Théorie de l'Information

Devoir # 4
Résolution

SIC-SICOM

Maria-João Rendas

1. Soit $\{X_n\}_{n=1,2,\ldots}$ une série stationnaire. Est-ce $\frac{1}{n}H(X_1^n)$ monotonne (en n)?
   Résolution
   Cette démonstration a été faite en cours. Nous la répétons ici.

   L'application de la règle de la chaîne pour l'entropie conduit à :

   $$\frac{1}{n}H(X_1^n) = \frac{1}{n}\sum_{i=1}^{n} H(X_i|X_1^{i-1})$$

   Comme la série est stationnaire nous pouvons changer l'origine de l'indice temporel de chaque terme de la série. En particulier

   $$\frac{1}{n}H(X_1^n) = \frac{1}{n}\sum_{i=1}^{n} H(X_n|X_{n-i-1}^{n-1}) \; .$$

   Comme le conditionnement diminue l'entropie, $H(X_n|X_{n-i-1}^{n-1}) \geq H(X_n|X_1^{n-1})$, et donc

   $$\frac{1}{n}H(X_1^n) \geq \frac{1}{n}\sum_{i=1}^{n} H(X_n|X_1^{n-1}) = H(X_n|X_1^{n-1}) \; . \tag{1}$$

   Comme (règle de la chaîne)

   $$\frac{1}{n}H(X_1^n) = \frac{1}{n}\left[H(X_1^{n-1}) + H(X_n|X_1^{n-1})\right] \; ,$$

   par l'inégalité précédante, equation (**??**),

   $$\frac{1}{n}H(X_1^n) \leq \frac{1}{n}\left[H(X_1^{n-1}) + \frac{1}{n}H(X_1^n)\right]$$

   et donc nous pouvons écrire

   $$(n-1)H(X_1^n) \leq nH(X_1^{n-1})$$

   et finalement nous obtenons

   $$\frac{1}{n}H(X_1^n) \leq \frac{1}{n-1}H(X_1^{n-1})$$

ce qui démontre que $\frac{1}{n}H(X_1^n)$ est monotonne non-croissante en n (*il y a une erreur dans la dernière équation dans le transparent 7 du cours 4, qui est corrigée ici*).

2. Considérez que l'on observe un de deux processus, X_1 ou X_2, mais on ne sait pas lequel. Quel est le taux d'entropie des observations?
Plus précisémment, soit $X_{11},X_{12},\ldots,X_{1n},\ldots$ un processus de Bernoulli de paramètre p et $X_{21},X_{22},\ldots,X_{2n},\ldots$ un autre processus de Bernoulli de paramètre q, θ une variabe aléatoire qui peut prendre les valeurs 1 et 2 avec probabilité $\frac{1}{2}$, et soit
$$Y_n = X_{\theta n}, \qquad n = 1,2,\ldots$$
   (a) Est-ce Y_n un processus stationnaire? (Remarquez que θ reste constante pour tout n.)
   (b) Est-ce Y_n une série i.i.d.?
   (c) Quel est le taux d'entropie de Y_n, $\overline{H}(Y)$? Est-ce que
$$-\frac{1}{n}\log p(y_1,\ldots,y_n) \overset{n\to\infty}{\longrightarrow} \overline{H}(Y) \ ?$$

Considérez maintenant que θ_n est un autre processus de Bernoulli, de paramètre $\frac{1}{2}$, et Z_n un autre processus d'observation
$$Z_n = X_{\theta_n n}, \qquad n = 1,2,\ldots$$
La valeur de θ n'est plus fixée comme pour Y_n, étant choisie, d'une manière indépendante, à chaque nouvelle observation. Répétez les questions *(a)* à *(c)* pour ce modèle.

Résolution
$$Y_n = X_{\theta n}, \qquad n = 1,2,\ldots$$
   (a) Considérons la loi de probabilité conjointe de m échantillons de Y_n:
$$p(Y_{t_1},Y_{t_2},\ldots,Y_{t_m})$$
Par la loi de la probablité totale,
$$\begin{aligned}
p(\{Y_{t_i}\}_{i=1}^m) &= \Pr(\theta=1)p(\{Y_{t_i}\}_{i=1}^m|\theta=1) + \Pr(\theta=2)p(\{Y_{t_i}\}_{i=1}^m|\theta=2)\\
&= \frac{1}{2}p(\{X_{1t_i}\}_{i=1}^m) + \frac{1}{2}p(\{X_{2t_i}\}_{i=1}^m)\\
&\overset{(a)}{=} \frac{1}{2}p(\{X_{1t_i+\delta}\}_{i=1}^m) + \frac{1}{2}p(\{X_{2t_i+\delta}\}_{i=1}^m)\\
&= \Pr(\theta=1)p(\{Y_{t_i+\delta}\}_{i=1}^m|\theta=1) + \Pr(\theta=2)p(\{Y_{t_i+\delta}\}_{i=1}^m|\theta=2)\\
&= p(\{Y_{t_i+\delta}\}_{i=1}^m)
\end{aligned}$$
ce qui démontre que la série est stationnaire. *(a)*: par stationnarité des séries X_1 et X_2.

(b) Calculons la loi de probabilité de Y_n:

$$\begin{aligned} p(Y_n=\alpha) &= \mathrm{Pr}(\theta=1)p(Y_n=\alpha|\theta=1)+\mathrm{Pr}(\theta=2)p(Y_n=\alpha|\theta=2) \\ &= \mathrm{Pr}(\theta=1)p(X_{1n}=\alpha)+\mathrm{Pr}(\theta=2)p(X_{2n}=\alpha) \end{aligned}$$

Comme mes séries X_1 et X_2 sont iid, nous pouvons déjà conclure que les probablités dans le membre droit ne dépéndent pas de n et donc que les variables Y_n sont identiquement distribuées.
Admettons que $X_{1n} \in \{a,b\}$ et $X_{2n} \in \{c,d\}$, et que $p(X_{1n}=a)=p$ et $p(X_{2n}=c)=q$. Alors

$$\begin{aligned} p(Y_n=a) &= \frac{1}{2}p(X_{1n}=a)+\frac{1}{2}p(X_{2n}=a)=\frac{1}{2}p \\ p(Y_n=b) &= \frac{1}{2}(1-p) \\ p(Y_n=c) &= \frac{1}{2}q \\ p(Y_n=d) &= \frac{1}{2}(1-q) \end{aligned}$$

Ce qui confirme ce que nous venons d'affirmer.
Pour étudier l'indépendance statistique, considérons les probabilités conjointes

$$\begin{aligned} p(Y_n=\alpha,Y_m=\beta) &= \mathrm{Pr}(\theta=1)p(X_{1n}=\alpha,X_{1m}=\beta)+\mathrm{Pr}(\theta=2)p(X_{2n}=\alpha,X_{2m}=\beta) \\ &= \frac{1}{2}p(X_{1n}=\alpha)p(X_{1m}=\beta)+\frac{1}{2}p(X_{2n}=\alpha)p(X_{2m}=\beta) \end{aligned}$$

où nous avons utilisé l'indépendance statistique des séries X_1 et X_2 pour factoriser chacune des probabiliés. Nous constatons donc que

$$\begin{aligned} p(Y_n=\alpha,Y_m=\beta) &= \frac{1}{2}\left(p(X_{1n}=\alpha)p(X_{1m}=\beta)+p(X_{2n}=\alpha)p(X_{2m}=\beta)\right) \\ &\neq p(Y_n=\alpha)p(Y_m=\beta) \\ &\quad = \frac{1}{2}\left(p(X_{1n}=\alpha)p(X_{2n}=\alpha)+p(X_{1m}=\beta)p(X_{2m}=\beta)\right) \end{aligned}$$

et que donc la série Y_n n'est pas i.i.d : ses valeurs sont "liés" par la valeur de θ.

(c) Nous rapellons d'abord que

$$\begin{aligned} H(\theta,Y_1^n) &= H(\theta)+H(Y_1^n|\theta) \\ &= H(Y_1^n)+H(\theta|Y_1^n) \end{aligned}$$

et donc que nous pouvons écrire

$$H(Y_1^n)=H(\theta)+H(Y_1^n|\theta)-H(\theta|Y_1^n)$$

Le taux d'entropie de Y_n est

$$\begin{aligned}
\overline{H}(Y) &= \lim_{n\to\infty} \frac{1}{n} H(Y_1^n) \\
&= \lim_{n\to\infty} \frac{1}{n} \left(H(\theta) + H(Y_1^n|\theta) - H(\theta|Y_1^n)\right) \\
&= \lim_{n\to\infty} \frac{1}{n} H(Y_1^n|\theta) \\
&\quad + \lim_{n\to\infty} \frac{1}{n} \left(H(\theta) - H(\theta|Y_1^n)\right) \\
&\overset{(b)}{=} \lim_{n\to\infty} \frac{1}{n} \frac{1}{2} \left(H(Y_1^n|\theta = 1) + H(Y_1^n|\theta = 2)\right) \\
&= \lim_{n\to\infty} \frac{1}{n} \frac{1}{2} \left(H(X_1) + H(X_2)\right) \\
&= \lim_{n\to\infty} \frac{1}{n} \frac{1}{2} \left(H_2(p) + H_2(q)\right)
\end{aligned}$$

où en (b) nous avons utilisé la definition d'entropie conditionnelle, et le fait que les entropies $H(\theta|Y_1^n) \leq H(\theta) < \infty$, et donc que la dernière limite est nulle. Dans cette équation $H_2(p)$ désigne l'entropie d'une variable aléatoire binaire avec loi de probabilité $(p, 1-p)$.

(d) Le processus Y_n n'est pas ergodique, et nous ne pouvons donc pas garantir la convergence. En fait, la variable aléatoire $\frac{1}{n} \log p(y_1,_l dots, y_n)$ tend vers l'entropie $H(X_1)$ quand $\theta = 1$ et vers $H(X_2)$ quand $\theta = 2$, car dans chaque cas nous avons une séquence i.i.d. Donc, elle converge en distribution vers une variable aléatoire qui prend ces deux valeurs avec égale probabilité.

Il est facile de montrer que

(a) La séquence Z_n est stationnaire.

(b) La séquence Z_n est i.i.d. (la dépendance statistique a été éliminée). Les Z_n sont des variable aléatoires iid, avec une loi de Bernoulli de paramètre $p' = \frac{p+q}{2}$ si $a = c$ et $b = d$.

(c) le taux d'entropie est simplement $H_2(p')$.

(d) Dans ce cas, les conditions de la propriété d'équi-répartition asymptotique sont satisfaites, et nous pouvons donc garantir la convergence en probabilité:

$$-\frac{1}{n} \log p(z_1, \ldots, z_n) \xrightarrow{n\to\infty} \overline{H}(Z)$$

avec probabilité 1.

3. Soit X_n un processus stationnaire. Lesquelles de ces affirmations sont vraies? Démontrez ou donnez un contre-exemple.

   (a) $H(X_n|X_0) = H(X_{n-1}|X_0)$

   (b) $H(X_n|X_0) \geq H(X_{n-1}|X_0)$

   (c) $H(X_n|X_1^{n-1}, X_{n+1})$ est non-croissante en n.

Résolution

(a) Cette équation est fausse en général. Les Chaînes de Markov e sont un exemple.

(b) Cette inégalité peut aussi être fausse. Un exemple est le cas d'une série périodique, de période m, telle que $X_{n+m} = X_n$, construite à partir de m variables aléaoies iid $\{X_0,X_1,\ldots,X_{m-1}\}$ avec entropie H. Dans ce cas

$$H(X_{km}|X_0) = 0, \text{ mais } H(X_{km-1}|X_0) = H$$

ce qui contredit l'inégalité.

(c) Cette affirmation est vraie. Par stationnarité

$$H(X_n|X_1^{n-1},X_{n+1}) = H(X_{n+1}|X_2^n,X_{n+2}) \geq H(X_{n+1}|X_1^n,X_{n+2})$$

car le conditionnement diminue l'entropie.

4. Montrez que pour des processus stationnaires

$$\lim_{n\to\infty} \frac{1}{2n} I(X_1^n; X_{n+1}^{2n}) \longrightarrow 0 \ .$$

Résolution
Nottons que

$$H(X_1^{2n}) = H(X_1^n,X_{n+1}^{2n}) = H(X_{n+1}^{2n}) + H(X_1^n|X_{n+1}^{2n})$$

et donc

$$H(X_1^n|X_{n+1}^{2n}) = H(X_1^{2n}) - H(X_{n+1}^{2n}) \ .$$

Alors,

$$\begin{aligned} \frac{1}{2n} I(X_1^n; X_{n+1}^{2n}) &= \frac{1}{2n}\left[H(X_1^n) - H(X_1^n|X_{n+1}^{2n})\right] \\ &= \frac{1}{2n}\left[H(X_1^n) - H(X_1^{2n}) + H(X_{n+1}^{2n})\right] \end{aligned}$$

Par stationnarité,

$$\frac{1}{2n} I(X_1^n; X_{n+1}^{2n}) = \frac{1}{2n}\left[2H(X_1^n) - H(X_1^{2n}))\right]$$

et donc

$$\lim_{n\to\infty} \frac{1}{2n} I(X_1^n; X_{n+1}^{2n}) = \lim_{n\to\infty} \frac{1}{n} H(X_1^n) - \lim_{n\to\infty} \frac{1}{2n} H(X_1^{2n}) = \overline{H} - \overline{H} = 0$$

Cet exercice montre que le passé n'apporte pas une information importante sur le futur distant.

5. **Résolution**

(a) La matrice de transition de cette Chaîne est (admettant l'ordre suivante pour les états de la chaîne : $\{RU, FT, PB\}$

$$P = \begin{bmatrix} 0.5 & 0.1 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.3 & 0.3 & 0.7 \end{bmatrix}$$

Le vecteur propre associé à la valeur propre unitaire (la distribution stationnaire) est

$$\mu = \begin{bmatrix} 1/6 \\ 1/3 \\ 1/2 \end{bmatrix} .$$

Nous avons donc

$$P\mu = \mu .$$

Cette densité indique que la personne mange un tiers du temps à France Télécom (et la moitié du temps à la Place Bermond).

(b) L'entropie de la distribution stationnaire est (par application de la définition)

$$H(\mu) = -\sum_{i=1}^{3} \mu_i \log \mu_i = 1.56 \text{ bits.}$$

(c) Le taux d'entropie de la chaîne est (par l'expression vue en cours)

$$\overline{H} = -\sum_{i,j} \mu_i P_{ji} \log P_{ji} = 1.258 \text{ bits} < H(\mu) .$$

Le taux d'entropie ($\overline{H}$) est strictement inférieur à l'entropie de la distribution stationnaire ($H(\mu)$) car les états successifs sont correlés (et pour une Chaîne de Markov *stationnaire*, $\overline{H} = H(X_n|X_{n-1}) \leq H(X_n) = H(\mu)$).