



PRÉPASIMPLE

prepasimple.fr

ESSEC

Mathématiques 2

2003

Corrigé de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Corrigé

Partie I — Introduction informatique

I.1)

On décompose chaque entier en somme de puissances distinctes de 2.

$$6 = 4 + 2 = 0 \times 16 + 0 \times 8 + 1 \times 4 + 1 \times 2 + 0.$$

Ainsi $6 = \overline{110}_{\text{deux}}$ et, en conservant les cinq chiffres demandés, $\text{bin}(6) = (0, 0, 1, 1, 0)$.

De même,

$$21 = 16 + 4 + 1 = 1 \times 16 + 0 \times 8 + 1 \times 4 + 0 \times 2 + 1,$$

d'où $\text{bin}(21) = (1, 0, 1, 0, 1)$.

I.2)

Question adaptée au programme actuel. Code adapté en Python.

Méthode 1 — Divisions successives par deux

Le reste de la division par 2 donne le chiffre de poids le plus faible. On le place à droite, puis on recommence avec le quotient entier.

Python

```
def bin(n):  
    L = [0, 0, 0, 0, 0]  
    for j in range(5):  
        L[4 - j] = n % 2  
        n = n // 2  
    return L
```

Aucun import n'est nécessaire. Au début de l'itération d'indice j , la variable n contient le quotient entier de l'entier initial par 2^j . Son reste modulo 2 est donc a_j , enregistré dans $L[4 - j]$. Après cinq itérations, tous les chiffres sont placés dans l'ordre $(a_4, a_3, a_2, a_1, a_0)$. Le quotient final est nul puisque l'entier initial est inférieur à 2^5 .

Méthode 2 — Extraction des poids décroissants

On examine successivement les poids 16, 8, 4, 2 et 1, en retirant du reste le poids retenu.

Python

```
def bin(n):
    L = [0, 0, 0, 0, 0]
    poids = 16
    reste = n
    for j in range(5):
        if reste >= poids:
            L[j] = 1
            reste = reste - poids
        poids = poids // 2
    return L
```

Avant chaque test, le reste est compris entre 0 et $2 \text{ poids} - 1$. Le chiffre correspondant vaut donc 0 ou 1 selon que le reste est inférieur ou supérieur au poids. Après soustraction éventuelle, le reste est inférieur au poids, ce qui prépare l'itération suivante. Au dernier poids, égal à 1, le reste devient nul. La liste renvoyée est donc la même écriture binaire.

La première méthode extrait les chiffres de droite à gauche ; la seconde reconstruit directement la décomposition par poids décroissants. Les deux fonctions ont été exécutées et vérifiées pour les 32 entiers autorisés.

I.3)

Question adaptée au programme actuel. Code adapté en Python.

Les deux premiers chiffres codent la famille, avec les indices 0, 1, 2, 3 pour trèfle, carreau, cœur et pique. Les trois derniers codent la valeur, avec les indices 0 à 7 pour sept, huit, neuf, dix, valet, dame, roi et as.

Voici le code complet nécessaire à l'exécution.

Python

```
def bin(n):  
    L = [0, 0, 0, 0, 0]  
    for j in range(5):  
        L[4 - j] = n % 2  
        n = n // 2  
    return L  
  
def carte(n):  
    fam = ["trèfle", "carreau", "cœur", "pique"]  
    val = ["sept", "huit", "neuf", "dix", "valet", "dame", "roi", "as"]  
    L = bin(n)  
    famille = fam[2 * L[0] + L[1]]  
    valeur = val[4 * L[2] + 2 * L[3] + L[4]]  
    print(valeur, "de", famille, "est la carte numéro", n)  
  
carte(6)  
carte(1)  
carte(21)
```

Pour 6, la liste (0, 0, 1, 1, 0) donne la famille d'indice 0 et la valeur d'indice 6 : c'est le roi de trèfle.

Pour 1, la liste (0, 0, 0, 0, 1) donne la famille d'indice 0 et la valeur d'indice 1 : c'est le huit de trèfle.

Pour la dame de cœur, l'indice de famille est $2 = 2 \times 1 + 0$ et l'indice de valeur est $5 = 4 \times 1 + 2 \times 0 + 1$. La liste est (1, 0, 1, 0, 1), donc le numéro est $16 + 4 + 1 = 21$.

Le code a été exécuté : il affiche respectivement « roi de trèfle est la carte numéro 6 »,

« huit de trèfle est la carte numéro 1 » et « dame de cœur est la carte numéro 21 ». La numérotation des 32 cartes a aussi été vérifiée : chaque couple famille-valeur apparaît une seule fois.

Partie II — Position du problème et recherche des fonctions solutions

II.1)

On vérifie séparément les quatre axiomes.

(i) Puisque $P(\Omega) = 1$, on a $i_\varphi(\Omega) = \varphi(1) = 0$.

(ii) Si $P(A) = P(\bar{A})$, l'identité $P(A) + P(\bar{A}) = 1$ donne $P(A) = 1/2$. Donc $i_\varphi(A) = \varphi(1/2) = 1$.

(iii) Si A et B sont indépendants et $P(A \cap B) > 0$, alors $P(A \cap B) = P(A)P(B)$ et les deux facteurs sont strictement positifs. On peut donc appliquer (2) :

$$i_\varphi(A \cap B) = \varphi(P(A)P(B)) = \varphi(P(A)) + \varphi(P(B)) = i_\varphi(A) + i_\varphi(B).$$

(iv) Si $P(A) = P(B) > 0$, l'application de la même fonction à ces deux réels donne $i_\varphi(A) = i_\varphi(B)$.

Les quatre axiomes sont ainsi satisfaits.

II.2)

La condition en 1 impose $\beta = 0$. La condition en $1/2$ impose ensuite

$$\alpha \ln(1/2) = 1, \quad \alpha = -\frac{1}{\ln(2)}.$$

Réciproquement, pour ces valeurs,

$$\varphi_{\alpha,\beta}(pq) = -\frac{\ln(pq)}{\ln(2)} = -\frac{\ln(p)}{\ln(2)} - \frac{\ln(q)}{\ln(2)}$$

pour tous $p, q \in]0, 1]$. Les deux normalisations sont également satisfaites. Il existe donc un unique couple de la forme recherchée :

$$(\alpha, \beta) = \left(-\frac{1}{\ln(2)}, 0 \right).$$

II.3)

Dans cette question, φ est seulement supposée continue au départ. Il faut établir sa dérivabilité avant toute dérivation de cette fonction.

II.3.a)

Fixons $p \in]0, 1]$. Le changement de variable affine $t = pq$ transforme les bornes $p/2, p$ en $1/2, 1$ et donne $dt = p dq$. Les fonctions intégrées sont continues sur les segments considérés. Ainsi,

$$\frac{1}{p} \int_{p/2}^p \varphi(t) dt = \int_{1/2}^1 \varphi(pq) dq.$$

Par (2), $\varphi(pq) = \varphi(p) + \varphi(q)$. La linéarité de l'intégrale donne donc

$$\frac{1}{p} \int_{p/2}^p \varphi(t) dt = \frac{1}{2} \varphi(p) + \int_{1/2}^1 \varphi(q) dq.$$

II.3.b)

Question adaptée au programme actuel : la dérivée en 1 est précisée comme une dérivée à gauche.

Notons $C = \int_{1/2}^1 \varphi(q) dq$ et $F(u) = \int_{1/2}^u \varphi(t) dt$ pour $u \in]0, 1]$. La continuité de φ et le théorème fondamental du calcul intégral assurent que F est dérivable, avec $F' = \varphi$; au point 1, on utilise une dérivée à gauche.

L'identité précédente donne

$$\varphi(p) = \frac{2}{p}(F(p) - F(p/2)) - 2C.$$

Le membre de droite est dérivable puisque $p > 0$. Cela établit d'abord la dérivabilité de φ .

On réécrit maintenant l'identité sous la forme

$$\int_{p/2}^p \varphi(t) dt = \frac{1}{2}p\varphi(p) + pC.$$

En dérivant par le théorème fondamental et la dérivation d'un produit,

$$\varphi(p) - \frac{1}{2}\varphi(p/2) = \frac{1}{2}\varphi(p) + \frac{1}{2}p\varphi'(p) + C.$$

Or (2) et (1) donnent $\varphi(p/2) = \varphi(p) + 1$. Le membre de gauche vaut donc $\frac{1}{2}\varphi(p) - \frac{1}{2}$.
En simplifiant,

$$-\frac{1}{2} = \frac{1}{2}p\varphi'(p) + C.$$

C'est exactement la relation demandée.

II.3.c)

La relation précédente donne

$$\varphi'(p) = \frac{-1 - 2C}{p}.$$

Posons $\alpha = -1 - 2C$. La fonction $p \mapsto \varphi(p) - \alpha \ln(p)$ a une dérivée nulle sur $]0, 1[$; elle est donc constante, d'après le théorème des accroissements finis. Par continuité en 1, la même constante β convient sur tout $]0, 1]$.

Ainsi $\varphi(p) = \alpha \ln(p) + \beta$. Les normalisations de (1), comme en II.2, imposent $\beta = 0$ et $\alpha = -1/\ln(2)$.

II.4)

Il existe une unique fonction continue sur $]0, 1]$ vérifiant (1) et (2) :

$$\varphi(p) = -\frac{\ln(p)}{\ln(2)}.$$

Cette fonction définit, par II.1, une incertitude respectant les quatre axiomes :

$i(A) = -\ln(P(A))/\ln(2)$ pour $P(A) > 0$.

L'unicité établie porte sur les fonctions continues vérifiant les identités pour tous $p, q \in]0, 1]$. Les quatre axiomes sur les seuls événements d'un espace probabilisé fixé ne suffisent pas, à eux seuls, à imposer ces identités pour toutes les probabilités réelles ni à garantir la continuité.

Partie III — Incertitude des événements

III.1)

Les 32 cartes sont équiprobables et une seule est la dame de cœur. Donc

$$P(A) = \frac{1}{32} = 2^{-5}, \quad i(A) = -\frac{\ln(2^{-5})}{\ln(2)} = 5.$$

III.2)

Chaque entier est décrit par une unique liste de n chiffres binaires, en ajoutant des zéros à gauche si nécessaire. Chaque chiffre a deux possibilités, donc $|E| = 2^n$. Plus précisément, $E = \{0, \dots, 2^n - 1\}$.

Le tirage étant uniforme,

$$P(A) = 2^{-n}, \quad i(A) = n.$$

III.3)

L'inclusion donne $0 < P(A) \leq P(B) \leq 1$. Comme $\ln$ est croissante et $\ln(2) > 0$,

$$i(A) = -\frac{\ln(P(A))}{\ln(2)} \geq -\frac{\ln(P(B))}{\ln(2)} = i(B).$$

Il y a égalité si et seulement si $P(A) = P(B)$, c'est-à-dire $P(B \setminus A) = 0$. Une inclusion stricte d'événements n'impose donc pas une inégalité stricte des incertitudes.

III.4)

Comme $\ln(x) \rightarrow -\infty$ quand $x \rightarrow 0^+$,

$$\lim_{x \rightarrow 0^+} \varphi(x) = +\infty.$$

L'incertitude associée à la réalisation d'un événement augmente sans borne lorsque sa probabilité tend vers zéro. La définition initiale ne donne pas de valeur finie à un événement de probabilité nulle.

Partie IV — Incertitude d'une variable aléatoire discrète

On pose $c = \ln(2) > 0$ pour alléger les calculs du corrigé.

IV.1)

La loi est finie et chacune des n valeurs a pour probabilité $1/n$. Ainsi,

$$H(U_n) = n h(1/n) = -\frac{\ln(1/n)}{c} = \frac{\ln(n)}{c}.$$

Cette formule vaut aussi pour $n = 1$, où l'entropie est nulle.

IV.2)

On a $h(1/4) = 1/2$ et $h(1/2) = 1/2$, donc

$$H(Y) = 2h(1/4) + h(1/2) = \frac{3}{2}.$$

Par IV.1, $H(U_2) = 1$ et $H(U_3) = \ln(3)/c$. Pour comparer les deux dernières valeurs sans approximation, $9 > 8$ donne $2 \ln(3) > 3 \ln(2)$, donc $\ln(3)/c > 3/2$.

La conclusion est

$$H(U_2) < H(Y) < H(U_3).$$

IV.3)

La fonction est continue sur $]0, 1]$. La croissance comparée $x \ln(x) \rightarrow 0$ lorsque $x \rightarrow 0^+$ montre que sa limite en 0 vaut $0 = h(0)$: h est continue sur $[0, 1]$.

Pour $0 < x \leq 1$, $\ln(x) \leq 0$, donc $h(x) \geq 0$. Ses seuls zéros sont 0 et 1, et $h(x) > 0$ sur $]0, 1[$.

Le taux d'accroissement à droite en 0 vaut

$$\frac{h(x) - h(0)}{x} = -\frac{\ln(x)}{c} \rightarrow +\infty.$$

Il n'existe donc pas de dérivée finie à droite en 0; la courbe admet une demi-tangente verticale en ce point.

Pour $x \in]0, 1]$,

$$h'(x) = -\frac{\ln(x) + 1}{c}, \quad h''(x) = -\frac{1}{cx} < 0.$$

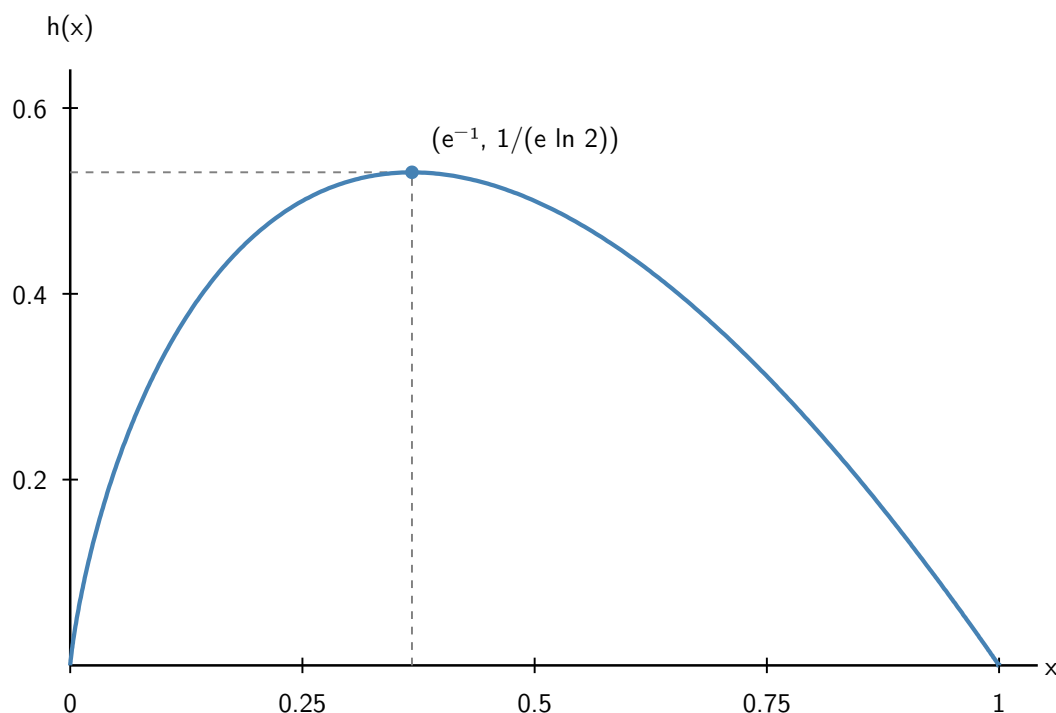
La fonction est strictement croissante sur $[0, e^{-1}]$, puis strictement décroissante sur $[e^{-1}, 1]$. Son maximum vaut

$$h(e^{-1}) = \frac{1}{ec}.$$

Elle est strictement concave sur $]0, 1[$. À l'extrémité 1, sa dérivée à gauche vaut $-1/c$ et sa tangente a pour équation $y = (1 - x)/c$. Le tableau résume les variations.

x	0	e^{-1}	1
$h(x)$	0	$1/(ec)$	0
Variation	croissante	maximum	décroissante

La courbe suivante est calculée à partir de la formule de h ; les axes ont la même échelle.



IV.4)

Notons $p_1, \dots, p_r$ les probabilités des valeurs d'un ensemble fini contenant toutes les valeurs de X . La somme

$$H(X) = \sum_{k=1}^r h(p_k)$$

est finie et tous ses termes sont positifs ou nuls. Donc $H(X) \geq 0$.

Si $H(X) = 0$, chaque terme est nul. Par IV.3, chaque p_k appartient alors à $\{0, 1\}$. Comme leur somme vaut 1, une et une seule de ces probabilités vaut 1. Ainsi X est quasi certaine, c'est-à-dire qu'il existe un réel a tel que $P(X = a) = 1$.

Réciproquement, une telle variable a une probabilité égale à 1 et les autres égales à 0; toutes les contributions à son entropie sont nulles. L'équivalence est établie.

IV.5)

La fonction h_2 est définie par $h_2(x) = h(x) + h(1 - x)$ sur $[0, 1]$.

IV.5.a)

La transformation $x \mapsto 1 - x$ est la réflexion par rapport à la droite verticale d'abscisse $1/2$. L'identité $h_2(x) = h_2(1 - x)$ signifie donc que la courbe de h_2 est symétrique par rapport à la droite $x = 1/2$.

IV.5.b)

Par continuité de h , h_2 est continue sur $[0, 1]$, avec $h_2(0) = h_2(1) = 0$. Pour $0 < x < 1$,

$$h'_2(x) = h'(x) - h'(1 - x) = \frac{\ln(1 - x) - \ln(x)}{c} = \frac{1}{c} \ln\left(\frac{1 - x}{x}\right).$$

Cette dérivée est strictement positive pour $x < 1/2$, nulle en $1/2$ et strictement négative pour $x > 1/2$. Ainsi h_2 croît strictement sur $[0, 1/2]$, puis décroît strictement sur $[1/2, 1]$.

$$h_2(1/2) = 2h(1/2) = 1.$$

Sa dérivée seconde vaut

$$h''_2(x) = -\frac{1}{c} \left(\frac{1}{1 - x} + \frac{1}{x} \right) = -\frac{1}{cx(1 - x)} < 0.$$

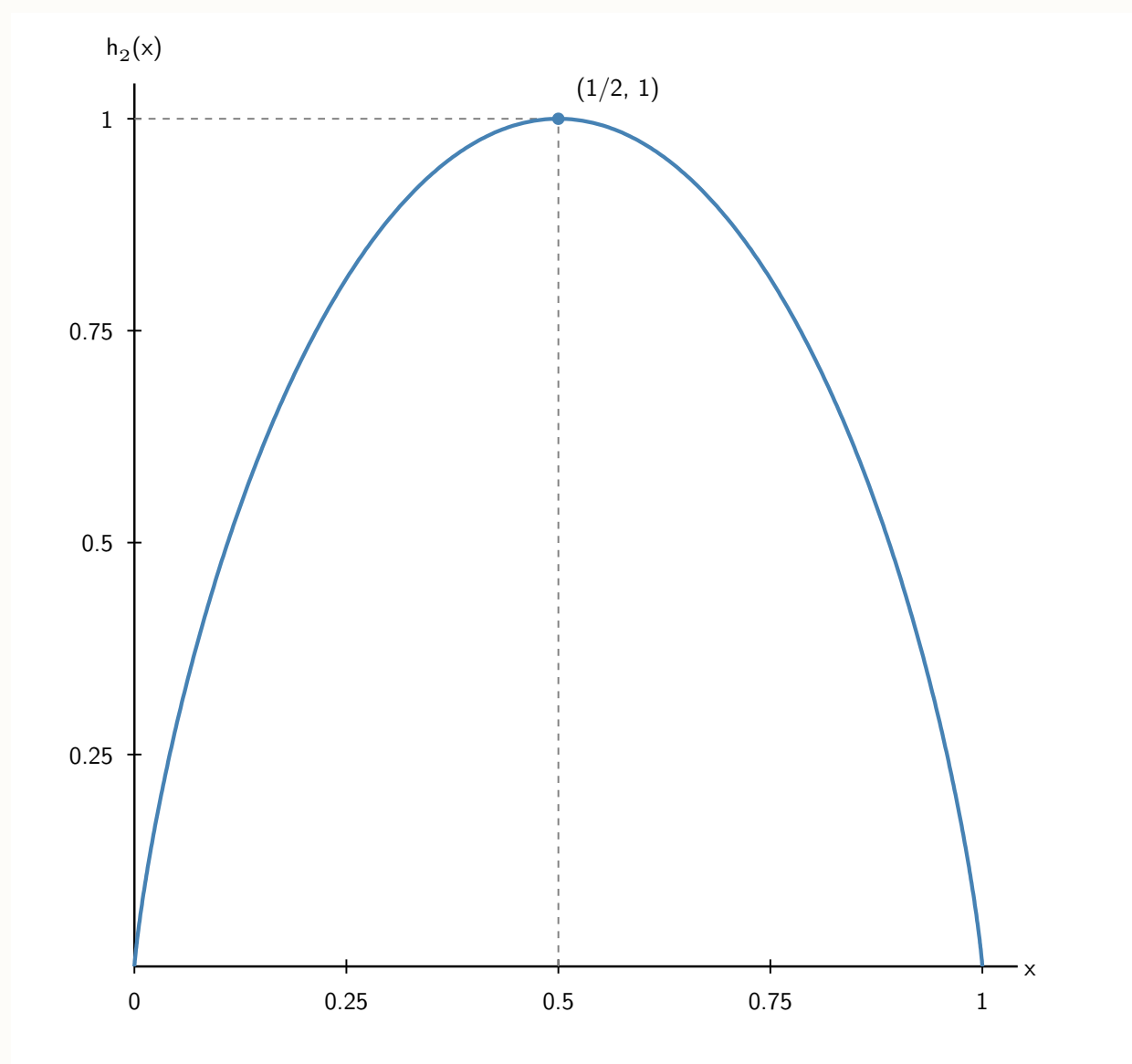
La fonction est strictement concave sur $]0, 1[$. Au bord gauche,

$$\frac{h_2(x)}{x} = \frac{h(x)}{x} + \frac{h(1 - x) - h(1)}{x} \rightarrow +\infty,$$

car le premier terme tend vers $+\infty$ et le second vers $-h'(1) = 1/c$. Par symétrie, le taux d'accroissement à gauche au bord droit tend vers $-\infty$. Les deux extrémités ont donc des demi-tangentes verticales; la tangente au sommet est horizontale.

x	0	$1/2$	1
$h_2(x)$	0	1	0
Variation	croissante	maximum	décroissante

Le graphe ci-dessous est calculé à partir de la formule de h_2 et présenté dans un repère orthonormé.



IV.5.c)

Méthode 1 — Étude de la fonction d'entropie binaire

Les probabilités de X sont p et $1 - p$, donc $H(X) = h_2(p)$. Par IV.5.b, le maximum unique de h_2 sur $[0, 1]$ est 1, atteint en $1/2$.

Ainsi $H(X) \leq 1$, avec égalité si et seulement si $p = 1/2$.

Méthode 2 — Inégalité des tangentes pour une fonction strictement concave

Pour $u \in]0, 1]$, la concavité de h donne

$$h(u) \leq h(1/2) + h'(1/2)(u - 1/2).$$

L'inégalité est stricte si $u \neq 1/2$, puisque h' est strictement décroissante. En l'appliquant à $u = p$ et à $u = 1 - p$, les termes linéaires se compensent, et

$$h(p) + h(1 - p) \leq 2h(1/2) = 1.$$

L'égalité exige $p = 1/2$, et cette valeur la réalise. La première approche utilise le graphe déjà étudié ; la seconde donne un argument de concavité réutilisable pour davantage de valeurs possibles.

IV.6)

La somme est impaire exactement pour les deux couples $(1, 0)$ et $(0, 1)$, qui correspondent à des événements disjoints. L'indépendance donne

$$p = P(Z = 1) = p_1(1 - p_2) + (1 - p_1)p_2 = p_1 + p_2 - 2p_1p_2.$$

Par définition, Z suit une loi de Bernoulli de paramètre p : $P(Z = 0) = 1 - p$ et $E(Z) = p$.

Enfin,

$$1 - 2p = 1 - 2p_1 - 2p_2 + 4p_1p_2 = (1 - 2p_1)(1 - 2p_2).$$

Les formules restent valides si l'un des paramètres de Bernoulli vaut 0 ou 1.

IV.7)

Méthode 1 — Récurrence sur le nombre d'épreuves

Notons r_n la probabilité d'obtenir un nombre impair de succès lors de n épreuves de Bernoulli indépendantes de paramètre p . Pour $n = 1$, $r_1 = p$, donc $1 - 2r_1 = 1 - 2p$.

La somme S_n des n premières épreuves suit la loi binomiale de paramètres n, p .

L'épreuve suivante est indépendante de S_n , et donc de l'indicatrice de sa parité, par le lemme des coalitions. L'application de IV.6 à cette indicatrice et à la nouvelle épreuve donne

$$r_{n+1} = r_n(1 - p) + (1 - r_n)p,$$

puis

$$1 - 2r_{n+1} = (1 - 2r_n)(1 - 2p).$$

Si $1 - 2r_n = (1 - 2p)^n$, cette dernière relation donne le résultat au rang suivant. Par récurrence, $1 - 2r_n = (1 - 2p)^n$ pour tout $n \geq 1$.

Comme X possède la même loi que S_n , la probabilité de parité impaire est la même. Donc

$$P(Z_n = 1) = \frac{1 - (1 - 2p)^n}{2}.$$

Méthode 2 — Formule du binôme et séparation des indices pairs et impairs

La formule du binôme donne

$$\sum_{k=0}^n (-1)^k \binom{n}{k} p^k (1-p)^{n-k} = ((1-p) - p)^n = (1-2p)^n.$$

Le membre de gauche est la somme des probabilités de $X = k$ pour les indices pairs, moins la somme pour les indices impairs. Il vaut donc

$P(X \text{ pair}) - P(X \text{ impair}) = 1 - 2P(Z_n = 1)$. On retrouve la même expression.

La récurrence explique comment une nouvelle épreuve modifie la parité ; le binôme donne un calcul direct à partir de la loi de X .

Puisque $-1 < 1 - 2p < 1$, on a $0 < P(Z_n = 1) < 1$. Par IV.5.c,

$$H(Z_n) = h_2\left(\frac{1 - (1 - 2p)^n}{2}\right) \leq 1.$$

L'égalité a lieu exactement lorsque $P(Z_n = 1) = 1/2$, donc lorsque $(1 - 2p)^n = 0$.

Puisque $n \geq 1$, cela équivaut à $p = 1/2$.

Partie V — Maximalité de l'entropie

V.1)

On fixe un entier $n \geq 2$ et on conserve $c = \ln(2) > 0$.

V.1.a)

Posons

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

Sur l'ouvert O , toutes les coordonnées $p_1, \dots, p_n$ sont strictement positives. La fonction h étant de classe C^2 sur $]0, 1[$ et p_n étant affine en les coordonnées, h_n est de classe C^2 sur O .

Pour $j \in \{1, \dots, n-1\}$,

$$\frac{\partial h_n}{\partial p_j} = h'(p_j) - h'(p_n) = \frac{1}{c} \ln \left(\frac{p_n}{p_j} \right).$$

Le gradient est nul si et seulement si $p_j = p_n$ pour chaque j . Comme la somme des n probabilités vaut 1, il existe un unique point critique,

$$P_* = (1/n, \dots, 1/n) \in O.$$

Tout extremum local dans l'ouvert O doit être un point critique ; il ne peut donc y avoir d'extremum en un autre point. Pour préciser la nature de P_* , on calcule la hessienne.

Pour chaque j , l'entrée diagonale vaut

$$\frac{\partial^2 h_n}{\partial p_j^2} = -\frac{1}{cp_j} - \frac{1}{cp_n}.$$

Pour $j \neq k$, l'entrée hors diagonale vaut

$$\frac{\partial^2 h_n}{\partial p_j \partial p_k} = -\frac{1}{cp_n}.$$

Notons J la matrice de taille $n-1$ dont toutes les entrées valent 1. Au point critique,

$$\nabla^2 h_n(P_*) = -\frac{n}{c}(I_{n-1} + J).$$

Le vecteur $\mathbf{1} = (1, \dots, 1)$ vérifie $J\mathbf{1} = (n-1)\mathbf{1}$: c'est donc un vecteur propre de la hessienne pour la valeur propre $-n^2/c$.

Pour tout vecteur $v = (v_1, \dots, v_{n-1})$ tel que

$$\sum_j v_j = 0,$$

on a $Jv = 0$ et la hessienne agit par multiplication par $-n/c$. Ce sous-espace a dimension $n-2$ et est supplémentaire de la droite engendrée par $\mathbf{1}$. Pour $n=2$, il est réduit au vecteur nul et la seule valeur propre est $-4/c$.

Toutes les valeurs propres sont strictement négatives. Le critère de la hessienne donne un maximum local strict en P_* . Il n'existe aucun autre extremum local sur O .

On peut également préciser que ce maximum est global. Le domaine O est convexe : toute combinaison convexe de deux de ses points conserve la positivité des coordonnées et celle de p_n . Pour un point $P \in O$ et un vecteur non nul $v \in \mathbb{R}^{n-1}$,

$${}^t v \nabla^2 h_n(P) v = -\frac{1}{c} \left(\sum_{j=1}^{n-1} \frac{v_j^2}{p_j} + \frac{(\sum_{j=1}^{n-1} v_j)^2}{p_n} \right) < 0.$$

Si $P \neq P_*$, on considère $g(t) = h_n(P_* + t(P - P_*))$ sur $[0, 1]$. La convexité du domaine assure que le segment y est inclus. On a $g'(0) = 0$ et $g''(t) < 0$; donc g' est strictement décroissante et $g'(t) < 0$ pour $t > 0$. Ainsi $g(1) < g(0)$.

En conclusion, h_n possède exactement un extremum sur O , un maximum global strict en P_* , de valeur $n h(1/n) = \ln(n)/c$.

V.1.b)

Méthode 1 — Inégalité de convexité rappelée dans l'énoncé

On a $(-h)''(x) = 1/(cx) > 0$ pour $0 < x < 1$, donc $-h$ est convexe sur $]0, 1[$. Sa continuité en 1 permet d'étendre les inégalités de convexité à l'extrémité 1, par passage à la limite ; elle est donc convexe sur $]0, 1]$.

Pour $p_1, \dots, p_n > 0$ de somme 1, la moyenne de ces nombres vaut $1/n$. L'inégalité rappelée donne

$$-h(1/n) \leq -\frac{1}{n} \sum_{k=1}^n h(p_k).$$

En multipliant par $-n$, on obtient

$$\sum_{k=1}^n h(p_k) \leq n h(1/n) = \frac{\ln(n)}{c}.$$

Méthode 2 — Somme des inégalités des tangentes

La concavité de h donne, pour chaque k ,

$$h(p_k) \leq h(1/n) + h'(1/n)(p_k - 1/n).$$

En sommant ces inégalités et en utilisant

$$\sum_k (p_k - 1/n) = 0,$$

on obtient à nouveau

$$\sum_k h(p_k) \leq n h(1/n) = \ln(n)/c.$$

La première approche applique directement l'inégalité fournie ; la seconde rend explicite la compensation du terme linéaire sous la contrainte de somme.

V.1.c)

Il faut tenir compte des probabilités éventuellement nulles, même lorsque $X(\Omega)$ est l'ensemble des n valeurs indiquées. Posons $p_k = P(X = x_k)$; alors $p_k \geq 0$ et

$$\sum_k p_k = 1.$$

Méthode 1 — Tangente au point uniforme, y compris au bord

Notons $a = 1/n \in]0, 1[$ et

$$d(u) = h(a) + h'(a)(u - a) - h(u), \quad u \in [0, 1].$$

Pour $u > 0$,

$$d'(u) = h'(a) - h'(u) = \frac{1}{c} \ln(u/a).$$

Ainsi d décroît strictement sur $]0, a[$, croît strictement sur $]a, 1]$ et atteint son minimum $d(a) = 0$ uniquement en a . De plus, par calcul direct,

$$d(0) = h(a) - a h'(a) = \frac{a}{c} > 0.$$

Donc $d(u) \geq 0$ sur tout $[0, 1]$, avec égalité exactement pour $u = a$. En sommant,

$$\sum_{k=1}^n d(p_k) = n h(a) - H(X) = \frac{\ln(n)}{c} - H(X) \geq 0.$$

L'égalité a lieu si et seulement si tous les $d(p_k)$ sont nuls, donc si et seulement si tous les p_k valent $1/n$. C'est précisément la loi uniforme.

Méthode 2 — Restriction au support de probabilité positive

Soit r le nombre de probabilités p_k strictement positives. Si $r = 1$, la variable est quasi certaine et $H(X) = 0 < \ln(n)/c$ puisque $n \geq 2$.

Si $r \geq 2$, les probabilités strictement positives sont toutes inférieures à 1 et définissent un point du domaine utilisé en V.1.a, avec r à la place de n . Cette question a établi un maximum global strict au point uniforme. Les probabilités nulles ne contribuant pas à $H(X)$,

$$H(X) \leq \frac{\ln(r)}{c} \leq \frac{\ln(n)}{c}.$$

Si $r < n$, la dernière inégalité est stricte. Si $r = n$, la première devient une égalité exactement lorsque toutes les probabilités valent $1/n$, par l'unicité du maximum établie en V.1.a.

Les deux méthodes donnent donc le même résultat et le même cas d'égalité. La première traite directement les bords ; la seconde distingue l'image de X de l'ensemble des valeurs de probabilité positive.

V.2)

On fixe $p \in]0, 1[$ et une variable G de loi géométrique de paramètre p , sur $\mathbb{N}^*$.

V.2.a)

La loi et l'espérance usuelles sont

$$p_k = p(1-p)^{k-1}, \quad m = E(G) = \frac{1}{p}.$$

Pour $k \geq 1$,

$$h(p_k) = -\frac{p_k}{c}(\ln(p) + (k-1)\ln(1-p)).$$

Ces termes sont positifs ou nuls. Les séries

$$\sum_k p_k$$

et

$$\sum_k (k-1)p_k$$

convergent, respectivement vers 1 et $m-1$, puisque G admet une espérance finie.

L'expression précédente montre donc que la série définissant $H(G)$ converge; aucune permutation de sommes infinies n'est nécessaire.

Par linéarité de séries convergentes,

$$H(G) = -\frac{\ln(p) + (m-1)\ln(1-p)}{c}.$$

Puisque $m = 1/p$, on obtient

$$H(G) = -\frac{\ln(p) + \frac{1-p}{p} \ln(1-p)}{\ln(2)} = \frac{h_2(p)}{p}.$$

V.2.b)

On commence par prouver (3), avec son cas d'égalité. Pour $x > 0$, posons $u(x) = x - 1 - \ln(x)$. On a $u'(x) = 1 - 1/x$, négatif sur $]0, 1[$ et positif sur $]1, +\infty[$. La fonction atteint donc son minimum $u(1) = 0$ uniquement en 1. Ainsi

$$\ln(x) \leq x - 1, \quad \text{avec égalité si et seulement si } x = 1.$$

On peut également obtenir cette inégalité par la concavité de $\ln$, dont la tangente en 1 est $x \mapsto x - 1$; la stricte concavité donne le même cas d'égalité.

Les nombres p_k et q_k étant strictement positifs, on applique (3) à p_k/q_k :

$$\ln(p_k) - \ln(q_k) \leq \frac{p_k}{q_k} - 1.$$

Pour établir l'inégalité d'entropie et son cas d'égalité, il faut contrôler les séries. Comme $E(X) = m$,

$$\sum_{k=1}^{\infty} q_k |\ln(p_k)| = -\ln(p) \sum_{k=1}^{\infty} q_k - \ln(1-p) \sum_{k=1}^{\infty} (k-1)q_k < +\infty.$$

De plus, $0 < q_k \leq 1$ et l'existence de $H(X)$ assurent

$$\sum_{k=1}^{\infty} q_k |\ln(q_k)| = c H(X) < +\infty.$$

Les deux séries logarithmiques sont donc absolument convergentes. Enfin,

$$\sum_k |p_k - q_k| \leq \sum_k (p_k + q_k) = 2.$$

En multipliant l'inégalité par q_k et en sommant d'abord jusqu'à un entier N , puis en faisant tendre N vers l'infini, on obtient

$$\sum_{k=1}^{\infty} q_k \ln(p_k) - \sum_{k=1}^{\infty} q_k \ln(q_k) \leq \sum_{k=1}^{\infty} (p_k - q_k) = 0.$$

La contrainte d'espérance donne

$$\sum_{k=1}^{\infty} q_k \ln(p_k) = \ln(p) + (m-1) \ln(1-p) = -c H(G).$$

Par ailleurs,

$$\sum_k q_k \ln(q_k) = -c H(X).$$

L'inégalité devient donc $c(H(X) - H(G)) \leq 0$, soit $H(X) \leq H(G)$.

Pour l'égalité, on conserve les écarts positifs ou nuls

$$D_k = p_k - q_k - q_k \ln(p_k/q_k) \geq 0.$$

Les convergences déjà établies autorisent leur somme, et

$$\sum_{k=1}^{\infty} D_k = c(H(G) - H(X)).$$

Si les deux entropies sont égales, cette série de termes positifs ou nuls a une somme nulle : chacun de ses termes est nul. Le cas d'égalité de (3), et $q_k > 0$, donnent $p_k/q_k = 1$ pour tout k . Les deux lois sont identiques.

Réciproquement, si $q_k = p_k$ pour tout k , les entropies sont égales. La loi géométrique est donc l'unique loi de la classe considérée qui maximise l'entropie à espérance fixée m .

Partie VI — Incertitude d'une variable aléatoire continue

Question adaptée au programme actuel : rectification du préambule de la partie VI. La fonction h est étendue à $[0, +\infty[$ par $h(0) = 0$ et $h(u) = -u \ln(u)/c$ pour $u > 0$. Cette extension est nécessaire car une densité peut prendre des valeurs supérieures à 1 ; le domaine $[0, 1]$ donné initialement ne permettait donc pas toutes les expressions demandées.

Contrairement à l'entropie d'une variable discrète, l'entropie ainsi définie pour une densité peut être négative, puisque $h(u) < 0$ lorsque $u > 1$.

VI.1)

On calcule l'entropie des lois normales en justifiant chaque fois la convergence absolue de l'intégrale.

VI.1.a)

Une densité de Y_0 est

$$f_{Y_0}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}.$$

Elle est strictement positive et

$$\ln(f_{Y_0}(x)) = -\frac{1}{2} \ln(2\pi) - \frac{x^2}{2}.$$

Ainsi,

$$h(f_{Y_0}(x)) = \frac{f_{Y_0}(x)}{2c} (\ln(2\pi) + x^2) \geq 0.$$

Les propriétés de la loi normale centrée réduite et le théorème de transfert donnent $\int_{\mathbb{R}} f_{Y_0}(x) dx = 1$ et $\int_{\mathbb{R}} x^2 f_{Y_0}(x) dx = E(Y_0^2) = 1$. L'intégrale de $|h(f_{Y_0})|$ est donc finie : $H(Y_0)$ existe.

Par linéarité d'intégrales convergentes,

$$H(Y_0) = \frac{\ln(2\pi) + 1}{2 \ln(2)} = \frac{\ln(2\pi e)}{2 \ln(2)}.$$

VI.1.b)

Méthode 1 — Utilisation du moment centré d'ordre deux

La densité de Y est

$$f_Y(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Posons $A = \ln(\sigma\sqrt{2\pi})$. Son logarithme vaut $-A - (x-m)^2/(2\sigma^2)$, donc

$$|h(f_Y(x))| \leq \frac{f_Y(x)}{c} \left(|A| + \frac{(x-m)^2}{2\sigma^2} \right).$$

Cette majorante est intégrable puisque $\int f_Y = 1$ et, par transfert, $\int (x-m)^2 f_Y(x) dx = V(Y) = \sigma^2$. L'entropie existe donc absolument.

On peut alors intégrer l'expression de $h(f_Y)$:

$$H(Y) = \frac{1}{c} \left(A + \frac{V(Y)}{2\sigma^2} \right) = \frac{\ln(\sigma\sqrt{2\pi}) + 1/2}{c}.$$

Ainsi,

$$H(Y) = \frac{\ln(2\pi e\sigma^2)}{2\ln(2)}.$$

Méthode 2 — Changement de variable affine et effet d'échelle

Le changement de variable $x = m + \sigma t$ donne $dx = \sigma dt$ et $f_Y(m + \sigma t) = f_{Y_0}(t)/\sigma$. Pour tout $u > 0$,

$$\sigma h(u/\sigma) = h(u) + \frac{u \ln(\sigma)}{c}.$$

Sur les segments transformés, cette identité permet de remplacer l'intégrande. Elle donne aussi

$$\sigma |h(f_{Y_0}(t)/\sigma)| \leq |h(f_{Y_0}(t))| + \frac{|\ln(\sigma)|}{c} f_{Y_0}(t).$$

Les deux termes de droite sont intégrables d'après VI.1.a et la normalisation de la densité. Le changement de variable s'étend donc aux intégrales impropres par passage à la limite et assure leur convergence absolue.

En intégrant l'identité exacte,

$$H(Y) = H(Y_0) + \frac{\ln(\sigma)}{c},$$

ce qui redonne la formule précédente. Le calcul par moments est direct ; le changement de variable montre que la translation ne modifie pas l'entropie et que l'échelle σ ajoute $\ln(\sigma)/c$.

VI.2)

Sur $]0, +\infty[$, la densité exponentielle est $f_0(x) = \lambda e^{-\lambda x}$. On peut la choisir nulle sur $] -\infty, 0]$, la valeur au point 0 ne modifiant aucun calcul d'intégrale.

VI.2.a)

Pour $x > 0$,

$$\ln(f_0(x)) = \ln(\lambda) - \lambda x,$$

et

$$|h(f_0(x))| \leq \frac{f_0(x)}{c} (|\ln(\lambda)| + \lambda x).$$

La loi exponentielle est normalisée et admet l'espérance $E(X_0) = 1/\lambda$. Par transfert, l'intégrale de cette majorante est $(|\ln(\lambda)| + 1)/c < +\infty$. Comme $h(f_0) = 0$ sur les réels négatifs, $H(X_0)$ existe absolument sur $\mathbb{R}$.

En intégrant,

$$H(X_0) = \frac{-\ln(\lambda) + \lambda E(X_0)}{c} = \frac{1 - \ln(\lambda)}{\ln(2)}.$$

En particulier, cette entropie est nulle pour $\lambda = e$, positive pour $0 < \lambda < e$ et négative pour $\lambda > e$. Il n'y a pas de contradiction avec IV.4, qui concernait les variables discrètes.

VI.2.b)

On choisit une densité f nulle pour $x \leq 0$, ce qui est possible puisque X est à valeurs

strictement positives. Alors $\int_0^{+\infty} f(x) dx = 1$ et, par l'existence de l'espérance et le transfert,

$$\int_0^{+\infty} x f(x) dx = \frac{1}{\lambda}.$$

Le logarithme de f_0 donne la majoration

$$|f(x) \ln(f_0(x))| \leq f(x)(|\ln(\lambda)| + \lambda x).$$

Le membre de droite est intégrable, donc l'intégrale croisée converge absolument. On peut calculer

$$\int_0^{+\infty} f(x) \ln(f_0(x)) dx = \ln(\lambda) - \lambda E(X) = \ln(\lambda) - 1.$$

Par VI.2.a, cela établit

$$H(X_0) = -\frac{1}{c} \int_0^{+\infty} f(x) \ln(f_0(x)) dx.$$

Pour comparer les entropies, considérons d'abord un point $x > 0$ tel que $f(x) > 0$. On a aussi $f_0(x) > 0$, donc (3) s'applique au quotient $f_0(x)/f(x)$:

$$f(x) \ln \left(\frac{f_0(x)}{f(x)} \right) \leq f_0(x) - f(x).$$

Si $f(x) = 0$, on définit le membre de gauche par 0, conformément à la continuité de $u \ln(u)$ en 0. L'inégalité devient $0 \leq f_0(x)$ et reste vraie. Ainsi aucune division par une densité nulle n'est effectuée.

Posons $g(x) = f(x) \ln(f_0(x)) - f(x) \ln(f(x))$, avec la convention $f(x) \ln(f(x)) = 0$ quand $f(x) = 0$. Sur chaque segment $[a, b] \subset]0, +\infty[$ évitant les éventuels points singuliers de la densité, l'intégration de l'inégalité donne

$$\int_a^b g(x) dx \leq \int_a^b (f_0(x) - f(x)) dx.$$

Lorsqu'un segment contient un point singulier, on le partage en sous-segments et on prend les limites correspondantes. L'intégrale de $f \ln(f_0)$ converge absolument, celle de $f \ln(f)$ converge par l'hypothèse d'existence de $H(X)$, et celles de f_0 et f convergent par normalisation. Il est donc légitime de faire tendre les bornes vers 0 et $+\infty$, en respectant les limites aux éventuels points singuliers.

On obtient

$$\int_0^{+\infty} g(x) dx \leq \int_0^{+\infty} (f_0(x) - f(x)) dx = 0.$$

Par la définition de $H(X)$ et l'identité croisée déjà établie,

$$\int_0^{+\infty} g(x) dx = -c H(X_0) + c H(X).$$

Ainsi $c(H(X) - H(X_0)) \leq 0$. Comme $c > 0$, la conclusion est

$$H(X) \leq H(X_0) = \frac{1 - \ln(\lambda)}{\ln(2)}.$$

Relevé des adaptations

- I.2 : code adapté en Python ; remplacement de la procédure par une fonction renvoyant la liste, conservation d'un emplacement à compléter, passage à des indices commençant en 0 et rappel du quotient entier et du reste.
- I.3 : code adapté en Python ; conversion du tableau des familles et des valeurs, correction cohérente des indices et utilisation de la liste renvoyée par bin. La numérotation des cartes est conservée.
- II.3.b : précision que la dérivée au point 1 est une dérivée à gauche ; il s'agit de la convention nécessaire à une fonction définie sur $]0, 1]$, sans changement du résultat demandé.
- Partie VI, préambule commun à VI.1.a, VI.1.b, VI.2.a et VI.2.b : rectification du domaine de h , étendu de $[0, 1]$ à $[0, +\infty[$ avec la même formule. Une densité peut dépasser 1, notamment pour les lois normales de petit écart type et les lois exponentielles de taux supérieur à 1. L'extension rend les expressions demandées bien définies et préserve toutes les questions.