le cnam

Codes et Automates finis

Cours 2 Problèmes de transmission

- Avant propos sur les probabilités discrètes

Le premier point est de préciser la notion d'événement. Nous ne développons pas ici de théorie mathématique précise et présentons quelques exemples significatifs de ce que nous souhaitons transmettre dans ce paragraphe sur les probabilités.
Si on lance un dé, on peut considérer l'événement A : "on observe la face contenant le numéro 1".
On peut aussi considérer l'événement B : "on observe une face contenant un nombre pair".
Si on lance une pièce de monnaie, elle peut retomber sur le côté pile ou sur le côté face. Les deux événements associés sont Pile : "la pièce retombe sur le côté pile" et Face : "la pièce retombe sur le côté face". On remarque que ces deux événements s'excluent mutuellement.
Un canal de transmission est un dispositif physique qui permet de faire passer une certaine information d'un lieu donné à un autre qui peut être très éloigné du premier. Dans le cas très simple où l'on cherche à envoyer une information élémentaire constituée d'un simple bit, représenté par le nombre 0 ou le nombre 1 pour fixer les idées, deux événements à considérer sont d'une part "on envoie 0 et on reçoit 0" et d'autre part "on envoie 0 et on reçoit 1". Dans ce dernier cas, la canal de transmission connaît une défaillance. Ces deux événements s'excluent là encore. Si on les nomme A et B, on peut noter cette propriété sous la forme $B = \overline{A}$.

- Probabilité d'un événement

Un événement A étant donné, on recommence l'expérience conduisant ou non à l'apparition de cet événement un grand nombre N de fois, pour savoir s'il survient ou pas. On compte le nombre de fois où l'événement A se produit. On étudie ensuite la limite du rapport entre le nombre de fois où l'événement A se produit et le nombre N d'expériences. Si cette limite existe quand N tend vers l'infini, elle définit la probabilité $P(A)$, notée parfois aussi P_A, de l'événement A. On a toujours $0 \leq P(A) \leq 1$. Pour deux événements A et $\overline{A}$ qui s'excluent mutuellement, on a $P(A) + P(\overline{A}) = 1$.
Dans le cas d'un dé équilibré, la probabilité de voir apparaître la face numéro j est donnée par $P(\{j\}) = \frac{1}{6}$. La probabilité de voir apparaître une face comportant un nombre pair vaut $P(\{2,4,6\}) = \frac{1}{2}$.
Dans le cas d'une pièce de monnaie ordinaire, la probabilité de voir apparaître le côté pile vaut $P(\text{Pile}) = \frac{1}{2}$ et celle de tomber sur le côté face est donnée par $P(\text{Face}) = \frac{1}{2}$. Ces deux événemts s'excluent mutuellement et $P(\text{Pile}) + P(\text{Face}) = 1$.
Pour un canal de transmission, on note p la probabilité de l'événement $A = \{0 \to 1\}$. La probabilité de l'événement contraire $\overline{A} = \{0 \to 0\}$ est donc donnée par $P(\overline{A}) = 1 - p$. Dans la pratique, le canal de transmission est défaillant avec une probabilité p qui est un nombre assez

petit. Des valeurs typiques sont par exemple $p = 10^{-2}$ ou $p = 10^{-3}$. Dans ce cas, le canal est défaillant en moyenne une fois sur cent ou une fois sur mille.

- Probabilités totales

La relation $P(A) + P(\overline{A}) = 1$ entre deux événements contradictoires A et $\overline{A}$ se généralise à l'ensemble des événements élémentaires $\{j\}$ [qui peuvent être éventuellement en nombre infini]. On a la relation $\sum_j P(\{j\}) = 1$.
Par ailleurs, si A et B sont deux événements quelconques, on note $A \cup B$ leur réunion et $A \cap B$ leur intersection. On a la relation $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
Pour le cas du lancer de dés et les événements A : "on observe une face contenant un nombre pair" et B : "on observe une face contenant un nombre premier", on vérifie la relation précédente sans difficulté.

- "Le hasard est égal"

Cette expression est utilisée dans la lettre de Blaise Pascal à Pierre de Fermat en 1654, document qui fonde les bases du calcul des probabilités. L'idée est que tous les événements élémentaires ont la même probabilité. Dans ce cas, la probabilité d'un événement A est le rapport entre le nombre de cas favorables où A survient, divisé par le nombre de tous les cas possibles. Nous écrivons $P(A) = \frac{\text{nombre de cas favorables où } A \text{ survient}}{\text{nombre de cas possibles}}$. Avec cette relation, le calcul des probabilités se transforme en un calcul de dénombrement.

- Événements indépendants

Deux événements A et B sont dits indépendants lorsque la réalisation de A ne dépend pas de la réalisation de B et réciproquement, lorsque la réalisation de l'événement B ne dépend pas de la réalisation de A. Pour deux événements indépendants, on a $P(A \cap B) = P(A)\, P(B)$.
Si on lance deux dés par exemple, l'événement "le premier dé affiche le nombre j" est indépendant de l'événement "le second dé affiche le nombre k". La probabilité de l'événement "le premier dé affiche le nombre j et le second dé affiche le nombre k" est donc égale au produit des probabilités, soit $\frac{1}{36}$ dans le cas de deux dés équilibrés.
Si on lance trois fois de suite la même pièce de monnaie, on peut faire l'hypothèse raisonnable que ces lancers sont indépendants et le résultat de chaque lancer est indépendant des deux autres. On a ainsi par exemple $P(\text{Pile, Face, Pile}) = \left(\frac{1}{2}\right)^2 = \frac{1}{8}$.
Pour un canal de transmission de probabilité de défaillance élémentaire égale à p, on a $P(\{00 \to 00\}) = (1-p)^2$, $P(\{00 \to 01\}) = P(\{00 \to 10\}) = p\,(1-p)$ et $P(\{00 \to 11\}) = p^2$.

- Probabilité conditionnelle

Il s'agit de calculer une probabilité dans le cas particulier où on a de l'information sur ce que est advenu au moment de l'expérience.
Si par exemple on lance deux dés et que la somme des valeurs affichées est égale à 6, on peut se poser la question de savoir quelle est la probabilité d'avoir observé le tirage $(3,3)$. Il y a maintenant cinq cas possibles seulement (et non 36 !) et un cas favorable parmi les cinq. La probabilité conditionnelle recherchée est donc égale à $\frac{1}{5}$.
De façon générale, la probabilité de l'événement A, sachant l'événement B, est notée $P(A|B)$. On a la relation $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Si les événements A et B sont indépendants, on a $P(A|B) = P(A)$ et $P(B|A) = P(B)$.
Réciproquement, si $P(A|B) = P(A)$ ou $P(B|A) = P(B)$, alors les deux événements A et B sont indépendants et $P(A \cap B) = P(A)\,P(B)$.

- Duplication de l'information avant la transmission

On souhaite envoyer le message $a = 0$. On commence par dupliquer cette information et on forme ainsi un mot de deux symboles : $u = 00$. On envoie ce mot à travers le canal de transmission et on reçoit un mot noté v, formé de deux symboles et qui prend l'une des quatre valeurs suivantes : $v \in \{00, 01, 10, 11\}$.
Quelle est la probabilité de réception de chacun de ces messages ? On fait l'hypothèse que les deux bits d'information sont modifiés de façon indépendante. La modification de l'un des bits d'information n'affecte pas ce qui peut advenir pour l'autre bit. La probabilité jointe est alors le produit des probabilités de chacun des événements. Ainsi,
$P(v = 00) = (1-p)^2$, $P(v = 01) = P(v = 10) = p\,(1-p)$ et $P(v = 11) = p^2$.
Cette duplication de l'information permet de détecter une erreur de transmission. En effet, si on reçoit $v = 01$ ou $v = 10$, on est certain de l'existence d'une erreur de transmission. Mais on ne sait alors rien sur le contenu du message envoyé !
On fera attention au fait que toutes les erreurs de transmission ne sont pas forcément détectées. En effet, si on reçoit $v = 00$, il est naturel de supposer que le message envoyé est $a = 0$. Mais avec la probabilité p^2, l'envoi du message $a = 1$ n'est pas à exclure. En effet, $a = 1$ se code $u = 11$. Avec deux erreurs de transmission, on peut recevoir $v = 00$.

- Contrôle de parité : ajout d'un bit de contrôle

On cherche dans cet exemple à envoyer un message a composé de trois bits. Il peut donc prendre *a priori* l'une des huit valeurs suivantes : $a \in \{000, 001, 010, 011, 100, 101, 110, 111\}$.
On forme le mot envoyé u en adjoignant au message a un "bit de contrôle" qui vaut 0 ou 1 selon que le nombre de "1" du message a est pair ou impair. On a ainsi le tableau suivant :

a	000	001	010	011	100	101	110	111
u	0000	0011	0101	0110	1001	1010	1100	1111

Chaque mot envoyé u comporte un nombre pair de fois le symbole "1". Si on reçoit un mot v qui comporte un nombre impair de "1", on est certain qu'il y a eu une erreur de transmission. Mais si deux erreurs de transmission se produisent, le nombre de "1" du mot reçu reste pair et cette double erreur n'est pas détectée.

- Codage et décodage

Rappelons sous forme symbolique la problématique générale d'un problème de transmission.

$$\underset{\text{message}}{a} \xrightarrow{\text{codage}} \underset{\text{mot envoyé}}{u} \xrightarrow{\text{canal bruité}} \underset{\text{mot reçu}}{v} \xrightarrow{\text{décodage}} \underset{\text{message décodé}}{\widetilde{a}}$$

Processus de transmission d'un message

Nous allons voir un exemple où le décodage $v \longmapsto \widetilde{a}$ est un processus "naturel", ce qui n'était pas le cas pour la simple duplication de l'information.

- Double répétition

Le message est $a = 0$ ou $a = 1$ composé d'un seul bit d'information. Le codage φ transforme le message a en $u = \varphi(a)$, mot ensuite envoyé à travers le canal. La double répétition consiste à dupliquer deux fois l'information : $\varphi(0) = 000$ et $\varphi(1) = 111$. Le mot envoyé u appartient à un ensemble à deux éléments composé de mots de trois bits: $u \in \{000, 111\}$.

On suppose ici que le canal est symétrique : lors de la transmission, les probabilités de défaillance $P(0 \longrightarrow 1)$ et $P(1 \longrightarrow 0)$ sont égales à un même nombre p tel que $0 < p < 1$. Le mot reçu v est alors un mot de trois bits tout à fait arbitraire ! Si on a envoyé $u = 000$ pour fixer les idées, on a le tableau suivant :

valeur du mot reçu v	commentaire	probabilité
000	transmission correcte	$(1-p)^3$
001	une erreur	$p\,(1-p)^2$
010	une erreur	$p\,(1-p)^2$
100	une erreur	$p\,(1-p)^2$
011	deux erreurs	$p^2\,(1-p)$
101	deux erreurs	$p^2\,(1-p)$
110	deux erreurs	$p^2\,(1-p)$
111	trois erreurs	p^3

Dans six cas sur huit, une erreur peut être détectée. Mais si on reçoit la chaîne $v = 111$ alors que $u = 000$ a été envoyé, on fera une erreur d'interprétation. Mais avec une probabilité p^3 très petite devant la probabilité de défaillance du canal. Par exemple, si $p = 10^{-2}$, le canal est défaillant une fois sur cent alors que l'erreur d'interprétation se produit en moyenne une fois sur un million.

On peut aussi constater [exercice !] que la somme des probabilités de la colonne de droite du tableau précédent est bien égale à 1.

Ayant pris en compte ces événements rares, le décodage $v \longrightarrow \widetilde{a}$ est très simple :

valeur du mot reçu v	message décodé $\widetilde{a}$
000	0
001	0
010	0
100	0
011	1
101	1
110	1
111	1

Attention ! Le fait de disposer d'une fonction de décodage comme celle du tableau ci-dessus n'empêche pas les erreurs ! Dans l'exemple précédent, la probabilité d'une erreur d'interprétation pour l'ensemble du processus, c'est à dire $\widetilde{a} \neq a$, est égale à

$p^3 + 3\,p^2\,(1-p) = 3\,p^2 - 2\,p^3 \simeq 3\,p^2$, soit de l'ordre de 0,0003 si $p = 10^{-2}$. Cette probabilité n'est pas nulle, mais elle est beaucoup plus petite que la probabilité p de défaillance. Le fait de

répéter deux fois l'information avant de l'envoyer a permis un gain de fiabilité.

- Nomenclature pour le codage

En général, le message est composé de k bits : $a \in (\mathbb{F}_2)^k$. Le mot envoyé u est composé de n bits : $u \in (\mathbb{F}_2)^n$, avec bien sûr $n \geq k$. Le nombre n est appelé "longueur du code" et le nombre $r = n - k$ le "nombre de bits de contrôle".
Le codage, ou fonction de codage, est une application φ définie de $(\mathbb{F}_2)^k$ et à valeurs dans $(\mathbb{F}_2)^n$. A tout message $a \in (\mathbb{F}_2)^k$, on associe de façon unique un mot envoyé $u = \varphi(a) \in (\mathbb{F}_2)^n$.

- Codage non linéaire de Hadamard

Avant d'étudier les codages linéaires, nous donnons dans ce chapitre d'introduction un exemple de codage non linéaire, simplement pour ne pas oublier que l'univers du codage est très riche !
On se place dans le cas $k = 2$ et $n = 4$. On construit l'ensemble des mots du code, c'est à dire $\{\varphi(a), a \in (\mathbb{F}_2)^k\} \subset (\mathbb{F}_2)^n$ *via* un processus relativement compliqué. On considère d'abord la matrice de Hadamard $H \equiv \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. A partir des deux lignes de H, on forme d'abord deux mots de quatre bits en faisant suivre les deux symboles par leurs opposés :
$(1, 1) \longrightarrow (1, 1, -1, -1)$ et $(1, -1) \longrightarrow (1, -1, -1, 1)$. Puis on prend les opposés de chacune de ces suites de quatre nombres : $(1, 1, -1, -1) \longrightarrow (-1, -1, 1, 1)$ et
$(1, -1, -1, 1) \longrightarrow (-1, 1, 1, -1)$. On obtient ainsi quatre mots formés des symboles ± 1 :
$(1, 1, -1, -1), (-1, -1, 1, 1), (1, -1, -1, 1), (-1, 1, 1, -1)$. On transforme ensuite le symbole "-1" en "0" pour ces quatre mots et on obtient quatre chaînes de quatre bits : 1100, 0011, 1001 et 0110.
On définit le codage φ par les images des quatre messages possibles dans $(\mathbb{F}_2)^k$:
$00 \longmapsto \varphi(00) = 1100$, $01 \longmapsto \varphi(01) = 0011$, $10 \longmapsto \varphi(10) = 1001$ et $11 \longmapsto \varphi(11) = 0110$.
On a alors la propriété suivante : deux mots différents de l'ensemble des mots du code (on dit aussi du "code") $\{1100, 0011, 1001, 0110\}$ diffèrent d'au moins deux bits. Un tableau à double entrée permet d'effectuer la vérification.

	1100	0011	1001	0110
1100	0	4	2	2
0011	4	0	2	2
1001	2	2	0	4
0110	2	2	4	0

- Un exemple de codage linéaire

On reste dans le cadre de messages de deux bits ($k = 2$) et d'un code formé de mots de quatre bits ($n = 4$). On donne une information partielle sur la fonction de codage $\varphi : (\mathbb{F}_2)^k \longrightarrow (\mathbb{F}_2)^n$: $\varphi(10) = 1010$ et $\varphi(01) = 0101$. On a simplement dupliqué les deux mots concernés.
On impose de plus à la fonction de codage d'être linéaire. Dans ce cas, l'image par la fonction de codage φ d'une combinaison linéaire de bits est égale à la combinaison linéaire des images avec les mêmes coefficients.
Avant d'écrire sous forme mathématique la condition précédente, nous devons d'abord introduire la somme $a + a'$ de deux couples de bits $a = (\varepsilon, \widetilde{\varepsilon}) \in (\mathbb{F}_2)^k$ et $a' = (\varepsilon', \widetilde{\varepsilon}') \in (\mathbb{F}_2)^k$:

$(\varepsilon, \widetilde{\varepsilon}) + (\varepsilon', \widetilde{\varepsilon}') = (\varepsilon + \varepsilon', \widetilde{\varepsilon} + \widetilde{\varepsilon}')$. Puis le produit $\alpha\, a$ du nombre $\alpha \in \mathbb{F}_2$ par le mot de deux lettres $a = (\varepsilon, \widetilde{\varepsilon}) \in (\mathbb{F}_2)^k$: $\alpha\,(\varepsilon, \widetilde{\varepsilon}) = (\alpha\,\varepsilon, \alpha\,\widetilde{\varepsilon})$. La linéarité prend alors la forme algébrique suivante : $\forall a, a' \in (\mathbb{F}_2)^k,\ \forall \alpha, \alpha' \in \mathbb{F}_2,\ \varphi(\alpha\, a + \alpha'\, a') = \alpha\, \varphi(a) + \alpha'\, \varphi(a')$. On peut aussi l'exprimer *via* les deux relations $\varphi(a+a') = \varphi(a) + \varphi(a')$, $\forall a, a' \in (\mathbb{F}_2)^k$ et $\varphi(\alpha\, a) = \alpha\, \varphi(a)$, $\forall \alpha \in \mathbb{F}_2$, $\forall a \in (\mathbb{F}_2)^k$.

En particulier, comme $(0, 0) = 0\,(1, 0)$, la linéarité impose la relation suivante
$\varphi(0, 0) = 0\,(1, 0, 1, 0) = (0, 0, 0, 0)$. On l'écrit avec une notation plus légère $\varphi(00) = 0000$.
De plus, comme $(1, 0) + (0, 1) = (1, 1)$, on a
$\varphi(1, 1) = \varphi(1, 0) + \varphi(0, 1) = (1, 0, 1, 0) + (0, 1, 0, 1) = (1, 1, 1, 1)$. On écrit cette relation sous la forme plus compacte $\varphi(11) = 1111$.

La fonction de codage φ est fialement définie par les quatre relations
$00 \longmapsto \varphi(00) = 0000$, $01 \longmapsto \varphi(01) = 0101$, $10 \longmapsto \varphi(10) = 1010$ et $11 \longmapsto \varphi(11) = 1111$.
On constate qu'elle est différente de celle introduite au paragraphe précédent. On particulier, la linéarité impose $00 \longmapsto 0000$. Cette relation n'est pas satisfaite pour le codage de Hadamard, qui n'est donc pas linéaire et est qualifié pour cette raison de "non linéaire".

On a pour le codage linéaire introduit dans ce paragraphe une propriété identique à celle du codage de Hadamard : deux mots différents du code $\{0000, 0101, 10101\,1111\}$ diffèrent d'au moins deux bits. Nous laissons la preuve en exercice au lecteur.

Exercices

- Dé truqué

On considère un dé à six faces numérotées de 1 à 6 tel que la probabilité de tomber sur la face numéro j est proportionnelle au nombre j.
Que valent les probabilités $P(\{j\})$ pour les différentes valeurs de j ? $\left[\frac{j}{21}\right]$

- Probabilités conditionnelles

On s'intéresse à une classe composé de filles et de garçons qui peuvent ou pas être germanistes.

	fille	garçon
apprennent l'allemand	10	7
n'apprennent pas l'allemand	4	9

Calculer de deux façons différentes la probabilité qu'une fille apprenne l'allemand. $\left[\frac{5}{7}\right]$

- Jeu de dés

On joue aux dés avec cinq dés de couleurs différentes qui ont chacun six faces. Les tirages sont équiprobables et indépendants.

a) Quel est le nombre total de tirages ?

b) Quelle est la probabilité d'un tirage donné ?

c) Quelle est la probabilité p de tirer une séquence qui comporte au moins trois faces identiques ? $\left[p = \frac{1}{6^5}\left(C_5^3.6.5.4 + C_5^3.6.5 + C_5^4.6.5 + 6\right) = \frac{1}{6^4}(200 + 50 + 25 + 1) = \frac{276}{6^4}\right]$

d) Quelle est la probabilité p' de tirer une séquence qui comporte au plus deux faces identiques ? $\left[p' = \frac{1}{6^5}\left(6.5.4.3.2 + C_5^2.6.5.4.3 + C_6^2.C_5^2.C_3^2.4\right) = \frac{1}{6^4}(120 + 600 + 300) = \frac{1020}{6^4}\right]$

e) Établir la cohérence des résultats pour les questions c) et d). $\left[p+p' = \frac{276+1020}{6^4} = \frac{1296}{6^4} = 1\right]$

- Paradoxe des anniversaires

On se place pour simplifier avec un calendrier qui ne contient aucune année bissextile.

a) Quelle est la probabilité que deux personnes choisies au hasard n'aient pas la même date d'anniversaire ?

b) Quelle est la probabilité que trois personnes choisies au hasard n'aient pas la même date d'anniversaire ?

c) Même question avec quatre personnes choisies au hasard.

d) On se donne un nombre N de personnes avec $2 \leq N \leq 365$. Quelle est la probabilité pour qu'aucune de ces N personnes n'aient la même date d'anniversaire ?

e) Application numérique. À partir de combien de personnes réunies ensemble a-t-on une probabilité supérieure à $\frac{1}{2}$ d'en compter deux qui ont la même date d'anniversaire ? [23]

- Transmission d'un mot de quatre bits (d'après Jacques Vélu)

On se donne un mot $M = 1010$ de longueur 4 composé de bits 0 ou 1. On envoie ce mot M à travers un canal bruité symétrique où les probabilités de transition $P(0 \longrightarrow 1)$ et $P(1 \longrightarrow 0)$ sont égales à un nombre p avec $0 < p < 1$.

a) Quelles est la probabilté de recevoir un mot de quatre lettres arbitraire ?

b) Quelles sont les probabilités p_0, p_1, p_2, p_3 et p_4 pour que le mot M soit transmis avec 0, 1, 2, 3 ou 4 erreurs ?

c) Montrer que les probabilités précédentes ne dépendent pas du mot M initialement choisi.

d) Calculer les probabilités précédentes de façon approchée pour $p = 10^{-2}$.

$[p_0 \simeq 0.96,\ p_1 \simeq 0.04,\ p_2 \simeq 6\ 10^{-4},\ p_3 \simeq 4\ 10^{-6},\ p_4 = 10^{-8}]$

e) Quelle est la somme de ces probabilités ? [1]

- Canal symétrique (d'après Jacques Vélu)

On se donne un canal binaire symétrique de bande passante 512 kilobits par seconde. La probabilité d'erreur dans la transmission d'un bit est de 1%.

Combien de bits faux sont transmis en moyenne au bout d'une heure ? [dix huit millions]

- Bit de parité pour un message de trois bits

On complète un message $a \in (\mathbb{F}_2)^3$ par un bit de parité pour obtenir un mot envoyé $u \in (\mathbb{F}_2)^4$. On a $u_j = a_j$ pour $1 \leq j \leq 3$ et $\sum_{j=1}^{4} u_j = 0$.

a) Montrer que pour un message qui subit un nombre impair de modifications, l'erreur est détectée.

b) Parmi les messages mal transmis, quelle est la proportion moyenne de ces messages qui ne sont pas détectés ? $\left[\frac{4}{11}\right]$