Théorie de l'information : une très rapide introduction

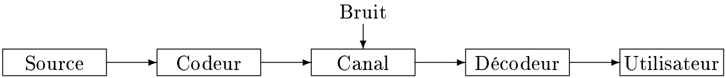
C. Poulliat

16 novembre 2020

Plan

Modèle de communication

Paradigme de Shannon : communications point-à-point



- source : voix, données capteurs, vidéos, etc,
- codeur : tout traitement effectué sur la source avant transmission,
- canal : vecteur de transport de l'information (liaison téléphonique ou satellite, canal magnétique etc) soumis à des perturbations (bruit, interférence),
- décodeur : traitement pour reproduire le plus fidèlement possible la source.

Modèle de communication

Paradigme de Shannon : communications point-à-point

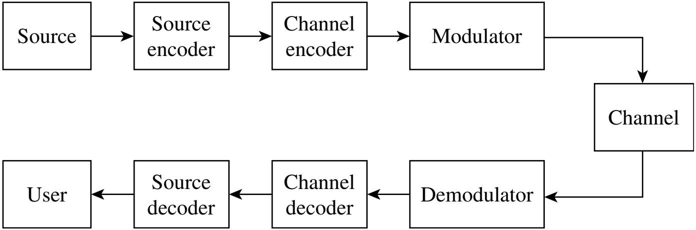
But

- Définir un cadre théorique qui permette d'analyser les performances d'un système de communication.
- Queles fonctions de traitement à l'émission et à la réception ?
- Comment les mettre en oeuvre ?

$\Rightarrow$ Shannon a montré que l'on pouvait séparer les problèmes de mise en forme de la source puis de transmission sur le canal,
$\Rightarrow$ Deux grandes problématiques : compression et codage de canal.

Modèle de communication

Séparation des traitements



Modèle de communication

Séparation des traitements

- source : suite de symboles d'un alphabet donné,
- codeur de source : représenter la source par une séquence binaire,
  $\Rightarrow$ combien de bit par symbols au minimum pour représenter fidèlement ma source (avec ou sans perte)
  $\Rightarrow$ théorie du codage de source avec ou sans perte
- codeur et décodeur de canal : introduire de la redondance pour pouvoir corriger des erreurs introduites par le canal et restituer le plus fidèlement,
  $\Rightarrow$ comment introduire de la redondance ? Quelle quantité ? Comment décoder ?
  $\Rightarrow$ théorie du codage de de canal
- décodeur source : restituer la source.

Notion de Source

Source d'information et stationnarité

- Une *source d'information* est représentée par une suite de symboles $(X_0, X_1, \cdots, X_n, \cdots)$.
  Chaque X_i est une variable aléatoire et donc $\{X_n\}_{n\in\mathbb{N}}$ définit un processus aléatoire. On suppose donc défini

$$p(X_0 = x_0, \cdots X_1 = x_1, \cdots, X_n = x_n), \ \forall n.$$

- Une source d'information est dîte *stationnaire* si toute distribution de probabilité de symboles est invariante par translation temporelle.

$$\forall n, \forall l, \forall (x_0, x_1, \cdots, x_n) \in \mathcal{X}^n,$$

$$p(X_0 = x_0, \cdots, X_n = x_n) = p(X_{0+l} = x_0, \cdots, X_{n+l} = x_n)$$

Source *sans mémoire*

Une stationnaire est dîte sans mémoire si les X_n sont produits de manière indépendante des symboles source passés, ie. $X_i, \ \forall i = 0 \cdots n-1$. On a ainsi

$$p(X_n | X_{n-1} \cdots X_0) = p(X_n)$$

On en déduit

$$p(X_0, \cdots, X_n) = p(X_0) \cdots p(X_n)$$

On a donc les symboles X_i indépendants et par stationnarité, les X_i sont identiquement distribués.

Source sans mémoire

Une source sans mémoire est définie par une suite de symboles $X_0, X_1, \cdots X_n$ indépendants et identiquement distribués (i.i.d.).

Source sans mémoire

Source discrète sans mémoire

X une variable aléatoire discrète à valeurs dans l'alphabet $\mathcal{X}$ de d.d.p. discrète $p(x) = Prob(X = x),\ x \in \mathcal{X}$. Une source discrète sans mémoire produit une séquence de symboles i.i.d à valeurs dans $\mathcal{X}$ et suivant $p(x)$.

Source continue sans mémoire

X une variable aléatoire pouvant prendre des valeurs réelles dans un intervalle $\mathcal{X} \subset \mathbb{R}$ de d.d.p. $f(X = x)$. Une source (absolument) continue sans mémoire produit une séquence de symboles i.i.d à valeurs dans $\mathcal{X}$ et suivant la d.d.p. $p(x)$.

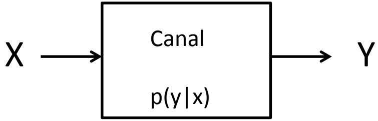
Exemples

symboles M-aires,

échantillons Gaussien

Notion de canal

Canal discret sans mémoire



- pour un canal discret sans mémoire, $p(y|x)$ est la probabilité de recevoir $Y = y$ sachant que l'on a émis $X = x$. $p(y|x)$ est appelée probabilité de transition du canal.
- si **X** est une v.a. M-aire et **Y** une v.a. N-aire, on appelle matrice de transition du canal la matrice de taille $N \times M$ donnée par

$$\Pi = (p(y|x))_{y,x}$$

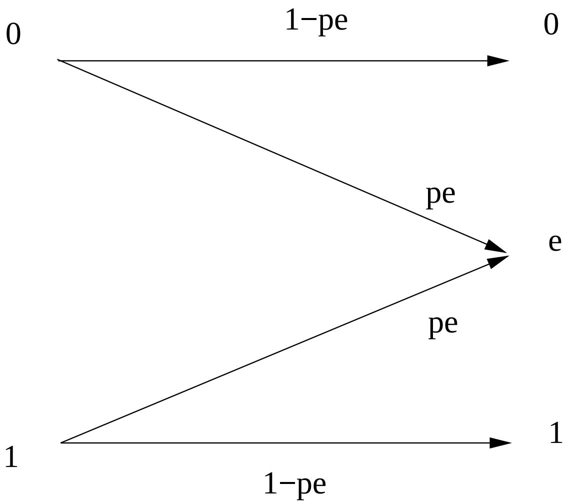
- la somme de chaque colonne de Π vaut 1 et on a

$$p_Y = \Pi p_X$$

avec $p_Y = (p(y))_y$ et $p_X = (p(x))_x$ vecteurs colonnes

Canal discret sans mémoire

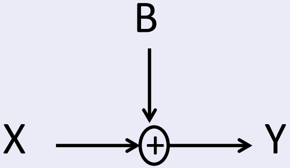
Exemple : le canal à effacement



Notion de Canal

Canal à bruit additif

canal additif



$$p(y|x) = p(z = y - x|x) = p_Z(y - x)$$

Z est une variable aléatoire indépendante de X

Canal continu sans mémoire

Canal à bruit additif Gaussien

canal additif Gaussien

- $Y = X + Z$ où X v.a. continue ou discrète à valeurs réelles et Z est un bruit blanc additif Gaussien de variance σ_Z^2,
- densité de transition :

$$p(y|x) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{(y-x)^2}{2\sigma_Z^2}}$$

Notion d'information propre

soit X une variable aléatoire discrète et $X = x$ un événement de probabilité $p(x)$

- une mesure de l'information $h(x)$ s'identifie à une mesure de l'inattendu, de l'improbable
  $\Rightarrow$ une information apportée par la réalisation de l'événement X sera d'autant plus importante que celle-ci est peu probable.
  $\Rightarrow$
  $$h(x) = f(\frac{1}{p(x)})$$
- Propriétés attendues :
  - $f(.)$ est une fonction croissante de $p(x)$,
  - $f(p) = 0$ quand $p \to 1$ (événement certain)
  - $f(p.q) = f(p) + f(q)$ (additivité de l'information pour des événements indépendants : $h(x \text{ et } y) = h(x) + h(y)$)

Notion d'information propre

Canal à bruit additif Gaussien

Unicité de l'information

La fonction $f(p) = -log(p)$ est la seule fonction qui soit à la fois positive, continue sur $[0, 1)$ et qui vérifie l'additivité des informations indépendantes. La base du logarithme est indifférente.

Définition (Self-information)

Soit X une variable aléatoire discrète et $X = x$ un événement de probabilité $p(x)$, on appelle information propre ou quantité d'information apportée par l'événement x, la quantité

$$h(x) = \log\left(\frac{1}{p(x)}\right) = -\log\left(p(x)\right)$$

Notion d'information propre

Canal à bruit additif Gaussien

Propriétés

- positivité : $h(x) \geq 0$
- additivité : soient x et y deux événements indépendants, alors

$$h(x \text{ et } y) = h(x) + h(y)$$

- unités :
  - $\log_e$: natural units, *nats*, Shannon,
  - $\log_2$: binary units, *bits*,

Exemple

- soit une source binaire $\{0, 1\}$ equidistribuée de symboles indépendants, l'information propre associée à chaque symbole binaire est $h(1/2) = 1$ bit ou Sh.
- soit une source M-aire $\{0, 1 \cdots, M-1\}$ equidistribué de symboles indépendants, l'information propre associée à chaque symbole est $h(1/M) = \log_2(M)$ bit ou Sh.

Entropie

X une variable aléatoire discrète à valeurs dans l'alphabet $\mathcal{X}$ de d.d.p. $p(x) = Prob(X = x), \ x \in \mathcal{X}$

Entropie

$$
\begin{aligned}
\mathbf{H}(X) &= -\sum_{x \in \mathcal{X}} p(X = x) \log_2 (p(X = x)) \\
&= -\mathbb{E}(\log_2 p(X)) \qquad (1)
\end{aligned}
$$

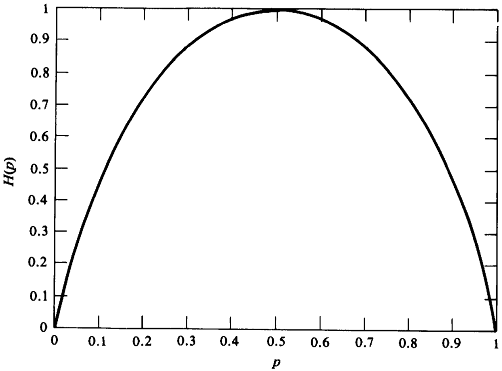
$\Rightarrow$ quantité d'information moyenne en bits/symbole

Propriétés

- $\mathbf{H}(X)$ est déterministe et c'est une fonction de $p(x)$,
- $\mathbf{H}(X) \geq 0$,
- $\mathbf{H}(X) = 0 \Leftrightarrow X$ est déterministe ,
- $\mathbf{H}(X) = \log_2 (M)$ pour distribution uniforme de symboles M-aire,
- invariance par équivalence ($Y = f(X)$ où $f(.)$ inversible).

Entropie

Entropie binaire



$$X \in \{0, 1\} \text{ avec } p(X = 1) = p$$

$$\mathbf{H}(X) = -p \log_2 (p) - (1 - p) \log_2 (1 - p) \triangleq H_2(p)$$

Entropie

Inégalité de Gibbs

Etant données 2 distributions de probabilité discrètes $(p_1, \cdots, p_n)$ et $(q_1, \cdots, q_n)$ sur un alphabet de même taille, alors on a l'inégalité suivante :

$$\sum_{i=1}^{n} p_i \log\left(\frac{q_i}{p_i}\right) \leq 0$$

avec égalité pour $p_i = q_i \; \forall i$.
Preuve : utiliser $\ln(x) \leq (x-1)$ avec $x = q_i/p_i$

Maximum d'entropie

L'entropie d'une source M-aire vérifie

$$\mathbf{H}(X) \leq \log_2(M)$$

avec égalité pour source uniforme Preuve : utiliser $q_i = \frac{1}{M}$

Entropie conjointe

Entropie conjointe

soient X et Y deux variables aléatoires discrètes

$$\begin{aligned}\mathbf{H}(X,Y) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(X=x,Y=y)\log_2\left(p(X=x,Y=y)\right) \\ &= -\mathbb{E}(\log_2 p(X,Y)) \qquad (2)\end{aligned}$$

Propriété

$$\mathbf{H}(X,Y) = \mathbf{H}(Y,X)$$

Entropie conditionnelle

soient X et Y deux variables aléatoires discrètes

Entropie de Y sachant $X = x$

$$\begin{aligned}\mathbf{H}(Y|X=x) &= -\sum_{y\in\mathcal{Y}} p(Y=y|X=x)\log_2\left(p(Y=y|X=x)\right) \\ &= -\mathbb{E}(\log_2 p(Y|X=x)) \qquad (3)\end{aligned}$$

Entropie conditionnelle

$$\begin{aligned}\mathbf{H}(Y|X) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(X=x,Y=y)\log_2\left(p(Y=y|X=x)\right) \\ &= \sum_{x\in\mathcal{X}} p(X=x)\mathbf{H}(Y|X=x) = \mathbb{E}(\mathbf{H}(Y|X=x)) \\ &= -\mathbb{E}(\log_2 p(Y|X)) \qquad (4)\end{aligned}$$

Entropie conditionnelle et conjointe

Propriétés

1. chain rule : $\mathbf{H}(X,Y) = \mathbf{H}(X) + \mathbf{H}(Y|X) = \mathbf{H}(Y) + \mathbf{H}(X|Y)$,
2. borne inf. : $\mathbf{H}(X,Y) \geq \mathbf{H}(X)$ ou $\mathbf{H}(Y)$
3. Conditionnement : $\mathbf{H}(X|Y) \leq \mathbf{H}(X)$
   égalité si X et Y indépendants
4. Décroissance par conditionnement :
   $\mathbf{H}(X_1|X_2,\cdots,X_n) \leq \cdots \leq \mathbf{H}(X_1|X_2,X_3) \leq \mathbf{H}(X_1|X_2) \leq \mathbf{H}(X_1)$
5. Encadrement (sous additivité de l'entropie) :
   $$\mathbf{H}(X,Y) \leq \mathbf{H}(X) + \mathbf{H}(Y) \leq 2\mathbf{H}(X,Y)$$
6. Entropie conjointe et conditionnement :
   $$\mathbf{H}(X,Y|Z) = \mathbf{H}(X|Z) + \mathbf{H}(Y|X,Z)$$
7. positivité : $\mathbf{H}(X|Y) \geq 0$
   égalité si $X = f(Y)$ où $f(.)$ déterministe

Entropie conditionnelle et conjointe ; cas générale

Entropie conjointe : borne de l'indépendance

Soit $X_1, X_2, \cdots X_n$ de loi conjointe $p(x_1, x_2, \cdots, x_n)$

$$\mathbf{H}(X_1, X_2, \cdots X_n) \leq \sum_{i=1}^{n} \mathbf{H}(X_i)$$

égalité si et seulement si les X_i sont indépendants

Chain rule : cas générale

Soit $X_1, X_2, \cdots X_n$ de loi conjointe $p(x_1, x_2, \cdots, x_n)$

$$\mathbf{H}(X_1, X_2, \cdots X_n) = \sum_{i=1}^{n} \mathbf{H}(X_i | X_{i-1}, \cdots X_1)$$

Entropie d'une variable aléatoire continue

Entropie différentielle

X une variable aléatoire continue définie par une densité de probabilité $f(x)$

Entropie différentielle

$$h(X) = -\int f(x) \log_2(f(x))\, dx. \tag{2}$$

NB : on ne peut pas interpréter $h(X)$ comme une mesure d'information ou d'incertitude dans le cas continue.

Changement de variable

soit $Y = f(X)$, par changement de variable, on a $h(X) \neq h(Y) = h(X)$ donc $h(X)$ n'est pas une information

Changement d'échelle

soit $Y = aX$, on a $h(X) \neq h(aX) = h(X) + \log(a)$ qui peut être négatif !

Entropie d'une variable aléatoire continue

Exemples

Loi uniforme sur $[a, b]$

$$f(x) = \frac{1}{b-a}$$

$$h(x) = \log(b-a)$$

Loi normale de moyenne μ et variance σ^2

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$h(x) = \frac{1}{2}\log 2\pi e + \log(\sigma)$$

Entropie de variables aléatoires continues

Entropie différentielle conjointe et conditionnelle

Entropie différentielle conjointe

Soit $X_1, X_2, \cdots, X_n$ avec $f(x_1, x_2, \cdots, x_n)$

$$h(X_1, X_2, \cdots, X_n) = -\int f(x_1, x_2, \cdots, x_n) \log_2(f(x_1, x_2, \cdots, x_n))\, dx_1\, dx_2 \cdots dx$$

Entropie différentielle conditionnelle

$$\mathbf{h}(X|Y) = -\int f(x,y) \log_2 (f(x|y))\, dxdy \qquad (7)$$

$$(8)$$

Information mutuelle

Information mutuelle

$$\begin{aligned}\mathbf{I}(X;Y) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(X=x,Y=y)\log_2\left(\frac{p(X=x)p(Y=y)}{p(X=x,Y=y)}\right) \\ &= -\mathbb{E}(\log_2\left(\frac{p(X)p(Y)}{p(X,Y)}\right)) \qquad (9)\end{aligned}$$

Propriétés

1. Positivité : $\mathbf{I}(X;Y) \geq 0$
2. Borne sup. : $\mathbf{I}(X;Y) \leq \min(\mathbf{H}(X), \mathbf{H}(Y))$
3. Symétrie : $\mathbf{I}(X;Y) = \mathbf{I}(Y;X)$
4. Information propre : $\mathbf{I}(X;X) = \mathbf{H}(X)$

Information mutuelle

Propriétés

1. Lien avec entropie et entropie conditionnelle :

$$\mathbf{I}(X;Y) = \mathbf{H}(X) - \mathbf{H}(X|Y) = \mathbf{H}(Y) - \mathbf{H}(Y|X)$$

2. Lien avec entropie et entropie conjointe :

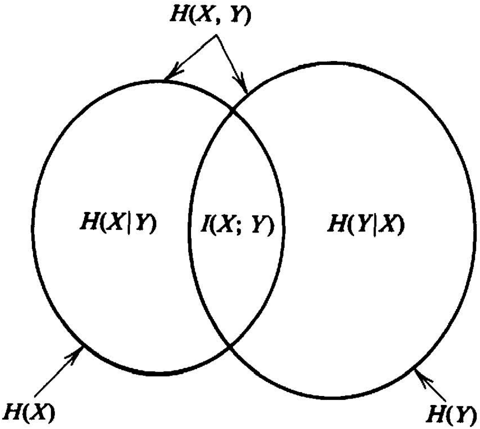
$$\mathbf{I}(X;Y) = \mathbf{H}(X) + \mathbf{H}(Y) - \mathbf{H}(X,Y)$$

3. Conditionnement : $\mathbf{I}(X;Y|Z) \triangleq \mathbf{H}(X|Z) - \mathbf{H}(X|Y,Z) \geq 0$
4. Chain rule :

$$\mathbf{I}(X_1, X_2, \cdots, X_n; Y) = \sum_{i=1}^{n} \mathbf{I}(X_i; Y|X_{i-1}, \cdots X_1)$$

Information mutuelle

Interprétation



Information mutuelle

Interprétation

Information mutuelle, cas continu

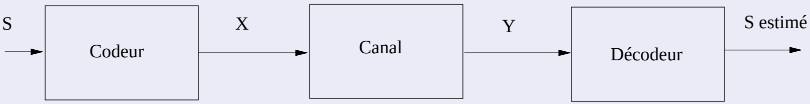
$$I(X;Y) = -\int f(x,y)\log_2\left(\frac{f(x)f(y)}{f(x,y)}\right)dxdy \geq 0 \quad (10)$$

Entropie différentielle conditionnelle

$$\begin{aligned} \mathbf{h}(X|Y) &= -\int f(x,y)\log_2\left(f(x|y)\right)dxdy && (11) \\ &= h(X,Y) - h(Y) && (12) \end{aligned}$$

Capacité d'un canal discret sans mémoire

Définition



- $X \in \mathcal{X}$, $Y \in \mathcal{Y}$
- Canal sans mémoire caractérisé par $p(Y|X)$

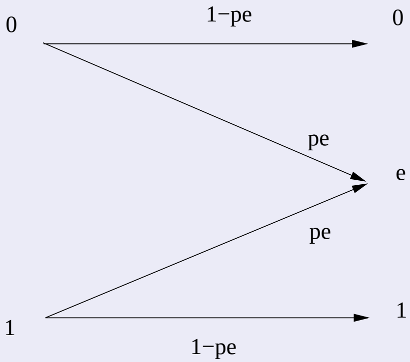
Définition

$$\begin{aligned} \mathbf{C} &= \max_{p(X)} \mathbf{I}(X;Y) \\ &= \max_{p(X)} \mathbf{H}(X) - \mathbf{H}(X|Y) = \max_{p(X)} \mathbf{H}(Y) - \mathbf{H}(Y|X) \end{aligned} \tag{5}$$

Max. atteint pour distribution uniforme pour les canaux *symétriques*.

Capacité d'un canal discret sans mémoire

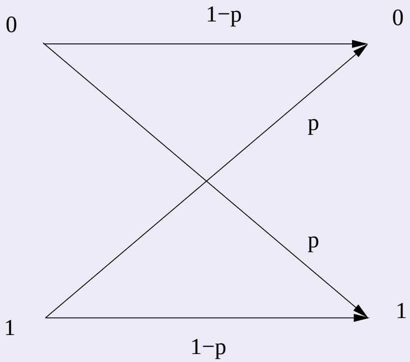
Canal à effacement (BEC)



$$\mathbf{C} = 1 - p_e \tag{6}$$

atteint pour une distribution d'entrée uniforme

Canal binaire symétrique (BSC)



$$\mathbf{C} \quad = \quad 1 - H_2(p) \tag{7}$$

atteint pour une distribution d'entrée uniforme

Théorème du codage de canal

Canal discret sans mémoire

Théorème du codage de canal (1/2)

Soit un canal discret sans mémoire de capacité **C**, on peut communiquer pour tout débit de transmission inférieur à C. En particulier, $\forall R < C$, il existe une séquence de codes $(2^{nR}, n)$ telle que la probabilité d'erreur bloc en sortie de décodage optimal soit arbitrairement petite pour n suffisamment grand.

Théorème du codage de canal (2/2)

Canal à temps discret et entrées/sorties continues

Théorème du codage de canal

- Extension au cas d'entrées ou de sorties continues.
- Les expressions précédentes mettent en jeu des densités de probabilités.
- Application principale : le cas du canal Gaussien.

Canal additif gaussien(AWGN)

$$\begin{array}{ccccc} & & B(\omega) & & \\ & & \downarrow & & \\ X(\omega) & \longrightarrow & \oplus & \longrightarrow & Y(\omega) \end{array} \tag{8}$$

- $X(\omega)$ tel que $\sigma_x^2 \leq P$
- $B(\omega) \sim \mathcal{N}(0, \sigma_b^2)$
-

$$\mathbf{C} = \frac{1}{2}\log_2\left(1+\sigma_x^2/\sigma_b^2\right) \text{ bits/symbol} \tag{9}$$

$$= \frac{1}{2}\log_2\left(1+2R_bE_b/N_0\right) \tag{10}$$

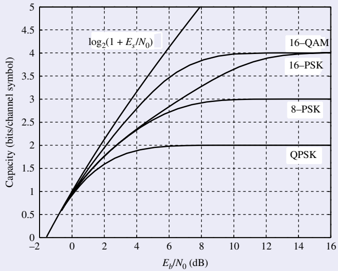
max. atteint pour $X(\omega) \sim \mathcal{N}(0, \sigma_x^2 = P)$

Canal additif gaussien à entrées binaires (BI-AWGN)

$$\begin{array}{ccccc} & & B(\omega) & & \\ & & \downarrow & & \\ X(\omega) \longrightarrow & & \bigoplus & & \longrightarrow Y(\omega) \end{array} \tag{11}$$

- $X(\omega) \in \mathcal{X} = \{0, 1\}$, avec $p(X = 1) = 1/2$
- $B(\omega) \sim \mathcal{N}(0, \sigma_b^2 = N_0/2)$
- canal symétrique : $p(y|x = +1) = p(-y|x = -1)$

Canal additif gaussien à entrées M-aire (CI-AWGN)



$$
\begin{array}{ccc}
 & B(\omega) & \\
 & \downarrow & \\
X(\omega) \longrightarrow & \bigoplus & \longrightarrow Y(\omega)
\end{array}
$$

- $X(\omega) \in \mathcal{X} = \{0, \ldots, M\}$, avec $p(X = x) = 1/M$
- $B(\omega) \sim \mathcal{CN}(0, \sigma_b^2 = N_0)$

Canal additif gaussien à entrées M-aire (CM-AWGN)

Comment calculer cette capacité pour une modulation donnée ? (1/2)

$$\begin{array}{ccccc} & & B(\omega) & & \\ & & \downarrow & & \\ X(\omega) & \longrightarrow & \bigoplus & \longrightarrow & Y(\omega) \end{array}$$

- $X(\omega) \in \mathcal{X} = \{0, \ldots, M\}$, avec $p(X = x) = 1/M$
- $B(\omega) \sim \mathcal{CN}(0, \sigma_b^2 = N_0)$

- Calculer la capacité revient à évaluer

$$\mathbf{C} = \mathbf{I}(X; Y)$$

- Termes à calculer : $\mathbf{H}(X)$, $\mathbf{H}(X|Y)$.

Canal additif gaussien à entrées M-aire (CM-AWGN)

Comment calculer cette capacité pour une modulation donnée ? (2/2)

- Calcul de $\mathbf{H}(X)$: $\mathbf{H}(X) = log2(M)$
- Calcul de $\mathbf{H}(X|Y)$:

$$\mathbf{H}(X|Y) = -\mathbb{E}(\log_2 p(X|Y)) = \mathbb{E}(h(X|Y))$$

or

$$p(X|Y) = \frac{p(Y|X)}{\sum_{x \in \mathcal{X}} p(Y|X = x)}$$

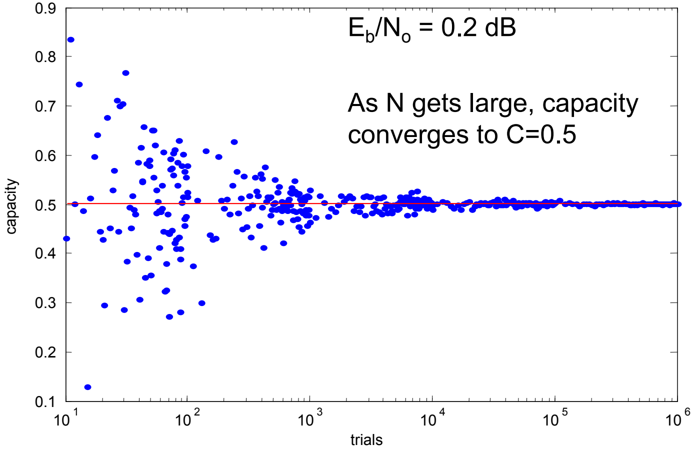
- Estimateur de $\mathbf{H}(X|Y)$: par ergodicité, on a

$$\mathbf{H}(X|Y) = \mathbb{E}(h(X|Y)) = \lim_{N \to +\infty} \frac{1}{N} \sum_n h(X(n)|Y(n))$$

- Méthode par Monte-Carlo :
  1. Tirer aléatoirement et uniformément des symboles issue de constellation,
  2. Calculer pour chaque couple $(X(n), Y(n))$, $h(X(n)|Y(n))$ et moyenner.

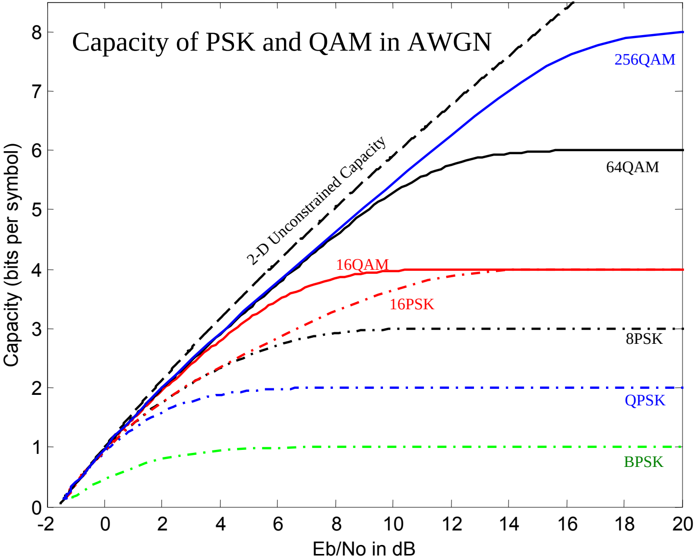
Canal additif gaussien à entrées M-aire (CM-AWGN)

Influence du nombre d'échantillons

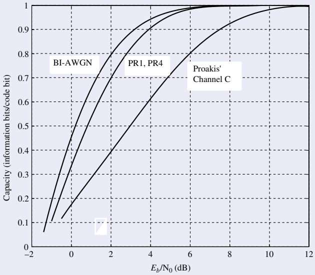


Canal additif gaussien à entrées M-aire (CM-AWGN)

Exemples



Canaux sélectif en fréquence à entrées binaires(BI-ISI)



$$y[n] = \sum_{k=0}^{L-1} h[k]x[n-k] + b[n] \qquad (12)$$

- $X \in \mathcal{X} = \{-1, +1\}$, avec $p(X = x) = 1/2$
- $B \sim \mathcal{N}(0, \sigma_b^2 = N_0/2)$

Comparer l'efficacité des systèmes de codage grâce à la capacité.