Théorie de l'Information
Notes de Cours (part 2)
2006-2007

SIC-SICOM

Maria-João Rendas

October 14, 2006

Chapter 2

Compréssion de données

2.1 Introduction

L'objectif de ce Chapitre est d'établir les limites fondamentaux de la compression de données, c'est à dire de la détermination de codes le plus efficaces possibles.

Nous començons par formuler mathématiquement l'opération de codage (et éventuelle compréssion). Considérons une source, X, qui émet des séquences x de symboles $\{x_i\}$ dans un alphabet $\mathcal{X}$, telle que nous représentons dans la Figure 2.1. Soit $c(x)$ le résultat de l'opération du *codeur* C (que nous admettons pour l'instant binaire) sur le message $x \in \mathcal{X}$. La séquence (binaire) $c(x)$ peut maintenant être enregistrée pour une ultérieure récupération/lecture, où servir à transmettre le message x à travers un canal de communication. Associé au codeur C, il doit exister un *décodeur*, D, qui reconstruit, à partir de la séquence binaire $c(x)$, le message initial $x \in \mathcal{X}$. Le codeur est donc une application

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

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

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

Nous désignerons l'ensemble $C(\mathcal{X})$ des mots (binaires) qui peuvent être engendrés par le code C, par *code*.

Codage sans pertes

Nous pouvons distinguer les méthodes de codage par le domaine de validité $\mathcal{Y} \subset \mathcal{X}$ de l'équation

$$d(c(x)) = x \,, \tag{2.1}$$

qui impose que le message décodé soit effectivement égal au message émis par la source. Si l'équation (2.1) se vérifie $\forall x \in \mathcal{X}$, ce qui implique que l'application C est

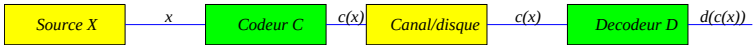


Figure 2.1: Codage/décodage.

inversible sur $\mathcal{X}$, nous dirons que le codages est *sans pertes*. Dans le cas contraire, nous dirons que C est un codeur *avec pertes*.

Remarque 1 *Nous pouvons déjà conclure qu'un code sans pertes doit vérifier la condition suivante:*

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

Codes de longueur fixe/variable

Une autre distinction importante concerne la longeur des séquences codées,[1]

$$n(x) = |c(x)|.$$

Si tous les éléments du code $C(\mathcal{X})$ ont la même longueur, nous dirons que C est un code *de longeur fixe*. Dans le cas contraire, nous parlerons d'un code *de longueur variable*. La longueur n des mots d'un code (binaire) sans pertes de longueur fixe doit nécessairement satisfaire

$$n \geq \log_2 |\mathcal{X}|. \tag{2.2}$$

Cependant, si nous acceptons que des pertes (c'est à dire, que des séquences distinctes $c_1 \neq c_2$ soient d écodées par le même message $d(c_1) = d(c_2) \in \mathcal{X}$), nous pouvons utiliser des mots de longueur inférieure à la borne de l'équation (2.2). Si la probabilité des messages pour lesquelles ces erreurs se produisent est très petite, la performance globale du code peut être acceptable. Pour pouvoir contrler cette probabilité, il faut utiliser une *caractérisation probabiliste* de la source.

Définition 1 *Plus petit ensemble δ-représentatif S_δ*

Soit X une variable aléatoire avec valeurs dans l'ensemble $\mathcal{X}$, 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 $1 - \delta$:

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

△

[1] Pour des séquences $x = x_1 \cdots x_n$, $n(x)$ désigne le nombre d'éléments de la séquence (sa *longueur*).

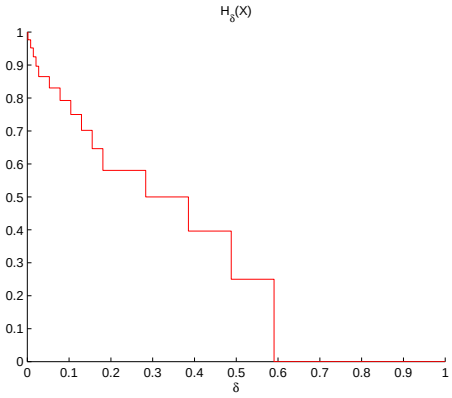


Figure 2.2: graphe de $H_\delta(X)/n$ ($n = 4$), $\delta \in [0, 1]$.

Définition 2 *Contenu δ-informatif de X* $H_\delta(X)$
Soit $X \sim p_X$, $X \in \mathcal{X}$ une variable aléatoire, et S_δ le plus petit sous-ensemble δ-informatif pour X. Le contenu δ-informatif de X est

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

△

La valeur de $H_\delta(X)$ nous indique le *nombre minimal de bits* d'un *code de longeur fixe* qui peut transmetre *sans erreur toutes les séquences de l'ensemble* S_δ, et qui a donc une probabilité d'erreur inférieure à δ.

Remarque 2 Si $\delta = 0$, $H_0(X)$ coincide avec la valeur maximale de l'entropie des variables aléatoires définies sur $\mathcal{X}$:

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

où $H(X)$ est l'entropie de Shannon de X introduite dans le Chapitre précédant.

Exemple 1 Nous illustrons les deux définitions précédantes, S_δ et $H_\delta(X)$.
Considérez une séquence de quatre variables binaires statistiquement indépendantes, qui prennent une des deux valeurs possibles avec probabilité $p = 0.8$. La Figure 2.2 montre la variation de $H_\delta(X)$ avec la valeur de δ sur l'intervalle unitaire. Pour cette exemple, le tableau suivant liste toutes les éléments de $\mathcal{X}$ par ordre croissante de probabilité ($Pr(1) = 0.8$)

$x_k \in \mathcal{X}$	$p(x_k)$	$H_\delta^k(X)$
0000	$(0.2)^4$	$\log_2(15) = 3.9069$
0001	$(0.2)^3 \cdot 0.8$	$\log_2(14) = 3.9069$
0010	$(0.2)^3 \cdot 0.8$	$\log_2(13) = 3.8074$
0100	$(0.2)^3 \cdot 0.8$	$\log_2(12) = 3.7004$
1000	$(0.2)^3 \cdot 0.8$	$\log_2(11) = 3.5850$
0011	$(0.2)^2 \cdot (0.8)^2$	$\log_2(10) = 3.4594$
0101	$(0.2)^2 \cdot (0.8)^2$	$\log_2(9) = 3.3219$
1001	$(0.2)^2 \cdot (0.8)^2$	$\log_2(8) = 3.1699$
0110	$(0.2)^2 \cdot (0.8)^2$	$\log_2(7) = 3.0000$
1010	$(0.2)^2 \cdot (0.8)^2$	$\log_2(6) = 2.8074$
1100	$(0.2)^2 \cdot (0.8)^2$	$\log_2(5) = 2.5850$
1110	$0.2 \cdot (0.8)^3$	$\log_2(4) = 2.0000$
1101	$0.2 \cdot (0.8)^3$	$\log_2(3) = 1.5850$
1011	$0.2 \cdot (0.8)^3$	$\log_2(2) = 1.0000$
0111	$0.2 \cdot (0.8)^3$	$\log_2(1) = 0$
1111	$(0.8)^4$	

La colonne à gauche liste par ordre croissante de probabilité les possibles séquences $x_k \in \mathcal{X}, k = 1, \ldots, 16$. La colonne centrale indique leur probabilité $p(x_k)$, fonction uniquement du nombre de zéros et 1's dans la séquence x_k. Les ensembles S_δ^k sont construits itérativement en enlevant les séquences par cet ordre (les moins probables avant les plus probables), avec l'initialisation $S^0 = \mathcal{X}$:

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

et donc leur taille est donnée par

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

La colonne à droite du tableau liste le logarithme de cette taille, i.e., les valeurs de $H_\delta^k(X)$. Finalement, la probabilité δ_k pour qu'une séquence ne soit pas dans S_δ^k est obtenue récursivement de la façon suivante:

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

La Figure 2.3 illustre la variation de $H_\delta(X)/n$ pour $n = 10$. Comparez avec la Figure précédante.

Ces exemples nous montrent que si nous admettons une *probabilité d'erreur* $\delta > 0$, nous pourrons transmettre les messages d'une source source avec *moins de bits que* $H_0(X)$. Les valeurs de H_δ obtenus dans les exemples précédants dépendent fortement de la longeur de la séquence binaire ($n = 4, 10$ dans les exemples). La Figure 2.4 montre que pour des valeurs de n grands (les trois courbes représentées correspondent aux valeurs de $n = 20, 50, 100$), le nombre de bits par symbole de la source,$H_\delta(X)/n$ tend vers une valeur constante, égale à $H(p)$, sauf dans les limites de l'interval unitaire : $\delta = 0$ (codage sans pertes) et $\delta = 1$ très grande probabilité d'erreur). C'est cela qui affirme le Théorème du codage source de Shannon, que nous énnonçons maintenant:

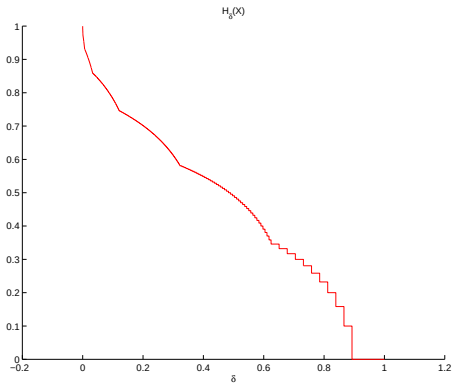


Figure 2.3: graphe de $H_\delta(X)/n$ ($n = 10$), $\delta \in [0, 1]$.

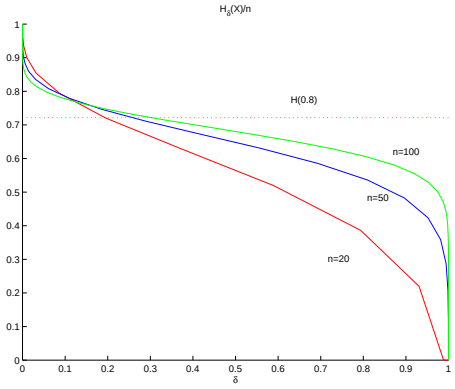


Figure 2.4: graphe de $H_\delta(X)/n$ ($n = 20, 50, 100$), $\delta \in [0, 1]$.

Théorème 1 *Théorème du codage source (Shannon)*
Soit X une source avec entropie $H(X)$. Alors

$$\forall \epsilon > 0, \forall \delta \in]0,1[, \exists n_0 : \forall n > n_0 \qquad \left| \frac{1}{n} H_\delta(X^{(n)}) - H \right| < \epsilon. \tag{2.3}$$

Dans cette équation, $X^{(n)}$ désigne l'ensemble de toutes les séquences de taille n dont les éléments sont des tirages statistiquement indépendants de la même variable aléatoire X. △

2.2 Propriété d'équi-répartition asymptotique

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

Par la loi des grands nombres (voir Théorème 3), une séquence *longue*, $x^{(n)}$ (de longueur n), de symboles statistiquement indépendants émis par une source avec des valeurs dans un alphabet discret $\mathcal{X} = \{1, \ldots, m\}$ de taille m, i.e. $|\mathcal{X}| = m$, contient un nombre $n_i(x)$ de occurrences de chaque symbole i dans $x^{(n)}$, $n_i(x) \simeq np(i)$. Sa probabilité est donc

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

L'information contenue dans la séquence est donc

$$\log \frac{1}{p\left(x^{(n)}\right)} \simeq n \sum_{i=1}^{m} p(i) \log \frac{1}{p(i)} \simeq nH(X), \tag{2.4}$$

où nous avons reconnu la définition de l'entropie, $H(X)$. Ceci explique le comportement de $\frac{1}{n} H_\delta(X)$ observé dans la Figure 2.4 à la fin de la section précédante.

Pour démontrer le Théorème de Shannon, nous allons faire appel à la notion d'ensemble typique, qui formalise la notion de *séquences typiques* sous-jacente à la dérivation de l'équation (2.4). La "typicité" de la séquence est liée au nombre n_i d'occurrences de chaque symbole. La définition formelle caractérise une séquence comme typique si l'information qu'elle contient diffère de $nH(X)$ de moins d'une quantité ϵ.

Définition 3 *Ensemble typique* $A_\epsilon^{(n)}$
Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes et identiquement distribuées (i.i.d.), avec loi de probabilité $p(x)$, $x \in \mathcal{X}$. L'ensemble ϵ-typique par rapport à p est le sous- ensemble de $\mathcal{X}^n$:

$$A_\epsilon^{(n)} = \left\{ x^{(n)} \in \mathcal{X}^n \ : \ p(x^{(n)}) \in \left[2^{-n(H(X)+\epsilon)}, 2^{-n(H(X)-\epsilon)} \right] \right\} \tag{2.5}$$

ou, d'une façon équivalente,

$$A_\epsilon^{(n)} = \left\{ x^{(n)} \in \mathcal{X}^n \ : \ \frac{1}{n} \log \frac{1}{p(x^{(n)})} \in [H(X) + \epsilon, H(X) - \epsilon] \right\} \tag{2.6}$$

△

Remarque 3 Par sa propre définition, eq. (2.5), les éléments de l'ensemble typique ont tous essenciellement la *même probabilité*. C'est ce fait que justifie le nom de la Propriété 1 que nous allons maintenant énoncer.

Nous verrons que quand n est très grand, cet ensemble typique contient presque toute la probabilité.

Propriété 1 *Propriété d'équi-répartition assymptotique*
Pour n suffisament large, une séquence $x^{(n)}$ de symboles statistiquement indépendants émis par une source X appartient presque surement à un sous-ensemble de $\mathcal{X}$ qui contient seulement $2^{nH(X)}$ él éments, chacun avec une probabilité proche de $2^{-nH(X)}$.
△

Cette Propriété est équivalente au Théorème de Shannon:

Théorème 2 *Codage source (version informelle)*
n variables $X_i \sim p, i = 1, \ldots, n$ statistiquement indépendantes et identiquement distribuées avec entropie $H(X)$ peuvent être codées avec un nombre de bits supérieur 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. △

La démonstration du théorème du codage de source est basée dans la Loi (faible) des grands nombres, que nous rappellons maintenant.

Théorème 3 *Loi (faible) des grands nombres*
Soient $X_i, i = 1, \ldots, n$, des variables aléatoires i.i.d., avec moyenne μ et variance σ^2. Soit

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

Alors,

$$\Pr\left\{ (X - \mu)^2 \geq \alpha \right\} \leq \frac{\sigma^2}{n\alpha}. \tag{2.7}$$

Nous pouvons énnoncer cette loi d'une autre façon équivalente comme

$$\forall \alpha' > 0, \ \ \forall \delta > 0, \ \exists n_0 : n > n_0 \qquad \Pr\{|X - \mu| \geq \alpha'\} \leq \delta. \tag{2.8}$$

Il suffit de prendre $\alpha' = \sqrt{\alpha}$, et $n_0 = \lceil \frac{\sigma^2}{\alpha\delta} \rceil^2$. △

[2] La notation $\lceil x \rceil$ désigne l'entier le plus petit plus grand où égal à x.

La Loi (faible) des grands nombres est une conséquence de l'inégalité de Chebyshev.

Théorème 4 *Inégalité de Chebychev*
Soit X une variable aléatoire non-négative et $\alpha > 0$. Alors

$$\Pr\{X \geq \alpha\} \leq \frac{\mathrm{E}[X]}{\alpha}. \tag{2.9}$$

△

La démonstration de cette inégalité est simple:

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

où les implications sont justifiées de la façon suivante:

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

De cette inégalité nous pouvons déduire imédiatement une inégalité concernant le moment centré d'ordre 2:

Théorème 5 *Inégalité de Chebychev (moment d'ordre 2)*
Soit X une variable aléatoire et $\alpha > 0$. Alors

$$\Pr\left\{(X - \mathrm{E}[X])^2 \geq \alpha\right\} \leq \frac{\sigma^2}{\alpha}. \tag{2.10}$$

△

Il sufit de prendre $X = (X - E[X])^2$ en (2.9).

La Loi faible des grands nombres, Propriété 3, découle si nous utilison le fait que la variance de la moyenne de n variables statistiquement indépendantes est égale à la variance individuelle de chaque variable divisée par n.

Nous revenons maintenant au principe d'équi-répartion asymptotique, Propriété 1. Comme les variables sont i.i.d., $p(x^{(n)}) = \prod_{i=1,\ldots,n} p(x_i)$, et une condition équivalente à (2.6) est encore

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

Considérons maintenant les variables aléatoires (i.i.d.)

$$Z_i = \log \frac{1}{p(X_i)}, i = 1, \ldots, n.$$

Ces variables ont une moyenne

$$\mathrm{E}\left[Z_i\right] = H(X),$$

et une variance que nous désignons par σ_Z^2. La condition (2.6) qui défini l'ensemble typique peut donc être écrite en fonction des variables Z_i:

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

Par la loi faible des grands nombres,nous sommes capables de calculer une borne inférieure pour la probabilité de cet événement :

$$\Pr\left\{\left(\frac{1}{n}\sum_{i=1}^{n} Z_i - H(X)\right)^2 \geq \epsilon^2\right\} \leq \frac{\sigma_Z^2}{n\epsilon^2} \rightarrow_{n\rightarrow\infty} 0, \tag{2.11}$$

ce qui démonstre la Propriété d'équi-répartition asymptotique: la probabilité de l'ensemble de séquences qui ont probabilité dans l'intervalle de l'équation (2.6) (l'ensemble typique $A_\epsilon^{(n)}$) est aussi près de 1 que l'on souhaite, il suffit pour cela de prendre n suffisament grand :

$$\Pr\left\{A_\epsilon^{(n)}\right\} = \Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n} Z_i - H(X)\right| < \epsilon\right\} \geq 1 - \frac{\sigma_Z^2}{n\epsilon^2} \rightarrow_{n\rightarrow\infty} 1 \,. \tag{2.12}$$

Nous pouvons re-écrire cette équation comme :

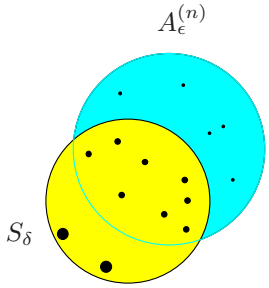
$$\Pr\left\{A_\epsilon^{(n)}\right\} \geq 1 - \delta(n,\epsilon), \qquad \delta(n,\epsilon) = \frac{\sigma_Z^2}{n\epsilon^2} \,, \tag{2.13}$$

où la définition de $\delta(n,\epsilon)$ est évidente.

Pour démontrer le Théorème de Shannon du codage source, il faut établir la relation entre la taille de l'ensemble ensemble typique et $H_\delta(X^{(n)}) = \log|S_\delta|$. Nous allons démontrer que pour toute valeur de la probabilité d'erreur δ, et pour toute valeur de ϵ, il existe un n_0 suffisament grand tel que $\forall n > n_0$, $H_\delta(X^{(n)}) - nH(X) \in [-n\epsilon, n\epsilon]$.

La démonstration est faite en deux étapes. Dans la première, nous allons démontrer que $\forall\epsilon > 0$, $\forall\delta \in [0,1]$, $\exists n_0$ tel que

$$\frac{1}{n}H_\delta(X^{(n)}) - H(X) < \epsilon, \qquad \forall n > n_0 \,. \tag{2.14}$$



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

Figure 2.5: Premier pas dans la démonstration.

Le deuxième pas établi que $\Pr\{S_\delta\} = 1-\delta, \forall n > n_0$ implique que sa taille H_δ doit satisfaire

$$H_\delta(X^{(n)}) > n\left(H(X)-\epsilon\right) \Rightarrow \frac{1}{n}H_\delta(X^{(n)}) - H(X) > -\epsilon, \tag{2.15}$$

et donc l'équation (2.3) est vraie.

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

L'ensemble typique $A_\epsilon^{(n)}$ n'est pas le meilleur ensemble pour compression, au sens que nous avons discuté dans la section précédante. Par exemple, il est simple de constater que dans les exemples de la section précédante la séquence la plus probable n'appartient pas à $A_\epsilon^{(n)}$. La taille de $A_\epsilon^{(n)}$ fourni donc une borne supérieure pour la taille de S_δ :

$$H_\delta(X^{(n)}) = \log|S_\delta| \leq \log|A_\epsilon^{(n)}|. \tag{2.16}$$

Nous allons montrer que S_δ doit être *petit*, en calculant une borne supérieure B_s pour la taille de $A_\epsilon^{(n)}$:

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

La borne supérieure est obtenue de la façon suivante:

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

$$= \quad 2^{-n(H+\epsilon)} \sum_{x^{(n)} \in A_\epsilon^{(n)}} 1 = 2^{-n(H+\epsilon)} |A_\epsilon^{(n)}|$$

$$\Leftrightarrow \quad |A_\epsilon^{(n)}| < 2^{n(H+\epsilon)} \tag{2.17}$$

où nous avons utilisé en (a) la borne inférieure pour la probabilité des séquences qui appartiennent à l'ensemble typique qui découle de sa définition (2.5). Si nous fixons n_0 tel que

$$\delta \geq \delta(n_0, \epsilon) = \frac{\sigma_Z^2}{\epsilon^2 n_0} \Leftrightarrow n_0 \geq \frac{\sigma_Z^2}{\epsilon^2 \delta}, \tag{2.18}$$

par l'équation (2.11), $\forall n > n_0$ l'ensemble typique a une probabilité supérieure à $1-\delta$:

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

$A_\epsilon^{(n)}$ satisfait donc la condition du Théorème de Shannon : pour toute valeur de $\epsilon > 0$ et de $\delta \in [0, 1]$, nous pouvons déterminer un n_0 (eq. (2.18)) tel que pour des séquences de longueur n plus grande que n_0 nous vérifions simultanément

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

Si nous utilisons la borne supérieure (2.17) dans l'équation (2.16), nous obtenons l'inégalité recherchée :

$$H_\delta(X^{(n)}) < n(H + \epsilon) \ . \tag{2.19}$$

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

La démonstration de la deuxième partie est faite par *contradiction*. Nous allons supposer qu'il existe un ensemble T, de taille plus petite que $2^{n(H-\epsilon)}$:

$$|T| < 2^{n(H(X)-\epsilon)}, \tag{2.20}$$

et qui a une probabilité supérieure à $1 - \delta$:

$$\Pr\{T\} \geq 1 - \delta \ , \tag{2.21}$$

pour tout n supérieur à un certain n_0. Nous allons voir qu'il est impossible de trouver n_0 de façon que (2.20) et (2.21) soient simultanément satisfaites pour tout $n > n_0$.

Soit $A_{\epsilon/2}^{(n)}$ l'ensemble$(\epsilon/2)$-typique. Nous pouvons décomposer la probabilité de l'ensemble T de la façon suivante [3] :

$$\Pr\{T\} = \Pr\left\{T \bigcap A_{\epsilon/2}^{(n)}\right\} + \Pr\left\{T \bigcap \overline{A_{\epsilon/2}^{(n)}}\right\} \tag{2.22}$$

Mais

$$\begin{aligned} \Pr\left\{T \bigcap A_{\epsilon/2}^{(n)}\right\} &= \sum_{x^{(n)} \in T, x^{(n)} \in A_{\epsilon/2}^{(n)}} p(x) \\ &\leq \max_{x \in A_{\epsilon/2}^{(n)}} p(x) \left|T \bigcap A_{\epsilon/2}^{(n)}\right| \\ &\leq 2^{-n(H-\epsilon/2)} |T| \leq 2^{-n(H-\epsilon/2)} 2^{n(H-\epsilon)} = 2^{-n\epsilon/2}. \end{aligned} \tag{2.23}$$

[3] Nous représentons par $\overline{A}$ le complément de l'ensemble A: $\overline{A} = \{x \notin A\}$.

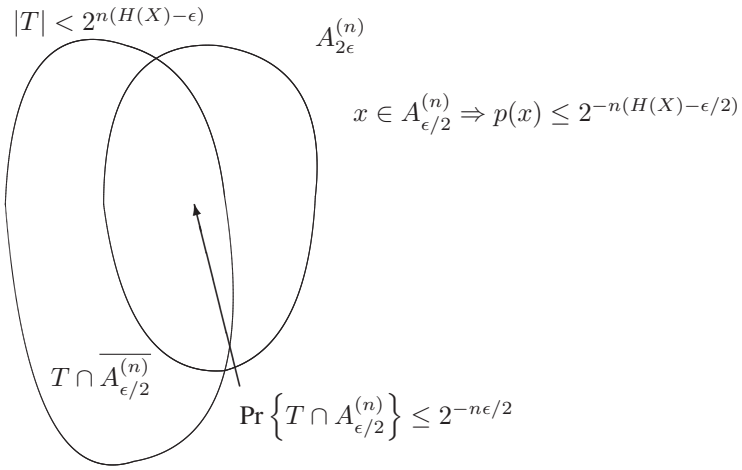


Figure 2.6: Illustration du calcul de la borne pour la probabilité de T.

Pour le deuxième terme de (2.22)

$$\Pr\left\{T \bigcap \overline{A_{\epsilon/2}^{(n)}}\right\} \leq \Pr\left\{\overline{A_{\epsilon/2}^{(n)}}\right\} \leq \frac{4\sigma_Z^2}{n\epsilon^2}\,, \tag{2.24}$$

où nous avons utilisé la borne (2.12) pour la probabilité de l'ensemble typique. Si nous utilisons les bornes (2.23) et (2.24) dans l'équation (2.22), nous obtenons

$$|T| \leq 2^{n(H(X)-\epsilon)} \Rightarrow \Pr\{T\} \leq 2^{-n\epsilon/2} + \frac{4\sigma_Z^2}{n\epsilon}, \tag{2.25}$$

ce qui montre que pour tout ensemble T de taille inférieure à $2^{n(H-\epsilon)}$ nous ne pouvons pas trouver un n_0 tel que pour tout $n > n_0$ la probabilité de T soit supérieur à $1-\delta$. (La Figure 2.7 illustre la variation avec n du membre droit de l'inégalité (2.25).)

La taille de S_δ doit donc satisfaire $|S_\delta| \geq 2^{n(H(X)-\epsilon)}$, ou encore

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

comme nous voulions démontrer.

Si n est suffisament large, le graphe de la fonction $H_\delta(X^{(n)})$ est donc bien compris dans une région horizontale autour de la valeur de l'entropie, comme nous avions affirmé, et vérifié numériquement dans la Figure 2.4 de la page 35.

Remarque 4 *La première partie du théorème nous dit qu'il suffit d'une petite tolérance $\delta \simeq 0$ aux erreurs pour que le nombre de bits par symbole ne doive pas excéder*

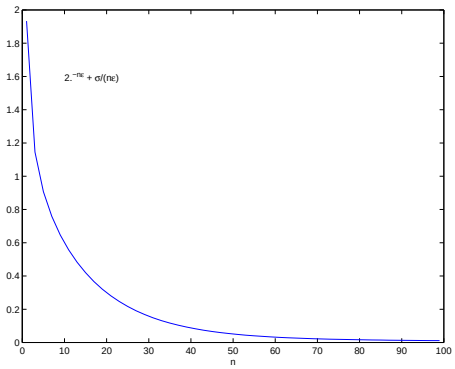


Figure 2.7: Borne supérieure dans léquation (2.25).

$H + \epsilon$. La deuxième partie montre que même si nous admettons une grande probabilité d'erreur $\delta \simeq 1$, le nombre de bits par symbole devra encore être à ϵ de l'entropie de la source. Ceci démontre le sens de l'entropie comme le **nombre moyen de bits (par symbole) nécessaires pour coder les symboles d'une source**.