



CENTRE SCOLAIRE OZANAM

Internat et externat pour lycéens et étudiants – Etudes encadrées et soutien scolaire – Stages intensifs de révision

60 rue Vauban 69006 LYON ☎ 04 78 52 27 99 / Fax : 04 78 52 11 15

✉ contact@ozanam-lyon.fr 🌐 www.ozanamlyon.fr

Concours BCE : Epreuve de Mathématiques II (option Scientifique)

Conception HEC Paris-ESCP Europe : 29 avril 2020

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'intéresse 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, \dots, x_n , considérées comme les réalisations de variables aléatoires discrètes X_1, \dots, X_n définies sur le même espace probabilisable (Ω, \mathcal{A}) à valeurs dans une partie B de \mathbb{N} .

L'espace probabilisable (Ω, \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, \dots, X_n sont mutuellement indépendantes, c'est-à-dire :

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

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

$$\forall i \in \llbracket 1, n \rrbracket, \forall x \in B, P^\theta ([X_i = x]) = P^\theta ([X = x]) \quad (2)$$

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

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

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

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

Partie 1. Développement en série

Dans les questions 1 à 4, x désigne un nombre réel strictement compris entre 0 et 1.

1. Justifier la convergence de la série $\sum_{k \geq 1} \frac{x^k}{k}$
2. Vérifier, pour tout $m \in \mathbb{N}^*$ et tout $t \in]0, 1[$, l'égalité :

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

3. Démontrer que :

l'intégrale $\int_0^x \frac{t^m}{1-t} dt$ tend vers 0 quand l'entier m tend vers l'infini

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

Dans les questions 5 à 8, indépendantes des précédentes, $(a_k)_{k \in \mathbb{N}}$ désigne une suite de nombres réels telle que la série $\sum_{k \geq 0} a_k c^k$ est absolument convergente pour un réel strictement positif c .

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

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

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

$$\left| \sum_{k=m+1}^{+\infty} a_k x^k \right| \leq M_m |x|^{m+1}$$

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

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

8. 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 $(x_1, \dots, x_n) \in B^n$, on note :

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

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, \dots, x_n .

Partie 2. 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 :

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

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

Soit $\theta \in \Theta$.

9. Reconnaître la loi de X lorsque (Ω, \mathcal{A}) est muni de la probabilité P^θ .

10. En déduire que \bar{X} est un estimateur sans biais du paramètre $1/\theta$.

11. Quel est le risque quadratique de cet estimateur ?

On note T la variable aléatoire $\frac{1}{n} \sum_{i=1}^n \frac{1}{X_i}$.

12. En utilisant le résultat de la question 4, justifier que :

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

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

Soit $(x_1, \dots, x_n) \in B^n$.

14. Justifier, pour tout $\theta \in \Theta$, l'égalité :

$$\ln(L((x_1, \dots, x_n), \theta)) = n \ln(\theta) - \left(n - \sum_{i=1}^n x_i \right) \ln(1 - \theta)$$

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

On note U la variable aléatoire $\frac{n}{\sum_{i=1}^n X_i}$.

16. Etablir, pour tout $\theta \in \Theta$ et tout entier $k \geq n$, l'égalité :

$$\frac{n}{k} = \theta - \theta^2 \left(\frac{k}{n} - \frac{1}{\theta} \right) + \int_{1/\theta}^{k/n} \left(\frac{k}{n} - t \right) \frac{2}{t^3} dt$$

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

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

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

Dans la question qui suit, on suppose que le nombre des observations est illimité. On dispose donc, pour estimer le paramètre θ , d'une autre 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 $T_n = \frac{1}{n} \sum_{i=1}^n \frac{1}{X_i}$ et $U_n = \frac{n}{\sum_{i=1}^n X_i}$

19. Etudier 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 :

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

Partie 3. 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 (Ω, \mathcal{A}) est muni de la probabilité P^θ .

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

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

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

21. En déduire que la statistique S est exhaustive.

Soit $k \in \llbracket 0, n \rrbracket$ et $(x_1, \dots, x_n) \in \{0,1\}^n$.

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

23. Etablir que pour tout $\theta \in \Theta$ l'égalité :

$$P_{[S=k]}^\theta([X_1 = 1]) = \frac{k}{n}$$

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
```

24. Décrire avec précision ce que représente une colonne de la matrice U .
25. Expliquer pourquoi les coefficients de Y fournissent une simulation d'une loi conditionnelle de vecteur (X_1, \dots, X_n) .
26. Commenter les résultats trouvés pour les coefficients de M .

A la suite du script précédent, on exécute l'instruction suivante :

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

27. Donner le format de la matrice C et indiquer la valeur de son coefficient $C(1,1)$.
28. A quelle valeur approchée peut-on s'attendre pour $C(1,2)$ et pour les autres coefficients non diagonaux de la matrice C ?
29. Quelle est la somme totale des coefficients de la matrice C ?

Partie 4. Inégalité de Rao-Blackwell

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

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

Soit T un estimateur sans biais du paramètre θ .

30. Démontrer que, pour tout $u \in s(B^n)$, l'espérance conditionnelle $E_{[S=u]}^\theta(T)$ existe et que sa valeur ne dépend pas de θ .
31. Justifier que $([S = u])_{u \in s(B^n)}$ est un système complet d'évènements.

Comme l'espérance conditionnelle $E_{[S=u]}^\theta(T)$ 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 :

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

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

On suppose que T admet une variance, quelle que soit la valeur du paramètre θ .

33. Justifier qu'il en est de même pour R et en utilisant les inégalités :

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

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

$$V^\theta(R) \leq V^\theta(T)$$

Un exemple d'estimateur sans biais optimal

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

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

Soit $u \in \mathbb{N}$ et $(x_1, \dots, x_n) \in \mathbb{N}^n$.

36. Vérifier que la probabilité conditionnelle $P_{[S=u]}^\theta([X_1 = x_1] \cap \dots \cap [X_n = x_n])$ ne dépend pas de θ .

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

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

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

et en prouver l'unicité à l'aide du résultat des questions 5 à 8.

39. En exploitant les résultats des questions 32 à 34, démontrer que, parmi les estimateurs sans biais de θ , l'estimateur $\frac{1}{n} \sum_{i=1}^n X_i$ est optimal, c'est-à-dire que son risque quadratique est inférieur ou égal à celui de tout autre estimateur sans biais de θ .