Cours 5 : Les séquences typiques et l'AEP

7 février 2020

Plan du cours

1. Source stationnaire, entropie par lettre ;
2. source markovienne ;
3. séquences typiques, AEP, propriétés ;
4. théorème de Shannon pour les sources vérifiant l'AEP ;
5. les sources sans mémoire vérifient l'AEP ;
6. les sources markoviennes stationnaires vérifient l'AEP.

1. Source stationnaire

Une source produit une suite de lettres dans l'alphabet $\mathcal{X}$. Pour décrire cette suite de lettres nous utiliserons une suite de variables aléatoires $X_1, X_2, \ldots$ à valeurs dans $\mathcal{X}$.

Ces variables n'ont pas nécessairement la même distribution.

Ces variables ne sont pas nécessairement indépendantes (mémoire).

Elle est caractérisée par les probabilités jointes :

$$\Pr\{(X_1, \ldots, X_n) = (x_1, \ldots, x_n)\} = p(x_1, \ldots, x_n), \quad n = 1, \ldots$$

Entropie par lettre

Rappel : On définit l'entropie des L premières lettres par

$$H(X_1, \ldots, X_L) = \sum_{x_1, \ldots, x_L} -p(x_1, \ldots, x_L) \log_2 p(x_1, \ldots, x_L)$$

L' *entropie par lettre* d'une source $\mathcal{X}$ est définie par

$$H(\mathcal{X}) = \lim_{L \to \infty} \frac{1}{L} H(X_1, \ldots, X_L),$$

si cette limite existe. On l'appelle aussi son taux d'entropie.

Entropie par lettre d'un processus stationnaire

Définition Une source est dite *stationnaire* si son comportement ne varie pas lorsque l'on décale l'observation dans le temps. Pour tous entiers positifs n et j, et tout $(x_1, \ldots, x_n) \in \mathcal{X}^n$

$$p_{X_1 \ldots X_n}(x_1, \ldots, x_n) = p_{X_{1+j} \ldots X_{n+j}}(x_1, \ldots, x_n)$$

Théorème 1. *Pour toute source stationnaire, les limites ci-dessous existent et sont égales*

$$H(\mathcal{X}) = \lim_{n \to \infty} \frac{1}{n} H(X_1, \ldots, X_n) = \lim_{n \to \infty} H(X_n \mid X_1, \ldots, X_{n-1}).$$

Preuve

1. Montrons d'abord que $\lim_{L\to\infty} H(X_L \mid X_1, \ldots, X_{L-1})$ existe pour un processus stationnaire.
   Pour tout L, on a (pourquoi ?)

$$\begin{aligned} H(X_L \mid X_1, \ldots, X_{L-1}) &\leq H(X_L \mid X_2, \ldots, X_{L-1}) \\ &= H(X_{L-1} \mid X_1, \ldots, X_{L-2}) \end{aligned}$$

   Ainsi, la suite $(H(X_i \mid X_1, \ldots, X_{i-1}))_{i>0}$ est décroissante.

   Comme elle est positive, elle est convergente.

Preuve

2. Pour tout $L > 0$, nous avons (pourquoi ?)

$$H(X_1, \ldots, X_L) = \sum_{i=1}^{L} H(X_i \mid X_1, \ldots, X_{i-1}).$$

Donc

$$\frac{1}{L} H(X_1, \ldots, X_L)$$

est la moyenne des $H(X_i \mid X_1, \ldots, X_{i-1})$ qui convergent.

Le théorème de Césaro permet de conclure :
si une suite $u_n \to \ell$, alors $1/n \sum_{i=1}^{n} u_i \to \ell$.

Processus dont l'entropie par lettre est indéfinie

$X_1, \ldots . X_n$ indépendants avec $p_i = p(X_i) = 1$ tel que

$$p_i = \begin{cases} 1/2 & 2^{2^{2k}} < i \leq 2^{2^{2k+1}} \\ 0 & 2^{2^{2k+1}} < i \leq 2^{2^{2k+2}} \end{cases}$$

Sur des segments exponentiellement longs $H(X_i) = 1$, suivant de segment exponentiellement encore plus longs tels que $H(X_i) = 0$.

On a $\frac{1}{n}H(X_1,\ldots,X_n) = \frac{1}{n}\sum_{i=1}^{n} H(X_i)$.

Posons $u_{2k+1} \overset{\text{def}}{=} \sum_{i\leq 2^{2^{2k+1}}} H(X_i)$ et $u_{2k} \overset{\text{def}}{=} \sum_{i\leq 2^{2^{2k}}} H(X_i)$ on a

$$\begin{aligned} u_{2k} - u_{2k-1} &= 0 \\ u_{2k+1} - u_{2k} &= 2^{2^{2k}}\left(2^{2^{2k}} - 1\right) \end{aligned}$$

et

$$\begin{aligned} \frac{2^{2^{2k}}\left(2^{2^{2k}} - 1\right)}{2^{2^{2k+1}}} \leq \quad & \frac{u_{2k+1}}{2^{2^{2k+1}}} \quad \leq 1 \\ 0 \leq \quad & \frac{u_{2k}}{2^{2^{2k}}} = \frac{u_{2k-1}}{2^{2^{2k}}} \quad \leq 2^{-2^{2k-1}} \end{aligned}$$

Donc cette moyenne oscille entre 0 et 1, sans limite.

2. Source markovienne invariante dans le temps

Définition Une source est dite *markovienne d'ordre 1* si pour tout entier positif n et tout $(x_1, \ldots, x_n) \in \mathcal{X}^n$

$$\Pr[X_n = x_n \mid X_1 = x_1, \ldots, X_{n-1} = x_{n-1}] = \Pr[X_n = x_n \mid X_{n-1} = x_{n-1}]$$

Elle est d'ordre s si

$$\Pr[X_n = x_n \mid X_1 = x_1, \ldots, X_{n-1} = x_{n-1}] = \Pr[X_n = x_n \mid X_{n-1} = x_{n-1} \ldots X_{n-s} = x_{n-s}]$$

Définition La source est dite *invariante dans le temps* si ces probabilités ne dépendent pas de n.

Nous noterons $p(x_2 \mid x_1) = \Pr[X_n = x_2 \mid X_{n-1} = x_1]$.

Théorème

Théorème L'entropie par lettre d'une source markovienne invariante dans le temps est égale à

$$H(\mathcal{X}) = H(X_2|X_1) = \sum_{x_1,x_2} -\lambda(x_1)p(x_2 \mid x_1)\log_2 p(x_2 \mid x_1)$$

où $\lambda(x), x \in \mathcal{X}$ est la *distribution stationnaire*.

Distribution stationnaire

Soit $\Pi = p(x_2|x_1)_{x_2,x_1}$. Le vecteur des probabilités de la distribution $V_n = (p(X_n = a_1), \ldots, p(X_n = a_k))$ vérifie

$$V_n = \Pi V_{n-1}$$

Si le processus admet une distribution stationnaire $\Lambda = (\lambda(x_1), \ldots, \lambda(x_k))$, il doit vérifier

$$\Lambda = \Pi \Lambda$$

On écrit

$$
\begin{aligned}
H(X) &= \lim_{L\to\infty} (H(X_L|X_{L-1}\dots X_1) \\
&= H(X_2|X_1) \\
&= \sum_{x_2,x_1} p(x_2|x_1)\lambda(x_1)\log_2 p(x_2|x_1)
\end{aligned}
$$

L'entropie de l'anglais et les modèles markoviens

1) approximation d'ordre 0 (toutes les lettres sont indépendantes)

$$H_0 = \log 27 \approx 4.76.$$

```
XFOML RXKHRJFFJUJ ZLPWCFWKCYJ
FFJEYVKCQSGXYD QPAAMKBZAACIBZLHJQD
```

2) approximation d'ordre 1 (les lettres sont choisies suivant la fréquence des lettres dans un texte en anglais)

$$H_1 \approx 4.03$$

```
OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI
ALHENHTTPA OOBTTVA NAH BRL
```

3) Approximation d'ordre 2 : même fréquence des couples de lettres que dans un texte en anglais

```
ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMMY
ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO
TIZIN ANDY TOBE SEACE CTISBE
```

4) Approximation d'ordre 3 : même fréquence des triplets de lettres que dans un texte en anglais

```
IN NO IST LAT WHEY CRATICT FROURE BERS GROCID
PONDENOME OF DEMONSTURES OF THE REPTAGIN IS
REGOACTIONA OF CRE
```

5) Modèle markovien d'ordre 3 de l'anglais

$$H_{1/3} \approx 2.8$$

```
THE GENERATED JOB PRIVIDUAL BETTER TRAND THE
DIPLAYED CODE, ABOVERY UPONDULTS WELL THE
CODERST IN THESTICAL IT DO HOCK BOTH MERG.
(INSTATES CONS ERATION. NEVER ANY OF PUBLE AND TO
THEORY. EVENTIAL CALLEGAND TO ELAST BENERATED IN
 WITH PIES AS WITH THE)
```

6) Approximation d'ordre 1 sur les mots

```
REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME
CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO
OF TO EXPERT GRAY COME TO FURNISHES THE LINE
MESSAGE HAD BE THESE
```

7) Modèle markovien d'ordre 1 sur les mots

```
THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH
WRITER THAT THE CHARACTER OF THIS POINT IS
THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE
TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED
```

$$H_{\text{anglais}} \approx 1.34$$

Utilisation du modèle markovien pour le codage arithmétique

Modèle sans mémoire

$$\Pr(x_1, x_2, \ldots, x_n) = \Pr(x_1)\Pr(x_2)\ldots\Pr(x_n)$$

Modèle markovien d'ordre 1

$$\Pr(x_1, x_2, \ldots, x_n) = \Pr(x_1)\Pr(x_2|x_1)\ldots\Pr(x_n|x_{n-1})$$

Modèle markovien d'ordre 2

$$\Pr(x_1, x_2, \ldots, x_n) = \Pr(x_1)\Pr(x_2|x_1)\Pr(x_3|x_1, x_2)\ldots\Pr(x_n|x_{n-2}, x_{n-1})$$

3. Séquences typiques

Soit une source constituée de la suite de variables aléatoires $X_1, X_2, \ldots, X_n, \ldots$ à valeur dans un alphabet $\mathcal{X}$. On suppose que l'entropie par lettre de cette source est définie

$$\mathcal{H} = H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n)$$

Définition Ensemble de séquences ε-typiques (de longueur n)

$$A_\varepsilon^{(n)} = \left\{ (x_1, \ldots, x_n) \in \mathcal{X}^n, \left| \frac{1}{n} \log_2 \frac{1}{p(x_1, \ldots, x_n)} - \mathcal{H} \right| \leq \varepsilon \right\}$$

Séquences typiques et probables

Les séquences typiques ne sont pas les plus probables !

Soit une source binaire produisant des 0 avec une probabilité $2/3$ et des 1 avec une probabilité $1/3$,

- la séquence ne comportant que des 1 est peu probable et n'est pas typique,
- la séquence ne comportant que des 0, bien qu'étant la plus probable, n'est pas typique.
- Les séquences typiques seront celles qui comportent environ $1/3$ de 1 (voir plus loin).

Asymptotic Equipartition Property

Définition [Asymptotic Equipartition Property]
Une source vérifie l'AEP si :

$$\forall \varepsilon > 0, \lim_{n \to \infty} \Pr\left[A_\varepsilon^{(n)}\right] = 1.$$

Propriétés des séquences typiques

Proposition 1. *Pour toute source vérifiant l'AEP*

1. $\frac{1}{n}\log_2 \frac{1}{p(x_1,\ldots,x_n)} \underset{n\to\infty}{\longrightarrow} \mathcal{H}$ *presque sûrement*
2. $|A_\varepsilon^{(n)}| \leq 2^{n(\mathcal{H}+\varepsilon)}$
3. *pour* n *suffisamment grand,* $|A_\varepsilon^{(n)}| \geq (1-\epsilon)2^{n(\mathcal{H}-\varepsilon)}$.

Autrement dit :

Il y a $\approx 2^{n\mathcal{H}}$ séqu. typiques, elles sont équiprobables de prob. $\approx 2^{-n\mathcal{H}}$.

Preuve

1. Par définition, on a

$$\left|\frac{1}{n}\log_2\frac{1}{p(x_1,\ldots,x_n)}-\mathcal{H}\right|\leq\varepsilon$$

avec une probabilité $\Pr[A_\varepsilon^{(n)}]$. Par conséquent

$$\mathcal{H}\quad-\quad\varepsilon\qquad\leq\qquad\frac{1}{n}\log_2\frac{1}{p(x_1,\ldots,x_n)}\qquad\leq\qquad\mathcal{H}\quad+\quad\varepsilon$$

avec une probabilité $\Pr[A_\varepsilon^{(n)}]$. Si la source vérifie l'AEP, $\Pr[A_\varepsilon^{(n)}]\to 1$.

Preuve

2. et 3. Soient $\varepsilon > 0$ et n un entier. Nous avons

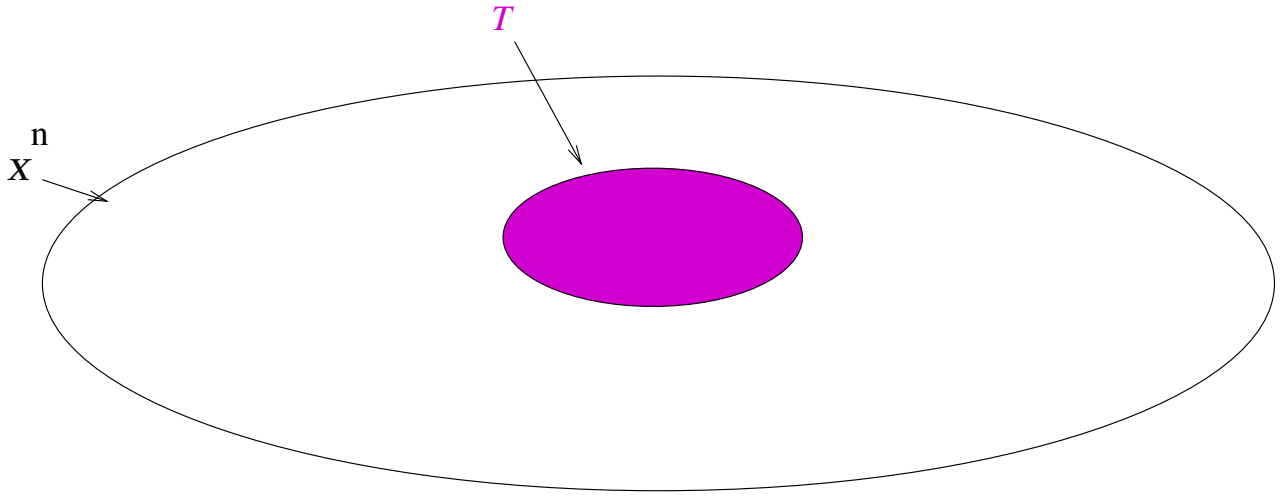
$$\begin{aligned} |A_\varepsilon^{(n)}| 2^{-n(\mathcal{H}+\varepsilon)} &\leq \sum_{(x_1,\ldots,x_n)\in A_\varepsilon^{(n)}} \Pr[X_1 = x_1, \ldots, X_n = x_n] \\ &= \Pr[A_\varepsilon^{(n)}] \leq 1, \end{aligned}$$

et donc $|A_\varepsilon^{(n)}| \leq 2^{n(\mathcal{H}+\varepsilon)}$. Et aussi

$$\Pr[A_\varepsilon^{(n)}] = \sum_{(x_1,\ldots,x_n)\in A_\varepsilon^{(n)}} \Pr[X_1 = x_1, \ldots, X_n = x_n] \leq |A_\varepsilon^{(n)}| 2^{-n(\mathcal{H}-\varepsilon)},$$

ce qui donne $|A_\varepsilon^{(n)}| \geq \Pr[A_\varepsilon^{(n)}] 2^{n(\mathcal{H}-\varepsilon)}$

Patates



Il y a un nombre exponentiellement faible de séquences ($2^{n(H+\epsilon)} \ll |\mathcal{X}|^n$) qui concentrent toute la masse de la distribution de probabilité.

Ensembles à haute probabilité

Théorème 2. *Soit une source vérifiant l'AEP. Soit* $B_\delta^{(n)}$ *t.q.*

$$\Pr\left\{B_\delta^{(n)}\right\} \geq 1-\delta$$

alors

1. *quel que soit* $\delta' > 0$,

$$\frac{1}{n}\log|B_\delta^{(n)}| > H - \delta', \text{ pour } n \text{ assez grand.}$$

2. $$\Pr\left\{A_\varepsilon^{(n)} \cap B_\delta^{(n)}\right\} \geq 1-\varepsilon-\delta \text{ pour } n \text{ assez grand..}$$

Preuve

Fait : *si* $\Pr(A) \geq 1-\varepsilon_1$ *et* $\Pr(B) \geq 1-\varepsilon_2$ *alors* $\Pr(A \cap B) > 1-\epsilon_1-\epsilon_2$

$$\begin{aligned} 1-\varepsilon-\delta &\leq \Pr\left\{A_\varepsilon^{(n)} \cap B_\delta^{(n)}\right\} \\ &= \sum_{x^n \in A_\varepsilon^{(n)} \cap B_\delta^{(n)}} p(x^n) \\ &\leq \sum_{x^n \in A_\varepsilon^{(n)} \cap B_\delta^{(n)}} 2^{-n(H-\varepsilon)} \\ &\leq |B_\delta^{(n)}| 2^{-n(H-\varepsilon)} \end{aligned}$$

On conclut en manipulant les $\log_2$.

4. Théorème de Shannon

Définition *Soit φ un codage de $\mathcal{X}$, sa longueur moyenne par lettre est définie par*

$$\mathcal{L}(\varphi) = \lim_{n\to\infty} \frac{1}{n} \sum_{x_1,\ldots,x_n} p(x_1,\ldots,x_n)|\varphi(x_1,\ldots,x_n)|,$$

lorsque cette limite existe.

***Théorème* 3. [Shannon]** *Soit une source discrète d'alphabet $\mathcal{X}$ et d'entropie par lettre $\mathcal{H}$ qui vérifie l'AEP.*

1. *Tout codage non ambigu φ de $\mathcal{X}$ vérifie $\mathcal{L}(\varphi) \geq \mathcal{H}$.*
2. *Il existe un codage non ambigu φ de $\mathcal{X}$ tel que $\mathcal{L}(\varphi) \leq \mathcal{H} + \varepsilon$, $\forall \varepsilon > 0$.*

Preuve

1. *Exercice : utiliser le théorème 2.*

Preuve

2. *Soit $\varepsilon > 0$, et $n > 0$ Pour tout entier n, soient*
   1. *F_n un code de longueur fixe minimale de $\mathcal{X}^n$*
   2. *$G_{n,\varepsilon}$ un code de longueur fixe minimale de $A_\varepsilon^{(n)}$.*
   3. *$\varphi_{n,\varepsilon}$ un code de $\mathcal{X}^n$ défini par*

$$\varphi_{n,\varepsilon}(x_1,\ldots,x_n) = \begin{cases} 0 \parallel G_{n,\varepsilon}(x_1,\ldots,x_n) & \text{si } (x_1,\ldots,x_n) \in A_\varepsilon^{(n)} \\ 1 \parallel F_n(x_1,\ldots,x_n) & \text{sinon} \end{cases}$$

Preuve (Suite)

En prenant n assez grand tel que $\Pr[A_\varepsilon^{(n)}] \geq 1 - \varepsilon$.

$$
\begin{aligned}
|\varphi_{n,\varepsilon}| &= \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n)|\varphi(x^n)| + \sum_{x^n \notin A_\varepsilon^{(n)}} p(x^n)|\varphi(x^n)| \\
&\leq \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n)(n(H+\varepsilon)+2) + \sum_{x^n \notin A_\varepsilon^{(n)}} p(x^n)(n\log|\mathcal{X}|+2) \\
&= \Pr[A_\varepsilon^{(n)}](n(H+\varepsilon)+2) + \Pr[\overline{A_\varepsilon^{(n)}}](n\log|\mathcal{X}|+2) \\
&\leq n(H+\varepsilon) + \varepsilon n(\log|\mathcal{X}|) + 2
\end{aligned}
$$

On conclut en faisant $n \to \infty$.

5. AEP des sources sans mémoire

Une source sans mémoire vérifie l'AEP.

On rappelle l'entropie :

$$\frac{1}{L}H(X_1,\ldots,X_L) = \frac{1}{L}(LH(X_1)) = H(X_1) = -\sum p(x_i)\log_2 p(x_i).$$

Source sans mémoire

Rappel : loi faible des grands nombres

Soient $Z_1, Z_2, \ldots, Z_n, \ldots$ une suite de variable aléatoire indépendantes de même distribution, d'espérance μ. Soit

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^{n} Z_i,$$

la moyenne des n premiers tirages.

Alors, pour tout ε :

$$\Pr\left\{|\bar{Z}_n - \mu| > \varepsilon\right\} \to 0$$

quand $n \to \infty$ (loi faible des grands nombres).

Un théorème

Soient $X_1, \ldots, X_n, \ldots$ *indépendantes de même distribution* $p(X)$ *alors*

$$-\frac{1}{n}\log p(X_1, \ldots, X_n) \quad = \quad -\frac{1}{n}\sum_{i=1}^{n}\log p(X_i)$$

Or (loi faible de grands nombres)

$$\Pr\left\{\left|-\frac{1}{n}\sum_{i=1}^{n}\log p(X_i) - \mathbb{E}\left[-\log p(X)\right]\right| \leq \varepsilon\right\} \to 1$$

Réécriture

$$\Pr\left\{\left|-\frac{1}{n}\sum_{i=1}^{n}\log p(X_i) - \mathbb{E}\left[-\log p(X)\right]\right| \leq \varepsilon\right\} \to 1$$

Or $\mathbb{E}\left[-\log p(X)\right] = H$. *Donc*

$$\Pr\left\{\left|-\frac{1}{n}\log p(X_n, \ldots, X_1) - H\right| \leq \varepsilon\right\} \to 1$$

ou encore, pour tout ϵ *:*

$$\Pr(A_n^{(\varepsilon)}) \to 1$$

quand $n \to \infty$.

Séquences typiques pour une source discrète sans mémoire

Soit $x = (x_1, \ldots, x_n)$ une séquence de longueur n, pour une source d'alphabet $\mathcal{X} = \{a_1, \ldots, a_k\}$ Soit $n_{a_i}(x)$ le nombre de fois où a_i apparaît dans x, et $n_{a_i}(x)/n$ la fréquence d'apparition.

La suite x est ε-typique de longueur n si et ssi

$$\left| \sum_{i=1}^{k} \left(\frac{n_{a_i}(x)}{n} - p(a_i) \right) \log p(a_i) \right| < \varepsilon$$

Les deux définitions sont équivalentes, dans le cas d'une source discrète sans mémoire.

6. Source Markovienne

Proposition 2. *Une source markovienne invariante dans le temps vérifie l'AEP.*

Preuve

Soit $\mathcal{X} = X_1, \ldots, X_l, \ldots$. *On a* $H(\mathcal{X}) = H(X_2|X_1)$. *On cherche à montrer que*

$$\Pr\left\{\left|-\frac{1}{n}\log p(X_1, X_2, \ldots, X_n) - H(X)\right| < \varepsilon\right\} \to 1$$

D'une part

$$\begin{aligned} -\frac{1}{n}\log_2 p(x_n, \ldots, x_1) &= -\frac{1}{n}\log_2 p(x_n|x_{n-1}) \ldots p(x_2|x_1)p(x_1) \\ &= -\frac{1}{n}\sum_{i=1}^{n}\log_2 p(x_i|x_{i-1}) - \frac{\log p(x_1)/p(x_1|x_0)}{n} \end{aligned}$$

Preuve (suite)

On a donc à considérer $-\frac{1}{n}\sum_{i=1}^{n}\log_2 p(x_i|x_{i-1})$. *Soient les variables aléatoires* $Z_i = -\log_2 p(X_i|X_{i-1})$. *Alors elles sont iid, et* $\bar{Z}_n$ *leur moyenne vérifie la loi des grands nombres :*

$$\Pr\left\{|\bar{Z}_n - \mu| < \varepsilon\right\} \to 1 \text{ où}$$

$$\mu = E(Z_i) = \sum_{x_i, x_{i-1}} -p(x_i, x_{i-1}) \log_2 p(x_i|x_{i-1}) = H(X_2|X_1) = H(\mathcal{X})$$

$$\text{Or : } \bar{Z}_n = -\frac{1}{n}\sum \log_2 p(x_1, \ldots, x_n) - \frac{\log p(x_1)/p(x_0|x_1)}{n}$$

$$\text{On a donc bien } \Pr\left\{\left|-\frac{1}{n}\log p(X_1, X_2, \ldots, X_n) - H(X)\right| < \varepsilon\right\} \to 1.$$