BANQUE COMMUNE D'ÉPREUVES

Code sujet : 283

Conception : HEC Paris – ESCP BS

OPTION SCIENTIFIQUE

MATHÉMATIQUES II

Mercredi 29 avril 2020, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
Les candidats sont invités à ***encadrer*** *dans la mesure du possible les résultats de leurs calculs.*
Aucun document n'est autorisé. ***L'utilisation de toute calculatrice et de tout matériel électronique est interdite.*** *Seule l'utilisation d'une règle graduée est autorisée.*
Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

Lorsque l'on cherche à estimer un paramètre inconnu à partir d'un échantillon de données, on appelle statistique exhaustive toute fonction de ces données qui résume à elle seule l'information que ces données fournissent sur le paramètre.
On donne ici une définition précise de cette notion d'exhaustivité dans le cas des échantillons de variables aléatoires discrètes, illustrée de plusieurs exemples qui en montrent l'intérêt.

On s'intéressera dans ce problème à l'estimation d'un paramètre réel inconnu θ appartenant à un intervalle Θ.

On dispose pour cela de plusieurs *observations* $x_1, \ldots, x_n$ considérées comme les réalisations de variables aléatoires discrètes $X_1, \ldots, X_n$ définies sur le même espace probabilisable $(\Omega, \mathcal{A})$, à valeurs dans une partie B de $\mathbb{N}$.

L'espace probabilisable $(\Omega, \mathcal{A})$ est muni d'une famille $(P^\theta)_{\theta\in\Theta}$ de probabilités indexées par le paramètre θ.

On fait, pour toutes les valeurs du paramètre θ, les trois hypothèses suivantes.

- Les variables aléatoires $X_1, \ldots, X_n$ sont mutuellement indépendantes, c'est-à-dire :

$$\forall (x_1, \ldots, x_n) \in B^n, \quad P^\theta\Big(\bigcap_{i=1}^{n} [X_i = x_i]\Big) = \prod_{i=1}^{n} P^\theta([X_i = x_i]) \tag{1}$$

- Les variables aléatoires $X_1, \ldots, X_n$ suivent toutes la même loi qu'une variable aléatoire de référence, notée X, à valeurs dans B, c'est-à-dire :

$$\forall i \in [\![1, n]\!], \quad \forall x \in B, \quad P^\theta([X_i = x]) = P^\theta([X = x]) \tag{2}$$

Tournez la page S.V.P.

• Tous les éléments de B sont des valeurs effectivement possibles de X, c'est-à-dire :

$$\forall x \in B, \quad P^\theta([X = x]) > 0 \tag{3}$$

On appelle *statistique* toute variable aléatoire S de la forme $\omega \longmapsto s\big(X_1(\omega), \ldots, X_n(\omega)\big)$, où s désigne une application définie sur B^n et à valeurs réelles. On note alors $S = s(X_1, \ldots, X_n)$.

Pour tout $\theta \in \Theta$, on note $E^\theta(S)$ l'espérance de S lorsque $(\Omega, \mathcal{A})$ est muni de la probabilité P^θ (si cette espérance existe). On note de même $V^\theta(S)$ la variance de S (si elle existe).

Partie 1 : développements en série

1. *Dans cette question*, x désigne un nombre réel strictement compris entre 0 et 1.

   a) Justifier la convergence de la série $\displaystyle\sum_{k\geq 1} \frac{x^k}{k}$.

   b) Vérifier, pour tout $m \in \mathbb{N}^*$ et tout $t \in]0,1[$, l'égalité :

   $$\frac{1}{1-t} = \frac{t^m}{1-t} + \sum_{k=0}^{m-1} t^k \ .$$

   c) Démontrer que l'intégrale $\displaystyle\int_0^x \frac{t^m}{1-t}\,\mathrm{d}t$ tend vers 0 quand l'entier m tend vers l'infini.

   d) En déduire la somme de la série $\displaystyle\sum_{k\geq 1} \frac{x^k}{k}$.

2. *Dans cette question*, indépendante de la précédente, $(a_k)_{k\in\mathbb{N}}$ désigne une suite de nombres réels telle que la série $\displaystyle\sum_{k\geq 0} a_k c^k$ est absolument convergente pour un réel strictement positif c.

   a) Justifier que la fonction $f : x \longmapsto a_0 + \displaystyle\sum_{k=1}^{+\infty} a_k x^k$ est bien définie sur le segment $[-c, +c]$.

   b) Pour un entier naturel m, on pose : $\quad M_m = \displaystyle\sum_{k=m+1}^{+\infty} |a_k|\, c^{k-m-1}$.

   Justifier, pour tout $x \in [-c, +c]$, l'inégalité :

   $$\Big| \sum_{k=m+1}^{+\infty} a_k x^k \Big| \leq M_m\, |x|^{m+1} \ .$$

   c) Justifier, pour tout $m \in \mathbb{N}^*$, le développement limité au voisinage de 0 :

   $$f(x) = a_0 + \sum_{k=1}^{m} a_k x^k \ + o(x^m) \ .$$

   d) Démontrer que si la fonction f est nulle sur l'intervalle $]0, +c]$, alors $(a_k)_{k\in\mathbb{N}}$ est la suite nulle.

Dans toute la suite du problème, pour tout $\theta \in \Theta$ et tout $(x_1, \ldots, x_n) \in B^n$, on note :

$$L(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} P^\theta([X_i = x_i]) \tag{4}$$

Cette quantité, qui s'écrit aussi $\prod_{i=1}^{n} P^\theta([X = x_i])$ d'après (2), est appelée la *vraisemblance* de la valeur θ du paramètre au vu des observations $x_1, \ldots, x_n$.

Partie II : estimateur du maximum de vraisemblance, un exemple

Dans cette partie, Θ est l'intervalle ouvert $]0, 1[$, B est égal à $\mathbb{N}^*$ et on a :

$$\forall x \in B, \quad P^\theta([X = x]) = (1-\theta)^{x-1}\,\theta.$$

On note $\overline{X}$ la variable aléatoire $\dfrac{1}{n}\sum_{i=1}^{n} X_i$.

3. Soit $\theta \in \Theta$.
   a) Reconnaître la loi de X lorsque $(\Omega, \mathcal{A})$ est muni de la probabilité P^θ.
   b) En déduire que $\overline{X}$ est un estimateur sans biais du paramètre $1/\theta$.
   c) Quel est le risque quadratique de cet estimateur ?

4. On note T la variable aléatoire $\dfrac{1}{n}\sum_{i=1}^{n} \dfrac{1}{X_i}$.
   a) En utilisant le résultat de la question 1.*d*, justifier que :

$$\forall \theta \in \Theta, \quad E^\theta(T) = \frac{\theta \ln(\theta)}{\theta - 1} .$$

   b) En déduire que T est un estimateur de θ dont le biais $b_\theta(T)$ est strictement positif.

5. Soit $(x_1, \ldots, x_n) \in B^n$.
   a) Justifier, pour tout $\theta \in \Theta$, l'égalité :

$$\ln\big(L(x_1, \ldots, x_n, \theta)\big) = n\ln(\theta) - \big(n - \sum_{i=1}^{n} x_i\big)\ln(1-\theta) .$$

   b) En déduire que, lorsque les x_i ne sont pas tous égaux à 1, le nombre $\dfrac{n}{\sum_{i=1}^{n} x_i}$ est l'unique valeur de θ qui maximise la vraisemblance $L(x_1, \ldots, x_n, \theta)$.

6. On note U la variable aléatoire $\dfrac{n}{\sum_{i=1}^{n} X_i}$.
   a) Établir, pour tout $\theta \in \Theta$ et tout entier $k \geq n$, l'égalité :

$$\frac{n}{k} = \theta - \theta^2\Big(\frac{k}{n} - \frac{1}{\theta}\Big) + \int_{1/\theta}^{k/n} \Big(\frac{k}{n} - t\Big)\,\frac{2}{t^3}\,\mathrm{d}t .$$

   b) En déduire que U est un estimateur de θ dont le biais $b_\theta(U)$ est donné par :

$$\forall \theta \in \Theta, \quad b_\theta(U) = \sum_{k=n}^{+\infty} P\Big(\Big[\sum_{i=1}^{n} X_i = k\Big]\Big) \int_{1/\theta}^{k/n} \Big(\frac{k}{n} - t\Big)\,\frac{2}{t^3}\,\mathrm{d}t .$$

   c) Justifier que $b_\theta(U)$ est strictement positif, quelle que soit la valeur du paramètre θ.

Tournez la page S.V.P.

7. *Dans cette question*, on suppose que le nombre des observations est illimité. On dispose donc, pour estimer le paramètre θ, d'une suite $(X_n)_{n\in\mathbb{N}^*}$ de variables aléatoires mutuellement indépendantes et de même loi.

   Pour tout entier $n \in \mathbb{N}^*$, on note $T_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{1}{X_i}$ et $U_n = \dfrac{n}{\displaystyle\sum_{i=1}^{n} X_i}$.

   Étudier la convergence des deux suites d'estimateurs $(T_n)_{n\in\mathbb{N}^*}$ et $(U_n)_{n\in\mathbb{N}^*}$ du paramètre θ.

Dans toute la suite du problème, on dit qu'une statistique $S = s(X_1,\dots,X_n)$ est *exhaustive* s'il existe une application g de $s(B^n)\times\Theta$ dans $\mathbb{R}_+$ et une application h de B^n dans $\mathbb{R}_+$ telles que :

$$\forall\theta\in\Theta,\ \forall(x_1,\dots,x_n)\in B^n,\ L(x_1,\dots,x_n,\theta) = g\big(s(x_1,\dots,x_n),\theta\big)\ h(x_1,\dots,x_n) \tag{5}$$

Partie III : statistique exhaustive, un exemple

Dans cette partie, on suppose que $B = \{0,1\}$, $\Theta =]0,1[$ et que, quel que soit $\theta\in\Theta$, les variables aléatoires $X_1,\dots,X_n$ suivent la loi de Bernoulli de paramètre θ, lorsque l'espace probabilisable $(\Omega,\mathcal{A})$ est muni de la probabilité P^θ.

On pose : $S = \displaystyle\sum_{i=1}^{n} X_i$.

8. a) Démontrer que la vraisemblance de n'importe quelle valeur $\theta\in\Theta$ du paramètre est donnée par :

$$\forall(x_1,\dots,x_n)\in\{0,1\}^n,\quad L(x_1,\dots,x_n,\theta) = \theta^{\left(\sum_{i=1}^{n} x_i\right)}\times(1-\theta)^{\left(\sum_{i=1}^{n}(1-x_i)\right)}.$$

   b) En déduire que la statistique S est exhaustive.

9. Soit $k\in[\![0,n]\!]$ et $(x_1,\dots,x_n)\in\{0,1\}^n$.

   a) Calculer la probabilité conditionnelle $P^\theta_{[S=k]}\big([X_1=x_1]\cap\dots\cap[X_n=x_n]\big)$ et vérifier que la loi conditionnelle du vecteur aléatoire $(X_1,\dots,X_n)$ sachant l'événement $[S=k]$ ne dépend pas du paramètre θ.

   b) Établir, pour tout $\theta\in\Theta$, l'égalité : $P^\theta_{[S=k]}([X_1=1]) = \dfrac{k}{n}$.

10. Le script *Scilab* suivant permet d'effectuer des simulations, qu'il place dans une matrice Y, dont il évalue ensuite la moyenne de chaque colonne.

```
--> theta=0.3;
--> N=100000;
--> n=10;
--> k=4;

--> U=grand(n,N,'bin',1,theta);
--> S=sum(U,'r'); // somme des lignes de U, colonne par colonne
--> K=find(S==k); // recherche des coefficients de S égaux à k
```

```
--> Y=U(1:n,K);

--> M=mean(Y,'c') // moyenne des colonnes de Y, ligne par ligne
ans  =
   0.4019917
   0.4042436
   0.4008908
   0.3962868
   0.4054947
   0.3953861
   0.3990892
   0.4002402
   0.3941851
   0.4021919
```

a) Décrire avec précision ce que représente une colonne de la matrice U.

b) Expliquer pourquoi les coefficients de Y fournissent une simulation d'une loi conditionnelle du vecteur $(X_1, \ldots, X_n)$.

c) Commenter les résultats trouvés pour les coefficients de M.

11. Á la suite du script précédent, on exécute l'instruction suivante :

```
--> C=Y*Y'/length(K);
```

a) Donner le format de la matrice C et indiquer la valeur de son coefficient C(1,1).

b) Á quelle valeur approchée peut-on s'attendre pour C(1,2) et pour les autres coefficients non diagonaux de la matrice C ?

c) Quelle est la somme totale des coefficients de la matrice C ?

Partie IV : inégalité de Rao-Blackwell

Dans cette partie, on reprend les hypothèses générales du préambule et on considère une statistique exhaustive $S = s(X_1, \ldots, X_n)$, au sens donné par (5).

On admet que, pour tout élément u de $s(B^n)$ et tout élément $(x_1, \ldots, x_n)$ de B^n, la probabilité conditionnelle $P^{\theta}_{[S=u]}\big([X_1 = x_1] \cap \cdots \cap [X_n = x_n]\big)$ ne dépend pas de θ.

12. Soit T un estimateur sans biais du paramètre θ.

a) Démontrer que, pour tout $u \in s(B^n)$, l'espérance conditionnelle $E^{\theta}_{[S=u]}(T)$ existe et que sa valeur ne dépend pas de θ.

b) Justifier que $\big([S = u]\big)_{u \in s(B^n)}$ est un système complet d'événements.

13. Comme l'espérance conditionnelle $E^{\theta}_{[S=u]}(T)$ ne dépend pas de la valeur de θ, on peut la noter $E_{[S=u]}(T)$ et définir une application r de B^n dans $\mathbb{R}$ par :

$$\forall\, (x_1, \ldots, x_n) \in B^n, \quad r(x_1, \ldots, x_n) = E_{[S=s(x_1,\ldots,x_n)]}(T)\ .$$

a) En utilisant la formule de l'espérance totale, démontrer que $R = r(X_1, \ldots, X_n)$ est un estimateur sans biais de θ.

b) On suppose que T admet une variance, quelle que soit la valeur du paramètre θ. Justifier qu'il en est de même pour R et en utilisant les inégalités

$$\big(E_{[S=u]}(T - \theta)\big)^2 \leq E_{[S=u]}\big((T - \theta)^2\big)$$

établir, pour tout $\theta \in \Theta$, l'inégalité (appelée inégalité de Rao-Blackwell) :

$$V^{\theta}(R) \leq V^{\theta}(T)\ .$$

Tournez la page S.V.P.

14. Un exemple d'estimateur sans biais optimal

Dans cette question uniquement, on suppose que $B = \mathbb{N}$, $\Theta =]0, +\infty[$ et que, pour tout $\theta \in \Theta$, la loi commune des variables aléatoires $X_1, \ldots, X_n$ sur l'espace probabilisé $(\Omega, \mathcal{A}, P^\theta)$ est la loi de Poisson de paramètre θ.

a) Justifier que la statistique $S = \sum_{i=1}^{n} X_i$ est exhaustive.

b) Soit $u \in \mathbb{N}$ et $(x_1, \ldots, x_n) \in \mathbb{N}^n$.
Vérifier que la probabilité conditionnelle $P^\theta_{[S=u]}\big([X_1 = x_1] \cap \cdots \cap [X_n = x_n]\big)$ ne dépend pas de θ.

c) Soit $u \in \mathbb{N}$.
Démontrer que chacune des variables aléatoires $X_1, \ldots, X_n$ suit une loi binomiale lorsque l'espace probabilisable $(\Omega, \mathcal{A})$ est muni de la probabilité $P^\theta_{[S=u]}$. Sont-elles indépendantes pour cette probabilité ?

d) Trouver une suite réelle $(\varphi_k)_{k\in\mathbb{N}}$ telle que

$$\forall \theta > 0, \quad \sum_{k=0}^{+\infty} \varphi_k \frac{(n\theta)^k}{k!} = \theta\, \mathrm{e}^{n\theta}$$

et en prouver l'unicité à l'aide du résultat de la question 2.

e) En exploitant le résultat de la question 13, démontrer que, parmi les estimateurs sans biais de θ, l'estimateur $\dfrac{1}{n}\sum_{i=1}^{n} X_i$ est optimal, c'est-à-dire que son risque quadratique est inférieur ou égal à celui de tout autre estimateur sans biais de θ.

IMPRIMERIE NATIONALE – 20 1070 – D'après documents fournis