PROBABILITES - 1ère partie

Plan

I : Définitions

1- Univers

On considère une expérience aléatoire (ou épreuve) dont le résultat ω (ou l'issue, ou la réalisation) appartient à un ensemble Ω, appelé **univers**. Le choix de Ω suppose une modélisation de l'expérience réalisée. Ω peut être :

❑ *fini* . Exemples :

- Lancer un dé et lire le numéro sorti : Ω = {1, 2, 3, 4, 5, 6}.
- Jouer au loto : Ω est par exemple l'ensemble de toutes les combinaisons possibles de 5 éléments parmi 49. Il possède $\binom{49}{5}$ éléments, soit 1 906 884.
- Lancer une pièce : Ω = {Pile, Face}

Ces ensembles font l'objet du programme de probabilité de CPGE 1ère année.

❑ *infini dénombrable*. On désigne ainsi les ensembles en bijection avec $\mathbb{N}$. En général, $\Omega = \mathbb{N}$ lui-même. On peut également choisir un tel Ω lorsque le nombre de résultats est fini, mais borné par un nombre inconnu. Les ensembles finis ou dénombrables sont appelés ensembles discrets. Exemples :

- Lancer une pièce jusqu'à ce qu'il apparaisse P. Au choix, ou bien $\Omega = \{F^n P \mid \cap \in \mathbb{N}\}$ où F^n désigne un suite de *n* F ; ou bien $\Omega = \mathbb{N}$ si on considère le résultat comme le nombre de lancers.
- Compter le nombre de voitures passant au péage de l'autoroute à Fontainebleau entre le 30 Juillet 0 h et le 1 Août 24 h. $\Omega = \mathbb{N}$
- Prendre un épi de blé et compter le nombre de grains. $\Omega = \mathbb{N}$.
- Se promener aléatoirement sur un axe de la façon suivante : i) On lance une pièce. Si elle tombe sur P, on avance d'une unité, sinon on recule d'une unité. ii) On relance la pièce. Si elle tombe sur P, on s'arrête. Sinon on recommence à l'étape i). Le résultat de l'épreuve est l'abscisse obtenue, lorqu'on s'arrête, $\Omega = \mathbb{Z}$.

Ces ensembles font l'objet du programme de probabilité de CPGE 2ème année.

❑ *infini non dénombrable*. En général, $\Omega = \mathbb{R}$ ou $\mathbb{R}^2$ ou un intervalle de $\mathbb{R}$. Exemples :

- Mesurer la durée écoulée jusqu'à observer la désintégration d'un élément radioactif. $\Omega = \mathbb{R}^+$.
- Chercher la proportion de pièces défectueuses fabriquées par une usine. Ω = [0,1].
- Lancer une pièce indéfiniment. Ω est l'ensemble des suites formées des lettres P et F, qu'on note parfois $\{P,F\}^{\mathbb{N}} = \{(x_i)_{i\in\mathbb{N}} \mid \forall\, i \in \mathbb{N},\, x_i = P \text{ ou } x_i = F\}$. Cet ensemble n'est pas dénombrable mais n'est pas non plus un intervalle de $\mathbb{R}$.

Ces ensembles ne font pas partie du programme de probabilité de CPGE.

Dans la suite de ce chapitre, Ω désigne un ensemble fini.

2- Evénement

a) Définition

On appelle événement une partie A *de* Ω.

A n'est autre qu'un ensemble de résultats, inclus dans l'ensemble Ω de tous les résultats possibles. L'ensemble des événements est donc l'ensemble des parties de Ω, noté $\mathcal{P}(\Omega)$.

Exemples :

- Tirer au dé un nombre pair ou un nombre premier. A = {2, 3, 4, 5, 6}
- Tirer au loto cinq nombres successifs. A = {(1, 2, 3, 4, 5) , (2, 3, 4 ,5, 6) , ... , (45, 46, 47, 48, 49)}
- Lancer une pièce et tomber sur Face. A = {Face}

L'événement A est **réalisé** si l'issue de l'épreuve appartient à A. En reprenant les trois exemples ci-dessus, donnons des exemples d'épreuves en indiquant respectivement si l'événement A est réalisé :

5	A réalisé
(2, 3, 5, 7, 8)	A non réalisé
Face	A réalisé

Ω s'appelle **événement certain**. En effet, par définition, l'issue de l'épreuve appartient à Ω et Ω est toujours réalisé au cours d'une épreuve. Pour une raison opposée, $\varnothing$ est appelé **événement impossible**.

Un **événement élémentaire** est constitué d'un singleton. Il est réduit à un élément. Ω est la réunion des événements élémentaires.

b) Opérations sur les événements

Les opérations sur les événements sont celles qu'on peut effectuer sur les parties de Ω, à savoir les opérations usuelles des ensembles.

❑ Réunion : $A \cup B$ est réalisé si et seulement si A ou B est réalisé, ("ou" est pris au sens inclusif, c'est-à-dire que A et B peuvent être réalisés simultanément)

❑ Intersection : $A \cap B$ est réalisé si et seulement si A et B sont réalisés simultanément. Si $A \cap B = \varnothing$, on dit que A et B sont **incompatibles**. Il est impossible de voir se réaliser simultanément ces deux événements.

On pourra vérifier que, I étant une famille d'indice quelconque, on a :

$$(\bigcap_{i \in I} A_i) \cup B = \bigcap_{i \in I} (A_i \cup B)$$

$$(\bigcup_{i \in I} A_i) \cap B = \bigcup_{i \in I} (A_i \cap B)$$

❑ Complémentaire : $\overline{A} = \complement A$ est l'ensemble des éléments de Ω qui ne sont pas dans A. Cet événement est réalisé si et seulement si A n'est pas réalisé. Pour cela, on l'appelle l'événement **contraire** de A. A et $\complement A$ sont incompatibles. On notera les propriétés suivantes dont la démonstration est laissée au lecteur :

$$\complement(A \cup B) = \complement A \cap \complement B$$

$$\complement(A \cap B) = \complement A \cup \complement B$$

$$\complement(\complement A) = A$$

On pourra vérifier que, I étant une famille d'indices quelconque, on a :

$$\complement\,(\bigcap_{i \in I} A_i) = \bigcup_{i \in I} \complement A_i$$

$$\complement\,(\bigcup_{i \in I} A_i) = \bigcap_{i \in I} \complement A_i$$

❑ Différence : A – B ou A \ B est l'ensemble des éléments de A qui ne sont pas dans B. (B n'est pas nécessairement inclus dans A). Cet événement est réalisé si et seulement si A est réalisé, mais pas B. On notera que :

$$A - B = A \cap \complement B$$

A – B et B sont incompatibles.

c) Système complet d'événements

$(A_i)_{i\in I}$ forme un système complet d'événements si et seulement si les trois conditions suivantes sont vérifiées :

i) $\forall\, i \in I,\ A_i \neq \emptyset$

ii) $\forall\, i \neq j,\ A_i \cap A_j = \emptyset$

iii) $\Omega = \bigcup_{i\in I} A_i$

Du point de vue ensembliste, $(A_i)_{i\in I}$ forme une partition de Ω.

On observe que les mêmes situations sont désignées par des mots différents suivant que l'on parle d'ensembles ou d'événements. Cela est dû au fait que, historiquement, les probabilités se sont développées indépendamment de la théorie des ensembles, à partir du XVIIème. La théorie des ensembles date de la fin du XIXème, et ce n'est qu'à partir du début du XXème que les probabilités ont connu un essor important, grâce au développement des théories des ensembles et de l'intégration. Voici un résumé de la correspondance entre le vocabulaire des ensembles et celui des probabilités.

ENSEMBLES	*PROBABILITES*
partie A de Ω	événement A de l'univers Ω
ensemble Ω	événement certain
ensemble vide ∅	événement impossible
singleton $\{x\}$	événement élémentaire $\{x\}$
ensembles disjoints	événements incompatibles
complémentaire	événement contraire
partition	système complet d'événements

3- Espace probabilisé

a) Equiprobabilité

Soit $\Omega = \{x_1, x_2, \dots, x_n\}$. On définit une loi équiprobable sur Ω, ou une loi uniforme, en posant que la probabilité est la même pour tous les événements élémentaires. La probabilité de l'événement certain étant égale à 1, cela nous amène à définir :

$$P(\{x_i\}) = \frac{1}{n}$$

$P(A) = \dfrac{\text{Card}(A)}{\text{Card}(\Omega)}$ pour tout événement A (nombre de cas favorables sur nombre de cas possibles), où Card(A) désigne le nombre d'éléments de A.

Exemples :

- Pour un dé équilibré, on posera $P(\{x\}) = \frac{1}{6}$ pour $x \in \{1, 2, 3, 4, 5, 6\}$
- Pour une pièce équilibrée, $P(\{\text{Pile}\}) = P(\{\text{Face}\}) = \frac{1}{2}$

- Au loto $P(\{x\}) = \dfrac{1}{1\,906\,884}$ pour chaque combinaison x de cinq chiffres parmi 49.

Le choix de l'équiprobabilité relève d'une modélisation d'une situation où une symétrie parfaite joue un rôle important (pièce ou dé parfaitement symétrique, boules jouant des rôles équivalents).

b) Probabilité quelconque

Plus généralement, on peut définir une probabilité quelconque sur $\Omega = \{x_1, ..., x_n\}$ à partir de la probabilité des événements élémentaires. Pour tout x_i élément de Ω, on pose $p_i = P(\{x_i\})$, où p_i est un élément de [0,1]. Pour toute partie A de Ω, on pose :

$$P(A) = \sum_{i \text{ tel que } x_i \in A} p_i = \sum_{i=1}^{n} p_i \, \mathbf{1}_A(x_i)$$

où $\mathbf{1}_A$ s'appelle fonction indicatrice de la partie A. Elle est définie de la façon suivante :

$$\begin{aligned} \mathbf{1}_A : \Omega &\to \{0,1\} \\ x &\to 1 \text{ si } x \in A \\ &\quad\ 0 \text{ si } x \notin A \end{aligned}$$

Par définition, on exige également que la probabilité de l'événement certain soit égale à 1. Il en résulte que les p_i doivent être de somme égale à 1.

$$P(\Omega) = 1 = \sum_{i=1}^{n} p_i$$

EXEMPLE :

- Pour une pièce truquée, on posera : $P(\{\text{Pile}\}) = p$ et $P(\{\text{Face}\}) = q = 1 - p$.

c) Définition abstraite

Les probabilités précédentes vérifient la propriété suivante : si A et B sont deux événements incompatibles (i.e. deux parties disjointes de Ω), alors $P(A \cup B) = P(A) + P(B)$. En effet :

$$\begin{aligned} P(A \cup B) &= \sum_{i=1}^{n} p_i \, \mathbf{1}_{A \cup B}(x_i) = \sum_{i=1}^{n} p_i \, (\mathbf{1}_A(x_i) + \mathbf{1}_B(x_i)) = \sum_{i=1}^{n} p_i \, \mathbf{1}_A(x_i) + \sum_{i=1}^{n} p_i \, \mathbf{1}_B(x_i) \\ &= P(A) + P(B) \end{aligned}$$

Pour des raisons de simplicité de présentation ou de rapidité de démonstration, il est commode de prendre cette propriété comme définition d'une probabilité :

DEFINITION

Soit Ω un univers fini. On appelle probabilité sur Ω une application P *de* $\mathcal{P}(\Omega)$ *dans* [0, 1] *telle que* :

$P(\Omega) = 1$

Si $A \cap B = \varnothing$, $P(A \cup B) = P(A) + P(B)$

(Ω, P) *s'appelle alors espace probabilisé.*

On généralise aussitôt cette propriété à une réunion finie quelconque. Pour toute famille $(A_i)_{i \in I}$ d'événements incompatibles, indicée par I ensemble fini[1], on a :

[1] Si Ω est infini, I est un ensemble fini ou infini dénombrable.

$$P(\bigcup_{i\in I} A_i) = \sum_{i\in I} P(A_i)$$

Cela se montre par récurrence sur le nombre d'éléments de I. Si I est réduit à un seul élément, la relation est triviale. Si I possède deux éléments, la relation résulte de la définition d'une probabilité. Supposons la propriété vérifiée pour les ensembles d'indices de cardinal p. Soit I un ensemble d'indice de cardinal $p + 1$. Soit j un élément quelconque de I, soit $J = I - \{j\}$, soit $A = \bigcup_{i\in J} A_i$ et $B = A_j$. A et B sont disjoints, et :

$$P(\bigcup_{i\in I} A_i) = P(A \cup B) = P(A) + P(B) = P(\bigcup_{i\in J} A_i) + P(A_j)$$

$$= \sum_{i\in J} P(A_i) + P(A_j) \qquad \text{en appliquant l'hypothèse de récurrence sur J}$$

$$= \sum_{i\in I} P(A_i)$$

La définition abstraite d'une probabilité permet de retrouver la définition par somme des probabilités des événements élémentaires. En effet, si on note $\Omega = \{x_1, ..., x_n\}$, et si on pose $p_i = P(\{x_i\})$, alors :

$$A = \bigcup_{i \text{ tel que } x_i \in A} \{x_i\} \qquad \text{union disjointe}$$

donc $$P(A) = \sum_{i \text{ tel que } x_i \in A} P(\{x_i\}) = \sum_{i \text{ tel que } x_i \in A} p_i$$

4- Propriétés des probabilités

La définition des probabilités permet d'en déduire les propriétés suivantes :

PROPOSITION :

i) $P(\emptyset) = 0$

ii) *Pour tout événement* A *et* B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

iii) *Pour tout événement* A, $P(\complement A) = 1 - P(A)$

iv) *Pour tout événement* A *et* B *tel que* A *soit inclus dans* B, *on a* :

$$P(A) \leq P(B) \text{ et } P(B - A) = P(B) - P(A)$$

Démonstration :

Nous donnons les démonstrations en utilisant la définition abstraite des probabilités. Dans le cas où Ω est fini, on peut également les vérifier en recourant à la définition des probabilités comme somme finie de probabilités élémentaires, mais ces propriétés sont également valides dans un espace probabilisé infini, où la probabilité n'est pas nécessairement définie de cette façon.

i) Ω et $\emptyset$ sont deux événements incompatibles, donc on a :

$$P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset)$$

d'où $P(\emptyset) = 0$.

ii) Pour tout événement A et B, incompatibles ou non, on a :

$A \cup B = B \cup (A - B)$
$A = (A \cap B) \cup (A - B)$
et les unions des membres de droites sont disjointes. D'où :
$P(A \cup B) = P(B) + P(A - B)$
$P(A) = P(A \cap B) + P(A - B)$
Le résultat s'en déduit en retranchant membre à membre.
Dans le cas d'une probabilité uniforme, on retrouve la relation :
$\text{Card}(A \cup B) = \text{Card}(A) + \text{Card}(B) - \text{Card}(A \cap B)$

iii) $\Omega = A \cup \complement A$ et cette union est disjointe. D'où :

$1 = P(\Omega) = P(A) + P(\complement A)$

iv) Si A est inclus dans B, alors $B = A \cup (B - A)$, union disjointe, d'où :
$P(B) = P(A) + P(B - A)$
et $P(B - A)$ est positif ou nul.

II : Probabilités conditionnelles

1- Exemples

On lance un dé. Il sort un nombre pair. Quelle est la probabilité d'avoir sorti un 2 ? Réponse : $\frac{1}{3}$.

Même question avec un dé pipé dont voici la loi de probabilité :

1	2	3	4	5	6
$\frac{2}{15}$	$\frac{3}{15}$	$\frac{2}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{3}{15}$

En moyenne, sur 15 tirages, 8 donnent un nombre pair, et sur ces 8, 3 tirages donnent le 2. On est donc amené à poser la probabilité cherchée égale à $\frac{3}{8}$.

Avec le dé ci-dessus, quelle est la probabilité de sortir un nombre pair, sachant qu'il est sorti un nombre supérieur ou égal à 3 ? Sur 15 tirages, en moyenne 10 donnent un nombre supérieur ou égal à 3, et sur ces 10, 5 donnent un nombre pair. La probabilité cherchée est donc $\frac{1}{2}$.

Quelle est la probabilité de sortir un nombre pair, sachant qu'il est sorti un nombre supérieur ou égal à 2 ? Sur 15 tirages, en moyenne 13 donnent un nombre supérieur ou égal à 2, et sur ces 13, 8 donnent un nombre pair. La probabilité cherchée est donc $\frac{8}{13}$.

Si l'on appelle A l'événement "tirer un nombre pair" et B l'événement tirer un nombre supérieur ou égal à 3 (respectivement 2), on remarque que la probabilité de A sachant que B est réalisé a été calculée comme étant égale à $\frac{P(A \cap B)}{P(B)}$. Cette situation est à rapprocher de la loi équiprobable sur un ensemble fini, où l'on calcule le quotient du nombre de cas favorable sur le nombre de cas possibles. Ici, on calcule le quotient de la probabilité de l'événement cherché $P(A \cap B)$ sur la probabilité de l'univers relatif $P(B)$.

DEFINITION

Soit B *un événement de probabilité non nul, d'un espace probabilisé* (Ω, P). *On définit la loi de probabilité sachant* B (*ou conditionnelle par rapport à* B) *par* :

$$P_B : \mathcal{P}(\Omega) \to [0,1]$$

$$A \to P_B(A) = \frac{P(A \cap B)}{P(B)}$$

On la note aussi $P(A \mid B)$

II est dit que P_B est une loi de probabilité. Prouvons le :

i) P_B est clairement définie de $\mathcal{P}(\Omega)$ dans $[0,1]$ car, puisque $A \cap B \subset B$, $P(A \cap B) \leq P(B)$.

ii) $P_B(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = 1$ car $\Omega \cap B = B$

iii) Si A_1 et A_2 sont disjoints, alors $A_1 \cap B$ et $A_2 \cap B$ le sont également et l'on a :

$$P_B(A_1 \cup A_2) = \frac{P((A_1 \cup A_2) \cap B)}{P(B)} = \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)}$$

$$= \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)} = P_B(A_1) + P_B(A_2)$$

Les propriétés i) à iv) du I-4) s'appliquent donc également aux probabilités conditionnelles.

La probabilité P_B est concentrée sur B. En effet, $P_B(B) = 1$, ou encore $P_B(\complement B) = 0$.

2- Formule des probabilités composées

❑ Cas de deux événements : II s'agit seulement de la formule suivante, qui résulte de la définition des probabilités conditionnelles.

$$P(A \cap B) = P(B) \times P_B(A)$$

Par hypothèse, on suppose $P(B) > 0$. On peut cependant étendre cette formule au cas où $P(B) = 0$ sans que $P_B(A)$ soit défini. En effet, on a également $P(A \cap B) = 0$ et il suffit d'attribuer la valeur nulle au membre de droite.

EXEMPLE :

Dans un groupe d'individus, il y a :

des fumeurs B

des non-fumeurs $\complement B$

des hommes A

des femmes $\complement A$

(On rappelle que fumer nuit gravement à la santé)

On donne :

i) $P(B) = \frac{1}{4}$ il s'agit de la proportion de fumeurs parmi la population totale (Cette probabilité est très imprécise. En fait, elle varie fortement avec l'âge. Si elle est de 0,4 pour les individus à 20 ans, elle n'est plus que de 0,2 à 60 ans).

ii) $P(A) = \frac{1}{2}$ il s'agit de la proportion d'hommes dans la population totale

iii) $P_B(A) = \frac{3}{5}$ il s'agit de la proportion d'hommes parmi les personnes qui fument

Calculer $P(A \cap B)$, $P(\complement A \cap B)$, $P(A \cap \complement B)$, $P(\complement A \cap \complement B)$, $P_B(\complement A)$, $P_A(B)$, $P_A(\complement B)$, $P_{\complement B}(A)$, $P_{\complement B}(\complement A)$, $P_{\complement A}(B)$, $P_{\complement A}(\complement B)$ et interpréter les résultats obtenus.

Réponses :

iv) $P(A \cap B) = P(B) \times P_B(A) = \frac{1}{4} \times \frac{3}{5} = \frac{3}{20}$ (proportion de personnes masculines qui fument parmi la population totale)

v) $P(A \cap \complement B) = P(A) - P(A \cap B) = \frac{1}{2} - \frac{3}{20} = \frac{7}{20}$ (proportion de personnes masculines qui ne fument pas parmi la population totale)

vi) $P(\complement A \cap B) = P(B) - P(A \cap B) = \frac{1}{4} - \frac{3}{20} = \frac{1}{10}$ (proportion de personnes féminines qui fument parmi la population totale)

vii) $P(\complement A \cap \complement B) = P(\complement A) - P(\complement A \cap B) = 1 - P(A) - P(\complement A \cap B) = 1 - \frac{1}{2} - \frac{1}{10} = \frac{2}{5}$

ou bien $= 1 - P(B) - P(A \cap \complement B) = 1 - \frac{1}{4} - \frac{7}{20} = \frac{2}{5}$ (proportion de personnes féminines qui ne fument pas parmi la population totale).

Ces données peuvent être résumées dans un tableau :

	hommes A	femmes $\complement A$	
fumeurs B	$\frac{3}{20}$	$\frac{1}{10} = \frac{2}{20}$	$\frac{5}{20} = \frac{1}{4}$
non fumeurs $\complement B$	$\frac{7}{20}$	$\frac{2}{5} = \frac{8}{20}$	$\frac{15}{20} = \frac{3}{4}$
	$\frac{10}{20} = \frac{1}{2}$	$\frac{10}{20} = \frac{1}{2}$	

Les autres probabilités ne posent plus de problèmes :

viii) $P_B(\complement A) = \frac{2}{5}$ (proportion de femmes parmi les fumeurs)

ix) $P_A(B) = \frac{3}{10}$ (proportion de fumeurs parmi les hommes)

x) $P_A(\complement B) = \frac{7}{10}$ (proportion de non-fumeurs parmi les hommes)

xi) $P_{\complement B}(A) = \frac{7}{15}$ (proportion d'hommes parmi les non-fumeurs)

xii) $P_{\complement B}(\complement A) = \frac{8}{15}$ (proportion de femmes parmi les non-fumeurs)

xiii) $P_{\complement A}(B) = \frac{1}{5}$ (proportions de fumeurs parmi les femmes)

xiv) $P_{\complement A}(\complement B) = \frac{4}{5}$ (proportion de non-fumeurs parmi les femmes)

❑ cas de plusieurs événements :
Soient A_1, A_2, ..., A_n, n événements. Alors, on a

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1) \times P_{A_1}(A_2) \times P_{A_1 \cap A_2}(A_3) \times ... \times P_{A_1 \cap A_2 \cap ... \cap A_{n-1}}(A_n).$$

Cela se montre aisément par récurrence en utilisant le fait que :

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1 \cap A_2 \cap ... \cap A_{n-1}) \times P_{A_1 \cap A_2 \cap ... \cap A_{n-1}}(A_n).$$

EXEMPLE :
une urne contient a boules blanches et b boules noires. On effectue des tirages avec remise. Lorsqu'on tire une boule noire, on la remplace par une boule blanche. Quelle est la probabilité de tirer n boules noires de suite ?

Soit A_i l'événement tirer une boule noire au i-ème coup. On a :

$$P(A_1) = \frac{b}{a+b}$$

$P_{A_1 \cap A_2 \cap ... \cap A_{i-1}}(A_i) = \frac{b-i+1}{a+b}$ puisqu'au i-ème tirage, il y a $i-1$ boules noires qui ont été remplacées par des blanches. La probabilité cherchée est donc :

$$P(A_1 \cap A_2 \cap ... \cap A_n) = \frac{b(b-1)(b-2)...(b-n+1)}{(a+b)^n}$$

3- Formule des probabilités totales

EXEMPLE :
On considère n usines fabriquant un même produit. Soit p_i la part de marché de la i-ème usine (proportion du nombre de pièces fabriquées par la i-ème usine sur le nombre de pièces fabriquées par toutes les usines). p_i est la probabilité pour que, prenant une pièce au hasard, elle provienne de la i-ème usine. Soit A_i cet événement. $P(A_i) = p_i$. On remarquera également que (A_1, A_2, ... ,A_n) forme un système complet d'événements.
Par ailleurs, chaque usine fabrique un certain nombre de pièces défectueuses. Soit q_i la proportion de pièces défectueuses fabriquées par la i-ème usine, relativement à sa propre production. Si B est l'événement "tirer une pièce défectueuse en choisissant une pièce au hasard", on a $q_i = P_{A_i}(B)$.
Nous prendrons comme exemple numérique :

$p_1 = 0,900$
$p_2 = 0,095$
$p_3 = 0,005$ (la première usine a quasiment le monopole du marché)
$q_1 = 0,01$
$q_2 = 0,05$
$q_3 = 0,20$

(la forte proportion de pièces défectueuses fabriquées par la troisième usine explique peut-être sa faible implantation ! !)
La question qu'on se pose est la suivante : quelle est la probabilité de tirer une pièce défectueuse en choisissant une pièce au hasard dans la production totale ? Autrement dit, calculer P(B).
La proportion de pièces défectueuses fabriquées par la i-ème usine, relativement au nombre total de pièces fabriquées par toutes les usines est $p_i q_i$. La probabilité cherchée est donc $\sum p_i q_i$. L'exemple numérique donne 0,01475 (de l'ordre d'une chance sur 100).

Ceci s'énonce dans le cas général de la façon suivante :

PROPOSITION (formule des probabilités totales)

Soit $(A_i)_{1\le i\le n}$ *un système complet d'événements d'un espace probabilisé* (Ω, P) *tels que, pour tout i,* $P(A_i) > 0$, *et* B *un autre événement de cet espace. Alors* :

$$P(B) = \sum_{i=1}^{n} P(A_i) \times P_{A_i}(B)$$

<u>Démonstration</u> :

$(A_i)_{1\le i\le n}$ formant un système complet d'événements, on a :

$$\Omega = \bigcup_{i=1}^{n} A_i \qquad \text{union disjointe.}$$

D'où
$$\begin{aligned} P(B) &= P(\Omega \cap B) \\ &= P\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap B\right) \\ &= P\left(\bigcup_{i=1}^{n} (A_i \cap B)\right) \qquad \text{cette union étant disjointe} \\ &= \sum_{i=1}^{n} P(A_i \cap B) \\ &= \sum_{i=1}^{n} P(A_i) \times P_{A_i}(B) \end{aligned}$$

4- Formule De Bayes (ou probabilité des causes)

Nous reprenons l'exemple précédent. On tire au hasard une pièce. Elle est défectueuse. Quelle est la probabilité qu'elle provienne de la *i*-ème usine ? On cherche $P_B(A_i)$. C'est égal à $\frac{P(B \cap A_i)}{P(B)}$. En remplaçant par les expressions trouvées au paragraphe précédant, on obtient :

PROPOSITION (Formule de Bayes)

Soit $(A_i)_{1\le i\le n}$ *un système complet d'événements d'un espace probabilisé* (Ω, P) *tels que, pour tout i,* $P(A_i) > 0$, *et* B *un autre événement de cet espace tel que* $P(B) > 0$. *Alors* :

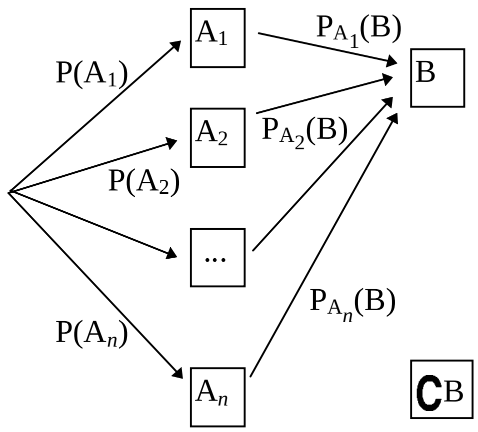
$$P_B(A_i) = \frac{P(B \cap A_i)}{P(B)} = \frac{P(A_i) \times P_{A_i}(B)}{\sum P(A_j) \times P_{A_j}(B)}$$

Le calcul donne :

❑ pour la première usine 0,61
❑ pour la deuxième usine 0,32
❑ pour la troisième usine 0,07

II peut paraître étonnant que la probabilité que la pièce défectueuse provienne de la troisième usine soit si faible, alors que cette usine est de mauvaise qualité, mais elle n'intervient que fort peu sur le marché. Par contre, la probabilité que la pièce défectueuse provienne de la première usine est très forte, car elle a presque le monopole de fabrication, malgré la qualité de sa fabrication.

La situation générale peut être visualisée par le schéma suivant :



La probabilité d'aller en B en passant par A_i est $P(B \cap A_i) = P(A_i) \times P_{A_i}(B)$

La probabilité d'aller en B est la somme des probabilités sur tous les chemins possibles. Cela donne la formule des probabilités totales :

$$P(B) = \sum_{i=1}^{n} P(A_i) \times P_{A_i}(B)$$

La probabilité d'être passé par A_i sachant qu'on est arrivé en B est le quotient de la probabilité du *i*-ème chemin sur la somme des probabilités de tous les chemins menant à B (un analogue du nombre de cas favorables sur le nombre de cas possibles). Cela donne la formule de Bayes :

$$P_B(A_i) = \frac{P(B \cap A_i)}{P(B)} = \frac{P(A_i) \times P_{A_i}(B)}{\sum P(A_j) \times P_{A_j}(B)}$$

5- Evénements indépendants

DEFINITION

Deux événements A *et* B *sont dits indépendants si* $P(A \cap B) = P(A) \times P(B)$

La justification de cette définition est la suivante. Si P(B) est non nul, et si A et B sont indépendants, alors on doit s'attendre à ce que P(A) et $P_B(A)$ soient égales. En effet, l'indépendance de A et B a pour conséquence que la connaissance de la réalisation de B n'a aucune influence sur celle de A. Or $P(A) = P_B(A) \Leftrightarrow P(A) = \dfrac{P(A \cap B)}{P(B)}$, d'où la définition.

EXEMPLE :

On considère généralement que les tirages successifs avec une pièce, un dé ou au loto sont indépendants. Les tirages précédents n'ont aucune influence sur les tirages suivants. Si la pièce est équilibrée, il y aura toujours une probabilité $\frac{1}{2}$ de tirer P. En particulier, le fait d'avoir joué toutes les semaines au loto depuis sa création sans jamais gagner n'augmente pas ses chances de gagner au prochain tirage (cela ne les diminue pas non plus d'ailleurs).

DEFINITION

Plusieurs événements $A_1, A_2, ..., A_n$ *sont indépendants si, pour toute sous-famille de p événements* $A_{i_1}, A_{i_2}, ... , A_{i_p}$ (*les indices* $i_1, ..., i_p$ *étant distincts*), *on a* :

$$P(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_p}) = P(A_{i_1})P(A_{i_2})...P(A_{i_p})$$

Insistons sur le fait que la propriété doit être vérifiée pour TOUTE sous-famille de p événements. On ne peut se contenter de vérifier l'indépendance des événements deux à deux.

EXEMPLE :

On lance deux dés. On considère les événements suivants :

A : le lancer du premier dé est pair
B : le lancer du deuxième dé est pair
C : la somme des deux lancers est paire

II est clair que les événements A et B, issus de deux dés différents sont indépendants, et que les trois événements A, B et C ne le sont pas. En effet la réalisation de A et B entraîne celle de C. Or :

$$P(A) = P(B) = P(C) = \frac{1}{2} \text{ (Pour P(C), réfléchir un peu)}$$

$$P(A \cap B) = \frac{1}{4} = P(A \cap C) = P(B \cap C)$$

En effet, les trois événements $A \cap B$, $B \cap C$ et $A \cap C$ sont identiques. Ainsi les événements sont indépendants deux à deux. On a également :

$$P(A \cap B \cap C) = \frac{1}{4}$$

car $A \cap B \cap C = A \cap B$. C n'apporte rien de plus. Donc, les trois événements ne sont pas indépendants, bien qu'ils le soient deux à deux.

6- Probabilité produit

On considère n univers finis Ω_i, $1 \leq i \leq n$. On réalise alors n expériences aléatoires x_i suposées indépendantes, la i-ème, ayant son issue dans Ω_i. Cela revient à faire une épreuve unique $(x_1, x_2, ..., x_n)$ dont l'issue appartient à l'espace $\Omega = \Omega_1 \times \Omega_2 \times ... \times \Omega_n$. Cet espace s'appelle l'espace produit des Ω_i. C'est un espace fini. Si Ω_i est munie d'une probabilité P_i, il est possible de munir Ω d'une probabilité P définie de la façon naturelle suivante :

$$P(\{(x_1, x_2, ..., x_n)\}) = P_1\{\{x_1\}\} \times P_2(\{x_2\}) \times ... \times P_n(\{x_n\})$$

P s'appelle la probabilité produit des P_i.

EXEMPLES :

- On lance un dé, et on tire une carte d'un jeu de 32 cartes. Alors, $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$, $\Omega_2 = \{c \mid c$ est une carte d'un jeu de 32 cartes$\}$. Ω est l'ensemble des couples formés d'un lancer de dé et d'un tirage de cartes : $\Omega = \{(d, c) \mid d \in \Omega_1, c \in \Omega_2\}$. On a $P_1(\{d\}) = \frac{1}{6}$, et $P_2(\{c\}) = \frac{1}{32}$. D'où $P(\{c, d\}) = \frac{1}{192}$.
- On lance un dé n fois, (ou bien n dés). Tous les espaces Ω_i sont égaux à $\{P, F\}$. $\Omega = \{P, F\}^n$. Toutes les lois P_i sont les mêmes : $P_i(\{\text{Pile}\}) = \frac{1}{2} = P_i(\{\text{Face}\})$. Pour chaque x élément de Ω, on a $P(\{x\}) = \frac{1}{2^n}$.

III : Variables aléatoires

l- Loi d'une variable aléatoire

Soit $(\Omega\ , P)$ un espace probabilisé fini et E un ensemble quelconque. On appelle variable aléatoire une application X de Ω dans E. Le plus souvent, $E = \mathbb{R}$ et X est appelée variable aléatoire réelle.

EXEMPLES :

- On lance un dé. On gagne un franc s'il tombe sur 5 ou 6, et on perd un franc sinon. Soit X le gain. On a $\Omega = \{1, 2, 3, 4, 5, 6\}$, et X est une application de Ω dans $\mathbb{R}$ définie par :

  $X(l) = X(2) = X(3) = X\{4) = -1$
  $X(5) = X(6) = 1$

La probabilité que le gain soit 1 est $\frac{2}{6}$ ou $\frac{1}{3}$; la probabilité qu'il soit –1 est $\frac{4}{6}$ ou $\frac{2}{3}$. Initialement nous avons une probabilité P sur Ω, or nous voyons que, naturellement, nous en déduisons une sur $\{-1,1\} = X(\Omega)$. Cette dernière probabilité s'appelle loi de X et sera notée P_X (ne pas confondre la notation avec celle d'une probabilité conditionnelle).

- On lance deux dés. Soit X la somme des numéros tirés. On a ici comme espace probabilisé Ω^2, où Ω est toujours {1, 2, 3, 4, 5, 6}, muni de la loi uniforme. X peut prendre les valeurs {2, 3, 4, 5, 6, ..., 11, 12}. Chercher la loi de X, c'est chercher la probabilité de sortie de chacun de ces totaux ; Le lecteur vérifiera que l'on a :

2	3	4	5	6	7	8	9	10	11	12
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

On a donc défini une probabilité P_X sur {2, 3, 4, ..., 12} La probabilité que X appartienne à $A = \{6, 7, 8\}$ est $\frac{16}{36}$, soit quasiment 1 chance sur 2.

Regardons, dans les exemples précédents, comment calculer $P_X(A)$ (qu'on interprète comme la probabilité que X appartienne à A). On réalise des expériences (tirer un dé, ou deux dés, ...) dont le résultat ω est dans Ω. On applique X à ω, et on regarde si $X(\omega)$ appartient à A ou non. Le premier cas correspond à l'événement recherché. La probabilité $P_X(A)$ de cet événement est donc égal à la probabilité de l'ensemble des ω tels que $X(\omega)$ appartienne à A. Or $\{\omega \mid X(\omega) \in A\}$ n'est autre que l'image réciproque de A par X, $X^{-1}(A)$.

DEFINITION

Soit X *une variable aléatoire définie sur un espace probabilisé* (Ω, P) *et* A *un intervalle de* $\mathbb{R}$*. On définit donc la loi de* X *sur* R *par* : $P_X(A) = P(X^{-1}(A)) = P(\{\omega \in \Omega \mid X(\omega) \in A\})$.

PROPOSITION

La loi P_X *d'une variable aléatoire définie sur un espace probabilisé* (Ω, P) *est une loi de probabilité sur* $X(\Omega)$.

Démonstration :

II faut vérifier les trois propriétés des probabilités.

i) Il est clair que P_X est à valeurs dans [0,1], puisque c'est le cas de P.

ii) $P_X(X(\Omega)) = P(X^{-1}(X(\Omega))) = P(\Omega) = 1$

iii) Soit A et B deux événements disjoints de $X(\Omega)$. Alors $P_X(A \cup B) = P(X^{-1}(A \cup B))$. Or $X^{-1}(A \cup B) = \{\omega \mid X(\omega) \in A \cup B\} = \{\omega \mid X(\omega) \in A\} \cup \{\omega \mid X(\omega) \in B\} = X^{-1}(A) \cup X^{-1}(B)$, donc :

$$P_X(A \cup B)) = P(X^{-1}(A) \cup X^{-1}(B))$$

Par ailleurs, A et B étant disjoints, il en est de même de $X^{-1}(A)$ et $X^{-1}(B)$ puisqu'un élément ω commun à ces deux parties serait tel que $X(\omega) \in A \cap B = \emptyset$. Donc :

$$P_X(A \cup B)) = P(X^{-1}(A)) + P(X^{-1}(B)) = P_X(A) + P_X(B)$$

On note $P_X(A) = P(X^{-1}(A)) = P(\{\omega \mid X(\omega) \in A\}) = P(X \in A)$. De même, si x est un élément de E, $P(X = \{x\}) = P(\{\omega \in \Omega \mid X(\omega) = x) = P(X^{-1}(\{x\})) = P_X(\{x\})$. Comme toute probabilité sur un espace fini, les valeurs $P(X = \{x\}) = P_X(\{x\})$ définissent sans ambiguïté la probabilité P_X sur $X(\Omega)$. Enfin, dans le cas d'une variable aléatoire réelle, on note, pour tout x réel :

$$P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x) = P(X^{-1}(]-\infty, x])) = P_X(]-\infty, x] \cap X(\Omega))$$

La variable aléatoire X permet de transférer la probabilité P sur Ω en une probabilité P_X sur $X(\Omega)$. Considérons maintenant une variable aléatoire X de Ω dans un ensemble E, et une application f de E dans un ensemble F.

❑ X permet de transférer la probabilité P sur Ω en une probabilité P_X sur $X(\Omega)$. $(X(\Omega), P_X)$ devient alors un espace probabilisé et g une variable aléatoire sur $X(\Omega)$ à valeurs dans F. g permet donc de transférer la probabilité P_X en une probabilité sur $g(X(\Omega))$.

❑ $Y = g \circ X$ est une variable aléatoire définie sur Ω à valeurs dans F, qui transfère la probabilité P de Ω en une probabilité P_Y sur $Y(\Omega) = g(X(\Omega))$

Vérifions que les deux procédés donnent le même résultat. Pour toute partie A de $Y(\Omega)$, on a :

$$P_Y(A) = P(Y^{-1}(A))$$

Or $Y^{-1}(A) = X^{-1}(g^{-1}(A))$ car :

$$\omega \in Y^{-1}(A) \Leftrightarrow Y(\omega) \in A \Leftrightarrow g(X(\omega)) \in A \Leftrightarrow X(\omega) \in g^{-1}(A) \Leftrightarrow \omega \in X^{-1}(g^{-1}(A))$$

Donc :

$$P_Y(A) = P(X^{-1}(g^{-1}(A))) = P_X(g^{-1}(A))$$

et cette dernière est bien la probabilité définie sur $g(X(\Omega))$ par le premier procédé.

2- Lois usuelles discrètes

a) Variable certaine

Soit (Ω, P) un espace probabilisé et k un réel donné. On appelle variable certaine une variable de la forme :

$$X : \Omega \to \mathbb{R}$$
$$\omega \to k$$

Sa loi est concentrée sur la valeur k. $P(X = k) = 1$.

EXEMPLE :

On lance une pièce. Si on tombe sur P, on gagne un franc (G = l), sinon on perd un franc (G = – l). La variable $X = G^2$ est une variable certaine, égale à l.

b) Loi de Bernoulli

Il s'agit de la loi notée $\mathcal{B}(p)$ d'une variable aléatoire qui ne prend que deux valeurs, habituellement 0 ou 1. Par exemple, si on lance une pièce, le fait de tomber sur Pile attribue à X la valeur 0, et sur Face, la valeur 1. On a :

$P(X = 1) = p$
$P(X = 0) = 1 - p$

Généralement, la valeur 1 de X est appelée succès, et la valeur 0 échec, p est élément de]0, 1[. Si $p = 0$ ou 1, on retrouve une loi certaine.

c) Loi binomiale

Il s'agit de la loi des variables aléatoires suivantes :

❑ On lance n pièces. On compte le nombre de Piles.
❑ On tire n boules d'une urne avec remise. On compte le nombre de boules blanches tirées.
❑ Schéma général : on répète n expériences de Bernoulli indépendantes, et on compte le nombre X de succès. Si X_i est le résultat 0 ou 1 de la i-ème expérience, $1 \leq i \leq n$, alors X est la somme des X_i.

On a :

$$P_X(\{k\}) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$\binom{n}{k}$ représente en effet le nombre de combinaisons de k éléments parmi n, représentant les succès. La probabilité de chacune de ces combinaisons est $p^k (1 - p)^{n-k}$.

Cette loi dépend des deux paramètres n et p et est notée $\mathcal{B}(n, p)$. La loi $\mathcal{B}(1, p)$ est identique à la loi de Bernoulli $\mathcal{B}(p)$.

❑ La loi binomiale, caractéristique des tirages avec remise, peut également parfois servir d'approximation lors de tirages sans remise. Considérons le cas de tirages de n boules sans remise dans une urne contenant N boules, dont m blanches. La loi de la variable aléatoire Y égale au nombre de boules blanches tirées se trouve en calculant le quotient du nombre de tirages favorables sur le nombre de tirages possibles.

$$P_Y(\{k\}) = P(Y = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} = \frac{m!\, n!\, (N - m)!\, (N - n)!}{k!\, (m - k)!\, (n - k)!\, (N - m - n + k)!\, N!}$$

Curieusement, cette expression est symétrique en n et m. Cette symétrie s'explique aisément si l'on imagine l'expérience suivante : une urne contient N boules. Deux personnes sont en présence. La première en choisit m sur lesquelles elle dessine une marque personnelle. Puis elle les remet dans l'urne. La deuxième en choisit alors n. Quelle est la probabilité qu'il y en ait k de marquées ? Il s'agit de la loi précédente (dite loi hypergeométrique). Il est évident que l'expérience donne la même loi si on permute les deux personnes, c'est-à-dire si on échange le rôle de n et m. Le nombre de boules marquées n'est autre que le nombre de boules communes au deux personnes.

Cette loi intervient également dans les sondages de n personnes différentes, dans une population totale de N personnes, dont m possèdent un caractère donné. On dénombre le nombre k de personnes dans l'échantillon choisi ayant le dit caractère.

Lorsque le nombre N de boules est grand, on peut s'attendre à ce qu'un tirage avec remise donne à peu près le même résultat qu'un tirage sans remise. La loi précédente doit donc permettre de retrouver une loi binomiale. Cherchons donc la limite de $P(Y = k)$ lorsque N tend vers l'infini et $\frac{m}{N}$ (proportion de boules blanches) tend vers p, élément de [0,1]. On a :

$$P(Y = k) = \frac{m!\, n!\, (N - m)!\, (N - n)!}{k!\, (m - k)!\, (n - k)!\, (N - m - n + k)!\, N!}$$
$$= \binom{n}{k} \frac{m!\, (N - m)!\, (N - n)!}{(m - k)!\, (N - m - n + k)!\, N!}$$
$$= \binom{n}{k} \frac{(N - m)(N - m - 1)...(N - m - n + k + 1)}{N(N - 1)(N - 2)...(N - n + k + 1)} \frac{m(m - 1)...(m - k + 1)}{(N - n + k)... (N - n + 1)}$$

Le terme général du premier quotient vaut, pour r variant de 0 à $n - k - 1$:
$$\frac{N - m - r}{N - r} = \frac{1 - m/N - r/N}{1 - r/N}$$

or r est borné donc $\frac{r}{N}$ tend vers 0, et $\frac{m}{N}$ tend vers p. La limite de chacun de ces termes est donc $1 - p$, et ils sont au nombre de $n - k$. On prouve de même que chacun des termes du second quotient tend vers p, et ils sont au nombre de k. La limite de $P(Y = k)$ est donc $\binom{n}{k} p^k (1 - p)^{n-k}$.

Cette propriété explique que, dans le cas d'un sondage où n est de l'ordre de 1000, alors que N est de l'ordre de 60 000 000, bien que ne procédant pas a priori à des tirages avec remise, la loi en présence est considérée comme une loi binomiale.

EXEMPLE :
46% des gens sont de groupe sanguin O. Sur 10 000 personnes, on en prend 20. Quelle est la probabilité d'en trouver 9 de groupe O ? Si on considère qu'on est dans le cadre d'une loi binomiale $\mathcal{B}(20, \frac{46}{100})$, alors :
$$P(X = 9) = \binom{20}{9} 0.46^9\, 0.54^{11} \approx 0{,}17634$$
alors que la valeur donnée par la loi hypergéométrique est 0,17651.

3- Espérance et variance

a) Espérance

On considère une loi de probabilité d'une variable aléatoire réelle X à valeur dans $\{x_i, i \in I\}$ où I est fini. Soit p_i la probabilité $P_X(\{x_i\}) = P(X = x_i)$. On s'intéresse à la valeur moyenne de X. On peut s'attendre à ce que, sur un grand nombre d'expérience, X prenne la valeur x_i avec une fréquence de plus en plus proche de p_i. Il est donc naturel de poser, comme valeur moyenne de X sur une expérience la quantité :

$$E(X) = \sum_{i \in I} p_i\, x_i = \sum_{i \in I} x_i\, P(X = x_i)$$

Cette quantité s'appelle espérance de X.

EXEMPLE :
Soit X la somme de deux dés. La loi de X est la suivante :

2	3	4	5	6	7	8	9	10	11	12
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Le calcul conduit à $E(X) = 7$, ce qui est naturel, vu la symétrie du tableau.

On remarquera que la définition de E(X) ne fait intervenir que les valeurs que prend X et sa loi sur $\mathbb{R}$, et que l'on peut oublier totalement l'espace Ω sur lequel est défini X. Nous pouvons cependant donner une expression de E(X) faisant intervenir Ω et la loi de probabilité P sur Ω.

$$E(X) = \sum_{\omega\in\Omega} X(\omega)\, P(\{\omega\})$$

En effet, soit $x_1, x_2, ..., x_n$ les valeurs que prend X. Alors les images réciproques $X^{-1}(\{x_1\})$, $X^{-1}(\{x_2\})$, ..., $X^{-1}(\{x_n\})$ forment un système complet d'événements de Ω : ils sont disjoints (car un élément ω ne peut avoir deux images différentes) et leur réunion est Ω (car tout ω admet comme image l'un des x_i). La somme sur Ω peut donc se calculer en sommant d'abord sur $X^{-1}(\{x_1\})$, puis sur $X^{-1}(\{x_2\})$, ..., puis sur $X^{-1}(\{x_n\})$, et en faisant la somme de ces n totaux :

$$\sum_{\omega\in\Omega} X(\omega)\, P(\{\omega\}) = \sum_{i=1}^{n} \sum_{\omega\in X^{-1}(\{x_i\})} X(\omega)\, P(\{\omega\})$$

D'autre part, lorsque ω appartient à $X^{-1}(\{x_i\})$, par définition, $X(\omega)$ est égal à x_i. D'où :

$$\sum_{\omega\in\Omega} X(\omega)\, P(\{\omega\}) = \sum_{i=1}^{n} \sum_{\omega\in X^{-1}(\{x_i\})} x_i\, P(\{\omega\}) = \sum_{i=1}^{n} x_i \sum_{\omega\in X^{-1}(\{x_i\})} P(\{\omega\})$$

$$= \sum_{i=1}^{n} x_i\, P(X^{-1}(\{x_i\}) = \sum_{i=1}^{n} x_i\, P(X = x_i) = E(X)$$

Cette formule est utile pour prouver le résultat suivant :

PROPOSITION
Soit $X : \Omega \to \{x_i,\ i \in I\} = E$ *une variable aléatoire définie sur l'espace probabilisé* (Ω, P) *et* $g : E \to \mathbb{R}$ *une fonction. Alors* $g \circ X$ (*notée également* $g(X)$) *est une variable aléatoire dont l'espérance se calcule de la façon suivante, au moyen de la loi de* X :

$$E(g(X)) = \sum_{i\in I} g(x_i)\, P(X = x_i)$$

<u>Démonstration</u> :
Elle est analogue à la démarche suivie précédemment :

$$E(g(X)) = \sum_{\omega\in\Omega} g(X(\omega))\, P(\{\omega\}) = \sum_{i=1}^{n} \sum_{\omega\in X^{-1}(\{x_i\})} g(x_i)\, P(\{\omega\}) = \sum_{i=1}^{n} g(x_i) \sum_{\omega\in X^{-1}(\{x_i\})} P(\{\omega\})$$

$$= \sum_{i=1}^{n} g(x_i)\, P(X^{-1}(\{x_i\}) = \sum_{i=1}^{n} g(x_i)\, P(X = x_i)$$

L'intérêt de cette formule est qu'il est inutile de rechercher la loi de $g(X)$ pour calculer son espérance. La loi de X suffit.

EXEMPLE :

Soit X la somme de deux dés. La loi de X est la suivante :

2	3	4	5	6	7	8	9	10	11	12
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

La formule démontrée permet d'écrire :

$$E((X-7)^2) = 25 \times \frac{1}{36} + 16 \times \frac{2}{36} + 9 \times \frac{3}{36} + 1 \times \frac{4}{36} + ... + 25 \times \frac{1}{36} = \frac{35}{6}$$

(Cette quantité mesure la dispersion de X autour de sa valeur moyenne et s'appelle variance de X).

PROPOSITION

L'espérance possède les propriétés suivantes :

i) $\forall a \in \mathbb{R}$, $\forall b \in \mathbb{R}$, $\forall$ X *et* Y *variables aléatoires sur* (Ω, P), *on a* :

$$E(aX + bY) = aE(X) + bE(Y).$$

Autrement dit, l'espérance est un opérateur linéaire sur l'espace vectoriel des variables aléatoires.

ii) $X \geq 0 \Rightarrow E(X) \geq 0$

ii bis) $X \geq Y \Rightarrow E(X) \geq E(Y)$

iii) $|E(X)| \leq E(|X|)$

iv) $|E(XY)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}$

v) *Si* X *est une variable certaine égale à k, alors* $E(X) = k$

<u>Démonstration</u> :

i) et ii) se déduisent facilement du fait que $E(X) = \sum_{\omega\in\Omega} X(\omega)\, P(\{\omega\})$, expression linéaire par rapport à X, et positive si X est positive.

ii bis) se déduit de ii) en considérant X – Y:

$$X - Y \geq \Rightarrow E(X - Y) \geq 0 \Rightarrow E(X) - E(Y) \geq 0 \Rightarrow E(X) \geq E(Y).$$

iii) se montre en remarquant que $-X \leq |X| \leq X$, donc $-E(X) \leq E(|X|) \leq E(X)$

iv) La quantité $\langle X,Y\rangle = E(XY)$ est une forme bilinéaire symétrique positive, et comme toute forme bilinéaire symétrique positive, elle vérifie l'inégalité de Schwarz :

$$\langle X,Y\rangle^2 \leq \langle X,X\rangle \langle Y,Y\rangle$$

v) On a $P(X = k) = 1$, donc $E(X) = k\, P(X = k) = k$

La variable aléatoire X – E(X) a une espérance nulle. On dit qu'elle est centrée.

b) Variance

On mesure la dispersion de la variable aléatoire X autour de son espérance $E(X) = \mu$ en calculant la quantité $E((X - \mu)^2)$. Il s'agit de la variance de X, notée V(X). Si l'on développe cette quantité, on obtient :

$$\begin{aligned} V(X) &= E((X-\mu)^2) \\ &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2) - 2\mu E(X) + E(\mu^2) \\ &= E(X^2) - 2\mu^2 + \mu^2 \qquad \text{car } E(X) = \mu \text{ et } E(\mu^2) = \mu^2. \end{aligned}$$

Donc $V(X) = E(X^2) - E(X)^2$.
C'est cette dernière quantité que l'on calcule le plus souvent pour obtenir la variance.

On vérifiera facilement que :

$$\forall a \in \mathbb{R}, \forall b \in \mathbb{R}, V(aX + b) = a^2 V(X).$$

On appelle écart-type la quantité $\sigma(X) = \sqrt{V(X)}$.

La variable aléatoire $\dfrac{X - E(X)}{\sigma(X)}$ a une espérance nulle. On dit qu'elle est centrée réduite.

c) Lois usuelles

❑ *Variable certaine* :
Soit $X = k$ une variable certaine égale à k. Le lecteur vérifiera sans difficulté que $E(X) = k$, $E(X^2) = k^2$, et $V(X) = 0$.
Réciproquement, il sera prouvé dans la paragraphe suivant que, si X admet une variance nulle, alors, il existe k tel que $P(X = k) = 1$.

❑ *Variable de Bernoulli* :
Elle admet la loi suivante :

X	0	1
P_X	$1 - p$	p

On a donc : $E(X) = p = E(X^2)$ (En effet, $X^2 = X$) D'où $V(X) = p - p^2 = p(1 - p) = pq$ avec $q = 1 - p$

Un exemple fréquent de variable de Bernoulli est celui des fonctions indicatrices. Soit (Ω, P) un espace probabilisé, et A un événement. On considère la variable aléatoire définie de la façon suivante :

$$\mathbf{1}_A : \Omega \to \{0, 1\}$$
$$\omega \to \begin{cases} 0 \text{ si } \omega \notin A \\ 1 \text{ si } \omega \in A \end{cases}$$

Cette variable aléatoire est la fonction indicatrice de la partie A. La loi de $\mathbf{1}_A$ est une loi de Bernoulli de paramètre p, avec :

$$\begin{aligned} p &= P(\mathbf{1}_A = 1) \\ &= P(\{\omega \in \Omega, \mathbf{1}_A(\omega) = 1\}) \\ &= P(A) \end{aligned}$$

Alors, $E(\mathbf{1}_A) = P(A)$

Voici quelques règles de calcul des fonctions indicatrices, dont la démonstration est laissée au lecteur :

$$\mathbf{1}_\Omega = 1$$
$$\mathbf{1}_\varnothing = 0$$
$$\mathbf{1}_{A \cap B} = \mathbf{1}_A \times \mathbf{1}_B$$
$$\mathbf{1}_{\complement A} = 1 - \mathbf{1}_A$$
$$1 - \mathbf{1}_{A \cup B} = (1 - \mathbf{1}_A)(1 - \mathbf{1}_B)$$
$$\mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A \cap B}$$

L'avant-dernière formule se généralise aisément à une réunion quelconque de n parties $A_1, ..., A_n$, de la façon suivante :

$$1 - \mathbf{1}_{\cup A_i} = (1 - \mathbf{1}_{A_1})(1 - \mathbf{1}_{A_2})...(1 - \mathbf{1}_{A_n})$$

En effet :

$$1 - \mathbf{1}_{\cup A_i} = \mathbf{1}_{\complement(\cup A_i)} = \mathbf{1}_{\cap \complement A_i} = \prod \mathbf{1}_{\cap \complement A_i} = \prod (1 - \mathbf{1}_{A_i})$$

Cela permet de démontrer la formule dite de Poincaré en développant le membre de droite, qui donne la probabilité d'une union quelconque d'événements :

$$1 - \mathbf{1}_{\cup A_i} = 1 - \sum_i \mathbf{1}_{A_i} + \sum_{i,j} \mathbf{1}_{A_i \cap A_j} - ... + (-1)^m \sum_{i_1<i_2<...<i_m} \mathbf{1}_{A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_m}} + ... + (-1)^n \mathbf{1}_{A_1 \cap A_2 \cap ... \cap A_n}$$

En prenant l'espérance des deux membres, on obtient :

$$P(\bigcup_{i=1}^{n} A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_n) - ...$$

$$+ (-1)^{m-1} \sum_{i_1<i_2<...<i_m} P(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_m}) + ... + (-1)^{n-1} P(A_1 \cap A_2 \cap ... \cap A_n)$$

❑ *Loi binomiale* :

Soit X_1 , ..., X_n n variables aléatoires de Bernoulli indépendantes (i.e. les événements correspondant ces variables sont indépendants), à valeur dans {0,1}, de paramètre p. Alors la somme $X = \sum_{i=1}^{n} X_i$ suit une loi binomiale $\mathcal{B}(n, p)$. X est égal au nombre de succès des n expériences. Rappelons que :

$E(X_i) = p$

$V(X_i) = p(1 - p)$

On en déduit que :

$$E(X) = \sum_{i=1}^{n} E(X_i) = np$$

Par ailleurs, il est démontré plus loin que la variance d'une somme de variables aléatoires indépendantes est égale à la somme des variances. D'où :

$$V(X) = \sum_{i=1}^{n} V(X_i) = np(1 - p)$$

On peut également démontrer ce dernier résultat directement. Calculons $E(X^2)$:

$$X^2 = \sum_{i=1}^{n} X_i^2 + 2 \sum_{i,j} X_iX_j$$

X_i^2 n'est autre que X_i, donc son espérance est p.

X_iX_j ne prend que les valeurs 0 ou 1. C'est une variable de Bernoulli, de paramètre p^2. En effet :

$$\begin{aligned} P(X_iX_j = 1) &= P(X_i = 1 \text{ et } X_j = 1) \\ &= P(X_i = 1) \times P(Xj = 1) \quad \text{(indépendance des variables)} \\ &= p^2 \end{aligned}$$

Donc $E(X_iX_j) = p^2$. Enfin, il y a $\frac{n(n-1)}{2}$ couples (i, j) tels que i soit strictement inférieur à j. Ainsi :

$$E(X^2) = np + n(n-1)p^2$$

$$\Rightarrow \quad V(X) = np + n(n-1)p^2 - (np)^2 = np(1-p)$$

❑ *Loi uniforme sur* $\{1, ..., n\}$:

La probabilité que $X = i$ vaut $\frac{1}{n}$. Donc :

$$E(X) = \sum_{i=1}^{n} \frac{i}{n} = \frac{n+1}{2}$$

$$E(X^2) = \sum_{i=1}^{n} \frac{i^2}{n} = \frac{(n+1)(2n+1)}{6}$$

$$V(X) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}$$

4- Inégalité de Bienaymé-Tchebychev

PROPOSITION

Soit X une variable aléatoire d'espérance μ *et d'écart-type* σ, *et soit* ε *un réel strictement positif. Alors* : $P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$

Démonstration :

Posons $Y = |X - \mu|$. On a $\sigma^2 = V(X) = E(Y^2)$. Notons $\mathbf{1}_{Y\geq\varepsilon}$ la fonction indicatrice de l'ensemble $\{\omega \in \Omega, Y(\omega) \geq \varepsilon\}$. Alors, on a :

$$Y^2 \geq \varepsilon^2\, \mathbf{1}_{Y\geq\varepsilon}$$

En effet, ou bien $Y(\omega) \geq \varepsilon$ et l'inégalité ci-dessus donne bien le même résultat, puisqu'alors $\mathbf{1}_{Y\geq\varepsilon}(\omega) = 1$; ou bien $Y(\omega) < \varepsilon$, est l'inégalité ci-dessus donne $Y^2 \geq 0$, puisque $\mathbf{1}_{Y\geq\varepsilon}(\omega) = 0$. Dans tous les cas, l'inégalité est vérifiée. On en déduit que :

$$\sigma^2 = E(Y^2) \geq E(\varepsilon^2\, \mathbf{1}_{Y\geq\varepsilon}) = \varepsilon^2\, E(\mathbf{1}_{Y\geq\varepsilon}) = \varepsilon^2\, P(Y \geq \varepsilon)$$

D'où le résultat.

Cette inégalité permet d'évaluer la probabilité de s'écarter de sa moyenne. Son avantage certain est qu'elle ne dépend pas de la loi de X. Cette loi peut être ignorée. Son inconvénient est que la majoration est grossière. Pour une loi binomiale $\mathcal{B}(20, \frac{1}{2})$ et $\varepsilon = 3$, on a $\sigma^2 = 5$, $\mu = 10$ et l'inégalité de Bienaymé-Tchebychev donne :

$$P(|X - 10| \geq 3) \leq \frac{5}{9} \approx 0{,}56$$

alors que le calcul donne $P(|X - 10| \geq 3) \approx 0{,}26$

Notons que l'argument utilisé sur la variable aléatoire Y^2 se généralise à n'importe quelle variable aléatoire Z positive ou nulle sous la forme :

$$\forall\, a > 0,\ E(Z) \geq a\, P(Z \geq a) \qquad \text{(inégalité de Markov)}$$

en remarquant que $Z \geq a\, \mathbf{1}_{Z\geq a}$ et en prenant l'espérance des deux membres.

Applications :

❑ Si $X \geq 0$ et $E(X) = 0$, alors $P(X = 0) = 1$.

En effet, pour tout $a > 0$, l'inégalité de Markov donne $0 \geq a\, P(X \geq a)$, ou encore $P(X \geq a) = 0$. Donc en prenant pour a la plus petite valeur strictement positive susceptible d'être prise par X, on obtient $P(X > 0) = 0$ et par passage au complémentaire, $P(X = 0) = 1$.

❑ Si $V(X) = 0$, alors il existe μ tel que $P(X = \mu) = 1$.

En effet, l'inégalité de Bienaymé-Tchebichev donne $P(\left|X - \mu\right| \geq \varepsilon) = 0$ pour tout $\varepsilon > 0$, ce qui permet de conclure que $P(\left|X - \mu\right| > 0) = 0$ comme ci-dessus, et par passage au complémentaire, $P(X = \mu) = 1$.

Dans le premier cas, on dit que X est presque sûrement nulle (ou presque partout nul), dans le second, que X est presque sûrement constante.

❑ Considérons une suite de variables aléatoires indépendantes X_i, $1 \leq i \leq n$ de même loi, d'espérance μ et d'écart-type σ. Appliquons l'inégalité de Bienaymé-Tchebichev à $Z_n = \dfrac{X_1 + ... + X_n}{n}$. On a $E(Z_n) = \mu$ et $V(Z_n) = \dfrac{\sigma^2}{n}$. Donc :

$$P(\left|Z_n - \mu\right| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

quantité qui tend vers 0 quand n tend vers l'infini. Cela exprime que la probabilité que Z_n appartienne à l'intervalle $]\mu - \varepsilon, \mu + \varepsilon[$ tend vers 1 quand n tend vers l'infini, et ceci, quel que soit le nombre ε choisi. Cette propriété s'appelle la loi faible des grands nombres.

IV : Couple de variables aléatoires

1-Définition

On considère un espace probabilisé (Ω, P) et deux variables aléatoires X et Y définies sur Ω. (X, Y) est un couple de variables aléatoires. C'est l'application de Ω à valeurs dans $\mathbb{R}^2$, définie naturellement par :

$$\omega \in \Omega \to (X(\omega), Y(\omega)) \in \mathbb{R}^2$$

Si X est à valeurs dans $\{x_i, i \in I\}$ et Y à valeurs dans $\{y_j, j \in J\}$ où I et J sont deux ensembles finis, alors (X,Y) est à valeurs dans $\{(x_i, y_j), (i, j) \in I \times J\}$. Pour déterminer la loi du couple, dite loi conjointe, il suffit donc de donner la probabilité des événements élémentaires $P(X = x_i \cap Y = y_j)$ noté également $P(X = x_i, Y = y_j)$. Quant à la probabilité d'une partie A quelconque de $\mathbb{R}^2$, on aura, comme pour une variable aléatoire à valeurs réelles :

$$P((X, Y) \in A) = P(\{\omega \in \Omega \mid (X(\omega), Y\,\omega)) \in A\}$$

$$= \sum_{(i,j) \text{ tels que } (x_i, y_j) \in A} P(X = x_i, Y = y_j)$$

EXEMPLE :

On lance deux dés, X est le plus grand numéro tiré, Y le plus petit. Quelle est la loi conjointe ? On la présentera sous la forme d'un tableau à double entrée.

X↓ Y→	1	2	3	4	5	6	loi de X
1	1/36	0	0	0	0	0	1/36
2	2/36	1/36	0	0	0	0	3/36
3	2/36	2/36	**1/36**	0	0	0	5/36
4	2/36	**2/36**	**2/36**	**1/36**	0	0	7/36
5	**2/36**	**2/36**	**2/36**	**2/36**	**1/36**	0	9/36
6	**2/36**	**2/36**	**2/36**	**2/36**	**2/36**	**1/36**	11/36
loi de Y	11/36	9/36	7/36	5/36	3/36	1/36	total 1

La convention est ici de mettre les valeurs prises par X dans la première colonne et celles prises par Y dans la première ligne. On prendra garde qu'on aurait pu prendre la convention inverse, ce qui aurait conduit à inverser le rôle des lignes et des colonnes dans toute la suite de ce paragraphe.

Quelle est la probabilité de l'événement X + Y > 5 ? Cette événement consiste en l'appartenance de (X, Y) à la partie A = $\{(x, y) \mid 1 \leq x \leq 6,\ 1 \leq y \leq 6,\ x + y > 5\}$. Les probabilités des événements élémentaires correspondants sont indiquées en gras dans le tableau ci-dessus. On trouve donc pour P(X + Y > 5), égal à P((X, Y) ∈ A), la valeur $\frac{26}{36}$ ou $\frac{13}{18}$.

2- Lois marginales

Un couple étant donné, on appelle lois marginales les lois de X et de Y. Les lois marginales ont été indiquées dans l'exemple précédent. On remarque qu'elles se trouvent facilement à partir de la loi conjointe par addition des lignes ou des colonnes. En effet :

$$X = x_i \Leftrightarrow \exists\, j \text{ tq } (X, Y) = (x_i, y_j)$$

donc $\{\omega \in \Omega \mid X(\omega) = x_i\} = \bigcup_{j\in J} \{\omega \in \Omega \mid X(\omega) = x_i, Y(\omega) = y_j\}$. Donc :

$$P(X = x_i) = \sum_{j\in J} P(X = x_i, Y = y_j\} \qquad \text{(somme des termes de la ligne } i\text{).}$$

De même,

$$p(Y = y_j\,) = \sum_{i\in I} P(X = x_i, Y = y_j\} \qquad \text{(somme des termes de la colonne } j\text{).}$$

On remarquera que la connaissance des lois marginales ne permet pas de reconstituer la loi du couple. On possède moins d'informations.

EXEMPLE :

X↓ Y→	0	1	loi X
0	p	$1/2 - p$	1/2
1	$1/2 - p$	p	1/2
loi de Y	1/2	1/2	total 1

Quelle que soit la valeur de p, entre 0 et $\frac{1}{2}$, la loi conjointe admet les mêmes lois marginales.

3- Lois conditionnelles

Soit (X,Y) un couple de variables aléatoires sur Ω. On suppose l'événement $Y = y_j$ réalisé. On peut alors munir Ω de la loi conditionnelle $P_{Y=y_j}$. La loi de X sur cet espace s'appelle loi de X sachant $Y = y_j$. On a donc :

$$P_{Y=y_j}(X = x_i) = P(X = x_i \mid Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$

La loi de X sachant $Y = y_j$ se trouve donc en divisant la j-ème colonne par $P(Y = y_j)$, qui n'est autre que la somme de cette j-ème colonne. Le calcul peut être fait pour toutes les valeurs de y_j, et le résultat présenté sous la forme d'un tableau à double entrée, lorsque les valeurs sont en nombre fini. On pourrait également chercher la loi de Y sachant que $X = x_i$.

Dans le cas du tirage de deux dés, où X représente le plus grand tirage et Y le plus petit, on obtient :

Loi de X↓ sachant que →	Y = 1	Y = 2	Y = 3	Y = 4	Y = 5	Y = 6
1	1/11	0	0	0	0	0
2	2/11	1/9	0	0	0	0
3	2/11	2/9	1/7	0	0	0
4	2/11	2/9	2/7	1/5	0	0
5	2/11	2/9	2/7	2/5	1/3	0
6	2/11	2/9	2/7	2/5	2/3	1

La loi de X sachant que $Y = j$ se lit dans la j-ème colonne

Loi de Y → sachant que ↓	1	2	3	4	5	6
X = 1	1	0	0	0	0	0
X = 2	2/3	1/3	0	0	0	0
X = 3	2/5	2/5	1/5	0	0	0
X = 4	2/7	2/7	2/7	1/7	0	0
X = 5	2/9	2/9	2/9	2/9	1/9	0
X = 6	2/11	2/11	2/11	2/11	2/11	1/11

La loi de Y sachant que $X = i$ se lit dans la i-ème ligne

Les lois conditionnelles présentent l'intérêt suivant. Si l'on connaît la loi de Y et les lois conditionnelles de X sachant Y, alors on peut reconstituer la loi du couple. En effet :

$$P(X = x_i, Y = y_j) = P(Y = y_j) \times P_{Y=y_j}(X = x_i)$$

4- Variables indépendantes

a) Cas de deux variables

Deux variables aléatoires X et Y à valeurs respectivement dans $\{x_i, i \in I\}$ et $\{y_j, j \in J\}$, sont dites indépendantes si, pour tout événement A et B :

$$P(X \in A, Y \in B) = P(X \in A) \times P(Y \in B)$$

En particulier, pour tout i et tout j, les événements $X = x_i$ et $Y = y_j$ sont indépendants. On a donc :

$$\forall i \in I, \forall j \in J, P(X = x_i, Y = y_j) = P(X = x_i) \times P(Y = y_j)$$

Connaissant les lois marginales, on est alors capable de reconstituer la loi du couple. Les lois conditionnelles de X sachant $Y = y_j$ sont identiques à la loi de X. Réciproquement, il suffit de vérifier les indépendances des événements $X = x_i$ et $Y = y_j$ pour conclure à l'indépendance de X et Y, car, en sommant sur les éléments d'un événement A et d'un événement B, on a retrouve la définition intiale. Le fait que deux variables soient indépendantes résulte soit d'une hypothèse portée sur le modèle (indépendance des lancers de deux dés...), soit de la vérification par le calcul de l'indépendance de X = x_i et Y = y_j.

EXEMPLE 1 :
Soit (X,Y) un couple à valeurs dans $\{1, ..., n\} \times \{1, ..., n\}$. On pose :

$$P(X = i, Y = j) = C \times \frac{i}{j}$$

où C est une constante telle que la somme des probabilités soit égale à 1. Les variables sont-elles indépendantes ?

La loi de X est :

$$P(X = i) = \sum_{j=1}^{n} P(X = i,\ Y = j) = \sum_{j=1}^{n} C \times \frac{i}{j} = CK \times i$$

où $K = \sum_{j=1}^{n} \frac{1}{j}$. C et K sont reliés par la relation suivante :

$$1 = \sum_{i=1}^{n} P(X = i) = CK \sum_{i=1}^{n} i$$

La loi de Y est :

$$P(Y = j) = \sum_{i=1}^{n} P(X = i,\ Y = j) = \sum_{i=1}^{n} C \times \frac{i}{j} = \frac{C}{j} \sum_{i=1}^{n} i = \frac{1}{Kj}$$

On a donc bien :

$$P(X = i) \times P(Y = j) = \frac{Ci}{j} = P(X = i, Y = j).$$

En fait, on peut prévoir l'indépendance des variables si la loi conjointe est de la forme suivante :

$$P(X = i, Y = j) = f(i)g(j)$$

En effet, soit $G = \sum_{j} g(j)$ et $F = \sum_{i} f(i)$. Alors F et G sont reliés par la relation :

$$FG = \sum_{i,j} f(i)g(j) = \sum_{i,j} P(X = i, Y = j) = 1$$

$$P(X = i) = \sum_{j} f(i)g(j) = f(i)G$$

$$P(Y = j) = \sum_{i} f(i)g(j) = g(j)F$$

D'où $P(X = i) \times P(Y = j) = FG \times f(i)g(j) = f\{i)g(j) = P(X = i, Y = j)$ et les variables sont bien indépendantes. Cette propriété apparaît dans le tableau donnant la loi du couple quand les lignes (ou les colonnes) sont proportionnelles entre elles.

EXEMPLE 2 :
Soit (X, Y) un couple à valeurs dans $\{1, ..., n\} \times \{1, ..., n\}$. On pose :

$$P(X = i, Y = j) = C(i + j) \text{ où C est une constante.}$$

Les variables sont-elles indépendantes ?

La loi de X est :

$$P(X = i) = \sum_{j=1}^{n} P(X = i,\ Y = j) = \sum_{j=1}^{n} C(i + j) = C(ni + \frac{n(n+1)}{2})$$

C vérifie la relation suivante : $1 = \sum_{i=1}^{n} P(X = i) = Cn^2(n + 1) = 1$

La loi de Y est, par symétrie :

$$P(Y = j) = C(nj + \frac{n(n+1)}{2})$$

On a donc :

$$P(X = i) \times P(Y = j) \neq P(X = i, Y = j).$$

Les variables ne sont pas indépendantes.

PROPOSITION

Soit $X : \Omega \to \mathbb{R}$ *et* $Y : \Omega \to \mathbb{R}$ *deux variables aléatoires indépendantes. Alors pour toute fonction f et g de* $\mathbb{R}$ *dans* $\mathbb{R}$, $f(X)$ *et* $g(Y)$ *sont indépendantes.*

Démonstration :

Pour toute valeur u_i et v_j que peuvent prendre $f(X)$ et $g(Y)$, on a :

$$\begin{aligned} P(f(X) = u_i,\ g(Y) = v_j) &= P(X \in f^{-1}(u_i),\ Y \in g^{-1}(v_j)) \\ &= P(X \in f^{-1}(u_i))\ P(Y \in g^{-1}(v_j)) \\ &= P(f(X) = u_i)\ P(g(Y) = v_j) \end{aligned}$$

b) Cas de *n* variables

n variables aléatoires X_1, X_2, ... X_n à valeurs respectivement dans $E_1 = \{x_{i_1},\ i_1 \in I_1\}$, ..., $E_n = \{x_{i_n},\ i_n \in I_n\}$, sont dites indépendantes si, pour tout événement A_1 de E_1, ..., A_n de E_n les événements $X_1 \in A_1$, ..., $X_n \in A_n$ sont indépendants, ou, de manière équivalente, si pour tout i_1, ..., tout i_n, les événements $X_1 = x_{i_1}$, ..., $X_n = x_{i_n}$ sont indépendants. Il suffit de vérifier que :

$$\forall\ i_1 \in I_1, \quad ..., \ \forall\ i_n \in I_n,\ P(X_1 = x_{i_1}, ..., X_n = x_{i_n}) = P(X_1 = x_{i_1}) ... P(X_n = x_{i_n})$$

car une relation analogue sera vérifiée pour toute sous-famille d'événements en sommant de manière adéquate les indices.

On peut vérifier les propriétés suivantes :

i) Toute sous-famille extraite de $(X_1, ..., X_n)$ est indépendante

ii) si f est une fonction de $E_1 \times ... \times E_k$ et g une fonction de $E_{k+1} \times ... \times E_n$ à valeurs dans $\mathbb{R}$, alors les deux variables $Y = f(X_1, ..., X_k)$ et $Z = g(X_{k+1}, ..., X_n)$ sont indépendantes.

5- Espérance d'une fonction d'un couple

a) Cas du produit XY :

On souhaite calculer E(XY), connaissant la loi du couple (X, Y). On a :

$$E(XY) = \sum_{\omega \in \Omega} X(\omega)Y(\omega)\ P(\{\omega\})$$

Nous reprenons ci-dessous un raisonnement comparable à celui que nous avons tenu pour une seule variable aléatoire dans III.3). Soient x_1, x_2 , ..., x_n les valeurs que prend X et y_1, y_2, ..., y_m les valeurs prises par Y. Alors les images réciproques $X^{-1}(\{x_1\})$, $X^{-1}(\{x_2\})$, ..., $X^{-1}(\{x_n\})$ forment un système complet d'événements de Ω : ils sont disjoints (car un élément *u* ne peut avoir deux images différentes) et leur réunion est Ω (car tout ω admet comme image l'un des x_i) . Il en est de même de $Y^{-1}(\{y_1\})$, $Y^{-1}(\{y_2\})$, ..., $Y^{-1}(\{y_n\})$. Et il en est encore de même de la famille $X^{-1}(\{x_i\}) \cap Y^{-1}(\{y_j\})$, $1 \leq i \leq n$, $1 \leq j \leq m$. On peut calculer la somme sur Ω en sommant d'abord sur chaque $X^{-1}(\{x_i\}) \cap Y^{-1}(\{y_j\})$, puis en faisant la somme de ces $n \times m$ totaux :

$$\sum_{\omega\in\Omega} X(\omega)Y(\omega)\, P(\{\omega\}) = \sum_{i,j} \sum_{\omega\in X^{-1}(\{x_i\})\cap Y^{-1}(\{y_j\})} X(\omega)Y(\omega)\, P(\{\omega\})$$

D'autre part, lorsque ω appartient à $X^{-1}(\{x_i\}) \cap Y^{-1}(\{y_j\})$, par définition, $X(\omega)$ est égal à s_i et $Y(\omega)$ à y_j. D'où :

$$\sum_{i,j} \sum_{\omega\in X^{-1}(\{x_i\})\cap Y^{-1}(\{y_j\})} X(\omega)Y(\omega)\, P(\{\omega\}) = \sum_{i,j} \sum_{\omega\in X^{-1}(\{x_i\})\cap Y^{-1}(\{y_j\})} x_i\, y_j\, P(\{\omega\})$$

$$= \sum_{i,j} x_i\, y_j \sum_{\omega\in X^{-1}(\{x_i\})\cap Y^{-1}(\{y_j\})} P(\{\omega\})$$

Enfin $\displaystyle\sum_{\omega\in X^{-1}(\{x_i\})\cap Y^{-1}(\{y_j\})} P(\{\omega\})$ n'est autre que :

$$P(X^{-1}(\{x_i\}) \cap Y^{-1}(\{y_j\})) = P(\{\omega\in \Omega \mid X(\omega) = x_i \cap Y(\omega) = y_j\}) = P(X = x_i, Y = y_j) .$$

Donc :

$$\sum_{\omega\in\Omega} X(\omega)Y(\omega)\, P(\{\omega\}) = \sum_{i,j} x_i\, y_j\, P(X = x_i,\, Y = y_j)$$

Ainsi :

$$E(XY) = \sum_{i,j} x_i\, y_j\, P(X = x_i,\, Y = y_j)$$

Examinons maintenant le cas où X et Y sont indépendantes. Alors, la formule obtenue devient :

$$E(XY) = \sum_{i,j} x_i\, y_j\, P(X = x_i)\, P(Y = y_j)$$

$$= \sum_{i} x_i\, P(X = x_i) \sum_{j} y_j\, P(Y = y_j)$$

$$= E(X)\, E(Y)$$

On peut donc énoncer la proposition suivante :

PROPOSITION

Soit X *et* Y *deux variables indépendantes, alors* E(XY) = E(X)E(Y)

Cette propriété se généralise par récurrence à un produit quelconque de variables aléatoires indépendants.

Si X et Y ne sont pas indépendantes, en général, il n'y a pas égalité.

EXEMPLE :

Soit $0 < p < \frac{1}{2}$

X↓ Y→	0	1	loi de X
0	p	$1/2 - p$	1/2
1	$1/2 - p$	p	1/2
loi de Y	1/2	1/2	total 1

XY suit une loi de Bernoulli de paramètre p. Donc E(XY) = p. Mais E(X) = E(Y) = $\frac{1}{2}$

On prendra garde que la proposition énonce une implication et non une équivalence. Il est tout à fait possible que E(XY) = E(X)E(Y) sans que X et Y soient indépendantes.

EXEMPLE :
On dispose de trois urnes. On y place trois boules au hasard. X est le nombre de boules dans l'urne 1, et Y est le nombre d'urnes vides. La loi conjointe est :

X↓ Y→	0	1	2	loi de X
0	0	6/27	2/27	8/27
1	6/27	6/27	0	12/27
2	0	6/27	0	6/27
3	0	0	1/27	1/27
loi de Y	2/9	6/9	1/9	total 1

$$\mathrm{E(X)} = \frac{12}{27} + 2 \times \frac{6}{27} + 3 \times \frac{1}{27} = \frac{4}{9} + \frac{4}{9} + \frac{1}{9} = 1$$

$$\mathrm{E(Y)} = \frac{6}{9} + 2 \times \frac{1}{9} = \frac{8}{9}$$

$$\mathrm{E(XY)} = \frac{6}{27} + 2 \times \frac{6}{27} + 6 \times \frac{1}{27} = \frac{24}{27} = \frac{8}{9} = \mathrm{E(X)E(Y)}$$

Pourtant, X et Y ne sont pas indépendantes.

b) Cas d'une fonction *h*(X,Y) :
Le raisonnement appliqué au produit se généralise de la façon suivante

$$\mathrm{E}(h(\mathrm{X},\mathrm{Y})) = \sum_{i,j} h(x_i, y_j)\, \mathrm{P}(\mathrm{X} = x_i, \mathrm{Y} = y_j)$$

Dans le cas de deux variables aléatoires indépendantes, on a :

$$\mathrm{E}(h(\mathrm{X},\mathrm{Y})) = \sum_{i,j} h(x_i, y_j)\, \mathrm{P}(\mathrm{X} = x_i)\, \mathrm{P}(\mathrm{Y} = y_j)$$

De plus, si $h(\mathrm{X}, \mathrm{Y}) = f(\mathrm{X})\, g(\mathrm{Y})$, on obtient :

$$E(f(X)\,g(Y)) = \sum_{i,j} f(x_i)\,g(y_j)\,P(X = x_i)\,P(Y = y_j) = E(f(X))\,E(g(Y))$$

ce qui découle également directement du fait que $f(X)$ et $g(Y)$ sont deux variables aléatoires indépendantes, et que l'espérance d'un produit de variables aléatoires indépendantes est le produit des espérances.

6- Somme de variables aléatoires

On s'intéresse à la loi de la somme Z = X + Y de deux variables aléatoires finies. L'événement $\{Z = z\}$ est égal à la réunion disjointe des événements $\{X = x_i, Y = y_j \mid x_i + y_j = z\}$. Donc :

$$P(Z = z) = \sum_{(i,\,j)\ \text{tq}\ x_i + y_j = z} P(X = x_i,\ Y = y_j)$$

Dans le cas de deux variables à valeurs dans $\mathbb{N}$, on obtient :

$$P(Z = n) = \sum_{i=1}^{n} P(X = i,\ Y = n - i)$$

Lorsque les variables sont indépendantes, cette formule devient :

$$P(Z = n) = \sum_{i=1}^{n} P(X = i)\,P(Y = n - i)$$

EXEMPLE 1 :
X et Y deux variables aléatoires indépendantes de même loi de Bernouilli de paramètre p :

$$P(X + Y = 0) = P(X = 0)\,P(Y = 0) = (1 - p)^2$$
$$P(X + Y = 1) = P(X = 0)\,P(Y = 1) + P(X = 1)\,P(Y = 0) = 2(1 - p)p$$
$$P(X + Y = 2) = P(X = 1)\,P(Y = 1) = p^2.$$

On reconnaît une loi binomiale $\mathcal{B}(2, p)$, ce qui n'est pas surprenant puisque la loi $\mathcal{B}(2, p)$ est précisément définie comme la loi de la somme de deux variables de Bernoulli indépendantes de même paramètre p.

EXEMPLE 2 :
On généralise ce résultat à la somme de deux variables aléatoires indépendantes suivant une loi binomiale de même paramètre. Soit X de loi $\mathcal{B}(n, p)$ et Y de loi $\mathcal{B}(m, p)$ deux variables aléatoires indépendantes. Alors X + Y admet pour loi une loi $\mathcal{B}(n + m, p)$. En effet :

$$P(X + Y = k) = \sum_{r} P(X = r)\,P(Y = k - r)$$

$$= \sum_{r} \binom{n}{r} p^r (1 - p)^{n-r} \binom{m}{k - r} p^{k-r} (1 - p)^{m-k+r}$$

$$= \sum_{r} \binom{n}{r}\binom{m}{k - r} p^k (1 - p)^{n+m-k}$$

Or $\sum_r \binom{n}{r}\binom{m}{k-r} = \binom{n+m}{k}$. En effet, $\binom{n+m}{k}$ est le nombre de parties de k éléments parmi $n + m$ et on peut dénombrer ce nombre de parties d'une autre façon. Il suffit de choisit r éléments parmi n, et $k - r$ parmi m. On obtient finalement :

$$P(X + Y = k) = \binom{n+m}{k} p^k (1-p)^{n+m-k}$$

qui est bien la loi binomiale annoncée. Ce résultat est parfaitement compréhensible si on imagine que X est le nombre de Pile en n lancers, et Y le nombre de P en m autres lancers. Alors, X + Y est le nombre de Pile en $n + m$ lancers.

7- Covariance

Début de partie réservée aux MPSI

On s'intéresse à la variance d'un somme :

$$\begin{aligned} V(X + Y) &= E((X + Y - E(X) - E(Y))^2) \\ &= E((X - E(X))^2 + (Y - E(Y))^2 + 2(X - E(X))(Y - E(Y))) \\ &= V(X) + V(Y) + 2\,E((X - E(X))(Y - E(Y))) \end{aligned}$$

On pose alors la covariance de X et Y comme égale à :

$$\text{cov}(X,Y) = E((X - E(X))(Y - E(Y)))$$

D'où :

$$V(X + Y) = V(X) + V(Y) + 2\,\text{cov}(X,Y)$$

(Cette formule est à rapprocher de ce qui ce passe sur $\mathbb{R}$ avec : $(x + y)^2 = x^2 + y^2 + 2xy$. Le rôle du double produit est joué ici par la covariance).

La formule ci-dessus se généralise à n variables de la façon suivante :

$$V(X_1 + ... + X_n) = V(X_1) + ... + V(X_n) + 2 \sum_{i,j} \text{cov}(X_i, X_j)$$

PROPOSITION

La covariance vérifie les propriétés suivantes :

i) $\text{cov}(X,X) = V(X)$

ii) $\text{cov}(X,Y) = E(XY) - E(X)E(Y)$

iii) $(X,Y) \to \text{cov}(X,Y)$ est bilinéaire

iv) $|\text{cov}(X,Y)| \leq \sqrt{V(X)}\sqrt{V(Y)}$

v) X *et* Y *indépendantes* $\Rightarrow \text{cov}(X, Y) = 0$

vi) X *et* Y *indépendantes* $\Rightarrow V(X + Y) = V(X) + V(Y)$

<u>Démonstration</u> :

i) est évident

ii) se montre en développant cov(X,Y) :

$$\begin{aligned} \text{cov}(X.Y) &= E((X - E(X))(Y - E(Y))) \\ &= E(XY - XE(Y) - YE(X) + E(X)E(Y)) \\ &= E(XY) - E(XE(Y)) - E(YE(X)) + E(E(X)E(Y)) \\ &= E(XY) - E(Y)E(X) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

iii) La vérification est laissée au lecteur

iv) résulte du fait qu'une forme bilinéaire symétrique positive vérifie l'inégalité de Schwarz.

v) Si X et Y sont indépendantes, alors E(XY) = E(X)E(Y). La propriété v) en résulte.

vi) est clairement équivalente à v).

On définit le coefficient de corrélation de X et Y comme étant égal à $\rho(X, Y) = \dfrac{\text{cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$. Ce nombre est compris entre –1 et 1, d'après la relation iv). Il joue un rôle analogue à celui du cosinus pour deux vecteurs.

Si X et Y sont indépendantes, alors le coefficient de corrélation est nul. On dit que les variables sont non corrélées. On prendra garde que cette condition n'est pas équivalente à l'indépendance de X et Y (pas plus que le fait que E(XY) = E(X)E(Y) n'entraîne l'indépendance de X et de Y).

Si $Y = aX + b$, alors $\text{cov}(X,Y) = aV(X)$ et le coefficient de corrélation vaut 1 si $a > 0$ et – 1 si $a < 0$.
Fin de partie réservée aux MPSI

◆