

2C7122

**Ecole Normale Supérieure Paris-Saclay
Ecole Normale Supérieure de Rennes**

**SECOND CONCOURS – ADMISSION EN CYCLE
MASTER MATHÉMATIQUES**

Session 2017

Epreuve de Mathématiques 2

Durée : 5 heures

« Aucun document n'est autorisé »

« L'usage de toute calculatrice est interdit »

Si, au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il le signale sur sa copie et poursuit sa composition en expliquant les raisons des initiatives qu'il est amené à prendre.

Sujet 1- Probabilités et statistiques

Sujet 2- Analyse numérique

Le candidat traitera au choix le sujet 1 ou le sujet 2 et rappellera sur sa copie le sujet qu'il a choisi de traiter.

Sujet 1 : Probabilités et Statistique

Préambule

Ce sujet est consacré à l'estimation des quantiles d'une loi de probabilité. Dans tout le sujet on considère $(X_n)_{n \in \mathbb{N}}$ des variables aléatoires réelles indépendantes et identiquement distribuées (i.i.d.) définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. La fonction de répartition de ces variables aléatoires est notée F et est supposée continue. Dans certaines questions il est supposé que la loi de ces variables est à densité, dans ce cas la densité est notée f . On se donne maintenant un réel $p \in [0, 1]$ et on cherche un réel θ_p vérifiant $F(\theta_p) = p$. Hormis certaines questions (question 5 de la partie 1 et question 4 de la partie 2), on suppose toujours que $p \in]0, 1[$ et qu'il existe un unique θ_p vérifiant $F(\theta_p) = p$. La partie 1 étudie l'estimation de θ_p à l'aide des statistiques d'ordre. Les parties 2 et 3 étudient l'estimation de θ_p utilisant un algorithme stochastique récursif : la partie 2 montre la convergence de l'estimateur et la partie 3 s'intéresse à la normalité asymptotique de ce dernier. Hormis quelques questions, ces trois parties sont indépendantes.

Nous allons commencer par préciser certaines notations et donner quelques résultats qui pourront être considérés comme connus. Sous réserve d'existence, l'espérance d'une variable aléatoire réelle Y est notée $\mathbb{E}[Y]$. De plus, si $\mathcal{G} \subset \mathcal{F}$ est une tribu, l'espérance conditionnelle de Y sachant \mathcal{G} est notée $\mathbb{E}[Y|\mathcal{G}]$. Pour tout $n \in \mathbb{N}^*$ on note \mathcal{F}_n la tribu $\sigma(\{X_1, \dots, X_n\})$ engendrée par $\{X_1, \dots, X_n\}$ et $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Dans tout le sujet on utilise la convention $\inf \emptyset = +\infty$ et $\sup \emptyset = -\infty$. On note $x \wedge y$ (respectivement $x \vee y$) le minimum (respectivement le maximum) entre deux réels x et y . On note également $[x]$ la partie entière de x . Pour une suite de variables aléatoires $(Y_n)_{n \in \mathbb{N}}$ on note

$$Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Y, \quad (\text{respectivement } Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} Y),$$

le fait que Y_n converge en loi (respectivement en probabilité) vers Y .

On rappelle que la fonction caractéristique d'une variable $Y \sim \mathcal{N}(m, \sigma^2)$ est donnée par

$$\phi_Y(t) = e^{imt - \frac{\sigma^2 t^2}{2}}.$$

On rappelle également le développement asymptotique suivant

$$\sum_{k=1}^n \frac{1}{k} = \ln(n) + \gamma + o(1)$$

avec γ la constante d'Euler.

Théorème 1 (Slutsky) Soient $(Y_n)_{n \in \mathbb{N}}$ et $(Z_n)_{n \in \mathbb{N}}$ deux suites de variables aléatoires à valeurs respectivement dans \mathbb{R}^p et \mathbb{R}^q . Si Y_n converge en loi vers Y et Z_n converge en probabilité vers une constante z , alors (Y_n, Z_n) converge en loi vers (Y, z) .

Rappelons enfin qu'un processus stochastique à valeurs réelles $(M_n)_{n \in \mathbb{N}}$ est une martingale par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$ si, pour tout $n \in \mathbb{N}$, M_n est \mathcal{F}_n -mesurable, intégrable et $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$ p.s. De même, $(M_n)_{n \in \mathbb{N}}$ est une sur-martingale par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$ si, pour tout $n \in \mathbb{N}$, M_n est \mathcal{F}_n -mesurable, intégrable et $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \leq M_n$ p.s.

Lemme 1 (Doob) Une sur-martingale minorée converge presque sûrement.

Si on considère $(M_n)_{n \in \mathbb{N}}$ une martingale de carré intégrable, i.e. $\mathbb{E}[M_n^2] < +\infty$ pour tout $n \geq 0$, on appelle crochet de la martingale $(M_n)_{n \in \mathbb{N}}$, et on note $(\langle M \rangle_n)_{n \in \mathbb{N}}$, le processus défini par $\langle M \rangle_0 = 0$ et, pour tout $n \geq 1$,

$$\langle M \rangle_n = \sum_{k=1}^n \mathbb{E}[(M_k - M_{k-1})^2 | \mathcal{F}_{k-1}].$$

Théorème 2 (TLC pour les martingales) Soit $(M_n)_{n \in \mathbb{N}}$ une martingale par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$ et soit $(a_n)_{n \in \mathbb{N}}$ une suite réelle, strictement positive, déterministe croissant vers l'infini. On suppose que

1. il existe $\lambda \geq 0$ déterministe tel que $a_n^{-1} \langle M \rangle_n$ converge en probabilité vers λ ,
2. il existe $\delta > 0$ tel que

$$\frac{1}{(a_n)^{1+\frac{\delta}{2}}} \sum_{k=1}^n \mathbb{E} \left[(M_k - M_{k-1})^{2+\delta} | \mathcal{F}_{k-1} \right] \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

Alors on a

$$\frac{1}{\sqrt{a_n}} M_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \lambda).$$

1 Estimation à l'aide des statistiques d'ordre

On note $(X_{(1,n)}, \dots, X_{(n,n)})$ le réarrangement croissant de (X_1, \dots, X_n) , c'est à dire que l'on a $X_{(1,n)} \leq \dots \leq X_{(n,n)}$ et il existe une permutation aléatoire de $\{1, \dots, n\}$, notée σ_n , telle que $(X_{(1,n)}, \dots, X_{(n,n)}) = (X_{\sigma_n(1)}, \dots, X_{\sigma_n(n)})$. Pour tous $n \in \mathbb{N}^*$ et $x \in \mathbb{R}$ on pose $S_n(x) := \sum_{i=1}^n \mathbb{1}_{X_i \leq x}$. Enfin, on note $(k(n))_{n \in \mathbb{N}^*}$ une suite d'entiers telle que, pour tout $n \in \mathbb{N}^*$ on a $1 \leq k(n) \leq n$ et $\lim_{n \rightarrow +\infty} \frac{k(n)}{n} = p$.

1. (a) Exprimer les événements aléatoires $\{X_{(k(n),n)} \leq x\}$ et $\{X_{(k(n),n)} > x\}$ à l'aide de $S_n(x)$.
(b) Montrer que

$$\lim_{n \rightarrow +\infty} \frac{S_n(x)}{k(n)} = \frac{F(x)}{p} \quad \text{p.s.}$$

- (c) En déduire que

$$\lim_{n \rightarrow +\infty} X_{(\lfloor np \rfloor + 1, n)} = \theta_p \quad \text{p.s.}$$

À défaut d'une convergence presque sûre, une convergence en probabilité sera appréciée.

2. En reprenant les raisonnements de la question 1), montrer que

$$\lim_{n \rightarrow +\infty} X_{(n,n)} = \inf\{x \in \mathbb{R}; F(x) = 1\} \quad \text{p.s.} \quad \text{et} \quad \lim_{n \rightarrow +\infty} X_{(1,n)} = \sup\{x \in \mathbb{R}; F(x) = 0\} \quad \text{p.s.}$$

On pourra distinguer selon que $\inf\{x \in \mathbb{R}; F(x) = 1\}$ est fini ou non.

3. On suppose dans cette question que la loi des $(X_n)_{n \in \mathbb{N}^*}$ est à densité, de fonction densité notée f , que f est continue en θ_p et que $f(\theta_p) > 0$. On suppose de plus que $k(n)$ vérifie le développement asymptotique suivant :

$$k(n) = np + o(\sqrt{n}).$$

(a) Soit $x \in \mathbb{R}$. Pour tout $n \in \mathbb{N}^*$ on pose $y_n = \theta_p + \frac{x}{\sqrt{n}}$ et $p_n = F(y_n)$. Montrer que

$$p_n - p = \frac{x}{\sqrt{n}} f(\theta_p) + o\left(\frac{1}{\sqrt{n}}\right).$$

(b) Montrer que pour tout $k \in \{1, \dots, n\}$ on a

$$\{\sqrt{n}(X_{(k,n)} - \theta_p) \leq x\} = \left\{V_n \geq \sqrt{n} \left(\frac{k}{n} - p_n\right)\right\}$$

$$\text{avec } V_n := \sqrt{n} \left(\frac{S_n(y_n)}{n} - p_n\right).$$

(c) Calculer la fonction caractéristique de V_n et montrer que

$$V_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} V,$$

$$\text{avec } V \sim \mathcal{N}(0, p(1-p)).$$

(d) En déduire que $\sqrt{n}(X_{(\lfloor np \rfloor + 1, n)} - \theta_p)$ converge en loi lorsque $n \rightarrow +\infty$ et donner sa limite.

4. Pour $\alpha \in]0, 1[$, on note a_α le quantile d'ordre $1 - \alpha/2$ de la loi $\mathcal{N}(0, 1)$. On pose $i_n = 1 \vee \lfloor np - \sqrt{n} a_\alpha \sqrt{p(1-p)} \rfloor$ et $j_n = n \wedge \lfloor np + \sqrt{n} a_\alpha \sqrt{p(1-p)} \rfloor$. Montrer que pour tout $i \in \{1, \dots, n\}$ on a

$$\mathbb{P}(\theta_p = X_{(i,n)}) = 0$$

et

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\theta_p \in [X_{(i_n,n)}, X_{(j_n,n)}]) = 1 - \alpha.$$

5. On s'intéresse dans cette question aux cas $p = 0$ et $p = 1$.

(a) Calculer les fonctions de répartition de $X_{(1,n)}$ et $X_{(n,n)}$ en fonction de F .

(b) On suppose dans cette question que les variables $(X_i)_{i \in \mathbb{N}^*}$ suivent une loi exponentielle de paramètre 1, dont la densité est donnée par

$$f(x) = e^{-x} \mathbb{1}_{x > 0}.$$

Trouver $\beta > 0$ et $(b_n)_{n \in \mathbb{N}^*}$ tels que $(n^\beta X_{(1,n)})_{n \in \mathbb{N}^*}$ et $(X_{(n,n)} - b_n)_{n \in \mathbb{N}^*}$ convergent en loi, lorsque $n \rightarrow +\infty$, vers des variables aléatoires à densité.

2 Estimation à l'aide d'un algorithme stochastique

On définit une suite de variables aléatoires réelles $(\hat{\theta}_n)_{n \in \mathbb{N}}$ en posant $\hat{\theta}_0 = 0$ et $\hat{\theta}_{n+1} = \hat{\theta}_n - \gamma_n (\mathbb{1}_{X_{n+1} \leq \hat{\theta}_n} - p)$ pour $n \in \mathbb{N}$. On suppose que $(\gamma_n)_{n \in \mathbb{N}}$ est une suite de réels positifs vérifiant

$$\sum_{n \in \mathbb{N}} \gamma_n = +\infty \quad \text{et} \quad \sum_{n \in \mathbb{N}} \gamma_n^2 < +\infty.$$

Le but de cette partie est de montrer que $\hat{\theta}_n$ tend vers θ_p presque sûrement. Dans toute la suite on pose $T_n := \hat{\theta}_n - \theta_p$.

1. Montrer que pour tout $n \in \mathbb{N}$ on a

$$\mathbb{E}[|T_{n+1}|^2 | \mathcal{F}_n] = |T_n|^2 - 2\gamma_n T_n (F(\hat{\theta}_n) - p) + \gamma_n^2 \mathbb{E}[|\mathbb{1}_{X_{n+1} \leq \hat{\theta}_n} - p|^2 | \mathcal{F}_n].$$

2. (a) Montrer que le processus $(Z_n)_{n \in \mathbb{N}}$ défini par

$$Z_n := |T_n|^2 + \sum_{k=0}^{+\infty} \gamma_k^2 - \sum_{k=0}^{n-1} \gamma_k^2 \mathbb{E} \left[|\mathbb{1}_{X_{k+1} \leq \hat{\theta}_k} - p|^2 | \mathcal{F}_k \right]$$

est une sur-martingale.

- (b) Montrer que $(|T_n|)_{n \in \mathbb{N}}$ converge presque sûrement.
 3. (a) Prouver que la série de terme général $\gamma_k T_k (F(\hat{\theta}_k) - p)$ est presque-sûrement convergente.
 (b) En utilisant le résultat précédent montrer que pour tout $\delta > 0$ on a

$$\mathbb{P}(\delta < \lim_{n \rightarrow +\infty} |T_n| < 1/\delta) = 0.$$

- (c) Montrer que $(\hat{\theta}_n)_{n \in \mathbb{N}}$ converge presque sûrement vers θ_p .
 4. On s'intéresse dans cette question au cas $p = 1$. On pose $x_F := \inf\{x \in \mathbb{R}; F(x) = 1\}$ et on suppose que $x_F > 0$ et $\sup\{x \in \mathbb{R}; F(x) = 0\} < 0$.
 (a) Pour tout $K \in \mathbb{R}^+$, on introduit la variable aléatoire $U^K = \inf\{n \in \mathbb{N}^*; \hat{\theta}_n > K\}$ à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$. Montrer que l'on a

$$\mathbb{E}[\hat{\theta}_{(n+1) \wedge U^K}] \geq \mathbb{E}[\hat{\theta}_{n \wedge U^K}] + \gamma_n (1 - F(K)) \mathbb{P}(U^K > n).$$

- (b) Si $K < x_F$, montrer que U^K est fini presque sûrement.
 (c) Si $x_F = +\infty$, montrer que $(\hat{\theta}_n)_{n \in \mathbb{N}}$ tend presque sûrement vers $+\infty$.
 (d) Si $x_F < +\infty$, montrer que $(\hat{\theta}_n)_{n \in \mathbb{N}}$ converge presque sûrement vers une variable aléatoire $\hat{\theta}_\infty$ bornée, donner des bornes explicites et montrer que $\mathbb{P}(\hat{\theta}_\infty \neq x_F) > 0$.

3 Vitesse de convergence de l'algorithme stochastique

On suppose dans cette partie que F est de classe C^2 et que $2f(\theta_p) - 1 > 0$. On considère à nouveau la suite $(\hat{\theta}_n)_{n \in \mathbb{N}}$ introduite dans la partie précédente pour laquelle on prend $\gamma_n = 1/(n + n_0)$ avec n_0 une constante fixée qui vérifie $2f(\theta_p) < n_0$. On a montré dans la partie 2 que $\hat{\theta}_n$ converge presque sûrement vers θ_p , le but de cette partie est de montrer la convergence suivante :

$$\sqrt{n} (\hat{\theta}_n - \theta_p) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(0, \frac{p(1-p)}{2f(\theta_p) - 1} \right). \quad (1)$$

On pose $Y_n := \mathbb{1}_{X_n \leq \hat{\theta}_{n-1}} - F(\hat{\theta}_{n-1})$, $\tau := f(\theta_p)$, $\delta_n := F(\hat{\theta}_n) - p - \tau(\hat{\theta}_n - \theta_p)$, $\alpha_n := 1 - \frac{\tau}{n+n_0}$, $\beta_n := \prod_{i=0}^n \alpha_i$ et $T_n := \hat{\theta}_n - \theta_p$.

1. Montrer que, pour tout $n \in \mathbb{N}^*$, T_n peut se réécrire

$$T_n = \beta_{n-1} T_0 - \beta_{n-1} M_n - \beta_{n-1} R_{n-1}$$

avec

$$M_n = \sum_{k=0}^{n-1} \left(\frac{1}{(k+n_0)\beta_k} \right) Y_{k+1} \quad \text{et} \quad R_n = \sum_{k=0}^n \left(\frac{1}{(k+n_0)\beta_k} \right) \delta_k.$$

2. Montrer que $(M_n)_{n \in \mathbb{N}}$ est une martingale et calculer son crochet.
 3. Montrer qu'il existe une constante $c > 0$ telle que $\lim_{n \rightarrow +\infty} n^\tau \beta_n = c$.
 4. (a) Pour tout réel $r < 1$, donner un équivalent de $\sum_{k=1}^n k^{-r}$ lorsque $n \rightarrow +\infty$.

(b) Montrer que l'on a, lorsque n tend vers l'infini,

$$\langle M \rangle_n \sim \frac{p(1-p)}{c^2(2\tau-1)} n^{2\tau-1}.$$

(c) Montrer qu'il existe $\delta > 0$ tel que

$$\sum_{k=0}^{n-1} \mathbb{E} \left[\left| \frac{1}{(k+n_0)\beta_k} Y_{k+1} \right|^{2+\delta} \middle| \mathcal{F}_k \right] = O(n^{2\tau-1-\delta(1-\tau)}).$$

(d) En déduire que l'on a

$$\sqrt{n}\beta_{n-1}M_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(0, \frac{p(1-p)}{2\tau-1} \right).$$

5. Pour $j \in \mathbb{N}$ et $\varepsilon > 0$ on pose $U^{j,\varepsilon} = \inf\{n \geq j; |T_n| > \varepsilon\}$.

(a) Montrer que pour ε suffisamment petit (indépendamment de j) et $n \geq j$ on a

$$\mathbb{E} \left[|T_{n+1}|^2 \mathbb{1}_{U^{j,\varepsilon} > n} \middle| \mathcal{F}_n \right] \leq \left(|T_n|^2 \left(1 - \frac{2f(\theta_p) - \eta}{n+n_0} \right) + \frac{1}{(n+n_0)^2} \right) \mathbb{1}_{U^{j,\varepsilon} > n},$$

avec $0 < \eta < 2f(\theta_p) - 1$. ε est maintenant fixé jusqu'à la fin de la partie de telle sorte que l'inégalité précédente est vraie pour tout $j \in \mathbb{N}$ et tout $n \geq j$.

(b) En déduire qu'il existe une constante $C > 0$ telle que, pour tout $n \in \mathbb{N}^*$ on a

$$\mathbb{E} [|T_n|^2 \mathbb{1}_{U^{j,\varepsilon} > n}] \leq \frac{C}{n}.$$

Indication : on pourra introduire les quantités $\tilde{\alpha}_n := 1 - \frac{2\tau-\eta}{n+n_0}$ et $\tilde{\beta}_n := \prod_{i=j}^n \tilde{\alpha}_i$.

(c) Montrer que

$$\lim_{n \rightarrow +\infty} n^{1/2-\tau} \mathbb{E} \left[\sum_{k=1}^n \frac{1}{k^{1-\tau}} |T_k|^2 \mathbb{1}_{U^{j,\varepsilon} = +\infty} \right] = 0.$$

(d) En déduire que $n^{1/2-\tau} \sum_{k=1}^{n-1} \left(\frac{1}{k^{1-\tau}} \right) \delta_k$ puis que $\sqrt{n}\beta_{n-1}R_{n-1}$ tendent vers 0 en probabilité lorsque n tend vers l'infini.

6. Montrer que le résultat (1) est vrai.

7. On suppose que l'on cherche à estimer θ_p à partir d'une très grande quantité de données (i.e. n très grand), ou bien que l'on a pas accès à toutes les données initialement et que de nouvelles données arrivent régulièrement. Dans ce cas, quel est l'intérêt de l'estimateur $\hat{\theta}_n$ par rapport à $X_{(\lfloor np \rfloor + 1, n)}$?