Chapitre 1 : Introduction à la théorie de l'information.

I. 1. Introduction

La théorie de l'information, sans précision, est le nom usuel désignant la théorie de l'information de Shannon, qui est une théorie probabiliste permettant de quantifier le contenu moyen en information d'un ensemble de messages, dont le codage informatique satisfait une distribution statistique précise.

Parmi les branches importantes de la théorie de l'information de Shannon, on peut citer :

- **le codage de l'information** : on s'intéresse ici aux moyens de formaliser l'information afin de pouvoir la manipuler (principalement pour la transmettre). On ne s'intéressera donc pas au contenu mais seulement à la forme.
- **la mesure quantitative de la redondance d'un texte**,
- **la compression de données :** la compression de données ou codage de source est l'opération informatique qui consiste à transformer une suite de bits A en une suite de bits B plus courte, contenant les mêmes informations, en utilisant un algorithme particulier. La décompression est l'opération inverse de la compression.

    - ❖ Avec un algorithme de compression **sans perte**, la suite de bits obtenue après les opérations successives de compression et de décompression est strictement identique à l'originale. Les algorithmes de compression sans perte sont utilisés pour de nombreux types de données notamment des documents, des archives, des fichiers exécutables ou des fichiers texte.
    - ❖ Avec un algorithme de compression **avec pertes**, la suite de bits obtenue après les opérations de compression et de décompression est différente de l'originale, mais l'information reste sensiblement la même. Les algorithmes de compression avec perte sont utilisés pour les images, le son et la vidéo.

- **la cryptographie**. : s'attachant à protéger des messages (assurant confidentialité, authenticité et intégrité) en s'aidant souvent de secrets ou clés.
    - **Confidentialité :** permet de protéger le contenu des informations sauvegardées ou transmises sur un réseau.

- **Intégrité :** le problème de l'intégrité est le contrôle du contenu. On vent pouvoir détecter toute modification, accidentelle ou intentionnelle, des données sauvegardées ou transmises.
- **Authenticité :** s'applique à la fois aux personnes et on parle dans ce cas d'**indentification**, et aux documents, ce qui correspond à l'authentification.

La théorie de l'information décrit les aspects les plus fondamentaux des systèmes de communication.

Cette théorie s'intéresse à la construction et à l'étude de modèles mathématiques à l'aide essentiellement de la théorie des probabilités. La théorie de l'information s'est faite de plus en plus précise et est devenue incontournable dans la conception tout système de communication.

Dans ce cours, nous étudierons certains de ces modèles mathématiques, qui, bien que considérablement plus simple que les sources et les canaux physiques, permettent de donner une bonne intuition de leur comportement.

2. Système de communication

La théorie des communications s'intéresse aux moyens de transmettre une information depuis une source jusqu'à un utilisateur (Figure 1).

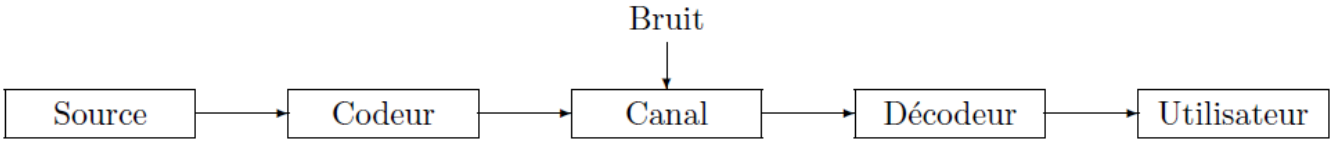

Figure 1 : Système de communication.

- **La source** peut-être de nature très variée. Il peut s'agir par exemple d'une voix, musique, image (fixe ou animée), une séquence de symboles binaires, texte…

- **Le canal,** c'est le support de transmission de l'information : peut-être par exemple, une liaison radio, fil, fibre optique, support magnétique ou optique, . . . Le canal sera généralement perturbé par un *bruit* qui dépendra de l'environnement et de la nature du

canal : perturbations électriques, ... Afin de rendre la sortie de la source compatible avec le canal, l'information doit être codée, donc,

- **Le codeur** représente l'ensemble des opérations effectuées sur la sortie de la source avant la transmission. Ces opérations peuvent être, par exemple, la modulation, la compression ou encore l'ajout d'une redondance pour combattre les effets du bruit. Enfin, l'information codée doit être décodée, donc :

- **Le décodeur** devra être capable, à partir de la sortie du canal de restituer de façon acceptable l'information fournie par la source.

Dans le but de simplifier l'étude d'un système de communication, les modèles de sources et les modèles de canaux sont séparés (la Figure 2) :

- Le but du **codeur de source** est de représenter la sortie de la source, ou information, en une séquence binaire, et cela de la façon la plus économique possible.
- Le but du **codeur de canal** et de son décodeur est de reproduire le plus fidèlement possible cette séquence binaire malgré le passage à travers le canal bruité.

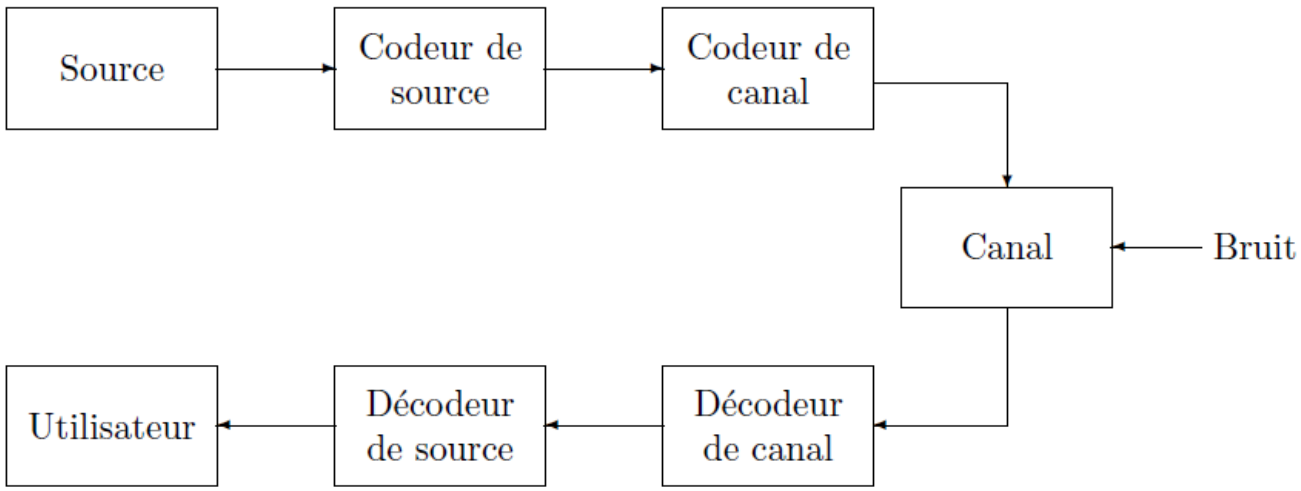

Figure 2 : Système de communication avec codeurs de source et de canal séparés.

Rappels de théorie des probabilités discrètes

1. Espace probabilisé discret

Une expérience aléatoire se décrit mathématiquement par la donnée de l'ensemble des résultats, ou issues, possibles de cette expérience, cet ensemble est appelé espace des épreuves (des évènements).

Nous noterons par U = {a1, . . ., aK}, l'espace des épreuves, et par u, la variable aléatoire dont la valeur est égale à l'issue de l'expérience.

Les différents résultats de l'expérience aléatoire sont alors : "u = a1", "u = a2", . . ., "u = aK". Une loi de probabilité sur U est la donnée des probabilités de chacun des résultats possibles : Pr(u = ak), 1 <= k <= K, que nous noterons P(ak). On a bien sur, P(ak)>= 0 pour tout k, et P(a1) + . . . + P(aK) = 1.

L'espace U muni d'une loi de probabilité P est appelé **espace probabilisé**.

2. Variable aléatoire

Une variable aléatoire (v.a. en abrégé) d'un espace probabilisé U = {a1, . . . , aK} est définie comme une application dont l'ensemble de départ est U et dont l'ensemble d'arrivée est quelconque.

Par exemple la v.a. u dont la valeur est égale à l'issue de l'expérience associée à U est l'application identité U→ U.

Une v.a. v à valeur dans l'ensemble des réels est appelée variable aléatoire réelle.

On peut définir la moyenne d'une telle v.a. par : $\mathrm{v} = \sum_{\mathrm{k}=1}^{\mathrm{K}} \mathrm{pu(ak)v(ak)}$

3. Espace probabilisé joint – Probabilités conditionnelles

3.1 Espace probabilisé joint

Pour modéliser un canal discret, nous considérons un espace des épreuves $X \times Y$, produit des deux ensembles $X = \{a1, \ldots, ak\}$, et $Y = \{b1, \ldots, bJ\}$. Le produit $X \times Y$ est l'ensemble des couples (ak, bj) pour tout k, $1 \leq k \leq K$ et tout j, $1 \leq j \leq J$. Le cardinal de $X \times Y$ est KJ.

Nous pouvons considérer cet ensemble comme un espace des épreuves et le munir d'une loi de probabilité, notée *PXY*, appelée loi de probabilité jointe de *X* et *Y* . L'espace probabilisé joint ainsi défini est noté *XY*.

3.1.1 Lois marginales

L'issue de l'expérience aléatoire est un couple, nous noterons *x* et *y* les variables aléatoires égales respectivement à la première et la seconde coordonnée de l'issue de l'expérience. La probabilité *PXY* (*ak, bj*) est donc la probabilité d'avoir simultanément *x* =*ak* et *y* = *bj* .

La probabilité d'un évènement étant égale à la somme des probabilités des issues réalisant cet évènement, la probabilité de l'évènement *x* = *ak* est donc

$$P_X(a_k) = \sum_{j=1}^{J} P_{XY}(a_k, b_j).$$

Cela définit une loi de probabilité *PX* sur *X*. De même la probabilité de l'évènement *y* = *bj* vaut

$$P_Y(b_j) = \sum_{k=1}^{K} P_{XY}(a_k, b_j).$$

Les deux lois de probabilité *PX* et *PY* ainsi définies sont appelées lois marginales de *PXY*.
Pour un espace probabilisé joint *XY*, avec *X* = {*a*1, . . . , *aK*} et *Y* = {*b*1, . . . , *bJ*}, la moyenne de la v.a. réelle *v* se définit par :

$$\bar{v} = \sum_{k=1}^{K} \sum_{j=1}^{J} P(a_k, b_j) v(a_k, b_j).$$

3.2 Probabilité conditionnelle

On suppose que $P(ak) > 0$, la probabilité conditionnelle pour que $y = bj$ sachant que $x = ak$, est définie par :

$$P_{Y|X}(b_j \mid a_k) = \frac{P_{XY}(a_k, b_j)}{P_X(a_k)}.$$

De façon symétrique, nous définissons la probabilité conditionnelle de $x = ak$ sachant $y = bj$ par :

$$P_{X|Y}(a_k \mid b_j) = \frac{P_{XY}(a_k, b_j)}{P_Y(b_j)}.$$

Les évènements $x = ak$ et $y = bj$ sont dit statistiquement indépendant si :

$$P_{YX}(a_k, b_j) = P_X(a_k) P_Y(b_j).$$

Si cette égalité est vraie pour tout couple de XY, alors les espaces X et Y sont dit statistiquement indépendant. Nous parlerons alors d'espace probabilisé produit.