Entropie conditionnelle

$$H(Y|X=x) = -\sum_{y\in\mathcal{A}_Y} p(y|x)log_2\left(p(y|x)\right)$$

$$H(Y|X) = -\sum_{x\in\mathcal{A}_X} p(x)H(Y|X=x)$$

$$\boxed{H(Y|X) = -\sum_{x\in\mathcal{A}_X}\sum_{y\in\mathcal{A}_Y} p(x,y)log_2\left(p(y|x)\right)}$$

Cas particulier

1. pour le couple (X,X) on a $H(X,X)=0$
$(\forall(x_i,x_j)\in\mathcal{A}_X^2)p(x_i|x_j)=\delta_{ij}$, et donc $(\forall x_j\in\mathcal{A}_X)H(X|x_j)=0$
2. pour un couple (X,Y) indépendant $H(Y|X)=H(Y)$:
$(\forall(x_i,y_j)\in\mathcal{A}_X\times\mathcal{A}_Y)p(y_j|x_i)=p(y_j)$, et donc
$(\forall x_i\in\mathcal{A}_X)H(Y|x_i)=H(Y)$

Entropie conditionnelle : règle du chaînage

$$\boxed{\begin{array}{l} H(X,Y) = H(X) + H(Y|X) \\ H(X,Y) = H(Y) + H(X|Y) \end{array}}$$

$$\begin{aligned} H(X,Y,Z) &= H(X,Y) + H(z|(X,Y)) \\ &= H(Z|(X,Y)) + H(Y|X) + H(X) \end{aligned}$$

généralisation

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

majoration de l'entropie conditionnelle

$$\boxed{\begin{array}{rcl} H(Y|X) & \leq & H(Y)) \\ H(X|Y) & \leq & H(X) \end{array}}$$

avec égalité ssi X et Y indépendants

Information mutuelle et entropie conditionnelle

attention l'entropie conditionnelle diminue en moyenne mais pour un y particulier on peut avoir $H(X|Y=y) > H(X)$

Exemple

	$X=0$	$X=1$
$Y=0$	0	$\frac{3}{4}$
$Y=1$	$\frac{1}{8}$	$\frac{1}{8}$

$H(X) = H\left(\mathcal{B}\left(\frac{1}{8}\right)\right) = 0.544 bits$

$H(X|Y=1) = 0$ et $H(X|Y=2) = 1 bit$

mais en moyenne

$$\begin{aligned} H(X|Y) &= \frac{3}{4}H(X|Y=1) + \frac{1}{4}H(X|Y=2) \\ &= \frac{1}{4} \\ &\leq H(X) \end{aligned}$$

Information mutuelle et entropie conditionnelle

$$\begin{aligned} H(X,Y) &= H(X)+H(Y)-I(X,Y) \\ H(X,Y) &= H(X)+H(Y|X) \end{aligned}$$

$$\boxed{\begin{aligned} I(X,Y) &= H(X)-H(X|Y) \\ &= H(Y)-H(Y|X) \end{aligned}}$$

Entropie conditionnelle nulle

1. dans quel cas $I(X,Y)=0$?
2. Comment interpréter $H(Y|X)=0$ ie $I(X,Y)=H(Y)$
   1. $H(Y|X) = -\sum_{x\in\mathcal{A}_X}\sum_{y\in\mathcal{A}_Y} p(x,y)log_2\left(\frac{p(x,y)}{p(x)}\right)$
   2. $H(Y|X)=0$ ssi $(\forall x\in\mathcal{A}_X)(\forall y\in\mathcal{A}_Y)\,(p(x,y)>0\Longrightarrow p(x,y)=p(x))$
   3. $(\forall x\in\mathcal{A}_X)\,(p(x)>0\Longrightarrow(\exists! y\in\mathcal{A}_Y))p(x,y)=p(x))$
   4. Y est une fonction déterministe de X

prévisions météorologiques

On teste un système de prévisions météorologiques pour lequel on a obtenu sur un an , les fréquences de résultats suivantes :

	temps : pluie	temps : soleil
temps prévu : pluie	$\frac{1}{12}$	$\frac{1}{6}$
temps prévu : soleil	$\frac{1}{12}$	$\frac{2}{3}$

1. quelle est la probabilité que le système donne une prévision incorrecte ?
2. Calculer l'information mutuelle entre le temps prévu et le temps effectif
3. Comparer à un système de prévisions qui prédit systématiquement le soleil

test de dépistage

Un test de dépistage pour une maladie est censé discriminer entre le situations $X \in \{C, \overline{C}\}$ en donnant un resultat $Y \in \{T^+, T^-\}$

	$Y = T^+$	$Y = T^-$
$X = C$	0.07	0.01
$X = \overline{C}$	0.03	0.89

On définit l'efficacité du test par $r = \frac{I(X,Y)}{H(X)}$

1. Calculer $H(X)$, $I(X, Y)$, $H(X|Y)$, r
2. Que signifie $r = 0$?, $r = 1$?

Chaine de communication

Source

Codeur source

Codeur Canal

La source délivre un message informatif
∗ numérique
∗ analogique : Conversion Analogique Numérique (échantillonnage et quantification)

La source numérique est modélisé de façon probabiliste.
∗ suite de v.a.r indépendentes et identiquement distribuées idd
∗ chaines de Markov

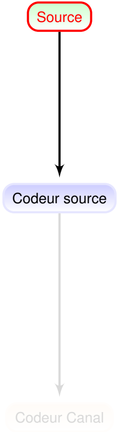
compression :
améliorer l'efficaité informationnelle en introduisant un codage préalable des signes
suppression de redondance dans le message issu de la source

taux de compression et réversibilité
codeurs avec perte :
taux de compression élévé au risque de ne pas revenir au message exact.
dégradation acceptable : MPEG, JPEG, MP3, ..
codeur sans perte : message après codage réversible quitte à avoir un taux de compression plus faible (fichier executable, ..)

Ajout de redondance pour permettre de corriger des erreurs dûes au canal,
lors de l'estimation de l'entrée du canal à partir de sa sortie

Problème de débit ? 2ème Théorème de Shannon :
on peut assurer le taux d'erreur aussi faible que voulu sans réduire à 0 le débit
pourvu que ce dernier soit inférieur à une limite liée aux caractéristiques du canal

Chaine de communication



La source délivre un message informatif
- ∗ numérique
- ∗ analogique : Conversion Analogique Numérique (échantillonnage et quantification)

La source numérique est modélisé de façon probabiliste.
- ∗ suite de v.a.r indépendentes et identiquement distribuées idd
- ∗ chaines de Markov

compression :
améliorer l'efficaité informationnelle en introduisant un codage préalable des signes
suppression de redondance dans le message issu de la source

taux de compression et réversibilité
- codeurs avec perte :
  taux de compression élévé au risque de ne pas revenir au message exact.
  dégradation acceptable : MPEG, JPEG, MP3, ..
- codeur sans perte : message après codage réversible quitte à avoir un taux de compression plus faible (fichier executable, ..)

Ajout de redondance pour permettre de corriger des erreurs dûes au canal,
lors de l'estimation de l'entrée du canal à partir de sa sortie

Problème de débit ? 2ème Théorème de Shannon :
on peut assurer le taux d'erreur aussi faible que voulu sans réduire à 0 le débit pourvu que ce dernier soit inférieur à une limite liée aux caractéristiques du canal

Chaine de communication

Source

Codeur source

Codeur Canal

La source délivre un message informatif
- ∗ numérique
- ∗ analogique : Conversion Analogique Numérique (échantillonnage et quantification)

La source numérique est modélisé de façon probabiliste.
- ∗ suite de v.a.r indépendentes et identiquement distribuées idd
- ∗ chaines de Markov

compression :
améliorer l'efficaité informationnelle en introduisant un codage préalable des signes
suppression de redondance dans le message issu de la source

taux de compression et réversibilité
- codeurs avec perte :
  taux de compression élévé au risque de ne pas revenir au message exact.
  dégradation acceptable : MPEG, JPEG, MP3, ..
- codeur sans perte : message après codage réversible quitte à avoir un taux de compression plus faible (fichier executable, ..)

Ajout de redondance pour permettre de corriger des erreurs dûes au canal,
lors de l'estimation de l'entrée du canal à partir de sa sortie

Problème de débit ? 2ème Théorème de Shannon :
on peut assurer le taux d'erreur aussi faible que voulu sans réduire à 0 le débit pourvu que ce dernier soit inférieur à une limite liée aux caractéristiques du canal

Chaine de communication

Modulateur → Canal → Démodulateur → Décodeur Canal → Décodeur Source → Récepteur

Bruit → Canal

Modulateur : adapte le message numérique à la nature physique du canal

Canal :
∗ medium de transmission (fibre, ligne de téléphone, liaison radio, ..)
∗ système de stockage (disque , RAM, support de QR code ..)

Fiabilité : pertubations

Démodulateur : Dual du modulateur :
pour un canal parfait, restitue idéalement la séquence numérique en entrée du modulateur

Décodeur Canal : Détecter et corriger les erreurs
limtes théoriques lié par
∗ au niveau de pertubation du canal
∗ à la quantité d'information produite par la source
∗ à taille du message transmis

Décodeur Source : Décompression
Vise à restituer le message source de façon exacte ou approchée.

Chaine de communication

Modulateur

adapte le message numérique à la nature physique du canal

Bruit → Canal

∗ medium de transmission (fibre, ligne de téléphone, liaison radio, ..)
∗ système de stockage (disque , RAM, support de QR code ..)

Fiabilité : pertubations

Démodulateur

Dual du modulateur :
pour un canal parfait, restitue idéalement la séquence numérique en entrée du modulateur

Décodeur Canal

Détecter et corriger les erreurs
limtes théoriques lié par
∗ au niveau de pertubation du canal
∗ à la quantité d'information produite par la source
∗ à taille du message transmis

Décodeur Source

Décompression
Vise à restituer le message source de façon exacte ou approchée.

Récepteur

Chaine de communication

Modulateur → adapte le message numérique à la nature physique du canal

Bruit → Canal

∗ medium de transmission (fibre, ligne de téléphone, liaison radio, ..)
∗ système de stockage (disque , RAM, support de QR code ..)

Fiabilité : pertubations

Démodulateur

Dual du modulateur :
pour un canal parfait, restitue idéalement la séquence numérique en entrée du modulateur

Décodeur Canal

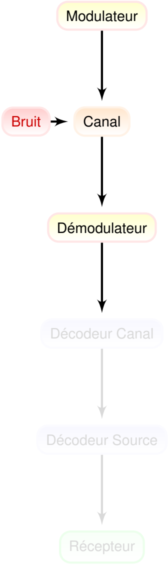
Détecter et corriger les erreurs
limtes théoriques lié par
- ∗ au niveau de pertubation du canal
- ∗ à la quantité d'information produite par la source
- ∗ à taille du message transmis

Décodeur Source

Décompression
Vise à restituer le message source de façon exacte ou approchée.

Récepteur

Chaine de communication



adapte le message numérique à la nature physique du canal

* medium de transmission (fibre, ligne de téléphone, liaison radio, ..)
* système de stockage (disque , RAM, support de QR code ..)

Fiabilité : pertubations

Dual du modulateur :
pour un canal parfait, restitue idéalement la séquence numérique en entrée du modulateur

Détecter et corriger les erreurs
limtes théoriques lié par
* au niveau de pertubation du canal
* à la quantité d'information produite par la source
* à taille du message transmis

Décompression
Vise à restituer le message source de façon exacte ou approchée.

Chaine de communication

Modulateur

adapte le message numérique à la nature physique du canal

Bruit → Canal

∗ medium de transmission (fibre, ligne de téléphone, liaison radio, ..)
∗ système de stockage (disque , RAM, support de QR code ..)

Fiabilité : pertubations

Démodulateur

Dual du modulateur :
pour un canal parfait, restitue idéalement la séquence numérique en entrée du modulateur

Décodeur Canal

Détecter et corriger les erreurs
limtes théoriques lié par
∗ au niveau de pertubation du canal
∗ à la quantité d'information produite par la source
∗ à taille du message transmis

Décodeur Source

Décompression
Vise à restituer le message source de façon exacte ou approchée.

Récepteur

Chaine de communication

Modulateur — adapte le message numérique à la nature physique du canal

Bruit → Canal
- ∗ medium de transmission (fibre, ligne de téléphone, liaison radio, ..)
- ∗ système de stockage (disque , RAM, support de QR code ..)

Fiabilité : pertubations

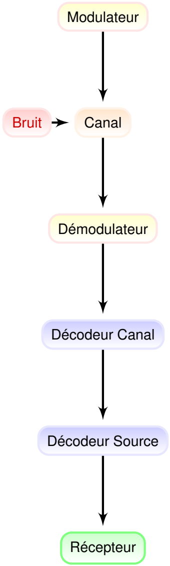
Démodulateur — Dual du modulateur :
pour un canal parfait, restitue idéalement la séquence numérique en entrée du modulateur

Décodeur Canal — Détecter et corriger les erreurs
limtes théoriques lié par
- ∗ au niveau de pertubation du canal
- ∗ à la quantité d'information produite par la source
- ∗ à taille du message transmis

Décodeur Source — Décompression
Vise à restituer le message source de façon exacte ou approchée.

Récepteur

Chaine de communication



adapte le message numérique à la nature physique du canal

∗ medium de transmission (fibre, ligne de téléphone, liaison radio, ..)
∗ système de stockage (disque , RAM, support de QR code ..)

Fiabilité : pertubations

Dual du modulateur :
pour un canal parfait, restitue idéalement la séquence numérique en entrée du modulateur

Détecter et corriger les erreurs
limtes théoriques lié par

- ∗ au niveau de pertubation du canal
- ∗ à la quantité d'information produite par la source
- ∗ à taille du message transmis

Décompression
Vise à restituer le message source de façon exacte ou approchée.

Définition d'un codage Source d'un seul état

- une **source S** émet des **symboles** d'un **alphabet source** On la considère comme une variable aléatoire $A_X = \{x_1, x_2, \cdots, x_N\}$ avec la loi de probabilité $\{p(X = x_1) = p_1, \cdots, p(X = x_N)\}$
- **Définition " code source"** C d'alphabet de codage $\mathcal{D} = \{a_1, \cdots a_D\}$ où $D = card(\mathcal{D})$.

$$\begin{array}{cccccc} C & : & \mathcal{A}_X & \longrightarrow & \mathcal{D}^\star \\ & & x & \longrightarrow & C(x) \end{array}$$

  où $\mathcal{D}^\star$ est l'ensemble des séquences de longueurs finies d'éléments de $\mathcal{D}$
- **Définition "mot de code "** A chaque symbole possible $x \in \mathcal{A}_X$, le mot de code associé est la séquence $C(x)$
- **Définition "le code"** est l'ensemble des mots de code $\{C(x) | x \in \mathcal{A}_X\}$
- **exemple** codage binaire $\mathcal{D} = \{0, 1\}$, , $\mathcal{D}^\star$ est l'ensemble des mots binaires de longueur finie

Codage source d'un seul état

- **Définition : code non singulier** On dit que le code est non singulier si $\mathcal{C}$ est injective
  à deux états distincts sont associés des mots de code distincts
  un mot de code est le mot de code d'un unique état
- un mot de $\mathcal{D}^*$ n'est pas nécessairement un mot de code

Codage source d'un seul état : compacité du code

- **Notation** on note $l(x_i) = l_i$ la longueur du mot de code $\mathcal{C}(x_i)$
- **Définition** On appelle longuer maximale du code $L = max_{i\in\{1,\cdots,N\}} l_i$
- **Définition " compacité du code "** on appelle **compacité du code** la longueur moyenne des mots du code. C'est l'espérance $E(l \circ X)$

$$\boxed{\nu = \sum_{x\in\mathcal{A}} p(x)l(x) = \sum_{i=1}^{i=N} p_i l_i}$$

- l'objectif du codage source est la compression, la suppression en partie de la redondance de la source. pour $x \in \mathcal{A}_X$ on définira la longueur de $C(x)$ à partir $p(X = x)$. L'objectif étant de minimiser la compacité du code. Quelles sont les limites à la compacité, autrement dit au choix des longueurs des mots de code.

Classification des sources d'information discrètes

- source sans mémoire : les symboles générés sont indépendant
- source avec mémoire
  - source du premier ordre : la mémoire se limite au dernier symbole émis
  - source d'ordre m : la mémoire tient compte des m symboles émis précédent

Codage source d'une suite d'états

- **Source simple** : source faite de copies indépendantes d'une même v.a X
  Une telle source génère des messages composés de réalisations indépendantes d'une même v.a X

- **Codage des séquence de longueur p d'une source simple**

$$\begin{array}{rccc} \mathcal{C}^p : & \mathcal{A}^p & \rightarrow & \mathcal{D}^* \\ & x_1 \cdots x_p & \rightarrow & \mathcal{C}(x_1) \ldots \mathcal{C}(x_p) \end{array}$$

2 approches du codage

- Combinatoire : retrouver le mot à partir de son codage (déchiffrabilité)
- Probabiliste : on s'intéresse à des propriétés ou a des grandeurs en moyenne

Notion de déchiffrabilité

Un code C est dit déchiffrable (ou non ambigu) si

$$(\forall m \in \mathcal{A}^*)(\forall m' \in \mathcal{A}^*)C(m) = C(m') \Longrightarrow m = m'$$

Un code qui n'est pas déchiffrable est dit ambigu

- si le code est déchiffrable, il existe une fonction de décodage
- Prouver qu'un code est ambigu : il suffit d'exhiber deux mots distincts qui ont le même codage
- il est plus difficile de prouver qu'un code est déchiffrable

Exemples

Pour chaque code dire s'il est déchiffrable

- $C_1 = \{0, 11, 101\}$
- $C_1 = \{00, 01, 001\}$
- $C_1 = \{0, 01, 10\}$
- $C_1 = \{0, 10, 000, 100\}$
- $C_1 = \{000100, 100101, 010101, 111000\}$
- $C_1 = \{0, 01, 11\}$

Dans un mot m' est préfixe d'un mot m s'il existe un mot m'' tel que

$$m = m' m''$$

Un code C vérifie la condition du préfixe si aucun mot de code n'est préfixe d'un autre

Code déchiffrable : méthodes classiques

méthodes classiques

- longueur unique imposée (ex code ASCII et UTF)
- séparateur
- condition du préfixe : un mot de code ne peut être le préfixe d'aucun autre mot de code
  Les codes utilisant cette technique sont dits instantanés.
  Un code instantané est déchiffrable
  Un code déchiffrable n'est pas nécessairement instantané
  $\{M_1 = 0, M_2 = 0000001\}$

Inégalité de Ktraft

Si D est le cardinal de l'alphabet utilisé pour le codage, il existe un code instantanné composé de mots de longueur $l_1, \ldots, l_N$ ssi

$$\sum_{i=1}^{i=N} \frac{1}{D^{l_i}} \leq 1$$

- CN : tout code vérifiant la condition du préfixe satisfait Kraft (dem : arbre N-aires)
- CS par récurrence : soit $L_k = \{l_1 \leq l_2 \leq \cdots \leq l_k\}$ tel que $K(L_k) < 1$ et $l_{k+1} \geq l_k$ tel que $K(L_k) \leq 1 - k^{-l_{k+1}}$, alors pour tout code préfixé d'assortiment de longueurs L_k il existe un mot de longueur l_{k+1} dont aucun préfixe n'est un mot de code.