Cours 5. Introduction au codage de canal

A. Désilles

21 avril 2010

Résumé

1. Rappels
2. Représentation mathématique d'un canal de transmission
   - Modèle probabiliste
   - Distribution conjointe
   - Distributions marginales
   - Distributions conditionnelles
3. Entropies associées à un couple émetteur-récepteur
   - Entropie conjointe
   - Entropie conditionnelle moyenne
   - Information mutuelle moyenne
   - Capacité d'un canal
   - Canal symétrique
4. Codage de canal : codage avec bruit
   - Règle de décodage d'un canal avec bruit.
   - Analyse de l'erreur de transmission
   - Notion de code de canal.
   - Second théorème de Shannon

Résumé

1. Rappels
2. Représentation mathématique d'un canal de transmission
   - Modèle probabiliste
   - Distribution conjointe
   - Distributions marginales
   - Distributions conditionnelles
3. Entropies associées à un couple émetteur-récepteur
   - Entropie conjointe
   - Entropie conditionnelle moyenne
   - Information mutuelle moyenne
   - Capacité d'un canal
   - Canal symétrique
4. Codage de canal : codage avec bruit
   - Règle de décodage d'un canal avec bruit.
   - Analyse de l'erreur de transmission
   - Notion de code de canal.
   - Second théorème de Shannon

Résumé

1. Rappels
2. Représentation mathématique d'un canal de transmission
   - Modèle probabiliste
   - Distribution conjointe
   - Distributions marginales
   - Distributions conditionnelles
3. Entropies associées à un couple émetteur-récepteur
   - Entropie conjointe
   - Entropie conditionnelle moyenne
   - Information mutuelle moyenne
   - Capacité d'un canal
   - Canal symétrique
4. Codage de canal : codage avec bruit
   - Règle de décodage d'un canal avec bruit.
   - Analyse de l'erreur de transmission
   - Notion de code de canal.
   - Second théorème de Shannon

Résumé

1. Rappels
2. Représentation mathématique d'un canal de transmission
   - Modèle probabiliste
   - Distribution conjointe
   - Distributions marginales
   - Distributions conditionnelles
3. Entropies associées à un couple émetteur-récepteur
   - Entropie conjointe
   - Entropie conditionnelle moyenne
   - Information mutuelle moyenne
   - Capacité d'un canal
   - Canal symétrique
4. Codage de canal : codage avec bruit
   - Règle de décodage d'un canal avec bruit.
   - Analyse de l'erreur de transmission
   - Notion de code de canal.
   - Second théorème de Shannon

Résumé

Modèle de communication

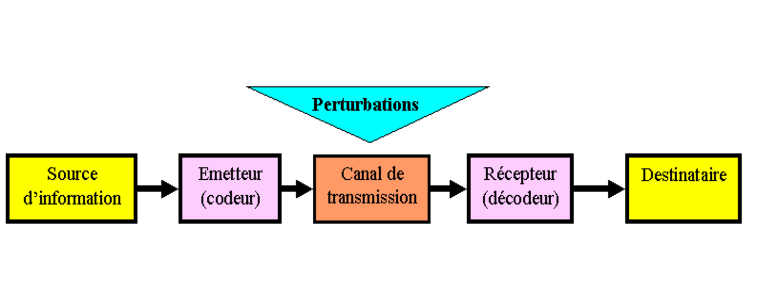


Figure: Paradigme de Shannon

Un canal

- Selon le modèle de communication de Shannon, le message émis par la source est soumis à l'entrée d'un canal de transmission ;
- Les perturbations présentes dans le canal modifient le signal initial ;
- Ces perturbations ne sont pas connues à l'avance par le destinataire ;
- Ainsi la transmission d'un message a également un caractère aléatoire.

Un canal

- Selon le modèle de communication de Shannon, le message émis par la source est soumis à l'entrée d'un canal de transmission ;
- Les perturbations présentes dans le canal modifient le signal initial ;
- Ces perturbations ne sont pas connues à l'avance par le destinataire ;
- Ainsi la transmission d'un message a également un caractère aléatoire.

Un canal

- Selon le modèle de communication de Shannon, le message émis par la source est soumis à l'entrée d'un canal de transmission ;
- Les perturbations présentes dans le canal modifient le signal initial ;
- Ces perturbations ne sont pas connues à l'avance par le destinataire ;
- Ainsi la transmission d'un message a également un caractère aléatoire.

Un canal

- Selon le modèle de communication de Shannon, le message émis par la source est soumis à l'entrée d'un canal de transmission ;
- Les perturbations présentes dans le canal modifient le signal initial ;
- Ces perturbations ne sont pas connues à l'avance par le destinataire ;
- Ainsi la transmission d'un message a également un caractère aléatoire.

Modèle probabiliste

- On associe une variable aléatoire X à la source d'information. Elle représente l'observation de symboles émis par la source.
- On associe une autre variable aléatoire Y au destinataire. Elle représente l'observation de symboles reçus.
- Le couple (X, Y) représente le processus de communication.
- Au couple de variables aléatoires (X, Y) on associe différents types de distributions de probabilité.

Modèle probabiliste

- On associe une variable aléatoire X à la source d'information. Elle représente l'observation de symboles émis par la source.
- On associe une autre variable aléatoire Y au destinataire. Elle représente l'observation de symboles reçus.
- Le couple (X, Y) représente le processus de communication.
- Au couple de variables aléatoires (X, Y) on associe différents types de distributions de probabilité.

Modèle probabiliste

- On associe une variable aléatoire X à la source d'information. Elle représente l'observation de symboles émis par la source.
- On associe une autre variable aléatoire Y au destinataire. Elle représente l'observation de symboles reçus.
- Le couple (X, Y) représente le processus de communication.
- Au couple de variables aléatoires (X, Y) on associe différents types de distributions de probabilité.

Modèle probabiliste

- On associe une variable aléatoire X à la source d'information. Elle représente l'observation de symboles émis par la source.
- On associe une autre variable aléatoire Y au destinataire. Elle représente l'observation de symboles reçus.
- Le couple (X, Y) représente le processus de communication.
- Au couple de variables aléatoires (X, Y) on associe différents types de distributions de probabilité.

Modèle probabiliste. Distribution conjointe.

Soit un couple de variables aléatoires discrètes X à valeurs dans $\{x_i\}_{i=1}^{n}$ et Y à valeurs dans $\{y_j\}_{j=1}^{m}$.
On définit **la distribution conjointe de probabilités** de manière suivante

$$P_{XY} : \{(x_i, y_j),\ (i,j) \in |[1,n]| \times |[1,m]|\} \to [0,1]$$

$$(x_i, y_j) \mapsto p(x_i, y_j) = P[X = x_i \text{ et } Y = y_j]$$

On a les propriétés suivantes :

$$\forall (i,j),\ 0 \leq p(x_i, y_j) \leq 1$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) = 1$$

Modèle probabiliste. Distribution conjointe. Exemple

Soit une source binaire X d'alphabet $\{0, 1\}$.

Supposons que à la sortie du canal on peut obtenir l'un des trois symboles : $\{0, 1, -1\}$. Notons Y la variable aléatoire associée.

Le symbole -1 signifie que le système n'a pas su déterminer le symbole à la réception.

La **la distribution conjointe de probabilités** peut s'écrire de manière suivante

$$P(X,Y) = \begin{array}{c|ccc} & 0 & 1 & \text{-1} \\ \hline 0 & 0.3 & 0.15 & 0.15 \\ 1 & 0.1 & 0.2 & 0.1 \end{array}$$

Modèle probabiliste. Distribution conjointe. Exemple

Soit une source binaire X d'alphabet $\{0, 1\}$.

Supposons que à la sortie du canal on peut obtenir l'un des trois symboles : $\{0, 1, -1\}$. Notons Y la variable aléatoire associée.

Le symbole -1 signifie que le système n'a pas su déterminer le symbole à la réception.

La **la distribution conjointe de probabilités** peut s'écrire de manière suivante

$$P(X,Y) = \begin{array}{c|ccc} & 0 & 1 & -1 \\ \hline 0 & 0.3 & 0.15 & 0.15 \\ 1 & 0.1 & 0.2 & 0.1 \end{array}$$

Modèle probabiliste. Distribution conjointe. Exemple

Soit une source binaire X d'alphabet $\{0, 1\}$.

Supposons que à la sortie du canal on peut obtenir l'un des trois symboles : $\{0, 1, -1\}$. Notons Y la variable aléatoire associée.

Le symbole -1 signifie que le système n'a pas su déterminer le symbole à la réception.

La **la distribution conjointe de probabilités** peut s'écrire de manière suivante

$$P(X, Y) = \begin{array}{c|ccc} & 0 & 1 & \text{-1} \\ \hline 0 & 0.3 & 0.15 & 0.15 \\ 1 & 0.1 & 0.2 & 0.1 \end{array}$$

Modèle probabiliste. Distribution conjointe. Exemple

Soit une source binaire X d'alphabet $\{0, 1\}$.

Supposons que à la sortie du canal on peut obtenir l'un des trois symboles : $\{0, 1, -1\}$. Notons Y la variable aléatoire associée.

Le symbole -1 signifie que le système n'a pas su déterminer le symbole à la réception.

La **la distribution conjointe de probabilités** peut s'écrire de manière suivante

$$P(X,Y) = \begin{array}{c|ccc} & 0 & 1 & \text{-1} \\ \hline 0 & 0.3 & 0.15 & 0.15 \\ 1 & 0.1 & 0.2 & 0.1 \end{array}$$

Modèle probabiliste. Distribution conjointe. Exemple

Soit une source binaire X d'alphabet $\{0, 1\}$.

Supposons que à la sortie du canal on peut obtenir l'un des trois symboles : $\{0, 1, -1\}$. Notons Y la variable aléatoire associée.

Le symbole -1 signifie que le système n'a pas su déterminer le symbole à la réception.

La **la distribution conjointe de probabilités** peut s'écrire de manière suivante

$$P(X, Y) = \begin{array}{c|ccc} & 0 & 1 & -1 \\ \hline 0 & 0.3 & 0.15 & 0.15 \\ 1 & 0.1 & 0.2 & 0.1 \end{array}$$

Modèle probabiliste. Distribution conjointe. Exemple

Soit une source binaire X d'alphabet $\{0,1\}$.

Supposons que à la sortie du canal on peut obtenir l'un des trois symboles : $\{0,1,-1\}$. Notons Y la variable aléatoire associée.

Le symbole -1 signifie que le système n'a pas su déterminer le symbole à la réception.

La **la distribution conjointe de probabilités** peut s'écrire de manière suivante

$$P(X,Y) = \begin{array}{c|ccc} & 0 & 1 & \text{-1} \\ \hline 0 & 0.3 & 0.15 P(X=0 \text{ et } Y=1) & 0.15 \\ 1 & 0.1 & 0.2 & 0.1 \end{array}$$

Distributions marginales

Étant donnée la distribution conjointe d'un couple de variables aléatoires (X, Y), $p(x_i, y_j)$, on définit les **distributions marginales** de X et de Y respectivement par les relations :

$$\begin{cases} P_X : \{x_i\}_{i=1}^{n} \to [0,1], \quad x_i \mapsto p_X(x_i) = \displaystyle\sum_{j=1}^{m} p(x_i, y_j) \\ \\ P_Y : \{x_j\}_{j=1}^{m} \to [0,1], \quad y_j \mapsto p_Y(y_j) = \displaystyle\sum_{i=1}^{n} p(x_i, y_j) \end{cases}$$

Distributions marginales

Étant donnée la distribution conjointe d'un couple de variables aléatoires (X, Y), $p(x_i, y_j)$, on définit les **distributions marginales** de X et de Y respectivement par les relations :

$$\begin{cases} P_X : \{x_i\}_{i=1}^{n} \longrightarrow [0,1], \quad x_i \mapsto p_X(x_i) = \sum_{j=1}^{m} p(x_i, y_j) \\ \\ P_Y : \{x_j\}_{j=1}^{m} \longrightarrow [0,1], \quad y_j \mapsto p_Y(y_j) = \sum_{i=1}^{n} p(x_i, y_j) \end{cases}$$

Distributions marginales

Étant donnée la distribution conjointe d'un couple de variables aléatoires (X, Y), $p(x_i, y_j)$, on définit les **distributions marginales** de X et de Y respectivement par les relations :

$$\begin{cases} P_X : \{x_i\}_{i=1}^{n} \to [0,1], \quad x_i \mapsto p_X(x_i) = \sum_{j=1}^{m} p(x_i, y_j) \\ \\ P_Y : \{x_j\}_{j=1}^{m} \to [0,1], \quad y_j \mapsto p_Y(y_j) = \sum_{i=1}^{n} p(x_i, y_j) \end{cases}$$

Distributions marginales. Exemple

Pour calculer par exemple $P[X = x_1]$ on effectue la somme de la première ligne de la matrice de distribution conjointe :

$$P[X = 0] =$$

	0	1	-1	
0	0.3	0.15	0.15	
1	0.1	0.2	0.1	

Pour calculer la distribution marginale de Y on effectue les sommes des éléments de colonnes de la matrice de distribution conjointe

Distributions marginales. Exemple

Pour calculer par exemple $P[X = x_1]$ on effectue la somme de la première ligne de la matrice de distribution conjointe :

$$P[X = 0] =$$

	0	1	-1	P_X
0	0.3	0.15	0.15	
1	0.1	0.2	0.1	

Pour calculer la distribution marginale de Y on effectue les sommes des éléments de colonnes de la matrice de distribution conjointe

Distributions marginales. Exemple

Pour calculer par exemple $P[X = x_1]$ on effectue la somme de la première ligne de la matrice de distribution conjointe :

$$P[X = 0] = p(x_1, y_1)$$

	0	1	-1	P_X
0	0.3	0.15	0.15	
1	0.1	0.2	0.1	

Pour calculer la distribution marginale de Y on effectue les sommes des éléments de colonnes de la matrice de distribution conjointe

Distributions marginales. Exemple

Pour calculer par exemple $P[X = x_1]$ on effectue la somme de la première ligne de la matrice de distribution conjointe :

$$P[X = 0] = p(x_1, y_1) + p(x_1, y_2)$$

	0	1	-1	P_X
0	0.3	0.15	0.15	
1	0.1	0.2	0.1	

Pour calculer la distribution marginale de Y on effectue les sommes des éléments de colonnes de la matrice de distribution conjointe

Distributions marginales. Exemple

Pour calculer par exemple $P[X = x_1]$ on effectue la somme de la première ligne de la matrice de distribution conjointe :

$$P[X = 0] = p(x_1, y_1) + p(x_1, y_2) + p(x_1, y_3)$$

	0	1	-1	P_X
0	0.3	0.15	0.15	
1	0.1	0.2	0.1	

Pour calculer la distribution marginale de Y on effectue les sommes des éléments de colonnes de la matrice de distribution conjointe

Distributions marginales. Exemple

Pour calculer par exemple $P[X = x_1]$ on effectue la somme de la première ligne de la matrice de distribution conjointe :

$$P[X = 0] = p(x_1, y_1) + p(x_1, y_2) + p(x_1, y_3) = 0.6$$

	0	1	-1	P_X
0	0.3	0.15	0.15	0.6
1	0.1	0.2	0.1	

Pour calculer la distribution marginale de Y on effectue les sommes des éléments de colonnes de la matrice de distribution conjointe

Distributions marginales. Exemple

Pour calculer par exemple $P[X = x_1]$ on effectue la somme de la première ligne de la matrice de distribution conjointe :

$$P[X = 0] = p(x_1, y_1) + p(x_1, y_2) + p(x_1, y_3) = 0.6$$

	0	1	-1	P_X
0	0.3	0.15	0.15	0.6
1	0.1	0.2	0.1	

Pour calculer la distribution marginale de Y on effectue les sommes des éléments de colonnes de la matrice de distribution conjointe

Distributions marginales. Exemple

Pour calculer par exemple $P[X = x_1]$ on effectue la somme de la première ligne de la matrice de distribution conjointe :

$$P[X = 0] = p(x_1, y_1) + p(x_1, y_2) + p(x_1, y_3) = 0.6$$

	0	1	-1	P_X
0	0.3	0.15	0.15	0.6
1	0.1	0.2	0.1	0.4

Pour calculer la distribution marginale de Y on effectue les sommes des éléments de colonnes de la matrice de distribution conjointe

Distributions marginales. Exemple

Pour calculer par exemple $P[X = x_1]$ on effectue la somme de la première ligne de la matrice de distribution conjointe :

$$P[X = 0] = p(x_1, y_1) + p(x_1, y_2) + p(x_1, y_3) = 0.6$$

	0	1	-1	P_X
0	0.3	0.15	0.15	0.6
1	0.1	0.2	0.1	0.4
P_Y				

Pour calculer la distribution marginale de Y on effectue les sommes des éléments de colonnes de la matrice de distribution conjointe

Distributions marginales. Exemple

Pour calculer par exemple $P[X = x_1]$ on effectue la somme de la première ligne de la matrice de distribution conjointe :

$$P[X = 0] = p(x_1, y_1) + p(x_1, y_2) + p(x_1, y_3) = 0.6$$

	0	1	-1	P_X
0	0.3	0.15	0.15	0.6
1	0.1	0.2	0.1	0.4
P_Y	0.4			

Pour calculer la distribution marginale de Y on effectue les sommes des éléments de colonnes de la matrice de distribution conjointe

$$P[Y = 0] = p(x_1, y_1) + p(x_2, y_1) = 0.4$$

Distributions marginales. Exemple

Pour calculer par exemple $P[X = x_1]$ on effectue la somme de la première ligne de la matrice de distribution conjointe :

$$P[X = 0] = p(x_1, y_1) + p(x_1, y_2) + p(x_1, y_3) = 0.6$$

	0	1	-1	P_X
0	0.3	0.15	0.15	0.6
1	0.1	0.2	0.1	0.4
P_Y	0.4	0.35		

Pour calculer la distribution marginale de Y on effectue les sommes des éléments de colonnes de la matrice de distribution conjointe

$$P[Y = -1] = p(x_1, y_1) + p(x_2, y_1) = 0.4$$

Distributions marginales. Exemple

Pour calculer par exemple $P[X = x_1]$ on effectue la somme de la première ligne de la matrice de distribution conjointe :

$$P[X = 0] = p(x_1, y_1) + p(x_1, y_2) + p(x_1, y_3) = 0.6$$

	0	1	-1	P_X
0	0.3	0.15	0.15	0.6
1	0.1	0.2	0.1	0.4
P_Y	0.4	0.35	0.25	

Pour calculer la distribution marginale de Y on effectue les sommes des éléments de colonnes de la matrice de distribution conjointe

$$P[Y = -1] = p(x_1, y_1) + p(x_2, y_1) = 0.4$$

Distributions conditionnelles

On associe à l'événement $Y = y_j$ une distribution conditionnelle $P[X|y_j]$ de X sachant y_j définie par

$$P : x_i \mapsto p(x_i|y_j) = P[X = x_i|Y = y_j] = \frac{p(x_i, y_j)}{p_Y(y_j)}$$

De façon analogue on définit la distribution conditionnelle $P[Y|x = x_i]$ de Y sachant x_i définie par

$$P : y_j \mapsto p(y_j|x_i) = P[Y = y_j|X = x_i] = \frac{p(x_i, y_j)}{p_X(x_i)}$$

Distributions conditionnelles

On associe à l'événement $Y = y_j$ une distribution conditionnelle $P[X|y_j]$ de X sachant y_j définie par

$$P : x_i \mapsto p(x_i|y_j) = P[X = x_i|Y = y_j] = \frac{p(x_i, y_j)}{p_Y(y_j)}$$

De façon analogue on définit la distribution conditionnelle $P[Y|x = x_i]$ de Y sachant x_i définie par

$$P : y_j \mapsto p(y_j|x_i) = P[Y = y_j|X = x_i] = \frac{p(x_i, y_j)}{p_X(x_i)}$$

Distributions conditionnelles. Exemple

Pour calculer $P[X = 0|Y = 0]$ on divise :

$$P[X = 0|Y = 0] =$$

	0	1	-1	P_X	0	1	-1
0	0.3	0.15	0.15				
1	0.1	0.2	0.1				
P_Y	0.4	0.35	0.25				
$P[X = 0\|Y]$							
$P[X = 1\|Y]$							

Distributions conditionnelles. Exemple

Pour calculer $P[X = 0|Y = 0]$ on divise :

$$P[X = 0|Y = 0] = \frac{P[X = 0 \text{ et } Y = 0]}{}$$

	0	1	-1	P_X	0	1	-1
0	0.3	0.15	0.15				
1	0.1	0.2	0.1				
P_Y	0.4	0.35	0.25				
$P[X = 0\|Y]$							
$P[X = 1\|Y]$							

Distributions conditionnelles. Exemple

Pour calculer $P[X=0|Y=0]$ on divise :

$$P[X=0|Y=0]=\frac{P[X=0 \text{ et } Y=0]}{P[Y=0]}$$

	0	1	-1	P_X	0	1	-1
0	0.3	0.15	0.15				
1	0.1	0.2	0.1				
P_Y	0.4	0.35	0.25				
$P[X=0\|Y]$							
$P[X=1\|Y]$							

Distributions conditionnelles. Exemple

Pour calculer $P[X = 0|Y = 0]$ on divise :

$$P[X = 0|Y = 0] = \frac{P[X = 0 \text{ et } Y = 0]}{P[Y = 0]} = \frac{0.3}{0.4} = \frac{3}{4}$$

	0	1	-1	P_X	0 1 -1
0	0.3	0.15	0.15		
1	0.1	0.2	0.1		
P_Y	0.4	0.35	0.25		
$P[X = 0\|Y]$	3/4				
$P[X = 1\|Y]$					

Distributions conditionnelles. Exemple

Pour calculer $P[X = 0|Y = 0]$ on divise :

$$P[X = 0|Y = 0] = \frac{P[X = 0 \text{ et } Y = 0]}{P[Y = 0]} = \frac{0.3}{0.4} = \frac{3}{4}$$

	0	1	-1	P_X	0	1	-1
0	0.3	0.15	0.15				
1	0.1	0.2	0.1				
P_Y	0.4	0.35	0.25				
$P[X = 0\|Y]$	3/4	3/7					
$P[X = 1\|Y]$							

Distributions conditionnelles. Exemple

Pour calculer $P[X=0|Y=0]$ on divise :

$$P[X=0|Y=0] = \frac{P[X=0 \text{ et } Y=0]}{P[Y=0]} = \frac{0.3}{0.4} = \frac{3}{4}$$

	0	1	-1	P_X	0	1	-1
0	0.3	0.15	0.15				
1	0.1	0.2	0.1				
P_Y	0.4	0.35	0.25				
$P[X=0\|Y]$	3/4	3/7	3/5				
$P[X=1\|Y]$							

Distributions conditionnelles. Exemple

Pour calculer $P[X = 0|Y = 0]$ on divise :

$$P[X = 0|Y = 0] = \frac{P[X = 0 \text{ et } Y = 0]}{P[Y = 0]} = \frac{0.3}{0.4} = \frac{3}{4}$$

	0	1	-1	P_X	0	1	-1
0	0.3	0.15	0.15				
1	0.1	0.2	0.1				
P_Y	0.4	0.35	0.25				
$P[X = 0\|Y]$	3/4	3/7	3/5				
$P[X = 1\|Y]$	1/4						

Distributions conditionnelles. Exemple

Pour calculer $P[X = 0|Y = 0]$ on divise :

$$P[X = 0|Y = 0] = \frac{P[X = 0 \text{ et } Y = 0]}{P[Y = 0]} = \frac{0.3}{0.4} = \frac{3}{4}$$

	0	1	-1	P_X	0	1	-1
0	0.3	0.15	0.15				
1	0.1	0.2	0.1				
P_Y	0.4	0.35	0.25				
$P[X = 0\|Y]$	3/4	3/7	3/5				
$P[X = 1\|Y]$	1/4	4/7					

Distributions conditionnelles. Exemple

Pour calculer $P[X = 0|Y = 0]$ on divise :

$$P[X = 0|Y = 0] = \frac{P[X = 0 \text{ et } Y = 0]}{P[Y = 0]} = \frac{0.3}{0.4} = \frac{3}{4}$$

	0	1	-1	P_X	0	1	-1
0	0.3	0.15	0.15				
1	0.1	0.2	0.1				
P_Y	0.4	0.35	0.25				
$P[X = 0\|Y]$	3/4	3/7	3/5				
$P[X = 1\|Y]$	1/4	4/7	2/5				

Distributions conditionnelles. Exemple

Pour calculer $P[X = 0|Y = 0]$ on divise :

$$P[X = 0|Y = 0] = \frac{P[X = 0 \text{ et } Y = 0]}{P[Y = 0]} = \frac{0.3}{0.4} = \frac{3}{4}$$

Pour calculer $P[Y = y_i|X = 0]$ on divise toute la premiere ligne de la matrice de probabilités conjointes par $P[X = 0]$:

$$P[Y = y_i|X = 0] = \frac{\color{red}{P[Y = y_i \text{ et } X = 0]}}{\color{blue}{P[X = 0]}}$$

	0	1	-1	P_X	0	1	-1
0	0.3	0.3	0.15	0.6	1/2	1/4	1/4
1	0.1	0.2	0.1				
P_Y	0.4	0.35	0.25				
$P[X = 0\|Y]$	3/4	3/7	3/5				
$P[X = 1\|Y]$	1/4	4/7	2/5				

Distributions conditionnelles. Exemple

Pour calculer $P[X=0|Y=0]$ on divise :

$$P[X=0|Y=0]=\frac{P[X=0 \text{ et } Y=0]}{P[Y=0]}=\frac{0.3}{0.4}=\frac{3}{4}$$

Pour calculer $P[Y=y_i|X=0]$ on divise toute la premiere ligne de la matrice de probabilités conjointes par $P[X=0]$:

$$P[Y=y_i|X=0]=\frac{P[Y=y_i \text{ et } X=0]}{P[X=0]}$$

	0	1	-1	P_X	0	1	-1
0	0.3	0.15	0.15	0.6	1/2	1/4	1/4
1	0.1	0.2	0.1	0.4	1/4	1/2	1/4
P_Y	0.4	0.35	0.25				
$P[X=0\|Y]$	3/4	3/7	3/5				
$P[X=1\|Y]$	1/4	4/7	2/5				

Distributions conditionnelles. Exemple

Pour calculer $P[X=0|Y=0]$ on divise :

$$P[X=0|Y=0] = \frac{P[X=0 \text{ et } Y=0]}{P[Y=0]} = \frac{0.3}{0.4} = \frac{3}{4}$$

Pour calculer $P[Y=y_i|X=0]$ on divise toute la premiere ligne de la matrice de probabilités conjointes par $P[X=0]$:

$$P[Y=y_i|X=0] = \frac{P[Y=y_i \text{ et } X=0]}{P[X=0]}$$

	0	1	-1	P_X	0	1	-1
0	0.3	0.15	0.15	0.6	1/2	1/4	1/4
1	0.1	0.2	0.1	0.4	1/4	1/2	1/4
P_Y	0.4	0.35	0.25				
$P[X=0\|Y]$	3/4	3/7	3/5				
$P[X=1\|Y]$	1/4	4/7	2/5				

Entropie conjointe

Définition

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{n} P(i,j)\log(P(i,j)).$$

Cette quantité représente l'information apportée en moyenne par l'observation d'un couple (x_i, y_j).

Entropie conjointe

Définition

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{n} P(i,j)\log(P(i,j)).$$

Cette quantité représente l'information apportée en moyenne par l'observation d'un couple (x_i, y_j).

Entropie conjointe : exemple

$$P(X,Y) = \begin{array}{c|ccc} & 0 & 1 & -1 \\ \hline 0 & 0.3 & 0.15 & 0.15 \\ 1 & 0.1 & 0.2 & 0.1 \end{array}$$

$$H(X,Y) \quad =$$

Entropie conjointe : exemple

$$P(X,Y) = \begin{array}{c|ccc} & 0 & 1 & -1 \\ \hline 0 & 0.3 & 0.15 & 0.15 \\ 1 & 0.1 & 0.2 & 0.1 \end{array}$$

$$H(X,Y) = -0.3\log_2(0.3)$$

Entropie conjointe : exemple

$$P(X,Y) = \begin{array}{c|ccc} & 0 & 1 & -1 \\ \hline 0 & 0.3 & 0.15 & 0.15 \\ 1 & 0.1 & 0.2 & 0.1 \end{array}$$

$$H(X,Y) = -0.3\log_2(0.3) - 2 * 0.15\log_2(0.15)$$

Entropie conjointe : exemple

$$P(X,Y) = \begin{array}{c|ccc} & 0 & 1 & -1 \\ \hline 0 & 0.3 & 0.15 & 0.15 \\ 1 & 0.1 & 0.2 & 0.1 \end{array}$$

$$\begin{aligned} H(X,Y) &= -0.3\log_2(0.3) - 2*0.15\log_2(0.15) \\ &= -2*0.1\log_2(0.1) \end{aligned}$$

Entropie conjointe : exemple

$$P(X,Y) = \begin{array}{c|ccc} & 0 & 1 & -1 \\ \hline 0 & 0.3 & 0.15 & 0.15 \\ 1 & 0.1 & \color{red}{0.2} & 0.1 \end{array}$$

$$\begin{aligned} H(X,Y) &= -0.3\log_2(0.3) - 2*0.15\log_2(0.15) \\ &= -2*0.1\log_2(0.1) \color{red}{- 0.2*\log_2(0.2)} \end{aligned}$$

Entropie conjointe : exemple

$$P(X,Y) = \begin{array}{c|ccc} & 0 & 1 & -1 \\ \hline 0 & 0.3 & 0.15 & 0.15 \\ 1 & 0.1 & 0.2 & 0.1 \end{array}$$

$$\begin{aligned} H(X,Y) &= -0.3\log_2(0.3) - 2*0.15\log_2(0.15) \\ &= -2*0.1\log_2(0.1) - 0.2*\log_2(0.2).2.47 \end{aligned}$$

Entropie conditionnelle moyenne

Définition

Alors l'**entropie conditionnelle moyenne** de X sachant Y est définie par

$$H(X|Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(i,j)\log(P(i|j)).$$

Entropie conditionnelle moyenne : exemple

	0	1	-1
0	0.3	0.15	0.15
1	0.1	0.2	0.1
$P[X=0\|Y]$	3/4	3/7	3/5
$P[X=1\|Y]$	3/4	3/7	3/5

$$H(X|Y) \quad = \quad .$$

Entropie conditionnelle moyenne : exemple

	0	1	-1
0	0.3	0.15	0.15
1	0.1	0.2	0.1
$P[X=0 \mid Y]$	3/4	3/7	3/5
$P[X=1 \mid Y]$	3/4	3/7	3/5

$$H(X|Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(i,j).$$

Entropie conditionnelle moyenne : exemple

	0	1	-1
0	0.3	0.15	0.15
1	0.1	0.2	0.1
$P[X=0\|Y]$	3/4	3/7	3/5
$P[X=1\|Y]$	1/4	4/7	2/5

$$H(X|Y) \quad = \quad -\sum_{i=1}^{n}\sum_{j=1}^{m} P(i,j)\log(P(i|j)).$$

Entropie conditionnelle moyenne : exemple

	0	1	-1
0	0.3	0.15	0.15
1	0.1	0.2	0.1
$P[X=0\|Y]$	3/4	3/7	3/5
$P[X=1\|Y]$	3/4	3/7	3/5

$$
\begin{aligned}
H(X|Y) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(i,j)\log(P(i|j)). \\
&= -0.3\log_2(3/4)
\end{aligned}
$$

Entropie conditionnelle moyenne : exemple

	0	1	-1
0	0.3	0.3	0.15
1	0.1	0.2	0.1
$P[X=0 \mid Y]$	3/4	3/7	3/5
$P[X=1 \mid Y]$	3/4	3/7	3/5

$$\begin{aligned} H(X|Y) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(i,j)\log(P(i|j)). \\ &= -0.3\log_2(3/4) - 0.15\log_2(3/7) \end{aligned}$$

Entropie conditionnelle moyenne : exemple

	0	1	-1
0	0.3	0.15	0.15
1	0.1	0.2	0.1
$P[X=0\|Y]$	3/4	3/7	3/5
$P[X=1\|Y]$	3/4	3/7	3/5

$$\begin{aligned} H(X|Y) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(i,j)\log(P(i|j)). \\ &= -0.3\log_2(3/4) - 0.15\log_2(3/7) - 0.15\log_2(3/5) \end{aligned}$$

Entropie conditionnelle moyenne : exemple

	0	1	-1
0	0.3	0.15	0.15
1	0.1	0.2	0.1
$P[X=0\|Y]$	$3/4$	$3/7$	$3/5$
$P[X=1\|Y]$	$1/4$	$3/7$	$3/5$

$$
\begin{aligned}
H(X|Y) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(i,j)\log(P(i|j)). \\
&= -0.3\log_2(3/4) - 0.15\log_2(3/7) - 0.15\log_2(3/5) \\
&= -0.1\log_2(1/4)
\end{aligned}
$$

Entropie conditionnelle moyenne : exemple

	0	1	-1
0	0.3	0.15	0.15
1	0.1	0.2	0.1
$P[X=0\|Y]$	3/4	3/7	3/5
$P[X=1\|Y]$	3/4	4/7	3/5

$$
\begin{aligned}
H(X|Y) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(i,j)\log(P(i|j)). \\
&= -0.3\log_2(3/4) - 0.15\log_2(3/7) - 0.15\log_2(3/5) \\
&= -0.1\log_2(1/4) - 0.2\log_2(4/7)
\end{aligned}
$$

Entropie conditionnelle moyenne : exemple

	0	1	-1
0	0.3	0.15	0.15
1	0.1	0.2	0.1
$P[X=0\|Y]$	3/4	3/7	3/5
$P[X=1\|Y]$	3/4	3/7	2/5

$$\begin{aligned} H(X|Y) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(i,j)\log(P(i|j)). \\ &= -0.3\log_2(3/4) - 0.15\log_2(3/7) - 0.15\log_2(3/5) \\ &= -0.1\log_2(1/4) - 0.2\log_2(4/7) - 0.1\log_2(2/5) \end{aligned}$$

Entropie conditionnelle moyenne : exemple

	0	1	-1
0	0.3	0.15	0.15
1	0.1	0.2	0.1
$P[X=0\|Y]$	3/4	3/7	3/5
$P[X=1\|Y]$	3/4	3/7	3/5

$$
\begin{aligned}
H(X|Y) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(i,j)\log(P(i|j)). \\
&= -0.3\log_2(3/4) - 0.15\log_2(3/7) - 0.15\log_2(3/5) \\
&= -0.1\log_2(1/4) - 0.2\log_2(4/7) - 0.1\log_2(2/5) \\
&= 0.91
\end{aligned}
$$

Information mutuelle moyenne

Définition

L'information mutuelle moyenne de X et Y est définie par

$$I(X;Y) = H(X) - H(X|Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(i,j)\log\left(\frac{P(i,j)}{P(x_i)P(y_j)}\right).$$

Interprétation

$I(X;Y)$ représente la diminution de l'incertitude et donc le gain de l'information sur X apporté par l'observation de Y.

Interprétation

$I(X;Y)$ représente la diminution de l'incertitude et donc le gain de l'information sur X apporté par l'observation de Y.

Interprétation

$I(X;Y)$ représente la diminution de l'incertitude et donc le gain de l'information sur X apporté par l'observation de Y.

On peut associer à un système de communication "source - canal - récepteur" différentes entropies :

- $H(X)$. **L'entropie de la source**
- $H(Y)$. **L'entropie du récepteur**
- $H(X, Y)$. **L'entropie conjointe "source-récepteur"**
- $H(X|Y)$. **L'entropie conditionnelle de la source, sachant le symbole reçu** Elle représente l'incertitude moyenne sur le symbole émis lorsqu'on connaît le symbole reçu.
- $H(Y|X)$. **L'entropie conditionnelle du récepteur, sachant le symbole émis** Elle représente l'incertitude moyenne sur le symbole reçu lorsqu'on connaît le symbole émis.
- $I(X; Y)$. **L'information mutuelle moyenne** Elle représente la quantité moyenne d'information par symbole transmise à travers le canal.

On peut associer à un système de communication "source - canal - récepteur" différentes entropies :

- $H(X)$. **L'entropie de la source**
- $H(Y)$. **L'entropie du récepteur**
- $H(X,Y)$. **L'entropie conjointe "source-récepteur"**
- $H(X|Y)$. **L'entropie conditionnelle de la source, sachant le symbole reçu** Elle représente l'incertitude moyenne sur le symbole émis lorsqu'on connaît le symbole reçu.
- $H(Y|X)$. **L'entropie conditionnelle du récepteur, sachant le symbole émis** Elle représente l'incertitude moyenne sur le symbole reçu lorsqu'on connaît le symbole émis.
- $I(X;Y)$. **L'information mutuelle moyenne** Elle représente la quantité moyenne d'information par symbole transmise à travers le canal.

On peut associer à un système de communication "source - canal - récepteur" différentes entropies :

- $H(X)$. **L'entropie de la source**
- $H(Y)$. **L'entropie du récepteur**
- $H(X, Y)$. **L'entropie conjointe "source-récepteur"**
- $H(X|Y)$. **L'entropie conditionnelle de la source, sachant le symbole reçu** Elle représente l'incertitude moyenne sur le symbole émis lorsqu'on connaît le symbole reçu.
- $H(Y|X)$. **L'entropie conditionnelle du récepteur, sachant le symbole émis** Elle représente l'incertitude moyenne sur le symbole reçu lorsqu'on connaît le symbole émis.
- $I(X; Y)$. **L'information mutuelle moyenne** Elle représente la quantité moyenne d'information par symbole transmise à travers le canal.

On peut associer à un système de communication "source - canal - récepteur" différentes entropies :

- $H(X)$. **L'entropie de la source**
- $H(Y)$. **L'entropie du récepteur**
- $H(X, Y)$. **L'entropie conjointe "source-récepteur"**
- $H(X|Y)$. **L'entropie conditionnelle de la source, sachant le symbole reçu** Elle représente l'incertitude moyenne sur le symbole émis lorsqu'on connaît le symbole reçu.
- $H(Y|X)$. **L'entropie conditionnelle du récepteur, sachant le symbole émis** Elle représente l'incertitude moyenne sur le symbole reçu lorsqu'on connaît le symbole émis.
- $I(X; Y)$. **L'information mutuelle moyenne** Elle représente la quantité moyenne d'information par symbole transmise à travers le canal.

On peut associer à un système de communication "source - canal - récepteur" différentes entropies :

- $H(X)$. **L'entropie de la source**
- $H(Y)$. **L'entropie du récepteur**
- $H(X, Y)$. **L'entropie conjointe "source-récepteur"**
- $H(X|Y)$. **L'entropie conditionnelle de la source, sachant le symbole reçu** Elle représente l'incertitude moyenne sur le symbole émis lorsqu'on connaît le symbole reçu.
- $H(Y|X)$. **L'entropie conditionnelle du récepteur, sachant le symbole émis** Elle représente l'incertitude moyenne sur le symbole reçu lorsqu'on connaît le symbole émis.
- $I(X; Y)$. **L'information mutuelle moyenne** Elle représente la quantité moyenne d'information par symbole transmise à travers le canal.

On peut associer à un système de communication "source - canal - récepteur" différentes entropies :

- $H(X)$. **L'entropie de la source**
- $H(Y)$. **L'entropie du récepteur**
- $H(X, Y)$. **L'entropie conjointe "source-récepteur"**
- $H(X|Y)$. **L'entropie conditionnelle de la source, sachant le symbole reçu** Elle représente l'incertitude moyenne sur le symbole émis lorsqu'on connaît le symbole reçu.
- $H(Y|X)$. **L'entropie conditionnelle du récepteur, sachant le symbole émis** Elle représente l'incertitude moyenne sur le symbole reçu lorsqu'on connaît le symbole émis.
- $I(X; Y)$. **L'information mutuelle moyenne** Elle représente la quantité moyenne d'information par symbole transmise à travers le canal.

Capacité de canal

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

Le maximum est pris sur toutes les distributions de probabilité possibles de la source.

Canal symétrique : définition

- Un canal symétrique est caractérisé par une propriété particulière de sa matrice de transition $P(Y|X)$.
- Toutes les lignes de cette matrice sont équivalentes entre elles par permutation et toutes les colonnes également.
- Voici un exemple d'une telle matrice

$$P(Y|X) = \begin{array}{c|ccc} & y_1 & y_2 & y_3 \\ \hline x_1 & 1/2 & 1/6 & 1/3 \\ x_2 & 1/6 & 1/3 & 1/2 \\ x_3 & 1/3 & 1/2 & 1/6 \end{array}$$

Canal symétrique : définition

- Un canal symétrique est caractérisé par une propriété particulière de sa matrice de transition $P(Y|X)$.
- Toutes les lignes de cette matrice sont équivalentes entre elles par permutation et toutes les colonnes également.
- Voici un exemple d'une telle matrice

$$P(Y|X) = \begin{array}{c|ccc} & y_1 & y_2 & y_3 \\ \hline x_1 & 1/2 & 1/6 & 1/3 \\ x_2 & 1/6 & 1/3 & 1/2 \\ x_3 & 1/3 & 1/2 & 1/6 \end{array}$$

Canal symétrique : définition

- Un canal symétrique est caractérisé par une propriété particulière de sa matrice de transition $P(Y|X)$.
- Toutes les lignes de cette matrice sont équivalentes entre elles par permutation et toutes les colonnes également.
- Voici un exemple d'une telle matrice

$$P(Y|X) = \begin{array}{c|ccc} & y_1 & y_2 & y_3 \\ \hline x_1 & 1/2 & 1/6 & 1/3 \\ x_2 & 1/6 & 1/3 & 1/2 \\ x_3 & 1/3 & 1/2 & 1/6 \end{array}$$

Canal symétrique : propriété importante

Proposition

Dans un canal symétrique l'entropie conditionnelle $H(Y|X)$ ne dépend pas de la distribution des probabilités de X. Elle est égale à

$$H(Y|X) = -\sum_{j=1}^{M} p(y_j|x_1) \cdot \log_2(p(y_j|x_1))$$

Cette proposition est démontrée dans le polycopié
Comme toutes les lignes de la matrice de transition sont équivalentes par permutations, la somme ci-dessus peut se faire le long de n'importe quelle ligne.

Canal symétrique : propriété importante

Proposition

Dans un canal symétrique l'entropie conditionnelle $H(Y|X)$ ne dépend pas de la distribution des probabilités de X. Elle est égale à

$$H(Y|X) = -\sum_{j=1}^{M} p(y_j|x_1) \cdot \log_2(p(y_j|x_1))$$

Cette proposition est démontrée dans le polycopié

Comme toutes les lignes de la matrice de transition sont équivalentes par permutations, la somme ci-dessus peut se faire le long de n'importe quelle ligne.

Canal symétrique : propriété importante

Proposition

Dans un canal symétrique l'entropie conditionnelle $H(Y|X)$ ne dépend pas de la distribution des probabilités de X. Elle est égale à

$$H(Y|X) = -\sum_{j=1}^{M} p(y_j|x_1) \cdot \log_2(p(y_j|x_1))$$

Cette proposition est démontrée dans le polycopié
Comme toutes les lignes de la matrice de transition sont équivalentes par permutations, la somme ci-dessus peut se faire le long de n'importe quelle ligne.

canal symétrique binaire

- Les deux alphabets, de source et de récepteur sont identiques et composés de deux caractères : $\Omega_X = \Omega_Y = \{0, 1\}$.
- La matrice de transition etc alors égale à

$$P(Y|X) = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}$$

où $\alpha \in [0, 1]$ est la probabilité d'erreur de transmission.

canal symétrique binaire

- Les deux alphabets, de source et de récepteur sont identiques et composés de deux caractères : $\Omega_X = \Omega_Y = \{0, 1\}$.
- La matrice de transition etc alors égale à

$$P(Y|X) = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}$$

où $\alpha \in [0, 1]$ est la probabilité d'erreur de transmission.

Capacité d'une canal symétrique

Proposition

Soit un couple émetteur-récepteur (X, Y) *avec les alphabets* $\Omega_X = \{x_i,\ i = 1, \ldots, n\}$ *et* $\Omega_X = \{x_i,\ i = 1, \ldots, m\}$. *Soit un canal symétrique de matrice de transition donnée*

$$P(Y|X) = \begin{pmatrix} p(y_1|x_1) & p(y_2|x_1) & \cdots & p(y_{m-1}|x_1) & p(y_m|x_1) \\ p(y_1|x_2) & \cdots & \cdots & \cdots & p(y_m|x_2) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ p(y_1|x_n) & \cdots & \cdots & \cdots & p(y_m|x_n) \end{pmatrix}$$

Alors la capacité de ce canal est

$$C = \log_2 m + \sum_{j=1}^{m} p(y_j|x_1)$$

Preuve

- On sait que $H(Y|X)$ ne dépend pas de la distribution de X pour un canal symétrique.
- Alors si l'on cherche à maximiser l'information transmise par le canal

$$I(Y,X) = H(Y) - H(Y|X)$$

- nous devons trouver la distribution de X qui maximise $H(Y)$ seulement puisque le deuxième terme de cette expression de dépend par de la distribution de X.

Preuve

- On sait que $H(Y|X)$ ne dépend pas de la distribution de X pour un canal symétrique.
- Alors si l'on cherche à maximiser l'information transmise par le canal

$$I(Y,X) = H(Y) - H(Y|X)$$

- nous devons trouver la distribution de X qui maximise $H(Y)$ seulement puisque le deuxième terme de cette expression de dépend par de la distribution de X.

Preuve

- On sait que $H(Y|X)$ ne dépend pas de la distribution de X pour un canal symétrique.
- Alors si l'on cherche à maximiser l'information transmise par le canal

$$I(Y,X) = H(Y) - H(Y|X)$$

- nous devons trouver la distribution de X qui maximise $H(Y)$ seulement puisque le deuxième terme de cette expression de dépend par de la distribution de X.

Preuve

- L'entropie atteint sa valeur maximale quand la variable aléatoire correspondante est distribuée uniformément.
- Montrons que si la source, X, est distribuée uniformément, alors la variable du récepteur, Y est aussi distribuée uniformément et donc $H(Y)$ est maximale.

Preuve

- L'entropie atteint sa valeur maximale quand la variable aléatoire correspondante est distribuée uniformément.
- Montrons que si la source, X, est distribuée uniformément, alors la variable du récepteur, Y est aussi distribuée uniformément et donc $H(Y)$ est maximale.

Preuve

- Supposons donc que $p(x_i) = \frac{1}{n}, \ \ \forall x_i \in \Omega_X$. Alors, en utilisant la règle de Bays, on trouve :

$$p(y_j) = \sum_{i=1}^{n} p(x_i, y_j) = \sum_{i=1}^{n} p(x_i)p(y_j|x_i) = \frac{1}{n}\sum_{i=1}^{n} p(y_j|x_i)$$

- Or, dans la matrice de transition d'un canal symétrique toutes les colonnes sont obtenues par permutation de l'une d'entre elles. Donc la somme

$$\sum_{i=1}^{n} p(y_j|x_i) = \sum_{i=1}^{n} p(y_1|x_i)$$

ne dépend pas de y_j.

- Ainsi on a montré que si la distribution de X est uniforme toutes les probabilités $p(y_j)$ sont égales entre elles. Donc dans ce cas $H(Y)$ est maximale et égale à $\log_2 m$.
- La formule de la proposition suit alors immédiatement de ce fait et de

Preuve

- Supposons donc que $p(x_i) = \dfrac{1}{n}, \quad \forall x_i \in \Omega_X$. Alors, en utilisant la règle de Bays, on trouve :

$$p(y_j) = \sum_{i=1}^{n} p(x_i, y_j) = \sum_{i=1}^{n} p(x_i)p(y_j|x_i) = \frac{1}{n}\sum_{i=1}^{n} p(y_j|x_i)$$

- Or, dans la matrice de transition d'un canal symétrique toutes les colonnes sont obtenues par permutation de l'une d'entre elles. Donc la somme

$$\sum_{i=1}^{n} p(y_j|x_i) = \sum_{i=1}^{n} p(y_1|x_i)$$

  ne dépend pas de y_j.

- Ainsi on a montré que si la distribution de X est uniforme toutes les probabilités $p(y_j)$ sont égales entre elles. Donc dans ce cas $H(Y)$ est maximale et égale à $\log_2 m$.
- La formule de la proposition suit alors immédiatement de ce fait et de

Preuve

- Supposons donc que $p(x_i) = \frac{1}{n}, \quad \forall x_i \in \Omega_X$. Alors, en utilisant la règle de Bays, on trouve :

$$p(y_j) = \sum_{i=1}^{n} p(x_i, y_j) = \sum_{i=1}^{n} p(x_i) p(y_j|x_i) = \frac{1}{n} \sum_{i=1}^{n} p(y_j|x_i)$$

- Or, dans la matrice de transition d'un canal symétrique toutes les colonnes sont obtenues par permutation de l'une d'entre elles. Donc la somme

$$\sum_{i=1}^{n} p(y_j|x_i) = \sum_{i=1}^{n} p(y_1|x_i)$$

ne dépend pas de y_j.

- Ainsi on a montré que si la distribution de X est uniforme toutes les probabilités $p(y_j)$ sont égales entre elles. Donc dans ce cas $H(Y)$ est maximale et égale à $\log_2 m$.
- La formule de la proposition suit alors immédiatement de ce fait et de

Preuve

- Supposons donc que $p(x_i) = \frac{1}{n}, \quad \forall x_i \in \Omega_X$. Alors, en utilisant la règle de Bays, on trouve :

$$p(y_j) = \sum_{i=1}^{n} p(x_i, y_j) = \sum_{i=1}^{n} p(x_i)p(y_j|x_i) = \frac{1}{n}\sum_{i=1}^{n} p(y_j|x_i)$$

- Or, dans la matrice de transition d'un canal symétrique toutes les colonnes sont obtenues par permutation de l'une d'entre elles. Donc la somme

$$\sum_{i=1}^{n} p(y_j|x_i) = \sum_{i=1}^{n} p(y_1|x_i)$$

ne dépend pas de y_j.

- Ainsi on a montré que si la distribution de X est uniforme toutes les probabilités $p(y_j)$ sont égales entre elles. Donc dans ce cas $H(Y)$ est maximale et égale à $\log_2 m$.
- La formule de la proposition suit alors immédiatement de ce fait et de

Exemple : capacité d'un canal sans bruit

Soit un canal sans bruit et une source d'alphabet $\Omega = \{x_i,\ i = 1, \dots, n\}$.

Dans ce cas $I(X;Y) = H(X) - H(X|Y) = H(X)$.

Alors

$$C = \max_{P(X)} I(X;Y) = \max_{P(X)} H(X) = \log n.$$

Exemple : capacité d'un canal sans bruit

Soit un canal sans bruit et une source d'alphabet $\Omega = \{x_i,\ i = 1, \ldots, n\}$.
Dans ce cas $I(X;Y) = H(X) - H(X|Y) = H(X)$.
Alors

$$C = \max_{P(X)} I(X;Y) = \max_{P(X)} H(X) = \log n.$$

Exemple : capacité d'un canal sans bruit

Soit un canal sans bruit et une source d'alphabet $\Omega = \{x_i,\ i = 1, \dots, n\}$.
Dans ce cas $I(X;Y) = H(X) - H(X|Y) = H(X)$.
Alors

$$C = \max_{P(X)} I(X;Y) = \max_{P(X)} H(X) = \log n.$$

Exemple : capacité d'un canal sans bruit

Soit un canal sans bruit et une source d'alphabet $\Omega = \{x_i,\ i = 1, \ldots, n\}$.
Dans ce cas $I(X;Y) = H(X) - H(X|Y) = H(X)$.
Alors

$$C = \max_{P(X)} I(X;Y) = \max_{P(X)} H(X) = \log n.$$

Problème de codage avec bruit

Étant donné un canal $X, Y, P(Y|X)$ de capacité C est il possible de transmettre des messages avec un débit aussi proche que possible de C et avec une probabilité d'erreur aussi petite que possible ?

Codage avec bruit : analyse du problème

- Codage sans bruit : la fonction de décodage est la transformation inverse de la fonction de codage ;
- En présence de bruit un message reçu peut correspondre à plusieurs messages en entrée
- La matrice de transition définit la loi de probabilité
- Il est donc nécessaire de définir une règle de décision permettant de décoder le message reçu

Codage avec bruit : analyse du problème

- Codage sans bruit : la fonction de décodage est la transformation inverse de la fonction de codage ;
- En présence de bruit un message reçu peut correspondre à plusieurs messages en entrée
- La matrice de transition définit la loi de probabilité
- Il est donc nécessaire de définir une règle de décision permettant de décoder le message reçu

Codage avec bruit : analyse du problème

- Codage sans bruit : la fonction de décodage est la transformation inverse de la fonction de codage ;
- En présence de bruit un message reçu peut correspondre à plusieurs messages en entrée
- La matrice de transition définit la loi de probabilité
- Il est donc nécessaire de définir une règle de décision permettant de décoder le message reçu

Codage avec bruit : analyse du problème

- Codage sans bruit : la fonction de décodage est la transformation inverse de la fonction de codage ;
- En présence de bruit un message reçu peut correspondre à plusieurs messages en entrée
- La matrice de transition définit la loi de probabilité
- Il est donc nécessaire de définir une règle de décision permettant de décoder le message reçu

Règle de décodage de canal

Définition

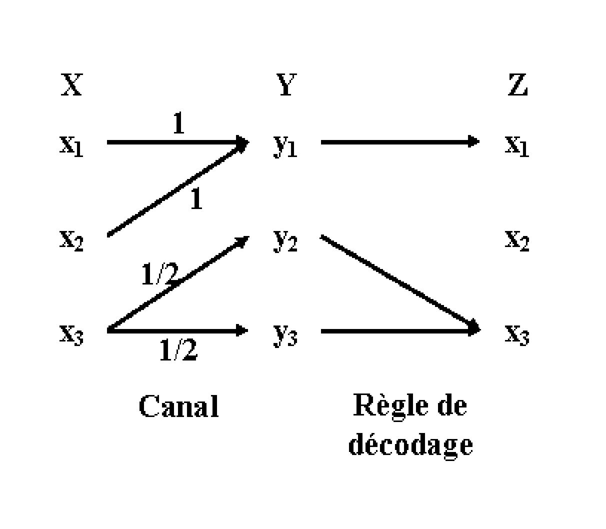
Soit un canal avec un alphabet d'entrée $\Omega_X = \{x_1, \ldots, x_n\}$ et un alphabet de sortie $\Omega_Y = \{y_1, \ldots, y_d\}$. La règle de décodage de canal est une fonction déterministe

$$g : \{y_1, \ldots, y_D\} \rightarrow \{x_1, \ldots, x_M\}$$

qui à chaque symbole reçu y_j associe un symbole de l'alphabet d'entrée $x_j^* = g(y_j)$.

Exemple arbitraire

Soient X, Y et $Z = g(Y)$ respectivement les symboles émis, reçu et décodé.



Règle de décodage

$$g(y_1) = x_1,\ g(y_1) = x_2,\ \ g(y_3) = x_2$$

Le seul cas où une erreur a lieu est l'émission de symbole x_2.
Ainsi la probabilité d'erreur est égale à la probabilité d'émission de x_2 donc à $1/4$.

Règle de décodage

$$g(y_1) = x_1, \; g(y_1) = x_2, \quad g(y_3) = x_2$$

Le seul cas où une erreur a lieu est l'émission de symbole x_2.
Ainsi la probabilité d'erreur est égale à la probabilité d'émission de x_2 donc à $1/4$.

Règle de décodage

$$g(y_1) = x_1, \; g(y_1) = x_2, \;\; g(y_3) = x_2$$

Le seul cas où une erreur a lieu est l'émission de symbole x_2.
Ainsi la probabilité d'erreur est égale à la probabilité d'émission de x_2 donc à $1/4$.

Erreur de transmission

Dans le cas général soit E l'événement correspondant à une erreur lors de transmission d'un seul symbole.

Soient les variables aléatoires X, Y et $Z = g(Y)$ représentant respectivement les symboles émis, reçu et décodé.

Comment peut on calculer $P[E]$?

Formule de Bays :

$$P(E) = \sum_{j=1}^{d} p(y_j) p(E|y_j)$$

Erreur de transmission

Dans le cas général soit E l'événement correspondant à une erreur lors de transmission d'un seul symbole.
Soient les variables aléatoires X, Y et $Z = g(Y)$ représentant respectivement les symboles émis, reçu et décodé.
Comment peut on calculer $P[E]$?
Formule de Bays :

$$P(E) = \sum_{j=1}^{d} p(y_j) p(E|y_j)$$

Erreur de transmission

Dans le cas général soit E l'événement correspondant à une erreur lors de transmission d'un seul symbole.

Soient les variables aléatoires X, Y et $Z = g(Y)$ représentant respectivement les symboles émis, reçu et décodé.

Comment peut on calculer $P[E]$?

Formule de Bays :

$$P(E) = \sum_{j=1}^{d} p(y_j)p(E|y_j)$$

Erreur de transmission

Dans le cas général soit E l'événement correspondant à une erreur lors de transmission d'un seul symbole.
Soient les variables aléatoires X, Y et $Z = g(Y)$ représentant respectivement les symboles émis, reçu et décodé.
Comment peut on calculer $P[E]$?
Formule de Bays :

$$P(E) = \sum_{j=1}^{d} p(y_j)p(E|y_j)$$

Erreur de transmission

- Supposons qu'on reçoit y_j. La règle de décodage va lui associer le caractère $x = g(y_j)$. L'erreur de transmission est alors équivalente à l'événement

$$X \neq g(y_j)$$

"le caractère émis, X, coïncide avec le caractère décodé, $g(y_j)$"

Erreur de transmission

- L'événement contraire, $X = g(y_j)$, correspond à la transmission correcte. Notons TC l'évènement de transmission correcte. Alors on a

$$\begin{aligned} P(E) &= 1 - P(TC) = 1 - \sum_{j=1}^{d} p(y_j)(p(TC|y_j) && (5.1) \\ &= 1 - \sum_{j=1}^{d} p(y_j)p(X = g(y_j)|y_j) && (5.2) \end{aligned}$$

- La probabilité d'erreur dépend ainsi non seulement des propriétés de la source et du canal mais aussi de la fonction de décodage !

Erreur de transmission

- L'événement contraire, $X = g(y_j)$, correspond à la transmission correcte. Notons TC l'évènement de transmission correcte. Alors on a

$$\begin{aligned} P(E) &= 1 - P(TC) = 1 - \sum_{j=1}^{d} p(y_j)(p(TC|y_j) && (5.1) \\ &= 1 - \sum_{j=1}^{d} p(y_j)p(X = g(y_j)|y_j) && (5.2) \end{aligned}$$

- **La probabilité d'erreur dépend ainsi non seulement des propriétés de la source et du canal mais aussi de la fonction de décodage !**

Observateur idéal

Définition

On appelle "Observateur idéal" la règle de décodage qui minimise la probabilité d'erreur.

Observateur idéal

- Ainsi la probabilité d'erreur de transmission se décompose de façon suivante

$$P(E) = 1 - \sum_{j=1}^{d} p(y_j)p(X = g(y_j)|y_j)$$

Dans cette somme les probabilités $p(y_j)$ ne dépendent pas du choix de la fonction g. Elles sont définies uniquement par la distribution de la source, X, et la matrice de transition du canal.

- Minimiser la probabilité d'erreur équivaut maximiser la probabilité de transmission correcte.
- On peut donc définir la règle de décodage de façon suivante : on associe à chaque y_j la valeur x_j^* qui maximise $p(X = x_i|y_j)$. On pose donc

$$g(y_j) = x_j* = arg \max_{i=1,\dots,n} p(x_i|y_j)$$

Ce choix rend maximal chaque terme de la somme et donc rend minimale la probabilité d'erreur.

Observateur idéal

- Ainsi la probabilité d'erreur de transmission se décompose de façon suivante
$$P(E) = 1 - \sum_{j=1}^{d} p(y_j)p(X = g(y_j)|y_j)$$
Dans cette somme les probabilités $p(y_j)$ ne dépendent pas du choix de la fonction g. Elles sont définies uniquement par la distribution de la source, X, et la matrice de transition du canal.
- Minimiser la probabilité d'erreur équivaut maximiser la probabilité de transmission correcte.
- On peut donc définir la règle de décodage de façon suivante : on associe à chaque y_j la valeur x_j^* qui maximise $p(X = x_i|y_j)$. On pose donc
$$g(y_j) = x_j* = arg \max_{i=1,\ldots,n} p(x_i|y_j)$$
Ce choix rend maximal chaque terme de la somme et donc rend minimale la probabilité d'erreur.

Observateur idéal

- Ainsi la probabilité d'erreur de transmission se décompose de façon suivante
$$P(E) = 1 - \sum_{j=1}^{d} p(y_j)p(X = g(y_j)|y_j)$$
Dans cette somme les probabilités $p(y_j)$ ne dépendent pas du choix de la fonction g. Elles sont définies uniquement par la distribution de la source, X, et la matrice de transition du canal.
- Minimiser la probabilité d'erreur équivaut maximiser la probabilité de transmission correcte.
- On peut donc définir la règle de décodage de façon suivante : on associe à chaque y_j la valeur x_j^* qui maximise $p(X = x_i|y_j)$. On pose donc
$$g(y_j) = x_j* = arg \max_{i=1,\dots,n} p(x_i|y_j)$$
Ce choix rend maximal chaque terme de la somme et donc rend minimale la probabilité d'erreur.

Observateur idéal

Proposition

La règle de décodage de l'observateur idéal est la fonction g_{oi} qui à tout $y_j, \quad j = 1, \ldots, d$ associe le symbole x_j^* qui maximise la probabilité

$$p(X = x_i | y_j), \quad i = 1, \ldots, n$$

Décodage de maximum de vraisemblance

Définition

La règle de décodage de maximum de vraisemblance est la fonction g_{mv} qui à tout $y_j, \quad j = 1, \ldots, d$ associe le symbole x_j^* qui maximise la probabilité

$$p(y_j | X = x_i), \quad i = 1, \ldots, n$$

Proposition

Si la distribution de la source est uniforme $(\forall\ i = 1, \ldots, n,\ \ p(x_i) = \frac{1}{n})$ alors les règles de l'observateur idéal et de maximum de vraisemblance sont équivalentes :

$$g_{oi}(y_j) = g_{mv}(y_j)$$

Décodage de maximum de vraisemblance

Définition

La règle de décodage de maximum de vraisemblance est la fonction g_{mv} qui à tout $y_j, \quad j = 1, \ldots, d$ associe le symbole x_j^* qui maximise la probabilité

$$p(y_j|X = x_i), \quad i = 1, \ldots, n$$

Proposition

Si la distribution de la source est uniforme ($\forall\ i = 1, \ldots, n,\ \ p(x_i) = \frac{1}{n}$) alors les règles de l'observateur idéal et de maximum de vraisemblance sont équivalentes :

$$g_{oi}(y_j) = g_{mv}(y_j)$$

Notion de code de canal

Définition

Soit un canal $X, Y, P(Y|X)$. Un code (n, l) pour ce canal est un couple (W, g) où

1. $W = \{w_1, \ldots, w_n\}$ est un ensemble de mots de longueur l dans l'alphabet d'entrée du canal Ω_X, appelés mots du code.
2. g est une règle de décodage

$$g : (\Omega_Y)^l \longrightarrow W$$

qui associe à toute séquence reçue de longueur l dans l'alphabet de sortie Ω_Y un des mots du code.

Les mots du code w_i peuvent être associés aux symboles de la source, x_i ou aux blocks de symboles de la source de longueur k (autrement dit aux symbole d'une extension de la source X).

Notion de code de canal

Définition

Soit un canal $X, Y, P(Y|X)$. Un code (n, l) pour ce canal est un couple (W, g) où

1. $W = \{w_1, \ldots, w_n\}$ est un ensemble de mots de longueur l dans l'alphabet d'entrée du canal Ω_X, appelés mots du code.
2. g est une règle de décodage

$$g : (\Omega_Y)^l \longrightarrow W$$

qui associe à toute séquence reçue de longueur l dans l'alphabet de sortie Ω_Y un des mots du code.

Les mots du code w_i peuvent être associés aux symboles de la source, x_i ou aux blocks de symboles de la source de longueur k (autrement dit aux symbole d'une extension de la source X).

Notion de code de canal

Définition

Soit un canal $X, Y, P(Y|X)$. Un code (n, l) pour ce canal est un couple (W, g) où

1. $W = \{w_1, \ldots, w_n\}$ est un ensemble de mots de longueur l dans l'alphabet d'entrée du canal Ω_X, appelés mots du code.
2. g est une règle de décodage

$$g : (\Omega_Y)^l \longrightarrow W$$

   qui associe à toute séquence reçue de longueur l dans l'alphabet de sortie Ω_Y un des mots du code.

Les mots du code w_i peuvent être associés aux symboles de la source, x_i ou aux blocks de symboles de la source de longueur k (autrement dit aux symbole d'une extension de la source X).

Notion de code de canal

Définition

Soit un canal $X, Y, P(Y|X)$. Un code (n, l) pour ce canal est un couple (W, g) où

1. $W = \{w_1, \ldots, w_n\}$ est un ensemble de mots de longueur l dans l'alphabet d'entrée du canal Ω_X, appelés mots du code.
2. g est une règle de décodage

$$g : (\Omega_Y)^l \longrightarrow W$$

   qui associe à toute séquence reçue de longueur l dans l'alphabet de sortie Ω_Y un des mots du code.

Les mots du code w_i peuvent être associés aux symboles de la source, x_i ou aux blocks de symboles de la source de longueur k (autrement dit aux symbole d'une extension de la source X).

Probabilité d'erreur conditionnelle

Définition

Soient un canal $X, Y, P(Y|X)$ et un code (n, l) pour ce canal (W, g). Pour chaque mot du code w_i on définit la probabilité d'erreur conditionnelle

$$\lambda_i^{(l)} = P[E|w_i]$$

Probabilité d'erreur moyenne et maximale

Probabilité d'erreur moyenne

Soient un canal $X, Y, P(Y|X)$ et un code (n, l) pour ce canal (W, g). On définit la probabilité d'erreur moyenne **(algébrique)** du code par :

$$\bar{\lambda}^{(l)} = \frac{1}{n} \sum_{i=1}^{n} \lambda_i^{(l)}.$$

Cette moyenne est définie en supposant que le choix d'un mot de code à transmettre est fait selon une distribution uniforme (tous les mots du code sont équiprobables)

Probabilité d'erreur maximale

$$\lambda^{(l)} = \max_{i=1,\dots,n} \lambda_i^{(l)}.$$

Probabilité d'erreur moyenne et maximale

Probabilité d'erreur moyenne

Soient un canal $X, Y, P(Y|X)$ et un code (n, l) pour ce canal (W, g). On définit la probabilité d'erreur moyenne **(algébrique)** du code par :

$$\bar{\lambda}^{(l)} = \frac{1}{n} \sum_{i=1}^{n} \lambda_i^{(l)}.$$

Cette moyenne est définie en supposant que le choix d'un mot de code à transmettre est fait selon une distribution uniforme (tous les mots du code sont équiprobables)

Probabilité d'erreur maximale

$$\lambda^{(l)} = \max_{i=1,\ldots,n} \lambda_i^{(l)}.$$

Probabilité d'erreur moyenne et maximale

Probabilité d'erreur moyenne

Soient un canal $X, Y, P(Y|X)$ et un code (n, l) pour ce canal (W, g). On définit la probabilité d'erreur moyenne **(algébrique)** du code par :

$$\bar{\lambda}^{(l)} = \frac{1}{n} \sum_{i=1}^{n} \lambda_i^{(l)}.$$

Cette moyenne est définie en supposant que le choix d'un mot de code à transmettre est fait selon une distribution uniforme (tous les mots du code sont équiprobables)

Probabilité d'erreur maximale

$$\lambda^{(l)} = \max_{i=1,\dots,n} \lambda_i^{(l)}.$$

Débit de communication d'un code

Définition

Soient un canal $X, Y, P(Y|X)$ et un code (n, l) pour ce canal (W, g). On définit le débit de communication du code par :

$$R = \frac{log_2(n)}{l}.$$

l'unité de mesure est Shannon par symbole transmis.

Dans l'hypothèse d'équiprobabilité des mots de code $\log_2(n)$ représente l'entropie associée au choix d'un mot du code à transmettre.

Débit de communication d'un code

Définition

Soient un canal $X, Y, P(Y|X)$ et un code (n, l) pour ce canal (W, g). On définit le débit de communication du code par :

$$R = \frac{log_2(n)}{l}.$$

l'unité de mesure est Shannon par symbole transmis.

Dans l'hypothèse d'équiprobabilité des mots de code $\log_2(n)$ représente l'entropie associée au choix d'un mot du code à transmettre.

Second théorème de Shannon

Théorème

Soit un canal $X, Y, P(Y|X)$ de capacité $C > 0$. Pour tout $R < C$ il est possible de trouver un code de canal avec un débit R et une probabilité d'erreur aussi petite que possible. Plus précisément, il existe une suite de codes $(M(l), l)$ tels que $M(l) = 2^{lR}$ telle que

$$\lim_{l \to \infty} \lambda^{(l)} = 0$$