Université des Sciences et Technologies de Lille
U.F.R. de Mathématiques Pures et Appliquées

IS Math 314 **Année 2008–2009**

Fiche n° 5

Ex 1. Médianes d'une loi

Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et X une v.a. réelle définie sur cet espace. On appelle *médiane de* X, tout réel m vérifiant :

$$P(X \geq m) \geq \frac{1}{2} \quad \text{et} \quad P(X \leq m) \geq \frac{1}{2}. \tag{1}$$

Cette double condition pouvant se réécrire $P_X([m, +\infty[) \geq \frac{1}{2}$ et $P_X(]-\infty, m]) \geq \frac{1}{2}$, il est clair que cette médiane si elle existe, ne dépend que de la loi P_X de X sous P. Il serait donc plus correct de parler de médiane de la loi de X sous P. Mais nous commettrons l'abus de langage habituel (comme pour la f.d.r., l'espérance, les moments,...) en parlant de médiane(s) de X.

1) Montrer que l'ensemble des médianes de X est un intervalle[1] de $\mathbb{R}$. On vérifiera que si $m' < m''$ sont deux médianes de X, tout $m \in [m', m'']$ est une médiane de X. Cet intervalle est *fermé* (vous pouvez le démontrer) et si on note a et b les bornes inférieure et supérieure de cet intervalle, l'intervalle $[a, b]$ est appelé *intervalle médian* de X.

2) Montrer que x est une médiane de X si et seulement si

$$P(X < x) \leq \frac{1}{2} \leq P(X \leq x). \tag{2}$$

3) En déduire que si F est continue sur $\mathbb{R}$, toute médiane m vérifie $F(m) = \frac{1}{2}$ et que s'il existe un intervalle médian $[a, b]$, c'est l'ensemble des solutions de l'équation $F(x) = \frac{1}{2}$.

4) On rappelle que l'*inverse généralisé* F^{-1} de F est donné par :

$$\forall u \in]0, 1[, \quad F^{-1}(u) := \inf\{t \in \mathbb{R};\ F(t) \geq u\}. \tag{3}$$

Montrer que $F^{-1}(1/2)$ est toujours une médiane de X et que c'est la plus petite des médianes de X.

Ex 2. Estimation de médiane

Sur l'espace probabilisé $(\Omega, \mathcal{F}, P)$, soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires réelles indépendantes de même loi de f.d.r. F. On suppose que l'inverse généralisé F^{-1} définie par (3) est continue au point $\frac{1}{2}$.

1. L'ensemble vide est un intervalle, mais on verra à la question 4 que l'ensemble des médianes de X n'est jamais vide.

1) Donner une condition suffisante sur F pour qu'il en soit ainsi.

On se propose d'estimer la plus petite médiane de la loi des X_i, donc $F^{-1}(\frac{1}{2})$ à partir de l'observation d'un échantillon de grande taille. On note F_n la f.d.r. empirique construite sur l'échantillon $X_1, \dots, X_n$ et on pose

$$M_n := F_n^{-1}\Big(\frac{1}{2}\Big) = \inf\big\{t \in \mathbb{R};\ F_n(t) \geq \frac{1}{2}\big\}.$$

Le but de cet exercice est de prouver que M_n converge presque-sûrement vers $F^{-1}(\frac{1}{2})$ lorsque n tend vers $+\infty$.

2) En écrivant $F(M_n) = F_n(M_n) + \big(F(M_n) - F_n(M_n)\big)$ et en comparant $F_n(M_n)$ avec $\frac{1}{2}$, montrer grâce au théorème de Glivenko-Cantelli, qu'il existe un $\Omega_1 \in \mathfrak{F}$ tel que $P(\Omega_1) = 1$ et

$$\forall \omega \in \Omega_1,\ \forall \varepsilon > 0,\ \exists N_1 = N_1(\omega, \varepsilon),\ \forall n \geq N_1, \quad F(M_n(\omega)) > \frac{1}{2} - \varepsilon. \tag{4}$$

3) De même en écrivant $F(M_n - \varepsilon) = F_n(M_n - \varepsilon) + \big(F(M_n - \varepsilon) - F_n(M_n - \varepsilon)\big)$, pour tout $\varepsilon > 0$, et en comparant $F_n(M_n - \varepsilon)$ avec $\frac{1}{2}$, montrer qu'il existe un $\Omega_2 \in \mathfrak{F}$ tel que $P(\Omega_2) = 1$ et

$$\forall \omega \in \Omega_2,\ \forall \varepsilon > 0,\ \exists N_2 = N_2(\omega, \varepsilon),\ \forall n \geq N_2, \quad F(M_n(\omega) - \varepsilon) < \frac{1}{2} + \varepsilon. \tag{5}$$

4) Déduire de ce qui précède l'existence d'un $\Omega_0 \in \mathfrak{F}$ de probabilité 1 tel que

$$\forall \omega \in \Omega_0,\ \forall \varepsilon > 0,\ \exists N_0 = N_0(\omega, \varepsilon),\ \forall n \geq N_0, \quad F^{-1}\Big(\frac{1}{2}-\varepsilon\Big) < M_n(\omega) < F^{-1}\Big(\frac{1}{2}+\varepsilon\Big)+\varepsilon.$$

Conclure sur la convergence p.s. de M_n.

5) À quel endroit a-t-on vraiment besoin de la convergence *uniforme* presque sûre du théorème de Glivenko-Cantelli ?

Ex 3. Test de Kolmogorov-Smirnov

On dispose de 10 résultats de simulation de nombres aléatoires sur $[0, 1]$ (obtenus par `rand` en scilab par exemple) :

0,134 0,628 0,789 0,905 0,250 0,563 0,790 0,470 0,724 0,569.

Etudiez si cet échantillon conduit à rejeter l'hypothèse nulle selon laquelle le tirage a bien eu lieu selon la loi uniforme sur $[0, 1]$.

Ex 4. Estimation du paramètre d'une loi de Poisson *(extrait de l'examen du 28 juin 2008)*

1) La variable aléatoire X suit une loi de Poisson de paramètre $\alpha > 0$. Vérifiez par le calcul que $\mathbf{E}X = \alpha$ et $\operatorname{Var} X = \alpha$.

2) On se propose d'estimer la valeur inconnue de α au vu d'un échantillon de grande taille $X_1, \dots, X_n$ de variables aléatoires indépendantes et de même loi que X. On note $\overline{X}$ la moyenne empirique de l'échantillon (cette moyenne dépend de n). Que pouvez vous dire du comportement de cette moyenne lorsque n tend vers $= +\infty$?

3) Justifiez la convergence en loi suivante :

$$\sqrt{\frac{n}{\overline{X}}}(\overline{X}-\alpha) \xrightarrow[n\to+\infty]{\text{loi}} Z, \quad Z \sim \mathfrak{N}(0,1). \tag{6}$$

4) En déduire un intervalle de confiance I au niveau 95% pour α.

5) Le tableau ci-dessous donne un échantillon observé de taille 100.

1	2	3	5	4	2	2	1	0	1	3	3	4	2	6	4	0	3	2	1
3	5	1	2	1	1	4	3	0	2	0	2	4	1	2	4	3	2	2	2
3	2	2	1	4	1	2	3	2	2	3	4	5	3	1	5	2	2	2	2
1	2	3	1	1	3	2	2	3	0	3	4	2	1	0	3	3	1	2	3
1	1	4	1	5	2	5	2	5	2	2	0	2	4	1	6	3	2	5	2

a) Reproduisez et complétez le tableau suivant. Utilisez le pour dessiner la fonction de répartition empirique associée à cet échantillon. Indiquez sur le dessin la hauteur de chaque saut.

Valeur	0	1	2	3	4	5	6
Fréquence	0,07						

b) Calculez la moyenne empirique et la variance empirique de l'échantillon en indiquant comme vous faites pour éviter de rentrer une à une les 100 valeurs dans votre calculatrice.

c) Donnez l'intervalle de confiance I calculé à partir de cet échantillon et l'intervalle de confiance J (toujours au niveau 95%) obtenu en appliquant le théorème limite central avec autonormalisation.

Ex 5. Vitesse moyenne *(extrait de l'examen du 15 juin 2006)*
On veut estimer par intervalle de confiance la vitesse moyenne des automobiles dans un certain virage d'une route à grand trafic. Pour cela on a enregistré à l'aide d'un radar les vitesses $X_1(\omega) = x_1, \dots, X_{400}(\omega) = x_{400}$ de 400 automobiles en une période de temps de 2 heures avec des conditions de circulation homogènes (météo, visibilité, densité de trafic,...). On a obtenu les statistiques suivantes :

$$\sum_{i=1}^{400} x_i = 35\ 200 \text{ km/h}, \qquad \sum_{i=1}^{400} x_i^2 = 3\ 107\ 600 \text{ (km/h)}^2.$$

L'homogénéité des conditions de trafic permet de supposer que les variables aléatoires $X_1, \dots, X_{400}$, dont on a ainsi observé une réalisation, sont indépendantes et de même loi. Proposez un intervalle de confiance au niveau 98% pour la vitesse moyenne $\mathbf{E}X_1$ en indiquant clairement quels résultats du cours légitiment les approximations faites. Les données numériques ci-dessus ont été « arrangées » pour vous permettre de faire facilement tous les calculs à la main si vous ne disposez pas d'une calculatrice.

Ex 6. Convergence en loi *(extrait de l'examen du 15 juin 2006)*
Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$, indépendantes et de même loi telle que $\mathbf{E}X_1^2 = 1$ et $\mathbf{E}X_1 = 0$. On suppose

de plus que pour tout i, $X_i(\omega)$ n'est jamais nul. On définit alors sur $(\Omega, \mathfrak{F}, P)$ les suites de variables aléatoires $(Y_n)_{n\geq 1}$ et $(Z_n)_{n\geq 1}$ par

$$Y_n := \sqrt{n}\frac{X_1 + \cdots + X_n}{X_1^2 + \cdots + X_n^2}, \quad Z_n := \frac{X_1 + \cdots + X_n}{(X_1^2 + \cdots + X_n^2)^{1/2}}.$$

Montrez que $(Y_n)_{n\geq 1}$ et $(Z_n)_{n\geq 1}$ convergent en loi quand n tend vers l'infini vers une v.a. gaussienne de loi $\mathfrak{N}(0, 1)$.