Sujet 1 : Probabilités et Statistiques

Préambule

Ce sujet est consacré à l'étude du comportement asymptotique d'extrêmes d'une suite $(X_k)_{k\geqslant 1}$ de variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$. La première partie présente des résultats obtenus par des techniques élémentaires sans aucune hypothèse sur la structure de dépendance des X_k. La deuxième partie traite du comportement asymptotique du maximum M_n du n-échantillon $X_1, \dots, X_n$ lorsque les X_k sont indépendantes et de même loi. La troisième partie est consacrée aux records de l'échantillon $X_1, \dots, X_n$ et aboutit à deux théorèmes limite pour le nombre S_n de ces records. La quatrième partie propose d'établir un théorème de caractérisation de la loi de X_1 par la suite des $\mathbf{E}\min(X_1, \dots, X_n)$.

Les 4 parties du problème sont indépendantes, mais il est conseillé d'essayer de les traiter dans l'ordre.

On précise maintenant quelques notations, rappels ou résultats que les candidats pourront considérer comme connus.

Définition 1. *Soient $(u_n)_{n\geqslant 1}$ une suite de réels strictement positifs à partir d'un certain rang et $(Y_n)_{n\geqslant 1}$ une suite de variables aléatoires réelles. On dit que*

a) $Y_n = O_P(u_n)$ si pour tout $\varepsilon \in]0, 1[$ il existe un entier n_0 et une constante C tels que pour tout $n \geqslant n_0$, $P(|Y_n| \leqslant C u_n) \geqslant 1 - \varepsilon$;

b) $Y_n = o_P(u_n)$ si $u_n^{-1}|Y_n|$ converge en probabilité vers 0.

Soit $(A_n)_{n\geqslant 1}$ une suite d'évènements sur le même espace probabilisable, chaque A_n étant défini par une propriété π_n. On notera $\{\pi_n \text{ i.s.}\}$ (lire π_n se réalise infiniment souvent) l'évènement $\limsup A_n$, autrement dit l'ensemble des évènements élémentaires ω pour lesquels $I(\omega) = \{n \in \mathbb{N} ; \omega \in A_n\} = \{n \in \mathbb{N} ; \pi_n(\omega) \text{ vraie}\}$ est une partie infinie de $\mathbb{N}$.

Soit $F : t \mapsto P(X \leqslant t)$ la fonction de répartition d'une variable aléatoire réelle X. On définit son inverse généralisé F^{-1} ou fonction quantile par

$$F^{-1} :]0, 1[\to \mathbb{R}, \quad u \mapsto F^{-1}(u) = \inf\{t \in \mathbb{R} ; F(t) \geqslant u\}. \tag{1}$$

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

Le lemme suivant pourra être utilisé sans le redémontrer dans des questions du type loi des grands nombres.

Lemme 2 (Kronecker)**.** *Soient $(b_k)_{k\geqslant 1}$ une suite de réels strictement positifs, croissante et de limite $+\infty$ et $(x_k)_{k\geqslant 1}$ une suite de réels, telles que la série*

$$\sum_{k=1}^{+\infty} \frac{x_k}{b_k} \text{ converge dans } \mathbb{R}.$$

Alors

$$\frac{1}{b_n}\sum_{k=1}^{n} x_k \xrightarrow[n\to+\infty]{} 0.$$

On rappelle enfin le théorème limite central de Liapounov, qui peut être vu comme un corollaire du théorème limite central de Lindeberg-Lévy.

Théorème 3 (Liapounov)**.** *Notons pour tout entier $n \geqslant 1$, $S_n = \sum_{k=1}^{n} Y_k$, où $(Y_k)_{k\geqslant 1}$ est une suite de variables aléatoires indépendantes (mais pas forcément de même loi) ayant toutes un moment absolu d'ordre 3 fini et telles que*

$$\frac{1}{(\operatorname{Var} S_n)^{3/2}}\sum_{k=1}^{n} \mathbf{E}|Y_k - \mathbf{E}Y_k|^3 \xrightarrow[n\to+\infty]{} 0.$$

Alors

$$\frac{S_n - \mathbf{E}S_n}{\sqrt{\operatorname{Var} S_n}} \xrightarrow[n\to+\infty]{\text{loi}} Z,$$

où Z est une variable aléatoire de loi gaussienne standard $\mathfrak{N}(0,1)$.

1 Généralités élémentaires

On définit la variable aléatoire $M_n = \max_{1\leqslant i\leqslant n} X_i$ par

$$\forall \omega \in \Omega, \quad M_n(\omega) = \max_{1\leqslant i\leqslant n} X_i(\omega).$$

1.1 Montrer que pour tout réel t,

$$P(M_n > t) \leqslant \sum_{i=1}^{n} P(X_i > t). \tag{2}$$

Si les X_i ont même loi, cette inégalité devient

$$P(M_n > t) \leqslant nP(X_1 > t). \tag{3}$$

1.2 En déduire que si les X_i ont même loi avec un moment absolu d'ordre $p > 0$ fini, $\max_{1 \leqslant i \leqslant n} |X_i| = O_P(n^{1/p})$.

1.3 Montrer qu'en fait, sous les mêmes hypothèses, $\max_{1 \leqslant i \leqslant n} |X_i| = o_P(n^{1/p})$.

1.4 On suppose dans cette question que les X_i sont des variables aléatoires positives de même loi exponentielle de paramètre $a > 0$. Leur fonction de survie G est donc définie par $G(t) = P(X_i > t) = \exp(-at)$ pour tout $t \geqslant 0$. Montrer que $M_n = O_P(\ln n)$.

1.5 Soit X une variable aléatoire gaussienne de loi $\mathfrak{N}(0,1)$. Montrer que

$$\forall t > 0, \quad P(|X| \geqslant t) \leqslant \exp\left(\frac{-t^2}{2}\right). \tag{4}$$

Indication : on pourra partir de $P(X \geqslant t) = \int_t^{+\infty} (2\pi)^{-1/2} \exp(-x^2/2)\,\mathrm{d}x$.

1.6 Soit $(X_k)_{k \geqslant 1}$ une suite de variables aléatoires gaussiennes de même loi $\mathfrak{N}(0,1)$. Utiliser (4) pour majorer $P(\max_{k \leqslant n} |X_k| \geqslant t)$ pour $t > 0$ et en déduire que :

$$\forall s > \sqrt{2}, \quad P\Big(\max_{k \leqslant n} |X_k| \geqslant s(\ln n)^{1/2}\Big) \xrightarrow[n \to +\infty]{} 0. \tag{5}$$

2 Comportement asymptotique de M_n dans le cas i.i.d.

Désormais on suppose, sauf mention explicite du contraire, que les X_i sont indépendantes et de même loi de fonction de répartition F.

2.1 Montrer que

$$\forall t \in \mathbb{R}, \quad P(M_n \leqslant t) = F(t)^n.$$

2.2 On suppose de plus que $F(t) < 1$ pour tout t réel.

2.2.1 Montrer que

$$\forall t \in \mathbb{R}, \quad P(M_n > t) \xrightarrow[n \to +\infty]{} 1, \tag{6}$$

autrement dit que M_n tend vers $+\infty$ en probabilité.

2.2.2 Donner un exemple simple montrant que sans l'hypothèse d'indépendance des X_i, (6) peut être fausse.

2.2.3 Montrer, sans utiliser l'indépendance, que si (6) est vraie, alors M_n converge presque sûrement vers $+\infty$.

2.3 On se propose maintenant de minorer $P(M_n > t)$.

2.3.1 Montrer que pour tout $\delta > 0$ et tous $n \geqslant 1$ et $t \in \mathbb{R}$ vérifiant : $nP(X_1 > t) \leqslant \delta$,

$$P(M_n > t) \geqslant \frac{1 - \mathrm{e}^{-\delta}}{\delta} nP(X_1 > t). \tag{7}$$

2.3.2 Si de plus $\delta \leqslant 1$, on a aussi sous les mêmes hypothèses,

$$P(M_n > t) \geqslant \Big(1 - \frac{\delta}{2}\Big) nP(X_1 > t). \tag{8}$$

2.3.3 En quoi (7) permet-elle d'affirmer que l'inégalité (3) est optimale ?

2.4 On suppose qu'il existe une suite de réels $(a_n)_{n\geqslant 1}$ et une suite de réels positifs $(b_n)_{n\geqslant 1}$ telles que

$$\forall t \in \mathbb{R}, \quad \lim_{n\to+\infty} nP(X_1 > a_n + b_n t) = u(t), \tag{9}$$

où u est une fonction définie sur $\mathbb{R}$ et à valeurs dans $\overline{\mathbb{R}}_+ = [0, +\infty]$. Montrer qu'alors

$$\forall t \in \mathbb{R}, \quad \lim_{n\to+\infty} P(M_n \leqslant a_n + b_n t) = \exp(-u(t)), \tag{10}$$

avec la convention d'écriture $\exp(-\infty) = 0$.

2.5 Si (9) est vérifiée et si les b_n sont strictement positifs, quelles conditions supplémentaires sur la fonction u permettent d'en déduire la convergence en loi de $b_n^{-1}(M_n - a_n)$?

2.6 Utiliser ce qui précède pour étudier brièvement la convergence en loi de $b_n^{-1}(M_n - a_n)$ pour chacun des exemples suivants.

2.6.1 X_1 suit la loi exponentielle de paramètre 1, $a_n = \ln n$, $b_n = 1$.

2.6.2 X_1 suit la loi uniforme sur $[0, 1]$, $a_n = 1$, $b_n = 1/n$.

2.6.3 X_1 suit une loi de Pareto de paramètre $p > 0$ donnée par $P(X_1 > t) = t^{-p}$ pour tout $t \geqslant 1$, $a_n = 0$, $b_n = n^{1/p}$.

2.7 Dans cette question, on suppose encore que les X_k sont à valeurs réelles (en particulier $\forall \omega \in \Omega$, $X_k(\omega) < +\infty$), mais on ne suppose plus les X_k indépendantes ni de même loi. On note F_k la fonction de répartition de X_k. Soit $(c_n)_{n\geqslant 1}$ une suite croissante de réels tendant vers $+\infty$. Montrer l'égalité d'évènements :

$$\{M_n > c_n \text{ i.s.}\} = \{X_n > c_n \text{ i.s.}\}. \tag{11}$$

En déduire que si

$$\sum_{k=1}^{+\infty} \big(1 - F_k(c_k)\big) < +\infty, \tag{12}$$

alors $P(M_n > c_n \text{ i.s.}) = 0$.

2.8 On revient au cas où les X_k sont indépendantes et de même loi de fonction de répartition F. On suppose de plus que F est *continue* sur $\mathbb{R}$ et que $F(t) < 1$ pour tout $t \in \mathbb{R}$. On choisit pour c_n le $(1 - 1/n)$-quantile de F, c'est-à-dire

$$c_n = \inf\{t \in \mathbb{R} \ ; \ F(t) \geqslant 1 - 1/n\}. \tag{13}$$

Montrer qu'alors $P(M_n > c_n \text{ i.s.}) = 1$. *Indications.* Montrer que c_n tend vers l'infini et calculer $F(c_n)$.

3 Records

Dans cette partie on suppose que les X_i sont indépendantes et de même loi de fonction de répartition *continue* sur $\mathbb{R}$. On se propose d'étudier le comportement asymptotique du nombre de *records* dans l'échantillon $X_1, \ldots, X_n$ quand n tend vers l'infini. Les deux premières questions permettent de voir que presque sûrement il n'y a pas d'ex aequo dans l'échantillon. On note R_n le rang de X_n parmi les X_i d'indice $i \leqslant n$. Ainsi

$$R_n = 1 + \sum_{k=1}^{n-1} \mathbf{1}_{\{X_k > X_n\}}. \tag{14}$$

Le nombre S_n de *records* dans l'échantillon $X_1, \ldots, X_n$ est alors donné par

$$S_n = \sum_{j=1}^{n} \mathbf{1}_{\{R_j = 1\}}. \tag{15}$$

La suite des *temps de records* $(T_n)_{n\geqslant 1}$ est définie par récurrence par

$$T_1 = 1 \text{ et pour } n \geqslant 1, \quad T_{n+1} = \inf\{j > T_n \ ; \ \forall i < j, X_i < X_j\}, \tag{16}$$

avec la convention $\inf \varnothing = +\infty$.

3.1 Montrer que si les variables aléatoires réelles X et Y sont indépendantes et si la fonction de répartition de X est continue sur $\mathbb{R}$, alors $P(X = Y) = 0$.

3.2 En déduire que presque sûrement la suite $(X_n)_{n\geqslant 1}$ ne présente pas de valeurs ex aequo.

3.3 Montrer que si la variable aléatoire réelle X a une fonction de répartition F continue sur $\mathbb{R}$, la variable aléatoire $F(X)$ suit la loi uniforme sur $]0, 1[$.

3.4 Pour tout entier $i \geqslant 1$, on note $U_i = F(X_i)$, F étant la fonction de répartition de chacune des X_i. Pour $n > 1$, soit $i_1, i_2, \ldots, i_n$ une permutation des entiers de $[\![1, n]\!] = [1, n] \cap \mathbb{N}$. Montrer en utilisant F et son inverse généralisé F^{-1} que

$$P(X_{i_1} < X_{i_2} < \cdots < X_{i_n}) = P(U_{i_1} < U_{i_2} < \cdots < U_{i_n}). \tag{17}$$

3.5 En déduire que

3.5.1 le vecteur aléatoire discret $(R_1, \ldots, R_n)$ suit la loi uniforme sur

$$D_n = \{1\} \times [\![1, 2]\!] \times [\![1, 3]\!] \times \cdots \times [\![1, n]\!];$$

3.5.2 les R_i sont indépendantes ;

3.5.3 pour tout $n \geqslant 1$, R_n suit la loi uniforme sur $[\![1, n]\!]$;

3.5.4 les R_i sont indépendantes ;

3.5.5 ni la loi de la suite $(S_n)_{n\geqslant 1}$ ni celle de $(T_n)_{n\geqslant 1}$ ne dépendent de la loi de X_1, tant que cette dernière est à fonction de répartition continue.

3.6 Déduire de la question précédente que pour tout *entier* $k \geqslant 2$,

$$P(T_2 = k) = \frac{1}{(k-1)k}. \tag{18}$$

Vérifier alors que $P(T_2 = +\infty) = 0$ et $\mathbf{E}T_2 = +\infty$.

3.7 Grâce à la question 3.5.3, on voit que

$$\mathbf{E}S_n = \ln n + \gamma + \varepsilon(n), \quad \lim_{n\to+\infty} \varepsilon(n) = 0, \tag{19}$$

où γ désigne la constante d'Euler. Établir un résultat analogue pour $\operatorname{Var} S_n$.

3.8 Montrer que

$$\sum_{k=2}^{+\infty} \frac{\mathbf{1}_{\{R_k=1\}} - P(R_k = 1)}{\ln k} \quad \text{converge p.s.} \tag{20}$$

En déduire que

$$\frac{S_n}{\ln n} \xrightarrow[n\to+\infty]{\text{p.s.}} 1. \tag{21}$$

3.9 Montrer que

$$\frac{S_n - \ln n}{\sqrt{\ln n}} \xrightarrow[n\to+\infty]{\text{loi}} \mathfrak{N}(0,1), \tag{22}$$

où $\mathfrak{N}(0,1)$ désigne la loi gaussienne standard.

4 Une caractérisation extrémale de la loi des X_i

Le but de cette partie est de prouver le théorème suivant.

Théorème 4. *Soient $(X_k)_{k\geqslant 1}$ et $(Y_k)_{k\geqslant 1}$ deux suites de variables aléatoires positives, indépendantes et de même loi. On suppose que $\mathbf{E}X_1 < +\infty$ et que*

$$\forall n \geqslant 1, \quad \mathbf{E}\min(X_1, \ldots, X_n) = \mathbf{E}\min(Y_1, \ldots, Y_n). \tag{23}$$

Alors X_1 et Y_1 ont même loi.

4.1 Soit Z une variable aléatoire positive. Vérifier que

$$\mathbf{E}Z = \int_0^{+\infty} P(Z > t)\,\mathrm{d}t. \tag{24}$$

4.2 Soit $G : t \mapsto P(X_1 > t)$ la fonction de survie de X_1. Calculer en fonction de G la fonction de survie de $\min(X_1, \ldots, X_n)$.

4.3 On note G^{-1} l'inverse généralisé de G défini par $G^{-1}(1) = 0$ et

$$\forall u \in]0,1[, \quad G^{-1}(u) = \sup\{t \in \mathbb{R}_+ \; ; \; G(t) > u\}. \tag{25}$$

Prouver la formule

$$\forall n \geqslant 1, \quad \int_0^{+\infty} G(t)^n \, \mathrm{d}t = \int_0^1 n u^{n-1} G^{-1}(u) \, \mathrm{d}u. \tag{26}$$

En particulier, on déduit du cas $n = 1$ et de l'intégrabilité de X_1 que G^{-1} est une fonction positive intégrable sur $[0,1]$ relativement à la mesure de Lebesgue.

4.4 Expliquer brièvement pourquoi si deux mesures positives bornées μ et ν sur la tribu borélienne de $[0,1]$ vérifient $\int_{[0,1]} h \, \mathrm{d}\mu = \int_{[0,1]} h \, \mathrm{d}\nu$ pour toute fonction h à valeurs réelles définie et continue sur $[0,1]$, alors $\mu = \nu$.

4.5 Démontrer le théorème 4.

4.6 Le théorème reste-t-il vrai si $\mathbf{E}X_1 = \mathbf{E}Y_1 = +\infty$?

Fin du sujet