1948 :Claude Shannon pose les fondemnents mathématiques des communications numériques

Théorie de l'information est une approche essentiellement probabiliste des problèmes liés à la mesure de la quantité d'information contenue dans un message, à son stockage et à sa transmission

Champ d'application : divers

Chaine de communication

Source → Codeur → Canal → Décodeur → Récepteur

Source : émission d'un message : qu'est-ce que l'information, comment la mesurer

Codeur :
- stabiliser l'information
- la sécuriser
- la comprimer sans perte
- la dissimuler
- assurer un débit de transmission

Canal :
- Bruit
- Limites physiques

Décodeur : repérer et corriger les erreurs

Récepteur

Notions abordées :
- Incertitude et information,
- Mesure de l'information,
- Entropie de Shanon
- Entropie de lois composées et transfert d'information
- Sources d'information et compression
- Canaux : capacité et codage canal
- Codage des sources discrètes
- Codes optimaux
- Canal Gaussien
- Canal discret (CBS)
- 2ème Th de Shannon
- Codes détecteurs et correcteurs
- Codage linéaire
- Hamming (distance, code)
- Décodage au sens du maximum de vraisemblance

Recherche d'information : exemple 1

Un mot de la langue Française est choisi au hasard dans le dictionnaire.
C'est un peu vague pour pouvoir déterminer ce mot.
Quels sont les renseignements les plus " informatifs" ? :

1. le mot commence par "A" (296 pages dans le petit Robert)
2. le mot commence par "W" (2 pages dans le petit Robert)

Recherche d'information : exemple 2

Un numéro est choisi au hasard un numéro parmi 1.
Est-il nécessaire de poser une question pour savoir quel numéro a été tiré ?
Quelle quantité d'information nous apporte le fait de connaitre à postériori le résultat du tirage ?

Information et incertitude

Au sens de Shannon, la théorie de l'information repose de façon essentielle sur l'existence d'une mesure objective de la quantité d'information contenue dans un message aléatoire.

La quantité d'information d'un message se définit comme une mesure de son imprévisibilité.

Erreur de prédiction de l'issue d'une expérience aléatoire

Dans une épreuve aléatoire, chaque éventualité se produit avec une probabilité connue à **priori** .
Prédire le résultat d'une expérience entraine une probabilité d'erreur. On dit que la prédiction est d'autant plus bonne que la probabilité d'erreur est faible.
La meilleure prédiction est l'état le plus probable : si la probabilité de cet état est p_{max}, la probabilité d'erreur est $1 - p_{max}$

Erreur de prédiction de l'issue d'une expérience de Bernoulli ($X \sim \mathcal{B}(p)$)

1. $p = 1$

   $1 - p = 0$ — 0

   $p = 1$ — 1

   Prédiction : $x = 1$,   Probabilité d'erreur : $P_e = 0$

2. $p = 0$

   $1 - p = 1$ — 0

   $p = 0$ — 1

   Prédiction : $x = 0$,   Probabilité d'erreur : $P_e = 0$

3. $p = 0.3$

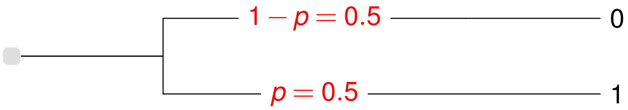
   $1 - p = .7$ — 0

   $p = .3$ — 1

   Prédiction : $x = 0$,   Probabilité d'erreur : $P_e = 0.3$

Unité d'incertitude : **le bit d'information**

$p = \frac{1}{2}$



Prédiction : incertitude maximale , Probabilité d'erreur : $P_e = 0.5$

On prend comme unité de mesure de l'incertitude, l'incertitude liée à cette expérience.
Cette unité s'appelle le **bit d'information**

on prendra soin de bien distinguer en informatique la notion de bit au sens binaire, et de bit d'information
Si dans les exemples précédents les issues possibles sont 0 et 1 (un bit au sens binaire), l'incertitude ne vaut 1 bit d'information que dans le cas $p = 0.5$.

Incertitude liée à la réalisation d'un état

l'incertitude $i(x)$ liée à un état x que peut prendre X est définie comme fonction la probabilité $p(x) = p(X = x)$: $i(x) = F(p(x))$ avec comme conditions

1. si $p(x) = 1$ alors $i(x) = 0$
2. i est positive
3. F est décroissante
4. si X et Y sont indépendantes $i(x, y) = i(x) + i(y)$ soit $F(p(x, y)) = F(p(x)p(y)) = F(p(x)) + F(p(y))$

$$\boxed{i(x) = -\alpha ln(p(x)) \ avec \ \alpha > 0}$$

exemple pour la décroissance de i : dans l'expérience " je choisis un mot dans le dictionnaire " considérons X la variable aléatoire " le mot choisit commence par la lettre .."

$p(X = a) > P(X = w) \Longrightarrow i(X = a) < i(X = w)$

ce qu'on peut généraliser à tout couple de lettres (α, β) de l'alphabet

Incertitude liée à la réalisation d'un état

L'incertitude liée à la variable $X \sim \mathcal{B}\left(p = \frac{1}{2}\right)$ étant définie égale à 1 bit d'information, on cherche alpha de telle sorte que la moyenne des incertitudes sur l'ensemble des états soit égale à 1. La probabilité de chacun des deux états étant égale à $\frac{1}{2}$ on a :

$$\frac{1}{2}\left(-\alpha ln\left(\frac{1}{2}\right)\right) + \frac{1}{2}\left(-\alpha ln\left(\frac{1}{2}\right)\right) = 1$$

soit $\alpha = \frac{1}{ln(2)}$
D'où la définition de l'entropie exprimée en bits d'information :

$$\boxed{i(X = x) = -log_2\left(p(X = x)\right)}$$

Entropie de la v.a X

Soit X une variable aléatoire à valeurs dans l'alphabet $\mathcal{A}$, On définit l'entropie de X , notée $H(X)$, comme étant la moyenne des incertitudes sur l'ensemble des états, mesurée en bits d'information

$$\boxed{H(X) = -\sum_{x \in \mathcal{A}} p(x) log_2\left(p(x)\right)}$$

On prendra comme convention $0 log_2(0) = 0$ (prolongement par continuité de $x log_2(x)$ en 0)

Entropie de la loi de Uniforme sur N états :

$$X \sim \mathcal{U}\{1, \cdots, n\}$$

$$\boxed{H(X) = log_2(N)}$$

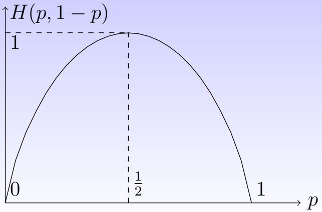
C'est l'entropie maximale possible sur les N états

Quelque soit la loi $\mathcal{L}$ de probabilité sur $\{1, \cdots, n\}$, si $X \sim \mathcal{L}\{1, \cdots, n\}$

$$H(X) \leq log_2(N)$$

Entropie de la loi de Bernouilli $X \sim \mathcal{B}(p)$

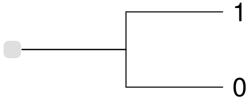
$$\boxed{H(X) = -p\log_2(p) - (1-p)\log_2(1-p)}$$



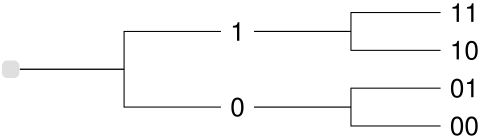
1. continue en p
2. nulle en $p = 0$ et $p = 1$
3. maximale et égale à 1 en $p = \frac{1}{2}$
4. strictement concave sur $[0, 1]$

Entropie d'une distribution 2-adique ($p(0) = \frac{1}{2}$, $p(1) = \frac{1}{2}$ quelque soit l'antériorité) : Exemples

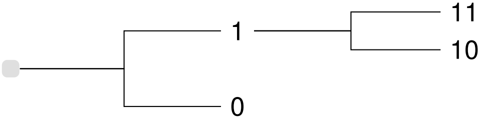
1. $H(X) = 1$, hauteur moyenne de l'arbre = 1



2. $H(X) =?$, hauteur moyenne de l'arbre = ?



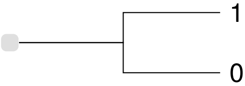
3. $H(X) =?$, hauteur moyenne de l'arbre = ?



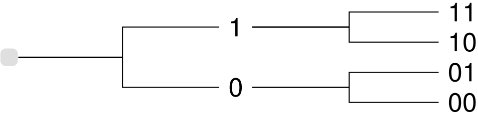
4. $H(X)$ lorsque $X \sim \mathcal{G}\left(\frac{1}{2}\right)$, hauteur moyenne de l'arbre = ?

Entropie d'une distribution non 2-adique : Exemples avec $p(0) = \frac{3}{4}$, $p(1) = \frac{1}{4}$ quelque soit l'antériorité

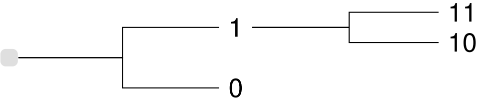
1. $H(X) = 0.811$, hauteur moyenne de l'arbre = 1



2. $H(X) = 1.887$, hauteur moyenne de l'arbre = 2



3. $H(X) = 1.014$, hauteur moyenne de l'arbre = 1.25



4. $H(X) = 4 \times 0.811 = 3.244$ lorsque $X \sim \mathcal{G}\left(\frac{1}{4}\right)$, hauteur moyenne de l'arbre > H(X)

Application à la construction d'un questionnaire

Soit une variable aléatoire X pouvant prendre 5 états $x_1, \cdots, x_5$ dont la loi est $p_1 = 0.3, p_2 = 0.2, p_3 = 0.2, p_4 = 0.15, p_5 = 0.15$. L'expérience étant réalisée on cherche à déterminer le résultat à l'aide de questions binaires.

1. calculer $H(X)$ (résultat : 2.27)
2. Construire un arbre représentant un questionnaire permettant de déterminer le résultat de l'expérience.
3. Calculer le nombre moyen de questions posées
4. qu'est-ce qu'un bon questionnaire ? comment le construire ?

L'Entropie est un minorant du nombre moyen de questions binaires à poser pour déterminer le résultat d'une expérience aléatoire

Entropie : propriétés

1. $0 \leq H(p_1, \cdots, p_N) \leq log_2(N)$
2. $H(p_1, \cdots, p_N) = 0$ lorsque la v.a X est déterministe
3. $H\left(\frac{1}{N}, \cdots, \frac{1}{N}\right) = log_2(N)$
   La loi uniforme est d'entropie maximale.
   L'incertitude moyenne (entropie) est égale à l'incertitude liée à chacune des réalisations.
   Exemple Si $N = 2^k$, $H(X) = k\ bits$:
   Exemple : l' octet (8 bits, 256 états) à quelle condition comporte-t-il 8 bits au sens de la théorie de l'information
4. l'association de plusieurs évènements fait décroitre l'entropie, la dissociation d'évènements accroit l'entropie