Mathématiques - 2 BCPST 1&2 - Lycée Michel Montaigne

DM N°16 - A remettre le mardi 6 mars 2012

« Estimation de paramètres »

PROBLÈME 1

Soit f la fonction définie sur $\mathbb{R}$ par :

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{\lambda}{\sqrt{1-x^2}} \text{ si } x \in]-1, 1[\quad \text{et} \quad f(x) = 0 \text{ sinon.}$$

1. Déterminer λ pour que f soit une densité de probabilité.

   Dans la suite du problème, on considère que cette condition est vérifiée et on note X une variable aléatoire réelle de densité f.

2. Déterminer F la fonction de répartition de X et tracer sa courbe représentative dans un repère orthonormé d'unités 4 cm.

3. Montrer que X admet une espérance et la calculer.

4. Montrer que X admet une variance et la calculer en utilisant le changement de variable : $x = \cos u$ que l'on justifiera de façon précise.

5. Déterminer et reconnaître la loi de la variable aléatoire $Y = \arcsin X$. Donner son espérance et sa variance.

6. Déterminer une densité de la variable aléatoire $Z = |X|$. Calculer son espérance et sa variance.

7. Soit $(\mathcal{C})$ un cercle de centre O et de rayon 1. Un point A est donné sur ce cercle $(\mathcal{C})$.
   On choisit au hasard un point B sur $(\mathcal{C})$ en supposant que l'angle $\theta = \left(\overrightarrow{OA}, \overrightarrow{OB}\right)$ est une variable aléatoire de loi Uniforme sur $[0, 2\pi]$. On note C le milieu de AB et L la variable aléatoire égale à la **distance** OC.
   (a) Montrer que $L = \left|\cos\left(\frac{\theta}{2}\right)\right|$.
   (b) Déterminer la loi, l'espérance et la variance de L.

PROBLÈME 2

Préliminaire

La plupart des expériences aléatoires conduisent à l'étude de variables aléatoires réelles obéissant à des lois dont le type est connu, mais qui dépendent d'un paramètre réel lié à l'expérience. Ce problème a pour objectif de donner des méthodes afin d'estimer la valeur numérique de ce paramètre, que nous désignerons dans la suite par la lettre θ.

Notations

Si X est une variable aléatoire, on notera, sous réserve d'existence, $\mathrm{E}(X)$ l'espérance de X et $\mathrm{Var}(X)$ sa variance ; F_X désignera la fonction de répartition de X et f_x une densité éventuelle. Soit n un entier naturel non nul, on appellera n-échantillon de X toute suite $(X_1, \dots, X_n)$ de variables aléatoires mutuellement indépendantes et suivant la même loi que X.
Si X est une variable aléatoire dont la loi dépend d'un paramètre θ, et $(X_1, \dots, X_n)$ un n-échantillon de X, une variable aléatoire T_n fonction de $(X_1, \dots, X_n)$ est un estimateur sans biais et convergent de θ si $\mathrm{E}(T_n) = \theta$ et $\lim_{n\to+\infty} \mathrm{Var}(T_n) = 0$. Le terme « estimateur » utilisé dans ce problème signifie par abus de langage que ces deux conditions sont vérifiées.
Enfin, on dira que l'estimateur T_n est meilleur que l'estimateur T'_n si pour tout n, entier naturel assez grand : $\mathrm{Var}(T_n) \leqslant \mathrm{Var}(T'_n)$.

Première partie : exemple introductif

Une population donnée contient une proportion inconnue, θ, d'individus possédant un certain caractère. On considère une variable aléatoire X suivant la loi de Bernoulli de paramètre θ. On prélève, avec remise, n individus de cette population, et on désigne par X_i la variable aléatoire associée au i-ème tirage. Enfin, pour tout n, entier naturel non nul, on note $T_n = \frac{X_1 + \ldots + X_n}{n}$

1. Rappelez les valeurs de $\mathrm{E}(X)$ et de $\mathrm{Var}(X)$.
2. Montrer que T_n est un estimateur de θ.

Deuxième partie

Soit θ un réel strictement positif ; soient X une variable aléatoire de loi uniforme sur l'intervalle $[0, \theta]$, et $(X_1, \ldots, X_n)$ un n-échantillon de X.

1. (a) Déterminer une densité f_X de la variable aléatoire X, ainsi que sa fonction de répartition F_X, puis exprimer, en fonction de θ, les valeurs de $\mathrm{E}(X)$ et de $\mathrm{Var}(X)$.

   (b) Pour tout n, entier naturel non nul, on pose $T_n = \frac{2}{n}(X_1 + \ldots + X_n)$; montrer que T_n est un estimateur de θ.

   On se propose dans la suite de cette deuxième partie, de construire d'autres estimateurs de θ liés à cette dernière variable aléatoire T_n.

2. Soit n un entier naturel non nul, on considère les deux variables aléatoires

$$Y = \max(X_1, \ldots, X_n) \quad \text{et} \quad Z = \min(X_1, \ldots, X_n)$$

   (a) Déterminer en fonction de θ les expressions de F_Y, f_Y, $\mathrm{E}(Y)$ et $\mathrm{Var}(Y)$.

   (b) Déterminer en fonction de θ les expressions de F_Z, f_Z, $\mathrm{E}(Z)$ et $\mathrm{Var}(Z)$.
   Indication : Pour déterminer F_Z, on pourra calculer $P(Z > z)$.

   (c) On remarque que $\mathrm{Var}(Z) = \mathrm{Var}(Y)$. Justifier cette égalité sans recourir au calcul.

3. Pour $n \in \mathbb{N}^*$, on considère la variable aléatoire $T'_n = \frac{n+1}{n} Y$.

   (a) Montrer que T'_n est un estimateur de θ.

   (b) T'_n est-il un meilleur estimateur que l'estimateur T_n défini à la question 1b ?

4. Pour $n \in \mathbb{N}^*$, on considère la variable aléatoire $T''_n = Y + Z$.

   (a) On rappelle que la covariance de Y et de Z, notée $\mathrm{Cov}(Y, Z)$ vérifie :

$$|\mathrm{Cov}(Y, Z)| \leqslant \sqrt{\mathrm{Var}(Y)\,\mathrm{Var}(Z)}$$

   En déduire que $\mathrm{Var}(T''_n) \leqslant 4\,\mathrm{Var}(Y)$.

   (b) Déduire du 4a et du 1b que T''_n est un estimateur de θ, meilleur que T_n.

Deuxième Partie : Loi de Pareto

1. **La loi $\boldsymbol{\gamma(p, \lambda)}$**

   p et λ désignent dans cette partie deux réels strictement positifs. On désigne par $I(p, \lambda)$ l'intégrale :

$$I(p, \lambda) = \int_0^{+\infty} \lambda^p x^{p-1} e^{-\lambda x} dx$$

   et on note $\Gamma(p) = I(p, 1)$.

   (a) Montrer que pour tout x, réel strictement positif,

$$x^{p-1} e^{-\lambda x} \leqslant x^{p-1}$$

   puis montrer que pour tout a réel, strictement positif, $\int_0^a x^{p-1} dx$ converge. En déduire la convergence de l'intégrale $I(p, \lambda)$ à la borne zéro.

(b) Montrer que pour x suffisamment grand,

$$0 \leqslant x^{p-1}e^{-\lambda x} \leqslant \frac{1}{x^2}$$

puis montrer que pour tout a réel, strictement positif, $\displaystyle\int_a^{+\infty} \frac{dx}{x^2}$ converge. En déduire la convergence de l'intégrale $I(p,\lambda)$ à la borne $+\infty$.

(c) En effectuant le changement de variable $x = \frac{u}{\lambda}$, montrer que $I(p,\lambda) = \Gamma(p)$.

(d) Calculer $\Gamma(1)$ et montrer à l'aide d'une intégration par parties, que $\Gamma(p+1) = p\Gamma(p)$. En déduire que pour tout n, entier naturel non nul, $\Gamma(n) = (n-1)!$

(e) Montrer que la fonction g définie par

$$\begin{cases} g(u) = 0 & \text{si } u < 0 \\ g(u) = \dfrac{1}{\Gamma(p)}\lambda^p u^{p-1}e^{-\lambda u} & \text{sinon} \end{cases}$$

est la densité de probabilité d'une variable aléatoire U, puis vérifier que $\mathrm{E}\,(U) = \frac{p}{\lambda}$ et $\mathrm{Var}\,(U) = \frac{p}{\lambda^2}$.

Notation : On dira dans la suite du problème qu'une telle variable aléatoire U suit la loi $\gamma(p,\lambda)$.

2. Soient θ un réel strictement positif et X la variable aléatoire de densité f_X définie comme suit :

$$\begin{cases} f_X(x) = 0 & \text{si } x < 1 \\ f_X(x) = \frac{1}{\theta}x^{-\frac{\theta+1}{\theta}} & \text{sinon} \end{cases} \qquad \textit{(loi de Pareto)}$$

(a) Vérifier que f_X est bien une densité de probabilité et déterminer F_X.

(b) Pour quelles valeurs de θ la variable aléatoire X admet-elle une espérance ? Une variance ? Calculer $\mathrm{E}\,(X)$ et $\mathrm{Var}\,(X)$ lorsque ces valeurs existent.

(c) Soit $Y = \ln(X)$. On admettra que Y est une variable aléatoire. Déterminer en fonction de θ les expressions de F_Y, f_Y, $\mathrm{E}\,(Y)$ et $\mathrm{Var}\,(Y)$.

(d) On considère un n-échantillon $(Y_1,\ldots,Y_n)$ de la variable aléatoire Y définie ci-dessus et on pose $S_n = Y_1 + \ldots + Y_n$ et $T_n = \frac{S_n}{n}$; montrer que T_n est un estimateur de θ.

(e) **On admettra le résultat de cours suivant :**

> Pour tout couple (U,V) de variables aléatoires indépendantes de densités respectives f_U et f_V la variable $W = U + V$ admet une densité f_W définie par :
>
> $$\forall w \in \mathbb{R}, \quad f_W(w) = \int_{-\infty}^{+\infty} f_U(w-t)f_V(t)dt$$

Vérifier que : si U et V sont à valeurs positives, alors :

$$\forall w \in \mathbb{R}, \quad f_W(w) = \int_0^w f_U(w-t)f_V(t)dt$$

(f) On considère deux variables aléatoires indépendantes suivant la même loi exponentielle de paramètre λ notée $\mathcal{E}(\lambda)$. Déterminer une densité de leur somme, puis, plus généralement, montrer qu'une densité de la variable aléatoire S_n définie au 2d est :

$$f_n(s) = \frac{s^{n-1}}{\theta^n(n-1)!}e^{-\frac{s}{\theta}} \quad \text{pour } s > 0 \quad \text{et 0 sinon.}$$

Quelle est la loi suivie par S_n ?