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

IS-Math314 **Année 2006-2007**

Fiche n° 6

Ex 1. Un produit commercialisé est présenté dans des boites sur lesquelles on peut lire : *contenance 500 grammes*. La quantité contenue dans une boite choisie au hasard est en réalité une variable aléatoire gaussienne X.
A partir des données fournies, construire un intervalle de confiance pour la moyenne de X, au niveau de confiance 90%. Construire également un intervalle de confiance pour l'écart-type de X, avec le même niveau de confiance.

Données (en grammes) : 490 ; 490 ; 490 ; 492 ; 492 ; 495 ; 497 ; 497 ; 502 ; 505.

Ex 2. Dans un laboratoire, on a mis au point un nouvel appareil de mesure du rayonnement émis par un objet porté à haute température. De par la construction de l'appareil, on sait que les résultats des mesures sont gaussiens, avec pour moyenne la température de l'objet. Leur écart-type σ est lié à la qualité de l'appareil, et pour l'évaluer, on effectue avec six mesures sur un objet dont on connait la température : 1600 degrés.

1) On note $X_1, \dots, X_6$ les résultats des 6 mesures effectuées. Ecrire l'expression de la moyenne empirique et de la variance empirique des X_i. En utilisant le théorème de Student, déterminer un intervalle de confiance pour σ au niveau de confiance 99%. Quel est le défaut de cette méthode ?

2) A partir de la variable aléatoire $W = \sum_{i=1}^{6} \left(\frac{X_i - 1600}{\sigma} \right)^2$ construire un meilleur intervalle de confiance pour σ au niveau de confiance 99%.

Données : 1597.43 , 1597.96 , 1590.26 , 1590.46 , 1612.31 , 1610.40
Pour simplifier les calculs, on a calculé la somme de ces valeurs : 9598.82
et la somme de leurs carrés : 15356680.34

Ex 3. *Estimateur du maximum de vraisemblance*

Estimer par maximum de vraisemblance le paramètre θ

a) d'une loi de Poisson ;

b) d'une loi géométrique ;

c) d'une loi exponentielle ;

d) de la loi uniforme sur $[0, \theta]$.

Ex 4. *Estimation de la durée d'une panne*

Des étudiants font un TP en utilisant n ordinateurs reliés à un serveur[1]. Chaque ordinateur envoie au serveur, à des instants aléatoires, des requêtes variées (exécution de commande, transmission de données, etc).

A l'instant $t = 0$ ce serveur tombe en panne. Il redémarre au bout d'une durée τ.

Les étudiants ne savent pas quelle est la durée τ de la panne. Mais pour i de 1 à n, ils constatent au bout de quel temps T_i après le début de panne la première requête de l'ordinateur i est acceptée par le serveur. Pour estimer la durée τ de la panne, ils disposent donc de n données $T_1(\omega), \cdots, T_n(\omega)$.

Les requêtes ayant lieu à des instants aléatoires, il est naturel de supposer que la durée $X_i = T_i - \tau$ entre le redémarrage du serveur et la première requête de l'ordinateur i suit une loi exponentielle. Le paramètre a de cette loi exponentielle sera supposé connu et identique pour toutes les machines (que représente-t-il ?). On supposera également que les v.a. $X_1, \cdots, X_n$ sont indépendantes.

1) Quelle est la loi de la v.a. $V = \inf(X_1, \cdots, X_n)$?

2) Estimer la durée inconnue τ à partir de $T_1, \cdots, T_n$ en utilisant la méthode du maximum de vraisemblance. On note W l'estimateur obtenu.

3) Calculer le biais de W. En déduire un estimateur sans biais Z de la durée de panne. Quelle est la variance de ce nouvel estimateur ?

4) Fabriquer un autre estimateur sans biais de τ, en utilisant cette fois la moyenne empirique $\overline{T}$ des T_i. Quelle est sa variance ?

5) Des deux estimateurs sans biais de τ, lequel est le meilleur ? N'y a-t-il pas contradiction ici avec l'inégalité de Cramer-Rao ? Pourquoi ?

Entraînement supplémentaire facultatif :

Ces quatre exercices sont liés et doivent être étudiés dans l'ordre proposé.

Ex 5. *Médianes d'un échantillon observé*

1) Soit $x_1, \ldots, x_n$ une suite finie de réels, pas forcément distincts. On appelle *médiane* de cette suite tout réel m tel qu'au moins la moitié des x_i vérifie $x_i \geq m$ et au moins la moitié des x_i vérifie $x_i \leq m$:

$$\operatorname{card}\big\{i \in \{1, \ldots, n\};\ x_i \geq m\big\} \geq \frac{n}{2} \quad \text{et} \quad \operatorname{card}\big\{i \in \{1, \ldots, n\};\ x_i \leq m\big\} \geq \frac{n}{2}. \tag{1}$$

Trouver toutes les médianes des suites finies suivantes :
a) 2 1 3 ; b) 4 2 5 7 ; c) 4 2 4 7.

2) Exprimer la double condition (1) à l'aide de la fonction de répartition F_n de la mesure $\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$ (c'est la fonction de répartition empirique construite sur l'échantillon observé $x_1, \ldots, x_n$). Illustrer graphiquement la recherche de la ou des médianes à partir du graphe de F_n, dans le cas où les $x_1, \ldots, x_n$ sont tous distincts, avec n impair puis avec n pair.

[1]Concrètement, un gros ordinateur sans écran installé dans un local verrouillé, climatisé, et sous alarme. Le seul point important ici est que les étudiants ont la possibilité d'utiliser le serveur, mais pas de le voir.

Ex 6. *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 :

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

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 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 [2] de $\mathbb{R}$. On vérifiera que si $m' < m''$ sont deux médianes de X, tout $m \in [m', m'']$ est une médiane de X.

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

$$P(X < x) \leq \frac{1}{2} \leq P(X \leq x), \tag{3}$$

ce qui peut s'écrire à l'aide de la f.d.r. F de X sous la forme

$$F(x-) \leq \frac{1}{2} \leq F(x). \tag{4}$$

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 $F(x) = \frac{1}{2}$.

3) On rappelle que l'*inverse généralisé* F^{-1} de F, déjà étudié dans le chapitre sur la simulation, est donné par :

$$\forall u \in]0, 1[, \quad F^{-1}(u) := \inf\{t \in \mathbb{R};\ F(t) \geq u\}. \tag{5}$$

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

Ex 7. *Estimation de médiane*

Sur l'espace probabilisé $(\Omega,_t F, P)$, soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires réelles indépendantes de même loi de f.d.r. F. On suppose que F^{-1} définie par (5) est continue au point $\frac{1}{2}$. 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 $X_1, \ldots, X_n$ 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] L'ensemble vide est un intervalle, mais on verra à la question 3 que l'ensemble des médianes de X n'est jamais vide.

1) 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 $\Omega_1 \in \mathcal{F}$ tel que $P(\Omega_1) = 1$ et

$$\forall\omega \in \Omega_1,\ \forall\varepsilon > 0,\ \exists N_1 = N_1(\omega,\varepsilon),\ \forall n \geq N_1, \quad F(M_n(\omega)) > \frac{1}{2} - \varepsilon. \tag{6}$$

2) 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 $\Omega_2 \in \mathcal{F}$ tel que $P(\Omega_2) = 1$ et

$$\forall\omega \in \Omega_2,\ \forall\varepsilon > 0,\ \exists N_2 = N_2(\omega,\varepsilon),\ \forall n \geq N_2, \quad F(M_n(\omega) - \varepsilon) < \frac{1}{2} + \varepsilon. \tag{7}$$

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

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

Conclure sur la convergence p.s. de M_n.

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

Ex 8. *Médianes d'une loi (suite)*

On se replace dans le cadre de l'exercice 6 : X est une v.a. réelle définie sur $(\Omega, \mathcal{F}, P)$ et on appelle *médiane de* X, tout réel m vérifiant $P(X \geq m) \geq \frac{1}{2}$ et $P(X \leq m) \geq \frac{1}{2}$.

1) On a prouvé que l'ensemble des médianes d'une loi est un intervalle. Montrer que cet intervalle est *fermé* : on notera a et b les bornes inférieure et supérieure de cet intervalle et on montrera que ce sont aussi des médianes. Pour b utiliser la continuité à gauche de $t \mapsto P(X \geq t)$. L'intervalle $[a, b]$ est appelé *intervalle médian* de X.

2) On note F la f.d.r. de X et F^{-1} son inverse généralisé. Vérifier que toute solution de l'équation $F(x) = \frac{1}{2}$ est une médiane de X. Montrer que si cette équation a *au plus* une solution, il y a une unique médiane qui est $F^{-1}(1/2)$.

3) Montrer qu'il y a au plus une médiane m telle que $F(m) > \frac{1}{2}$ et donner les deux cas de figure correspondant à cette situation.

Ex 9. *Estimation du paramètre de localisation d'une loi de Cauchy*

La loi de Cauchy $\mathrm{Cau}(a, b)$ de paramètres a et b ($a \in \mathbb{R}$, $b \in \mathbb{R}_+^*$) a pour densité :

$$f(t) = \frac{1}{\pi b} \frac{1}{1 + \big(\frac{t-a}{b}\big)^2}.$$

1) Montrer que cette loi a une unique médiane que l'on calculera en fonction de a.

2) Déduire de l'exercice 7 un estimateur p.s. convergent de a basé sur l'observation d'un échantillon $X_1, \ldots, X_n$ de la loi $\mathrm{Cau}(a, b)$.