Entropie et information mutuelle

`Exercice 1` `Définition axiomatique de l'entropie`

Étant donnée une distribution de probabilité $p_1, \dots, p_n$, on cherche une fonction $H(p_1, \dots, p_n)$ quantifiant « l'incertitude » associée à cette distribution. On pose les conditions suivantes.

(A1) $H(p_1, \dots, p_n)$ est maximum pour $p_1 = \dots = p_n = 1/n$.

(A2) H est une fonction symétrique en ses arguments.

(A3) $H(p_1, \dots, p_n) \geq 0$ avec égalité lorsqu'un des p_i vaut 1.

(A4) $H(p_1, \dots, p_n, 0) = H(p_1, \dots, p_n)$.

(A5) $H(\frac{1}{n}, \dots, \frac{1}{n}) \leq H(\frac{1}{n+1}, \dots, \frac{1}{n+1})$.

(A6) la fonction H est continue.

(A7) pour des entiers m et n, $H(\frac{1}{mn}, \dots, \frac{1}{mn}) = H(\frac{1}{m}, \dots, \frac{1}{m}) + H(\frac{1}{n}, \dots, \frac{1}{n})$.

(A8) pour $p = p_1 + \dots + p_m$ et $q = q_1 + \dots + q_n$, où tous les p_i, q_i sont positifs. Si p et q sont strictement positifs tels que $p + q = 1$, on a

$$H(p_1, \dots, p_m, q_1, \dots, q_n) = H(p, q) + pH(p_1/p, \dots, p_m/p) + qH(q_1/q, \dots, q_n/q).$$

1. Justifier les différents axiomes.

2. Montrer que la fonction $g(n) = H(\frac{1}{n}, \dots, \frac{1}{n})$ est de la forme $g(n) = A \log n$ pour un certain A.

3. En déduire la forme de $H(p, 1-p)$ quand p est rationnel.

4. Conclure.

`Exercice 2` `Quasi-distance`

Soit E l'ensemble des variables aléatoires (sur un espace probabilisé donné) à valeurs dans un ensemble fini donné $\mathcal{X} = \{x_1, \dots, x_N\}$. Pour tout $(X, Y) \in E \times E$, on pose

$$d(X, Y) = H(X|Y) + H(Y|X).$$

1. L'application d ainsi définie est-elle une distance sur E ? Quelle condition n'est pas satisfaite ?

2. À quelle condition a-t-on $d(X, Y) = 0$?

`Exercice 3` `Compenser le biais d'une pièce truquée`

Soient $(X_1, \dots, X_n)$, n variables aléatoires de Bernoulli indépendantes de même paramètre $p \in (0, 1)$. Sans connaître p, on voudrait utiliser les $(X_1, \dots, X_n)$ pour générer une suite $(Y_1, \dots, Y_K) = f(X_1, \dots, X_n)$ de variables aléatoires de Bernoulli indépendantes de paramètre $\frac{1}{2}$. La longueur $K \geq 0$ de la suite produite peut dépendre de l'entrée $(X_1, \dots, X_n)$, et la stratégie $f \colon \{0, 1\}^n \to \{0, 1\}^*$ est d'autant meilleure que $\mathbf{E}[K]$ est élevé.

1. Trouver une stratégie non triviale (i.e. $\mathbf{E}[K] > 0$) pour $n = 2$.

2. Donner une bonne stratégie pour $n = 4$.

3. Montrer que toute stratégie vérifie $\mathbf{E}[K] \leq nH(p)$, avec $H(p) = -p \log(p) - (1-p) \log(1-p)$.

`Exercice 4` `Simulation parfaite à l'aide d'une pièce de monnaie`

Soit $\mathcal{X}$ un ensemble fini et P une loi de probabilité sur $\mathcal{X}$. Le but de cet exercice est de simuler une variable aléatoire X selon la probabilité P exactement à l'aide d'une pièce de monnaie (non biaisée). On veut aussi minimiser le nombre moyen de lancers, c'est-à-dire $\mathbf{E}[T]$, où T est le nombre de lancers effectués pour une simulation.

1. Montrer que toute stratégie peut être représentée par un arbre binaire (éventuellement infini) dont les feuilles sont étiquetées par des symboles de $\mathcal{X}$. Quelle condition doit être vérifiée pour que le nombre de lancements soit presque sûrement fini et que la loi simulée soit bien P ?

2. Étant donnée une stratégie, exprimer le nombre moyen de lancers comme l'entropie d'une variable aléatoire que l'on explicitera, et en déduire que nécessairement $\mathbf{E}[T] \geq H(X)$.

3. Donner une condition nécessaire et suffisante sur P pour que cette borne puisse être atteinte.

4. Dans le cas général, proposer une stratégie qui garantisse au moins $H(X) \leq \mathbf{E}[T] \leq H(X)+2$.

5. En déduire une interprétation asymptotique de l'entropie pour la simulation de suites i.i.d.

`Exercice 5` `Algorithmes de tri`

On considère un algorithme de tri par comparaisons que l'on suppose capable de trier toute séquence de N éléments (deux à deux distincts) en effectuant au plus K comparaisons. Le but de l'exercice est d'établir, en utilisant l'entropie, que

$$K = \Omega(N \log N) \text{ lorsque } N \to \infty.$$

Puisque seul l'ordre des éléments importe, on peut considérer l'entrée comme une simple permutation inconnue (*i.e.* aléatoire) Σ de $\{1, \dots, N\}$ dont il s'agit de calculer l'inverse Σ^{-1}. Le résultat de la $k^{\text{ème}}$ comparaison ($1 \leq k \leq K$) est la variable binaire $T_k = \mathbf{1}_{\{\Sigma(I_k) \leq \Sigma(J_k)\}}$, où le choix des éléments $1 \leq I_k, J_k \leq N$ peut dépendre de $T_1, \dots, T_{k-1}$.

Justifier l'égalité $H(\Sigma) = H(T_1, \dots, T_K)$ et conclure.

`Exercice 6` `Test d'hypothèse`

Soient P et Q deux distributions de probabilité sur un espace fini $\mathcal{U}$. On souhaite distinguer P et Q à l'aide d'un échantillon de taille k, c'est-à-dire à partir du résultat de k tirages indépendants dans $\mathcal{U}$. Le test est défini par un ensemble $A \subseteq \mathcal{U}^k$: si l'échantillon $(U_1, \dots, U_k) \in A$, alors le test retourne P sinon, il retourne Q.

On désire que la probabilité d'erreur du test soit au plus ϵ si P est la vraie distribution ($P(A) \geq 1 - \epsilon$) et minimiser la probabilité d'erreur lorsque Q est la vraie distribution. On pose

$$\beta(k, \epsilon) = \min\{Q(A) \mid A \subseteq \mathcal{U} \text{ et } P(A) \geq \epsilon\}.$$

Pour ce faire, on définit l'entropie relative de deux distributions p et q, ou distance de Kullback-Leibler, par

$$D(p||q) = \mathbf{E}_p\left[\log \frac{P(U)}{Q(U)}\right] = \sum_{u \in \mathcal{U}} p(u) \log \frac{p(u)}{q(u)},$$

avec les conventions $0 \log \frac{0}{0} = 0$, $0 \log \frac{0}{q} = 0$ et $p \log \frac{p}{0} = +\infty$.

1. Montrer que $d(p||q) \geq 0$ et que $D(p||q) = 0$ si et seulement si $p = q$.

On suppose désormais que pour tout $u \in \mathcal{U}$, $p(u)q(u) > 0$, et on définit pour tout $\delta > 0$ et tout $n \in \mathbb{N}$ l'ensemble $A_\delta^{(n)}(p||q)$ par l'ensemble des suites $(u_1, \dots u_n) \in \mathcal{U}^n$ telles que

$$2^{-n(D(p||q)+\delta)} \leq \frac{q(u_1, \dots, u_n)}{p(u_1, \dots, u_n)} \leq 2^{-n(D(p||q)-\delta)}.$$

2. Montrer que pour tout $(u_1, \dots, u_n) \in A_\delta^{(n)}(p||q)$,

$$D(p||q) - \delta \leq \frac{1}{n} \log \frac{p(u_1, \dots, u_n)}{q(u_1, \dots, u_n)} \leq D(p||q) + \delta.$$

3. Montrer que pour tout $\epsilon > 0$ et tout n suffisamment grand, $P(A_\delta^{(n)}(p||q)) \geq 1 - \epsilon$.

4. En déduire pour que $\epsilon > 0$ et tout n suffisamment grand,

$$(1-\epsilon)2^{-n(D(p||q)+\delta)} \leq Q(A_\delta^{(n)}(p||q)) \leq 2^{-n(D(p||q)-\delta)}.$$

5. Conclure en montrant que pour tout $\epsilon \in (0,1)$,

$$\lim_{k\to\infty} \frac{1}{k} \log \beta(k,\epsilon) = -\sum_{u\in\mathcal{U}} p(u) \log \frac{p(u)}{q(u)}.$$