Mathématiques 1

TSI **2017**

CONCOURS CENTRALE•SUPÉLEC

4 heures Calculatrices autorisées

Objectif du problème

Ce problème aborde la question de la mesure quantitative de l'information.

La première partie étudie quelques propriétés de la fonction logarithme népérien utiles pour les autres parties du problème.

La deuxième partie vise à construire les fonctions permettant de modéliser l'information contenue dans les événements de probabilité non nulle d'un espace probabilisé. **L'information contenue dans un tel événement correspond intuitivement à l'effet de surprise provoqué par la réalisation de cet événement.**

La troisième partie aborde la notion d'entropie d'une variable aléatoire discrète à valeurs entières.

La quatrième partie aborde l'entropie d'un couple de variables aléatoires et la notion d'entropie conditionnelle.

L'exemple traité dans la cinquième partie établit le lien entre la quantité d'information contenue dans un message aléatoire et le nombre minimal de questions que le récepteur du message doit poser à son émetteur pour pouvoir identifier sans ambiguïté l'une des réalisations de ce message.

Notations

- Pour $x \in \mathbb{R}^{+*}$, $\ln x$ désigne le logarithme népérien de x.
- Pour $n \in \mathbb{N}$, $[\![0, n]\!]$ désigne les entiers naturels compris entre 0 et n.

I Questions préliminaires

I.A – Représenter graphiquement la fonction logarithme népérien.

I.B – Démontrer que, pour tout $x \in \mathbb{R}^{+*}$, $\ln(x) \leqslant x - 1$ et que $\ln(x) = x - 1$ si et seulement si $x = 1$.

I.C – Donner une interprétation graphique de ces deux résultats.

I.D – Montrer que la fonction g définie sur $[0, 1]$ par $g(0) = 0$ et $\forall x \in]0, 1]$, $g(x) = x\ln(x)$ est continue sur $[0, 1]$ et dérivable sur $]0, 1]$. Représenter graphiquement la fonction g.

II Mathématisation de l'effet de surprise

Soit (Ω, P) un espace probabilisé fini. On convient de modéliser la quantité d'information contenue dans les événements de probabilité non nulle par une fonction S définie par

$$\forall A \in \mathcal{P}(\Omega) \text{ tel que } P(A) \neq 0 \qquad S(A) = f(P(A))$$

où $f\colon]0, 1] \to \mathbb{R}$ vérifie les contraintes suivantes :

i. $f(1) = 0$
ii. f est décroissante sur $]0, 1]$
iii. $\forall (p, q) \in]0, 1]^2 \quad f(pq) = f(p) + f(q)$
iv. f est continue sur $]0, 1]$

II.A – Quelle est la quantité d'information de l'événement certain ? Interpréter en terme d'effet de surprise.

II.B – Que peut-on dire de la quantité d'information contenue dans l'événement $A \cap B$ lorsque A et B sont indépendants ? Interpréter en terme d'effet de surprise.

II.C – Donner un exemple de fonction f vérifiant les quatre contraintes i, ii, iii et iv.

II.D – On se propose de déterminer l'ensemble des fonctions vérifiant ces quatre contraintes.

Soit f une telle fonction.

II.D.1) Soit $p \in]0, 1]$. Établir, à l'aide d'un changement de variable, l'égalité

$$\frac{1}{p}\int_{p/2}^{p} f(t)\,\mathrm{d}t = \frac{1}{2}f(p) + \int_{1/2}^{1} f(u)\,\mathrm{d}u$$

II.D.2) En déduire que f est dérivable sur $]0, 1]$.

II.D.3) Dans cette question, on fixe $p \in]0, 1]$. En dérivant par rapport à q l'égalité iii, démontrer l'existence d'un réel a indépendant de p tel que $f'(p) = a/p$. Préciser la valeur de a.

II.D.4) L'égalité $f'(p) = a/p$ étant vraie quel que soit p dans $]0,1]$, déterminer l'ensemble des fonctions f vérifiant les quatre contraintes i, ii, iii et iv.

II.D.5) Montrer que parmi ces fonctions, il en existe une et une seule vérifiant en plus l'égalité $f(1/\mathrm{e}) = 1$.

Cette fonction, notée h dans la suite du problème, correspond au choix d'une unité particulière (le logon) pour mesurer la quantité d'information.

Que vaut $\lim\limits_{p\to 0^+} h(p)$? Interpréter ce résultat.

II.D.6) On réalise l'expérience aléatoire consistant à effectuer deux lancers successifs d'un dé équilibré à six faces. On considère les évènements suivants

- E : « le numéro sorti lors du premier lancer est pair » ;
- M : « le maximum des deux numéros sortis est égal à 4 » ;
- N : « la somme des deux numéros sortis est égale à 7 ».

Ordonner les quantités d'information contenues dans chacun de ces trois événements. Interpréter en terme d'effet de surprise.

III Entropie d'une variable aléatoire

III.A – Dans cette sous-partie, toutes les variables aléatoires considérées sont définies sur un même univers fini Ω et prennent leurs valeurs dans $[\![0,n]\!]$.

Si X est une telle variable, on note $p_k = P(X = k)$. On définit l'entropie de X par

$$H(X) = -\sum_{k=0}^{n} p_k \ln(p_k)$$

en convenant que $p_k \ln(p_k)$ vaut 0 lorsque $p_k = 0$.

III.A.1) Interpréter $H(X)$ comme une espérance, puis en terme de quantité d'information.

III.A.2) Montrer que $H(X) \geqslant 0$ et que $H(X) = 0$ si et seulement si X est une variable aléatoire certaine, c'est-à-dire

$$\exists i \in [\![0,n]\!] \qquad \text{tel que} \qquad p_i = 1 \quad \text{et} \quad \forall j \neq i, \quad p_j = 0$$

III.A.3)

a) X_0 est une variable aléatoire suivant la loi uniforme sur $[\![0,n]\!]$. Calculer $H(X_0)$.

b) En appliquant l'inégalité de la question I.B à un nombre réel x bien choisi, démontrer que

$$\forall k \in [\![0,n]\!] \qquad -p_k \ln(p_k) + p_k \ln\left(\frac{1}{n+1}\right) \leqslant \frac{1}{n+1} - p_k$$

c) En déduire que $H(X) \leqslant H(X_0)$ avec égalité si et seulement si X suit la même loi que X_0 (pour le cas d'égalité on pourra utiliser le cas d'égalité de la question I.B). Interpréter ce résultat en terme de quantité moyenne d'information.

III.B – Dans cette sous-partie, on s'intéresse à des variables aléatoires discrètes définies sur un espace probabilisé (Ω, P) et prenant leurs valeurs dans $\mathbb{N}^*$. Si X est une telle variable pour laquelle $P(X = k)$ est noté p_k, on dit qu'elle est d'entropie finie si la série $\sum p_k \ln(p_k)$ est absolument convergente et on définit alors son entropie par

$$H(X) = -\sum_{k=1}^{+\infty} p_k \ln(p_k)$$

en convenant à nouveau que $p_k \ln(p_k)$ vaut 0 lorsque $p_k = 0$.

III.B.1) Pour $p \in]0,1[$, X_1 est une variable aléatoire suivant la loi géométrique de paramètre p.

Rappeler les valeurs de $P(X_1 = k)$ et de l'espérance de X_1 (aucune démonstration n'est demandée).

Démontrer que X_1 est d'entropie finie et que $H(X_1) = -\dfrac{1-p}{p}\ln(1-p) - \ln(p)$.

III.B.2) Dans cette question et la suivante, X est une variable aléatoire à valeurs dans $\mathbb{N}^*$ d'espérance finie.

On note $E(X) = \sum\limits_{k=1}^{+\infty} k p_k$. On se propose de démontrer que X est d'entropie finie.

a) Quelle est la limite de p_k lorsque k tend vers $+\infty$?

b) En déduire que $\lim\limits_{k\to+\infty} \sqrt{p_k}\ln(p_k) = 0$, puis qu'il existe un entier k_0 tel que $\forall k \geqslant k_0 \quad 0 \leqslant -\sqrt{p_k}\ln(p_k) \leqslant 1$.

c) Soit $k \geqslant k_0$. Montrer que

— si $p_k \leqslant \dfrac{1}{k^3}$, alors $0 \leqslant -p_k \ln(p_k) \leqslant \dfrac{1}{k^{3/2}}$;

— si $p_k \geqslant \dfrac{1}{k^3}$, alors $0 \leqslant -p_k \ln(p_k) \leqslant 3p_k \ln(k)$.

d) Soit $k \geqslant 1$. Justifier que $\ln(k) \leqslant k$, puis que la série $\sum\limits_{k\geqslant 1} \left(\dfrac{1}{k^{3/2}} + 3p_k \ln(k)\right)$ converge.

e) Conclure.

III.B.3) Dans cette question, on suppose en plus que $E(X) \leqslant 1/p$, p étant un réel de l'intervalle $]0,1[$. On veut montrer que $H(X) \leqslant H(X_1)$ (entropie d'une variable aléatoire suivant la loi géométrique de paramètre p dont la valeur a été calculée à la question III.B.1).

Pour $k \in \mathbb{N}^*$, on note $p_k = P(X = k)$ et $q_k = P(X_1 = k)$.

a) Justifier que la série $\sum\limits_{k=1}^{+\infty} (k-1)p_k$ converge et exprimer sa somme en fonction de $E(X)$.

b) Justifier la convergence de la série $\sum p_k \ln(q_k)$ et démontrer que

$$\sum_{k=1}^{+\infty} p_k \ln(q_k) = \ln(p) + (E(X) - 1)\ln(1-p)$$

c) Démontrer que

$$-H(X_1) \leqslant \sum_{k=1}^{+\infty} p_k \ln(q_k)$$

d) En déduire que

$$H(X) - H(X_1) \leqslant \sum_{k=1}^{+\infty} p_k \ln\left(\frac{q_k}{p_k}\right)$$

puis que

$$H(X) \leqslant H(X_1)$$

On pourra utiliser l'inégalité démontrée dans la question I.B.

Interpréter ce résultat en terme de quantité moyenne d'information.

IV Quatrième partie

Dans cette partie, m et n sont des entiers naturels non nuls. (X, Y) et (X', Y') sont deux couples de variables aléatoires discrètes. X et X' sont à valeurs dans $[\![0, n]\!]$, Y et Y' sont à valeurs dans $[\![0, m]\!]$. Pour $i \in [\![0, n]\!]$ et $j \in [\![0, m]\!]$, on note $p_i = P(X = i)$, $q_j = P(Y = j)$, $\lambda_{ij} = P(X = i; Y = j)$ et $\lambda'_{ij} = P(X' = i; Y' = j)$.

On suppose que, pour tout $(i, j) \in [\![0, n]\!] \times [\![0, m]\!]$, $\lambda_{ij} \neq 0$ et $\lambda'_{ij} \neq 0$.

Notations

On définit l'entropie du couple (X, Y) par

$$H(X, Y) = -\sum_{i=0}^{n} \sum_{j=0}^{m} \lambda_{ij} \ln(\lambda_{ij})$$

On définit l'information entre les couples (X, Y) et (X', Y') par

$$K(X, Y, X', Y') = -\sum_{i=0}^{n} \sum_{j=0}^{m} \lambda_{ij} \ln\left(\frac{\lambda'_{ij}}{\lambda_{ij}}\right)$$

IV.A – Propriétés de l'information entre deux couples

IV.A.1) Rappeler les valeurs de $\sum\limits_{i=0}^{n} \sum\limits_{j=0}^{m} \lambda_{ij}$ et de $\sum\limits_{i=0}^{n} \sum\limits_{j=0}^{m} \lambda'_{ij}$ et en déduire que

$$K(X, Y, X', Y') = -\sum_{i=0}^{n} \sum_{j=0}^{m} \lambda_{ij} \left(\ln\left(\frac{\lambda'_{ij}}{\lambda_{ij}}\right) - \frac{\lambda'_{ij}}{\lambda_{ij}} + 1\right)$$

IV.A.2) À l'aide de l'inégalité de la question I.B, établir que $K(X, Y, X', Y') \geqslant 0$, et que l'égalité a lieu si et seulement si les deux couples (X, Y) et (X', Y') ont la même loi conjointe.

IV.A.3) On suppose que les deux variables X' et Y' sont indépendantes, que X' suit la même loi que X et que Y' suit la même loi que Y.

Démontrer que $K(X, Y, X', Y') = H(X) + H(Y) - H(X, Y)$. Déduire de ce qui précède que

$$H(X, Y) \leqslant H(X) + H(Y) \tag{IV.1}$$

Donner une condition nécessaire et suffisante pour que cette inégalité soit une égalité.

Remarque — L'inégalité (IV.1) a été obtenue en supposant, pour tout $(i, j) \in [\![0, n]\!] \times [\![0, m]\!]$, $\lambda_{ij} \neq 0$ et $\lambda'_{ij} \neq 0$. On admet qu'elle reste vraie même en dehors de cette condition

IV.B – Entropie conditionnelle

On définit l'entropie conditionnelle de Y sachant X par $H_X(Y) = H(X, Y) - H(X)$.

Elle mesure l'incertitude restant sur la valeur de Y lorsque la valeur de X est connue.

IV.B.1) Montrer que $H_X(Y) \leqslant H(Y)$. Interpréter cette inégalité.

IV.B.2) On considère $m + 1$ réels $a_0, a_1, ..., a_m$ compris entre 0 et 1.

a) Dans cette question, on suppose $(a_0, a_1, ..., a_m) \in]0, 1]^{m+1}$.

Démontrer que, pour tout $j \in [\![0, m]\!]$, $\ln(a_j) \leqslant \ln(a_0 + a_1 + \cdots + a_m)$.

En déduire l'inégalité

$$\sum_{j=0}^{m} g(a_j) \leqslant g\left(\sum_{j=0}^{m} a_j\right) \tag{IV.2}$$

La fonction g a été définie dans la partie I.

b) L'inégalité (IV.2) reste-t-elle vraie si $(a_0, a_1, ..., a_m) \in [0, 1]^{m+1}$?

c) Montrer que l'inégalité (IV.2) est une égalité si et seulement s'il existe au plus un indice $j \in [\![0, m]\!]$ pour lequel $a_j \neq 0$.

IV.B.3) Montrer que, pour tout $i \in [\![0, n]\!]$, $\displaystyle\sum_{j=0}^{m} g(\lambda_{ij}) \leqslant g(p_i)$. En déduire que $H_X(Y) \geqslant 0$.

V Une application

Un jeu oppose deux joueurs A et B. Une urne contient 2016 boules indiscernables au toucher numérotées de 0 à 2015. Le joueur A tire une boule au hasard dans l'urne. On note Y la variable aléatoire égale au numéro de la boule tirée par le joueur A. Le joueur B pose alors au joueur A une série de N questions amenant la réponse « oui » ou « non » lui permettant de déterminer sans ambiguïté la valeur prise par Y. Le but de cette partie est de trouver la valeur minimale de N si B procède par dichotomies successives.

V.A – Déterminer l'entropie de Y.

V.B – La première question posée par B est « Y est-il compris entre 1008 et 2015 ? » La réponse fournie par A lui permet de positionner Y par rapport 1008. Si la réponse de A est « oui », B posera $X_1 = 1$ et cherchera ensuite à positionner Y par rapport à 1512. Si la réponse à la première question est « non », B posera $X_1 = 0$ et cherchera à positionner Y par rapport à 504. Il continue selon le même procédé.

Expliquer pourquoi le joueur B finira par trouver la valeur de Y. Au bout de combien de questions ?

V.C – On se propose de donner une interprétation de ce résultat en terme d'entropie. Les variables aléatoires $(X_1, X_2, ..., X_N)$ sont renseignées par les réponses données par A à la première, la deuxième, ..., la N-ième question.

On pose $X = \displaystyle\sum_{i=1}^{N} X_{N+1-i} 2^{i-1}$. On admet que X est une variable aléatoire.

Montrer qu'elle prend ses valeurs dans $[\![0, 2^N - 1]\!]$. On ne cherchera pas la loi de X.

V.D – Démontrer que $H(X) \leqslant N \ln(2)$.

On pourra utiliser l'inégalité vue à la question III.A.3.

V.E – Expliquer en langage courant pourquoi $H_Y(X) = 0$. En déduire que $H(X, Y) = H(Y)$.

V.F – Montrer que $H_X(Y) \geqslant \ln(2016) - N \ln(2)$.

V.G – En combien de questions le joueur B est-il certain de pouvoir trouver à coup sûr la valeur de Y ? Comparer au résultat de la question V.B.

• • • FIN • • •