Capacité d'un canal – Second Théorème de Shannon

Plan du cours

1. Canaux discrets sans mémoire, exemples ;
2. Capacité ;
3. Canaux symétriques ;
4. Codage de canal ;
5. Second théorème de Shannon ; esquisse de preuves.

1. Canal discret sans mémoire

Définition Un canal discret est défini par la donnée de
- un alphabet d'entrée $X = \{a_1, \ldots, a_K\}$
- un alphabet de sortie $Y = \{b_1, \ldots, b_J\}$
- une loi de transition $P_{Y|X}$, i.e. une matrice stochastique

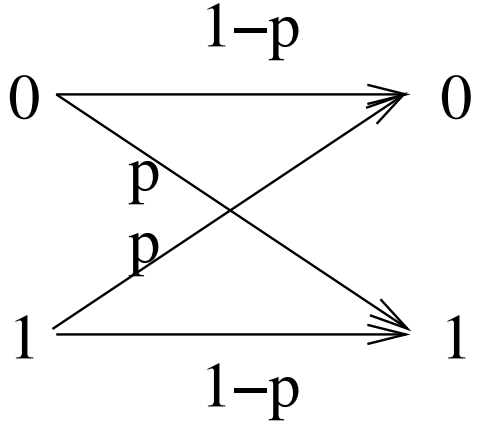
$$\Pi = \begin{pmatrix} \mathbf{P}(b_1 \mid a_1) & \ldots & \mathbf{P}(b_J \mid a_1) \\ \vdots & \ddots & \vdots \\ \mathbf{P}(b_1 \mid a_K) & \ldots & \mathbf{P}(b_J \mid a_K) \end{pmatrix}$$

Le canal est *sans mémoire* si pour tout $(x_1, \ldots, x_n)$ transmis et $(y_1, \ldots, y_n)$ reçu, on a

$$\mathbf{P}(y_1, \ldots, y_n | x_1, \ldots, x_n) = \mathbf{P}(y_1|x_1) \ldots \mathbf{P}(y_n|x_n).$$

Nous étudierons principalement les canaux sans mémoire.

Exemple – Canal binaire symétrique

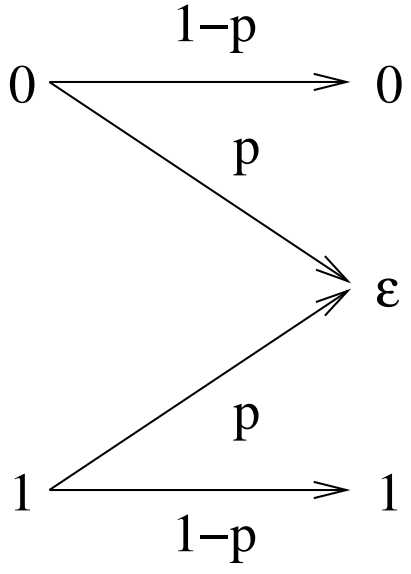


La matrice stochastique est

$$\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}.$$

p est appelé *probabilité de transition* ou *probabilité d'erreur* du canal.

Canal à effacement



$$\Pi = \begin{pmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{pmatrix}.$$

2. Capacité

La capacité d'un canal est définie par l'information mutuelle maximale entre une variable aléatoire X à valeurs sur l'alphabet d'entrée du canal et sa sortie correspondante Y par le canal :

$$C \stackrel{\text{def}}{=} \sup_X I(X;Y) \text{ avec}$$

$$X \stackrel{\text{canal}}{\rightsquigarrow} Y$$

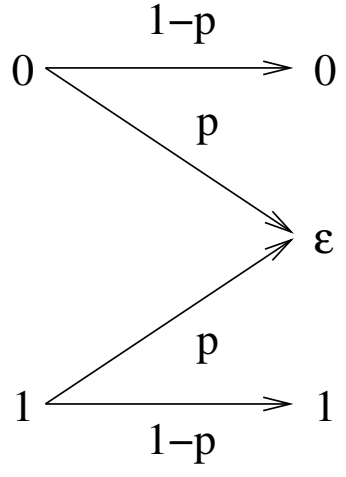
Capacité

On remarquera que $I(X;Y)$ peut s'écrire en fonction des seules lois de transition et d'émission :

$$I(X;Y) = \sum_{x,y} \mathbf{P}(y \mid x)\mathbf{P}(x) \log_2 \frac{\mathbf{P}(y \mid x)}{\mathbf{P}(y)}$$

$$\mathbf{P}(y) = \sum_{x} \mathbf{P}(y \mid x)\mathbf{P}(x).$$

Capacité du canal binaire à effacement



$$C \quad = \quad \max_{p(x)} I(X;Y) = \max_{p(x)} (H(Y) - H(Y|X))$$

Noter que $H(Y|X) = \mathbf{P}(X=0)h(p) + \mathbf{P}(X=1)h(p) = h(p)$ avec $h(p) \stackrel{\text{def}}{=} -p\log_2 p - (1-p)\log_2(1-p)$.

Capacité du canal binaire à effacement (II)

Posons $a \stackrel{\text{def}}{=} \mathbf{P}(X = 1)$, on a :

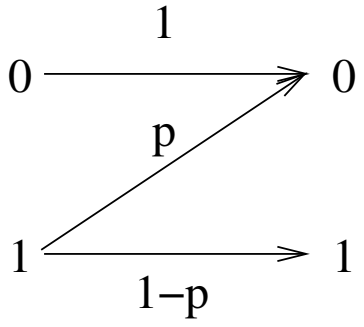
$$\begin{aligned} \mathbf{P}(Y = 1) &= a(1-p) \\ \mathbf{P}(Y = 0) &= (1-a)(1-p) \\ \mathbf{P}(Y = \epsilon) &= ap + (1-a)p = p \end{aligned}$$

$$\begin{aligned} H(Y) &= -a(1-p)\log a(1-p) \\ &\quad -(1-a)(1-p)\log(1-a)(1-p) - p\log p \\ &= (1-p)h(a) + h(p) \end{aligned}$$

Donc

$$C = \max_a (1-p)h(a) = 1-p$$

Canal en Z



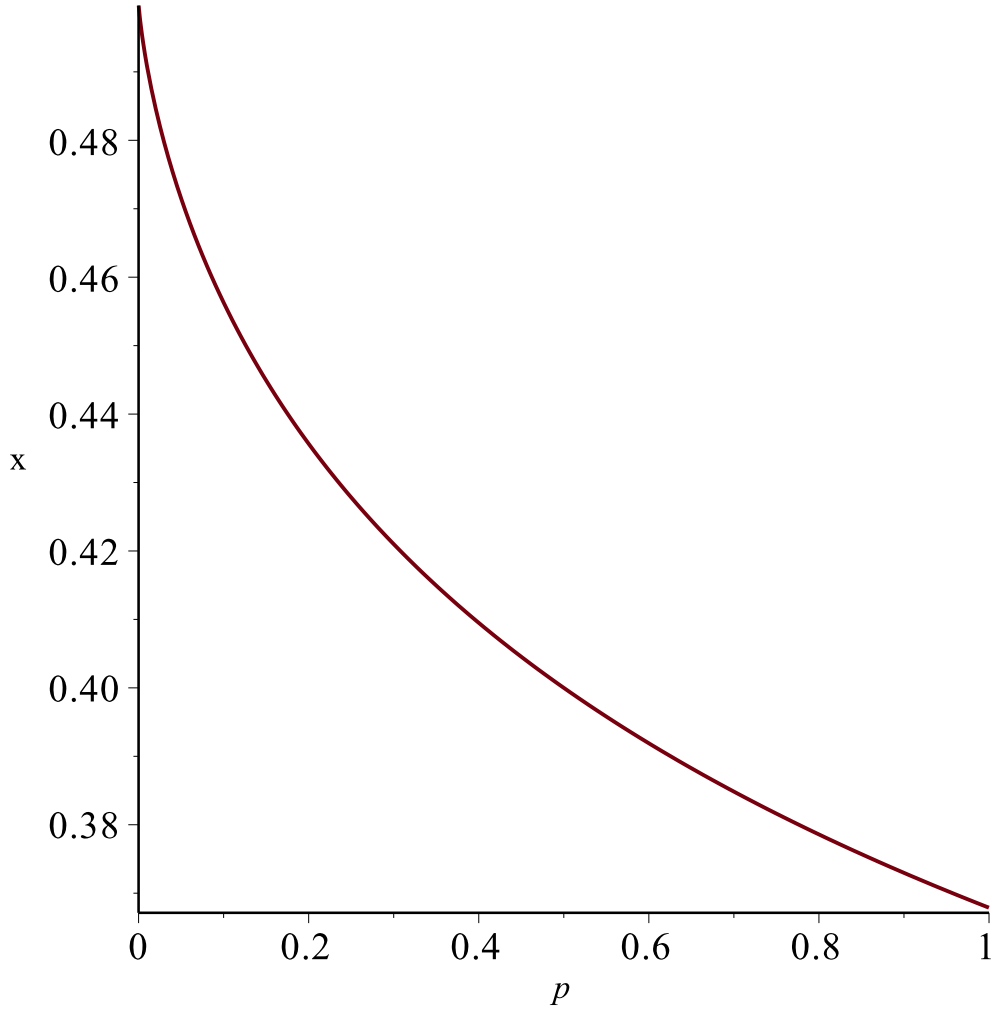
La matrice stochastique est

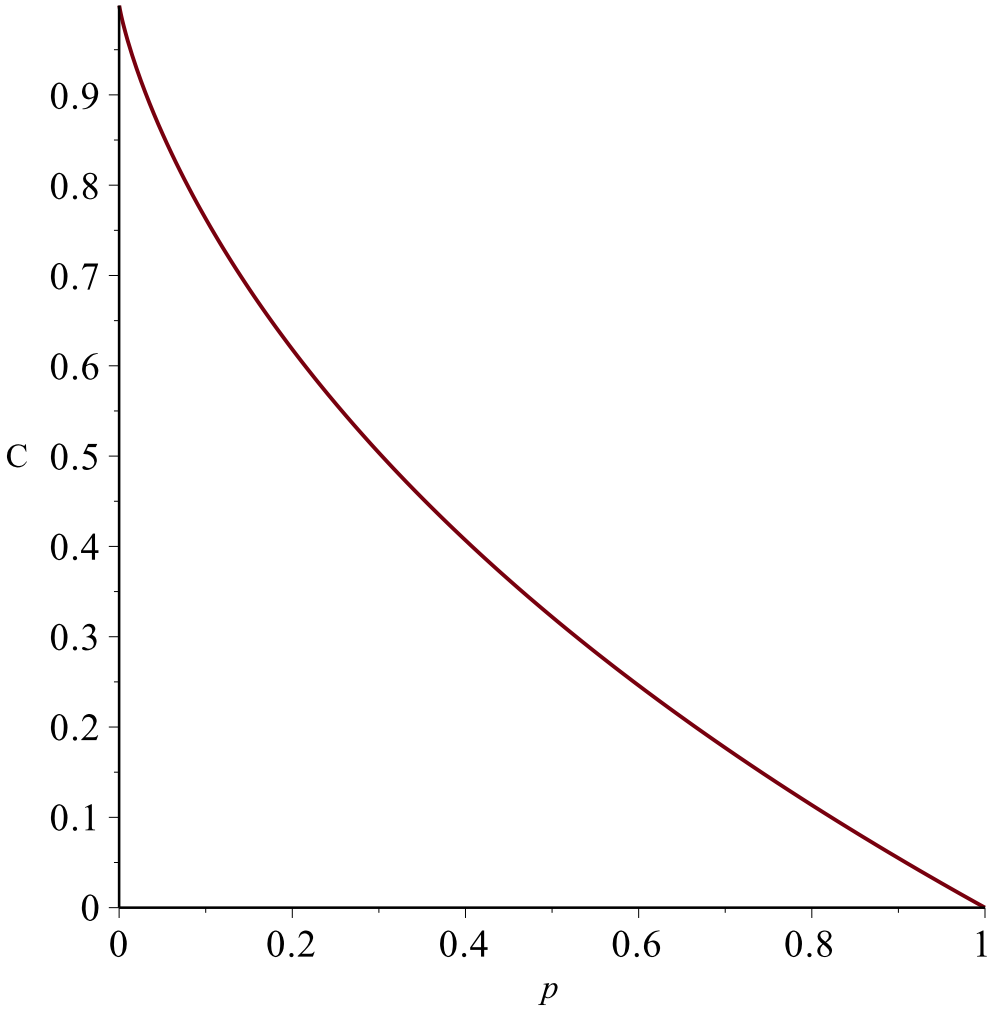
$$\begin{pmatrix} 1 & 0 \\ p & 1-p \end{pmatrix}.$$

Pour une distribution $x \stackrel{\text{def}}{=} \mathbf{P}(X=1)$, on a

$$I(X;Y) = h(x(1-p)) - xh(p)$$

Maximum atteint en $x = \left((1-p)\left(1+2^{\frac{h(p)}{1-p}}\right)\right)^{-1}$


0.48
0.46
0.44
x
0.42
0.40
0.38
0
0.2
0.4
0.6
0.8
1
p


C
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
p

Canaux symétriques

Définition Un canal discret est dit *symétrique* si les lignes et les colonnes de sa matrice stochastique sont égales à une permutation près.

Proposition 1. *Dans un canal symétrique* $H(Y|X)$ *ne dépend pas de* X*.*

$$H(Y|X) = -\sum_y P(y \mid x) \log_2 P(y \mid x).$$

Preuve

$$
\begin{aligned}
H(Y \mid X) &= -\sum_{x,y} P(x,y) \log_2 P(y \mid x) \\
&= -\sum_{x} P(x) \sum_{y} P(y \mid x) \log_2 P(y \mid x) \\
&= \sum_{x} P(x) H(\Pi) = H(\Pi)
\end{aligned}
$$

où $H(\Pi) = -\sum_y P(y \mid x) \log_2 P(y \mid x)$ est indépendent de x.

Capacité d'un canal symétrique

donc

$$\begin{aligned} C &= \sup_X I(X;Y) \\ &= \sup(H(Y) - H(Y \mid X)) \\ &= \sup(H(Y)) - H(\Pi) \\ &\leq \log_2 |Y| - H(\Pi). \end{aligned}$$

L'entropie est maximisée quand la distribution de Y est uniforme. On a que Y est uniforme quand X est uniforme pour un canal symétrique.

Capacité d'un canal symétrique (II)

Proposition 2. *La capacité d'un canal fortement symétrique est atteinte pour une loi d'émission uniforme et vaut*

$$C = \log_2 |Y| - H(\Pi)$$

Exemple

Capacité du canal binaire symétrique :

$$C = 1 - h(p)$$

A comparer avec la capacité du canal binaire à effacement :

$$C = 1 - p$$

3. Codage de canal

Codage de canal

Nous considérons un canal discret $\mathcal{T} = (X, Y, \Pi)$

Définition Un *code en bloc* de *longueur* n et de *cardinal* M est M séquences de n lettres de X. Nous parlerons de code (M, n). Le *taux de transmission* d'un code est égal à

$$R = \frac{\log_2 M}{n}$$

Un code va permettre de « coder » une quantité d'information égale à $\log_2 M$ bits, par unité de temps.

R est aussi le nombre de bits transmis par usage du canal.

Un *codeur* est une procédure qui associe à toute séquence binaire finie une séquence finie de lettres de X.

Performance d'un code – Décodage

Soit $\mathcal{C}$ un code en bloc (M, n) utilisé dans un canal discret (X, Y, Π)

Définition Un *algorithme de décodage* de $\mathcal{C}$ est une procédure qui a tout bloc de n lettres de Y associe un mot de code de $\mathcal{C}$.

L'événement « mauvais décodage » pour un algorithme de décodage et un canal donné est défini par :

> *Un mot de code* $\mathbf{x} \in \mathcal{C} \subset X^n$ *est transmis à travers le canal, le mot* $\mathbf{y} \in Y^n$ *est reçu et est décodé en* $\tilde{\mathbf{x}} \neq \mathbf{x}$.

Définition Le *taux d'erreur* de $\mathcal{C}$ (dans le canal considéré) noté $P_e(\mathcal{C}, \mathbf{x})$ est la probabilité de mauvais décodage quand $\mathbf{x}$ est transmis

Exemples pour le canal binaire symétrique

Code à répétition de longueur 3

$$C = \{000, 111\}$$

Code de parité de longueur 4

$$C = \{0000, 0011, 0101, 0110, 1001, 1010, 1100, 1111\}$$

Code de Hamming de longueur 7

$$\begin{aligned} C \quad = \quad & \{0000000, 1101000, 0110100, 0011010, \\ & 0001101, 1000110, 0100011, 1010001, \\ & 1111111, 0010111, 1001011, 1100101, \\ & 1110010, 0111001, 1011100, 0101110\} \end{aligned}$$

Décodage du code de Hamming

La distance de Hamming $d(x, y)$ est

$$d(x, y) = |\{i; x_i \neq y_i\}|$$

On vérifie que tous les mots du code de Hamming sont au moins à distance trois les uns des autres. Donc les boules de rayon 1 centrées sur les mots de code ne s'intersectent pas.

On vérifie aussi que tout mot de l'espace ambiant $\{0, 1\}^7$ est à distance au plus 1 d'un mot de code :

$$16(1 + 7) = 2^4 \times 2^3 = 2^7.$$

Algorithme de décodage : pour y reçu, retourner le mot de code x à distance au plus 1 de y.

Second théorème de Shannon

Théorème 1. *Soit un canal discret sans mémoire de capacité C. Pour tout $R < C$, il existe une suite de codes en bloc avec un algorithme de décodage associé $(\mathcal{C}_n(M,n))_{n>0}$ de taux de transmission R_n telle que*

$$\lim_{n\to\infty} R_n = R \quad \textit{et} \quad \lim_{n\to\infty} \sup_{\mathbf{x}\in\mathcal{C}_n} P_e(\mathcal{C}_n, \mathbf{x}) = 0$$

Théorème 2. *Soit un canal discret sans mémoire de capacité C. Tout code $\mathcal{C}$ de taux de transmission $R > C$ vérifie $\frac{1}{M}\sum_{\mathbf{x}\in\mathcal{C}} P_e(\mathcal{C}, \mathbf{x}) > K(C,R)$, où $K(C,R) > 0$ dépend du canal et du taux de transmission mais est indépendant de la longueur du code.*

Exposant d'erreur

On peut même montrer une version plus forte du théorème de Shannon : lorsque n croit, pour la « plupart » des codes en bloc de longueur n et de taux de transmission R, le taux d'erreur dans un canal de capacité C s'écrit

$$\sup_x P_e(\mathcal{C}, x) \approx \mathbf{e}^{-nE(R)}$$

où $E(R)$ est appelé l'*exposant d'erreur*. Il dépend du canal et du taux de transmission, mais pas de n et vérifie

$$E(R) > 0 \text{ si } R < C$$

Les séquences simultanément typiques

Définition [Ensemble simultanément typique] Soient $(X^{(n)}, Y^{(n)})$ un couple de v.a. prenant ses valeurs dans un ensemble discret $\mathcal{A}^n \times \mathcal{B}^n$, $p(\mathbf{x}, \mathbf{y})$ la loi conjointe de $(X^{(n)}, Y^{(n)})$, $p(\mathbf{x})$ et $p(\mathbf{y})$ les lois de $X^{(n)}$ et $Y^{(n)}$ respectivement. L'ensemble des séquences simultanément typiques $\mathcal{T}_\epsilon^{(n)}$ est donné par

$$\begin{aligned}
\mathcal{T}_\epsilon^{(n)} \quad = \quad & \{(\mathbf{x}, \mathbf{y}) \in \mathcal{A}^n \times \mathcal{B}^n : \\
& \left| \frac{1}{n} \left(-\log_2 p(\mathbf{x}) - H(X^{(n)}) \right) \right| < \epsilon \qquad (1) \\
& \left| \frac{1}{n} \left(-\log_2 p(\mathbf{y}) - H(Y^{(n)}) \right) \right| < \epsilon \qquad (2) \\
& \left| \frac{1}{n} \left(-\log_2 p(\mathbf{x}, \mathbf{y}) - H(X^{(n)}, Y^{(n)}) \right) \right| < \epsilon \Bigg\} \qquad (3)
\end{aligned}$$

Théorème 3. *Soient (X_i, Y_i) une suite de couples de v.a. i.i.d à valeurs dans $\mathcal{A} \times \mathcal{B}$ de même loi que (X, Y). On définit $\mathcal{T}_\epsilon^{(n)}$ à partir de $(X^{(n)}, Y^{(n)})$ avec $X^{(n)} \stackrel{\text{def}}{=} (X_1, X_2, \ldots, X_n)$ et $Y^{(n)} \stackrel{\text{def}}{=} (Y_1, Y_2, \ldots, Y_n)$. Alors*

1. $\mathbf{Prob}(\mathcal{T}_\epsilon^{(n)}) > 1 - \epsilon$ *pour n suffisamment grand.*
2. $|\mathcal{T}_\epsilon^{(n)}| \leq 2^{n(H(X,Y)+\epsilon)}$.
3. *Soient $\left(\tilde{X}^{(n)}, \tilde{Y}^{(n)}\right)$ un couple de variables aléatoires vectorielles indépendantes avec $\tilde{X}^{(n)} \sim X^{(n)}$ et $\tilde{Y}^{(n)} \sim Y^{(n)}$. Alors,*

$$\mathbf{Prob}\left\{\left(\tilde{X}^{(n)}, \tilde{Y}^{(n)}\right) \in \mathcal{T}_\epsilon^{(n)}\right\} \leq 2^{-n(I(X;Y)-3\epsilon)}.$$

Par ailleurs, pour n suffisamment grand

$$\mathbf{Prob}\left\{\left(\tilde{X}^{(n)}, \tilde{Y}^{(n)}\right) \in \mathcal{T}_\epsilon^{(n)}\right\} \geq (1-\epsilon)2^{-n(I(X;Y)+3\epsilon)}.$$

Preuve du point 3.

$$
\begin{aligned}
p\left\{\left(\tilde{X}^{(n)},\tilde{Y}^{(n)}\right)\in\mathcal{T}_{\epsilon}^{(n)}\right\} &= \sum_{(\mathbf{x},\mathbf{y})\in\mathcal{T}_{\epsilon}^{(n)}} p(\mathbf{x})p(\mathbf{y}) \\
&\leq |\mathcal{T}_{\epsilon}^{(n)}|2^{-n(H(X)-\epsilon)}2^{-n(H(Y)-\epsilon)} \\
&\leq 2^{(nH(X,Y)+\epsilon)}2^{-n(H(X)-\epsilon)}2^{-n(H(Y)-\epsilon)} \\
&= 2^{-n(I(X;Y)-3\epsilon)}.
\end{aligned}
$$

Preuve du point 3. (II)

$$
\begin{aligned}
p\left\{\left(\tilde{X}^{(n)},\tilde{Y}^{(n)}\right)\in\mathcal{T}_{\epsilon}^{(n)}\right\} &= \sum_{(\mathbf{x},\mathbf{y})\in\mathcal{T}_{\epsilon}^{(n)}} p(\mathbf{x})p(\mathbf{y}) \\
&\geq |\mathcal{T}_{\epsilon}^{(n)}|2^{-n(H(X)+\epsilon)}2^{-n(H(Y)+\epsilon)} \\
&\geq (1-\epsilon)2^{n(H(X,Y)-\epsilon)}2^{-n(H(X)+\epsilon)}2^{-n(H(Y)+\epsilon)} \\
&= (1-\epsilon)2^{-n(I(X;Y)+3\epsilon)}.
\end{aligned}
$$

La partie directe du théorème de Shannon

Argument essentiel : choix aléatoire du code !

On commence par choisir une probabilité $\mathbf{P}$ sur l'alphabet d'entrée $\mathcal{A}$ du canal. On choisit ensuite un code de longueur n de rendement R (on suppose ici que Rn est entier) en choisissant aléatoirement (sans remise) 2^{nR} mots de $\mathcal{A}^n$ au hasard suivant la distribution $\mathbf{P}^{(n)}$ sur $\mathcal{A}^n$ donnée par

$$\mathbf{P}^{(n)}(x_1, x_2, \ldots, x_n) = \Pi_{i=1}^{n} \mathbf{P}(x_i).$$

Décodage par ensemble typique

$\mathbf{x}$ mot transmis et $\mathbf{y}$ le mot reçu. On désigne par $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^{2^{nR}}$ les 2^{nR} mots de code.

1. calculer les 2^{nR} probabilités $\mathbf{P}(\text{mot reçu} = \mathbf{y}, \text{ mot émis} = \mathbf{x}^s)$ pour $s \in \{1, \ldots, 2^{nR}\}$.
2. si plus d'un couple ou moins d'un $(\mathbf{x}^i, \mathbf{y})$ est ϵ-simultanément typique $\rightarrow \ll$ échec au décodage $\gg$.
3. Sinon sortir le seul $\mathbf{x}_s$ t.q. $(\mathbf{x}_s, \mathbf{y})$ soit simultanément typique.

Analyse du décodeur

Ce décodeur peut donc échouer pour deux raisons :
- le "vrai couple" $(\mathbf{x}, \mathbf{y})$ n'est pas simultanément typique (évènement $\mathcal{E}_0$),
- il existe au moins un des $2^{nR} - 1$ couples $(\mathbf{x}^s, \mathbf{y})$ qui soit simultanément typique pour $\mathbf{x}^s \neq \mathbf{x}$, (évènement $\mathcal{E}_1$).

$$\begin{aligned} \mathbf{Prob}(\text{erreur au décodage}) &= \mathbf{Prob}(\mathcal{E}_0 \cup \mathcal{E}_1) \\ &\leq \mathbf{Prob}(\mathcal{E}_0) + \mathbf{Prob}(\mathcal{E}_1) \end{aligned}$$

La probabilité d'une erreur de type 1

$$\begin{aligned}\mathbf{Prob}(\mathcal{E}_0) &= \mathbf{Prob}\left((X^{(n)}, Y^{(n)}) \text{ n'est pas typique}\right) \\ &= 1 - \mathbf{Prob}(\mathcal{T}_\epsilon^{(n)})\end{aligned}$$

En utilisant le point 1. du théorème 3, nous avons que $1 - \mathbf{Prob}(\mathcal{T}_\epsilon^{(n)}) \leq \epsilon$.

La probabilité d'une erreur de type 2

$$\begin{aligned}\mathbf{Prob}(\mathcal{E}_1) &= \mathbf{Prob}\left(\cup_{s:\mathbf{x}^s\neq\mathbf{x}}\{(\mathbf{x}^s,\mathbf{y}) \text{ est typique}\}\right)\\ &\leq \sum_{s:\mathbf{x}^s\neq\mathbf{x}} \mathbf{Prob}\left((\mathbf{x}^s,\mathbf{y}) \text{ est typique}\right)\end{aligned}$$

$(\tilde{X}^{(n)},\tilde{Y}^{(n)})$ où $\tilde{X}^{(n)}\sim X^{(n)}$ $\tilde{Y}^{(n)}\sim X^{(n)}$ et $(\tilde{X}^{(n)},\ \tilde{Y}^{(n)})$ indépendants

$$\mathbf{Prob}\left((\mathbf{x}^s,\mathbf{y}) \text{ est typique}\right) = \mathbf{Prob}\left((\tilde{X}^{(n)},\tilde{Y}^{(n)}) \text{ est typique}\right)$$

La probabilité d'une erreur de type 2 (II)

$$\mathbf{Prob}\left\{\left(\tilde{X}^{(n)},\tilde{Y}^{(n)}\right) \text{ est typique}\right\} \leq 2^{-n(I(X;Y)-3\epsilon)}.$$

Par conséquent

$$\begin{aligned}\mathbf{Prob}(\mathcal{E}_1) &\leq (2^{nR}-1)2^{-n(I(X;Y)-3\epsilon)} \\ &\leq 2^{-n(I(X;Y)-R-3\epsilon)}.\end{aligned}$$

Fin de la preuve

$$\mathbf{Prob}(\text{erreur au décodage}) \leq \epsilon + 2^{-n(I(X;Y)-R-3\epsilon)}.$$

Fin : choisir X t.q. $C = I(X;Y)$.