École Normale Supérieure
FIMFA — Statistique 2016–2017

DEVOIR I POUR LE 7 OCTOBRE

Les statistiques d'ordre $X_{1:n} \leq X_{2:n} \leq X_{n:n}$ d'un n-échantillon $X_1, \ldots, X_n$ d'observations indépendantes identiquement distribuées sont formées par le réarrangement croissant (convention) de l'échantillon. Quand n est clair d'après le contexte on peut les noter $X_{(1)} \leq \ldots \leq X_{(n)}$.

Exercice 1
Vérifier que la loi jointe des statistiques d'ordre est absolument continue par rapport à la loi de l'échantillon.
On suppose que X est une variable aléatoire réelle, absolument continue de densité continue. Montrer que l'échantillon est presque sûrement formé de valeurs deux à deux distinctes. Donner la densité de la loi jointe des statistiques d'ordre.

Exercice 2
Si la loi des X_i définie par sa fonction de répartition F, admet une densité f, quelle est la densité de la loi de $X_{k:n}$ pour $1 \leq k \leq n$?

Exercice 3
Montrer que conditionnellement à $X_{k:n} = x$, la suite

$$(X_{i:n} - X_{k:n})_{i=k+1,\ldots,n}$$

est distribuée comme les statistiques d'ordre d'un $n-k$ échantillon de la loi d'excès au dessus de x (fonction de survie $\overline{F}(x+\cdot)/\overline{F}(x)$) avec la convention $\overline{F} = 1 - F$).

Exercice 4
Si $X_1, \ldots, X_n$ est un échantillon i.i.d. de la loi exponentielle d'espérance 1 (densité $\mathbb{I}_{x>0}\mathrm{e}^{-x}$), et $X_{n:n} \geq X_{n-1:n} \geq \ldots \geq X_{1:n}$ les statistiques d'ordre associées, montrer que :

— avec la convention $X_{0:n} = 0$, les écarts $(X_{i:n} - X_{i-1:n})_{1 \leq i \leq n}$ (*spacings*) forment une collection de variables aléatoires indépendantes ;

— $X_{i:n} - X_{i-1:n}$ est distribuée selon une loi exponentielle d'espérance $\frac{1}{n+1-i}$.

Exercice 5
Dans cet exercice, $X_1, \ldots, X_n$ est un échantillon i.i.d. de la loi exponentielle d'espérance 1 (densité $\mathbb{I}_{x>0}\mathrm{e}^{-x}$), et $X_{n:n} \geq X_{n-1:n} \geq \ldots \geq X_{1:n}$ les statistiques d'ordre associées, et $(k_n)_n$ est une suite croissante d'entiers qui tend vers l'infini, telle que k_n/n tende vers une limite finie (éventuellement nulle).
Montrer que $\frac{X_{k_n:n} - \mathbb{E}X_{k_n:n}}{\sqrt{\mathrm{var}(X_{k_n:n})}}$ converge en loi vers une Gaussienne centrée réduite.

Suggestion : pensez à utiliser la version Lindeberg-Feller du TCL.

Exercice 6
Si f est une fonction croissante sur $[a, b]$, on définit son inverse généralisée $f^{\leftarrow}$ par

$$f^{\leftarrow}(y) = \inf\{x : x \in [a, b],\ f(x) \geq y\} \qquad \text{pour } y \in [f(a), f(b)]\,.$$

Montrer que si $(f_n)_n$ est une suite de fonctions croissantes sur $[a, b] \subset \mathbb{R}$ qui converge *simplement* vers f une autre fonction croissante sur $[a, b]$ en tout point de continuité de f, alors la suite $(f_n^{\leftarrow}(y))$ converge simplement vers $f^{\leftarrow}(y)$ en tout $y \in [f(a), f(b)]$ où $f^{\leftarrow}$ est continue.

Exercice 7
Soit F une fonction de répartition, on définit la fonction quantile associée $F^{\leftarrow}$ par

$$F^{\leftarrow}(p) = \inf\{x : F(x) \geq p\} \qquad \text{pour } p \in]0, 1[\,.$$

Vérifier que si deux fonctions de répartitions ont même fonction quantiles alors elles sont égales.
Dans cette définition F est une fonction de répartition, $F^{\leftarrow}$ la fonction quantile associée.
Montrer que pour tout $x \in \mathbb{R}, p \in]0, 1[$,

1. $F^{\leftarrow}(p) \leq x \Leftrightarrow p \leq F(x)$.
2. $F \circ F^{\leftarrow}(p) \geq p$ avec égalité si et seulement si il existe x tel que $F(x) = p$.
   Si $F \circ F^{\leftarrow}(p) > p$ alors $F^{\leftarrow}$ est discontinue en p.
3. $F^{\leftarrow} \circ F(x) \leq x$.
   Si $F^{\leftarrow} \circ F(x) < x$, alors il existe $\epsilon > 0$ tel que $F(x - \epsilon) = F(x)$.
4. $(F \circ G)^{\leftarrow} = G^{\leftarrow} \circ F^{\leftarrow}$

Comment la convergence en distribution se traduit elle sur les fonctions quantiles ?

Exercice 8
Si $Y_{1:n} \leq Y_{2:n} \leq \ldots \leq Y_{n:n}$ forme les statistiques d'ordre d'un n-échantillon de la loi exponentielle d'espérance 1, et si $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ désigne les statistiques d'ordre d'un n-échantillon d'une loi de fonction de répartition F qui admet une densité partout positive, montrer que

$$(X_{1:n}, X_{2:n}, \ldots, X_{n:n}) \sim (F^{\leftarrow}(1 - \exp(-Y_{1:n})), \ldots, F^{\leftarrow}(1 - \exp(-Y_{n:n})))\ .$$

Exercice 9
La fonction quantile empirique $F_n^{\leftarrow}$ est la fonction quantile associée à la fonction de répartition empirique F_n. Si les points de l'échantillon sont deux à deux distincts

$$F_n^{\leftarrow}(p) = X_{k:n} \qquad \text{pour } \frac{k-1}{n} < p \leq \frac{k}{n}\,.$$

Soit F une fonction de répartition qui est dérivable en $F^{\leftarrow}(p)$ de dérivée non nulle notée $f(p)$ pour une valeur $p \in]0, 1[$. Montrer que

$$\sqrt{n}\,(F_n^{\leftarrow}(p) - F^{\leftarrow}(p)) + \sqrt{n}\frac{1}{f(p)}\,(F_n(F^{\leftarrow}(p)) - p) = o_P(1)\,.$$

Quelle est la loi limite de $\sqrt{n}\,(F_n^{\leftarrow}(p) - F^{\leftarrow}(p))$?