A. PROBABILITÉS ET STATISTIQUES

Aucun document personnel n'est autorisé

L'usage de toute calculatrice est interdit

Notations et rappels

On note $\mathbb{R}$ le corps des réels et $\mathbb{C}$ celui des complexes. Le complémentaire d'une partie $A \subset \mathbb{R}$ est noté A^c. La partie réelle d'un nombre complexe z est notée $\mathrm{Re}(z)$.

1) Les variables aléatoires de ce problème sont à valeurs dans $\mathbb{R}$ et définies sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. L'opérateur d'espérance est noté $\mathbb{E}$.

Si X est une variable aléatoire, sa loi sur $\mathbb{R}$ est notée P_X et sa fonction caractéristique f_X. Par définition $f_X(t) = \mathbb{E}(e^{itX})$, $t \in \mathbb{R}$. La loi P_X d'une variable aléatoire X est *dégénérée* s'il existe $r \in \mathbb{R}$ tel que $P_X(\{r\}) = 1$.

Si $m \in \mathbb{R}$ et $\sigma \in [0, +\infty[$, la Loi Normale de moyenne m et de variance σ^2 est notée $\mathcal{N}(m, \sigma^2)$. Sa fonction caractéristique est $t \longmapsto e^{imt - \sigma^2 t^2/2}$, $t \in \mathbb{R}$.

Si $(X_n)_{n \geq 1}$ et X sont des variables aléatoires telles que X_n converge en loi vers X quand $n \to +\infty$, nous notons $P_{X_n} \to P_X$. Une suite de lois de probabilité $(P_n)_{n \geq 1}$ sur $\mathbb{R}$ est *tendue* si $\forall \varepsilon > 0$, il existe un compact $K \subset \mathbb{R}$ tel que $P_n(K^c) \leq \varepsilon$, pour tout $n \geq 1$. Rappelons les faits suivants :

– Si $(X_n)_{n \geq 1}$ et X sont des variables aléatoires, alors $P_{X_n} \to P_X$ si et seulement si f_{X_n} converge simplement vers f_X.

– Une suite tendue de lois de probabilité admet une mesure de probabilité comme valeur d'adhérence en loi. Par ailleurs, toute suite de lois de probabilité convergeant en loi est tendue.

– Si $(X_n)_{n \geq 1}$ sont des variables aléatoires telles que $(f_{X_n})_{n \geq 1}$ converge simplement dans un voisinage de 0 vers une fonction continue en 0, alors $(P_{X_n})_{n \geq 1}$ est tendue et par conséquent admet une sous-suite convergeant en loi vers une loi de probabilité.

2) On note *log* l'application définie sur $\mathbb{C}\backslash\{0\}$ et à valeurs dans l'espace quotient $\mathbb{C}/2i\pi$ telle que si $z = re^{i\theta}$, avec $r > 0$ et $\theta \in \mathbb{R}$, alors $\log z = \log r + i\theta \mod (2i\pi)$. Dans ce cas :

$$\forall (z_1, z_2) \in (\mathbb{C}\backslash\{0\})^2, \log(z_1 z_2) = \log z_1 + \log z_2 \mod (2i\pi).$$

Si $z \in \mathbb{C}\backslash]-\infty, 0]$, nous notons $Arg(z)$ l'unique réel $-\pi < \theta < \pi$ tel que $z = |z|e^{i\theta}$. Mentionnons qu'aucune connaissance en Analyse de la Variable Complexe n'est requise.

Thème du problème
Le Théorème Central Limite indique que si $(X_n)_{n\geq 1}$ sont des variables aléatoires indépendantes, de même loi et admettant un moment d'ordre 2, alors $n^{-1/2}(X_1 + \cdots + X_n - n\mathbb{E}(X_1))$ converge en loi vers $\mathcal{N}(0, \mathrm{Var}(X_1))$.

Nous envisageons la question du comportement en loi de sommes de variables aléatoires indépendantes, lorsque l'on ne suppose plus qu'elles ont la même loi ou un moment d'ordre 2. La partie I concerne des préliminaires. La partie II considère tout d'abord une situation de non-identique distribution. La partie III présente le problème de la "convergence des types". La partie IV introduit ensuite les lois infiniment divisibles et les lois stables, puis la partie V débute une étude plus précise des lois infiniment divisibles.

Les parties III, IV et V sont indépendantes de la partie II. La partie V est indépendante de la partie III.

I Préliminaires

1. Soit X une variable aléatoire. Les trois questions suivantes sont indépendantes.

a) Montrer que f_X est une fonction uniformément continue sur $\mathbb{R}$.

b) On suppose que X admet un moment d'ordre p, avec $p \geq 1$. Vérifier que f_X est de classe C^p, puis établir l'inégalité :

$$\left| f_X(t) - \sum_{k=0}^{p-1} \frac{(it)^k}{k!} \mathbb{E}[X^k] \right| \leq \frac{|t|^p}{p!} \mathbb{E}[|X|^p], \ t \in \mathbb{R}.$$

c) Montrer que si f_X est de module constant dans un voisinage de 0, alors P_X est dégénérée. On pourra considérer f_X en des valeurs u et v rationnellement indépendantes.

2. Soient $(X_n)_{n\geq 1}$ et X des variables aléatoires vérifiant $P_{X_n} \to P_X$, quand $n \to +\infty$.

a) Soit $J \subset \mathbb{R}$ un intervalle borné. Montrer que pour tout entier $n \geq 1$ et tous réels t et h :

$$|f_{X_n}(t+h) - f_{X_n}(t)| \leq \sup_{x \in J} |e^{ihx} - 1| + 2P_{X_n}(J^c).$$

b) Pour $h > 0$, posons $\delta(h) = \sup_{n\geq 1,\ t\in\mathbb{R}} |f_{X_n}(t+h) - f_{X_n}(t)|$. Montrer que $\delta(h) \to 0$ quand $h \to 0$.

c) Soit $I \subset \mathbb{R}$ un intervalle borné. En découpant par exemple I en un nombre fini d'intervalles de longueur plus petite que h, établir que f_{X_n} converge vers f_X uniformément sur I.

3. Définissons la fonction $\xi : \mathbb{R} \to \mathbb{R}$ par $\xi(t) = \sin(t)/t$, si $t \neq 0$, et $\xi(0) = 1$. Soit X une variable aléatoire.

a) Vérifier que :

$$\left[\inf_{t\in\mathbb{R},\ |t|\geq 1} (1 - \xi(t))\right] > 0.$$

b) Fixons $t > 0$. Montrer que :

$$\frac{1}{t}\int_0^t [1 - \mathrm{Re}(f_X(u))]\ du = \int_{\mathbb{R}} (1 - \xi(tx))\ \ dP_X(x).$$

c) En déduire qu'il existe une constante $M > 0$ telle que pour tout $t > 0$:

$$P_X\left(\left[-\frac{1}{t}, \frac{1}{t}\right]^c\right) \leq \frac{M}{t}\int_0^t [1 - \mathrm{Re}(f_X(u))]\ du.$$

II Cas d'une loi non uniforme : un exemple

Soient $(X_n)_{n\geq 1}$ des variables aléatoires indépendantes et admettant chacune un moment d'ordre 3. On fait l'hypothèse que $\mathbb{E}(X_n) = 0$ pour tout $n \geq 1$. On pose ensuite $\sigma_n^2 = \mathbb{E}(X_n^2)$ et on suppose que $\sigma_1^2 > 0$. On introduit également $V_n \geq 0$ tel que $V_n^2 = \sum_{k=1}^n \sigma_k^2$.

1. Vérifier que $\forall n \geq 1$, $0 < V_n < +\infty$. On impose dans toute la suite la condition :

$$\lim_{n\to+\infty} \frac{1}{V_n^3}\sum_{k=1}^n \mathbb{E}(|X_k|^3) = 0.$$

Prouver que $V_n \to +\infty$ quand $n \to +\infty$.

2. Soit t un réel fixé jusqu'à la fin de la partie II.

a) Etablir que pour tous entiers $1 \leq k \leq n$:

$$\left|f_{X_k}\left(\frac{t}{V_n}\right) - 1\right| \leq \frac{t^2}{2}\frac{\sigma_k^2}{V_n^2} \text{ et } \left|f_{X_k}\left(\frac{t}{V_n}\right) - 1 + \frac{t^2}{2}\frac{\sigma_k^2}{V_n^2}\right| \leq \frac{|t|^3}{6}\frac{\mathbb{E}(|X_k|^3)}{V_n^3}.$$

b) Montrer que $V_n^{-1}\sup_{1\leq k\leq n}\sigma_k \to 0$, quand $n \to +\infty$. En déduire que :

$$\sup_{1\leq k\leq n} |f_{X_k}(t/V_n) - 1| \to 0,\ \text{quand } n \to 0.$$

3. On pose $S_n = X_1 + \cdots + X_n$.

a) Montrer que pour tout nombre complexe z tel que $|z| \leq 1/2$, il existe $\alpha(z)$ tel que $|\alpha(z)| \leq 1$ et $\log(1 + z) = z(1 + \alpha(z)z) \mod (2i\pi)$.

b) Montrer que si n est assez grand, il existe des $(\alpha_k)_{1\leq k\leq n}$ tels que $|\alpha_k| \leq 1$ et :

$$\log(f_{S_n/V_n}(t)) = \sum_{k=1}^{n}\left(f_{X_k}\left(\frac{t}{V_n}\right)-1\right) + \sum_{k=1}^{n}\alpha_k\left(f_{X_k}\left(\frac{t}{V_n}\right)-1\right)^2 \quad \text{mod } (2i\pi).$$

c) Vérifier que :

$$\sum_{k=1}^{n}|\alpha_k|\left|f_{X_k}\left(\frac{t}{V_n}\right)-1\right|^2 \leq \frac{t^2}{2}\sup_{1\leq k\leq n}\left|f_{X_k}\left(\frac{t}{V_n}\right)-1\right|.$$

d) En déduire que S_n/V_n converge en loi vers $\mathcal{N}(0,1)$.

III Convergence des types et application

Soient $(X_n)_{n\geq 1}$, X et X' des variables aléatoires. Supposons qu'il existe des suites réelles $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ telles que $P_{X_n} \to P_X$ et $P_{a_nX_n+b_n} \to P_{X'}$, quand $n \to +\infty$. On fait l'hypothèse que les lois P_X et $P_{X'}$ sont non dégénérées.

1. Montrer sur un exemple que la suite $(a_n)_{n\geq 1}$ peut ne pas converger.

2. On suppose à présent que $a_n > 0$, pour tout $n \geq 1$.

a) Montrer que $|f_{X_n}(a_nt)| \to |f_{X'}(t)|$ et $|f_{X_n}(t)| \to |f_X(t)|$, uniformément en $t \in I$, si I est un intervalle borné.

b) Supposons qu'il existe une sous-suite d'entiers $(\varphi(n))_{n\geq 1}$ telle que $a_{\varphi(n)} \to +\infty$, lorsque $n \to +\infty$. Montrer que $|f_{X_{\varphi(n)}}(t)| \to |f_{X'}(0)|$, pour tout $t \in \mathbb{R}$. En déduire que la suite $(a_n)_{n\geq 1}$ est bornée.

c) On suppose qu'il existe deux valeurs d'adhérence finies $a' < a''$ de la suite $(a_n)_{n\geq 1}$. Montrer que pour tout $t \in \mathbb{R}$, on a $|f_X(a't)| = |f_X(a''t)| = |f_{X'}(t)|$. Etablir une contradiction, puis que $(a_n)_{n\geq 1}$ converge.

d) D'après la question précédente, la suite $(a_n)_{n\geq 1}$ converge vers un réel noté a. Montrer que $f_{X_n}(a_nt) \to f_X(at)$ uniformément en $t \in I$, où I est un intervalle borné. Soit $\varepsilon > 0$. Prouver qu'il existe $\delta > 0$ tel que e^{itb_n} converge uniformément sur $[0,\delta]$ vers une fonction $\theta(t)$ telle que $|\theta(t)-1| \leq \varepsilon$.

e) En déduire que la suite $(b_n)_{n\geq 1}$ converge (on pourra effectuer une intégration bien choisie). Conclure qu'il existe des réels a et b, avec $a > 0$, tels que $a_n \to a$ et $b_n \to b$, puis que $P_{X'} = P_{aX+b}$.

3. Soient $(X_n)_{n\geq 1}$ des variables aléatoires indépendantes, de même loi et ayant un moment d'ordre 2. Soient des suites réelles $(\gamma_n)_{n\geq 1}$ et $(\delta_n)_{n\geq 1}$, avec $\gamma_n > 0$ pour tout $n \geq 1$ et posons :

$$Y_n = \frac{1}{\gamma_n}(X_1 + \cdots + X_n) + \delta_n.$$

On suppose qu'il existe une variable aléatoire X telle que $P_{Y_n} \to P_X$ et que P_{X_1} et P_X sont non dégénérées. Montrer qu'il existe une constante $c > 0$ telle que $\gamma_n \sim c\sqrt{n}$, quand $n \to +\infty$.

IV Lois infiniment divisibles et lois stables

La loi P_X d'une variable aléatoire X est dite *infiniment divisible* si pour tout $n \geq 1$, il existe des variables aléatoires indépendantes et de même loi $(X_{k,n})_{1\leq k\leq n}$ telles que $P_X = P_{X_{1,n}+\cdots+X_{n,n}}$.

1.

a) Soient $m \in \mathbb{R}$ et $\sigma \in [0, +\infty[$. Montrer que $\mathcal{N}(m, \sigma^2)$ est infiniment divisible.

b) Soient deux réels $\lambda > 0$ et $s \neq 0$. Une variable aléatoire Y suit une loi de Poisson $\mathcal{P}(\lambda, s)$ de paramètre λ avec saut s si $P_Y(\{ns\}) = e^{-\lambda}\lambda^n/n!$, $n \geq 0$. Montrer dans ce cas que $f_Y(t) = e^{\lambda(e^{its}-1)}$, $t \in \mathbb{R}$. En déduire que $\mathcal{P}(\lambda, s)$ est infiniment divisible.

c) Soient un réel $L \geq 0$ et X une variable aléatoire telle que $P_X([-L, L]) = 1$. Supposons P_X infiniment divisible. Pour tout $n \geq 1$, écrivons $P_X = P_{X_{1,n}+\cdots+X_{n,n}}$, avec $(X_{k,n})_{1\leq k\leq n}$ des variables aléatoires indépendantes et de même loi. Montrer que :

$$\forall n \geq 1, \; P_{X_{1,n}}([-L/n, L/n]) = 1, \text{ puis que } \mathrm{Var}(X_{1,n}) \leq L^2/n^2.$$

En déduire que P_X est dégénérée.

2. Pour tout $n \geq 1$, soient $(X_{k,n})_{1\leq k\leq n}$ des variables aléatoires indépendantes et de même loi. On pose $S_n = X_{1,n} + \cdots + X_{n,n}$. Soit X une variable aléatoire telle que $P_{S_n} \to P_X$. Fixons un entier $k \geq 1$ et écrivons :

$$S_{kn} = Y_{1,n} + \cdots + Y_{k,n}, \text{ avec } Y_{l,n} = X_{(l-1)n+1,kn} + \cdots + X_{ln,kn}, \text{ pour } 1 \leq l \leq k.$$

a) Montrer que les $(Y_{l,n})_{1\leq l\leq k}$ sont indépendantes et de même loi. On note P_n leur loi commune. Montrer que la suite (P_n) est tendue.

b) En déduire que P_X est infiniment divisible.

c) Montrer qu'il existe un voisinage V de 0 dans $\mathbb{R}$ et $\delta > 0$ tels que $\mathrm{Re}(f_{S_n}(t)) \geq \delta$, pour $t \in V$ et tout $n \geq 1$.

d) Prouver que si $t \in V$, alors $\mathrm{Arg}(f_{S_n}(t)) = n\mathrm{Arg}(f_{X_{1,n}}(t))$, pour tout $n \geq 1$. On pourra par exemple montrer que, pour tout $n \geq 1$, cette relation est vraie dans un voisinage V_n de 0, puis qu'elle reste vraie dans V.

e) Montrer que $f_{X_{1,n}}$ converge simplement vers 1 sur V. En déduire que la suite $(P_{X_{1,n}})_{n\geq 1}$ est tendue, puis que $X_{1,n}$ converge en loi vers 0 et que $f_{X_{1,n}}$ converge vers 1 uniformément sur tout compact.

La loi P_X d'une variable aléatoire X est dite *stable* si pour tout $n \geq 1$, il existe des constantes réelles $a_n > 0$ et b_n telles que si $(X_k)_{1 \leq k \leq n}$ sont indépendantes et de même loi P_X, alors $P_{X_1+\cdots+X_n} = P_{a_n X + b_n}$.

Soient $(X_n)_{n\geq 1}$ des variables aléatoires indépendantes, de même loi et des suites réelles $(C_n)_{n\geq 1}$ et $(D_n)_{n\geq 1}$, avec $C_n > 0$ pour tout $n \geq 1$. Posons :

$$U_n = \frac{1}{C_n}(X_1 + \cdots + X_n) - D_n. \tag{1}$$

3. Montrer que si la loi P_X d'une variable aléatoire X est stable, alors X est limite en loi de sommes de type (1).

4. Soit X une variable aléatoire telle que $P_{U_n} \to P_X$. On suppose les lois P_X et P_{X_1} non dégénérées.

a) Montrer que P_X est infiniment divisible. En déduire que $C_n \to +\infty$ et $D_n/n \to 0$, quand $n \to +\infty$.

b) Soit $k \geq 1$ un entier fixé. Pour tout entier $1 \leq i \leq k$, on pose $S_n^{(i)} = X_{(i-1)n+1} + \cdots + X_{in}$. Montrer l'existence d'une constante $F_{n,k}$ telle que :

$$\left(\frac{S_n^{(1)}}{C_n} - D_n\right) + \cdots + \left(\frac{S_n^{(k)}}{C_n} - D_n\right) = \frac{C_{nk}}{C_n} U_{nk} + F_{n,k}.$$

c) L'entier k étant toujours fixé, montrer que les suites $(C_{nk}/C_n)_{n\geq 1}$ et $(F_{n,k})_{n\geq 1}$ convergent. Etablir que P_X est stable.

V Fonction caractéristique d'une loi infiniment divisible

Soit X une variable aléatoire telle que P_X est infiniment divisible. Par définition, pour tout $n \geq 1$, il existe une variable aléatoire X_n telle que $f_X = (f_{X_n})^n$.

1. Montrer que $|f_X(t)| > 0$ pour tout $t \in \mathbb{R}$. En utilisant la partie IV, vérifier qu'il existe un voisinage V de 0 dans $\mathbb{R}$ et une suite de fonctions $(t \longmapsto \eta_n(t))$ définies sur $\mathbb{R}$ et tendant vers 0 uniformément sur tout compact, tels que :

$$\log(f_X(t)) = n(f_{X_n}(t) - 1)(1 + \eta_n(t)) \mod (2i\pi), \text{ si } t \in \mathbb{R}$$

et

$$\log|f_X(t)| + iArg(f_X(t)) = n(f_{X_n}(t) - 1)(1 + \eta_n(t)), \text{ si } t \in V.$$

On introduit la mesure $m_n = nP_{X_n}$. En déduire l'expression :

$$\log(f_X(t)) = (1 + \eta_n(t)) \int_{\mathbb{R}} (e^{itx} - 1)\, dm_n(x) \mod (2i\pi),\ t \in \mathbb{R}.$$

2.

a) Montrer que si $t \in V$:

$$\lim_{n \to +\infty} n(1 - \mathrm{Re}(f_{X_n}(t))) = -\log |f_X(t)|.$$

b) Soit $L > 0$ tel que $[0, 1/L] \subset V$. Montrer qu'il existe une constante M telle que l'on ait l'inégalité :

$$\limsup_{n \to +\infty} m_n\left([-L, L]^c\right) \leq ML \int_0^{1/L} |\log |f_X(u)|| \, du.$$

c) Montrer qu'il existe une constante β avec $0 < \beta < \infty$, telle que $\cos x \leq 1 - \beta x^2$, pour $|x| \leq 1$. Soit $\varepsilon > 0$, avec $\varepsilon \in V$. Déduire de l'égalité $n(1 - \mathrm{Re}(f_{X_n}(\varepsilon))) = \int_{\mathbb{R}} (1 - \cos \varepsilon u) \, dm_n(u)$ que :

$$\limsup_{n \to +\infty} \int_{-1/\varepsilon}^{1/\varepsilon} x^2 \, dm_n(x) < +\infty.$$

3. Pour $n \geq 1$, on définit $I_n = \int_{\mathbb{R}} x^2(1 + x^2)^{-1} \, dm_n(x) > 0$ et

$$\beta_n = \int_{\mathbb{R}} x(1 + x^2)^{-1} \, dm_n(x),$$

ainsi que la mesure de probabilité P_n sur $\mathbb{R}$ telle que :

$$P_n(B) = \frac{1}{I_n} \int_B \frac{x^2}{1 + x^2} \, dm_n(x), \text{ pour tout borélien } B.$$

a) On introduit la fonction suivante sur $\mathbb{R} \times \mathbb{R}$:

$$\zeta(x, t) = \left(e^{itx} - 1 - \frac{itx}{1 + x^2}\right) \frac{1 + x^2}{x^2}, \text{ si } x \neq 0, \text{ et } \zeta(0, t) = -t^2/2.$$

Montrer que ζ est continue sur $\mathbb{R} \times \mathbb{R}$ et que :

$$\forall A > 0, \quad \sup_{x \in \mathbb{R}, \, |t| \leq A} |\zeta(x, t)| < +\infty.$$

b) Vérifier l'égalité :

$$\log(f_X(t)) = \left[(1 + \eta_n(t)) I_n \int_{\mathbb{R}} \zeta(x, t) \, dP_n(x)\right] + i(1 + \eta_n(t))\beta_n t \mod (2i\pi).$$

c) Supposons à présent que ci-dessus le premier terme entre crochets du membre de droite converge uniformément sur tout compact. Montrer qu'alors (β_n) converge.

4.

a) Montrer que si $\liminf_{n\to+\infty} I_n = 0$, alors P_X est dégénérée.

b) On suppose que $\liminf_{n\to+\infty} I_n > 0$. Vérifier que la suite $(I_n)_{n\geq 1}$ est bornée, puis que la suite $(P_n)_{n\geq 1}$ est tendue.

c) En déduire que si P_X est infiniment divisible, alors il existe des constantes réelles β, σ et une mesure finie μ ne chargeant pas l'origine telles que :

$$\log(f_X(t)) = i\beta t - \frac{\sigma^2 t^2}{2} + \int_{\mathbb{R}} \zeta(x,t)\, d\mu(x) \mod (2i\pi).$$