Cours 1. Introduction. Formalisme mathématique

A. Désilles

18 mars 2009

Résumé

1. Plan du cours
2. Les origines de la théorie de l'information
3. Modèle mathématique du paradigme de Shannon
4. Mesure de l'information
   - Information propre
   - Entropie d'une variable aléatoire

Résumé

1. Plan du cours
2. Les origines de la théorie de l'information
3. Modèle mathématique du paradigme de Shannon
4. Mesure de l'information
   - Information propre
   - Entropie d'une variable aléatoire

Résumé

1. Plan du cours
2. Les origines de la théorie de l'information
3. Modèle mathématique du paradigme de Shannon
4. Mesure de l'information
   - Information propre
   - Entropie d'une variable aléatoire

Résumé

Résumé

Plan du cours

Plan du cours

Plan du cours

Plan du cours

Plan du cours

Plan du cours

Cours 1	Introduction. Concepts principaux
Cours 2	Modélisation mathématique de l'information
Cours 3	Théorèmes fondamentaux de la théorie de l'information
Cours 4	Algorithmes de codage. Partie I
Cours 5	Algorithmes de codage. Partie II
Cours 6	Applications à la compression des données.

Objectifs du cours

- Introduire les concepts principaux de la théorie de l'information
- Présenter les principaux domaines d'application
- Initier aux techniques de base de codage

Objectifs du cours

- Introduire les concepts principaux de la théorie de l'information
- Présenter les principaux domaines d'application
- Initier aux techniques de base de codage

Objectifs du cours

- Introduire les concepts principaux de la théorie de l'information
- Présenter les principaux domaines d'application
- Initier aux techniques de base de codage

Un peu d'histoire

- **1838** Le premier télégraphe électrique, conçu par S. Morse fonctionne.
- Chaque lettre de l'alphabet est représentée par une séquence de
  - **points** (courent électrique de courte durée)
  - **traits** (courent de longue durée)
  - **espaces** (absence de courent)
- **Fait curieux** : la lettre **"E"** est représentée de la façon la plus courte possible : un point.

Un peu d'histoire

- **1838** Le premier télégraphe électrique, conçu par S. Morse fonctionne.
- Chaque lettre de l'alphabet est représentée par une séquence de
  - **points** (courent électrique de courte durée)
  - **traits** (courent de longue durée)
  - **espaces** (absence de courent)
- Fait curieux : la lettre "E" est représentée de la façon la plus courte possible : un point.

Un peu d'histoire

- **1838** Le premier télégraphe électrique, conçu par S. Morse fonctionne.
- Chaque lettre de l'alphabet est représentée par une séquence de
  - **points** (courent électrique de courte durée)
  - **traits** (courent de longue durée)
  - **espaces** (absence de courent)
- **Fait curieux** : la lettre **"E"** est représentée de la façon la plus courte possible : un point.

Un peu d'histoire

- **1838** Le premier télégraphe électrique, conçu par S. Morse fonctionne.
- Chaque lettre de l'alphabet est représentée par une séquence de
  - **points** (courent électrique de courte durée)
  - **traits** (courent de longue durée)
  - **espaces** (absence de courent)
- **Fait curieux** : la lettre **"E"** est représentée de la façon la plus courte possible : un point.

Un peu d'histoire

- **1838** Le premier télégraphe électrique, conçu par S. Morse fonctionne.
- Chaque lettre de l'alphabet est représentée par une séquence de
  - **points** (courent électrique de courte durée)
  - **traits** (courent de longue durée)
  - **espaces** (absence de courent)
- **Fait curieux :** la lettre **"E"** est représentée de la façon la plus courte possible : un point.

Un peu d'histoire

- **1838** Le premier télégraphe électrique, conçu par S. Morse fonctionne.
- Chaque lettre de l'alphabet est représentée par une séquence de
  - **points** (courent électrique de courte durée)
  - **traits** (courent de longue durée)
  - **espaces** (absence de courent)
- **Fait curieux :** la lettre **"E"** est représentée de la façon la plus courte possible : un point.

Un peu d'histoire

- **1843** Construction de ligne télégraphique entre Washington et Baltimore
- Les lignes sont enterrées
- Des problèmes électriques causent des erreurs de transmission

Un peu d'histoire

- **1843** Construction de ligne télégraphique entre Washington et Baltimore
- Les lignes sont enterrées
- Des problèmes électriques causent des erreurs de transmission

Un peu d'histoire

- **1843** Construction de ligne télégraphique entre Washington et Baltimore
- Les lignes sont enterrées
- Des problèmes électriques causent des erreurs de transmission

Incertitude dans un processus de communication

- On ne connaît pas à l'avance le message émis. $\Rightarrow$ 1ère source d'incertitude.
- On ne connaît pas les erreurs qui ont été commises lors de la transmission. $\Rightarrow$ 2ème source d'incertitude.

Question

Comment quantifier l'incertitude ?

Incertitude dans un processus de communication

- On ne connaît pas à l'avance le message émis. $\Rightarrow$ 1ère source d'incertitude.
- On ne connaît pas les erreurs qui ont été commises lors de la transmission. $\Rightarrow$ 2ème source d'incertitude.

Question

Comment quantifier l'incertitude ?

Incertitude dans un processus de communication

- On ne connaît pas à l'avance le message émis. $\Rightarrow$ 1ère source d'incertitude.
- On ne connaît pas les erreurs qui ont été commises lors de la transmission. $\Rightarrow$ 2ème source d'incertitude.

Question

Comment quantifier l'incertitude ?

Incertitude dans un processus de communication

- On ne connaît pas à l'avance le message émis. $\Rightarrow$ 1ère source d'incertitude.
- On ne connaît pas les erreurs qui ont été commises lors de la transmission. $\Rightarrow$ 2ème source d'incertitude.

Question

Comment quantifier l'incertitude ?

Information = Incertitude ?

- En théorie de la communication le terme d'information est associé avec l'incertitude qui accompagne l'émission et la transmission d'un message.
- Ainsi, mesurer l'information signifie mesurer l'incertitude.

Problématique

La théorie de l'information fournit un formalisme rigoureux qui permet de modéliser, quantifier et analyser la notion de l'information.

Information = Incertitude ?

- En théorie de la communication le terme d'information est associé avec l'incertitude qui accompagne l'émission et la transmission d'un message.
- Ainsi, mesurer l'information signifie mesurer l'incertitude.

Problématique

La théorie de l'information fournit un formalisme rigoureux qui permet de modéliser, quantifier et analyser la notion de l'information.

Information = Incertitude ?

- En théorie de la communication le terme d'information est associé avec l'incertitude qui accompagne l'émission et la transmission d'un message.
- Ainsi, mesurer l'information signifie mesurer l'incertitude.

Problématique

La théorie de l'information fournit un formalisme rigoureux qui permet de modéliser, quantifier et analyser la notion de l'information.

Information = Incertitude ?

- En théorie de la communication le terme d'information est associé avec l'incertitude qui accompagne l'émission et la transmission d'un message.
- Ainsi, mesurer l'information signifie mesurer l'incertitude.

Problématique

La théorie de l'information fournit un formalisme rigoureux qui permet de modéliser, quantifier et analyser la notion de l'information.

Théorie de l'information et applications

- Le codage est une mise en application de la théorie de l'information. La théorie de codage développe les techniques pratiques de stockage et de transmission d'informations fiables.
- Compression des données. La théorie de l'information permet d'établir les limites théoriques de compression sans pertes.
- Cryptographie. Les concepts de la théorie de l'information permettent de définir et d'évaluer la sécurité des systèmes cryptographiques

Théorie de l'information et applications

- Le codage est une mise en application de la théorie de l'information. La théorie de codage développe les techniques pratiques de stockage et de transmission d'informations fiables.
- Compression des données. La théorie de l'information permet d'établir les limites théoriques de compression sans pertes.
- Cryptographie. Les concepts de la théorie de l'information permettent de définir et d'évaluer la sécurité des systèmes cryptographiques

Théorie de l'information et applications

- Le codage est une mise en application de la théorie de l'information. La théorie de codage développe les techniques pratiques de stockage et de transmission d'informations fiables.
- Compression des données. La théorie de l'information permet d'établir les limites théoriques de compression sans pertes.
- Cryptographie. Les concepts de la théorie de l'information permettent de définir et d'évaluer la sécurité des systèmes cryptographiques

C. Shannon (1916-2001)

- Etudes en génie électrique et mathématiques, dans les années 30
- Deuxième Guerre Mondiale : C. Shannon travaille sur la cryptographie pour les services secrets américains
- 1948 : Publication dans son article " A Mathematical Theory of Communications" des fondements de la théorie de l'Information
- Années 40-70 : travail au MIT et pour les laboratoires Bell

C. Shannon (1916-2001)

- Etudes en génie électrique et mathématiques, dans les années 30
- Deuxième Guerre Mondiale : C. Shannon travaille sur la cryptographie pour les services secrets américains
- 1948 : Publication dans son article " A Mathematical Theory of Communications" des fondements de la théorie de l'Information
- Années 40-70 : travail au MIT et pour les laboratoires Bell

C. Shannon (1916-2001)

- Etudes en génie électrique et mathématiques, dans les années 30
- Deuxième Guerre Mondiale : C. Shannon travaille sur la cryptographie pour les services secrets américains
- 1948 : Publication dans son article " A Mathematical Theory of Communications" des fondements de la théorie de l'Information
- Années 40-70 : travail au MIT et pour les laboratoires Bell

C. Shannon (1916-2001)

- Etudes en génie électrique et mathématiques, dans les années 30
- Deuxième Guerre Mondiale : C. Shannon travaille sur la cryptographie pour les services secrets américains
- 1948 : Publication dans son article " A Mathematical Theory of Communications" des fondements de la théorie de l'Information
- Années 40-70 : travail au MIT et pour les laboratoires Bell

Modèle de communication

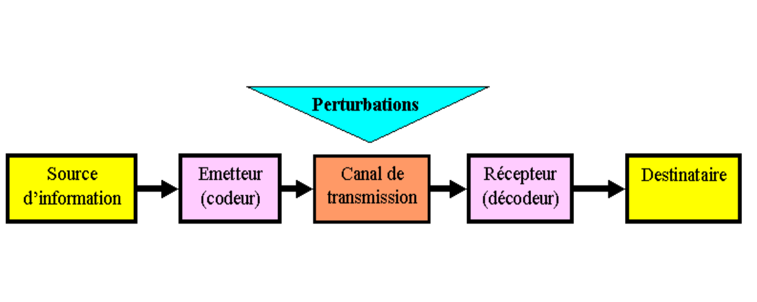


Figure: Paradigme de Shannon

Paradigme de Shannon

- la **source d'information** choisit un message M parmi un certain nombre de messages possibles ;
- l'**émetteur** transforme le message en signal S compatible physiquement avec le mode de transmission choisi.
- le signal S est alors soumis à l'entrée d'un **canal de transmission** ;
- lors de la transmission des perturbations peuvent transformer le signal envoyé ; on parle alors **de bruit de canal** ;
- à la sortie du canal le signal $\tilde{S}$ est soumis au décodeur qui le transforme en message $\tilde{M}$ lisible par le **destinataire**.

Paradigme de Shannon

- la **source d'information** choisit un message M parmi un certain nombre de messages possibles ;
- l'**émetteur** transforme le message en signal S compatible physiquement avec le mode de transmission choisi.
- le signal S est alors soumis à l'entrée d'un **canal de transmission** ;
- lors de la transmission des perturbations peuvent transformer le signal envoyé ; on parle alors **de bruit de canal** ;
- à la sortie du canal le signal $\tilde{S}$ est soumis au décodeur qui le transforme en message $\tilde{M}$ lisible par le **destinataire**.

Paradigme de Shannon

- la **source d'information** choisit un message M parmi un certain nombre de messages possibles ;
- l'**émetteur** transforme le message en signal S compatible physiquement avec le mode de transmission choisi.
- le signal S est alors soumis à l'entrée d'un **canal de transmission** ;
- lors de la transmission des perturbations peuvent transformer le signal envoyé ; on parle alors **de bruit de canal** ;
- à la sortie du canal le signal $\tilde{S}$ est soumis au décodeur qui le transforme en message $\tilde{M}$ lisible par le **destinataire**.

Paradigme de Shannon

- la **source d'information** choisit un message M parmi un certain nombre de messages possibles ;
- l'**émetteur** transforme le message en signal S compatible physiquement avec le mode de transmission choisi.
- le signal S est alors soumis à l'entrée d'un **canal de transmission** ;
- lors de la transmission des perturbations peuvent transformer le signal envoyé ; on parle alors **de bruit de canal** ;
- à la sortie du canal le signal $\tilde{S}$ est soumis au décodeur qui le transforme en message $\tilde{M}$ lisible par le **destinataire**.

Paradigme de Shannon

- la **source d'information** choisit un message M parmi un certain nombre de messages possibles ;
- l'**émetteur** transforme le message en signal S compatible physiquement avec le mode de transmission choisi.
- le signal S est alors soumis à l'entrée d'un **canal de transmission** ;
- lors de la transmission des perturbations peuvent transformer le signal envoyé ; on parle alors **de bruit de canal** ;
- à la sortie du canal le signal $\tilde{S}$ est soumis au décodeur qui le transforme en message $\tilde{M}$ lisible par le **destinataire**.

Caractère aléatoire d'une source

- Du point de vue du destinataire le message émis n'est pas connu à l'avance.
- Il peut donc être considéré comme étant aléatoire.
- On peut définir une variable aléatoire X associée à l'observation d'un symbole émis par la source.
- L'alphabet $\Omega_X = \{x_1, \ldots, x_2\}$ représente alors le support (l'ensemble de valeurs possibles de X)
- L'ensemble de probabilités $P_X = \{p_i, \; i = 1, \ldots, n\}$ d'émission des symboles définit la distribution de probabilités de X.

Caractère aléatoire d'une source

- Du point de vue du destinataire le message émis n'est pas connu à l'avance.
- Il peut donc être considéré comme étant aléatoire.
- On peut définir une variable aléatoire X associée à l'observation d'un symbole émis par la source.
- L'alphabet $\Omega_X = \{x_1, \ldots, x_2\}$ représente alors le support (l'ensemble de valeurs possibles de X)
- L'ensemble de probabilités $P_X = \{p_i, \ i = 1, \ldots, n\}$ d'émission des symboles définit la distribution de probabilités de X.

Caractère aléatoire d'une source

- Du point de vue du destinataire le message émis n'est pas connu à l'avance.
- Il peut donc être considéré comme étant aléatoire.
- On peut définir une variable aléatoire X associée à l'observation d'un symbole émis par la source.
- L'alphabet $\Omega_X = \{x_1, \ldots, x_2\}$ représente alors le support (l'ensemble de valeurs possibles de X)
- L'ensemble de probabilités $P_X = \{p_i, \ i = 1, \ldots, n\}$ d'émission des symboles définit la distribution de probabilités de X.

Caractère aléatoire d'une source

- Du point de vue du destinataire le message émis n'est pas connu à l'avance.
- Il peut donc être considéré comme étant aléatoire.
- On peut définir une variable aléatoire X associée à l'observation d'un symbole émis par la source.
- L'alphabet $\Omega_X = \{x_1, \ldots, x_2\}$ représente alors le support (l'ensemble de valeurs possibles de X)
- L'ensemble de probabilités $P_X = \{p_i,\ i = 1, \ldots, n\}$ d'émission des symboles définit la distribution de probabilités de X.

Caractère aléatoire d'une source

- Du point de vue du destinataire le message émis n'est pas connu à l'avance.
- Il peut donc être considéré comme étant aléatoire.
- On peut définir une variable aléatoire X associée à l'observation d'un symbole émis par la source.
- L'alphabet $\Omega_X = \{x_1, \ldots, x_2\}$ représente alors le support (l'ensemble de valeurs possibles de X)
- L'ensemble de probabilités $P_X = \{p_i,\ i = 1, \ldots, n\}$ d'émission des symboles définit la distribution de probabilités de X.

Caractère aléatoire d'un canal de transmission

- 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.

Caractère aléatoire d'un canal de transmission

- 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.

Caractère aléatoire d'un canal de transmission

- 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.

Caractère aléatoire d'un canal de transmission

- 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 d'une source d'information

Une source d'information X est décrite par un couple (Ω_X, P_X) où Ω_X est un alphabet fini et P_X est une distribution de probabilités sur Ω_X.

Exemple

Soit une source binaire, produisant des symboles 0 et 1 avec les probabilités respectives p et $q = 1 - p$. Nous avons ici l'alphabet constitué de deux symboles $\Omega = \{0, 1\}$ et $P[X = 0] = p, \quad P[X = 1] = q$.

Modèle d'une source d'information

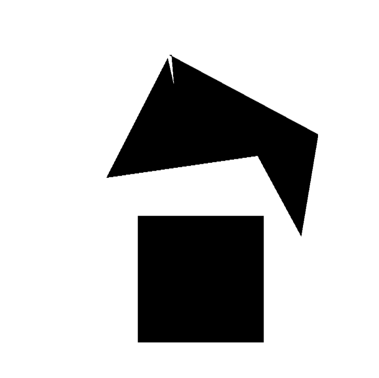
Une source d'information X est décrite par un couple (Ω_X, P_X) où Ω_X est un alphabet fini et P_X est une distribution de probabilités sur Ω_X.

Exemple

Soit une source binaire, produisant des symboles 0 et 1 avec les probabilités respectives p et $q = 1 - p$. Nous avons ici l'alphabet constitué de deux symboles $\Omega = \{0, 1\}$ et $P[X = 0] = p, \quad P[X = 1] = q$.

Un autre exemple

Cette image est un message d'une source binaire : 0 pour le noir et 1 pour le blanc



Un texte est aussi une source d'informations

Un texte français est un message d'une source utilisant l'alphabet latin. La distribution de probabilité de cette source est définie par les fréquences d'apparition des lettres, établies empiriquement.

Lettre	e	s	a	i	t	n	r	u
Probabilité (%)	14.715	7.948	7.636	7.529	7.244	7.095	6.553	6.311

Modèle du couple émetteur-récepteur

- 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 du couple émetteur-récepteur

- 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 du couple émetteur-récepteur

- 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 du couple émetteur-récepteur

- 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

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

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

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

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

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

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

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

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

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

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

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

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 :

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

Distributions marginales

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 :

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

Distributions conditionnelles

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

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

Information propre.

- Que cherche-t-on à mesurer exactement ?
- L'incertitude que l'on a à prévoir un événement.
- On joue à "pile ou face" ?
- Comment quantifiez vous votre difficulté à deviner le résultat ?
- Et si on lance un dé à 6 faces ? Est ce plus difficile à deviner ?
- Et si on vous donnait une indication ?
- Par exemple, que le résultat est pair.
- Combien d'information cette indication vous a apporté ?
- Intuitivement, on pourrait dire : "plus l'événement est rare, plus il est difficile à prévoir"

Information propre.

- Que cherche-t-on à mesurer exactement ?
- L'incertitude que l'on a à prévoir un événement.
- On joue à "pile ou face" ?
- Comment quantifiez vous votre difficulté à deviner le résultat ?
- Et si on lance un dé à 6 faces ? Est ce plus difficile à deviner ?
- Et si on vous donnait une indication ?
- Par exemple, que le résultat est pair.
- Combien d'information cette indication vous a apporté ?
- Intuitivement, on pourrait dire : "plus l'événement est rare, plus il est difficile à prévoir"

Information propre.

- Que cherche-t-on à mesurer exactement ?
- L'incertitude que l'on a à prévoir un événement.
- On joue à "pile ou face" ?
- Comment quantifiez vous votre difficulté à deviner le résultat ?
- Et si on lance un dé à 6 faces ? Est ce plus difficile à deviner ?
- Et si on vous donnait une indication ?
- Par exemple, que le résultat est pair.
- Combien d'information cette indication vous a apporté ?
- Intuitivement, on pourrait dire : "plus l'événement est rare, plus il est difficile à prévoir"

Information propre.

- Que cherche-t-on à mesurer exactement ?
- L'incertitude que l'on a à prévoir un événement.
- On joue à "pile ou face" ?
- Comment quantifiez vous votre difficulté à deviner le résultat ?
- Et si on lance un dé à 6 faces ? Est ce plus difficile à deviner ?
- Et si on vous donnait une indication ?
- Par exemple, que le résultat est pair.
- Combien d'information cette indication vous a apporté ?
- Intuitivement, on pourrait dire : "plus l'événement est rare, plus il est difficile à prévoir"

Information propre.

- Que cherche-t-on à mesurer exactement ?
- L'incertitude que l'on a à prévoir un événement.
- On joue à "pile ou face" ?
- Comment quantifiez vous votre difficulté à deviner le résultat ?
- Et si on lance un dé à 6 faces ? Est ce plus difficile à deviner ?
- Et si on vous donnait une indication ?
- Par exemple, que le résultat est pair.
- Combien d'information cette indication vous a apporté ?
- Intuitivement, on pourrait dire : "plus l'événement est rare, plus il est difficile à prévoir"

Information propre.

- Que cherche-t-on à mesurer exactement ?
- L'incertitude que l'on a à prévoir un événement.
- On joue à "pile ou face" ?
- Comment quantifiez vous votre difficulté à deviner le résultat ?
- Et si on lance un dé à 6 faces ? Est ce plus difficile à deviner ?
- Et si on vous donnait une indication ?
- Par exemple, que **le résultat est pair**.
- **Combien** d'information cette indication vous a apporté ?
- Intuitivement, on pourrait dire : "plus l'événement est rare, plus il est difficile à prévoir"

Information propre.

- Que cherche-t-on à mesurer exactement ?
- L'incertitude que l'on a à prévoir un événement.
- On joue à "pile ou face" ?
- Comment quantifiez vous votre difficulté à deviner le résultat ?
- Et si on lance un dé à 6 faces ? Est ce plus difficile à deviner ?
- Et si on vous donnait une indication ?
- Par exemple, que **le résultat est pair**.
- Combien d'information cette indication vous a apporté ?
- Intuitivement, on pourrait dire : "plus l'événement est rare, plus il est difficile à prévoir"

Information propre.

- Que cherche-t-on à mesurer exactement ?
- L'incertitude que l'on a à prévoir un événement.
- On joue à "pile ou face" ?
- Comment quantifiez vous votre difficulté à deviner le résultat ?
- Et si on lance un dé à 6 faces ? Est ce plus difficile à deviner ?
- Et si on vous donnait une indication ?
- Par exemple, que **le résultat est pair**.
- **Combien** d'information cette indication vous a apporté ?
- Intuitivement, on pourrait dire : "plus l'événement est rare, plus il est difficile à prévoir"

Information propre.

- Que cherche-t-on à mesurer exactement ?
- L'incertitude que l'on a à prévoir un événement.
- On joue à "pile ou face" ?
- Comment quantifiez vous votre difficulté à deviner le résultat ?
- Et si on lance un dé à 6 faces ? Est ce plus difficile à deviner ?
- Et si on vous donnait une indication ?
- Par exemple, que **le résultat est pair**.
- **Combien** d'information cette indication vous a apporté ?
- Intuitivement, on pourrait dire : "plus l'événement est rare, plus il est difficile à prévoir"

Un premier exemple : probabilités et information

- Reprenons l'exemple du dé.
- Imaginons que le résultat est $X = 2$. Soit A l'événement associé. $P(A) = \frac{1}{6}$.
- Est ce difficile à deviner ? Nous avons une chance de $\frac{1}{6}$ d'avoir la réponse correcte.
- Imaginons que l'on nous dit que le résultat du lancé est un nombre pair. Soit B l'événement associé. On a $P(B) = \frac{1}{2}$.
- Cette information, augmente notre chance de réussite.
- En effet, nous avons maintenant, la probabilité de $P(A|B) = \frac{1}{3}$ de deviner le résultat.

Un premier exemple : probabilités et information

- Reprenons l'exemple du dé.
- Imaginons que le résultat est $X = 2$. Soit A l'événement associé. $P(A) = \dfrac{1}{6}$.
- Est ce difficile à deviner ? Nous avons une chance de $\dfrac{1}{6}$ d'avoir la réponse correcte.
- Imaginons que l'on nous dit que le résultat du lancé est un nombre pair. Soit B l'événement associé. On a $P(B) = \dfrac{1}{2}$.
- Cette information, augmente notre chance de réussite.
- En effet, nous avons maintenant, la probabilité de $P(A|B) = \dfrac{1}{3}$ de deviner le résultat.

Un premier exemple : probabilités et information

- Reprenons l'exemple du dé.
- Imaginons que le résultat est $X = 2$. Soit A l'événement associé. $P(A) = \frac{1}{6}$.
- Est ce difficile à deviner ? Nous avons une chance de $\frac{1}{6}$ d'avoir la réponse correcte.
- Imaginons que l'on nous dit que le résultat du lancé est un nombre pair. Soit B l'événement associé. On a $P(B) = \frac{1}{2}$.
- Cette information, augmente notre chance de réussite.
- En effet, nous avons maintenant, la probabilité de $P(A|B) = \frac{1}{3}$ de deviner le résultat.

Un premier exemple : probabilités et information

- Reprenons l'exemple du dé.
- Imaginons que le résultat est $X = 2$. Soit A l'événement associé. $P(A) = \frac{1}{6}$.
- Est ce difficile à deviner ? Nous avons une chance de $\frac{1}{6}$ d'avoir la réponse correcte.
- Imaginons que l'on nous dit que le résultat du lancé est un nombre pair. Soit B l'événement associé. On a $P(B) = \frac{1}{2}$.
- Cette information, augmente notre chance de réussite.
- En effet, nous avons maintenant, la probabilité de $P(A|B) = \frac{1}{3}$ de deviner le résultat.

Un premier exemple : probabilités et information

- Reprenons l'exemple du dé.
- Imaginons que le résultat est $X = 2$. Soit A l'événement associé. $P(A) = \frac{1}{6}$.
- Est ce difficile à deviner ? Nous avons une chance de $\frac{1}{6}$ d'avoir la réponse correcte.
- Imaginons que l'on nous dit que le résultat du lancé est un nombre pair. Soit B l'événement associé. On a $P(B) = \frac{1}{2}$.
- Cette information, augmente notre chance de réussite.
- En effet, nous avons maintenant, la probabilité de $P(A|B) = \frac{1}{3}$ de deviner le résultat.

Un premier exemple : probabilités et information

- Reprenons l'exemple du dé.
- Imaginons que le résultat est $X = 2$. Soit A l'événement associé. $P(A) = \frac{1}{6}$.
- Est ce difficile à deviner ? Nous avons une chance de $\frac{1}{6}$ d'avoir la réponse correcte.
- Imaginons que l'on nous dit que le résultat du lancé est un nombre pair. Soit B l'événement associé. On a $P(B) = \frac{1}{2}$.
- Cette information, augmente notre chance de réussite.
- En effet, nous avons maintenant, la probabilité de $P(A|B) = \frac{1}{3}$ de deviner le résultat.

Premier exemple : conclusion

L'observation d'un événement aléatoire A nous apporte une quantité d'information $h(A)$ qui est

- non négative : $h(A) \geq 0$
- nulle si l'événement est certain : $P(A) = 1 \Rightarrow h(A) = 0$
- d'autant plus grande que l'événement est rare ou incertain : $P(A_1) < P(A_2) \Rightarrow h(A_1) > h(A_2)$
- additive pour les événements indépendants : $h(A \cap B) = h(A) + h(B)$

Premier exemple : conclusion

L'observation d'un événement aléatoire A nous apporte une quantité d'information $h(A)$ qui est

- non négative : $h(A) \geq 0$
- nulle si l'événement est certain : $P(A) = 1 \;\Rightarrow\; h(A) = 0$
- d'autant plus grande que l'événement est rare ou incertain : $P(A_1) < P(A_2) \;\Rightarrow\; h(A_1) > h(A_2)$
- additive pour les événements indépendants : $h(A \cap B) = h(A) + h(B)$

Premier exemple : conclusion

L'observation d'un événement aléatoire A nous apporte une quantité d'information $h(A)$ qui est

- non négative : $h(A) \geq 0$
- nulle si l'événement est certain : $P(A) = 1 \;\Rightarrow\; h(A) = 0$
- d'autant plus grande que l'événement est rare ou incertain : $P(A_1) < P(A_2) \;\Rightarrow\; h(A_1) > h(A_2)$
- additive pour les événements indépendants : $h(A \cap B) = h(A) + h(B)$

Premier exemple : conclusion

L'observation d'un événement aléatoire A nous apporte une quantité d'information $h(A)$ qui est

- non négative : $h(A) \geq 0$
- nulle si l'événement est certain : $P(A) = 1 \;\Rightarrow\; h(A) = 0$
- d'autant plus grande que l'événement est rare ou incertain : $P(A_1) < P(A_2) \;\Rightarrow\; h(A_1) > h(A_2)$
- additive pour les événements indépendants : $h(A \cap B) = h(A) + h(B)$

Information propre. Définition

Information propre

Soient $\Omega = \{\omega_1, \ldots, \omega_m\}$ un alphabet discret et X la variable aléatoire associée. Pour tout événement $A \subset \Omega$ la quantité d'information propre de A est définie par

$$h(A) = -\log_2(P(A)).$$

Entropie d'une source. Idée.

Soit une source X d'alphabet $\Omega_X = \{x_1, \ldots, x_m\}$ et de distribution de probabilité $P_X = \{p_1, \ldots, p_n\}$.
L'information associé à l'observation de chaque symbole x_i est $-\log(p_i)$

Quantité d'information moyenne

L'entropie d'une telle source représente la quantité moyenne d'information propre associée à l'observation de chacun des symboles possibles.

Entropie d'une source. Idée.

Soit une source X d'alphabet $\Omega_X = \{x_1, \ldots, x_m\}$ et de distribution de probabilité $P_X = \{p_1, \ldots, p_n\}$.
L'information associé à l'observation de chaque symbole x_i est $-\log(p_i)$

Quantité d'information moyenne

L'entropie d'une telle source représente la quantité moyenne d'information propre associée à l'observation de chacun des symboles possibles.

Entropie d'une source. Idée.

Soit une source X d'alphabet $\Omega_X = \{x_1, \ldots, x_m\}$ et de distribution de probabilité $P_X = \{p_1, \ldots, p_n\}$.
L'information associé à l'observation de chaque symbole x_i est $-\log(p_i)$

Quantité d'information moyenne

L'entropie d'une telle source représente la quantité moyenne d'information propre associée à l'observation de chacun des symboles possibles.

Entrepie d'une source. Définition.

Entropie d'une source

Soient $\Omega_X = \{x_1, \ldots, x_m\}$ l'alphabet fini d'une source et X la variable aléatoire associée t.q. $P[\omega_i] = p_i, \quad i = 1, \ldots, m$. On appelle **entropie** ou encore **quantité moyenne d'information** de la source la quantité

$$H(X) = H(p_1, p_2, \ldots, p_n) = E[h(x)] = -\sum_{i=1}^{m} p_i \log(p_i)$$

L'unité de mesure de cette quantité est le "bit par symbole".

Exemple. Entropie d'une source binaire

Soit une source émettant des symboles 0 avec la probabilité p et 1 avec la probabilité $q = 1 - p$.

$$H_2(p) = -p\log(p) - (1-p)\log(1-p).$$

Lorsque les deux symboles sont équiprobables ($p = 1/2$) :

$$H_2(1/2) = -1/2\log(1/2) - 1/2\log(1/2) = 1.$$

Nous pouvons interpréter ce résultat comme suit : *lorsque les symboles d'une source binaire sont équiprobables, il faut un bit par symbole en moyenne.*

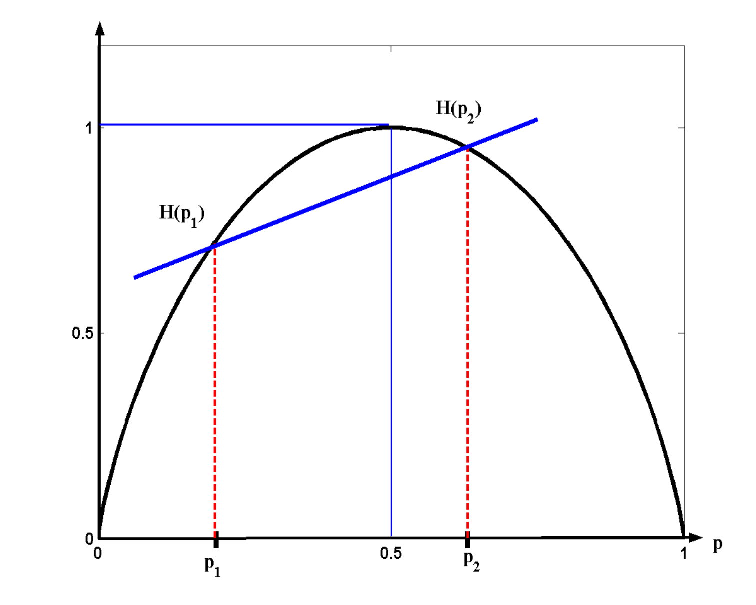


Figure: Fonction d'entropie d'une source binaire

Soit $H_2(p),\ p \in [0,1]$ l'entropie d'une source binaire.

1. $H_2(p)$ est une fonction continue sur $]0,1[$ telle que

$$\lim_{p \to 0+} H_2(p) = 0 = \lim_{p \to 1-} H_2(p)$$

2. $H_2(p)$ est positive sur $H_2(p),\ p \in [0,1]$
3. $H_2(p)$ est symétrique par rapport à $p_0 = 0.5$ et atteint son maximum en p_0 t.q. $H_2(0.5) = 1$.
4. $H_2(p)$ est strictement concave :

$$\forall (p_1, p_2) \in [0,1],\ p_1 \neq p_2, \quad \forall \lambda \in]0,1[$$

$$H_2(\lambda p_1 + (1-\lambda)p_2) > \lambda H_2(p_1) + (1-\lambda) H_2(p_2)$$

Propriétés de l'entropie d'une source binaire

Soit $H_2(p),\ p \in [0,1]$ l'entropie d'une source binaire.

1. $H_2(p)$ est une fonction continue sur $]0,1[$ telle que

$$\lim_{p\to 0+} H_2(p) = 0 = \lim_{p\to 1-} H_2(p)$$

2. $H_2(p)$ est positive sur $H_2(p),\ p \in [0,1]$
3. $H_2(p)$ est symétrique par rapport à $p_0 = 0.5$ et atteint son maximum en p_0 t.q. $H_2(0.5) = 1$.
4. $H_2(p)$ est strictement concave :

$$\forall (p_1, p_2) \in [0,1],\ p_1 \neq p_2, \quad \forall \lambda \in]0,1[$$

$$H_2(\lambda p_1 + (1-\lambda)p_2) > \lambda H_2(p_1) + (1-\lambda) H_2(p_2)$$

Propriétés de l'entropie d'une source binaire

Soit $H_2(p)$, $p \in [0,1]$ l'entropie d'une source binaire.

1. $H_2(p)$ est une fonction continue sur $]0,1[$ telle que

$$\lim_{p \to 0+} H_2(p) = 0 = \lim_{p \to 1-} H_2(p)$$

2. $H_2(p)$ est positive sur $H_2(p)$, $p \in [0,1]$
3. $H_2(p)$ est symétrique par rapport à $p_0 = 0.5$ et atteint son maximum en p_0 t.q. $H_2(0.5) = 1$.
4. $H_2(p)$ est strictement concave :

$$\forall (p_1, p_2) \in [0,1], \; p_1 \neq p_2, \quad \forall \lambda \in]0,1[$$

$$H_2(\lambda p_1 + (1-\lambda)p_2) > \lambda H_2(p_1) + (1-\lambda) H_2(p_2)$$

Propriétés de l'entropie d'une source binaire

Soit $H_2(p),\ p \in [0,1]$ l'entropie d'une source binaire.

1. $H_2(p)$ est une fonction continue sur $]0,1[$ telle que

$$\lim_{p \to 0+} H_2(p) = 0 = \lim_{p \to 1-} H_2(p)$$

2. $H_2(p)$ est positive sur $H_2(p),\ p \in [0,1]$
3. $H_2(p)$ est symétrique par rapport à $p_0 = 0.5$ et atteint son maximum en p_0 t.q. $H_2(0.5) = 1$.
4. $H_2(p)$ est strictement concave :

$$\forall (p_1, p_2) \in [0,1],\ p_1 \neq p_2, \quad \forall \lambda \in]0,1[$$

$$H_2(\lambda p_1 + (1-\lambda)p_2) > \lambda H_2(p_1) + (1-\lambda) H_2(p_2)$$

Interprétations

- **Entropie comme mesure d'information**
- Entropie comme mesure de nombre de bits pour le codage

Interprétations

- Entropie comme mesure d'information
- Entropie comme mesure de nombre de bits pour le codage

Interprétations. Exemple

Soit une source S définie par

X	a	b	c	d	e
P(X)	0.3	0.2	0.2	0.15	0.15

L'entropie de cette source, calculée selon la définition est

Interprétations. Exemple

Soit une source S définie par

X	a	b	c	d	e
P(X)	0.3	0.2	0.2	0.15	0.15

L'entropie de cette source, calculée selon la définition est

Interprétations. Exemple

Soit une source S définie par

X	a	b	c	d	e
P(X)	0.3	0.2	0.2	0.15	0.15

L'entropie de cette source, calculée selon la définition est

$$H(X) = -\sum_{i=1}^{5} p_i \log(p_i) =$$

Interprétations. Exemple

Soit une source S définie par

X	a	b	c	d	e
P(X)	0.3	0.2	0.2	0.15	0.15

L'entropie de cette source, calculée selon la définition est

$$H(X) = -\sum_{i=1}^{5} p_i \log(p_i) = \; -\left(0.3 \log(0.3)\right.$$

Interprétations. Exemple

Soit une source S définie par

X	a	b	c	d	e
P(X)	0.3	0.2	0.2	0.15	0.15

L'entropie de cette source, calculée selon la définition est

$$H(X) = -\sum_{i=1}^{5} p_i \log(p_i) = -(0.3 \log(0.3) - 2 \cdot 0.2 \log(0.2)$$

Interprétations. Exemple

Soit une source S définie par

X	a	b	c	d	e
P(X)	0.3	0.2	0.2	0.15	0.15

L'entropie de cette source, calculée selon la définition est

$$H(X) = -\sum_{i=1}^{5} p_i \log(p_i) = -(0.3\log(0.3) - 2\cdot 0.2\log(0.2) - 2\cdot 0.15\log(0.15)$$

Interprétations. Exemple

Soit une source S définie par

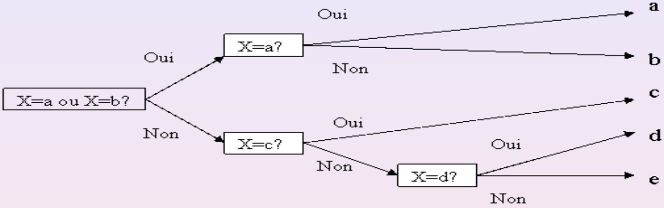
X	a	b	c	d	e
P(X)	0.3	0.2	0.2	0.15	0.15

L'entropie de cette source, calculée selon la définition est

$$H(X) = -\sum_{i=1}^{5} p_i \log(p_i) = -(0.3\log(0.3) - 2\cdot 0.2\log(0.2) - 2\cdot 0.15\log(0.15)$$

$$H(X) \simeq 2.27.$$

Un jeu



x	a	b	c	d	e
P(x)	0.3	0.2	0.2	0.15	0.15
Nb(x)	2	2	2	3	3

x	a	b	c	d	e
P(x)	0.3	0.2	0.2	0.15	0.15
Nb(x)	2	2	2	3	3

Le nombre moyen de questions nécessaires pour deviner le symbole est

$$\bar{N}b = 2(0.3 + 0.2 + 0.2) + 3(0.15 + 0.15) = 2.3.$$

Un code

a	Oui-Oui	11
b	Oui-Non	10
c	Non-Oui	01
d	Non-Non-Oui	001
e	Non-Non-Non	000