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

Samira Gama

Titre :

Statistiques d'ordre

Membres du Comité d'Examen :

Pr.	Djamel Meraghni	UMKB	Président
Pr.	Abdelhakim Necir	UMKB	Encadreur
Dr.	Imane Benelmir	UMKB	Examinateur

Juin 2020

Dédicace

Ce modeste travail est dédié :

A mes chers parents.

A mes fréres et sœurs.

A ma famille.

A mes amis et collégues et tous ceux qui m'ont aidé.

REMERCIEMENTS

Tout d'abord, je remercie Dieu tout puissant de nous avoir donné la force et la capacité d'effectuer ce travail.

A mes parents qui nous ont enseigné l'amour du travail, pour leurs sacrifices, leurs affections et soutien moral.

Ma sincère reconnaissance s'adresse à **Pr Abdelhakim Necir** pour avoir accepté d'être le directeur de ce mémoire, sa disponibilité, son aide, ainsi que pour ses précieux conseils, et de me avoir accueilli dans son service à bras ouverts et d'avoir naitre en me l'amour de cette spécialité. Sa rigueur, son amour du travail et ses larges compétences pratiques ne sont plus à prouver.

Je tiens aussi à remercier également tous les membres de jury, pour avoir accepter d'évaluer notre travail.

Mes remerciements les plus sincères vont également aux enseignants qui ont participé à mon formation.

A mes chères amies qui ont toujours été présents et fidèles.

Enfin, pour toute personne qui a contribué, de prés ou de loin, à l'élaboration de ce mémoire.

Veuillez bien trouver ici l'expression de mes sincères remerciements.

Table des matières

Introduction générale

En raison de l'importance des statistiques d'ordre dans de nombreuses théories et son rôle efficace dans de beaucoup des applications, cette mémoire est une compilation des travaux de recherche liés au statistique d'ordre et du plan de travail suivant :

***Chapitre* 1 :** Dans ce chapitre, nous parlerons du rôle des statistiques d'ordre dans beaucoup de théories différentes. Nous examinerons également les applications des statistiques d'ordre en créant une simple liste de paramètres dans les quells les statistiques d'ordre peuvent jouer un rôle important.

***Chapitre* 2 :** Dans ce chapitre, on passe en revue la définition et quelques propriétés et des théories des statistiques d'ordr et des quantiles, nous avons présenté la simulation des statistiques d'ordre.

***Chapitre* 3 :** Dans ce chapitre, nous nous intéressons à la simulation des statistiques d'ordre et utilisons trois méthodes de simulation à l'aide de R, la première méthode est méthode de la fonction de répartition empirique, la deuxième méthode est méthode dircte et la troisième méthode est représentation de Reny.

Chapitre 1

Domaines d'application des statistiques d'ordre

Une branche de la statistique connue sous le nom des statistiques d'ordre joue un rôle prépondérant dans beaucoup des théories par exemple : la théorie du moment L et théorie des valeurs extrêmes et d'estimation des quantiles...

L'étude de l'ordre statistique est l'étude des statistiques des variables aléatoires ordonnées (triées) et des échantillons.

Ce chapitre présente une très brève introduction aux statistiques d'ordre pour fournir une base pour les chapitres suivants. Une exposition complète de l'ordre, les statistiques sont fournies par David (1981), et une approche orientée R est décrite dans divers contextes par Baclawski (2008).

Il y a des années, l'étude des statistiques d'ordre était une curiosité. L'apparence de volume édité de Sarhan et Greenberg en 1962, et le traité de Herbert Aron David sur le sujet en 1970 et sa révision ultérieure en 1981 ont changé tout cela. Jusqu'à présent, le livre cité par le professeur David a été utilisé à la fois comme texte et comme en quête de plus en plus et introduction aux statistiques d'ordre maintenant.

En plus de Sarhan et Herbert Aron David, nous trouvons aussi. Balakrishnan, Gnanadesikan, Gnedenko et Biometrika... ils ont reçu le plus grand crédit pour enrichir et développer les statistiques d'ordre.

1.1 Les applications des statistiques d'ordre

Voici une brève liste des paramètres dans les quels les statistiques d'ordre peuvent avoir un rôle important :

1. ***Estimations d'emplacement robustes* :** Supposons que n mesures indépendantes sont disponibles, et nous souhaitons estimer leur moyenne commune supposée. Il a longtemps il a été reconnu que la moyenne de l'échantillon, bien qu'attrayante à plusieurs égards, souffre d'une extrême sensibilité aux valeurs aberrantes et aux violations de modèles. Estimations sur la base de la médiane ou de la moyenne des statistiques de l'ordre central sont moins sensibles hypothèses de modèle. Une application particulièrement connue de cette observation est la pratique acceptée d'utiliser des moyens ajustés (en ignorant les scores les plus élevés et les plus bas) dans l'évaluation des performances olympiques de patinage artistique.

2. ***Détection des valeurs aberrantes* :** Si l'on est confronté à un ensemble de mesures et est soucieux de déterminer si certains ont été mal faits ou signalés, l'attention se concentre naturellement sur certaines statistiques d'ordre de l'échantillon. Généralement les plus grands un ou deux et / ou les plus petits sont considérés comme les plus susceptibles valeurs aberrantes. En règle générale, nous posons des questions comme les suivantes : si l'observation s'étaient iid, quelle est la probabilité que la statistique d'ordre le plus récent soit aussi grande que la valeur étrangement grande que nous avons observée ?

3. ***Échantillonnage censuré* :** Expériences de test de durée de vie de Consister, dans les quelles un nombre fixe en éléments sont mis à l'essai et l'expérience est terminée dès qu'un le nombre prescrit r a échoué. Les durées de vie observées sont donc $X_{1,n} \leq ... \leq X_{r,n}$, tandis que les durées de vie $X_{r+1,n} \leq ... \leq X_{n,n}$ restent non observées. Par exemple : Cinquante machines chères démarrent dans un expérience pour déterminer la durée de vie prévue d'une machine. Si comme on peut l'espérer, ils sont assez fiables, il faudrait énormément de temps pour attendre tous machines à échouer. Au lieu de cela, de grandes économies de temps et de machines peuvent être effectuées si nous basons nos estimations sur les premiers temps d'échec (c.-à-d les premiers ordres statistiques de l'échantillon conceptuel des temps de défaillance indépendantes et identiquement distribuées (iid)). Notez que nous pouvons bien retirer les machines non défaillantes de l'environnement de test pour les vendre et / ou modification. Seuls les articles défectueux sont endommagés (et, ou) détruits. Tel les procédures sont particulièrement importantes du point de vue des essais destructifs (combien de moteurs

Mercedes-Benz devez-vous brûler en cas de panne d'huile expériences ?). D'un point de vue éthique, une pression encore plus grande est exercée chercheur médical dans ses efforts pour se renseigner sur l'efficacité de traitements avec un minimum d'échecs réels (patients).

4. ***En attendant le grand*** **:** Inondations désastreuses et tremblements de terre des tructeursse reproduisent à travers l'histoire. La construction du barrage s'est longtemps concentrée sur Inondations de 100 ans. Vraisemblablement, les barrages sont construits assez gros et assez solides pour gérer tout débit d'eau à rencontrer, sauf pour un niveau prévu pour ne se produisent qu'une fois tous les 100 ans. Les architectes en Californie sont particulièrement concernés par la construction conçue pour résister au «grand», présumant un tremblement de terre d'une force énorme, peut-être un «tremblement de terre de 100 ans». Que l'on soit d'accord ou non avec la philosophie du désastre centenaire, il est évident que les concepteurs de barrages et de gratte-ciel et même les niches, devraient être préoccupé par la distribution de statistiques d'ordre important es séquence dépendante, éventuellement non distribuée de façon identique.

5. ***Résistance des matériaux*** **:** L'adage selon lequel une chaîne n'est pas plus forte que son maillon le plus faible sous-tend une grande partie de la théorie de la résistance des matériaux, ce sont des fils, des feuilles ou des blocs. En considérant le potentiel de défaillance dans des sections finement petites du matériau, on est rapidement conduit à la résistance distributions associées aux limites de distribution des minima d'échantillons. De bien sûr, si nous nous en tenons à la chaîne finie avec n maillons, sa force serait la minimum des points forts de ses n composants, encore une statistique d'ordre.

6. ***Fiabilité :*** L'exemple d'un cordon composé de n fils peut être étendu pour nous conduire à des applications de fiabilité des statistiques d'ordre. C'est possible cette défaillance d'un fil entraînera la rupture du cordon (le maillon le plus faible), mais plus probablement le cordon fonctionnera aussi longtemps que k (un nombre inférieur à n) les fils restent intacts. En tant que tel, il est un exemple de A : de n système Souvent discuté dans les paramètres de fiabilité. En ce qui concerne la défaillance des pneus, L'automobile est souvent un exemple de système 4 sur 5 (rappelez-vous la pièce de rechange).Empruntant à la terminologie des systèmes électriques, le système n sur n est également connu sous le nom de système en série. Toute défaillance d'un composant est désastreuse. Le 1 sur du système n est connu comme un système parallèle, il fonctionnera aussi longtemps que l'un des les composants survivent. La durée de vie du système k sur n est clairement $X_{n-k+1,n}$, le $(n-k+1)$ est la plus grande des durées de vie des composants ou de manière équivalente, la jusqu'à ce que moins de k composants fonctionnent. Autres plus compliqués les structures du système peuvent être envisagées mais en fait ils peuvent être considérés comme hiérarchies peut-être compliquées de sous-systèmes parallèles et en série, et l'étude de la durée de vie du système impliquera nécessairement des distributions de statistiques.

7. ***Contrôle de qualité :*** Prenez une chaise confortable et regardez la production quotidienne des barres chocolatées Snickers passent sur le tapis roulant. Chaque barre chocolatée doit peser $2,1$ onces, juste un peu plus que le poids indiqué sur l'emballage. Peu importe dans quelle mesure la verseuse de bonbons vient d'être ajustée au début du changement, des fluctuations

mineures se produiront et des aberrations potentiellement importantes pourraient être rencontré (si une cacahuète se coince dans la valve de contrôle). Nous devons être vigilants dysfonctionnements corrigibles entraînant une variation déraisonnable du poids de la barre chocolatée. Entrez l'homme de contrôle qualité avec ses graphiques $\bar{X}$ et R ou sa médiane et R graphiques. Un échantillon de friandises est pesé toutes les heures et une attention particulière est porté eaux statistiques d'ordre des poids ainsi obtenus. Si la médiane (ou peut-être la moyenne) est loin du valeur cible, il faut fermer la ligne. Soit nous sommes tourner des barres maigres et entendra des enfants mécontents de six ans, ou nous sommes tourner des barres en surpoids et gaspiller de l'argent (nous allons donc entendre des mécontents la gestion). L'attention est également concentrée sur la plage d'échantillonnage, le plus grand moins le plus petit poids. S'il est trop important, le processus est hors de contrôle et la fluctuation du poids des barres chocolates causera probablement des problèmes plus loin. Encore une fois, nous nous arrêtons et cherchons à identifier une cause corrigible avant de redémarrer le ligne snickers.

8. ***Sélection du meilleur*** **:** Les essais sur le terrain de variétés de mais ont impliqué expériences menées pour déterminer laquelle de plusieurs variétés est la plus productive. De toute évidence, nous sommes préoccupés par le maximum d'un ensemble de probablement pas variables réparties de manière identique dans un tel paramètre. La situation n'est pas différente celui discuté précédemment dans le contexte de l'identification des valeurs aberrantes. Dans la situation actuelle, la valeur aberrante (la meilleure variété) est cependant bonne et mérite

rétention (plutôt que d'être rejetée ou actualisée comme ce serait le cas dans le paramètre aberrant habituel). Un autre exemple en biologie dans lequel l'ordre les statistiques jouent clairement un rôle dans l'élevage sélectif par abattage. Ici peut-étrilles 10% supérieurs en ce qui concerne la viande des animaux dans chaque génération sont conservé à des fins de reproduction. Une brève discussion de ce sujet sera trouvée dans la section 8.6.

9. ***Inégalité de mesure*** **:** La répartition des revenus en Bolivie (où un peu d'individus gagnent la majeure partie de l'argent) est nettement plus inégalitaire que Suède (où la fiscalité progressive a un effet de nivellement). Comment fait-on rendre ces déclarations précises ? L'approche habituelle implique des statistiques d'ordre des distributions de revenus correspondantes. Le dispositif particulier utilisé est appelé une courbe de Lorenz. Il résume le pourcentage du revenu total revenant à les p pour cent les plus pauvres de la population pour diverses valeurs de p. Mathématiques ce n'est que l'intégrale mise à l'échelle de la fonction quantile empirique, un fonction avec saut $X_{i,n}$au point $\frac{i}{n}, i = 1, ..., n$ (où n est le nombre de revenus individuels dans la population). Un haut degré de convexité dans la courbe de Lorenz signale une forte inégalité des revenus distribution

10. ***Records olympiques*** **:** Le saut en longueur de Bob Beamon en 1968 Livre des records olympiques. Peu d'autres enregistrements durent aussi longtemps. Si les meilleures perfora chaque Jeux Olympiques ont été modélisés comme indépendants de manière identique des variables aléatoires distribuées, les enregistrements deviendraient de plus en plus rare avec

le temps. Tel n'est pas le cas. L'explication la plus simple implique l'amélioration et l'augmentation des populations. Ainsi le saut en hauteur de 1964 champion était le meilleur de, disons, N_1 cavaliers de calibre international actifs. Dans 1968, il y avait plus de cavaliers de haut calibre probablement de plus haut calibre. Alors on recherche, très probablement une séquence d'aléas non distribués de façon identique variables. Mais en tout cas, nous nous concentrons sur les maxima, c'est-à-dire sur certaines statistiques d'ordre. Plus de détails sur les valeurs d'enregistrement et en particulier les valeurs d'enregistrement l'amélioration des populations se trouve au chapitre 9. Bien sûr ces modèles rendre la survie du record de Beamon encore plus problématique.

11. ***Caractérisations et qualitéde l′ justement :*** La distribution exponentielle est famous pour son soi-disant manque de mémoire. Le modèle habituel implique une ampoule ouaté appareil électronique. L'argument veut qu'une ampoule qui a été en service 20 heures n'est ni plus ni moins susceptible d'échouer dans la minute suivante qu'un qui a été en service pendant disons 5 heures ou même pour ce compteur, qu'une marque nouvelle ampoule. Une situation de distribution aussi curieuse se reflète dans les statistiques d'ordre à partir d'échantillons exponentiels. Par exemple, si $X_1, ..., X_n$ sont exponentielles iid, alors leurs espacements $(X_i - X_{i-1})$ sont à nouveau exponentiels et remarquablement sont indépendants bosse. Ce n'est que dans le cas de variables aléatoires exponentielles que ces espacements des propriétés sont rencontrés. Une vaste littérature de caractérisations exponentielles et les tests de qualité d'ajustement associés se sont développés en conséquence. On remarque au passage

que la plupart des tests d'adéquation à toute distribution parentale impliquent implicitement der les statistiques, car ils se concentrent souvent sur les écarts entre le quantile empirique fonction et la fonction quantile hypothétique.

Chapitre 2

Statistique d'ordre

Ce chapitre est principalement consacré à l'étude des propriétés des statistiques d'ordre associées à un échantillon. On obtient une réalisation des statistiques d'ordre en rangeant par ordre croissant les réalisations de l'échantillon initial.

Les statistiques d'ordre jouent un rôle de plus en plus important dans la pratique. Nous commençons dans ce chapitre par donner les définitions et quelques propriétés des statistiques d'ordre puis nous étudions les lois exactes et asymptotiques des statistiques d'ordres , simulation et quelques théories des statistiques d'ordres.

2.1 Statistique d'ordre

Définition 2.1.1 :

Soit $(X_1, ..., X_n)$ *est une suite des variables indépendantes et identiquement distribuées(iid) de loi de probabilité* F *continue, on appelle statistique d'ordre associée à l'échantillon* $(X_1, ..., X_n)$, *la suite notée* $(X_{1,n}, ..., X_{n,n})$ *telle que :*

$$X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}, \textit{ presque surement.}$$

• Pour tout $k = 1, ..., n$ la variable aléatoire $X_{k,n}$ s'appelle étant la k^{ieme}statistique d'ordre.

• La première statistique d'ordre $X_{1,n}$ est toujours le minimum de l'échantillon et $X_{n,n}$ le maximum, en d'autre termes

$$X_{1,n} = \min(X_1, ..., X_n) \text{ et } X_{n,n} = \max(X_1, ..., X_n).$$

La fonction de répartition de $X_{1,n}$ est donnée par :

$$\begin{aligned} F_{X_{1,n}}(x) &= P(X_{1,n} \leq x) = 1 - P(X_{1,n} > x) \\ &= 1 - P(\min(x_1, ..., x_n) > x) \\ &= 1 - P(\prod_{i=1}^{n} x_i > x) \\ &= 1 - \prod_{i=1}^{n} [1 - P(x_i \leq x)] = 1 - \prod_{i=1}^{n} [1 - F_X(x)] . \end{aligned}$$

D'où déduire

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

et la densité associée est donnée par :

$$f_{X_{1,n}}(x) = n [1 - F(x)]^{n-1} f(x).$$

La fonction de répartition de $X_{n,n}$ est donnée par :

$$\begin{aligned} F_{X_{n,n}}(x) &= P(X_{n,n} \leq x) = P(\max(x_1, ..., x_n) \leq x) \\ &= P(x_1 \leq x, ..., x_n \leq x) \\ &= P(\prod_{i=1}^{n} [X_i \leq x]) = \prod_{i=1}^{n} P(X_i \leq x) = \prod_{i=1}^{n} F(x), \end{aligned}$$

ainsi

$$F_{X_{n,n}}(x) = (F(x))^n,$$

et la densité associée est donnée par :

$$f_{Xn,n}(x) = n(F(x))^{n-1}f(x).$$

De ces résultats, nous en tirons la conclusion que le maximum $X_{n,n}$ est une variable aléatoire dont la fonction de répartition correspondante est F^n. La fonction de répartition de X n'étant pas souvent connue, il n'est généralement pas possible de déterminer la distribution du maximum à partir de ce résultat. On s'intéresse alors à la distribution asymptotique du maximum en faisant tendre n vers l'infini. On a :

$$\lim_{n \to +\infty} F_{X_{n,n}} = \lim_{n \to +\infty} [F(x)]^n = \begin{cases} 1 & \text{si } F(x) = 1, \\ 0 & \text{si } F(x) < 1. \end{cases}$$

On constate que la distribution asymptotique du maximum, déterminée en faisant n tendre vers l'infini, donne une loi dégénérée (ils prennent des valeurs de 0 et 1 seulement).

Lemme 2.1.1 :

1. *Si F est continue, alors presque sûrement on a $X_{1,n} < ... < X_{n,n}$.*

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

On suppose F continue, la loi de σ_n est la loi uniforme sur S_n. De plus la permutation σ_n est indépendante de la statistique d'ordre.

Preuve.

1. Il suffit de vérifier que $P(\exists i \neq j$ tel que $X_i = X_j) = 0$.

   On a

$$P(\exists i \neq j \text{ tel que } X_i = X_j) \leq P(\exists i \neq j \text{ tel que } F(X_i) = F(X_j))$$
$$\leq \sum_{i \neq j} P(F(X_i) = F(X_j)).$$

   Les variables $F(X_i)$ et $F(X_j)$ sont pour des variables aléatoires indépendantes de loi uniforme sur $[0,1]$. On déduit que

$$P(F(X_i) = F(X_j) = \int_{[0,1]^2} 1_{\{u=v\}} dudv = 0.$$

   Donc presque sûrement pour tous $i \neq j$ on a $X_i \neq X_j$.

2. Soit $\sigma_n \in S_n$. on a $P(\sigma_n = \sigma) = P(X_{\sigma_n(1)} < ... < X_{\sigma_n(n)})$.

   Les variables $(X_{\sigma_n(1)}, ..., X_{\sigma_n(n)})$ sont indépendantes et de même loi. En particulier le vecteur $(X_{\sigma_n(1)}, ..., X_{\sigma_n(n)})$ a même loi que $(X_1, ..., X_n)$. Il vient

$$P(\sigma_n = \sigma) = P(X_1 < ... < X_n).$$

   Le membre de droite est indépendant de σ. La loi de σ_n est donc la loi uniformesur S_n, et on a $P(\sigma_n = \sigma) = \frac{1}{n!}$. Soit g une fonction de $\mathbb{R}^n$ dans R, mesurable bornée.

■

Comme le vecteur $(X_{\sigma_n(1)}, ..., X_{\sigma_n(n)})$ a même loi que $(X_1, ..., X_n)$, on a

$$E\left[1_{\{\sigma_n=\sigma\}} g(X_{1,n}, ..., X_{n,n})\right] = E\left[1_{\left\{X_{\sigma_n(1)} < ... < X_{\sigma_n(n)}\right\}} g(X_{\sigma_n(1)}, ..., X_{\sigma_n(n)})\right]$$
$$= E\left[1_{\{X_1 < ... < X_n\}} g(X_1, ..., X_n)\right].$$

On en déduit, en sommant sur $\sigma_n \in S_n$, que

$$E\left[g(X_{1,n},...,X_{n,n})\right] = n!E\left[1_{\{X_1<...<X_n\}}g(X_1,...,X_n)\right].$$

Enfin, on remarque que

$$E\left[1_{\{\sigma_n=\sigma\}}g(X_{1,n},...,X_{n,n})\right] = P(\sigma_n=\sigma)E\left[g(X_{1,n},...,X_{n,n})\right]$$

Cela implique que la permutation σ_n est indépendante de la statistique d'ordre.

2.1.1 Propriétés générales

Soit $(X_1,...,X_n)$ est une suite de variables iid :

1. $f_{X_{1,n},...,X_{n,n}}(x_1,...,x_n) = \begin{cases} n!\prod\limits_{i=1}^{n} f(x_i) & \text{si } -\infty < x_1 < ... < x_n < +\infty. \\ 0 & \text{sinon.} \end{cases}$
2. $F_{X_{k,n}}(x) = \sum\limits_{i=k}^{n} C_n^i \left[F(x)\right]^i \left[1-F(x)\right]^{n-i}, \forall x \in \mathbb{R}.$
3. $f_{X_{k,n}}(x) = nC_{n-1}^{k-1}\left[F(x)\right]^{k-1}\left[1-F(x)\right]^{n-k}$, telle que $C_{n-1}^{k-1} = \frac{(n-1)!}{(k-1)!(n-k)!}$.
4. $F_{X_{r,n},X_{s,n}}(x,y) = \sum\limits_{s=j}^{n}\sum\limits_{r=i}^{s}\frac{n!}{r!(s-r)!(n-s)!}\left[F(x)\right]^r\left[F(y)-F(x)\right]^{s-r}\left[1-F(y)\right]^{n-s}, \forall x,y \in \mathbb{R}.$
5. $f_{X_{r,n},X_{s,n}}(x,y) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}\left[F(x)\right]^{r-1}f(x)\left[F(y)-F(x)\right]^{s-r-1}f(y)\left[1-F(y)\right]^{n-s}, \forall x,y \in \mathbb{R}.$

Preuve.

1. On rappelle que, puisque $f(x)$ est la dérivée de $F(x)$, alors

$$\begin{aligned}
f(x) &= \lim_{\delta x\to 0}\frac{F(x+\delta x)-F(x)}{\delta x} = \lim_{\delta x\to 0}\frac{P(x\leq X<x+\delta x)}{\delta x}\\
&= \frac{P(x_1\leq X_{1,n}<x_1+\delta x_1,...,x_n\leq X_{n,n}<x_n+\delta x_n)}{\delta x_1...\delta x_n},\ \text{quand } \delta x_i\to 0\\
&= \frac{n!P(x_1\leq X_{1,n}<x_1+\delta x_1)...P(x_n\leq X_{n,n}<x_n+\delta x_n)}{\delta x_1...\delta x_n}\\
&= n!\frac{P(x_1\leq X_{1,n}<x_1+\delta x_1)}{\delta x_1}...\frac{P(x_n\leq X_{n,n}<x_n+\delta x_n)}{\delta x_n}\\
&\boxed{f_{X_{1,n},...,X_{n,n}}(x_1,...,x_n) = n!\prod_{i=1}^{n} f(x_i),\ \text{avec } -\infty<x_1\leq...\leq x_n<+\infty.}
\end{aligned}$$

2. Le calcule de la fonction de répartition $F_k(x)$ de $X_{k,n}$ est immédiat :

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

3. La fonction de répartition $F_k(x)$ de $X_{k,n}$ est donnée par $\sum_{i=k}^{n} C_n^i\left[F(x)\right]^i\left[1-F(x)\right]^{n-i}$, $\forall x\in$ $\mathbb{R}$.La densité de $X_{k,n}$ est donnée par $nC_{n-1}^{k-1}\left[F(x)\right]^{k-1}\left[1-F(x)\right]^{n-k}$. Ça peut être montré par le fait que

$$f_{X_{k,n}}(x) = \frac{d}{dp}\sum_{i=k}^{n} C_n^i\left[p\right]^i\left[1-p\right]^{n-i} = nC_{n-1}^{k-1}\left[p\right]^{k-1}\left[1-p\right]^{n-k}.$$

4. La fonction de distribution jointe de$(X_{r,n}, X_{s,n})$ est :

$$F_{X_{r,n},X_{s,n}}(x,y) = P(\{X_{r,n} \leq x\} \cap \{X_{s,n} \leq y\}), (x,y) \in \mathbb{R}.$$

On a deux cas :

$\mathbf{1}^{iére}$ ***cas*** : $x \geq y$

$$\begin{aligned} F_{X_{r,n},X_{s,n}}(x,y) &= P(X_{r,n} \leq x, X_{s,n} \leq y) = P(X_{s,n} \leq y) F_{X_{r,n},X_{s,n}}(x,y) \\ &= F_{X_{r,n},X_{s,n}}(y). \end{aligned}$$

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

$$\begin{aligned} F_{X_{r,n},X_{s,n}}(x,y) &= P(X_{r,n} \leq x, X_{s,n} \leq y) \\ &= P(\text{au moin } r \text{ de } X_1, ..., X_n \text{sont inférieurs à } x \text{ et au moins } s \text{ de } X_1, .., X_n \\ &\quad \text{sont inférieurs à } y) \\ &= \sum_{s=j}^{n} \sum_{r=i}^{s} \frac{n!}{r!(s-r)!(n-s)!} [F(x)]^r [F(y) - F(x)]^{s-r} [1 - F(y)]^{n-s}, \forall (x,y) \in \mathbb{R}. \end{aligned}$$

5. La fonction de répartition $F_{X_{r,n},X_{s,n}}(x,y)$ de $(X_{r,n}, X_{s,n})$ est donnée par

$$\sum_{s=j}^{n} \sum_{r=i}^{s} \frac{n!}{r!(s-r)!(n-s)!} [F(x)]^r [F(y) - F(x)]^{s-r} [1 - F(y)]^{n-s}, \forall (x,y) \in \mathbb{R},$$

donc la densité de $(X_{r,n}, X_{s,n})$ est :

$$\begin{aligned} f_{X_{r,n},X_{s,n}}(x,y) &= \frac{d}{dp} \sum_{s=j}^{n} \sum_{r=i}^{s} \frac{n!}{r!(s-r)!(n-s)!} [F(x)]^r [F(y) - F(x)]^{s-r} [1 - F(y)]^{n-s} \\ &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} [F(x)]^{r-1} f(x) [F(y) - F(x)]^{s-r-1} f(y) [1 - F(y)]^{n-s}. \end{aligned}$$

■

Remarque 2.1.1 :

• *Pour* $k = 1, ..., n$ *on a*

$$F_{X_{k,n}}(x) = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n+1-k)} \int_0^{F(x)} y^{k+1}(1-y)^{n-k}dy.$$

C'est,

$$F_{X_{k,n}}(x) = F_{\beta(k,n+1-k)}(F(x)).$$

En particulier, si $X \in [0,1]$

$$X_{k,n} \in \beta(k, n+1-k).$$

$$\frac{n!}{(k-1)!(n-k)!} = \frac{1}{\beta(k, n+1-k)}.$$

2.1.2 uniforme des statistiques d'ordre

Soit $U_1, ..., U_n$ un échantillon aléatoire de la distribution uniforme $[0,1]$ et $X_1, ..., X_n$ échantillon aléatoire d'une population de distribution $F(x)$. De plus, laissez $U_{1,n} \leq ... \leq U_{n,n}$ et $X_{1,n} \leq ... \leq X_{n,n}$ être les statistiques d'ordre obtenues à partir de ces échantillons.

Plus précisément, lorsque $F(x)$ est la transformation intégrale de probabilité continue $U = F(x)$ produit une distribution uniforme standard. Ainsi, lorsque $F(x)$ est nous avons continu $F(X_{i,n}) \stackrel{d}{=} U_{i,n}, 1 < i < n$

$$X_{1,n}, ..., X_{n,n} \stackrel{d}{=} (F^{-1}(U_1)), ..., F^{-1}(U_{n,n})).$$

où $\stackrel{d}{=}$ doit être lu comme «a la même distribution que».

2.1.3 La fonction de répartition empirique

Définition 2.1.2 :

On appelle fonction de répartition empirique associée à un échantillon $X_1, ..., X_n$ la fonction F_n définie par :

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{\{X_k \leq x\}} = \frac{1}{n}Card\{k \in \{1, ..., n\} : X_k \leq x\}$$
$$= \begin{cases} 0 & si \;\; x < X_{1,n}, \\ \frac{1}{n} & si \;\; X_{1,n} < x < X_{2,n}, \\ \frac{2}{n} & si \;\; X_{2,n} < x < X_{3,n}, \\ . & ... \\ . & ... \\ \frac{k}{n} & si \;\; X_{k,n} < x < X_{k+1,n}, \\ . & ... \\ . & ... \\ 1 & si \;\; x \geq X_{n,n}. \end{cases}$$

où $\mathbf{1}_{\{X_k \leq x\}}$: c'est la fonction indicatrice, telle que

$$\mathbf{1}_{\{X_k \leq x\}} = \begin{cases} 0 & \text{si} \;\; X_k > x, \\ 1 & \text{si } X_k \leq x. \end{cases}$$

2.1.4 Convergence de la fonction de répartition empirique

Théorème 2.1.1 . [04]

Soit $(X_n, n \in N)$ une suite de variables aléatoires réelles, indépendantes et de méme loi F alors :

$$\lim_{n \to +\infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0,$$

Convergence presque sûre.

2.1.5 Convergence presque sure des statistiques d'ordre

Proposition 2.1.1 . [04]

Soit F une fonction de répartition, telle que $x_F \leq +\infty$ et $(k(n), n \in N)$ une suite d'entiers naturels non décroissante, telle que :

$$\lim_{n \to +\infty} \frac{k(n)}{n} = c, \;\; c \in [0, 1] \, .$$

a) $X_{(k(n),n)} \overset{p.s.}{\to} x_F$ avec $c = 0$ respectivement $(c = 1)$.

b) Si $c \in (0, 1)$ alors, $x(c)$ est la solution unique de l'équation $F(x) = c$.

2.1.6 Limiter les distributions le maximum et le minimum

$F_{X_{1,n}}(x) = P(X_{1,n} \leq x) = 1 - [1 - F(x)]^n \, .$

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

Lorsque n tend vers l'infini, nous obtenons

$$\lim_{n \to +\infty} F_{X_{1,n}}(x) = \lim_{n \to +\infty} 1 - [1 - F(x)]^n = \begin{cases} 0 & \text{si } F(x) = 0, \\ 1 & \text{si } F(x) > 1. \end{cases}$$

$$\lim_{n \to +\infty} F_{X_{n,n}}(x) = \lim_{n \to +\infty} [F(x)]^n = \begin{cases} 0 & \text{si } F(x) = 1, \\ 1 & \text{si } F(x) < 1. \end{cases}$$

2.1.7 Les quantiles

Les quantiles d'une variable aléatoire discrète (entière) ou continue (réelle) sont les valeurs que prend la variable pour des valeurs de probabilité p $(0 < p < 1)$. On les appelles encore fractile, et ce sont les valeurs réciproques de la fonction de répartition de la loi de probabilité considérée.

Les quantiles d'un échantillon statistique des nombres sont les valeurs remarquables permettant de diviser le jeu de ces données ordonnées c'est à dire triées en intervalle consécutifs contenant le même nombre de données.

Fonction des quantiles

Définition 2.1.3 :

L'inverse généralisé de la fonction de répartition appelée la fonction des quantiles "notée Q" tele que :

$$Q(t) = F^{-1}(t) = \inf\{s \in \mathbb{R} : F(s) \geq t\}, t \in \left]0, 1\right[$$

Quantile d'ordre p

Définition 2.1.4 :

Soit X une variable aléatoire réelle de loi F, soit $p \in \left]0, 1\right[$. On appelle p-quantile de F toute réel $(Q_P(F) = \xi_p(F))$ vérifiant

$$F(Q_P(F) - 0) \leq p \leq F(Q_P(F) + 0).$$

est défini comme :

$$Q_P = x_p = F^{-1}(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\}.$$

Usuellement, un $1/2$-quantile est appellée la médiane de F, les $k/4$-quantile sont appellés quartiles de F, les $k/10$-quantile sont appellés déciles de F, ect...

- Si la fonction F est continu et strictement croissant, le p-quantile $Q_P(F)$ est unique et vérifie

$$F^{-1}(p) = Q_P(F), \text{ pour tout } p \in \left[a, b\right[.$$

• S'il existe un intervalle $[a, b[$ contenu dans $\mathbb{R}$ tel que

$$\forall x \in [a, b[\,,\ \ F(x) = p,$$

avec $p \in \,]0, 1[\,$, alors tout point $x \in [a, b]$ est un p-quantile. C'est le seul cas où il n'ya pas unicité du p-quantile.

• Si la fonction F est discontinuité en x_0, C'est-à-dire si $F(x_0 - 0) < F(x_0 + 0)$, alors $Q_P(F)$ est unique et égal à x_0 pour tout $p \in \,]F(x_0 - 0), F(x_0)[\,$. la figure 1 illustre ces divers cas.

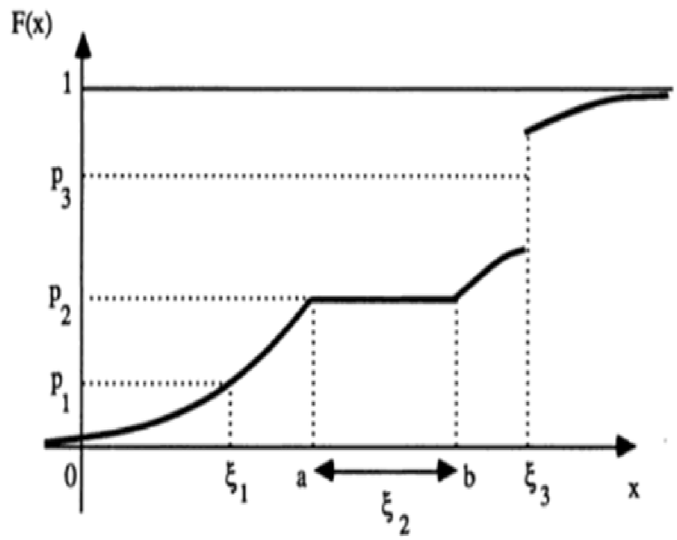

figure.1.

Exemple 2.1.1 :

soit $(X_{1,n}, X_{2,n}, ..., X_{5,n})$ est un échantillon rangé de taille $n = 5$, la médiane empirique est $X_{3,n}$, c'est la valeur qui partage l'effectif de l'échantillon en 2 parties de même effectif.

Quantile empirique

Définition 2.1.5 :

Soit $X_1, ..., X_n$ un échantillon d'une loi F. soit $p \in]0,1[$. on appelle p-quantile empirique, et on note $Q_n(p)$ est donnée par :

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

La fonction quantile empirique de la queue correspondante est définie par :

$$U_n(t) = F_n^{\leftarrow}(1 - \frac{1}{t}) \text{ avec } 0 < t < \infty.$$

où $F^{\leftarrow}$: Inverse généralisée de la fonction de répartition.

On peut représenté la fonction des quantiles empirique en terme des statistiques d'ordre

$$Q_n(t) = \begin{cases} X_{j,n} & \text{si } \frac{j-1}{n} < t \leq \frac{j}{n}, \\ X_{jn,n} & \text{si } 0 < t < 1. \end{cases}$$

Lemme 2.1.2 *Soit F une fonction de distribution. La fonction $F^{-1}(t), \ 0 < t < 1$, croissante et gauche-continue, et satisfait :*

i $F^{-1}(F(x)) \leq x$, $-\infty < x < \infty$.

ii $F(F^{-1}(t)) \geq t$, $0 < t < 1$.

Par conséquent

iii $F(x) \geq t$ si et seulement si $x \geq F^{-1}(t)$.

Correspondant à un échantillon $\{X_1, ..., X_n\}$ d'observations sur F, l'échantillon pth quantile est défini comme le premier quantile de la fonction de distribution d'échantillon F_n, c'est-à-dire $F_n^{-1}(p)$. En ce qui concerne le pth quantile de l'échantillon comme estimateur de ξ_P, nous le désignons par $\hat{\xi}_{Pn}$, ou simplement par $\hat{\xi}_{Pn}$ quand cela est commode.

La statistique d'ordre étant équivalente à la fonction de distribution d'échantillon F_n, son rôle est fondamentale même si elle n'est pas toujours explicite. Ainsi, par exemple, la moyenne de l'échantillon peut être considérée comme la moyenne des statistiques d'ordre, et le premier quantile d'échantillon peut être exprimé comme

$$\hat{\xi}_{Pn} \begin{cases} X_{n,np} & \text{si } np \text{ est un entier,} \\ X_{n,[np]+1} & \text{si } np \text{ n'est pas un entier.} \end{cases}$$

Comportement asymptotique d'un quantile d'ordre centrale

Définition 2.1.6 *Soient n variables aléatoires $X_1, ..., X_n$ iid de fonction de répartition commune F et de densité f.*

La statistique d'ordre $X_{(k(n),n)}$est dite centrale si $\exists p$ tel que :

$$\lim_{n\to\infty} \sqrt{n}(\frac{k(n)}{n} - p) = 0$$

Théorème 2.1.2 *Soit $0 < p < 1$ et supposons que F possède une dérivée au voisinage de x_p avec $0 < f(x_p) < \infty$ et f continue au point x_p, alors :*

$$\sqrt{n}(\hat{X}([np]+1, n) - x_p) \xrightarrow{L} N(0, \frac{p(1-p)}{[f(F^{-1}(p))]^2}).$$

2.1.8 Fonctions des statistiques d'ordre

Nous considérons ici les statistiques qui peuvent être exprimées en fonction des statistiques d'ordre. Une variété de procédures raccourcies pour des estimations rapides des paramètres de localisation ou d'échelle, ou pour des tests rapides d'hypothèse sont connexes, sont fournies sous la forme de fonctions linéaire de statistique d'ordre, c'est-à-dire statistique de la forme

$$\sum_{i=1}^{n} c_{ni} X_{(i,n)}.$$

Nous appelons ces statistiques des " L-estimations ". Par exemple, la moyenne ajustée qui est un concurrent populaire de $\bar{X}$ pour une estimation robuste de l'emplacement. La théorie de la distribution asymptotique des statistiques prend des formes très différentes, selon le caractère du coefficients $\{c_{ni}\}$.

Les représentations de $\overline{X}$ et $\hat{Q}_n$ en termes des statistiques d'ordre sont un peu artificielles sur d'autre part, pour de nombreuses statistiques utiles, les représentations les plus naturelles et efficaces sont en termes de statistique d'ordre. Les exemples sont les valeurs extrêmes $X_{1,n}$ et $X_{n,n}$ et l'échantillon gamma $X_{n,n} - X_{1,n}$.

2.1.9 Distribution conditionnelle de statistiques d'ordre

Proposition 2.1.2 :

Soit $X_1, ..., X_n$ n variable aléatoire (iid) de densité commune (cd) f. Alors pour $1 \leq r \leq s \leq n$, la densité conditionnelle $f_{(X_{r,n}/X_{s,n})}$ de $X_{r,n}$ sachant l'événement $\{X_{s,n} = y\}$ est donnée par :

$$f_{(X_{r,n}/X_{s,n})}(x/y) = \frac{(s-1)!}{(r-1)!(s-r-1)!} [F(x)]^{r-1} f(x) [F(y) - F(x)]^{s-r-1} [F(y)]^{1-s}, \text{ avec } -\infty < x < y < \infty.$$

Preuve.

On rappelle que

$$f_{(X_{r,n}/X_{s,n})}(x/y) = \frac{f_{(X_{r,n},X_{s,n})}(x,y)}{f_{X_{s,n}}(y)},$$

alors pour $1 \leq r < s \leq n$, on a :

$$\begin{aligned} f_{(X_{r,n}/X_{s,n})}(x/y) &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} [F(x)]^{r-1} f(x) [F(y) - F(x)]^{s-r-1} f(y) [1 - F(y)]^{n-s} \\ &\times \frac{(s-1)!(n-s)!}{n!} [F(y)]^{1-s} [f(y)]^{-1} [1 - F(y)]^{s-n} \\ &= \frac{(s-1)!}{(r-1)!(s-r-1)!} [F(x)]^{r-1} f(x) [F(y) - F(x)]^{s-r-1} [F(y)]^{1-s}, \end{aligned}$$

■

avec $-\infty < x < y < \infty$.

Théorème 2.1.3 :

Soit $X_1, ..., X_n$ un échantillon aléatoire d'une population absolument continue avec $F(x)$ et la fonction de densité $f(x)$, et soit $X_{1,n} \leq ... \leq X_{n,n}$ désignent les statistiques d'ordre obtenues à partir de cette échantillon. Alors la distribution conditionnelle de $X_{j,n}$, étant donné que $X_{i,n} = x_i$ pour $i < j$, est la même que la distribution de la $(j - i)$ statistique d'ordre obtenu à partir d'un échantillon de taille $n - i$ d'une population dont la distribution est simplement $F(x)$ tronquée à gauche en x_i.

Preuve.

D'après la fonction de densité marginale de $X_{i,n}$ et la fonction de densité conjointe de $X_{i,n}$ et $X_{j,n}$, nous avons la fonction de densité conditionnelle de $X_{j,n}$, étant donné que $X_{i,n} = x_i$ comme

$$\begin{aligned} f_{Xj,n}(x_j|X_{i,n} = x_i) &= \frac{f_{Xi,n},\ldots,_{X_{j,n}}(x_i, x_j)}{f_{Xi,n}(x_i)} \\ &= \frac{(n-i)!}{(j-i-1)!(n-j)!}\left\{\frac{F(x_j)-F(x_i)}{1-F(x_i)}\right\}^{j-i-1} \times \left\{\frac{1-F(x_j)}{1-F(x_i)}\right\}^{n-j} \frac{f(x_j)}{1-F(x_i)} \end{aligned}$$

Ici $i < j \leq n$ et $x_i \leq x_j < \infty$. Le résultat suit facilement en réalisant que $\frac{\{F(x_j)-F(x_i)\}}{\{1-F(x_i)\}}$ et$\frac{f(x_j)}{\{1-F(x_i)\}}$ sont le cd f et la fonction de densité de la population dont la distribution est obtenue en tronquant la distribution $F(x)$ à gauche en x_i. ∎

Théorème 2.1.4 :

Soit $X_1, ..., X_n$un échantillon aléatoire d'une population absolument continue avec $F(x)$ et la fonction de densité $f(x)$, et soit $X_{1,n} \leq ... \leq X_{n,n}$ désignent les statistiques d'ordre obtenues à partir de cet échantillon. Alors la distribution conditionnelle de $X_{i,n}$, étant donné que $X_{j,n} = x_j$ pour $j > i$, est la même que la distribution de la statistique du ième ordre dans un échantillon de taille $j-1$ d'une population dont la distribution est simplement $F(x)$ tronquée sur la droite à x_j.

Preuve.

De la fonction de densité marginale de $X_{i,n}$ et de la fonction de densité conjointe de $X_{i,n}$ et $X_{j,n}$, nous avons la fonction de densité conditionnelle de $X_{i,n}$, étant donné que $X_{j,n} = x_j$, comme

$$\begin{aligned} f_{X_{i,n}}(x_i|X_{j,n} = x_j) &= \frac{f_{X_{i,n}X_{j,n}}(x_i, x_j)}{f_{X_{j,n}}(x_j)} \\ &= \frac{(j-i)!}{(j-i-1)!(i-1)!}\left\{\frac{F(x_i)}{F(x_j)}\right\}^{i-1} \times \left\{\frac{F(x_j)-F(x_i)}{F(x_j)}\right\}^{j-i-1} \frac{f(x_i)}{F(x_j)}. \end{aligned}$$

■

Ici $1 \leq i < j$ et $-\infty < x_i \leq x_j$. La preuve est complétée en notant que $\frac{F(x_i)}{F(x_j)}$ et $\frac{f(x_i)}{F(x_j)}$ sont les fonctions $F(x)$ et densité de la population dont la distribution est obtenue en tronquant la distribution $F(x)$ à droite en x_j.

2.1.10 Espacements

Propriétés des espacements uniformes

Soit $U_1, ..., U_n$ une va's iid uniformes sur $[0,1]$, avec les so associées $U_{1,n} \leq ... \leq U_{n,n}$. les statistiqes S_i définies par

$$S_i = U_{i,n} - U_{i-1,n},\ 1 \leq i \leq n+1,$$

où par convention $U_{0,n} = 0$, $U_{n+1,n} = 1$, s'appellent les espacements uniformes cet échantillon.

Théorème 2.1.5 :

1. *Les espacements $S_1, ..., S_n$ sont uniformément distribuées sur le simplex*

$$A_n = \left\{ (x_1, ..., x_n) : x_i \geq 0, \sum_{i=1}^{n} x_i \leq 1 \right\}.$$

2. *Soit $E_1, ..., E_{n+1}$ une suit des va's iid exponentielle de paramétre* 1.

$$\{S_1, ..., S_{n+1}\} \stackrel{d}{=} \left\{ \frac{E_1}{\sum_{i=1}^{n+1} E_i}, ..., \frac{E_{n+1}}{\sum_{i=1}^{n+1} E_i} \right\}.$$

2.1.11 Simulation des statistiques d'ordres

Dans cette partie, nous discuterons des quelques méthodes de simulation des statistiques d'ordre à partir d'une distribution $F(x)$. Tout d'abord, il convient de mentionner qu'une manière simple de simuler l'ordre la statistique consiste à générer un échantillon pseudo-aléatoire à partir de la distribution $F(x)$ puis à trier l'échantillon grâce à un algorithme efficace comme le tri rapide. Cette méthode générale (étant peuvent être évités dans de nombreux cas en utilisant certains des propriétés de distribution à établir maintenant).

Par exemple, si nous souhaitons générer l'échantillon complet $x_{1,1}, ..., x_{n,n}$ ou même un type échantillon censuré iI $x_{1,n}, ..., x_{r,n}$ de la distribution exponentielle standard. Cela peut être fait simplement en générant un échantillon pseudo-aléatoire $y_1, ..., y_r$ à partir de l'exponentielle standard la distribution d'abord, puis la définition

$$x_{i,n} = \sum_{j=1}^{i} \frac{y_j}{n-j+1}; i = 1, ..., r.$$

La raison en est la suivante :

Théorème 2.1.6 :

Soit $X_{1,n} \leq ... \leq X_{n,n}$ les statistiques d'ordre de l'exposant standard distribution partielle. Ensuite les variables aléatoires $Z_1, ..., Z_n$, où

$$Z_i = (n-i+1)(X_i - X_{i-1}),$$

avec $X_{0,n} \equiv 0$, sont statistiquement indépendants et ont également des distributions exponentielles standard.

Preuve.

Notez que la fonction de densité conjointe de $X_1, ..., X_n$ est

$$f_{1,2,...,n,n}(x_1, ..., x_n) = n!exp(-\sum_{i=1}^{n} x_i),\ 0 \leq x_1 < ... < x_n < \infty.$$

Considérons maintenant la transformation

$$Z_1 = nX_{1,n}, Z_2 = (n-1)(X_{2,n} - X_{1,n}), ..., Z_n = X_{n,n} - X_{n-1,n},$$

ou la transformation équivalente

$$X_{1,n} = \frac{Z_1}{n}, X_{2,n} = \frac{Z_1}{n} + \frac{Z_2}{n-1}, ..., X_{n,n} = \frac{Z_1}{n} + \frac{Z_2}{n-1}, ..., +Z_n.$$

Après avoir noté le jacobien de cette transformation, c'est $\frac{1}{n!}$ et que $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} z_i$, nous obtenir immédiatement la fonction de densité conjointe de $Z_1, ..., Z_n$ est

$$f_{Z_1,...,Z_n}(z_1, ..., z_n) = exp(-\sum_{i=1}^{n} z_i), 0 \leq z_1, ..., z_n < \infty.$$

Si nous souhaitons générer des statistiques d'ordre à partir de la distribution uniforme $(0, 1)$, nous pouvons utilisez les deux théorèmes suivants et évitez de trier à nouveau. Par exemple, si nous avons seulement besoin la statistique du *ième* ordre $u_{i,n}$, elle peut simplement être générée comme une observation pseudo-aléatoire distribution $B(i, n-i+1)$. ■

Théorème 2.1.7 :

Pour la distribution uniforme $(0, 1)$, *les variables aléatoires* $V_1 = \frac{U_i}{Uj}$ *et* $V_2 = Uj$, $1 \leq i < j \leq n$, *sont statistiquement indépendants,* V_1 *et* V_2 *ayant* $B(i, j-i)$ *et les distributions* $B(j, n-j+1)$, *respectivement.*

Preuve.

La fonction de densité conjointe de $U_{i,n}$ et $U_{j,n}(1 \leq i < j \leq n)$ être

$$f_{i,j,n}(u_i, u_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} u_i^{i-1}(u_j - u_i)^{j-i-1}(1-u_j)^{n-j},\ 0 < u_i < u_j < 1.$$

Maintenant, lors de la transformation $V_1 = \frac{Ui,n}{Uj,n}$ et $V_2 = U_{j,n}$ et en notant que le Jacobien de cette transformation est v_2, nous dérivons la fonction de densité conjointe de V_1 et V_2 pour être

$$\begin{aligned} f_{V_1,V_2}(v_1, v_2) &= \frac{(j-1)!}{(i-1)!(j-i-1)!} v_i^{i-1}(1-v_1)^{j-i-1} \\ &= \frac{n!}{(j-1)!(n-j)!} v_2^{j-1}(1-v_2)^{n-j},\ 0 < v_1 < 1,\ 0 < v_2 < 1. \end{aligned}$$

D'après l'équation ci-dessus, il est clair que les variables aléatoires $v_1$et v_2 sont statistiquement indépendants, et aussi qu'ils sont distribués en $B(i, j-i)$ et $B(j, n-j+1)$, respectivement. ∎

Théorème 2.1.8 :

Pour la distribution uniforme $(0, 1)$, *les variables aléatoires*

$$V_1^* = \frac{U_{1,n}}{U_{2,n}}; V_2^* = \left(\frac{U_{2,n}}{U_{3,n}}\right)^2; ...; V_{n-1}^* = \left(\frac{U_{n-1,n}}{U_{n,n}}\right)^{n-1}.$$

et $V_n^* = U_{n,n}^n$ *sont toutes des variables aléatoires uniformes* $(0, 1)$ *indépendantes.*

Preuve.

Soit $X_{1,n} \leq ... \leq X_{n,n}$ désignent les statistiques d'ordre de l'expo standard distribution inductive.

Ensuite, en utilisant les faits que $X = -logU$ a une norme distribution exponentielle et que $-logu$ est une fonction décroissante monotone dans u, nousavoir immédiatement $X_i \overset{d}{=} -logU_{n-i+1,n}$.

L'équation ci-dessus donne

$$V_i^* = \left(\frac{U_{i,n}}{U_{i+1,n}}\right) \stackrel{d}{=} \left(\frac{e^{-X_{n-i+1,n}}}{e^{-X_{n-i,n}}}\right) = exp[-i(X_{n-i+1,n} - X_{n-i,n})] \stackrel{d}{=} exp(-Y_{n-i+1,n}),$$

en utilisant le théorème ci-dessus, où Y_i sont des variables aléatoires exponentielles standard indépendant esaptes. Les méthodes de simulation d'un ordre uniforme qui viennent d'être décrites peuvent également être utilisées pour générer facilement des statistiques d'ordre à partir de n'importe quelle distribution connue $F(x)$ pour laquelle $F^{-1}(.)$ est relativement facile à calculer. On peut simplement obtenir les statistiques d'ordre $x_{1,n}, ..., x_{n,n}$ à partir du distribution requise $F(.)$ en fixant $x_{i,n} = F^{-1}(u_{i,n})$. $\blacksquare$

Chapitre 3

Simulation des statistiques d'ordres à l'aide du R

Dans ce chapitre nous allons utiliser trois méthodes à l'aide du **R,** pour générer un échantillon ordonné. Soit $(U_1, ..., U_n)$ un échantillon suit la loi uniforme sur $[1, 0]$ de fonction de répartition $F(x) = x$, $x \in [1, 0]$.

3.1 Méthode de la fonction de répartition empirique

On peut représenté la fontion de répartition empirique en terme des statistique d'ordre comme suit :

$$F_n(x) = \frac{1}{n}\sum_{j=1}^{n} 1_{\{X_{j,n} \leq x\}} = \begin{cases} 0 & \text{si } x < X_{1,n}, \\ \frac{j-1}{n} & \text{si } \ X_{j-1,n} \leq x < X_{j,n}, 2 \leq j \leq n, \\ 1 & \text{si } \ x > X_{n,n}. \end{cases}$$

Par conséquent les statistiques d'ordre sont la solution de cette équation

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

le code suivant permet de résodre cette équation, afin de simuler un échantillon ordonné issue d'un échantillon uniforme standard.

$N = 10$

$XO = nu\,\mathrm{m}\,eric(N)$

$U = runif(N)$

$FN = function(x, U)\,\{$

$Y = nu\,\mathrm{m}\,eric(N)$

$for(j\ in\ 1 : N)\,\{$

$if(U\,[j] <= x)Y\,[j] = 1$

$else$

$Y\,[j] = 0\ \}$

$mean(Y)\ \}$

$for(j\ in\ 1 : N)\,\{$

$H = function(x)\,\{FN\}\,(x, U) - (j/N)\ \}$

$X = uniroot(H,\ lower = 0,\ upper = 1)$

$XO\,[j] = X\$root\ \}$.

3.2 Méthode dircte

Soit $(U_1, ..., U_n)$ un échantillon suit la loi uniforme sur $[1, 0]$ et $(U_{1,n}, ..., U_{n,n})$ la statistique d'ordre associée.

le code suiant nous donne une suite ordonnée de taille $N = 200$.

$N = 200$

$U = runif(N)$

$UO = sort$

3.3 Représentation de Reny

soit $(U_1, ..., U_n)$ un échantillon suit la loi uniforme sur $[1, 0]$et $(U_{1,n}, ..., U_{n,n})$ la statistique d'ordre associée, et soit $E_1, ..., E_n$ un échantillon de taille $n + 1$ suit la loi exponentiellede paramétre 1, alors :

$$U_{i,n} \stackrel{d}{=} \frac{E_1 + ... + E_i}{E_i + ... + E_{n+1}}, \; i = 1, ..., n.$$

$N = 200$

$U = runif(N)$

$UO = sort(U)$

$US = nu\,\mathrm{m}\,eric(N)$

$E = rexp(N + 1)$

$S = 0$

$for(i\ in\ 1 : N)\{$

$S = S + E[i]$

$US[i] = S/sum(E)\ \}$

$qqplot(UO, US)$

$ks.test(UO, US)$

L'exécution de ces commende nous donne FIG1, cette figure montre que le quanti je associée à l'échantillon ordonné par la représentation de Reny et le quantile relative à l'échantillon ordonné par la méthode directe sont alignés.

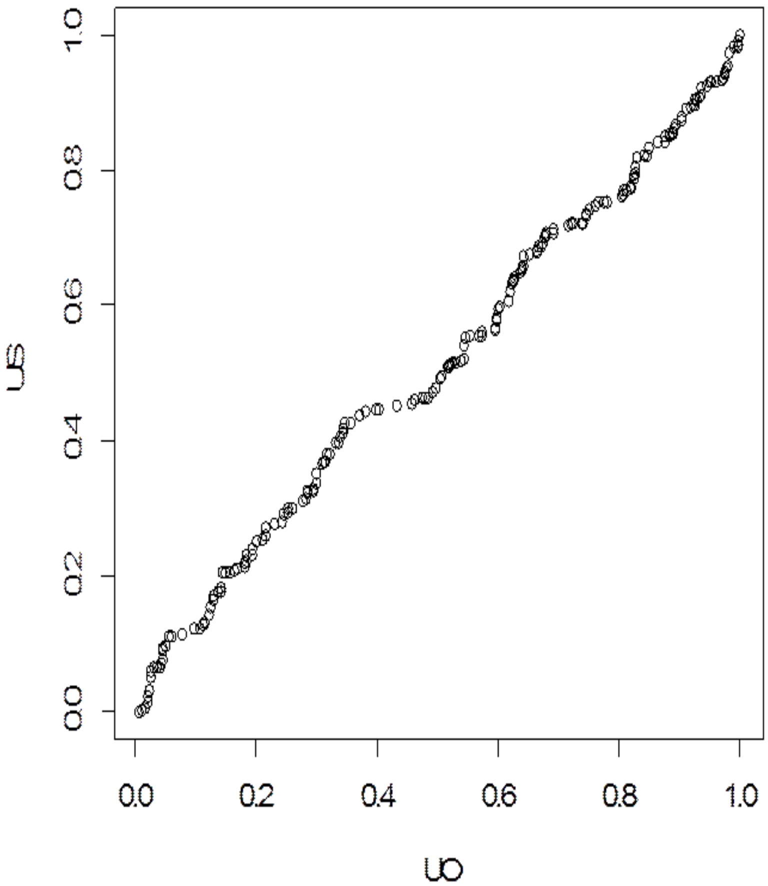

FIG1−Comparaison des statistique d'ordre entre la méthode directe et représentation de Reny

Chapitre 4

Conclusion

Dans ce travail, nous avons appris tous les concepts des statistiques d'ordre et les avons simulés.

Nous nous basons sur trois chapitres

Tout d'abord, nous avons identifié les scientifiques les plus importants qui ont joué un rôle majeur dans son développement, puis nous avons brièvement énuméré les paramètres dans lesquels les statistiques d'ordre pourraient jouer un rôle important.

Deuxièmement, nous avons identifié tous les concepts et informations de base dans les statistiques d'ordre, ainsi que des simulations à travers des théories appliquées.

Et le troisième est le côté application R, et nous l'avons utilisé de trois méthodes différentes R.

Bibliographie

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

[2] Gut, A. Intermediate course in probability [electronic resource]. Springer.

[3] Capéraà, P., & Van Cutsem, B. (1988). Méthodes et modèles en statistique non paramétrique : exposé fondamental (Vol. 1). Presses Université Laval.

[4] J.F.Delmas, B.Jourdain (2006). Modéles aléatoires : applications aux sciences de l'ingénieur et du vvant. Springer, pages 302-341

[5] Statistiques d'ordre https ://www.math.u-bordeaux.fr/~mchabano/Agreg/ProbaAgreg1213-COURS1-Statordre.pdf

[6] Samah, B. (2010). Mémoire determination du nombre de statistique d'ordre extremes.Biskra.

ملخص

احصاء الترتيب تعد من بين الأدوات الأساسية في الإحصائيات والاستدلالات. تتمثل الحالات الخاصة المهمة لإحصائيات الترتيب في الحد الأدنى والحد الأقصى لقيمة العينة. إن الهدف من هذه الأطروحة هو تقديم احصاء الترتيب بالإضافة الى خصائصها كما نعطي محاكاة لها و تطبيقاتها.

<u>الكلمات المفتاحية:</u> احصاء الترتيب, محاكاة.

RÉSUMÉ

La statistique d'ordre est l'un des principaux outils de la statistique inférentielle. Les cas particuliers des statistiques d'ordres sont les valeurs minimale et maximale d'échantillon. Le but de ce mémoire est de faire une synthèse sur le sujet en mettant en évidence les outils de bases de ce domaine. En outre nous avons enrichi notre travail par des simulations et des applications.

<u>Mots clés</u> : Statistiques d'ordre, simulations.

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. The aim of this memory is to summarize the subject by highlighting the basic tools in this field. In addition, we have enriched our work with simulations and applications.

<u>Key words:</u> Order statistics, simulations.