

HEC Paris – ESSEC

Mathématiques I

Mathématiques approfondies

Sujet adapté – 2012

Les parties du sujet contenant des éléments n'étant plus au programme depuis 2022 ont été adaptées pour convenir aux nouveaux programmes de la filière Mathématiques approfondies en vigueur depuis 2023.

Table des matières

1	Quelques inégalités de concavité	2
	<i>Thèmes abordés : fonctions d'une variable ; intégrales généralisées.</i>	2
2	Entropie dans le cas continu	3
	<i>Thèmes abordés : variables à densité ; intégrales généralisées.</i>	3
3	Entropie dans le cas discret	3
	<i>Thèmes abordés : probabilités discrètes ; fonctions de plusieurs variables ; suites ; Python.</i>	3
4	Entropie et taux de rendement asymptotique	5
	<i>Thèmes abordés : convergences de variables ; probabilités discrètes.</i>	5

Le problème a pour objet la mise en évidence de quelques propriétés de l'entropie de variables aléatoires discrètes ou à densité. La dernière partie utilise, dans un exemple, certaines des propriétés établies dans le problème.

On suppose que toutes les variables aléatoires introduites dans le problème sont définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. La notation $\exp$ désigne la fonction exponentielle.

1 Quelques inégalités de concavité

Thèmes abordés : fonctions d'une variable ; intégrales généralisées.

1. Soit h la fonction de $]0, 1[$ dans $\mathbb{R}$ définie par

$$h(x) = -x \ln x - (1-x) \ln(1-x).$$

- Montrer que la fonction h est positive et concave sur $]0, 1[$.
 - Montrer que h est prolongeable en une fonction continue sur le segment $[0, 1]$. Ce prolongement est-il de classe $\mathcal{C}^1$ sur $[0, 1]$?
 - Tracer la courbe représentative de h dans le plan rapporté à un repère orthonormé.
2. Justifier, pour tout réel $u > 0$, l'inégalité $\ln u \leq u - 1$. Pour quelles valeurs de u a-t-on $\ln u = u - 1$?
3. Soit d la fonction de $]0, 1[^2$ dans $\mathbb{R}$ définie par

$$d(x, y) = x \ln \left(\frac{y}{x} \right) + (1-x) \ln \left(\frac{1-y}{1-x} \right).$$

Montrer que $d(x, y) \leq 0$ et préciser les couples (x, y) de $]0, 1[^2$ pour lesquels $d(x, y) = 0$.

4. On considère trois fonctions ℓ , r et f vérifiant les hypothèses suivantes :
- ℓ est définie et de classe $\mathcal{C}^1$ sur $\mathbb{R}$, à valeurs réelles et concave sur $\mathbb{R}$ (on note ℓ' sa fonction dérivée) ;
 - r est définie et continue sur $\mathbb{R}$, à valeurs réelles ;
 - f est définie et continue sur $\mathbb{R}$, à valeurs positives ou nulles, et $\int_{-\infty}^{+\infty} f(x) dx = 1$;
 - les intégrales $\int_{-\infty}^{+\infty} r(x)f(x) dx$ et $\int_{-\infty}^{+\infty} \ell(r(x))f(x) dx$ sont convergentes.
- Établir, pour tout couple (x, y) de $\mathbb{R}^2$, l'inégalité $\ell(x) - \ell(y) \leq \ell'(y)(x - y)$.
 - Montrer, pour tout réel y , l'inégalité

$$\int_{-\infty}^{+\infty} \ell(r(x))f(x) dx \leq \ell(y) + \ell'(y) \left(\int_{-\infty}^{+\infty} r(x)f(x) dx - y \right).$$

- En déduire l'inégalité

$$\int_{-\infty}^{+\infty} \ell(r(x))f(x) dx \leq \ell \left(\int_{-\infty}^{+\infty} r(x)f(x) dx \right).$$

5. Soit $(x_n)_{n \in \mathbb{N}^*}$ une suite réelle et $(p_n)_{n \in \mathbb{N}^*}$ une suite de réels positifs ou nuls vérifiant $\sum_{n=1}^{+\infty} p_n = 1$, telles que les séries $\sum_{n \geq 1} r(x_n)p_n$ et $\sum_{n \geq 1} \ell(r(x_n))p_n$ soient convergentes. Établir l'inégalité

$$\sum_{n=1}^{+\infty} \ell(r(x_n))p_n \leq \ell \left(\sum_{n=1}^{+\infty} r(x_n)p_n \right).$$

2 Entropie dans le cas continu

Thèmes abordés : variables à densité ; intégrales généralisées.

On note $\mathcal{F}$ l'ensemble des fonctions f définies et continues sur $\mathbb{R}$, à valeurs strictement positives, vérifiant $\int_{-\infty}^{+\infty} f(x) dx = 1$ et telles que l'intégrale $\int_{-\infty}^{+\infty} f(x) \ln(f(x)) dx$ soit convergente.

Pour toute variable aléatoire X ayant pour densité un élément f de $\mathcal{F}$, on définit l'entropie $H(X)$ de X par

$$H(X) = - \int_{-\infty}^{+\infty} f(x) \ln(f(x)) dx.$$

6. On note Z une variable aléatoire qui suit la loi normale centrée réduite et φ sa densité continue.
 - a) Justifier l'existence de l'entropie $H(Z)$ de Z et la calculer.
 - b) Soit $(a, b) \in \mathbb{R}_+^* \times \mathbb{R}$ et X une variable aléatoire qui admet pour densité un élément f de $\mathcal{F}$. Montrer que la variable aléatoire $Y = aX + b$ admet une densité appartenant à $\mathcal{F}$ et que $H(Y) = H(X) + \ln a$.
 - c) En déduire l'entropie d'une variable aléatoire suivant la loi normale d'espérance μ et d'écart-type $\sigma > 0$.
7. Dans cette question, on considère les couples (f, g) de $\mathcal{F}^2$ pour lesquels l'intégrale $\int_{-\infty}^{+\infty} f(x) \ln\left(\frac{g(x)}{f(x)}\right) dx$ est convergente. On pose alors

$$D(f, g) = - \int_{-\infty}^{+\infty} f(x) \ln\left(\frac{g(x)}{f(x)}\right) dx.$$

- a) Montrer que $D(f, g) \geq 0$.
- b) On suppose que $D(f, g) = 0$. Établir l'égalité

$$\int_{-\infty}^{+\infty} \left(\ln\left(\frac{g(x)}{f(x)}\right) + 1 - \frac{g(x)}{f(x)} \right) f(x) dx = 0.$$

En déduire que $f = g$.

3 Entropie dans le cas discret

Thèmes abordés : probabilités discrètes ; fonctions de plusieurs variables ; suites ; Python.

8. Dans cette question, N désigne un entier supérieur ou égal à 2. Soit X une variable aléatoire à valeurs dans $\{1, \dots, N\}$. On pose, pour tout $k \in \{1, \dots, N\}$, $p_k = \mathbb{P}(X = k)$. L'entropie $H(X)$ de X est définie par $H(X) = - \sum_{k=1}^N p_k \ln(p_k)$. S'il existe un entier $k \in \{1, \dots, N\}$ tel que $p_k = 0$, on pose par convention $p_k \ln(p_k) = 0$. On note h_N la fonction de $]0, 1[^N$ dans $\mathbb{R}$ définie par

$$h_N(x) = h_N(x_1, \dots, x_N) = - \sum_{k=1}^N x_k \ln(x_k).$$

- a) Calculer, en tout point x de $]0, 1[^N$, le gradient $\nabla h_N(x)$ et la matrice hessienne $\nabla^2 h_N(x)$ de h_N .

- b) Montrer que, pour l'optimisation de h_N sous la contrainte $\sum_{k=1}^N x_k = 1$, il existe un unique point critique x^* que l'on précisera.
- c) En utilisant la question 5 ou l'égalité de Taylor-Lagrange à l'ordre 1, montrer que h_N admet en x^* un maximum global sous la contrainte $\sum_{k=1}^N x_k = 1$.
- d) Parmi les variables aléatoires à valeurs dans $\{1, \dots, N\}$, quelle est la loi de celles qui ont la plus grande entropie ?

On note $\mathcal{S}$ l'ensemble des suites réelles strictement positives $(p_n)_{n \in \mathbb{N}^*}$ telles que $\sum_{n=1}^{+\infty} p_n = 1$. Soit X une variable aléatoire à valeurs dans $\mathbb{N}^*$ vérifiant, pour tout $n \in \mathbb{N}^*$, $\mathbb{P}(X = n) = p_n$, avec $(p_n)_{n \in \mathbb{N}^*} \in \mathcal{S}$. On appelle entropie de X , sous réserve de convergence de la série $\sum_{n \geq 1} p_n |\ln(p_n)|$, le réel

$$H(X) = \sum_{n=1}^{+\infty} p_n |\ln(p_n)|.$$

9. Soit $(p_n)_{n \in \mathbb{N}^*}$ une suite de $\mathcal{S}$ telle que la série $\sum_{n \geq 1} n p_n$ soit convergente.
- a) Justifier l'existence d'un entier n_0 tel que, pour tout entier $n \geq n_0$, on ait $\sqrt{p_n} |\ln(p_n)| \leq 1$.
- b) Établir, pour tout $n \geq n_0$ tel que $p_n \leq \frac{1}{n^3}$, l'inégalité $p_n |\ln(p_n)| \leq \frac{1}{n^{3/2}}$.
- c) En déduire que, pour tout $n \geq n_0$, on a $p_n |\ln(p_n)| \leq \max \left\{ \frac{1}{n^{3/2}}, 3p_n \ln n \right\}$.
- d) Montrer que la série $\sum_{n \geq 1} p_n |\ln(p_n)|$ est convergente. Que peut-on en conclure sur l'entropie d'une variable aléatoire à valeurs dans $\mathbb{N}^*$ possédant une espérance ?
10. Soit $\theta \in]0, 1[$ et $(p_n)_{n \in \mathbb{N}^*}$ la suite de $\mathcal{S}$ définie par : pour tout $n \in \mathbb{N}^*$, $p_n = \theta(1 - \theta)^{n-1}$. Soit X une variable aléatoire à valeurs dans $\mathbb{N}^*$ qui vérifie, pour tout $n \in \mathbb{N}^*$, $\mathbb{P}(X = n) = p_n$.
- a) Reconnaître la loi de X ; préciser son espérance, puis calculer son entropie.
- b) Écrire une fonction Python `X(theta)` permettant de simuler X .

```
import random

def X(theta):
    n = 1
    while random.random() > theta:
        n += 1
    return n
```

- c) Soit Y une variable aléatoire à valeurs dans $\mathbb{N}^*$ ayant une espérance égale à celle de X . Pour tout $n \in \mathbb{N}^*$, on pose $q_n = \mathbb{P}(Y = n)$. On suppose que $(q_n)_{n \in \mathbb{N}^*} \in \mathcal{S}$ et que la série $\sum_{n \geq 1} q_n \ln \left(\frac{p_n}{q_n} \right)$ est convergente. Établir l'égalité

$$H(Y) - H(X) = \sum_{n=1}^{+\infty} q_n \ln \left(\frac{p_n}{q_n} \right).$$

- d) Déterminer le signe de $H(Y) - H(X)$. Conclusion.

4 Entropie et taux de rendement asymptotique

Thèmes abordés : convergences de variables ; probabilités discrètes.

11. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires à valeurs réelles, définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, qui converge en probabilité vers une variable aléatoire X .
- Montrer que, pour tout $n \in \mathbb{N}^*$, l'application Z_n définie sur Ω par $Z_n : \omega \mapsto \exp(X_n(\omega))$ est une variable aléatoire. De même, on note Z la variable aléatoire $Z : \omega \mapsto \exp(X(\omega))$.
 - Soit ε et α deux réels strictement positifs. Justifier l'existence d'un réel s tel que $\mathbb{P}(|X| \geq s) < \alpha$.
 - Soit K_1, K_2 et K_3 trois éléments de $\mathcal{A}$. Montrer que $\mathbb{P}(K_1 \cup K_2 \cup K_3) \leq \mathbb{P}(K_1) + \mathbb{P}(K_2) + \mathbb{P}(K_3)$; en déduire l'inégalité

$$\mathbb{P}(|Z_n - Z| \geq \varepsilon) \leq \mathbb{P}(|X| \geq s) + \mathbb{P}(|X_n - X| \geq 1) + \mathbb{P}(|X_n - X| \geq \varepsilon \exp(-1 - s)).$$

- Conclure.

On considère une succession de courses hippiques entre N chevaux participants ($N \geq 2$) numérotés $1, 2, \dots, N$. Pour tout $n \in \mathbb{N}^*$, on note G_n la variable aléatoire égale au numéro du cheval gagnant de la n -ième course.

On suppose que les variables aléatoires $G_1, G_2, \dots, G_n, \dots$, à valeurs dans $\{1, \dots, N\}$, sont définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, mutuellement indépendantes et de même loi. On suppose qu'il n'y a qu'un seul gagnant par course. On pose, pour tout $k \in \{1, \dots, N\}$ et tout $n \in \mathbb{N}^*$, $p_k = \mathbb{P}(G_n = k)$, avec $0 < p_k < 1$.

Pour tout $k \in \{1, \dots, N\}$, on note c_k ($c_k > 1$) la cote du cheval k ; ainsi, un parieur qui a misé un montant m_k sur le cheval k perdra sa mise quelle que soit l'issue de la course, mais recevra la somme $m_k c_k$ si le cheval k est gagnant. On suppose que les cotes $c_1, c_2, \dots, c_N$ sont fixes au cours du temps.

À l'occasion de la première course, un parieur dispose d'une somme monétaire $r_0 > 0$ qu'il souhaite répartir en totalité entre les N chevaux dans les proportions respectives $f_1, f_2, \dots, f_N$, où, pour tout $k \in \{1, \dots, N\}$, $0 < f_k < 1$. À l'issue de cette première course, le parieur dispose d'une somme monétaire $R_1 = r_0 M_1$, avec $M_1 > 0$.

À l'occasion de la deuxième course, ce parieur réinvestit en totalité la somme R_1 entre les N chevaux dans les mêmes proportions $f_1, f_2, \dots, f_N$. À l'issue de cette deuxième course, le parieur dispose d'une somme monétaire $R_2 = R_1 M_2$, avec $M_2 > 0$, et ainsi de suite. La richesse monétaire R_n acquise au terme de n courses est donc

$$R_n = r_0 \prod_{i=1}^n M_i.$$

On définit, pour tout $n \in \mathbb{N}^*$, le taux de rendement moyen des paris par

$$T_n = \left(\frac{R_n}{r_0} \right)^{1/n} - 1.$$

-
12. a) Justifier que $(M_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires indépendantes, à valeurs strictement positives et de même loi.
- b) On suppose que la variable aléatoire $\ln(M_1)$ admet une espérance $\mathbb{E}(\ln(M_1))$ et une variance $\mathbb{V}(\ln(M_1))$. Montrer que la suite de variables aléatoires $(T_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers une variable certaine τ que l'on exprimera en fonction de $\mathbb{E}(\ln(M_1))$. Le réel τ est le taux de rendement asymptotique des paris.
13. La stratégie du parieur consiste à choisir les proportions $f_1, f_2, \dots, f_N$ qui maximiseraient τ . On rappelle que les proportions $f_1, f_2, \dots, f_N$ sont constantes au cours du temps.
- a) Montrer que $\tau = \exp \left(\sum_{k=1}^N p_k \ln(f_k c_k) \right) - 1$.
- b) En déduire la stratégie optimale du parieur et la valeur optimale de τ associée à ses paris.
- c) On suppose dans cette question que $\sum_{k=1}^N \frac{1}{c_k} = 1$. Montrer que $\sum_{k=1}^N p_k \ln(p_k c_k) \geq 0$. Dans quel cas le parieur ne dispose-t-il d'aucune stratégie lui permettant de s'assurer un taux de rendement asymptotique optimal strictement positif?