CHAPITRE I

ESPACES DE PROBABILITES

A - Rappels de cours

1. Construction de l'espace de probabilités

En termes d'ensembles, le hasard se formalise à partir de :

- **l'expérience aléatoire** (c'est à dire une épreuve dont les résultats sont soumis au hasard) et l'ensemble $\Omega=\{\omega\}$ de tous les résultats qu'elle génère.

  C'est, par exemple, le cas du jet d'un dé dont le résultat conduit au choix au hasard d'un numéro parmi les six faces possibles, c'est à dire d'un point ω dans l'espace $N_6^* = \{1,2,3,4,5,6\}$.

  C'est aussi le cas de l'observation de la durée de vie d'un équipement qui conduit au choix au hasard d'un nombre réel ω dans l'espace R^{+*}.

  Ainsi Ω peut-il être discret ou continu, et au delà unidimensionnel ou multidimensionnel.

- les sous-ensembles $A \in P(\Omega)$ qui vérifient une propriété donnée et qui forment les **événements aléatoires**.

  Par exemple, « une face paire » obtenue suite au jet d'un dé, événement A qui s'identifie avec le sous-ensemble A=$\{2,4,6\}$ de N_6^*.

  Autre exemple, « fonctionner plus de 1000 h » pour le cas d'un équipement, ce qui constitue un événement A décrit par le sous-ensemble $\{x \in R^{+*} / x \geq 1000\ h\}$.

- **la probabilité** p applicable aux événements aléatoires ci-dessus et définie de façon axiomatique par toute application de $P(\Omega)$ sur $[0,1]$ vérifiant les conditions :

a) $\mathrm{P(A)} \geq 0, \forall A \in P(\Omega)$

b) $\mathrm{P}(\Omega) = 1$

c) $\mathrm{P}(\bigcup_i A_i) = \sum_i p(A_i)$ si $A_i \cap A_j = \varnothing, \forall (i,j), i \neq j$

Le triplet $(\Omega, P(\Omega), p)$ ainsi formé est un exemple d'espace dit *mesurable*.

2. La probabilité uniforme

Parmi toutes les probabilités possibles, la probabilité uniforme p_0 qui correspond à l'équiprobabilité des résultats est la plus courante. Concrètement, cette probabilité est définie par $p_0(A) = \dfrac{card(A)}{card(\Omega)}$, c'est à dire le rapport entre le nombre de cas favorables et le nombre de cas possibles.

Cette notion de probabilité uniforme sur les espaces discrets est extensible aux espaces continus suivant le schéma $\frac{card(A)}{card(\Omega)} --> \frac{l(A)}{l(\Omega)} --> \frac{aire(A)}{aire(\Omega)} --> \frac{volume(A)}{volume(\Omega)} --> ..$

3. Indépendance et incompatibilité

Dans le cadre de la mise en œuvre des formules ensemblistes concernant les probabilités, les deux définitions ci-dessous, sont particulièrement utiles et importantes compte tenu de la confusion possible entre indépendance et incompatibilité :

- A_1 et A_2 sont des **événements indépendants** si la réalisation de l'un n'a aucune influence sur la réalisation de l'autre et réciproquement.
- A_1 et A_2 sont des **événements incompatibles** s'ils ne peuvent pas avoir lieu simultanément (soit, en termes ensemblistes ($A_1 \cap A_2 = \varnothing$).

4. Probabilités conditionnelles

La notion d'événements et de probabilités conditionnelles complète les éléments ci-dessus. Plus précisément, la probabilité de l'événement « *B sachant la réalisation de l'événement A* », notée « *B/A* » et lue « *B si A* », est définie par la formule $p(B/A) = \frac{p(A \cap B)}{p(A)}$.

L'expérience du tirage sans remise de deux boules dans une urne comprenant, par exemple, deux boules rouge et une boule noire, illustre assez bien cette notion d'événement conditionnel. Désignant respectivement par A et B, les deux événements « *1ère boule tirée rouge* » et « *2ème boule tirée rouge* », on a immédiatement p(A)=2/3. Par contre, l'embarras est plus grand quand on passe à la seconde boule. Si on ignore le résultat du 1er tirage, p(B) qui résulte d'un compromis entre une première boule tirée rouge et une première boule tirée noire , reste égale en absence d'informations à 2/3 (c'est à dire le même résultat que dans l'hypothèse d'un tirage avec remise).

L'événement conditionnel « *B sachant connu le résultat du premier tirage* » lève totalement cette ambiguïté. Si cette première boule tirée est rouge, on a par exemple, p(B/A)= $\frac{1}{2}$, puisqu'avant le second tirage, il ne reste plus qu'une boule rouge et une boule noire dans l 'urne.

5. Les formules utiles sur les probabilités

Plus largement, les formules ci-dessous, revêtent un intérêt pratique de première importance :

- $p(\bigcup_i A_i) = \sum_i p(A_i)$ pour des événements A_i incompatibles deux à deux, c'est à dire vérifiant $A_i \cap A_j = \varnothing, \forall (i,j) / i \neq j$ (*théorème dit* des *probabilités totales*).
- $p(\bigcap_i A_i) = p(A_1).p(A_2 / A_1).p(A_3 / A_1 \cap A_2)...p(A_n / A_1 \cap A_2 \cap ... \cap A_{n-1})$

(*théorème dit des probabilités composées*).

- $p(\bigcap_i A_i) = \prod_i A_i$, lorsque les A_i sont indépendants deux à deux.
- $p(A) = \sum_{i=1}^{i=n} p(H_i).p(A/H_i)$, les H_i étant supposées former un ensemble de causes incompatibles deux à deux et d'union certaine (*formule dite des probabilités totales*).
- $p(H_i / A) = \dfrac{p(H_i).p(A/H_i)}{\sum_{i=1}^{i=n} p(H_i).p(A/H_i)}$, les H_i étant supposées former ici encore un ensemble de causes incompatibles deux à deux et d'union certaine (*formule de* BAYES).

Assurant la liaison entre des probabilités « *à priori* » formulées relativement à l'apparition des causes H_i et les probabilités « *à posteriori* » desdites causes après observation d'un certain événement A, la formule de BAYES permet ainsi d'améliorer quantitativement la connaissance des aléas en fonction de l'expérimentation. Comme indiqué plus loin, elle forme une équation clef en intelligence artificielle et dans les systèmes experts.

6. Introduction aux processus aléatoires

Parmi toutes les applications nombreuses de la formule des probabilités totales, il convient de signaler l'écriture des **équations de transitions** des *processus aléatoires* et de procéder à ce sujet, à une première initiation à ces derniers.

De façon générale, on s'intéresse ici à des systèmes physiques dont les états aléatoires $E_1, E_2, .., E_i, .., E_m$ (ils peuvent être en nombre infini) sont recensés à des instants discrets $t_0, t_1, ..., t_j, ..., t_n$ ou continus t, t+Δt,… les *probabilités dites d'états* que le système soit dans l'état E_j à l'instant t_n ($\forall (j,n)$), étant notées $p_j(t_n)$.

Considérant deux instants de recensements successifs, et les *probabilités de transition* entre tous couples d'états entre ces instants, soient $p_{ij}(t_n)$ ces probabilités de passer de l'état E_i à l'instant t_{n-1} à l'état E_j à l'instant t_n, la formule des probabilités totales nous enseigne immédiatement que $p_j(t_n) = \sum_{i=1}^{i=m} p_{ij}(t_n).p_i(t_{n-1})$. La même relation s'écrit $p_j(t+\Delta t) = \sum_{i=1}^{i=m} p_{ij}(t).p_i(t)$ dans l'hypothèse d'instants de recensement continus t, t+Δt.

Ces relations qui forment les équations de transition du processus, revêtent une formulation simplifiée lorsqu'on suppose que les probabilités de transitions entre les états sont indépendantes de l'instant considéré (en d'autre termes $p_{ij}(n) \equiv p_{ij}, \forall (i,j)$).

On a en effet, $p_j(t_n) = \sum_{i=1}^{i=m} p_{ij}.p_j(t_{n-1})$, ($p_j(t+\Delta t) = \sum_{i=1}^{i=m} p_{ij}.p_{j-1}(t)$, dans le cas continu), relation dont l'écriture matricielle est :

$$P(n) = M.P(n-1) \text{ en posant } P(n) = \begin{bmatrix} P_1(n) \\ P_2(n) \\ . \\ P_m(n) \end{bmatrix} \text{ et } M = (p_{ij}).$$

On constate immédiatement que la seule connaissance des conditions initiales et des probabilités de transition suffit à maîtriser l'évolution de ces processus (qui sont dits de « MARKOV »), puisque de proche en proche on a alors P(n) = $M^n.P(0)$.

C'est la même chose lorsque les instants de recensement sont continus. En effet, la relation P(t+ $\Delta t) = M.P(t)$ qui s'écrit aussi $\dfrac{P(t+\Delta t)-P(t)}{\Delta t} = (M-I).P(t)$, forme lorsque $\Delta t \to 0$, un système différentiel du premier ordre $\dfrac{dP(t)}{dt} = \lim\limits_{\Delta t \to 0}(\dfrac{M-I}{\Delta t}).P(t)$, système dont les solutions sont connues dès lors qu'on fixe les conditions initiales P(0).

B - $\boxed{\text{Applications}}$

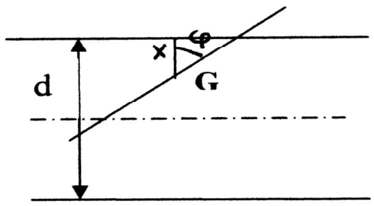
1. Sur l'extension de la probabilité uniforme dans les espaces continus

1.1 Le problème de l'aiguille de BUFFON (1777)

Enoncé : Le plan étant quadrillé par un réseau de droites parallèles distantes d'une longueur donnée d, on y jette une aiguille de longueur l (on supposera que l $\prec d$).

Quelle est la probabilité que l'aiguille coupe une ligne du réseau ?

Solution : Désignant par A l'événement « *l'aiguille coupe une ligne du réseau* », nous allons reprendre ici la formalisation correspondante de l'espace de probabilités et de l'événement considéré.



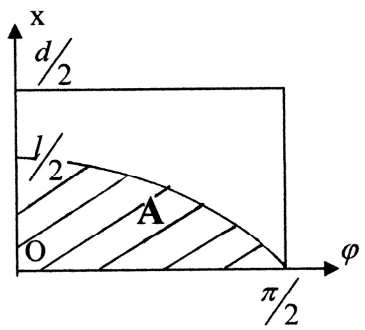
- On part d'une **expérience aléatoire** qui est le jet d'une aiguille et qui aboutit à un résultat formé par la position de ladite aiguille dans le plan.

Pour des raisons de symétrie, cette position peut être repérée par rapport à la droite la plus proche (D)

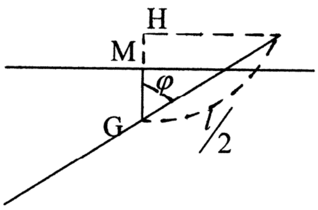
- Se résumant à un paramètre de translation tel la distance x du milieu G de l'aiguille à la droite (D) et à un paramètre de rotation, tel l'angle φ défini dans le schéma ci-dessus, le **résultat** d'un jet conduit en définitive à un couple :

$$\omega = (x,\varphi) \text{ où } 0 \le x \le \frac{d}{2} \text{ et } 0 \le \varphi \le \frac{\pi}{2}$$

(du moins en réduisant le champ de variations de φ compte tenu des symétries).



Ainsi $\Omega = \{\omega\}$ peut-il être imagé dans R ^{2} et suivant un repère (φ, x) par le pavé $\left[0,\frac{\pi}{2}\right]\times\left[0,\frac{d}{2}\right]$ représenté ci-contre. Dans ce pavé, et admettant l'extension de la **probabilité uniforme** sur R ^{2}, on peut donc écrire que p(A)= $\dfrac{aireA}{aire\Omega}$.



• Or comme on peut le constater sur le schéma ci-contre, **l'événement aléatoire** A considéré « *l'aiguille coupe une ligne du réseau* », est caractérisé par A=$\{(x,\varphi)/GM \leq GH\}$, soit, mathématiquement, A=$\left\{(x,\varphi)/x \leq \frac{l}{2}.\cos\varphi\right\}$.

Ainsi, revenant au rapport des aires représentées en page 6, a-t-on en conclusion :

$$p(A) = \frac{\int_0^{\frac{\pi}{2}} \frac{l}{2}.\cos\varphi.d\varphi}{\frac{\pi.d}{4}} = \frac{2l}{\pi.d}.$$

• En réalisant un grand nombre de fois l'expérience, on obtient une fréquence $f_A = \frac{N_A}{N}$ de l'événement A qui converge vers p(A) lorsque N$\rightarrow +\infty$ (ce phénomène de stabilité lorsque le nombre des épreuves augmente indéfiniment forme un résultat connu sous le nom de *loi empirique des grands nombres*).

Pour N grand, on a donc $f_A \approx \frac{2l}{\pi.d}$, soit $\pi \approx \frac{2l}{f_A.d}$. Or en réalisant l'expérience 5000 fois par exemple, on arrive ainsi à une valeur très proche de π soit effectivement 3,1416.

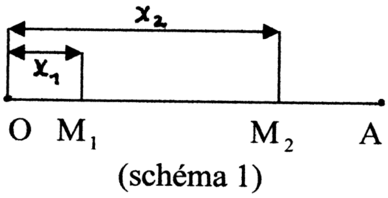
Le miracle est ici d'avoir retrouvé par un processus expérimental une **évaluation empirique** de π, convergence de résultat validant ainsi toute la démarche suivie et notamment le concept d'extension de la probabilité uniforme à un espace continu.

Le lecteur pourra consulter sur ce sujet plusieurs sites internet qui proposent par simulation interactive une illustration de ce résultat spectaculaire (entrer « *aiguille de BUFFON* » comme mot clef). A noter enfin qu'une autre version de ce problème existe sous la forme d'un quadrillage en carrés constitué de droites parallèles horizontales et verticales (carrés de côté a, par exemple), l'expérience consistant ici encore à y jeter une aiguille de longueur l supposée égale à a (pour simplifier les choses). Un tel problème est traité également sur certains sites.

1.2 Un exercice simple utilisant la probabilité uniforme

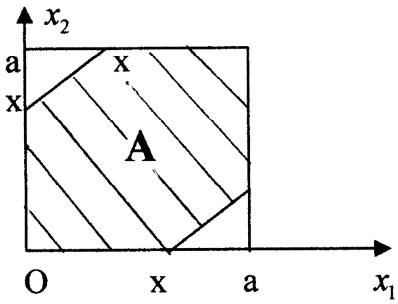
Enoncé : On choisit deux points au hasard sur un segment de longueur a. Quelle est la probabilité que ces deux points soient éloignés d'une distance inférieure ou égale à x (x donné et tel que $0 \leq x \leq a$).

Solution : L'**expérience aléatoire** qui est ici le choix de deux points génère un couple (M_1, M_2) que l'on peut caractériser par $\omega = (x_1, x_2) \in [0,a]^2$ (cf. schéma 1 ci-dessous). Ainsi $\Omega = [0,a]^2$.



(schéma 1)

Quant à l'**événement aléatoire** A dont on recherche la probabilité, il se caractérise mathématiquement par la condition d(M_1, M_2)$\leq x$, soit $\{(x_1,x_2)/|x_1 - x_2| \leq x\}$.



(schéma 2)

Dans ces conditions et raisonnant dans un repère $(0, x_1, x_2)$, on a une image de A et de Ω représentée ci-contre.

Admettant opérer un **choix uniforme** de ces deux points, il ressort dès lors que :

$$p(A) = \frac{aireA}{aire\Omega} = \frac{a^2 - (a-x)^2}{a^2} = \frac{x(2a-x)}{a^2}$$

Ce résultat est maximal quand x = a, ce à quoi on pouvait s'attendre fort logiquement.

2. Comment utiliser les formules sur les probabilités ?

De façon générale on cherchera à formuler les données et les questions sous forme ensembliste et en faisant preuve de précision et de rigueur.

2.1 De la différence entre indépendance et incompatibilité

Enoncé : On considère un jeu de 52 cartes et on procède au tirage successif de deux cartes.

Dans la première expérience, le tirage est conduit avec la remise de la première carte tirée avant le tirage de la seconde carte.

Dans la seconde expérience, les deux cartes sont tirées simultanément ce qui équivaut à un tirage dans lequel la première carte tirée n'est pas remise dans le jeu avant le tirage de la seconde carte.

Quelles sont, pour chacune des hypothèses ci-dessus, les probabilités des événements ci-dessous :

a) les deux cartes tirées sont des cœurs.
b) il y a au moins un cœur parmi ces deux cartes tirées.

Solution : L'expérience est formée ici par le tirage des deux cartes. Il est proposé de considérer les événements A_1 « *la première carte tirée est un cœur* » et A_2 : « *la seconde carte tirée est un cœur* ». Les probabilités cherchées sont donc immédiatement $p(A_1 \cap A_2)$ et $p(A_1 \cup A_2)$.

- Dans le cas d'un **tirage avec remise**, les événements A_1 et A_2 sont indépendants puisque le tirage de la première carte n'influe pas sur le résultat du tirage suivant (cette carte est remise dans le jeu). On a donc $p(A_1 \cap A_2)$=p(A_1).p(A_2).

Or p(A_1)=p(A_2)=$\frac{1}{4}$, en définitive, la réponse cherchée est $p(A_1 \cap A_2) = \frac{1}{16}$.

Par contre, si A_1 et A_2 sont indépendants, il ne saurait être question d'en conclure qu'ils sont aussi incompatibles car on peut tirer deux fois de suite un cœur et d'ailleurs $A_1 \cap A_2 \neq \emptyset$.

Dans ces conditions $p(A_1 \cup A_2) =$ p(A_1)+p(A_2)-p($A_1 \cap A_2$), soit numériquement $p(A_1 \cup A_2) = \frac{1}{4} + \frac{1}{4} - \frac{1}{16} = \frac{7}{16}$.

- Dans le cas d'un **tirage avec remise,** il n'y a plus l'indépendance entre A_1 et A_2, si bien que $p(A_1 \cap A_2)$=p(A_1).p(A_2 / A_1), avec p(A_1)=$\frac{1}{4}$ et p(A_2 / A_1)=$\frac{12}{51}$. Ainsi :

$$p(A_1 \cap A_2) = \frac{1}{4} \cdot \frac{12}{51} = \frac{1}{17}.$$

Par contre, cette modification d'hypothèse n'a pas d'effet sur l'incompatibilité et on a toujours $p(A_1 \cup A_2)$ = p(A_1)+p(A_2)-p($A_1 \cap A_2$) $= \frac{1}{4} + \frac{1}{4} - \frac{1}{17} = \frac{15}{34}$.

- Il est toujours difficile d'admettre que p(A_2)=$\frac{1}{4}$, quelle que soit l'hypothèse émise c'est à dire qu'il y ait ou non replacement de la première carte tirée avant le tirage de la seconde carte. En fait, la seule considération de l'événement A_2 et non pas de l'événement conditionnel A_2 / A_1 conduit à faire abstraction des tirages antérieurs et c'est là une explication de l'invariance susmentionnée car ces derniers peuvent aller dans un sens ou son contraire.

La *formule des probabilités totales* fournit ici une justification mathématique puisqu'en effet P(A_2) $= p(A_1).p(A_2 / A_1) + p(A_1^c) / p(A_2 / A_1^c) = \frac{1}{4} \cdot \frac{12}{51} + \frac{3}{4} \cdot \frac{13}{51} = \frac{1}{4}$ (c.q.f.d).

2.2 Une évaluation probabiliste de e

Enoncé : Une secrétaire étourdie enferme au hasard n lettres différentes dans n enveloppes d'adresses différentes. Quelle est la probabilité pour qu'une lettre au moins arrive à son véritable destinataire ?

Solution : Considérons les événements notés $A_i (1 \leq i \leq n)$ définis par « *la lettre i est dans la bonne enveloppe* ». La question posée est d'évaluer p$(\bigcup_{i=1}^{i=n} A_i)$.

Or $p(A_1 \cup A_2)$ = p(A_1)+p(A_2)-p($A_1 \cap A_2$).

Par extension à l'union de trois événements, on a aussi $p(A_1 \cup A_2 \cup A_3) = p(A_1 \cup A_2) + p(A_3) - p((A_1 \cup A_2) \cap A_3)$, soit en définitive :

$$p(A_1 \cup A_2 \cup A_3) = p(A_1) + p(A_2) + p(A_3) - p(A_1 \cap A_2) - p(A_1 \cap A_3) - p(A_2 \cap A_3) + p(A_1 \cap A_2 \cap A_3)$$

S'agissant de l'union de n événements, on a :

$$p(\bigcup_{i=1}^{i=n} A_i) = \sum_{i=1}^{i=n} p(A_i) - \sum_j \sum_{k(k \succ j)} p(A_j \cap A_k) + \sum_i \sum_j \sum_k p(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1} p(\bigcap_{i=1}^{i=n} A_i)$$

On constate immédiatement que p(A$_i$) $= \frac{1}{n}$ et que $\forall (i, j / i \neq j), p(A_i \cap A_j) = \frac{1}{n(n-1)}$.

Plus généralement p$(\bigcap_{i=1}^{i=n} A_i) = \frac{1}{n!}$.

D'autre part dans la somme $\sum_{i=1}^{i=n} p(A_i)$ on dénombre n termes. De même, la somme $\sum_{i=1}^{i=n} \sum_{j=1 (j \neq i)}^{j=n} p(A_i \cap A_j)$ comporte C_n^2 termes, et ainsi de suite pour les sommes triples……

En définitive, la probabilité cherchée est égale à :

$$p(\bigcup_{i=1}^{i=n} A_i) = n.\frac{1}{n} - C_n^2.\frac{1}{(n-1)} + ...(-1)^{n-1}.\frac{1}{n!}$$, soit après simplifications :

$$p(\bigcup_{i=1}^{i=n} A_i) = S_n = 1 - \frac{1}{2!} + \frac{1}{3!} + ...(-1)^{n-1}.\frac{1}{n!}$$. On en déduit que $\lim_{n\to\infty} S_n = 1 - \frac{1}{e}$.

En fait, cette convergence est très rapide puisque S_{10}, par exemple, est déjà très proche de la valeur limite $1 - \frac{1}{e}$.

La rapidité de cette convergence permet ainsi de comprendre pourquoi la probabilité qu'une lettre au moins parvienne à son destinataire reste sensiblement la même que n soit égal à 10 ou à 10000, ce qui est un résultat étonnant mais incontestable au plan mathématique.

3. Portée pratique des formules de BAYES et des probabilités totales

La formule de BAYES est utilisée notamment pour calculer les probabilités des causes dans les diagnostics (maladies, pannes...).

3.1 Valeur diagnostique d'un test de dépistage.

Enoncé : On choisit au hasard un individu dans la population et on le soumet à un test de dépistage. On considère les événements A « l'individu est atteint par la maladie », T^+ « le test est positif », et T^- « le test est négatif ».

On suppose que p(A)=p et que $p(T^+/A) = p(T^-/A^c) = q$, A^c désignant le complémentaire de A (p correspond à la proportion d'individus atteints par la maladie, où $0 \le p \le 1$, et q est une mesure de la fiabilité du test garantie par le laboratoire, où q est supposée être comprise entre 0,9 et 1).

Quelle est la valeur diagnostique de ce test, c'est à dire la probabilité qu'un individu dont le test est positif soit réellement atteint ?

Solution : De la formule de BAYES, il ressort que :

$$p(A/T^+) = \frac{p(A).p(T^+/A)}{p(A).p(T^+/A) + p(A^c).p(T^+/A^c)}$$

Or $p(A^c)=1-p$ et $p(T^+/A^c) = 1 - p(T^-/A^c) = 1 - q$.

Ainsi obtient-on en définitive, le résultat attendu, à savoir :

$$p(A/T^+) = \frac{p.q}{p.q + (1-p).(1-q)} = \frac{p.q}{2.p.q + 1 - p - q}$$

Comme on pouvait s'y attendre, la valeur diagnostique susmentionnée, dépend de la fiabilité du test, mais aussi de la proportion d'individus qui sont atteints par la maladie dans la population considérée.