



PRÉPASIMPLE

prepasimple.fr

ESSEC

Mathématiques 2 — Option économique

2019

Mathématiques appliquées

Mathématiques appliquées — Prépa ECG

Annales de concours · Sujet

On étudie la notion d'entropie d'une variable aléatoire, puis l'information que deux variables aléatoires s'apportent mutuellement. La première partie concerne les variables à densité; les deuxième et troisième parties concernent les variables discrètes.

Toutes les variables aléatoires intervenant dans le problème sont définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Pour toute variable aléatoire Y , on note $E(Y)$ son espérance lorsqu'elle existe.

Première partie : Entropie différentielle d'une variable à densité

1. La fonction logarithme de base 2, notée $\log_2$, est définie sur $\mathbb{R}_+^*$ par $\log_2(x) = \frac{\ln x}{\ln 2}$.

1. a) Montrer que, pour tout $(x, y) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$, on a $\log_2(xy) = \log_2 x + \log_2 y$.

1. b) Vérifier que, pour tout réel α , $\log_2(2^\alpha) = \alpha$.

1. c) Montrer que la fonction $\log_2$ est concave sur $\mathbb{R}_+^*$.

2. Soit X une variable aléatoire réelle à densité, et soit f une densité de X . On appelle support de f l'ensemble $I = \{x \in \mathbb{R} : f(x) > 0\}$, et on suppose que I est un intervalle de $\mathbb{R}$ d'extrémités a et b , avec $a < b$, ces extrémités pouvant être finies ou infinies. L'entropie différentielle de X est, sous réserve de convergence absolue, le réel

$$h(X) = - \int_a^b f(x) \log_2(f(x)) dx.$$

Montrer que $h(X) = -E(\log_2(f(X)))$. Dans cette expression, on peut attribuer la valeur 0 à la fonction $\log_2 f$ hors de I , ensemble auquel X appartient avec probabilité nulle.

3. Soit X une variable aléatoire de densité f , de support I , intervalle de $\mathbb{R}$ d'extrémités a et b . On suppose que X admet une entropie différentielle.

3. a) Soit $c \in \mathbb{R}$ et soit $Y = c + X$.

3. a) i) Déterminer une densité de Y .
3. a) ii) Justifier l'existence de l'entropie différentielle $h(Y)$ et la déterminer en fonction de $h(X)$.
3. b) Soit $\alpha > 0$ et soit $Z = \alpha X$.
3. b) i) Déterminer une densité de Z .
3. b) ii) Justifier l'existence de l'entropie différentielle $h(Z)$ et la déterminer en fonction de $h(X)$.
4. On détermine dans cette question l'entropie différentielle de quelques variables aléatoires suivant des lois classiques.
4. a) Soit $a > 0$. On considère une variable aléatoire X de loi uniforme sur $[0, a]$.
4. a) i) Donner une densité de X .
4. a) ii) Justifier l'existence de $h(X)$ et la déterminer.
4. a) iii) Déterminer une condition nécessaire et suffisante sur a pour que $h(X) > 0$.
4. b) On considère une variable aléatoire Y de loi normale centrée réduite. Montrer que Y admet une entropie différentielle et que $h(Y) = \frac{1}{2} \log_2(2\pi e)$.
4. c) On considère une variable aléatoire Z de loi exponentielle de paramètre $\lambda > 0$. Justifier l'existence de $h(Z)$ et la déterminer.
4. d) Soit f définie sur $\mathbb{R}$ par $f(x) = \frac{1}{2} \lambda e^{-\lambda|x|}$, où $\lambda > 0$.

4. d) i) Montrer que f est une densité de probabilité sur $\mathbb{R}$.
4. d) ii) Soit W une variable aléatoire de densité f . Justifier l'existence de $h(W)$ et la déterminer.
5. On dit qu'un couple (X, Y) de variables aléatoires est un couple gaussien centré si, pour tout $(\alpha, \beta) \in \mathbb{R}^2$, $\alpha X + \beta Y$ est une variable de loi normale centrée au sens suivant : il existe $\gamma \in \mathbb{R}$ et une variable Z de loi normale centrée réduite tels que $\alpha X + \beta Y$ ait la même loi que γZ . Le cas $\gamma = 0$ désigne la variable constante nulle. On considère un tel couple (X, Y) et on note σ^2 la variance de X . On suppose $\sigma^2 > 0$.
5. a) Montrer que X suit une loi normale centrée.
5. b) Calculer $h(X)$.
5. c) On suppose désormais que X et Y suivent la même loi normale centrée de variance σ^2 . On admet que les propriétés usuelles de l'espérance des variables discrètes se généralisent aux variables aléatoires quelconques, sous les conditions d'existence requises.
5. c) i) Montrer que $E(XY)$ existe.
5. c) ii) Montrer que, pour tout $\lambda \in \mathbb{R}$, $\lambda^2 E(Y^2) + 2\lambda E(XY) + E(X^2) \geq 0$.
5. c) iii) En déduire que $E(XY)^2 \leq E(X^2)E(Y^2)$.
5. c) iv) On pose $\rho = \frac{E(XY)}{\sigma^2}$. Montrer que $\rho \in [-1, 1]$.
5. c) v) Que vaut ρ si X et Y sont indépendantes ?
5. d) On suppose $|\rho| < 1$. On appelle entropie jointe du couple (X, Y) le réel défini par
- $$h(X, Y) = \log_2 \left(2\pi e \sigma^2 \sqrt{1 - \rho^2} \right).$$

5. d) i) À quelle condition $h(X, Y)$ est-elle nulle ?
5. d) ii) L'information mutuelle de X et Y est définie par $I(X, Y) = h(X) + h(Y) - h(X, Y)$. Calculer $I(X, Y)$.
5. d) iii) Montrer que $I(X, Y) \geq 0$.
5. d) iv) Quelle est la limite de $I(X, Y)$ quand ρ tend vers 1 ?

Deuxième partie : Généralités sur l'entropie des variables discrètes

Soit A un ensemble fini non vide. On dit que la loi de X est à support A si X est à valeurs dans A et si, pour tout $x \in A$, $\mathbb{P}(X = x) > 0$.

6. Soit X une variable aléatoire de loi à support $\{0, 1, \dots, n\}$, où $n \in \mathbb{N}$. On appelle entropie de X le réel

$$H(X) = - \sum_{k=0}^n \mathbb{P}(X = k) \log_2(\mathbb{P}(X = k)).$$

6. a) On définit $g : \{0, 1, \dots, n\} \rightarrow \mathbb{R}$ par $g(k) = \log_2(\mathbb{P}(X = k))$. Montrer que $H(X) = -E(g(X))$.
6. b) Montrer que $H(X) \geq 0$.
6. c) Soit $p \in]0, 1[$. On suppose que X suit la loi de Bernoulli $\mathcal{B}(p)$.
6. c) i) Calculer $H(X)$ en fonction de p . On note ψ la fonction qui à p associe $H(X)$.
6. c) ii) Montrer que ψ est concave sur $]0, 1[$.
6. c) iii) Déterminer la valeur de p où ψ est maximale.

6. d) On suppose que X a pour support $\{0, 1, 2, 3\}$ et que $\mathbb{P}(X = 0) = \frac{1}{2}$,
 $\mathbb{P}(X = 1) = \frac{1}{4}$, $\mathbb{P}(X = 2) = \mathbb{P}(X = 3) = \frac{1}{8}$. Calculer $H(X)$.
7. On souhaite écrire une fonction Python calculant l'entropie d'une variable aléatoire X dont la loi a pour support $A = \{0, 1, \dots, n\}$, où $n \in \mathbb{N}$. La liste ou le tableau P vérifie $P[k] = \mathbb{P}(X = k)$ pour $0 \leq k \leq n$. Compléter la fonction ci-dessous, d'argument P , qui renvoie l'entropie de X .

Python

```
import numpy as np
```

```
def entropie(P):
```

```
    ...
```

La commande `len(P)` donne le nombre d'éléments de P ; `np.log(x)` donne le logarithme népérien d'un réel strictement positif x .

On souhaite maintenant démontrer quelques inégalités concernant l'entropie.

8. On commence par une inégalité générale, appelée inégalité de Jensen. Dans cette question, une fonction φ est dite convexe sur $\mathbb{R}_+$ lorsque, pour $u, v \geq 0$ et $t \in [0, 1]$,
 $\varphi(tu + (1 - t)v) \leq t\varphi(u) + (1 - t)\varphi(v)$.
8. a) Soit $N \geq 2$. Soit X une variable aléatoire de loi à support $\{x_1, \dots, x_N\}$, où les x_i sont des éléments distincts de $\mathbb{R}_+$. On pose $\mathbb{P}(X = x_i) = p_i$. Montrer que, pour tout $i \in \{1, \dots, N\}$, $0 < p_i < 1$.

On désire démontrer par récurrence la propriété $\mathcal{P}(N)$: pour toute fonction φ convexe sur $\mathbb{R}_+$, si X est une variable aléatoire de loi à support $A \subset \mathbb{R}_+$ avec $\text{Card}(A) = N$, alors $E(\varphi(X)) \geq \varphi(E(X))$.

8. b) Montrer que $\mathcal{P}(2)$ est vraie.

8. c) Soit $N \geq 3$. On suppose que $\mathcal{P}(N-1)$ est vérifiée. Soit X de loi à support

$A = \{x_1, \dots, x_N\}$, où les x_i sont des éléments distincts de $\mathbb{R}_+$. On pose

$$\mathbb{P}(X = x_i) = p_i \text{ et, pour } 1 \leq i \leq N-1, \quad p'_i = \frac{p_i}{1 - p_N}.$$

8. c) i) Montrer que

$$\sum_{i=1}^{N-1} p'_i = 1$$

et $0 < p'_i < 1$ pour $1 \leq i \leq N-1$.

8. c) ii) Soit Y de loi à support $\{x_1, \dots, x_{N-1}\}$ telle que $\mathbb{P}(Y = x_i) = p'_i$. Montrer que

$$\sum_{i=1}^{N-1} p'_i \varphi(x_i) \geq \varphi \left(\sum_{i=1}^{N-1} p'_i x_i \right).$$

8. c) iii) Montrer que $E(\varphi(X)) \geq \varphi(E(X))$.

8. d) Montrer que, si φ est concave sur $\mathbb{R}_+$, alors $E(\varphi(X)) \leq \varphi(E(X))$.

9. Soit X de loi à support $\{0, 1, \dots, n\}$. Pour $0 \leq k \leq n$, on pose $p_k = \mathbb{P}(X = k)$.

9. a) Montrer que

$$\sum_{k=0}^n p_k \log_2 \left(\frac{1}{(n+1)p_k} \right) \leq \log_2 \left(\sum_{k=0}^n \frac{p_k}{(n+1)p_k} \right) = 0.$$

9. b) Montrer que

$$\sum_{k=0}^n p_k \log_2((n+1)p_k) = \log_2(n+1) - H(X).$$

9. c) Montrer que $H(X) \leq \log_2(n+1)$.

9. d) On suppose que X suit la loi uniforme sur $\{0, 1, \dots, n\}$. Calculer $H(X)$.

10. Soient X et Y deux variables aléatoires de même loi à support $\{0, 1, \dots, n\}$. On suppose en outre X et Y indépendantes.

10. a) Montrer que

$$\mathbb{P}(X = Y) = \sum_{k=0}^n (\mathbb{P}(X = k))^2.$$

10. b) Pour $k \in \{0, 1, \dots, n\}$, on pose $v(k) = \mathbb{P}(X = k)$. Montrer que

$$2^{E(\log_2(v(X)))} \leq E(2^{\log_2(v(X))}) = E(v(X)).$$

10. c) En déduire que $2^{-H(X)} \leq \mathbb{P}(X = Y)$.

10. d) Donner un exemple de loi où l'inégalité précédente est une égalité.

Troisième partie : Entropie jointe et information mutuelle de deux variables discrètes

Soient X et Y de lois à support $\{0, 1, \dots, n\}$. On appelle entropie jointe de X et Y le réel

$$H(X, Y) = - \sum_{k=0}^n \sum_{j=0}^n \mathbb{P}([X = k] \cap [Y = j]) \log_2 (\mathbb{P}([X = k] \cap [Y = j])),$$

avec la convention $0 \log_2 0 = 0$. Cette convention s'applique aussi aux termes d'entropie conditionnelle. Dans une somme définissant une information mutuelle, un terme de probabilité jointe nulle est également égal à 0.

11. a) On définit $g : \{0, 1, \dots, n\}^2 \rightarrow \mathbb{R}$ en posant $g(k, j) = \log_2 (\mathbb{P}([X = k] \cap [Y = j]))$ lorsque cette probabilité est strictement positive et $g(k, j) = 0$ sinon. Montrer que $H(X, Y) = -E(g(X, Y))$.

11. b) Montrer que $H(X, Y) = H(Y, X)$.

11. c) Pour $0 \leq k \leq n$, on pose

$$H(Y | X = k) = - \sum_{j=0}^n \mathbb{P}_{[X=k]}(Y = j) \log_2 (\mathbb{P}_{[X=k]}(Y = j)).$$

On appelle entropie conditionnelle de Y sachant X le réel

$$H(Y | X) = \sum_{k=0}^n \mathbb{P}(X = k) H(Y | X = k).$$

Montrer que $H(X, Y) = H(X) + H(Y | X)$.

11. d) Montrer que, pour tout couple de variables aléatoires X et Y de lois à support $\{0, 1, \dots, n\}$,

$$H(X) - H(X | Y) = H(Y) - H(Y | X).$$

12. On considère deux variables aléatoires de lois à support $\{0, 1, 2, 3\}$ dont la loi conjointe est donnée par le tableau suivant. La colonne k et la ligne j donnent $\mathbb{P}([X = k] \cap [Y = j])$.

$j \backslash k$	0	1	2	3
0	1/8	1/16	1/32	1/32
1	1/16	1/8	1/32	1/32
2	1/16	1/16	1/16	1/16
3	1/4	0	0	0

12. a) Déterminer la loi de X et montrer que $H(X) = 7/4$.
12. b) Déterminer la loi de Y et calculer $H(Y)$.
12. c) Montrer que $H(X | Y) = 11/8$.
12. d) Que vaut $H(Y | X)$?
12. e) Calculer $H(X, Y)$.
13. Soient X et Y de lois à support $\{0, 1, \dots, n\}$. On appelle information mutuelle de X et Y le réel

$$I(X, Y) = \sum_{k=0}^n \sum_{j=0}^n \mathbb{P}([X = k] \cap [Y = j]) \log_2 \left(\frac{\mathbb{P}([X = k] \cap [Y = j])}{\mathbb{P}(X = k)\mathbb{P}(Y = j)} \right).$$

13. a) Montrer que $I(X, Y) = I(Y, X)$.
13. b) Montrer que $I(X, Y) = H(X) - H(X | Y)$.
13. c) Montrer que $I(X, X) = H(X)$.
13. d) Que vaut $I(X, Y)$ si X et Y sont indépendantes ?
14. Soient X et Y de lois à support $\{0, 1, \dots, n\}$. On fixe $k \in \{0, 1, \dots, n\}$ et, pour $0 \leq j \leq n$, on pose

$$p_j = \frac{\mathbb{P}([X = k] \cap [Y = j])}{\mathbb{P}(X = k)}.$$

On note $J_k = \{j \in \{0, \dots, n\} : p_j > 0\}$. Pour $j \in J_k$, on pose

$$x_j = \frac{\mathbb{P}(X = k)\mathbb{P}(Y = j)}{\mathbb{P}([X = k] \cap [Y = j])}.$$

14. a) Montrer que

$$\sum_{j=0}^n p_j = 1.$$

14. b) Soit J une variable aléatoire à valeurs dans J_k , de loi $\mathbb{P}(J = j) = p_j$ pour $j \in J_k$. On pose $Z_k = x_J$. Si plusieurs x_j sont égaux, leurs probabilités s'additionnent pour donner la loi de Z_k . Montrer que $E(\log_2 Z_k) \leq 0$. On pourra montrer d'abord que $E(Z_k) \leq 1$.

14. c) En déduire que $I(X, Y) \geq 0$.