A. PROBABILITÉS ET STATISTIQUES

Aucun document personnel n'est autorisé

L'usage de calculatrices électroniques de poche à alimentation autonome, non imprimantes et sans document d'accompagnement, est autorisé, une seule à la fois étant admise sur la table ou le poste de travail.

Les parties B, C et D sont indépendantes entre elles ; la partie E peut être traitée indépendamment des autres en admettant le résultat de la question D(1)b

Notations et rappels

Pour toute fonction de la variable réelle f, on note $f(x-)$ la limite à gauche de f au point x, lorsque celle-ci existe.

Pour toute partie A non vide de $\mathbb{R}^d$, muni par exemple de sa norme euclidienne, on note ∂A la frontière de A, $diam(A)$ le diamètre de A (c'est-à-dire la plus grande distance possible entre deux éléments de A) et $\mathbf{1}_A(x)$ la fonction indicatrice de A au point x.

On note $x \vee y$ le plus grand des deux réels x et y, et $x \wedge y$ le plus petit.

On se donne un espace probabilisé $(\Omega, \mathcal{F}, P)$. Une propriété est vraie $P-p.s.$ si elle est vraie pour tout $\omega \in \Omega$ hors d'un ensemble N tel que $P(N) = 0$.

Les variables aléatoires rencontrées dans ce problème sont, sauf mention expresse du contraire, définies sur Ω et à valeur dans la droite réelle $\mathbb{R}$ ou dans un espace $\mathbb{R}^d$; en particulier, elles sont toujours presque sûrement finies. Pour toute variable aléatoire (en abrégé v.a.) X, on note P_X la loi de X. Si X est à valeurs dans $\mathbb{R}$, on note F_X sa fonction de répartition, c'est-à-dire que pour tout $x \in \mathbb{R}$, $F_X(x) = P(X \leq x)$. Quand il n'y aura pas d'ambiguïté, on pourra simplifier l'écriture en posant $F_X = F$. On rappelle que F_X détermine la loi de X : par exemple, si $F_X(x) = (1-e^{-x})\mathbf{1}_{\mathbb{R}^+}(x)$, la loi de X est la loi exponentielle de paramètre 1.

On rappelle que si $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires, X_n converge en loi vers la v.a.

X (ce que l'on note $X_n \xrightarrow{\mathcal{L}} X$) si et seulement si la suite de lois P_{X_n} converge étroitement vers la loi P_X, et que cela est équivalent à la convergence de $F_{X_n}(x)$ vers $F_X(x)$ en tout point x où F_X est continue.

On rappelle aussi que la fonction caractéristique φ d'une variable aléatoire réelle est définie par $\varphi(t) = E(e^{itX})$, $t \in \mathbb{R}$.

Une procédure de test est définie comme suit : on considère une variable aléatoire S de loi a priori inconnue, et on fait une hypothèse H_0 qui détermine la loi de S (par exemple, supposons que S est la somme de n variables indépendantes ; si H_0 dit que ces variables sont toutes de loi de Bernoulli de paramètre $1/2$, H_0 détermine la loi de S comme étant la loi binomiale de paramètres n et $1/2$). Pour $\alpha \in]0,1[$, une région de rejet de H_0 au niveau α est une partie de Ω de la forme $\{S \in E\}$ telle que, si l'hypothèse H_0 est vraie, $P(S \in E) = \alpha$.

A. Dans cette partie, on suppose que X est une v.a. réelle.

1. Montrer que F_X est continue à droite et admet une limite à gauche en tout point.
2. On note $G_X(t) = \inf\{x, F_X(x) \geq t\}$. Montrer que G_X est une application croissante, continue à gauche et admettant en tout point une limite à droite. Montrer l'égalité

$$G_X(t) \leq x \Longleftrightarrow F_X(x) \geq t$$

3. Soit U une v.a. de loi uniforme sur $[0,1]$. Montrer que la v.a. $G_X(U)$ a même loi que X
4. On suppose dans cette question que F_X est continue.
   (a) Montrer que G_X est une inverse à droite de F_X.
   (b) Montrer que $H_X(t) = \sup\{x, F_X(x) = t\}$ définit également une inverse à droite de F_X.
   (c) Montrer que $F_X(x) \leq t \Longleftrightarrow x \leq H_X(t)$, et en déduire la loi de la v.a. $F_X(X)$.

B. Dans cette partie, $X_1, \cdots, X_n$ désignent des v.a. réelles, indépendantes, de même loi et de fonction de répartition F. Pour $x \in \mathbb{R}$ fixé, et pour tout entier $1 \leq i \leq n$, on note $\xi_i = \mathbf{1}_{[X_i \leq x]}$ et

$$F^n(x) = \frac{1}{n}\sum_{i=1}^{n} \xi_i.$$

1. Montrer que $\xi_1, \cdots, \xi_n$ sont des v.a. indépendantes dont on précisera la loi, l'espérance et la variance en fonction de F et x.
2. Répondre à la même question, mais avec les v.a. $\zeta_i = \mathbf{1}_{[X_i < x]}$.
3. Montrer que, pour tout $x \in \mathbb{R}$, $\lim_{n\to+\infty} F^n(x) = F(x)$ $P-p.s.$, et $\lim_{n\to+\infty} F^n(x-) = F(x-)$ $P-p.s.$
4. On se propose à présent de démontrer que la convergence prouvée à la question précédente est en fait uniforme en x ; pour cela, on se fixe un entier positif M, et on définit une subdivision de $\mathbb{R}$ comme suit :

   $x_0 = -\infty$, $x_{M+1} = +\infty$, et pour $1 \leq k \leq M$, $x_k = \inf\{x, F(x) \geq k/M\}$.

(a) Montrer que la proposition :

$$\forall k \leq M, \lim_{n\to+\infty} |F^n(x_k) - F(x_k)| \vee |F^n(x_{k+1}-) - F(x_{k+1}-)| = 0$$

est vraie $P-p.s.$

(b) Montrer la relation

$$\sup_{x\in\mathbb{R}} |F^n(x) - F(x)| \leq \sup_{k\leq M} (|F^n(x_k) - F(x_k)| \vee |F^n(x_{k+1}-) - F(x_{k+1}-)|) + \frac{1}{M}.$$

(c) Conclure.

C. Dans cette partie, on reprend les notations de la partie précédente, et on suppose en outre que la fonction de répartition F est continue. On s'intéresse à la loi de la v.a. $S_n = \sup_{x\in\mathbb{R}} |F^n(x) - F(x)|$.

1. H désignant l'inverse à droite de F définie à la partie A(4)b, montrer que

$$S_n = \sup_{x\in\mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq H(F(x))\}} - F(x) \right| \qquad P-p.s.,$$

puis que

$$S_n = \sup_{y\in[0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{F(X_i) \leq y\}} - y \right| \qquad P-p.s.$$

2. On considère n v.a. $U_1, \cdots, U_n$ indépendantes, de même loi uniforme sur [0,1]. Déduire de ce qui précède que la v.a.

$$D_n = \sup_{y\in[0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{U_i \leq y\}} - y \right|$$

a même loi que S_n.

3. On suppose dans cette question que $n = 2$.

(a) Montrer que, pour tout $0 < \alpha < 1/2$,

$$P(S_2 \geq 1 - \sqrt{\alpha/2}) = \alpha.$$

(b) On veut faire le test de l'hypothèse H_0 : "F est la fonction de répartition de la loi exponentielle de paramètre 1". Montrer que $E = \{X_1 \vee X_2 < \ln(10/9)\} \cup \{X_1 \wedge X_2 > \ln 10\}$ est une région de rejet de H_0 au niveau $\alpha = 0,02$. Interpréter.

D. On s'intéresse dans cette partie au problème suivant : si $(\mu_n)_n$ est une suite de lois de probabilité convergeant étroitement vers la probabilité μ, existe-t-il des v.a. X_n de loi μ_n telles que X_n converge presque sûrement vers une v.a. X de loi μ ?

1. Dans cette question, les μ_n sont les lois de variables réelles X_n, et μ est la loi d'une v.a réelle X, toutes définies sur l'espace probabilisé $(\Omega, \mathcal{F}, P)$. On utilise les fonctions G_{X_n} et G_X définies à la question A2.

(a) Montrer que $G_{X_n}(t) \to G_X(t)$ pour λ-presque tout $t \in [0,1]$.

(b) Conclure qu'il existe, sur l'espace de probabilité $[0, 1]$ muni de ses boréliens et de la mesure de Lebesgue λ, une suite de variables aléatoires $(Z_n)_n$ et une variable aléatoire Z telles que Z_n soit de loi μ_n, Z de loi μ, et $Z_n \to Z$ presque sûrement.

2. Proposer une critique de l'argument suivant :

"On suppose que $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y$. Alors d'après la question précédente on peut trouver des v.a. X'_n, X', Y'_n et Y', de même loi que X_n, X, Y_n et Y respectivement, et telles qu'on ait $X'_n \to X'$ et $Y'_n \to Y'$ presque sûrement. Par suite, on a convergence presque sûre, donc aussi en loi, du couple (X'_n, Y'_n) vers le couple (X', Y'), et on en déduit la convergence en loi de (X_n, Y_n) vers (X, Y)."

3. On se propose dans cette question de prolonger le résultat obtenu en D(1)b à des variables aléatoires à valeurs dans un espace $\mathbb{R}^d$, $d > 1$. Pour cela, on considère donc des lois de probabilité μ_j sur $\mathbb{R}^d$, et on suppose qu'on a convergence étroite de μ_j vers une probabilité μ quand $j \to +\infty$.

On construit ensuite une suite $(\mathcal{A}^n)_{n\geq 1}$ de partitions infinies de $\mathbb{R}^d$, $\mathcal{A}^n = \{A^n_k\}_{k\geq 1}$, vérifiant les propriétés suivantes :

- Pour tout $A^n_k \in \mathcal{A}^n$, il existe l tel que $A^n_k \subset A^{n-1}_l$.
- Si $k \leq k'$ et $A^n_k \subset A^{n-1}_l$ et $A^n_{k'} \subset A^{n-1}_{l'}$, alors $l \leq l'$.
- Pour tous j, n et k, $\mu_j(\partial A^n_k) = 0$ et $\mu(\partial A^n_k) = 0$.
- Pour tous n et k, $diam(A^n_k) \leq 1/2^{n-1}$.

Les deux premières propriétés indiquent que chaque élément de la partition d'ordre n est inclus dans un élément de la partition d'ordre $n-1$, et que l'ordre dans lequel on numérote les éléments de la partition d'ordre n est cohérent avec celui dans lequel on a numéroté les éléments de la partition d'ordre $n-1$.

A chaque entier $j \geq 1$, et à chaque élément A^n_k de $\mathcal{A}^n$, on associe l'intervalle réel $B^{n,j}_k = [\alpha^{n,j}_k, \beta^{n,j}_k[$ défini par :

- $\alpha^{n,j}_1 = 0$;
- pour tout k, $\alpha^{n,j}_{k+1} = \beta^{n,j}_k$;
- $\beta^{n,j}_k - \alpha^{n,j}_k = \mu_j(A^n_k)$.

De manière analogue, $B^n_k = [\alpha^n_k, \beta^n_k[$ est défini par :

- $\alpha^n_1 = 0$;
- pour tout $k \geq 1$, $\alpha^n_{k+1} = \beta^n_k$;
- $\beta^n_k - \alpha^n_k = \mu(A^n_k)$.

Enfin, pour tous n et k, on fixe un $x^n_k \in A^n_k$

(a) Montrer que, pour tous entiers n et j, $\lim\limits_{k\to\infty} \beta^{n,j}_k = 1$ et $\lim\limits_{k\to\infty} \beta^n_k = 1$.

(b) Montrer que pour tous entiers n et k,

$$\lim_{j\to\infty} (\beta^{n,j}_k - \alpha^{n,j}_k) = \beta^n_k - \alpha^n_k$$

et en déduire que $\lim\limits_{j\to\infty} \alpha^{n,j}_k = \alpha^n_k$ et $\lim\limits_{j\to\infty} \beta^{n,j}_k = \beta^n_k$.

Pour tout $y \in [0,1[$ et tout couple (n,j) d'entiers, il existe donc un unique $k \geq 1$ tel que $y \in B_k^{n,j}$. On définit alors une variable aléatoire $X^{n,j}$ sur $[0,1[$ muni de la mesure de Lebesgue λ, à valeurs dans $\mathbb{R}^d$ et constante sur chaque $B_k^{n,j}$ en posant $X^{n,j}(y) = x_k^n$ si $y \in B_k^{n,j}$ (x_k^n est l'élément fixé précédemment de manière arbitraire dans le A_k^n auquel est associé $B_k^{n,j}$).

(c) Montrer que, pour tous n, p et j entiers, et pour tout $y \in [0,1[$,

$$\|X^{n,j}(y) - X^{n+p,j}(y)\| \leq \frac{1}{2^{n-1}}.$$

La suite $(X^{n,j}(y))_{n\geq 1}$ est donc une suite de Cauchy dans $\mathbb{R}^d$. On note $X^j(y)$ sa limite.

(d) Montrer que, pour tout n entier et pour tout $y \in [0,1[-(\bigcup_k\{\alpha_k^n\}\cup\{1\})$, la suite $(X^{n,j}(y))_{j\geq 1}$ est constante à partir d'un certain rang.

On note alors $X^{n,\infty}(y)$ la limite de cette suite.

(e) Montrer que, pour n et m entiers,

$$\|X^{n,\infty}(y) - X^{m,\infty}(y)\| \leq \frac{1}{2^{n-1}} + \frac{1}{2^{m-1}}.$$

(f) Déduire de ce qui précède que $X^{n,\infty}$ converge $\lambda - p.s.$ vers une v.a. X^∞ quand $n \to \infty$, puis que X^j converge $\lambda - p.s.$ vers X^∞ quand $j \to \infty$.

(g) Montrer que si f est une fonction continue et bornée, définie sur R^d et à valeurs dans $\mathbb{R}$, on a pour tout entier j

$$\lim_{n\to\infty} \int_0^1 f(X^{n,j}(y))\lambda(dy) = \int f d\mu_j.$$

En déduire que X^j est de loi μ_j ; déterminer de même la loi de X^∞.

(h) Conclure.

E. Cette partie vise essentiellement à donner une application du résultat vu en D(1)b.

Si Z est une v.a. réelle, on dit que Z^* est une symétrisée de Z si la loi de Z^* est la même que celle de $Z - Z'$ où Z' est une v.a. de même loi que Z, et indépendante de Z.

1. Soit Z une variable aléatoire réelle, et φ sa fonction caractéristique. Calculer la fonction caractéristique φ^* de Z^*, où Z^* est une symétrisée de Z. Montrer que la loi de Z^* ne dépend que de la loi de Z et que, si $b \in \mathbb{R}$, Z^* est une symétrisée de la v.a. $Z + b$.

2. On se donne désormais une suite de réels $(b_n)_{n\geq 1}$, une suite de réels strictement positifs $(a_n)_{n\geq 1}$ et une suite de variables aléatoires réelles $(Z_n)_{n\geq 1}$. On suppose jusqu'à la fin du problème que

$$\frac{Z_n}{a_n} - b_n \xrightarrow{\mathcal{L}} U$$

quand $n \to +\infty$, où U est une variable aléatoire réelle non $P - p.s.$ constante.

Si Z_n^* et U^* désignent respectivement des symétrisées de Z_n et U, montrer que

$$\frac{Z_n^*}{a_n} \xrightarrow{\mathcal{L}} U^*$$

quand $n \to +\infty$.

3. On suppose désormais qu'il existe une suite de réels $(\beta_n)_{n\geq 1}$, une suite de réels strictement positifs $(\alpha_n)_{n\geq 1}$ et une variable aléatoire réelle non $P-p.s.$ constante U' telles que

$$\frac{Z_n}{\alpha_n} - \beta_n \xrightarrow{\mathcal{L}} U'$$

quand $n \to +\infty$.

Montrer qu'il existe, sur un espace de probabilité adéquat, des variables aléatoires X_n de même loi que Z_n et T de même loi que U telles que

$$\frac{X_n}{a_n} - b_n \to T \qquad p.s.$$

Montrer de même qu'il existe des variables aléatoires Y_n de même loi que Z_n^* et V de même loi que U^* telles que

$$\frac{Y_n}{a_n} \to V \qquad p.s.$$

4. En déduire que si a est une valeur d'adhérence (éventuellement infinie) de la suite $(a_n/\alpha_n)_{n\geq 1}$, et si U'^* est une symétrisée de U', alors U'^* a même loi que aV

En déduire que $a \neq 0$, $a \neq +\infty$, et que

$$a = \lim_{n\to+\infty} \frac{a_n}{\alpha_n}.$$

5. Soit enfin une valeur d'adhérence b (éventuellement infinie) de la suite

$$\left(\frac{a_n b_n - \alpha_n \beta_n}{a_n}\right)_{n\geq 1}.$$

Déduire de la question précédente qu'il existe une sous-suite $(n_k)_{k\geq 1}$ pour laquelle

$$\frac{Z_{n_k}}{\alpha_{n_k}} - \beta_{n_k} \xrightarrow{\mathcal{L}} aU + ab.$$

En déduire que la loi de U' est celle de $aU + ab$, et que b est en fait la limite de la suite

$$\left(\frac{a_n b_n - \alpha_n \beta_n}{a_n}\right)_{n\geq 1}.$$

6. Résumer le résultat obtenu dans cette partie.