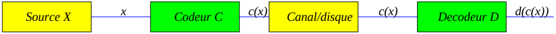
Théorème de Shannon (codage source)

Un codeur (binaire) est une application

$$C: \begin{array}{ccc} \mathcal{X} & \rightarrow & \{0,1\}^\star \\ x & \rightarrow & c(x) \end{array},$$

Le décodeur D est une application des séquences binaires dans l'alphabet $\mathcal{X}$:

$$D: \begin{array}{ccc} \{0,1\}^\star & \rightarrow & \mathcal{X} \\ c(x) & \rightarrow & d(c(x)) \end{array}$$



Codage sans pertes

Si $\forall x \in \mathcal{X}$

$$d(c(x)) = x \ ,$$

(l'application C est inversible sur $\mathcal{X}$)
nous dirons que le codage est sans pertes.
Dans le cas contraire, nous dirons que C est un codeur avec pertes.

$\Rightarrow$ Un code sans pertes doit vérifier :

$$|C(\mathcal{X})| = |\mathcal{X}|.$$

Codes de longueur fixe et variable

Longueur d'une séquence :

$$x = x_1 \cdots x_n \Rightarrow |x| = n$$

C est un code de longueur fixe si

$$\forall c \in C(\mathcal{X}), |c| = n$$

Sinon, C est de longueur variable.
C binaire, sans pertes, de longueur fixe $\Rightarrow$

$$|c| \geq \log_2 |\mathcal{X}|.$$

Si nous acceptons des pertes (erreurs)

$$d(c(x_1)) = d(c(x_2)), \qquad x_1 \neq x_2 \in \mathcal{X}$$

nous pouvons utiliser $|c| < \log_2 |\mathcal{X}|$.
Pour garantir $\Pr\{erreurs\} \ll 1 \Rightarrow$
caractérisation probabiliste de la source.

Plus petit ensemble δ-représentatif S_δ

X : variable aléatoire $X \in \mathcal{X}$, et loi p_X : $X \sim p_X$.
S_δ est le plus petit sous-ensemble de $\mathcal{X}$ avec probabilité plus grande ou égale à $1 - \delta$:

$$S_\delta = \arg \min_{S \subset \mathcal{X}, \Pr(S) \geq 1-\delta} |S|.$$

Contenu δ-informatif de X $H_\delta(X)$

$X \sim p_X$, $X \in \mathcal{X}$, S_δ le plus petit sous-ensemble δ-informatif pour X.
Le contenu δ-informatif de X est

$$H_\delta(X) = \log |S_\delta|.$$

$H_\delta(X)$ indique le *nombre minimal de bits* d'un *code C de longueur fixe* qui peut transmetre *sans erreur toutes les séquences de l'ensemble* S_δ.
C a donc $\Pr\{erreur\} < \delta$.

$H_0(X)$ est égal à la valeur maximale de $H(X)$, $X \in \mathcal{X}$:

$$H_0(X) = \log |\mathcal{X}| \geq H(X),$$

Exemple

Séquence binaire de longueur 4 :
$x_1 x_2 x_3 x_4$, $x_i \in \{0, 1\}$, $\Pr\{x_i = 0\} = 0.2$.

$x_k \in \mathcal{X}$	$p(x_k)$	$H_\delta^k(X)$
0000	$(0.2)^4$	$\log_2(15) = 3.9$
0001	$(0.2)^3 \cdot 0.8$	$\log_2(14) = 3.9$
0010	$(0.2)^3 \cdot 0.8$	$\log_2(13) = 3.8$
0100	$(0.2)^3 \cdot 0.8$	$\log_2(12) = 3.7$
1000	$(0.2)^3 \cdot 0.8$	$\log_2(11) = 3.6$
0011	$(0.2)^2 \cdot (0.8)^2$	$\log_2(10) = 3.4$
0101	$(0.2)^2 \cdot (0.8)^2$	$\log_2(9) = 3.3$
1001	$(0.2)^2 \cdot (0.8)^2$	$\log_2(8) = 3.2$
0110	$(0.2)^2 \cdot (0.8)^2$	$\log_2(7) = 3.0$
1010	$(0.2)^2 \cdot (0.8)^2$	$\log_2(6) = 2.8$
1100	$(0.2)^2 \cdot (0.8)^2$	$\log_2(5) = 2.6$
1110	$0.2 \cdot (0.8)^3$	$\log_2(4) = 2.0$
1101	$0.2 \cdot (0.8)^3$	$\log_2(3) = 1.6$
1011	$0.2 \cdot (0.8)^3$	$\log_2(2) = 1.0$
0111	$0.2 \cdot (0.8)^3$	$\log_2(1) = 0$
1111	$(0.8)^4$	

Plus petit ensemble δ-représentatif

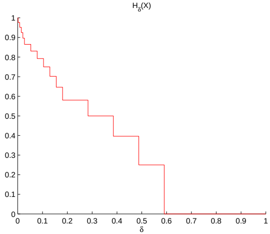
$$S^0 = \mathcal{X} \qquad S_\delta^{k-1} = S_\delta^k \bigcup \{x_k\}, k = 1, 2, \ldots, 16,$$

Taille :

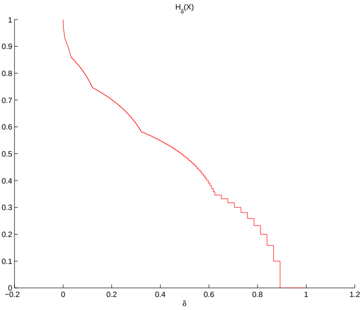
$$|S_\delta^k| = |S_\delta^{k-1}| - 1, k = 1, 2, \ldots, 16 \qquad |S_\delta^0| = 16.$$

Probabilité d'erreur : $\delta_k = \Pr\left\{x_i \notin S_\delta^k\right\}$

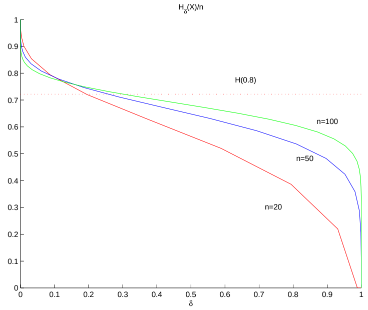
$$\delta_0 = 1, \qquad \delta_{k+1} = \delta_k - p(x_{k+1}), k = 1, 2, \ldots, 16.$$



graphe de $H_\delta(X)/n$ $(n = 4)$



graphe de $H_\delta(X)/n$ $(n = 10)$



graphe de $H_\delta(X)/n$ $(n = 20, 50, 100)$

Théorème du codage source (Shannon)

Soit X une source avec entropie $H(X)$. Alors

$$\forall \epsilon > 0, \quad \forall \delta \in]0,1[, \quad \exists n_0 : \\ \forall n > n_0 \qquad \left| \frac{1}{n} H_\delta(X^{(n)}) - H \right| < \epsilon.$$

$X^{(n)}$: séquences de longueur n dont les éléments sont des tirages statistiquement indépendants de la même variable aléatoire X.

Propriété d'équi-répartition asymptotique

"Tous les événements qui peuvent se produire sont essenciellement équiprobables"

$x^{(n)}, n \gg 1, \mathcal{X} = \{1, \ldots, m\}, |\mathcal{X}| = m$

$$n_i(x) = |\{x_i = i\}|$$

Loi des grands nombres $\Rightarrow$ $n_i \simeq np(i)$.

$\Rightarrow$

$$p(x^{(n)}) = \prod_{k=1}^{n} p(x_k) = \prod_{i=1}^{m} p(i)^{n_i(x)} \simeq \prod_{i=1}^{m} p(i)^{np(i)}$$

$$\log \frac{1}{p\left(x^{(n)}\right)} \simeq n \sum_{i=1}^{m} p(i) \log \frac{1}{p(i)} \simeq nH(X)$$

Ensemble typique $A_\epsilon^{(n)}$

$X_1, X_2, \ldots, X_n$: variables aléatoires indépendantes et identiquement distribuées (i.i.d.), $X_i \sim p(x)$, $x \in \mathcal{X}$.
L'ensemble ϵ-typique par rapport à p est le sous-ensemble de $\mathcal{X}^n$:

$$A_\epsilon^{(n)} = \left\{ x^{(n)} \in \mathcal{X}^n \ : \ p(x^{(n)}) \in \left[2^{-n(H(X)+\epsilon)}, 2^{-n(H(X)-\epsilon)} \right] \right\}$$

$\Leftrightarrow$

$$A_\epsilon^{(n)} = \left\{ x^{(n)} \in \mathcal{X}^n \ : \ \frac{1}{n} \log \frac{1}{p(x^{(n)})} \in [H(X)+\epsilon, H(X)-\epsilon] \right\}$$

Les séquences dans $A_\epsilon^{(n)}$ ont toutes “ la *même probabilité*”.

Propriété d'équi-répartition assymptotique

Pour $n \gg 1$, $x^{(n)}$ (symboles iid d'une source X) appartient *presque surement* à un sous-ensemble de $\mathcal{X}$ qui *contient seulement* $2^{nH(X)}$ éléments, chacun avec une *probabilité proche* de $2^{-nH(X)}$.

Équivalente au Théorème de Shannon:

Codage source (version informelle)
n variables $X_i \sim p, i = 1, \ldots, n$ iid, avec entropie $H(X)$, peuvent être codées avec un nombre de bits non inférieur à $nH(X)$ avec une probabilité d'erreur négligeable; si un nombre de bits inférieur à $nH(X)$ est utilisé, la probabilité d'erreur sera près de 1.

Loi (faible) des grands nombres

X_i, $i = 1, \ldots, n$, i.i.d., moyenne μ et variance σ^2. Soit

$$X = \frac{1}{n} \sum_{i=1} X_i.$$

Alors,

$$\Pr\left\{(X-\mu)^2 \geq \alpha\right\} \leq \frac{\sigma^2}{n\alpha}$$

$\Leftrightarrow$

$$\forall \alpha' > 0,\ \ \forall \delta > 0,\ \exists n_0 : n > n_0 \qquad \Pr\left\{|X-\mu| \geq \alpha'\right\} \leq \delta$$

$$(\alpha' = \sqrt{\alpha}, \text{ et } n_0 = \lceil \tfrac{\sigma^2}{\alpha\delta} \rceil)$$

Inégalité de Chebychev

$X \geq 0$, $\alpha > 0$. Alors

$$\Pr\{X \geq \alpha\} \leq \frac{\mathrm{E}[X]}{\alpha}.$$

Démonstration :

$$\Pr\{X \geq \alpha\} = \sum_{x \geq \alpha} p(x) \overset{(a)}{\Leftrightarrow} \Pr\{X \geq \alpha\} \leq \sum_{x \geq \alpha} \frac{x}{\alpha} p(x)$$

$$\overset{(b)}{\Leftrightarrow} \Pr\{X \geq \alpha\} \leq \sum_{x \in \mathcal{X}} \frac{x}{\alpha} p(x) = \frac{\mathrm{E}[X]}{\alpha}$$

(a): car $\frac{x}{\alpha} \geq 1$ (b): car les termes ajoutés sont positifs.

Inégalité de Chebychev $\Rightarrow$

Inégalité de Chebychev (moment d'ordre 2)

X variable aléatoire, $\alpha > 0$. Alors

$$\Pr\left\{(X - \mathrm{E}[X])^2 \geq \alpha\right\} \leq \frac{\sigma^2}{\alpha}.$$

(prendre $X = (X - E[X])^2$)

Principe d'équi-répartion asymptotique

x_i i.i.d. $\Rightarrow p(x^{(n)}) = \prod_{i=1,\ldots,n} p(x_i)$, $A_\epsilon^{(n)}$ est défini par

$$\frac{1}{n}\sum_{i=1}^{n} \log \frac{1}{p(x_i)} \in [H(X)-\epsilon, H(X)+\epsilon] \,.$$

Soient

$$Z_i = \log \frac{1}{p(X_i)}, i = 1, \ldots, n, \text{(i.i.d.)}, \quad \mathrm{E}\,[Z_i] = H(X), \quad \mathrm{var}(Z_i) = \sigma_Z^2$$

$A_\epsilon^{(n)} \Leftrightarrow$

$$\left| \frac{1}{n}\sum_{i=1}^{n} Z_i - H(X) \right| \le \epsilon \Leftrightarrow \left(\frac{1}{n}\sum_{i=1}^{n} Z_i - H(X) \right)^2 \le \epsilon^2 \,.$$

Borne supérieure de probabilité

$$\Pr\left\{\left(\frac{1}{n}\sum_{i=1}^{n} Z_i - H(X)\right)^2 \geq \epsilon^2\right\} \leq \frac{\sigma_Z^2}{n\epsilon^2} \to_{n\to\infty} 0$$

$$\Pr\left\{A_\epsilon^{(n)}\right\} = \Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n} Z_i - H(X)\right| < \epsilon\right\} \geq 1 - \frac{\sigma_Z^2}{n\epsilon^2} \to_{n\to\infty} 1\ .$$

$\Leftrightarrow$

$$\Pr\left\{A_\epsilon^{(n)}\right\} \geq 1 - \delta(n,\epsilon), \qquad \delta(n,\epsilon) = \frac{\sigma_Z^2}{n\epsilon^2}\ ,$$

Démonstration Théorème de Shannon du codage source

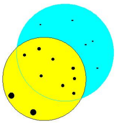
$$|A_\epsilon^{(n)}| \leftrightarrow H_\delta(X^{(n)}) = \log|S_\delta|$$

$$\forall \epsilon > 0, \forall \delta \in]0,1[, \exists n_0 \text{ tel que } \forall n > n_0$$
$$H_\delta(X^{(n)}) - nH(X) \in [-n\epsilon, n\epsilon]$$

Deux étapes

1. $\forall \epsilon > 0$, $\forall \delta \in [0,1]$, $\exists n_0$ tel que
$$\frac{1}{n} H_\delta(X^{(n)}) - H(X) < \epsilon, \qquad \forall n > n_0 \; .$$
2. $\Pr\{S_\delta\} = 1 - \delta, \forall n > n_0 \Rightarrow$
$$H_\delta(X^{(n)}) > n\,(H(X) - \epsilon) \Rightarrow \frac{1}{n} H_\delta(X^{(n)}) - H(X) > -\epsilon,$$

$\frac{1}{n}H_\delta(X^{(n)}) - H(X) < \epsilon$



S_δ

- $\Pr\{S_\delta\} = 1 - \delta$

$A_\epsilon^{(n)}$

- $\Pr\left\{A_\epsilon^{(n)}\right\} \geq 1 - \delta(n, \epsilon)$
- $|A_\epsilon| \leq 2^{n(H+\epsilon)}$

$\left|A_\epsilon^{(n)}\right|$ borne supérieure de $|S_\delta|$:

$$H_\delta(X^{(n)}) = \log|S_\delta| \leq \log|A_\epsilon^{(n)}|.$$

Nous allons montrer que $\exists B_s$:

$$|A_\epsilon^{(n)}| \leq B_s \Rightarrow H_\delta(X^{(n)}) \leq \log B_s.$$

$|A_\epsilon^{(n)}| < 2^{n(H+\epsilon)}$

$$
\begin{aligned}
1 \geq \Pr\left\{A_\epsilon^{(n)}\right\} &= \sum_{x^{(n)} \in A_\epsilon^{(n)}} p(x^{(n)}) \\
&\stackrel{(a)}{>} \sum_{x^{(n)} \in A_\epsilon^{(n)}} 2^{-n(H+\epsilon)} \\
&= 2^{-n(H+\epsilon)} \sum_{x^{(n)} \in A_\epsilon^{(n)}} 1 = 2^{-n(H+\epsilon)} |A_\epsilon^{(n)}| \\
&\Leftrightarrow |A_\epsilon^{(n)}| < 2^{n(H+\epsilon)}
\end{aligned}
$$

(a) la borne inférieure pour $x^{(n)} \in A_\epsilon^{(n)}$

Fixons n_0 tel que

$$\delta \geq \delta(n_0, \epsilon) = \frac{\sigma_Z^2}{\epsilon^2 n_0} \Leftrightarrow n_0 \geq \frac{\sigma_Z^2}{\epsilon^2 \delta}$$

Alors, $\forall n > n_0$

$$\Pr\left\{A_\epsilon^{(n)}\right\} \geq 1 - \delta(n, \epsilon) \geq 1 - \delta(n_0, \epsilon) \geq 1 - \delta$$

$$\Pr\left\{A_\epsilon^{(n)}\right\} \geq 1 - \delta, \qquad \text{et} \qquad \log |A_\epsilon^{(n)}| \leq n\,(H(X) + \epsilon)\ .$$

$\Rightarrow$

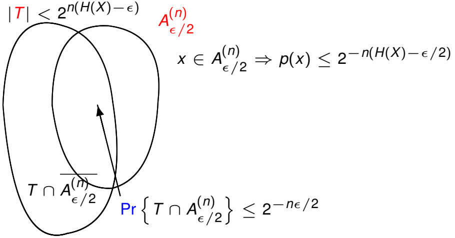
$$H_\delta(X^{(n)}) < n(H + \epsilon)$$

$\frac{1}{n}H_\delta(X^{(n)}) - H(X) > -\epsilon$

Par *contradiction*.
Admettons $\exists T$, tel que $\forall n > n_0$

$$|T| < 2^{n(H(X)-\epsilon)}, \qquad \Pr\{T\} \geq 1-\delta$$

$\Rightarrow$ impossible de trouver n_0.



$$\Pr\{T\} = \Pr\left\{T \cap A_{\epsilon/2}^{(n)}\right\} + \Pr\left\{T \cap \overline{A_{\epsilon/2}^{(n)}}\right\}$$

Mais

$$\begin{aligned}
\Pr\left\{T \bigcap A_{\epsilon/2}^{(n)}\right\} &= \sum_{x^{(n)} \in T, x^{(n)} \in A_{\epsilon/2}^{(n)}} p(x) \\
&\leq \max_{x \in A_{\epsilon/2}^{(n)}} p(x) \left|T \bigcap A_{2\epsilon}^{(n)}\right| \\
&\leq 2^{-n(H-\epsilon/2)} |T| \leq 2^{-n(H-\epsilon/2)} 2^{n(H-\epsilon)} = 2^{-n\epsilon/2}
\end{aligned}$$

$$\Pr\left\{T \cap \overline{A_{\epsilon/2}^{(n)}}\right\} \leq \Pr\left\{\overline{A_{\epsilon/2}^{(n)}}\right\} \leq \frac{4\sigma_Z^2}{n\epsilon^2}$$

$\Rightarrow$

$$|T| \leq 2^{n(H(X)-\epsilon)} \Rightarrow \Pr\{T\} \leq 2^{-n\epsilon/2} + \frac{4\sigma_Z^2}{n\epsilon} \rightarrow_{n\rightarrow\infty} 0$$

$\Rightarrow$

$|S_\delta| \geq 2^{n(H(X)-\epsilon)}$, ou encore

$$\frac{1}{n} H_\delta(X^{(n)}) - H(X) > -\epsilon$$

$$\frac{1}{n} H_\delta(X^{(n)}) - H(X) > -\epsilon$$

et

$$H_\delta(X^{(n)}) < n(H + \epsilon)$$

$\Rightarrow \forall \epsilon > 0, \forall \delta \in]0, 1[, \exists n_0$ tel que $\forall n > n_o$

$$\left| \frac{1}{n} H_\delta(X^{(n)}) - (H + \epsilon) \right| < \epsilon$$