 0.05, ce qui nous permet de valider le résultat théorique de la fonction de répartition empirique d'un point de vue pratique.

3.2 Simulation des espacements

3.2.1 Simulation des espacements uniformes

Soit $(U_1, U_2, \ldots, U_n)$ un échantillon de v.a.'s i.i.d. uniformément distribuées sur $[0, 1]$. Le code suivant permet de générer des espacements uniformes par deux méthodes.

Génération à travers la méthode directe

```
n=200
U=numeric(n)
D=numeric(n+1)
for(i in 1 :n){U[i]=runif(1)}
UO=sort(U)
D[1]=UO[1]
for(i in 2 :n){D[i]=UO[i]-UO[i-1]}
D[n+1]=1-UO[n]
```

Génération à travers le résultat du théorème (2.1.4)

```
E=numeric(n+1)
Y=numeric(n+1)
for(i in 1 :n+1){E[i]=rexp(1)}
S=sum(E)
for(i in 1 :n+1){Y[i]=E[i]/S}
qqplot(D,Y)
ks.test(D,Y)
```

La figure (3.4) permet de comparer la distributions de l'échantillons

$$Y = \left(\frac{E_1}{\sum\limits_{i=1}^{n+1} E_i}, \frac{E_2}{\sum\limits_{i=1}^{n+1} E_i}, \ldots, \frac{E_{n+1}}{\sum\limits_{i=1}^{n+1} E_i} \right)$$

avec la distribution de l'échantillon des espacements uniforme $D = (D_1, D_2, \ldots, D_{n+1})$. On remarque que la majorité des points sont alignés le long de la première bissectrice ce qui indique la présence d'une identité de distribution de ces deux échantillons.

De plus, la $p - value = 0.7945$ du test de Kolmogorov-Smirnov est supérieur à 0.05 donc on accepte l'hypothèse nulle, cela confirme le résulats du théorème 2.1.4.

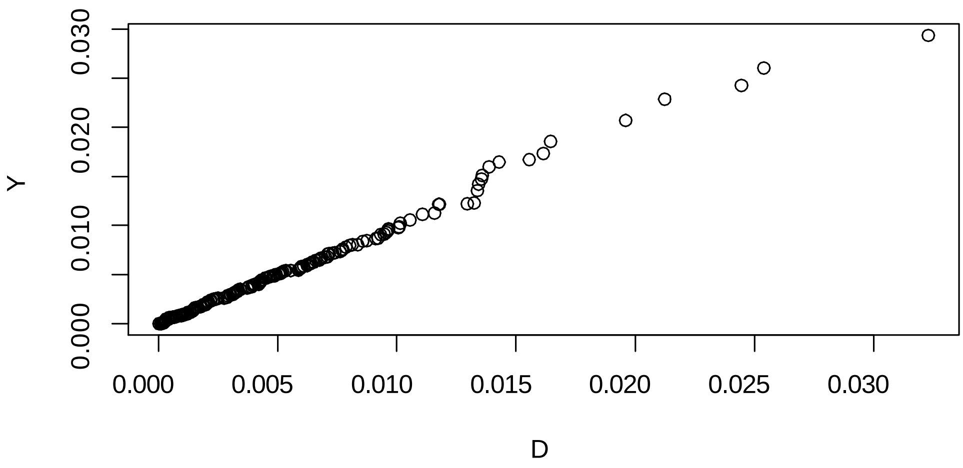

FIG. 3.4 – Comparaison des distributions des espacements uniformes

3.2.2 Simulation des espacements exponentiels

Soit $(X_1, X_2, \ldots, X_n)$ un échantillon de n variables aléatoires i.i.d. de la loi exponentielle de paramètre $\lambda > 0$.

Génération à travers la méthode directe

Le code suivant permet de générer un échantillon des espacements exponentiels des statistiques d'ordre relatives à un échantillon de v.a.'s iid de loi exponentielle de paramètre $\lambda = 2$ et de taille $n = 500$.

```
n=500
lambda=2
X=numeric(n)
D=numeric(n)
for(i in 1 :n){X[i]=rexp(1,lambda)}
XO=sort(X)
D[1]=XO[1]
for(i in 2 :n){D[i]=XO[i]-XO[i-1]}
```

3.2.3 Génération à travers le résultst du théorème de Sukhatme (1937)

D'après le théorème (2.2.4), on peut observer que

$$X_1 n^{-1}, X_2(n-1)^{-1}, \dots, X_n$$

et les espacements exponentiels $\tilde{D}_1, \tilde{D}_2, \dots, \tilde{D}_n$ ont la même distribution.

Pour générer des espacements exponentiels à travers ce résultat, on utilise le code suivant.

```
n=500
lambda=2
X=numeric(n)
D=numeric(n)
E=numeric(n)
for(i in 1 :n){X[i]=rexp(1,lambda)
E[i]=X[i]/(n-i+1)}
XO=sort(X)
D[1]=XO[1]
for(i in 2 :n){D[i]=XO[i]-XO[i-1]}
qqplot(D,E)
ks.test(D,E)
```

La figure (3.5) représente les quantiles de $E = (X_1 n^{-1}, X_2(n-1)^{-1}, \dots, X_n)$ en fonction des quantiles de $D = (\tilde{D}_1, \tilde{D}_2, \dots, \tilde{D}_n)$. On peut observer clairement que la majorité des points sont alignés donc on peut dire que les distributions des deux échantillons est identiques. De plus, vue la $p-value$ qui est égale à 0.3291, le test de Kolmogorov-Smirnov accepte l'hypothèse nulle d'égalité des distributions à un seuil de signification 5%.

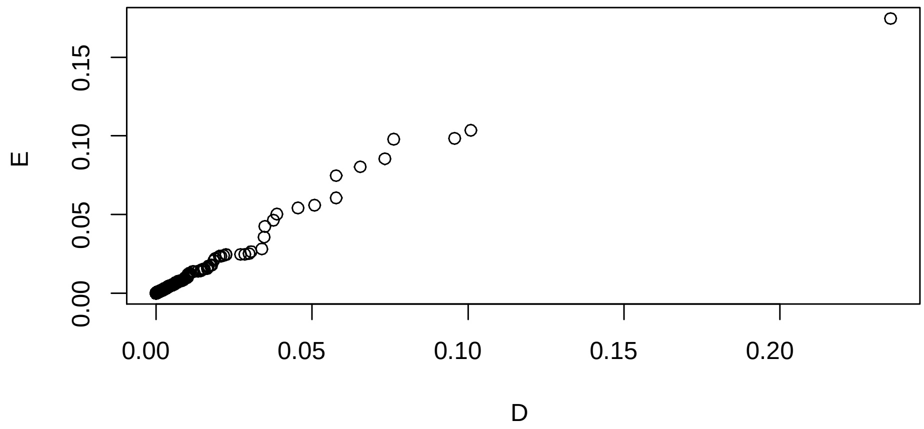

Fig. 3.5 – Comparaison des distributions des espacements exponentiels à travers le théorème de Sukhatme

Conclusion

La réalisation de ce mémoire m'a été très bénéfique, il m'a permit d'enrichir et d'approfondir mes connaissances sur la théorie des statistiques d'ordre et leurs espacements qui est d'une importance majeur en statistique, ça ma permis de connaitre les propriétés des variables d'un échantillons ordrdonné et comment obtenir leurs distributions.

On conclusion, on peut dire que la distribution d'un échantillon de variables aléatoire désordonnées et la distribution des statistiques d'ordre relatives sont très différentes. Aussi, les espacements des statistiques d'ordre ne sont pas moins importants que les statistiques d'ordre, en effet leurs distributions et ses propriétés changent d'un espacement à un autre, à titre d'exemple les espacements uniformes sont identiquement distribués et les espacements exponentiels sont indépendants et non identiquement distribués.

Bibliographie

[1] Ahsanullah, M., Nevzorov, V. B., & Shakil, M. (2013). An Introduction to Order Statistics. Atlantis Press.

[2] Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1992). A First Course in Order Statistics. Wiley, New York.

[3] Casella, G., & Berger, R. L. (2002). Statistical Inference. Thomson Learning.

[4] David, H. A. (1970). Order Statistics, (1st ed.), Wiley, New York.

[5] David, H. A. (1981). Order Statistics, (2nd ed.), Wiley, New York.

[6] David, H. A., & Nagaraja, H. N. (2003). Order Statistics, Third Edition. Wiley.

[7] Devroye, L. (1986). Non-Uniform Random Variate Generation. Springer-Verlag, New York.

[8] Gardes, L. (2020). Théorie des Valeurs Extrêmes. Université de Strasbourg.

[9] Malmquist, S. (1950). On a proprety of order statistics from a rectangular distribution. Skandinavish Aktuarietidkrift 33, 214-222.

[10] Pyke, R. (1965). Spacings. Journal of the Royal Statistical Society : Series B (Methodological), 27(3), 395-436.

[11] Rényi, A. (1953). On the Theory of Order Statistics. Acta Mathematica Hungarica, 4(3–4) : 191–231. 18, 41, 78.

[12] Riffi, M. I. (2006). On The Spacings of Exponential and Uniform Order Statistics. Al-Aqsa University J.

[13] Shahbaz, M. Q., Ahsanullah, M., Shahbaz, S. H., & Al-Zahrani, B. M. (2016). Ordered Random Variables : Theory and Applications. Atlantis Press.

[14] Sukhatme, P.V. (1937). Tests of significance for sample of the chi square population with two degrees of freedom. Ann, Eugen 8, 52-56.

Annexe A : Tests de comparaison de deux distributions

Test graphique par le diagramme quantile-quantile

Le diagramme quantile-quantile (ou le qqplot) est un outil graphique permettant d'évaluer la pertinence de l'ajustement d'une distribution donnée à un modèle théorique. Le terme de quantile-quantile provient du fait que l'on compare la position de certains quantiles dans la population observée avec leur position dans la population théorique.

Le diagramme quantile-quantile permet également de comparer deux distributions que l'on estime identique. Sur ce graphe, l'axe des ordonnées porte les quantiles x_i^* de la distribution observée, tandis que l'axe des abscisses porte les quantiles x_i correspondants de la loi théorique. Le nuage des points (x_i, x_i^*) s'aligne sur la première bissectrice lorsque la distribution théorique proposée est une bonne représentation des observations.

Test de Kolmogorov-Smirnov

Soient $(X_1, X_2, \ldots, X_n)$ et $(Y_1, Y_2, \ldots, Y_m)$ deux échantillons de v.a.'s identiquement distribuées de taille n et m respectivement. Le principe du test de Kolmogorov-Smirnov est de comparer deux distributions au moyen de la fonction de répartition.

On note par F_X la fonction de répartition de l'échantillon $(X_1, X_2, \ldots, X_n)$ et par F_Y la fonction de répartition de l'échantillon $(Y_1, Y_2, \ldots, Y_m)$.

On veut tester les hypothèses suivantes

$$\begin{cases} H_0 : F_X = F_Y \\ H_1 : F_X \neq F_Y \end{cases}$$

Si la "$p - value$" est inférieure à un seuil de signification α qu'on s'est donné (en général, 5%), alors on rejete H_0.

La fonction de test de Kolmogorov-Smirnov en R logiciel est "ks.test", c'est dans le paquet "dgof" et "stats".

Annexe B : Abréviations et Notations

v.a.	:	variable aléatoire.
i.i.d.	:	indépendantes et identiquement distribuées.
$E(.)$	:	espérance.
$u^{(m)} = E(X^m)$	:	moment d'ordre m.
$Var(.)$	:	variance.
$Cov(.,.)$	:	covariance.
$e^x = \exp(x)$	:	fonction exponentielle.
$\mathbb{R}$	:	ensemble des nombres réelles.
$\mathbb{N}$	:	ensemble des nombres naturelles.
n	:	nombre appartenir l'ensemble $\mathbb{N}$.
$\stackrel{d}{=}$	:	égalité en distribution.
$\mathbb{I}_A$	:	fonction indicatrice de A.
$C_n^r = \dfrac{n!}{r!(n-r)!}$	:	nombre de combinaisons de r éléments parmi n éléments (sansremise).
$f(x)$	:	fonction de masse de probabilité ou fonction de densité de probabilité.

$F(x) = P(X \leq x)$	:	fonction de répartition
$F(x-)$	:	$\mathbb{P}(X < x)$
$F_n(x) = \dfrac{1}{n}\sum\limits_{i=1}^{n} \mathbb{I}_{\{X_i \leq x\}}$	:	fonction de répartition empirique.
$F^{\leftarrow}(x) = \inf\{x : F(x) \geq \alpha\},\ 0 < \alpha < 1$	:	fonction inverse généralisée de F
$I_p(a,b) = \dfrac{\int_0^p t^{a-1}(1-t)^{b-1}dt}{B(a,b)}$	:	fonction bêta incomplète.
$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$	:	fonction bêta.
$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt$	:	fonction gamma.

Résumé

L'objectif de ce travail est de présenter les différentes propriétés et distributions des statistiques d'ordre et leurs espacements, qu'ils ont un large éventail d'application dans les sciences statistiques. Tout d'abord, nous présentons quelques définitions des statistiques d'ordre, puis nous passons graduellement à ses distributions dans le cas continu et discret. S'ensuit la théorie des espacements entre les statistiques d'ordre successives, dont nous illustrons leurs propriétés et distributions. Nous terminons ce travail, par des simulations des échantillons ordonnés et leurs espacements à l'aide du logiciel R.

Mots-clés : Statistique d'ordre, Distribution, Espacements uniforme, Espacements exponentiels.

ملخص

الهدف من هذا العمل هو تقديم الخصائص والتوزيعات المختلفة لإحصاءات المرتبة وتباعدها ، والتي لها نطاق واسع من التطبيقات في علوم الاحصاء. أولا نقدم بعض تعريفات احصاءات المرتبة ثم ننتقل تدريجياً إلى توزيعاتها في الحالة المستمرة والمنفصلة. يلي ذلك الجزء النظري للتباعد بين إحصاءات المرتبة المتتالية، والذي نوضح خصائصه وتوزيعاته. ننهي هذا العمل، من خلال محاكاة العينات المرتبة وتباعدها باستخدام برنامج R.

الكلمات المفتاحية : الإحصاءات المرتبة، التوزيع، التباعد الموحد ، التباعد الأسي .

Abstract

The aim of this work is to present the different properties and distributions of order statistics and their spacings, which have a wide range of application in the statistics sciences. First, we introduce some definitions of order statistics, and then we move gradually to its distributions in the continuous and discrete case. This is followed by the theory of spacings between successive order statistics, we illustrate their properties and distributions. We finish this work, by some simulations of the ordered samples and their spacings using the R software.

Keywords : Order statistics, Distribution, Uniform spacings, Exponential spacings.