Illustrations de la notion d'entropie dans les deux théorèmes de Claude Shannon en théorie de l'information

Joël Le Roux, leroux@essi.fr, April 2002

13 février 2004

Table des matières

1 Résumé

Le développement des théorèmes de Shannon est illustré dans le cas d'un bruit modifiant un message binaire transmis par un canal binaire symétrique. Ce développement ne débute pas par les propriétés de la notion d'entropie, qui bien sûr apparaîtront au cours des calculs, et évite d'utiliser l'entropie conjointe. L'outil principal dans ce développement est l'approximation de la loi binomiale fondée sur la formule de Stirling.

2 Introduction

Les démonstrations des théorèmes de Shannon me paraissent abstraites et difficiles à comprendre pour un bon nombre d'étudiants dans le domaine des transmissions numériques qui aimeraient bien avoir une idée intuitive de ces théroèmes sans pour autant chercher à devenir des experts en théorie de l'information. Il est peut être utile de présenter une illustration de ces démonstrations en évitant l'utilisation de notions difficiles et peu intuitives comme l'information mutuelle, l'entropie jointe ou l'entropie conditionnelle.

L'objectif de ce cours est d'essayer d'illustrer dans le cas le plus simple (messages binaires et canal binaire sans mémoire) le concept d'entropie et plus spécialement son utilisation dans les deux théorèmes de C. Shannon en codage de source (section 3) et en codage de canal (section 4). L'outil principal utilisé dans ce développemnt est l'approximation de densités de probabilités de loi binomiales par la formule de Stirling, comme l'a fait L. Boltzmann dans son interprétation statistique de l'entropie. Cette approche a aussi utilisée par D. MacKay dans son excellente présentation.

Il est peut-être utile de rappeler le rôle historique fondamental de l'entropie dans le développement de la science (sans citer tous les domaines où cette notion est un outil important). Rudolf Clausius (1865)a inventé la notion d'entropie dans le domaine de la thermodynamique. Il a dérivé le mot du grec "$\eta\tau\rho\omega\pi\eta$" qui signifie "changement". Ludwig Boltzmann (1877) a donné une interprétation de ce concept en termes de probabilités.

Max Plank (1901) a utilisé cette interprétation statistique pour modéliser la radiation du corps noir, ce qui l'a conduit à la découverte de la mécanique quantique, découverte qui fut ensuite enrichie par Albert Einstein (1905) qui lui aussi fonda son développement sur la travail de Boltzmann.

Claude Shannon (1948) a lui aussi trouvé son inspiration dans le travail de L. Boltzmann dans la création de la théorie de l'information, et dans l'établissement des théorèmes fondamentaux sur les bornes inférieures portant sur la compression de messages (codage de source), et la borne suppérieure donnant le nombre maximum d'erreurs qu'on peut accepter dans la transmission d'un message de telle sorte que le message original puisse être reconstitué intégralement. La borne inférieure du codage de source est atteinte dans le codage arithmérique de J. Rissanen et G. Langdon (1978) ; et les performances des turbocodes (C. Berrou *et al.*, 1993) sont proches de la borne supérieure de codage de canal.

3 Illustration du premier théorème de Shannon dans un cas simple

On considère l'émission d'un message $B(\ell)$ de longueur L composé de données binaires aléatoires indépendantes : des 'uns' avec la probabilité p inférieures à $1/2$ et des 'zéros' avec la probabilité $(1-p)$ (fig. 1).

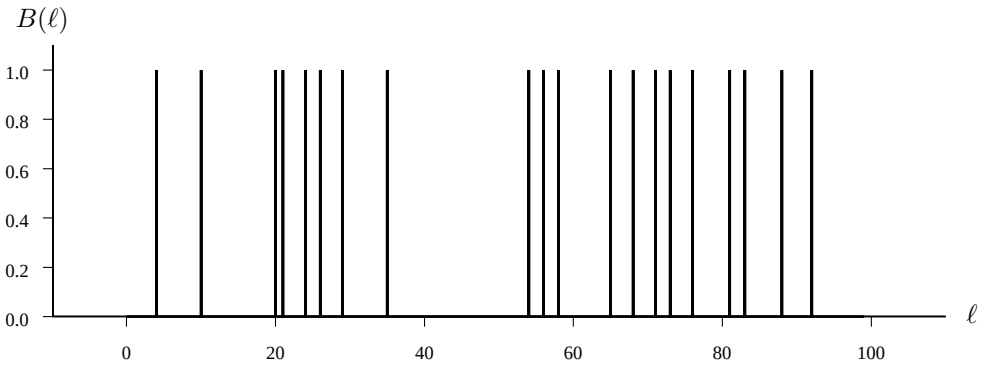

Figure 1: Exemple de message composé de '1' avec la probabilité $p = 0.2$ et de '0' avec la probabilité $(1-p) = 0.8$.

D'après la loi des grands nombres, les messages émis ont la propriété suivante : dans un message de longueur L, il y a à peu près pL '1's et $(1-p)L$ '0's (fig. 2).

L'idée sur laquelle est fondé le premier théorème est qu'il suffit de coder les messages comportant pL '1's et $(1-p)L$ '0's parce que les autres messages n'apparaissent pratiquement jamais.

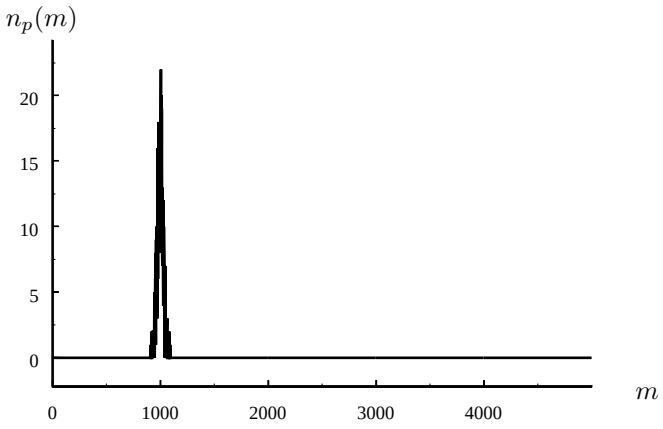

Figure 2: Histogramme du nombre de "1"s dans un message de longueur $L = 5000$ quand la probabilité p d'émission d'un "1" est 0.2; ce nombres est presque toujours entre 900 et 1100 ; d'après la loi des grands nombres, quand la longueur L est grande l'histogramme de ce nombre m divisé par L tend vers une distribution de Dirac en $m/L = p$.

3.1 Interpretation du premier théorème

Le premier théorème affirme qu'il est possible de coder la séquence $B(\ell)$ avec seulement

$$L\left((1-p)\log_2 \tfrac{1}{1-p} + p\log_2 \tfrac{1}{p}\right) \text{ bits}$$

au lieu des L bits nécessaires au codage des 2^L mots différents. La quantité

$$H_B(p) = (1-p)\log_2 \frac{1}{1-p} + p\log_2 \frac{1}{p}, \tag{1}$$

est l'entropie de la séquence $B(\ell)$. C. Shannon suit dans son développement l'idée de L. Boltzmann qui suivait une démarche similaire afin de compter le nombre de molécules en mouvement dans un volume de gaz donné. Nous allons illustrer les étapes principales du développement :

1. Nous donnerons la distribution des messages probables et calculerons une approximation (section 3.2);

2. Nous estimerons leur nombre et nous comparerons ce nombre au nombre de tous les messages possibles (section 3.3) ;

3. Nou déduirons de ce nombre les bases des méthodes de compression (section 3.4).

3.2 Approximation par la formule de Stirling du nombre de messages possibles

D'après la loi binomiale, le nombre de messages composé de m uns et de $L-m$ zéros est donnée par (fig 3)

$$n_p(m) = \frac{L!}{m!(L-m)!}(1-p)^{L-m}p^m, \tag{2}$$

lorsque la probabilité d'occurence d'un 'un' est p.

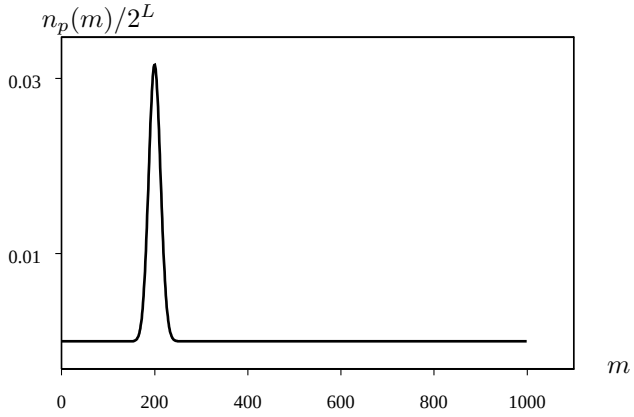

Figure 3: Densité de probabilité du nombre de "1"s dans des messages de longueur $L = 1000$ lorsque la probabilité p d'émettre un "1" est 0.2; m est presque toujours entre 180 et 220; D'après la loi des grands nombres, lorsque le nombre de données , L est grand , cette densité tend vers une loi gaussienne de moyenne $p \times L$ et de variance $Lp(1-p)$, la densité de m/L tend vers une distribution de Dirac en $\frac{m}{L} = p$.

3.2.1 Approximation fondée sur la formule de Stirling

D'après la formule de Stirling,

$$m! \simeq \sqrt{2\pi m}\left(\frac{m}{e}\right)^m, \tag{3}$$

$n_p(m)$ devient

$$n_p(m) \simeq \sqrt{\frac{L}{2\pi m(L-m)}}\left(\frac{L}{e}\right)^L \left(\frac{e(1-p)}{L-m}\right)^{L-m}\left(\frac{ep}{m}\right)^m, \tag{4}$$

ou bien

$$\boxed{n_p(m) \simeq \sqrt{\frac{1}{2\pi L\frac{m}{L}(1-\frac{m}{L})}}\left(\frac{1-p}{1-\frac{m}{L}}\right)^{L(1-\frac{m}{L})}\left(\frac{p}{\frac{m}{L}}\right)^{L\frac{m}{L}}.} \tag{5}$$

En nommant

$$\frac{m}{L} = q, \tag{6}$$

l'éq. (5) devient

$$n'_p(q) = n_p(m) \simeq \frac{1}{\sqrt{2\pi Lq(1-q)}}\left(\frac{1-p}{1-q}\right)^{L(1-q)}\left(\frac{p}{q}\right)^{Lq}, \tag{7}$$

ou, exprimé en termes de logarithmes

$$\begin{aligned}\log_e n'_p(q) &\simeq \log_e n_p(m) = \log_e \frac{1}{\sqrt{2\pi Lq(1-q)}} \\ &+ L\left[(1-q)\left(\log_e(1-p) - \log_e(1-q)\right) + q\left(\log_e(p) - \log_e(q)\right)\right].\end{aligned} \tag{8}$$

3.2.2 Approximation gaussian

D'après la loi des grands nombres, cette loi est proche de la loi gaussienne de moyenne Lp et de variance $Lp(1-p)$. Ceci se vérifie en remplaçant q par $p+\varepsilon$:

$$\begin{aligned}n''_p(\varepsilon) &= \log_e n'_p(q) \simeq -\frac{1}{2}\log_e\left[2\pi L(p+\varepsilon)(1-p-\varepsilon)\right] \\ &+ L\left[(1-p-\varepsilon)\left(\log_e(1-p) - \log_e(1-p-\varepsilon)\right) + (p+\varepsilon)\left(\log_e(p) - \log_e(p+\varepsilon)\right)\right],\end{aligned} \tag{9}$$

$$n''_p(\varepsilon) \simeq -\frac{1}{2}\log_e\left(2\pi Lp(1-p)\right) + \left(-\frac{\varepsilon}{2(1-p)} + \frac{\varepsilon}{2p}\right) + L\left((1-p-\varepsilon)\frac{\varepsilon}{1-p} - (p+\varepsilon)\frac{\varepsilon}{p}\right). \quad (10)$$

Le second terme de l'eq. (10) peut être négligé quand L est grand

$$n''_p(\varepsilon) \simeq -\frac{1}{2}\log_e\left(2\pi Lp(1-p)\right) - L\frac{\varepsilon^2}{2p(1-p)}. \quad (11)$$

Le logarithme de la densité de $\frac{m}{L}$ est (fig 4)

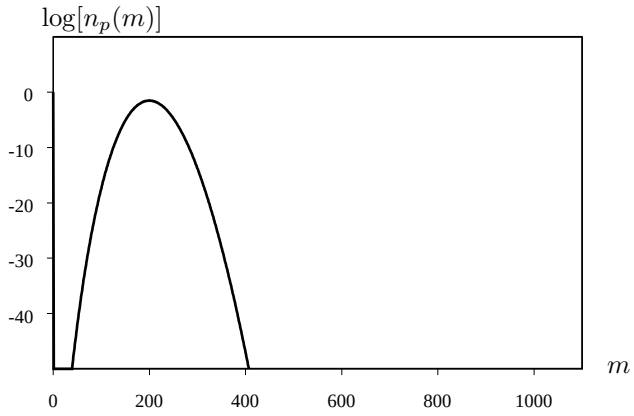

Figure 4: Densité de probabilité du nombre de "1"s dans des messages de longueur $L = 1000$ lorsque la probabilité p d'émettre un "1" est 0.2; mêmes données que précédemment présentées sur une échelle logarithmique. La forme parabolique de la fonction montre la validité de l'approximation par une loi gaussienne.

$$\boxed{\frac{1}{2}\log_e\frac{L}{2\pi p(1-p)} - \frac{L\left(\frac{m}{L}-p\right)^2}{2p(1-p)}.} \quad (12)$$

Quand L est grand le premier terme de la somme (11) peut aussi être négligé. La plupart des séquences ont un nombre de uns compris entre $L(p-\delta)$ et $L(p+\delta)$ où δ peut être aussi petit que l'on veut. Quand L est grand, la probabilité que m soit en dehors de ce domaine tend vers zéro (voir les fig. 5 et 6).

3.2.3 Inégalité de Bienaymé Chebyshev:

Il est peut-être intéressant de rappeler cette inégalité formulée dans le cas de l'approximation considérée ici :

$$\text{probability that} \quad \left(\left|\frac{m}{L} - p\right| > s\right) < \frac{p(1-p)}{s^2 L}. \quad (13)$$

Quand L est grand, la probabilité que m soit en dehors de ce domaine décroit au moins aussi vite que $1/L$. Dans le cas particulier considéré ici, cette probabilité décroit bien plus vite, comme $e^{-2(1-2p)L}$ (voir la section 3.3.2).

3.3 Quel est le nombre de message ayant exactement ou à peu près pL uns et $(1-p)L$ zéros ?

D'après la loi binomiale, pL étant entier, le nombre de messages ayant pL '1's et $(1-p)L$ '0's est :

$$n_{1/2}(p) = \frac{L!}{(pL)![(1-p)L]!}, \quad (14)$$

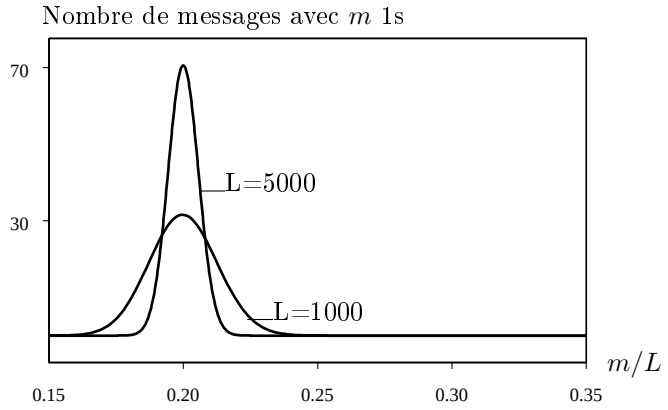

Figure 5: Densité de probability du nombre de "1"s dans des messages de longueur $L = 1000$ et $L = 5000$ lorsque la probabilité p d'émettre un "1" est 0.2.

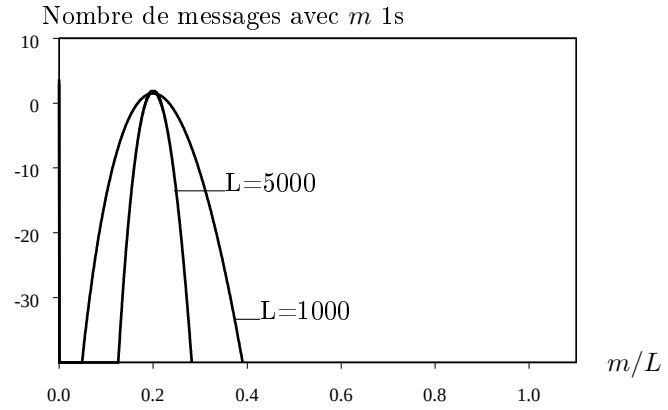

Figure 6: Densité de probabilité du nombre de "1"s dans des messages de longueur $L = 1000$ et $L = 5000$ quand la probabilité p d'émettre un "1" est 0.2; mêmes données que précédemment présentées sur une échelle logarithmique. La dérivée seconde de la parabole est $-\frac{L}{2p(1-p)}$.

alors que le nombre total de messages possibles est 2^L. Nous avons

$$\sum_{pL=0}^{L} n_{1/2}(p) = \sum_{pL=0}^{L} \frac{L!}{(pL)![(1-p)L]!} = 2^L. \tag{15}$$

3.3.1 Approximation fondée sur la formule de Stirling

D'après la formule de Stirling, ce nombre peut être approximé par

$$n_{1/2}(p) = \frac{\sqrt{2\pi L}\left(\frac{L}{e}\right)^L}{\sqrt{2\pi Lp}\left(\frac{Lp}{e}\right)^{Lp}\sqrt{2\pi L(1-p)}\left(\frac{L(1-p)}{e}\right)^{L(1-p)}}. \tag{16}$$

Il peut s'écrire

$$n_{1/2}(p) = \frac{1}{\sqrt{2\pi Lp(1-p)}p^{Lp}(1-p)^{L(1-p)}}, \tag{17}$$

ou bien

$$n_{1/2}(p) = \frac{1}{\sqrt{2\pi Lp(1-p)}2^{-L(p\log_2 p+(1-p)\log_2(1-p))}}, \tag{18}$$

ou encore

$$n_{1/2}(p) = 2^{-L(p\log_2 p+(1-p)\log_2(1-p))-\frac{1}{2}\log_2(2\pi Lp(1-p))}. \tag{19}$$

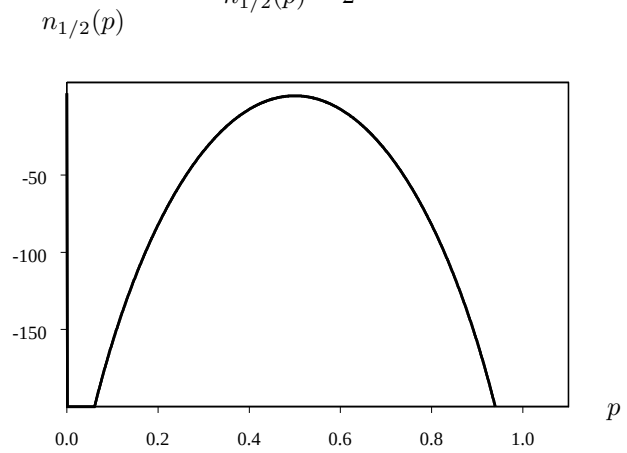

Figure 7: Proportion de messages de longueur L comportant exactement m uns, (échelle logarithmique); cette proportion est $2^{L(H_B-1)}$.

Quand L est grand, le terme prépondérant dans l'exposant est proportionnel à l'entropie (fig. 7 et 8)

$$\boxed{LH_B(p) = L\left(-p\log_2 p - (1-p)\log_2(1-p)\right).} \tag{20}$$

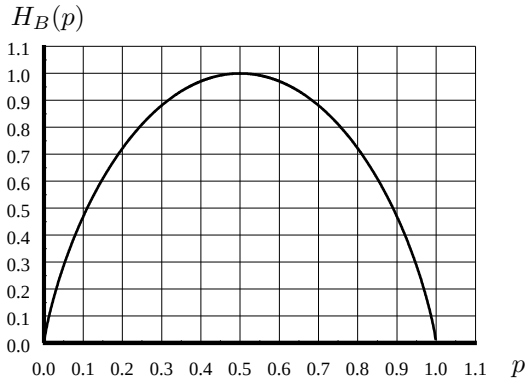

Figure 8: Fonction entropie.

Le nombre de messages avec Lp uns est inférieur à $2^{L(H_B(p)+\delta)}$ où δ peut être aussi petit que l'on veut. $L(H_B(p)+\delta)$ bits sont suffisants pour les coder.

3.3.2 Le nombre de bits donné par le premier théorème est suffisant pour coder tous les messages pour lesquels la probabilité de 'uns' est plus petite que p

Nous considérons des messages pour lesquels la probabilité de 'uns' est

$$\sum_{Lx=0}^{Lp} n_{1/2}(x) = \sum_{Lx=0}^{Lp} \frac{L!}{Lx!L(1-x)!}, \tag{21}$$

où x est dans le domaine $[0,1]$ et Lx est un entier.

Nous voulons montrer que cette probabilité est bornée par

$$\sum_{Lx=0}^{Lp} n_{1/2}(x) \leq n_{1/2}(p)(1+\delta), \tag{22}$$

où δ peut être rendu aussi petit qu'on le souhaite en choisissant L suffisamment grand. Quand L est suffisamment grand, nous pouvons utiliser l'approximation suivante par une gaussienne :

$$n_{1/2}(x) \simeq \sqrt{\frac{2L}{\pi}} e^{-2L\left(x-\frac{1}{2}\right)^2}. \tag{23}$$

Quand x décroit à partir de p ($x < p$), cette fonction décroit extrêmement rapidement (si p n'est pas trop près de 0.5)

$$n_{1/2}(x) \simeq \sqrt{\frac{2L}{\pi}} e^{-2L\left(p-\frac{1}{2}\right)^2} e^{-2L\left[\left(x-\frac{1}{2}\right)^2-\left(p-\frac{1}{2}\right)^2\right]}, \tag{24}$$

ou

$$n_{1/2}(x) \simeq n_{1/2}(p) e^{-2L[x-p](x-1+p)} < n_{1/2}(p) e^{2(1-2p)L[x-p]}. \tag{25}$$

Par conséquent,

$$\sum_{Lx=0}^{Lp} n_{1/2}(x) < \frac{1-e^{-2Lp(1-2p)}}{1-e^{-2(1-2p)L)}} n_{1/2}(p). \tag{26}$$

ou bien pour L grand :

$$\sum_{Lx=0}^{Lp} n_{1/2}(x) < (1+e^{-2(1-2p)L}) n_{1/2}(p). \tag{27}$$

Ainsi les messages correspondant à une probabilité inférieure à p peuvent être négligés car leur nombre est très petit en comparaison du nombre de messages correspondant à une probabilité proche de p: la décroissance de $e^{-2L(x-p)^2}$ est très rapide lorsque $(x-p)$ décroit (fig 9 et 10). Cette approximation nous sera utile pour l'illustration du deuxième théorème dans la section 4.

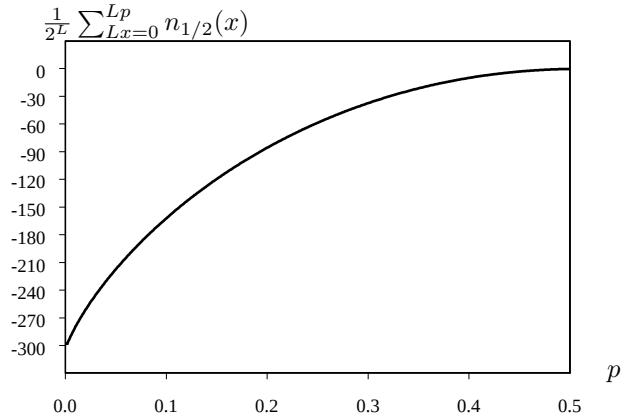

Figure 9: Proportion de messages où la probabilité de 'uns' est comprise entre 0 et p (échelle logarithmique) en fonction de p pour $L = 1000$. Ce nombre est comparable au nombre de messages ayant à peu près p 'uns', $2^{LH(p)}$.

3.4 Le principle des techniques de compression

Pour comprimer des messages,les messages improbables (ceux qui n'ont pas à peu près p uns, (ou si l'on préfère plus de $p+\delta$ uns comme dans la section 3.3.2) sont écartés (fig 11), et les $2^{LH_B(p)}$

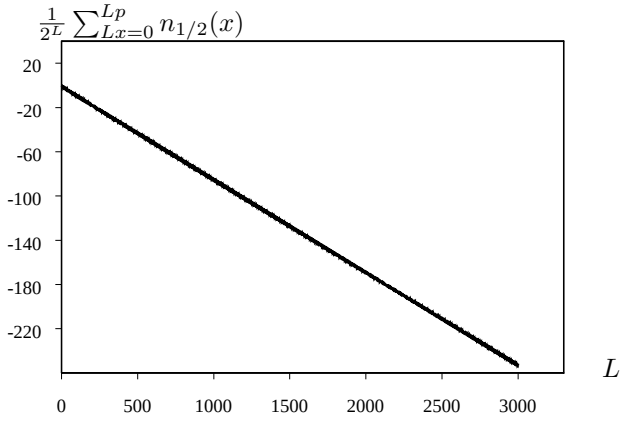

Figure 10: Proportion de messages où la probabilité de 'uns' est comprise entre 0 et $p = 0.2$ (échelle logarithmique) en fonction de L. Ce nombre est comparable au nombre de messages ayant à peu près p 'uns'.

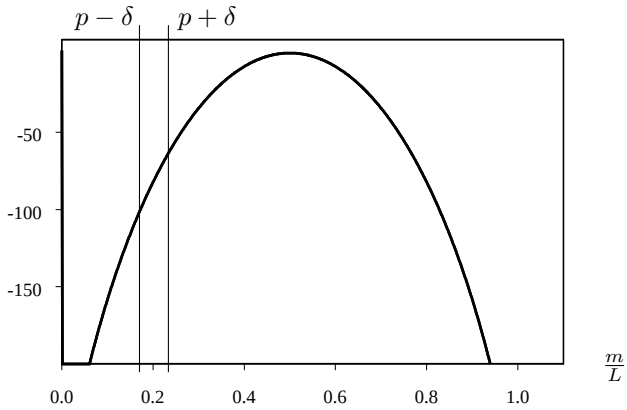

Figure 11: Proportion de messages de longueur L ayant exactement m uns, échelle logarithmique ; il n'est pas nécessaire de coder les messages ayant plus de $p + \delta$ uns ou moins de $p - \delta$ uns car ils sont très peu probables.

messages restants sont affectés d'un numéro, par exemple en utilisant le codage arithmétique de Rissanen et Langdon.

Ces messages acceptables sont appelés messages *"typiques"*. Etant donné le nombre de messages différents de cette forme, il suffit d'une longueur légèrement supérieure à $LH_B(p)$ pour les coder en bits.

3.4.1 Approximation de l'entropie pour les petites valeurs de p

Il peut être utile d'avoir à l'esprit cette approximation qui en donne un ordre de grandeur. Quand p est petit,

$$LH_B(p) = Lp \log_2 \frac{L}{Lp} - L(1-p) \log_2 e \log_e(1-p), \tag{28}$$

$$\boxed{LH_B(p) \simeq Lp \left(\log_2 \frac{L}{Lp} + \log_2 e \right).} \tag{29}$$

Pour chaque Lp 'un's du message, le nombre de bits nécessaire au codage est donné par le nombre de bits nécessaire pour coder la longueur moyenne séparant deux 'uns', soit $\frac{L}{Lp}$ plus $\log_2 e$.

4 Illustration du second théorème de Shannon dans un cas simple

Un message d'entropie 1 ($p_M = 1/2$) et de longueur M peut être rallongé en lui adjoignant un syndrome afin de construire un message de longueur L. Le syndrome se déduit du message original par des operations déterministes. Il sera utilisé pour corriger les erreurs de transmission.

Le message redondant de longueur L ne contient pas nécessairemnt le texte explicite du message original de longueur M. Il peut s'obtenir par le choix de 2^M mots parmi les 2^L mots qui peuvent être transmis. Si le message codé contient explicitement le message original, on dit que le code est "systematique", mais ceci réduit bien sûr le nombre des codes possibles.

L'entropie du message étendu est $H_M = M/L$: il y a 2^M messages et l'entropie H_M d'un sous ensemble de 2^M éléments pris parmi 2^L éléments est telle que

$$2^M = 2^{LH_M}. \tag{30}$$

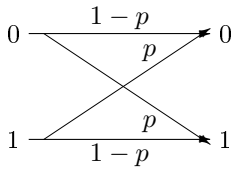

Figure 12: Canal binaire symétrique avec une probabilité d'erreur p.

Ce message est transmis à un récepteur et modifié par un bruit binaire indépendant du message original. (fig 12). Cette transmission est caractérisée par une probabilité d'erreur p, ou une entropie $H_B = -p\log_2 p - (1-p)\log_2(1-p)$.

Dans la section 4.1,nous donnons une interprétation du théorème dans ce cas simple ; dans la section 4.2 nous donnons la formulation correspondante du problème ; dans la section 4.3 nous déduisons le nombre maximum de messages possibles et dans la section 4.4 nous montrons finalement que nous pouvons disposer d'un nombre possible de messages aussi proche qu'on le désire de cette borne.

4.1 Interpretation du second theorem

Le second théorème de Shannon dit que si la somme des entropies du message M et du bruit independant B, soit $H_M + H_B$ vérifie

$$H_M + H_B < 1, \tag{31}$$

alors il est possible de trouver une méthode pour coder M de telle sorte que il est presque toujours possible de reconstruire exactement le message M à partir du message reçu et perturbé par le bruit. La longueur de la redondance introduite dans le message, $L - M$ doit être suffisante pour décrire le bruit, et ainsi pour décoder le message original (fig 13):

$$L - M > LH_B, \tag{32}$$

ou

$$\boxed{\frac{M}{L} < 1 - H_B.} \tag{33}$$

En bon mathématicien, Claude Shannon montre qu'il existe certainement une méthode pour effectuer le codage permettant le décodage sans erreur, mais il ne propose aucune piste pour la trouver ! Il ne dit pas s'il est possible de trouver des codeurs efficaces pour lesquels le décodage ne sera pas excessivement complexe... On peut déduire de la démonstration de Shannon que la

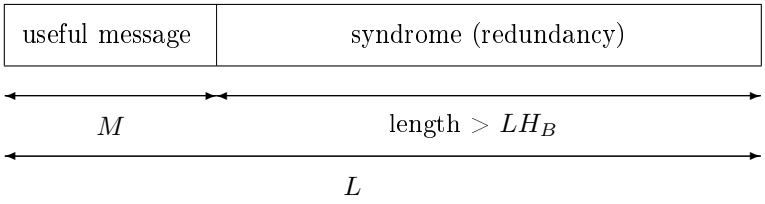

Figure 13: Allongement du message par adjonction d'un syndrome de longueurs suffisante pour que les erreurs de transmission puissent être corrigées. La longueur minimale du syndrome est proportionnelle à l'entropie du bruit B (nous supposons que l'entropie du message origianl de longueur M est égale à 1, et qu'en conséquence il ne peut pas être comprimé.

plupart des codes sont probablement de bons codes, car en moyenne, ils atteignent l'objectif désiré ; cependant, la complexité de leur décodage empêche d'envisager leur utilisation.

Trouver une méthode de codage pertinente, et plus particulièrement une méthode de décodage raisonnablement complexe demeure un problème ouvert. Il a fallu attendre quarante cinq ans pour voir une proposition de codage et de décodage dont les performances sont proches de la borne de Shannon : les turbocodes inventés par Claude Berrou et ses collègues ... Les vérifications expérimentales sont convaincantes, mais il ne semble pas qu'il y ait pour le moment une justification théorique rigoureuse des performances des turbocodes.

La démonstration de Shannon est abstraite, élégante et concise. Le lecteur peut se référer à plusieurs ouvrages (voir par exemple les références bibliographiques à la fin du document.) Ici le but est seulement d'illustrer dans un cas simple les points principaux de cette démonstration qui ne me paraissent pas intuitifs ; ceci peut peut-être aider des étudiants ou des personnes intéressées par le sujet qui ne dominent pas bien des notions qui sont probablement considérées comme évidentes par les experts de la théorie de l'information et sur lesquelles ceux-ci insistent rarement. Le point central du théorème est le suivant :

4.2 Formulation du problème dans un cas simple

On émet des messages longs (longueur L). Ils sont reçus avec au plus Lp bits faux. Les 2^M messages possibles sont codés par 2^M mots parmi les 2^L mots possibles ; si le rapport $\frac{M}{L}$ est plus petit que la borne (33) fonction de l'entropie du bruit, alors il est très peu probable qu'un message codé et entâché de Lp erreurs puisse être avec un des autes messages qui aurait pu être émis ; "*confondu*" signifiant que la distance de Hamming entre deux messages est inférieure à Lp.

4.2.1 Codage aléatoire

Le nombre de codeurs possibles est très grand : un codeur transforme deux des 2^M messages en deux mots de code differents parmi les 2^L possibles. Le nombre de codeurs est

$$\frac{(2^L)!}{(2^{L-M})!} \text{ or } \sqrt{\frac{2^L}{2^{L-M}}} \frac{2^{L2^L-(L-M)2^{L-M}}}{e^{2^L-2^{L-M}}}.$$

Shannon suppose que tous ces codeurs peuvent être choisis avec la même probabilité. Ici nous prenons un codeur au hasard. La probabitité qu'un des 2^L mots est un mot du code est

$$\frac{2^M}{2^L},$$

car il y a 2^L mots differents et 2^M messages possibles. Quand L augmente pour un M fixé, cette proportion décroît rapidement. Il sera possible de trouver des mots de code de telle sorte que la distance entre deux mots de code soit plus grande qu'un seuil donné, ce qui permettra d'éviter la confusion entre ces mots de code.

4.2.2 Critère de décodage

La probabilité d'une erreur de transmission est p. A la réception un mot de code est reconnu comme un des 2^M messages si la distance de Hamming entre le mot de code correspondant à ce dernier et le mot reçu est inférieure à Lp. Un mot reçu correspondant à l'émission d'un mot du code est certainement reconnu correctement car il y a au plus Lp erreurs dans le mot reçu. Mais il faut encore trouver les conditions garantissant qu'il n'y a pas d'erreur à la reconnaissance (de confusion entre deux messages). Nous supposerons que le récepteur ne reçoit que des mots de code altérés par du bruit.

4.3 Une borne supérieure sur le nombre de messages possibles de longueur L

Si la probabilité d'erreur est au plus p, il est possible de trouver un code tel qu'un mot de longueur L (supposée très grande) puisse être utilisé pour coder $2^{L(1-H_B(p))}$ messages : du fait des erreurs de probabilité p, chacun des 2^M messages peut être transformé en un des 2^{LH_B} messages altérés possibles à la reception : pour éviter les confusions, M doit vérifier

$$2^M 2^{LH_B} \leq 2^L. \tag{34}$$

Lorsqu'il n'y a pas de bruit de transmission, $(p = 0, H_B(p) = 0)$, il est possible de coder 2^L messages ; si $p = \frac{1}{2}$ et $H_B(p) = 1$, il n'est pas possible de transmettre d'information par ce canal.

> Il ne peut pas y avoir plus de $2^{L-LH_B(p)}$ messages differents, car il y a $2^{LH_B(p)}$ configurations du bruit et 2^L mots de code possibles.

Un plus grand nombre de messages impliquerait nécessairement des erreurs à la reconnaissance. Nous montrons maintenant que le nombre de messages peut être aussi proche que l'on veut de cette borne.

4.4 Interpretation fondée sur la formule de Stirling

Quand L est grand, de mots de code différents pris au hasard parmi les 2^M ont en moyenne $\frac{L}{2}$ bits identiques. La distribution du nombre de bits identiques entre deux des 2^L mots suit la loi binomiale

$$n_{1/2}(x) = \frac{L!}{(xL)![(1-x)L]!}. \tag{35}$$

Après réception d'un mot du code, le mot de code erroné correspondant a en moyenne $\frac{L}{2}$ bits en commun avec n'importe lequel des mots de code ; leur nombre suit toujours la loi binomiale (35).

On suppose qu'il y a 2^M mots de code différents : nous voulons voir si un des $(2^M - 1)$ autres mots du code, après reception, peut être confondu avec celui qui a été émis, c'est à dire si la distance de Hamming entre un des $(2^M - 1)2^{LH_B}$ mots reçus et le M-th est inférieure à $L(p + \varepsilon)$.

La probabilité que cette distance est inférieure à Lp est

$$Q = \sum_{xL=0}^{Lp} \frac{1}{2^L} n_{1/2}(x) = \sum_{xL=0}^{Lp} \frac{L!}{xL!(L-xL)!} \left(\frac{1}{2^L}\right), \tag{36}$$

qu'on peut approximer par

$$Q = \sum_{xL=0}^{Lp} \frac{1}{\sqrt{2\pi L(1-x)x}} \frac{1}{(1-x)^{L(1-x)}} \frac{1}{(x)^{Lx}} \left(\frac{1}{2^L}\right). \tag{37}$$

Nous avons vu dans la section 3.3.2, eq. (22 - 27), que l'ordre de grandeur de cette probabilté d'erreur est donnée par (fig. 14)

$$Q \simeq \frac{n_{1/2}(p)}{2^L} \simeq \frac{2^{LH_B}}{2^L}. \tag{38}$$

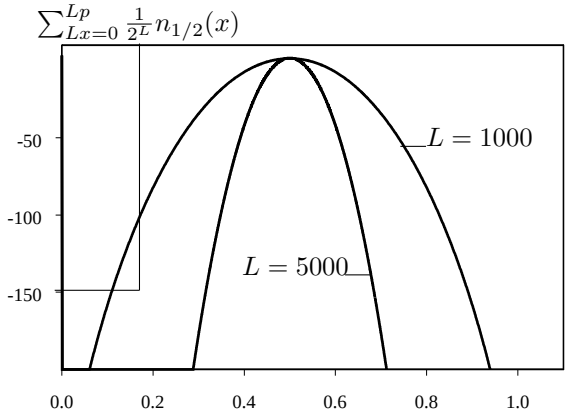

Figure 14: Probabilité d'erreur entre deux mots (échelle logarithmique) : Quand L est suffisamment grand, la probabilité de confondre un message avec un autre peut être rendue aussi petite qu'on le souhaite. Si cette probabilité décroit plus vite que 2^M quand L augmente, $\frac{M}{L}$ restant constant, il sera presque toujours possible de reconnaitre le mot émis.

On peut déduire de cette formule une borne sur la probabilité qu'aucun des $2^M - 1$ mots est à une distance plus petite que Lp de ce mot : La probabilité qu'il y ait au moins une erreur est bornée par

$$S = \sum_{1}^{2^M-1} Q \simeq 2^M Q \simeq 2^{M+LH_B-L}. \tag{39}$$

Nous supposons que la condition (33) est vérifiée :

$$M + LH_B - L < 0. \tag{40}$$

Si L est suffisamment grand, la fraction $\frac{M}{L}$ restant constante, il existe un α negatif tel que

$$S \simeq 2^{M+LH_B-L} < 2^\alpha, \tag{41}$$

$$\frac{M}{L} < 1 - H_B + \frac{\alpha}{L}. \tag{42}$$

Si L augmente, la probabilité d'erreur décroit lorsque la condition (33) :

$$H_M = \frac{M}{L} < 1 - H_B, \tag{43}$$

est vérifiée. Cette borne est la *capacité* du canal. Il peut être intéressant de montrer la redondance

$$\frac{L}{M} = \frac{1}{1 - H_B}, \tag{44}$$

nécessaire pour vérifier la borne de Shannon (fig 15).

5 Conclusion

Nous avons proposé deux illustrations simples des théorèmes de Shannon fondées sur l'utilisation de la formule de Stirling.

Même si le résultats de Shannon sont bien plus généraux, et en dépit de l'inélégance des développements, nous espérons que cette présentation peut aider à la compréhension des aspects concrets de ces théorèmes. Les suggestions d'amélioration et les corrections sont bienvenues, envoyez un mail à leroux@essi.fr.

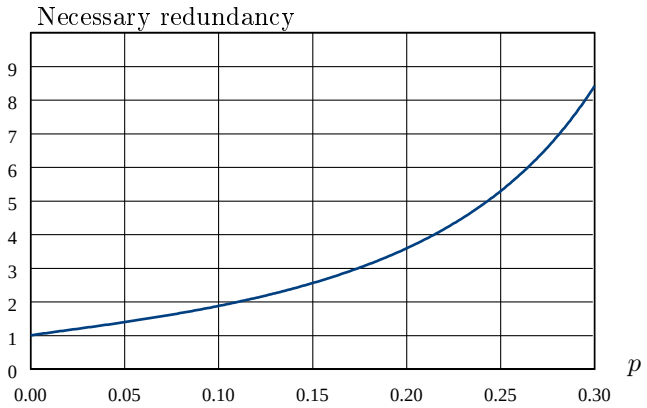

Figure 15: Redondance correspondant à la borne de Shannon en fonction de la probabilté d'erreur de transmission.

6 Bibliographie

6.1 Quelques references historiques

R. Clausius, "Ueber verschiedene für die anwendung bequeme formen der Hauptgleichungen der mechanischen Wärmetheorie", (On different forms, convenient for application, of the main equations of the mechanical heat theory) Annalen der physik und chemie, band CXX5, no 7, 1865, pp 353-400.

L. Boltzmann, "Uber die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung, respective den Sätzen über das Wärmegleichgewicht," (On the Relation Between the Second Law of the Mechanical Theory of Heat and the Probability Calculus with Respect to the Theorems on Thermal Equilibrium), Sitzb. d. Kaiserlichen Akademie der Wissenschaften, mathematich-naturwissen Cl. LXXVI, Abt II, 1877, pp. 373-435.

M. Planck, "Uber des Gesetz der Energieverteilung im Normalspectrum", "On the Law of Energy Distribution in Normal Spectra", Annalen der Physik, 4, 1901, pp 553-563. (french translation : A propos de la loi de distribution de l'énergie dans le spectre normal, Sources et évolution de la physique quantique, textes fondateurs, J. Leite-Lopes et B. Escoubès, edts, Masson, 1995. pp. 20-27.)

A. Einstein, "Uber einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt," ("On a Heuristic Viewpoint Concerning the Production and Transformation of Light") Annalen der Physik, 17, 1905, pp. 132-148. (french translation : Un point de vue heuristique concernant la production et la transformation de la lumière, Sources et évolution de la physique quantique, textes fondateurs, J. Leite-Lopes et B. Escoubès, edts, Masson, 1995. pp. 28-40.)

C. E. Shannon, "A mathematical theory of communication," Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

L. Brillouin, "Science and Information theory", Academic Press, 1962.

R. G. Gallager,"The work of Claude Shannon," IEEE Trans. on IT, nov. 2001.

6.2 References imprimées

R.G. Gallager, "Information theory and reliable communication", Wiley, 1968.

T. M. Cover and J.A. Thomas, "Elements of information theory", Wiley, 1991.

G. Battail, "Théorie de l'information, application aux techniques de communications", Masson, 1997 (in french).

J. Rissanen and G.G. Langdon, "Arithmetic coding", IBM J. Res. Develop., Vol. 23, No. 2, pp. 149-162, March 1979.

J. Rissanen and G.G. Langdon, "Universal modeling and coding", IEEE Trans. on Information Theory, Vol. 27, No. 1, pp. 12-23, January 1981.

C. Berrou, A. Glavieux and P. Thihimajshima, "Near Shannon limit error-correcting coding and decoding : turbo codes", Proc. 1993, Int. Conf. Comm., pp 1064-1070.

C. Berrou and A. Glavieux, "Near Shannon limit error-correcting coding and decoding : turbo codes", IEEE Trans. Comm., Oct. 1996, pp. 1261-1271.

6.3 Sites web

Articles de Shannon :

http: //cm.bell-labs.com/cm/ms/what/shannonday/paper.html

The courses of Marc Uro (Institute of telecommunications, Evry, France), in french :

http: //www-sim.int-evry.fr/~ uro/old.htm

David J.C. MacKay, "Information Theory, Inference and Learning Algorithms", Cavendish Laboratory, Cambridge, Great Britain, January 1995 :

http: //www.inference.phy.cam.ac.uk/mackay/info-theory/course.html

http: //www.inference.phy.cam.ac.uk/mackay/itprnn/book.html#book

Explanation of Stirling's formula on the page of B. Gourevitch about π (in french) :

http: //membres.lycos.fr/bgourevitch/mathematiciens/moivre/moivre.html

Une traduction en anglais d'un des articles de Boltzmann :

http: //www.essi.fr/ ~ leroux/boltztrad.ps