L'entropie en probabilité, approche élémentaire. Une correction.

Préliminaire.

1. • On a $\lim_{x\to 0^+} x\ln x = 0 = h(0)$ (par croissances comparées). Puis h est continue sur $]0,+\infty[$ (par exemple comme produit de fonctions continues sur cet intervalle). Ainsi $\boxed{h \text{ est continue sur } \mathbb{R}^+}$.

   • Si $x>0$, on a : $\dfrac{h(x)-h(0)}{x} = \ln x \underset{x\to 0}{\longrightarrow} -\infty$. Il en résulte que h n'est pas dérivable en 0 (mais son graphe admet une tangente « verticale » au point d'abscisse 0).

   • Les antécédents de 0 par h sont 0 et 1.

2. La fonction ln a une dérivée seconde strictement négative sur $]0,+\infty[$: c'est donc une fonction **strictement** concave. Son graphe est donc situé au dessous de ses tangentes et ne rencontre ces droites qu'au point de tangence. Comme $y = x-1$ est l'équation de la tangente au graphe du logarithme naturel au point d'abscisse 1, on peut conclure que, pour tout $x>0$, on a $\ln x \leqslant x-1$ avec égalité si et seulement si $x=1$.

   Autre idée. Etudier la fonction $x \mapsto x-1-\ln x\ldots$

Partie 1. Exemples.

3. On a $\boxed{H(U_n) = -\sum_{i=1}^{n} \frac{1}{n}\ \ln\frac{1}{n} = \ln n}$.
4. Notons $X(\Omega) = \{x_1,\ldots,x_n\}$.

   • Si X est presque sûrement constante, il existe $i_0 \in \{1,\ldots,n\}$ tel que $P(X=x_{i_0}) = 1$. On a alors $P(X=x_i)=0$ pour tout $i\neq i_0$ dans $\{1,\ldots,n\}$ et l'entropie de X est alors :

   $$H(X) = -\sum_{i=1}^{n} h(P(X=x_i)) = -h(1) = 0.$$

   • Réciproquement, supposons $H(X)=0$. Ainsi $\sum_{i=1}^{n} h(P(X=x_i)) = 0$. Chaque $P(X=x_i)$ est dans $[0,1]$, donc chaque $h(P(X=x_i))$ est négatif. Il en résulte que :

   $$h(P(X=x_i)) = 0 \ \text{ pour tout } i \text{ dans } \{1,\ldots,n\}.$$

   Ainsi, pour i dans $\{1,\ldots,n\}$, on a $P(X=x_i)\in h^{-1}(0)$. Mais 0 a deux antécédents par h qui sont 1 et 0. On ne peut avoir $P(X=x_i)=0$ pour tout i dans $\{1,\ldots,n\}$: il existe donc i_0 dans $\{1,\ldots,n\}$ tel que $P(X=x_{i_0})=1$.

   Conclusion. $\boxed{H(X)=0 \text{ si et seulement si } X \text{ est presque sûrement constante}}$.

5. • La fonction g est continue sur $[0,1]$ et dérivable au moins sur $]0,1[$ par composition. Pour $x \in]0,1[$,

$$g'(x) = -h'(x) + h'(1-x) = -(\ln(x)+1) + (\ln(1-x)+1) = \ln\left(\frac{1-x}{x}\right)$$

Ainsi, pour $x \in]0,1[$, on a : $g'(x) \geqslant 0 \Leftrightarrow \dfrac{1-x}{x} \geqslant 1 \Leftrightarrow \dfrac{1}{2} \geqslant x$, avec égalité si et seulement si $x = \frac{1}{2}$. On peut alors sans peine établir le tableau de variation de g en précisant que $g(0) = g(1) = 0$ et que $g(\frac{1}{2}) = \ln 2$ est un **maximum global strict** de g.

• On a $P(X=1) = p, P(X=0) = 1-p$ donc $H(X) = -h(p) - h(1-p) = g(p) \leqslant \ln 2$ avec égalité si, et seulement si, $p = \dfrac{1}{2}$.

6. On a $(Z=1) = (\text{« } X_1 + X_2 \text{ est impair »}) = [(X_1 = 1) \cap (X_2 = 0)] \cup [(X_1 = 0) \cap (X_2 = 1)]$. Ainsi (**union disjointe**) :

$$p = P(Z=1) = P((X_1 = 1) \cap (X_2 = 0)) + P(X_1 = 0) \cap (X_2 = 1)$$

Puis, par **indépendance**, de X_1 et X_2 :

$$p = P(X_1 = 1)P(X_2 = 0) + P(X_1 = 0)P(X_2 = 1) = p_1(1-p_2) + (1-p_1)p_2.$$

Remarque. On peut aussi rédiger avec la formule des probabilités totales et le système complet $((X_1 = 1), (X_1 = 0))$.

Conclusion. $\boxed{p = P(Z=1) = p_1 + p_2 - 2p_1p_2, \quad \mathbb{E}(Z) = p, \quad (1-2p) = (1-2p_1)(1-2p_2)}$.

7. Le cas $n = 1$ est évident. Soit $n \in \mathbb{N}^*$, on suppose le résultat au rang n.
Soit X qui suit $\mathcal{B}(n+1, p)$.
On sait que X **est la somme de variables de Bernoulli indépendantes** $X_1, X_2, \ldots, X_{n+1}$ **de paramètre** p, on peut écrire $X = X_1 + X_2 + \ldots + X_{n+1} = X_1 + Y$ où Y est une variable qui suit $\mathcal{B}(n,p)$.
Soit Z_n et Z_{n+1} les variables de Bernoulli telle que $P(Z_n = 1) = P(\text{« } Y \text{ est impair »})$ et $P(Z_{n+1} = 1) = P(\text{« } X \text{ est impair »})$.
On a ainsi $(Z_{n+1} = 1) = (\text{« } X_1 + Z_n \text{ est impair »})$.
Or Z_n et X_1 sont indépendantes ; la question précédente et l'hypothèse de récurrence donnent :

$$1 - 2P(Z_{n+1} = 1) = (1-2p)(1 - 2P(Z_n = 1)) = (1-2p)(1-2p)^n = (1-2p)^{n+1}$$

D'où le résultat par récurrence.
On a **d'après I.1c** : $H(Z_n) \leqslant \ln 2$ et égalité si et seulement si $P(Z_n = 1) = \frac{1}{2}$ c'est-à-dire $(1-2p)^n = 0$, $\boxed{H(Z_n) = \ln 2 \text{ si et seulement si } p = 1/2}$.

Partie 2. Propriétés de l'entropie.

8. On a $m = E(X_0) = \dfrac{1}{p}$. Soit $k \in \mathbb{N}^*$. On a $p_k = P(G = k) = (1-p)^{k-1}p$ donc :

$$h(p_k) = p_k \ln(p_k) = \ln(p)p_k - (k-1)\ln(1-p)p_k.$$

Les séries $\sum_{k\geqslant 1} p_k$ et $\sum_{k\geqslant 1} kp_k$ étant convergentes d'après le cours, de sommes respectives 1 et $\mathbb{E}(X_0) = \frac{1}{p}$, $H(X_0)$ existe et :

$$\begin{aligned} H(X_0) &= -\ln(p)\sum_{k=1}^{+\infty} p_k - \ln(1-p)\sum_{k=1}^{+\infty}(k-1)p_k = -\ln(p) - \ln(1-p)\left(\mathbb{E}(X_0)-1\right) \\ &= -\ln(p) - \ln(1-p)\left(\frac{1}{p}-1\right) \end{aligned}$$

Conclusion. $\boxed{H(X_0) = -\left[\ln(p) + \ln(1-p)\left(\frac{1}{p}-1\right)\right]}$.

9. Soit k dans $\mathbb{N}^*$. On a : $q_k \ln p_k = q_k[(k-1)\ln(1-p)+\ln p] = q_k \ln p + kq_k\ln(1-p) - q_k\ln(1-p)$. Or la série $\sum_{k\geqslant 1} q_k$ converge de somme 1 et la série $\sum_{k\geqslant 1} kq_k$ converge de somme $\mathbb{E}(X)$. Ainsi $\sum_{k\geqslant 1} q_k\ln(p_k)$ est convergente et :

$$-\sum_{k=1}^{+\infty} q_k\ln(p_k) = -\ln(p) - \ln(1-p)\mathbb{E}(X) + \ln(1-p) = H(X_0),$$

puisque $\mathbb{E}(X_0) = \mathbb{E}(X) = \frac{1}{p}$.

Conclusion. $\boxed{H(X_0) = -\sum_{k=1}^{+\infty} q_k\ln(p_k)}$.

10. • Soit $k \in \mathbb{N}^*$. La question 2 du préliminaire appliquée à $x = \frac{p_k}{q_k} > 0$ donne :

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

On multiplie par q_k pour obtenir :

$$\boxed{-h(q_k) + q_k\ln(p_k) \leqslant p_k - q_k}.$$

• Par somme, (les séries sont convergentes) :

$$H(X) - X(X_0) \leqslant \sum_{k=1}^{+\infty} p_k - \sum_{k=1}^{+\infty} q_k = 1 - 1 = 0.$$

Conclusion. $\boxed{H(X) \leqslant H(X_0)}$.

11. Si $H(X) = H(X_0)$ alors on a :

$$H(X) - H(X_0) - \sum_{k=1}^{+\infty} p_k - \sum_{k=1}^{+\infty} q_k = 0$$

donc, selon la question précédente, $\sum_{k=0}^{+\infty} \underbrace{-h(q_k) + q_k\ln(p_k) \leqslant p_k - q_k}_{\leqslant 0} = 0$. Il en résulte que, pour tout $k \in \mathbb{N}^*$,

$$\boxed{-h(q_k) + q_k\ln(p_k) = p_k - q_k}.$$

Ainsi $\ln\frac{p_k}{q_k} = \frac{p_k}{q_k} - 1$ donc $\frac{p_k}{q_k} = 1$ (préliminaire...).

Conclusion. $\boxed{\text{Si } H(X) = H(X_0) \text{ alors } X \text{ suit la même loi que } X_0}$.

Partie 3. Entropie d'un couple.

12. **Propriétés de l'information entre deux couples**

a. La famille $([X=i]\cap[Y=j])_{(i,j)\in I_n\times I_m}$ forme un système complet d'événements. Il en résulte que $\boxed{\displaystyle\sum_{i=1}^n\sum_{j=1}^m\lambda_{i,j}=1}$. De même $\boxed{\displaystyle\sum_{i=1}^n\sum_{j=1}^m\mu_{i,j}=1}$.

b. On a :

$$\begin{aligned}-\sum_{i=1}^n\sum_{j=1}^m\lambda_{i,j}\left(\ln\frac{\mu_{i,j}}{\lambda_{i,j}}-\frac{\mu_{i,j}}{\lambda_{i,j}}+1\right) &= -\sum_{i=1}^n\sum_{j=1}^m\lambda_{i,j}\ln\frac{\mu_{i,j}}{\lambda_{i,j}}+\underbrace{\sum_{i=1}^n\sum_{j=1}^m\mu_{i,j}}_{=1}-\underbrace{\sum_{i=1}^n\sum_{j=1}^m\lambda_{i,j}}_{=1}\\ &= K(X,Y,X',Y')\end{aligned}$$

Conclusion. $\boxed{\displaystyle K(X,Y,X',Y')=-\sum_{i=1}^n\sum_{j=1}^m\lambda_{i,j}\left(\ln\frac{\mu_{i,j}}{\lambda_{i,j}}-\frac{\mu_{i,j}}{\lambda_{i,j}}+1\right)}$.

c. Pour tout $x>0$, on a vu que $\ln x\leqslant x-1$. Il en résulte que, pour tout $(i,j)\in I_n\times I_m$, puisque chaque $\dfrac{\mu_{i,j}}{\lambda_{i,j}}$ est strictement positif, on a :

$$\ln\frac{\mu_{i,j}}{\lambda_{i,j}}-\frac{\mu_{i,j}}{\lambda_{i,j}}+1\leqslant 0.$$

Comme $\lambda_{i,j}>0$, pour tout $(i,j)\in I_n\times I_m$, l'expression de la question précédente permet d'affirmer que $\boxed{K(X,Y,X',Y')\geqslant 0}$.

d. • Supposons que les couples (X,Y) et (X',Y') ont même loi conjointe. On a alors $\lambda_{i,j}=\mu_{i,j}$ pour tout $(i,j)\in I_n\times I_m$. Il en résulte que $K(X,Y,X',Y')=0$.

• Réciproquement, supposons que $K(X,Y,X',Y')=0$. Pour tout $(i,j)\in I_n\times I_m$, on a vu que :

$$-\lambda_{i,j}\left(\ln\frac{\mu_{i,j}}{\lambda_{i,j}}-\frac{\mu_{i,j}}{\lambda_{i,j}}+1\right)\geqslant 0.$$

Ainsi, $K(X,Y,X',Y')$ est une somme de termes positifs qui est nulle : chacun des termes de la somme est nul. Il en résulte que, pour tout $(i,j)\in I_n\times I_m$, on a

$$\ln\frac{\mu_{i,j}}{\lambda_{i,j}}-\frac{\mu_{i,j}}{\lambda_{i,j}}+1=0.$$

La question 2 du préliminaire nous affirme alors que $\dfrac{\mu_{i,j}}{\lambda_{i,j}}=1$ pour tout $(i,j)\in I_n\times I_m$: les couples (X,Y) et (X',Y') ont même loi conjointe.

Conclusion. $K(X,Y,X',Y')=0$ si et seulement si les couples (X,Y) et (X',Y') ont même loi conjointe.

e. i. D'après les hypothèses, X' et Y' sont indépendantes, de même loi que X et Y respectivement. Ainsi, pour tout $(i,j)\in I_n\times I_m$, on a :

$$\mu_{i,j}=p_iq_j.$$

Il vient alors :

$$\begin{aligned}
K(X,Y,X',Y') &= -\sum_{i=1}^{n}\sum_{j=1}^{m}\lambda_{i,j}\ln\frac{\mu_{i,j}}{\lambda_{i,j}} = -\sum_{i=1}^{n}\sum_{j=1}^{m}\lambda_{i,j}(\ln\mu_{i,j}-\ln\lambda_{i,j})\\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m}\lambda_{i,j}(\ln p_i+\ln q_j)+\underbrace{\sum_{i=1}^{n}\sum_{j=1}^{m}\lambda_{i,j}\ln\lambda_{i,j}}_{=H(X,Y)}\\
&= -\sum_{i=1}^{n}\left(\sum_{j=1}^{m}\lambda_{i,j}\right)\ln p_i-\sum_{j=1}^{m}\left(\sum_{i=1}^{n}\lambda_{i,j}\right)\ln q_j+H(X,Y)
\end{aligned}$$

Or la famille $([X=i])_{1\leqslant i\leqslant n}$ forme un système complet d'événements, ce qui implique que, pour tout j dans I_m, on a :

$$\sum_{i=1}^{n}\lambda_{i,j}=\sum_{i=1}^{n}P(X=i,Y=j)=P(Y=j)=q_j.$$

De même, $\displaystyle\sum_{j=1}^{m}\lambda_{i,j}=p_i$, pour tout i dans I_n. Il en résulte que :

$$K(X,Y,X',Y') = \underbrace{-\sum_{i=1}^{n}p_i\ln p_i}_{=H(X)}+\underbrace{\left(-\sum_{j=1}^{m}q_j\ln q_j\right)}_{H(Y)}+H(X,Y)$$

Conclusion. $\boxed{K(X,Y,X',Y')=H(X)+H(Y)-H(X,Y)}$.

ii. On a : $H(X,Y)-H(X)+H(Y)=-K(X,Y,X',Y')\leqslant 0$ d'après la question III.1c.

Ainsi $\boxed{H(X,Y)\leqslant H(X)+H(Y)\quad(\clubsuit)}$.

Cette inégalité est une égalité si et seulement si $K(X,Y,X',Y')=0$ i.e. si et seulement si $\boxed{\text{les couples } (X,Y) \text{ et } (X',Y') \text{ ont même loi conjointe}}$.

13. **Entropie conditionnelle**

a. D'après la question III.2e.ii, on a :

$$H(Y\mid X)=H(X,Y)-H(X)\leqslant H(X)+H(Y)-H(X)=H(Y).$$

Conclusion. $\boxed{H(Y\mid X)\leqslant H(Y)}$.

b. • Pour tout i dans I_m on a : $\ln a_i\leqslant\ln(a_1+\cdots+a_m)$. Il en résulte que, pour tout i dans I_n, on a :

$$h(a_i)\leqslant a_i\ln(a_1+\cdots+a_m).$$

Par somme membre à membre, on obtient :

$$\sum_{i=1}^{m}h(a_i)\leqslant\sum_{i=1}^{m}a_i\ln(a_1+\cdots+a_m)=\ln(a_1+\cdots+a_m)\sum_{i=1}^{m}a_i=h\left(\sum_{i=1}^{m}a_i\right).$$

Conclusion. $\boxed{\sum_{i=1}^{m}h(a_i)\leqslant h\left(\sum_{i=1}^{m}a_i\right)\quad(\spadesuit)}$.

• Si tous les a_i sont nuls, (♠) est encore vraie de manière évidente (les deux membres valent 0). Si l'un des a_i est non nul, on a encore, pour tout i dans I_n, l'inégalité :

$$h(a_i) \leqslant a_i \ln(a_1 + \cdots + a_m).$$

Ainsi $\boxed{(♠) \text{ est encore vraie si } a_1, \ldots, a_n \text{ sont dans } [0,1]}$.

• Supposons qu'il existe deux indices i dans I_n tels que $a_i \neq 0$. Par exemple $a_1 > 0$ et $a_2 > 0$. On a alors :

$$\ln(a_1) < \ln(a_1 + \cdots + a_m).$$

Ainsi (♠) est une inégalité stricte.

Si maintenant tous les a_i sont nuls, on a vu que (♠) est une égalité. Enfin, supposons, par exemple, que $a_1 > 0$ et que tous les autres a_i sont nuls. On a cette fois :

$$h(a_i) = a_i \ln(a_1 + \cdots + a_m)$$

pour tout i dans I_m : (♠) est une égalité.

Conclusion. Il y a égalité dans (♠) si et seulement s'il existe au plus un indice i dans I_m tel que $a_i \neq 0$.

c. • Soit i dans I_n. Selon la question précédente, on a :

$$\sum_{j=1}^{m} h(\lambda_{i,j}) \leqslant h\left(\sum_{j=1}^{m} \lambda_{i,j}\right) = h(p_i).$$

• Il en résulte que :

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} h(\lambda_{i,j}) \geqslant -\sum_{i=1}^{n} h(p_i) = H(X)$$

donc $\boxed{H(Y \mid X) = H(X,Y) - H(X) \geqslant 0}$.

d. • On a, par hypothèse, $H(Y \mid X) = 0$ donc $\displaystyle\sum_{i=1}^{n} \underbrace{\left(\sum_{j=1}^{m} h(\lambda_{i,j}) - h(p_i)\right)}_{=\alpha_i} = 0$. D'après la question III.2b, chaque α_i est positif. Ainsi pour tout i dans I_n, $\alpha_i = 0$ donc

$$\sum_{j=1}^{m} h(\lambda_{i,j}) = h(p_i) = h\left(\sum_{j=1}^{m} \lambda_{i,j}\right).$$

Pour tout i dans I_n, selon III.2b, il existe au plus un seul indice j dans I_n tel $\lambda_{i,j}$ non nul. En fait, si $p_i > 0$, il en existe exactement un, sinon p_i serait nul.

Conclusion. Pour tout $i \in I_n$ tel que $p_i > 0$, il existe un unique j dans I_m tel que $\lambda_{i,j} > 0$. On pose $j = \alpha(i)$.

• On a ainsi défini une fonction $\alpha : I_n \to I_m$ et on a :

$$\begin{aligned} P(Y = \alpha(X)) &= \sum_{i=1}^{n} P([Y = \alpha(X)] \cap [X = i]) = \sum_{i=1}^{n} P([Y = \alpha(i)] \cap [X = i]) \\ &= \sum_{i=1}^{n} \lambda_{i,\alpha(i)} \end{aligned}$$

Mais, pour tout i dans I_n, si $p_i = 0$ on a $\lambda_{i,j} = 0$ et si $p_i > 0$, alors $\lambda_{i,j} = 0$ pour $j \neq \alpha(i)$. Il en résulte que :

$$P(Y = \alpha(X)) = \sum_{i=1}^{n} \lambda_{i,\alpha(i)} = \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_{i,j} = 1.$$

Conclusion. $\boxed{\text{On a } Y = \alpha(X) \text{ presque sûrement}}$.

• Ce résultat est naturel : si on n'a aucune incertitude sur la valeur de Y, celle de X étant connue, c'est que Y est presque sûrement une fonction de X...

Fin de la correction
