PROBABILITES - 2ème partie

Plan

I : Probabilité sur un univers quelconque

1- Ensemble dénombrable

Nous étendons ci-dessous les notions de probabilité sur un univers fini au cas d'un ensemble Ω quelconque. Nous commençons par le cas où Ω est dénombrable, i.e. qui est en bijection avec $\mathbb{N}$. Cela signifie qu'on peut décrire Ω sous la forme $\{a_n, n \in \mathbb{N}\}$, les a_n étant distincts.

❑ Toute partie A d'un ensemble Ω dénombrable est finie ou dénombrable. Il suffit en effet de parcourir les éléments a_0, a_1, a_2, ..., a_n, ... de Ω et d'appeler b_0 le premier élément de la liste appartenant à A, b_1 l'élément suivant, b_2 l'élément suivant, etc... A est fini ou dénombrable suivant que la liste des b_i s'arrête ou se prolonge indéfiniment.

❑ Toute image d'un ensemble dénombrable par une application f est finie ou dénombrable. Il suffit en effet de parcourir les éléments de $f(\Omega)$: $f(a_0)$, $f(a_1)$, ..., $f(a_n)$, ... et de poser $b_0 = f(a_0)$, b_1 égal au premier $f(a_n)$ différent de b_0, b_2 le $f(a_n)$ suivant différent de b_0 et b_1, etc...

❑ La réunion de deux ensembles dénombrables est dénombrable. En effet, si ces deux ensembles sont $\{a_n, n \in \mathbb{N}\}$ et $\{b_n, n \in \mathbb{N}\}$, on peut obtenir leur réunion de la façon suivante :

i) ranger les éléments dans l'ordre $a_0, b_0, a_1, b_1, \ldots, a_n, b_n, \ldots$

ii) éliminer de la liste les éléments identiques (l'un des b_i peut en effet être égal à l'un des a_j) pour n'en garder qu'un exemplaire.

iii) énumérer la liste finale obtenue en désignant ses éléments par $c_0, c_1, c_2, \ldots, c_n, \ldots$

EXEMPLE :
$\mathbb{Z}$ est dénombrable. Il est en effet la réunion de $\mathbb{N}$ et de $\{-n, n \in \mathbb{N}^*\}$. Or $\{-n, n \in \mathbb{N}^*\}$ au moyen de la bijection $p \in \mathbb{N} \to -p-1$. On peut donner explicitement une bijection entre $\mathbb{N}$ et $\mathbb{Z}$:

$$f : \mathbb{N} \to \mathbb{Z}$$
$$n \to n/2 \text{ si } n \text{ est pair}$$
$$-\frac{n+1}{2} \text{ si } n \text{ est impair}$$

❑ Le produit de deux ensembles dénombrables est dénombrable. Montrons-le d'abord dans le cas de $\mathbb{N}^2$. Il suffit d'énumérer ses éléments dans l'ordre suivant :

$c_0 = (0,0)$				
$c_1 = (1,0)$	$c_2 = (0,1)$			
$c_3 = (2,0)$	$c_4 = (1,1)$	$c_5 = (0,2)$		
$c_6 = (3,0)$	$c_7 = (2,1)$	$c_8 = (1,2)$	$c_9 = (0,3)$	
$c_{10} = (4,0)$	$c_{11} = (3,1)$	$c_{12} = (2,2)$	$c_{13} = (1,3)$	$c_{14} = (0,4)$
...				
$c_{n(n-1)/2} = (n-1,0)$	$c_{n(n-1)/2+1} = (n-2, 1)$	...		$c_{n(n+1)/2-1} = (0, n-1)$
...				

Le cas général du produit $\{(a_n, b_m), n \in \mathbb{N}, m \in \mathbb{N}\}$ se montre exactement de la même façon en remplaçant ci-dessus les couples (n, m) par (a_n, b_m).
En particulier $\mathbb{Q}$ est dénombrable. En effet $\mathbb{Q}^+$ peut s'injecter dans $\mathbb{N}^2$ au moyen d'une application du type $\frac{p}{q} \to (p,q)$, où p et q sont choisis sans facteur commun. De même pour $\mathbb{Q}^-$.

A titre indicatif, voici une bijection curieuse entre $\mathbb{Q}^{+*}$ et $\mathbb{N}$. On définit la fonction f de $\mathbb{N}$ dans $\mathbb{N}^*$ de la façon suivante :

$$f(0) = 1 \text{ et } \forall n, f(2n+1) = f(n), f(2n+2) = f(n) + f(n+1)$$

de sorte que les valeurs de f sont :

$$1, 1, 2, 1, 3, 2, \underset{n}{\underset{\uparrow}{3}}, \underset{n+1}{\underset{\uparrow}{1}}, 4, 3, 5, 2, 5, \underset{2n+1}{\underset{\uparrow}{3}}, \underset{2n+2}{\underset{\uparrow}{4}}, 1, 5, 4, 7, 3, 8 \ldots$$

Les valeurs successives de $\frac{f(n)}{f(n+1)}$ sont :

$$1, \frac{1}{2}, 2, \frac{1}{3}, \frac{3}{2}, \frac{2}{3}, 3, \frac{1}{4}, \frac{4}{3}, \frac{3}{5}, \frac{5}{2}, \frac{2}{5}, \frac{5}{3}, \frac{3}{4}, 4, \frac{1}{5}, \frac{5}{4}, \frac{4}{7}, \frac{7}{3}, \frac{3}{8} \ldots$$

On montre que tous les rationnels positifs apparaissent une fois et une seule dans cette liste, de sorte que l'application $n \to \frac{f(n)}{f(n+1)}$ forme une bijection de $\mathbb{N}$ dans $\mathbb{Q}^{+*}$.

La suite précédente peut également être générée par la suite récurrence suivante[1] :

$x_0 = 0$

$\forall\ n \geq 0,\ x_{n+1} = \varphi(x_n)$ avec $\varphi(x) = \dfrac{1}{1 + 2\lfloor x \rfloor - x}$, où $\lfloor x \rfloor$ désigne la partie entière de x.

2- Probabilité sur un espace dénombrable

Ω est l'ensemble des résultats possibles d'une épreuve aléatoire. La situation où Ω est dénombrable se rencontre lorsque ces résultats peuvent être en nombre infini et peuvent être énumérés. On peut également choisir un tel Ω infini dénombrable lorsque le nombre de résultats d'une expérience aléatoire est fini, mais borné par un nombre indéterminé. Les ensembles finis ou dénombrables sont appelés ensembles discrets.

EXEMPLES :

- Lancer une pièce jusqu'à ce qu'il apparaisse P(ile). Au choix, ou bien $\Omega = \{F^nP \mid n \in \mathbb{N}\}$ où F^n désigne un suite de n F(aces) ; ou bien directement $\Omega = \mathbb{N}^*$ si on considère le résultat comme le nombre de lancers. Si on envisage la possibilité que Pile n'apparaisse jamais, on ajoutera un élément supplémentaire pour cette éventualité, de sorte que $\Omega = \{F^nP \mid n \in \mathbb{N}\} \cup \{F^\infty\}$ ou bien directement $\Omega = \mathbb{N}^* \cup \{\infty\}$
- 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 déterminant le sens du déplacement. Si elle tombe sur Pile, on avance d'une unité, sinon on recule d'une unité. ii) On relance la pièce décidant si on poursuit le mouvement. Si elle tombe sur Pile, on s'arrête. Sinon on recommence à l'étape i). Le résultat de l'épreuve est l'abscisse obtenue, lorqu'on s'arrête. $\Omega = \mathbb{Z}$. On peut montrer que l'éventualité d'un mouvement se poursuivant indéfiniment est nulle.

DEFINITION

On appelle événement une partie A *de* Ω. *L'ensemble des événements est ici l'ensemble des parties de* Ω, *noté* $\mathcal{P}(\Omega)$.

EXEMPLES :

- Lancer une pièce et obtenir Pile pour la première fois entre le 15ème et le 20ème lancer. Suivant l'univers Ω choisi, ou bien $A = \{F^nP \mid 15 \leq n \leq 20\}$; ou bien $A = \{15,16,17,18,19,20\}$
- Compter entre 10 000 et 20 000 voitures au péage de Fontainebleau. $A = \{n \in \mathbb{N} \mid 10\,000 \leq n \leq 20\,000\}$.
- Compter plus de 30 grains de blé sur un épi. $A = \{n \in \mathbb{N} \mid n > 30\}$.
- Terminer la promenade aléatoire sur l'axe avec une abscisse positive ou nulle. $A = \mathbb{N}$.

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

FFFFFFP	A non réalisé
16 258 voitures	A réalisé

[1] Aimeric Malter, Dierk Schleicher, Don Zagler, *New looks at old number theory*, American Math. Monthly, **128**, n°3 (mars 2013), 243-264

59 grains de blé	A réalisé
abcisse –5	A non réalisé

Ω s'appelle **événement certain**. $\varnothing$ est appelé **événement impossible**. Un **événement élémentaire** est constitué d'un singleton. Si $A \cap B = \varnothing$, on dit que les événements A et B sont **incompatibles**. $\complement A$ s'appelle l'événement **contraire** de A.

On peut définir une probabilité sur Ω à partir de la probabilité des événements élémentaires. Pour tout a_n élément de Ω, on pose $p_n = P(\{a_n\})$, où p_n est un élément de [0,1]. Pour toute partie A de Ω, on pose :

$$P(A) = \sum_{n \text{ tel que } a_n \in A} p_n = \sum_{i=0}^{\infty} p_n \, \mathbf{1}_A(a_n)$$

où $\mathbf{1}_A$ est la fonction indicatrice de la partie A, définie comme suit :

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

On notera que, A pouvant contenir une infinité d'éléments, P(A) est défini a priori par la somme d'une série et non par une somme finie. Par ailleurs, nous admettrons que la somme de cette série ne dépend pas de l'ordre dans lequel on somme ses éléments, ce qui signifie que la façon dont on a numéroté les éléments de Ω est sans importance.

La probabilité de l'événement certain étant égal à 1, les p_n doivent être de somme égale à 1.

EXEMPLE :

- On lance une pièce jusqu'à obtenir Pile. Le résultat est le nombre de lancers. On a donc comme ensemble $\Omega = \mathbb{N}^*$, ou même $\mathbb{N}^* \cup \{\infty\}$ si on envisage la possibilité de ne jamais tomber sur Pile. On pose alors, pour n variant élément de $\mathbb{N}^*$:

$$p_n = \frac{1}{2^n}$$

$$p_\infty = 0$$

La valeur de p_∞ est nécessairement nulle. En effet :

$$p_\infty = P(\{\infty\}) = P(\complement \mathbb{N}^*) = 1 - P(\mathbb{N}^*) = 1 - \sum_{n=1}^{\infty} \frac{1}{2^n} = 1 - 1 = 0$$

La probabilité que Pile arrive au bout d'un nombre pair de coups (événement A) est :

$$P(A) = \sum_{n=0}^{\infty} p_{2n} = \sum_{n=0}^{\infty} \frac{1}{4^n} = \frac{1}{3}$$

3- Cas général

Si Ω n'est pas dénombrable (par exemple $\Omega = \mathbb{R}$ ou $\mathbb{R}^2$), on ne peut définir une probabilité sur Ω de la façon précédente. Il est même en général exclu qu'on puisse définir une probabilité sur n'importe quelle partie A de Ω tant ces parties peuvent être compliquées (comment décrire une partie générale de $\mathbb{R}$?). Pour ce faire, on limite les événements A à décrire non pas $\mathcal{P}(\Omega)$ tout entier, mais seulement une famille particulière $\mathcal{T}$ de parties de Ω. On souhaite que cette famille vérifie les propriétés suivantes :

i) $\Omega \in \mathcal{T}$

ii) Si $A \in \mathcal{T}$, alors $\complement A$ aussi

iii) Si $(A_n)_{n\in \mathbb{N}}$ est une famille dénombrable d'éléments de $\mathcal{T}$, alors $\bigcup_{n=0}^{\infty} A_n$ aussi

On dit que $\mathcal{T}$ est une **tribu**. On a alors également :

iv) $\varnothing \in \mathcal{T}$

v) Si (A_n) est une famille finie d'éléments de $\mathcal{T}$, alors $\bigcup A_n$ aussi

vi) Si (A_n) est une famille finie ou dénombrable d'éléments de $\mathcal{T}$, alors $\bigcap A_n$ aussi

En effet, iv) résulte directement de ii) en prenant A = Ω. v) résulte de iii) en adjoignant à la famille finie une infinité dénombrable de parties vides. vi) résulte de iii) ou v) en appliquant ii), le complémentaire d'une réunion étant une intersection, et le complémentaire du complémentaire redonnant l'ensemble lui-même :

$$\bigcap_n A_n = \complement(\bigcup_n \complement A_n)$$

Bref, l'ensemble des événements est stable par les opérations suivantes : passage au complémentaire, réunion finie ou dénombrable, intersection finie ou dénombrable.

Par exemple, si $\Omega = \mathbb{R}$, on prend pour $\mathcal{T}$ la tribu engendrée par les intervalles, i.e. la plus petite famille contenant les intervalles et vérifiant les propriétés ii) et iii) sans qu'on cherche davantage à décrire cette famille.

Un exemple encore moins simple est donné par le lancer successif d'une pièce jusqu'à ce qu'un certain événement soit observé. En pratique, la pièce est lancé un nombre fini de fois, mais ce nombre n'est peut-être pas borné. Aussi est-il commode de prendre $\Omega = \{0,1\}^{\mathbb{N}}$ (ensemble des suites de 0 ou 1), dont on peut montrer qu'il est non dénombrable. On prendra comme événement au moins tout ce qui dépend d'un nombre fini de lancers, à savoir toute partie de la forme :

$$\{(x_k)_{k\in \mathbb{N}}, \exists\, n \in \mathbb{N}, \exists\, A \subset \{0,1\}^n, (x_k)_{k\leq n} \in A\}$$

Ci-dessus, l'événement est le fait que les n premiers lancers, n étant donné, vérifient une certaine propriété caractérisée par une partie A donnée. On prend alors pour $\mathcal{T}$. la plus petite tribu contenant ces événements. On peut montrer que $\mathcal{T}$. est différent de $\mathcal{P}(\Omega)$ et qu'il est possible de définir une probabilité sur $\mathcal{T}$. correspondant à la modélisation attendue d'un lancer de pièces, mais qu'il n'est pas possible d'étendre cette probabilité à $\mathcal{P}(\Omega)$. On ne précisera pas davantage à quoi peut ressembler $\mathcal{T}$..

DEFINITION

Soit Ω *un ensemble quelconque muni d'une tribu* $\mathcal{T}$. *On dit que* P *est une probabilité sur* $(\Omega, \mathcal{T})$ *si* P *est une application de* $\mathcal{T}$ *dans* [0, 1] *telle que* :

def-i) P(Ω) = 1

def-ii) *Si* $(A_n)_{n\in\mathbb{N}}$ *est une famille d'éléments de* $\mathcal{T}$ *incompatibles (événements disjoints deux à deux), alors* : $P(\bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty} P(A_n)$

$(\Omega, \mathcal{T}, P)$ *s'appelle alors espace probabilisé.*

Dans ii), on accepte de prendre des réunions dénombrables car se limiter à des réunions finies est trop restrictif, comme le montre les propriétés v) et vi) du paragraphe suivant. Le cas infini dénombrable est également nécessaire pour retrouver la définition par somme des probabilités des événements élémentaires dans le cas où Ω est dénombrable. En effet, si on note $\Omega = \{a_n, n \in \mathbb{N}\}$, et si on pose $p_n = P(\{a_n\})$, alors :

$A = \bigcup_{n \text{ tel que } a_n \in A} \{a_n\}$ union disjointe éventuellement infinie dénombrable

donc $P(A) = \sum_{n \text{ tel que } a_n \in A} P(\{a_n\}) = \sum_{n \text{ tel que } a_n \in A} p_n$ comme au ii)

Voici un exemple de probabilité sur un ensemble non dénombrable.

EXEMPLE :
On prend $\Omega = [0,1]$, et pour tout a et b tels que $0 \leq a \leq b \leq 1$. On peut montrer qu'il existe une tribu $\mathcal{T}$ contenant tous les intervalles et une probabilité P définie sur $\mathcal{T}$ telle :

$\forall (a, b) \in \Omega^2, P([a, b]) = b - a$

On remarque que, pour $a = b$, $P(\{a\}) = 0$ et que P ne peut être définie point par point, comme dans le cas dénombrable. C'est aussi le cas de la loi normale, vue en Terminale et définie par une intégrale.

4- Propriétés des probabilités

On retrouve ci-dessous des propriétés vues en première année dans le cas fini. D'autres sont nouvelles.

PROPOSITION :

i) $P(\varnothing) = 0$

def-ii bis) *Si* $(A_k)_{1\leq k\leq n}$ *est une famille finie d'éléments de* $\mathcal{T}$ *incompatibles, alors* :

$$P(\bigcup_{k=1}^{n} A_n) = \sum_{k=1}^{n} P(A_n)$$

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

$$P(A \cup B) \leq P(A) \cup P(B)$$

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* :

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

v) *Pour toute suite croissante* $(A_n)_{n\in\mathbb{N}}$ *d'événements* (i.e. $A_n \subset A_{n+1}$), *on a* :

$$P(\bigcup_{n=0}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$$

vi) *Pour toute suite décroissante* $(A_n)_{n\in\mathbb{N}}$ *d'événements* (i.e. $A_{n+1} \subset A_n$), *on a* :

$$P(\bigcap_{n=0}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$$

vii) *Pour toute suite quelconque* $(A_n)_{n\in\mathbb{N}}$ *d'événements* :

$$P(\bigcup_{n=0}^{\infty} A_n) \leq \sum_{n=0}^{\infty} P(A_n) \qquad \text{(\textit{propriété de sous-additivité})}$$

Démonstration :

i) Appliquer def-ii) avec les A_n tous égaux à $\varnothing$.

def-ii bis) Appliquer def-ii en complétant la famille finie par une infinité dénombrable d'ensembles vides.

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.

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.

v) Posons $A = \bigcup_{n=0}^{\infty} A_n$ et $B_0 = A_0$, $B_n = A_n - A_{n-1}$ pour $n \geq 1$. Alors, on a $A = \bigcup_{n=0}^{\infty} B_n$. En effet, soit x élément de $\bigcup_{n=0}^{\infty} B_n$. Alors x est élément d'un des B_n, donc de A_n, donc est élément de A. Inversement, si x est élément de A, alors soit n le plus petit indice tel que x appartienne à A_n. x n'est donc pas dans A_{n-1}. x est alors dans B_n, donc dans leur réunion. De plus, les B_n sont deux à deux incompatibles puisque, si $k < n$, $B_k \subset A_k \subset A_{n-1}$ or $B_n \cap A_{n-1} = \varnothing$ donc $B_n \cap B_k = \varnothing$.

On montre de même que $A_n = \bigcup_{k=0}^{n} B_k$, union disjointe.

On a alors :

$$P(A) = P(\bigcup_{n=0}^{\infty} B_n) = \sum_{n=0}^{\infty} P(B_n) \qquad \text{car l'union est disjointe}$$

$$= \lim_{n\to\infty} \sum_{k=0}^{n} P(B_k) = \lim_{n\to\infty} P(\bigcup_{k=0}^{n} B_k) = \lim_{n\to\infty} P(A_n)$$

vi) Il n'est pas difficile de voir que si $(A_n)_{n\in \mathbf{N}}$ est une suite décroissante, alors $(\complement A_n)_{n\in \mathbf{N}}$ est une suite croissante. Posons $A = \bigcap\limits_{n=0}^{\infty} A_n$. On a :

$$
\begin{aligned}
P(A) &= 1 - P(\complement A) \\
&= 1 - P(\complement(\bigcap_{n=0}^{\infty} A_n)) \\
&= 1 - P(\bigcup_{n=0}^{\infty} \complement A_n) \\
&= 1 - \lim_{n\to\infty} P(\complement A_n) \qquad \text{d'après v)} \\
&= \lim_{n\to\infty} 1 - P(\complement A_n) \\
&= \lim_{n\to\infty} P(A_n)
\end{aligned}
$$

vii) Ici, la série $\sum P(A_n)$ peut diverger, auquel cas $\sum\limits_{n=0}^{\infty} P(A_n) = +\infty$. La suite $B_n = \bigcup\limits_{k=0}^{n} A_k$ est croissante, et $\bigcup\limits_{n=0}^{\infty} B_n = \bigcup\limits_{n=0}^{\infty} A_n$. Donc :

$$P(\bigcup_{n=0}^{\infty} A_n) = P(\bigcup_{n=0}^{\infty} B_n) = \lim_{n\to\infty} P(B_n) = \lim_{n\to\infty} P(\bigcup_{k=0}^{n} A_k)$$

Mais $P(\bigcup\limits_{k=0}^{n} A_k) \leq P(\bigcup\limits_{k=0}^{n-1} A_k) + P(A_n)$ en utilisant la propriété $P(A \cup B) \leq P(A) + P(B)$ vue en ii), et par récurrence, $P(B_n) = P(\bigcup\limits_{k=0}^{n} A_k) \leq \sum\limits_{k=0}^{n} P(A_k) \leq \sum\limits_{n=0}^{\infty} P(A_n)$. Donc, en passant à la limite :

$$P(\bigcup_{n=0}^{\infty} A_n) \leq \sum_{n=0}^{\infty} P(A_n)$$

EXEMPLE :

Soit une pièce truquée ayant la probabilité $p \in [0,1]$ de tomber sur Pile. On considère une expérience aléatoire consistant à lancer la pièce jusqu'à obtenir Pile, où à la lancer indéfiniment s'il n'y a que des Faces qui apparaissent. Soit A_n l'événement : Pile apparaît au-delà du n-ème lancer ou bien n'apparaît jamais. A_n est aussi l'événement : les n premiers lancers sont des Faces. On a :

$$P(A_n) = (1 - p)^n$$

$$A_{n+1} \subset A_n$$

L'événement $A = \bigcap\limits_{n=0}^{\infty} A_n$ est l'événement : Pile n'apparaît jamais. C'est aussi l'événement : on tire indéfiniment des Faces. Comme la suite des événements (A_n) décroît, on a :

$$P(A) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} (1 - p)^n = \begin{cases} 0 \text{ si } p > 0 \\ 1 \text{ si } p = 0 \end{cases}$$

Si $p = 0$, on ne sera guère surpris de voir que la probabilité de ne tirer que des Faces est 1. Dans le cas contraire, si $p > 0$, aussi petit soit-il, la probabilité que A se produise est nulle. Conceptuellement, A est un événement envisageable (il est différent de $\emptyset$), mais sa probabilité est nulle. On dit que A est presque sûrement impossible. Son événement contraire, consistant à obtenir un moment ou un autre un Pile, est dit presque sûrement certain.

5- Probabilités conditionnelles

On rappelle rapidement les notions vues en première années, en les généralisant éventuellement au cas dénombrable :

DEFINITION

Soit B *un événement de probabilité non nul, d'un espace probabilisé* $(\Omega, \mathcal{T}, P)$. *On définit la loi de probabilité sachant* B (*ou conditionnelle par rapport à* B) *par* :

$$P_B : \mathcal{T} \to [0,1]$$
$$A \to P_B(A) = \frac{P(A \cap B)}{P(B)}$$

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

On a donc $P(A \cap B) = P(B) \times P_B(A)$ (formule des probabilités composées). Dans le cas où $P(B) = 0$, on peut donner à l'égalité précédente la signification $0 = 0$ même si $P_B(A)$ n'est pas définie, car $P(A \cap B) \leq P(B)$ donc on a aussi $P(A \cap B) = 0$.

P_B est une loi de probabilité. En effet :

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

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

iii) Si les $(A_n)_{n\in\mathbb{N}}$ sont deux à deux incompatibles, alors les $A_n \cap B$ le sont également et l'on a :

$$P_B\left(\bigcup_{n=0}^{\infty} A_n\right) = \frac{P\left(\left(\bigcup_{n=0}^{\infty} A_n\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup_{n=0}^{\infty} (A_n \cap B)\right)}{P(B)} =$$

$$= \frac{\sum_{n=0}^{\infty} P(A_n \cap B)}{P(B)} = \sum_{n=0}^{\infty} P_B(A_n)$$

Soit I une famille finie ou dénombrable. $(A_i)_{i\in I}$ 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.

PROPOSITION (formule des probabilités totales)

Soit $(A_i)_{i\in I}$ *un système complet d'événements d'un espace probabilisé* $(\Omega, \mathcal{T}, P)$ *tels que, pour tout* i, $P(A_i) > 0$, *et* B *un autre événement de cet espace. Alors* :

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

En effet :

<u>Démonstration</u> :

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

$$\Omega = \bigcup_{i\in I} A_i \qquad \text{union disjointe finie ou dénombrable}$$

D'où $\quad P(B) = P(\Omega \cap B)$

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

$$= P(\bigcup_{i\in I} (A_i \cap B)) \qquad \text{cette union étant disjointe}$$

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

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

La formule reste valide pour une famille quelconque $(A_i)_{i\in I}$ d'événements deux à deux incompatibles (toujours avec I fini ou dénombrable) à condition que $\sum\limits_{i\in I} P(A_i) = 1$. Il suffit, dans la démonstration précédente de vérifier qu'on a toujours $P(\Omega \cap B) = P((\bigcup\limits_{i\in I} A_i) \cap B)$. Soit D l'événement contraire de $\bigcup\limits_{i\in I} A_i$. On a $\Omega = \bigcup\limits_{i\in I} A_i \cup D$ (union disjointe), donc $P(\Omega) = \sum\limits_{i\in I} P(A_i) + P(D)$ donc $P(D) = 0$ et a fortiori $P(D \cap B) = 0$, donc on a bien :

$$P(\Omega \cap B) = P((\bigcup_{i\in I} A_i) \cap B) + P(D \cap B) = P((\bigcup_{i\in I} A_i) \cap B)$$

On déduit de la formule des probabilités totales la formule suivante :

PROPOSITION (Formule de Bayes)

Soit $(A_i)_{i\in I}$ *un système complet d'événements d'un espace probabilisé* $(\Omega, \mathcal{T}, P)$ *tels que, pour tout i,* $P(A_i) > 0$, *et* B *un autre événement de cet espace tel que* $P(B) > 0$. *Alors* :

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

EXEMPLE :

Pour tout $n \geq 1$, on considère une urne U_n contenant une boule blanche et $n - 1$ boules noires. On choisit une urne au hasard, l'urne n étant choisie avec la probabilité $\frac{1}{2^n}$ (n est par exemple le nombre de lancers d'une pièce à effectuer pour obtenir Pile). L'événement A_n est réalisé si l'urne U_n est choisie. Puis on tire au hasard une boule dans l'urne choisie. L'événement B est réalisé si on tire une boule blanche. Ayant réalisé successivement ces deux choix, un expérimentateur annonce qu'il a effectivement tiré une boule blanche. On demande la probabilité qu'il ait choisi l'urne U_1.

On a ici $P_{A_n}(B) = \dfrac{1}{n}$ pour tout n, donc :

$$P(B) = \sum_{n=1}^{\infty} P(A_n) \times P_{A_n}(B) = \sum_{n=1}^{\infty} \frac{1}{n2^n} = -\ln(\frac{1}{2}) = \ln(2)$$

Donc $P_B(A_1) = \dfrac{P(A_1)\ P_{A_1}(B)}{P(B)} = \dfrac{1}{2} \times 1 \times \dfrac{1}{\ln(2)} = \dfrac{1}{2\ln(2)} \approx 0{,}72$

Voici ci-dessous une simulation en Python permettant de tester ce résultat :

```
# On lance une pièce jusqu'à ce qu'elle tombe sur Pile.
# Soit n le nombre de lancers.
# On tire alors une boule dans une urne contenant n boules dont
une seule
# blanche. On cherche la probabilité de tirer une boule
blanche,
# ainsi que, une boule blanche étant tirée, la probabilité
qu'elle
# provienne de la première urne.

import numpy as np

# Simulation d'un lancer de pièce. Pour tout entier n,
# np.random.randint(n) donne un nombre aléatoire entre 0 et n-1
compris
# selon une loi uniforme.
def piece():
    return(np.random.randint(2))

# On lance la pièce jusqu'à tomber sur Pile (= 1)
def lancePiece():
    n=1
    while piece()<>1:
        n+=1
    return(n)

# Simulation d'un tirage d'urne ayant n boules. On peut
convenir que la
# boule blanche est la boule n°0
def urne(n):
    return(np.random.randint(n))

# N = nombre de tirages
# A = nombre de fois où l'urne 0 est choisie
# B = nombre de boules blanches tirées
# AB = nombre de fois où l'urne 0 est choisie et où la boule
blanche est
#      choisie
def simule(N):
    B=0
    A=0
    AB=0
    for i in range(N):
        n=lancePiece()
        m=urne(n)
```

```
        if n==1:
            A+=1
            if m==0:
                B+=1
                AB+=1
        elif m==0:
            B+=1
    pB=float(B)/N   # conversion de l'entier B en flottant car le
                    # comportement de / diffère suivant les versions de
                    # Python
    pA=float(A)/N
    pAsachantB=float(AB)/B
    print("proba de tirer la blanche = "+str(pB))
        # proba théorique : ln(2)
    print("proba de tirer l'urne 0 = "+str(pA))
        # proba théorique : 1/2
    print("proba de tirer l'urne 0 sachant qu'on a tiré la blanche = "+str(pAsachantB))
        # proba théorique : 1/(2ln(2))
```

DEFINITION

Une famille finie ou dénombrable d'événements $(A_i)_{i\in I}$ *sont indépendants si, pour toute sous-famille de p événements* $A_{i_1}, A_{i_2}, \dots, A_{i_p}$ (*les indices* $i_1, \dots, i_p$ *étant distincts*), *on a* :

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

Un exemple typique de famille dénombrable d'événements indépendants est donné par le lancer d'une pièce indéfiniment, avec A_n l'événement "le n-ème tirage est Pile".

Si A et B sont indépendants, on a $P_B(A) = P(A)$.

II : Variables aléatoires

1- Définition

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé et X une fonction de Ω dans $\mathbb{R}$. On se limitera dans la suite au cas où X prend un nombre fini ou dénombrable de valeurs. Soit x élément de $X(\Omega)$. On note $\{X = x\}$ l'ensemble $\{\omega \in \Omega, X(\omega) = x\} = X^{-1}(\{x\})$. C'est une partie de Ω. On pourra en donner la probabilité si cette partie est un élément de la tribu $\mathcal{T}$. D'où la définition :

DEFINITION

On appelle variable aléatoire discrète définie sur un espace probabilisé $(\Omega, \mathcal{T}, P)$ *une fonction* X *de* Ω *dans* $\mathbb{R}$ *telle que* $X(\Omega)$ *est fini ou dénombrable, et telle que l'image réciproque par* X *de tout élement de* $X(\Omega)$ *est élement de la tribu* $\mathcal{T}$.

On note $P(X = x)$ ou $P_X(\{x\})$ la quantité $P(\{\omega \in \Omega, X(\omega) = x\}) = P(X^{-1}(\{x\}))$, et d'une manière générale, si A est une partie de $X(\Omega)$, on note $P(X \in A)$ ou $P_X(A)$ la quantité :

$$P(\{\omega \in \Omega, X(\omega) \in A\}) = P(X^{-1}(A))$$

P_X s'appelle la loi de X.

EXEMPLE :

Soit $\Omega = \mathbb{N}$ et P définie par $P(\{n\}) = \frac{1}{2^{n+1}}$. Soit X l'application qui à n associe son chiffre des unités. Quelle est la loi de X ? X prend ses valeurs dans $\{0, 1, 2, ..., 9\}$ et, pour k dans cet ensemble, on a :

$$P_X(\{k\}) = \sum_{n=0}^{\infty} P(\{10n+k\}) = \sum_{n=0}^{\infty} \frac{1}{2^{10n+k+1}} = \frac{1}{2^{k+1}} \sum_{n=0}^{\infty} \frac{1}{2^{10n}} = \frac{1}{2^{k+1}} \frac{1}{1-\frac{1}{2^{10}}}$$

On peut vérifier que $\sum_{k=0}^{9} P_X(\{k\}) = 1$.

PROPOSITION

P_X *est une loi de probabilité sur* $X(\Omega)$ *muni de la tribu de l'ensemble de ses parties*.

Démonstration :

On a $P_X(X(\Omega)) = P(\Omega) = 1$. Par ailleurs, soit $(A_n)_{n\in\mathbb{N}}$ une famille d'événements incompatibles sur $X(\Omega)$. Alors :

$$P_X(\bigcup_{n=0}^{\infty} A_n = P(X^{-1}(\bigcup_{n=0}^{\infty} A_n)$$

Or $X^{-1}(\bigcup_{n=0}^{\infty} A_n = \bigcup_{n=0}^{\infty} X^{-1}(A_n)$. En effet :

$$\omega \in X^{-1}(\bigcup_{n=0}^{\infty} A_n)$$

$\Leftrightarrow \quad X(\omega) \in \bigcup_{n=0}^{\infty} A_n$

$\Leftrightarrow \quad \exists\, n,\ X(\omega) \in A_n$

$\Leftrightarrow \quad \exists\, n,\ \omega \in X^{-1}(A_n)$

$\Leftrightarrow \quad \omega \in \bigcup_{n=0}^{\infty} X^{-1}(A_n)$

Donc :

$$P_X(\bigcup_{n=0}^{\infty} A_n) = P(\bigcup_{n=0}^{\infty} X^{-1}(A_n))$$

Par ailleurs, les A_n étant deux à deux incompatibles, il en est de même des $X^{-1}(A_n)$. En effet, pour n différent de m :

$$\omega \in X^{-1}(A_n) \cap X^{-1}(A_m) \Leftrightarrow X(\omega) \in A_n \cap A_m = \emptyset \quad \text{impossible}$$

Donc :

$$P_X(\bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty} P(X^{-1}(A_n)) = \sum_{n=0}^{\infty} P_X(A_n)$$

On a défini la loi de X à partir de la loi de probabilité sur Ω. Il existe une réciproque que nous admettrons. On peut se donner à priori la loi de X et en déduire une loi de probabilité P sur Ω de façon que X soit une variable aléatoire discrète dont la loi se déduit de P. Plus précisément, Si X prend ses valeurs dans $\{x_n,\ n \in I\}$, I étant fini ou dénombrable et les x_n étant distincs et si (p_n) est telle que $\sum_{n\in I} p_n = 1$, alors il existe une probabilité sur Ω telle que $P(X = x_n) = p_n$.

Comme nous le verrons, dans la plupart des cas, on a rarement besoin d'expliciter Ω et P. On raisonne directement sur P_X.

2- Loi de Poisson, convergence de la loi binomiale vers la loi de Poisson

Considérons la situation suivante, pendant un intervalle de temps [0, T] :

- Des personnes arrivent l'une après l'autre devant un guichet. La variable aléatoire X est le nombre de personnes qui arrivent dans l'intervalle de temps considéré.
- Un détecteur de particules élémentaires compte le nombre X de particules lors d'une mesure se déroulant pendant l'intervalle de temps considéré
- Un appareil est susceptible de subir de interventions ou des réparations, suite à des pannes ou des anomalies de fonctionnement. Pendant l'intervalle de temps considéré, on compte le nombre X d'interventions.

Dans chaque cas, on s'intéresse à la loi de X. Nous dirons qu'un phénomène est observé dans l'intervalle de temps $[t_1, t_2]$ si une personne arrive au guichet pendant cet intervalle de temps, ou si une particule est détectée, ou si une anomalie de fonctionnement de l'appareil survient. Nous supposerons que la probabilité qu'un phénomène soit observé pendant l'intervalle de temps $[t_1, t_2]$ ne dépend que de la longueur $t_2 - t_1$ de l'intervalle et est indépendant de ce qui se passe avant t_1 ou après t_2. Cette hypothèse suppose une régularité des phénomènes au cours du temps (pas d'heure de pointe, pas de mémoire de la durée de vie passée dans le cas de particule radioactive, pas d'usure de la machine augmentant le risque de panne au cours du temps). Nous appellerons λ le nombre moyen de phénomènes observés pendant l'intervalle de temps considéré.

Divisons [0, T] en n = 100 intervalles de longueur égales $[t_i, t_{i+1}]$. n est choisi assez grand de façon qu'il soit très peu probable que deux phénomènes soient observés dans le même intervalle $[t_i, t_{i+1}]$. Ainsi, dans chacun des n intervalles, ou bien on observe un phénomène, ou bien on n'observe aucun phénomène. Soit p_n la probabilité d'observation d'un phénomène et X_n le nombre d'événements observés. La loi de X_n est alors une loi binomiale $\mathcal{B}(n, p_n)$. L'hypothèse de la loi binomiale repose sur le fait que deux phénomènes ne peuvent se produire dans le même intervalle $[t_i, t_{i+1}]$. Cette hypothèse est d'autant mieux vérifiée que la longueur de l'intervalle est petite, et donc que n est grand. A la limite, quand n tend vers $+\infty$, p_n vers 0 de façon que np_n (nombre moyen d'événements observés dans le cas de X_n) tende vers λ (nombre moyen d'événements observés dans le cas de X). Nous allons en déduire une expression de $P(X = k)$. Pour tout n :

$$\begin{aligned}
P(X_n = k) &= \binom{n}{k} p_n^{\,k} (1 - p_n)^{n-k} \\
&= \frac{n(n-1)...(n-k+1)}{k!} p_n^{\,k} (1 - p_n)^{n-k} \\
&= \frac{1}{k!} \prod_{i=0}^{k-1} (np_n - ip_n) \times (1 - p_n)^{n-k} \\
&\sim \frac{1}{k!} \lambda^k \times \exp((n-k)\ln(1-p_n)) \qquad \text{car, pour tout } i,\ np_n - ip_n \to \lambda
\end{aligned}$$

Par ailleurs :

$$\exp((n-k)\ln(1-p_n)) = \exp((n-k)(-p_n + o(p_n))) = \exp(-\lambda + o(1)) \to \exp(-\lambda)$$

A la limite, on trouve donc :

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

k peut prendre maintenant des valeurs entières positives arbitraires. On reconnaît d'ailleurs dans $\dfrac{\lambda^k}{k!}$ le développement en série de e^λ. La somme des $P(X = k)$ pour k variant de 0 à $+\infty$ et donc bien 1.

La loi que nous venons de découvrir s'appelle loi de Poisson $\mathcal{P}(\lambda)$. Elle ne dépend que du paramètre λ. Nous avons également mis en évidence le théorème suivant de convergence de la loi binomiale vers la loi de Poisson.

PROPOSITION

Soit $\lambda > 0$ *et* (X_n) *une suite des variables aléatoires de loi binomiale* $\mathcal{B}(n, p_n)$, *où* (p_n) *est une suite telle* $\lim\limits_{n\to\infty} np_n = \lambda$. *Alors* :

$$\forall\, k \in \mathbb{N},\ \lim_{n\to\infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

En pratique, on peut approximer la loi binomiale par la loi de Poisson dès que $p < \dfrac{1}{10}$, $n > 30$, $\lambda = np < 16$.

EXEMPLE :

Une usine fabrique 1000 pièces dont 3 % sont défectueuses. On en tire 100. Quelle est la probabilité d'en avoir 5 défectueuses ?

❑ Soit X le nombre de pièces défectueuses. Si le tirage a lieu sans remise, X suit de façon exacte une loi dite hypergéométrique. La probabilité cherchée est :

$$P(X = 5) = \frac{\binom{30}{5}\binom{970}{95}}{\binom{1000}{100}} \approx 0{,}1024$$

❑ Si les tirages se font avec remise, on a une loi binomiale $\mathcal{B}(100, \frac{3}{100})$, d'où une évaluation de $P(X = 5)$:

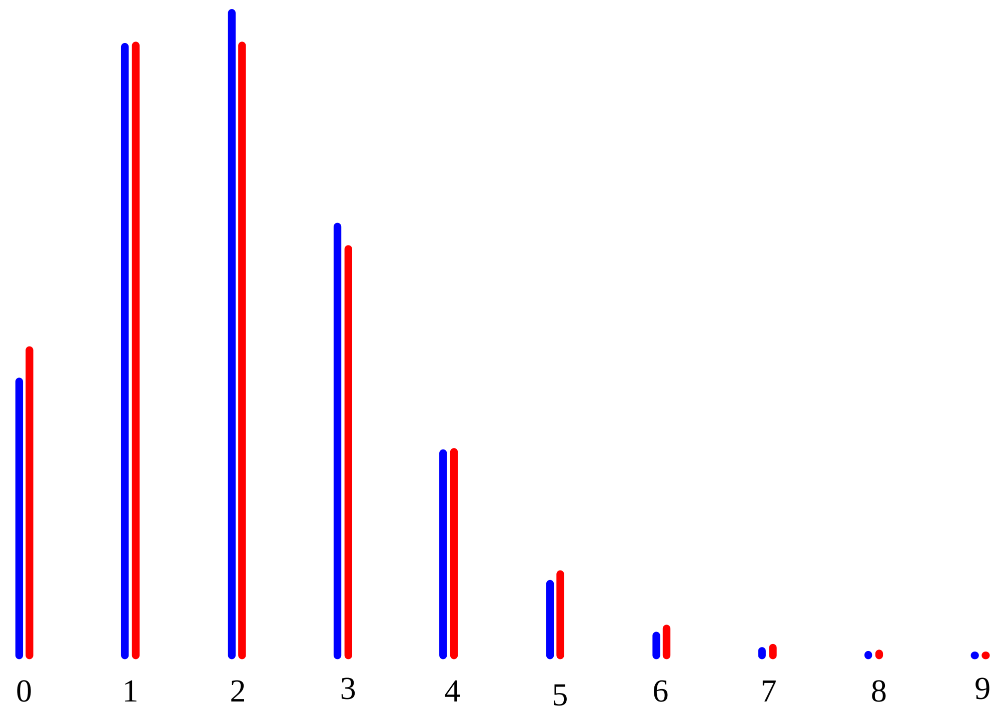
$$P(X = 5) = \binom{100}{5} 0{,}03^5 \times 0{,}97^{95} \approx 0{,}1013$$

❑ 100 étant lui-même élevé, $p = 0.03$ étant faible, et $\lambda = np = 3$ étant raisonnable, on peut approximer la loi binomiale par la loi de Poisson $\mathcal{P}(3)$. D'où une nouvelle évaluation :

$$P(X = 5) = \frac{3^5}{5!} e^{-3} \approx 0{,}1008$$

C'est évidemment cette dernière quantité qui est le plus facile à calculer. Avec deux chiffres significatifs, le résultat obtenu est correct. Les deux autres décimales n'ont été données que pour permettre d'apprécier la différence de précision entre les diverses lois.

Ci-dessous, on donne une représentation graphique en histogramme, avec $n = 20$ et $p = 0{,}1$, pour $0 \le k \le 9$. Les bâtons bleus représentent la répartition de la loi binomiale et les bâtons rouges ceux de la loi de Poisson.



3- Loi géométrique

Cette loi intervient dans les situations suivantes :

❑ On lance une pièce jusqu'à ce qu'on tombe sur pile. On cherche la loi du nombre de lancers.

❑ On joue au loto jusqu'à ce qu'on gagne le gros lot. On cherche la loi du nombre de parties.

❑ Plus généralement, on répète la même expérience de Bernoulli (à deux issues, échec ou succès), jusqu'à obtenir un succès. On cherche la loi du nombre X d'expériences. X peut éventuellement prendre la valeur +∞ si le succès ne se produit jamais.

Soit p la probabilité d'un succès et $1 - p$ celle d'un échec. Alors :

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

k varie de 1 à +∞. La somme des $P(X = k)$ pour k élément de $\mathbb{N}^*$ vaut 1 si p appartient à]0,1[. Cela signifie aussi que $P(X = \infty) = 0$, et donc qu'il est certain que le succès se produira au bout d'un temps fini (c'est encourageant pour les joueurs de loto) mais peut-être long (quelques dizaines de milliers d'années en moyenne pour le gros lot du joueur de loto !).

Deux cas dégénérés peuvent être considérés :

$p = 1$. Alors X est la variable certaine égale à 1 (succès à la première expérience).

$p = 0$. Alors X est la variable certaine +∞. $P(X = +\infty) = 1$ (échec certain indéfiniment).

La loi géométrique caractérise les phénomènes sans mémoire et sans usure. Calculons d'abord $P(X > k)$.

$$P(X > k) = \sum_{n=k+1}^{\infty} (1 - p)^{n-1} p = (1 - p)^k$$

Cette valeur se retrouve d'ailleurs directement en considérant que $X > k$ signifie k échecs de suite. On en tire ensuite les probabilités conditionnelles suivantes :

$$\begin{aligned}
P(X > n + k \mid X > k) &= \frac{P(X > n + k \cap X > k)}{P(X > k)} \\
&= \frac{P(X > n + k)}{P(X > k)} \qquad \text{car } X > n + k \cap X > k \Leftrightarrow X > n + k \\
&= \frac{(1-p)^{n+k}}{(1-p)^k} = (1-p)^n \\
&= P(X > n)
\end{aligned}$$

Ainsi, sachant que X est supérieur à k, la loi de X – k est la même que la loi de X. Cela signifie donc que l'on peut oublier les k premières expériences et recommencer à zéro ; la loi du nombre d'expériences jusqu'au succès est inchangée. Cette règle est rarement connue, et contraire à l'intuition courante. La plupart des gens pense que le fait de lancer un dé cinq fois sans que le n° 6 sorte rend sa probabilité de sortie au coup suivant plus grande. Ou encore, le fait de jouer au loto sans gagner depuis des semaines encourage certaines personnes à persévérer en pensant que cela va bientôt être leur tour de gagner. Il n'en est rien. Elles ont autant de chance de gagner ou de perdre qu'avant, et cela indépendamment de leurs gains ou pertes passés.

Inversement, soit X une variable aléatoire sur $\mathbb{N}^*$ vérifiant :

$$\forall k > 0, \forall n > 0, P(X > n + k \mid X > k) = P(X > n).$$

Montrons que X suit une loi géométrique. On a :

$$P(X > n) = \frac{P(X > n + k \cap X > k)}{P(X > k)} = \frac{P(X > n + k)}{P(X > k)}$$

Donc $P(X > n + k) = P(X > n)\, P(X > k)$

En particulier :

$$P(X > n + 1) = P(X > n)\, P(X > 1)$$

Posons $p = P(X = 1)$. D'où $P(X > 1) = 1 - P(X = 1) = 1 - p$, et $P(X > n)$ est une suite géométrique de raison $1 - p$. Comme $P(X > 0) = 1$, on a donc :

$$P(X > n) = (1-p)^n$$

et $P(X = n) = P(X > n - 1) - P(X > n) = (1-p)^{n-1} - (1-p)^n = (1-p)^{n-1}\, p$

Ce qui prouve que X est bien de loi géométrique.

EXEMPLE :

Un exemple de phénomène sans mémoire se rencontre dans la désintégration des atomes radioactifs. Il est expérimentalement constaté que, si on dispose à l'instant initial d'un nombre N grand de radionucléides d'un certain type, il existe une constante λ telle que, au bout d'un temps t quelconque, il ne reste plus que $Ne^{-\lambda t}$ atomes non désintégrés. λ est la proportion d'atomes qui se désintègre par unité de temps. Ce comportement peut s'interpréter par la modélisation suivante. Considérons un intervalle de temps τ très court. La proportion d'atomes se désintégrant pendant la durée τ est $\lambda\tau$. Supposons donc qu'à chacun de ces intervalles de temps τ successifs, le radionucléide tire à pile ou face avec une probabilité $p = \lambda\tau$ de se désintégrer. Soit X le nombre d'intervalles de temps τ au bout duquel il se désintègre. Si les tirages sont indépendants, i.e. si le radionucléide est sans mémoire, X suit une loi géométrique de paramètre p. On a donc $P(X > n) = (1 - p)^n$. Si $t = n\tau$, le nombre d'atomes qui ne se sont pas désintégrés suit une loi binomiale $\mathcal{B}(N, (1 - p)^n)$, en supposant que la désintégration d'un atome est indépendante de celle d'un autre, i.e. il n'y a pas de fission induite d'un atome à l'autre. Dans ce cas, l'espérance du nombre d'atomes non désintégrés au bout du temps t est :

$$N(1-p)^n = N \exp(n \ln(1-p))$$

Si on fait tendre n vers $+\infty$, de sorte que $\tau = \dfrac{t}{n}$ tend vers 0, ainsi que $p = \lambda\tau = \dfrac{\lambda t}{n}$, alors on a :

$$N(1-p)^n = N\exp(-np + no(p)) = N\exp(-np + no(\frac{1}{n})) = N\exp(-np + o(1)) \to Ne^{-\lambda t}$$

On retrouve la loi de décroissance initiale. Par ailleurs, la durée de vie moyenne d'un radionucléide est $\mathbf{E}(X)\tau = \frac{\tau}{p} = \frac{1}{\lambda}$.

On rencontre, suivant les conventions adoptées, des variantes de la loi géométrique.

- On la définit parfois sur $\mathbb{N}$ par : $P(X = k) = (1-p)^k\, p$
- Les rôles de p et $1-p$ sont parfois inversés.

Il convient donc de réfléchir à la situation proposée avant d'appliquer mécaniquement les résultats de ce paragraphe.

4- Fonction de répartition

Soit X une variable aléatoire. On appelle fonction de répartition de X la fonction $F_X : \mathbb{R} \to [0,1]$ définie par : $F_X(t) = P(X \leq t)$.

PROPRIETES

i) F_X *est croissante*

ii) $\lim\limits_{t\to-\infty} F_X(t) = 0$

iii) $\lim\limits_{t\to+\infty} F_X(t) = 1$

<u>Démonstration</u> :

i) Si $t \leq u$, $\{\omega \in \Omega, X(\omega) \leq t\} \subset \{\omega \in \Omega, X(\omega) \leq u\}$ donc :

$$F_X(t) = P(X \leq t) \leq P(X \leq u) = F_X(u)$$

ii) Il suffit de montrer que, pour tout suite (t_n) décroissante vers $-\infty$, on a $\lim\limits_{n\to\infty} F_X(t_n) = 0$ (caractérisation séquentielle d'une limite). Pour cela, remarquons que les événements $\{X \leq t_n\}$ sont décroissants pour l'inclusion et d'intersection vide. Donc :

$$0 = P(\varnothing) = P(\bigcap_{n=0}^{\infty} \{X \leq t_n\}) = \lim_{n\to\infty} P(X \leq t_n) = F_X(t_n)$$

iii) Il suffit de montrer que, pour tout suite (t_n) croissante vers $+\infty$, on a $\lim\limits_{n\to\infty} F_X(t_n) = 1$ Pour cela, remarquons que les événements $\{X \leq t_n\}$ sont croissants pour l'inclusion et de réunion Ω. Donc :

$$1 = P(\Omega) = P(\bigcup_{n=0}^{\infty} \{X \leq t_n\}) = \lim_{n\to\infty} P(X \leq t_n) = F_X(t_n)$$

5- Couple de variables aléatoires

La notion de couple de variables aléatoires est identique à celle vue dans le cas fini, sauf que les sommes qui interviennent peuvent être des séries, voire des séries doubles.

X et Y étant des variables aléatoires discrètes, on définit la loi conjointe du couple par la donnée des $P(X = x, Y = y)$, $x \in X(\Omega)$, $y \in Y(\Omega)$. Les lois de X et de Y sont appelées lois marginales. On peut déterminer les lois marginales à partir de la loi conjointe en utilisant le fait que l'événement $\{X = x\}$ est la réunion disjointe et finie ou dénombrable des événements $\{X = x, Y = y\}$, $y \in Y$:

$$P(X = x) = \sum_{y \in Y(\Omega)} P(X = x, Y = y)$$

de même :

$$P(Y = y) = \sum_{x \in X(\Omega)} P(X = x, Y = y)$$

Mais la réciproque est fausse car $\{X = x, Y = y\} = \{X = x\} \cap \{Y = y\}$ et d'une façon générale, on ne peut retrouver $P(A \cap B)$ à partir de $P(A)$ et $P(B)$, sauf si les événements sont indépendants.

Si pour tout x et tout y, les événements $\{X = x\}$ et $\{Y = y\}$ sont indépendants, on dit que les variables aléatoires discrètes X et Y sont indépendantes. On a dans ce cas :

$$P(X = x, Y = y) = P(X = x)\, P(Y = y)$$

et plus généralement, pour toute partie A incluse dans $X(\Omega)$ et toute partie B incluse dans $Y(\Omega)$, on a :

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

En effet, pour tout x, on a, en tenant compte du fait que l'événement $\{X = x, Y \in B\}$ est la réunion disjointe des événements $\{X = x, Y = y\}$ quand y parcourt B :

$$\begin{aligned} P(X = x, Y \in B) &= \sum_{y \in B} P(X = x, Y = y) \\ &= \sum_{y \in B} P(X = x)\, P(Y = y) \\ &= P(X = x) \sum_{y \in B} P(Y = y) \\ &= P(X = x)\, P(Y \in B) \end{aligned}$$

puis, en tenant compte du fait que l'événement $\{X \in A, Y \in B\}$ est la réunion disjointe des événements $\{X = x, Y \in B\}$ quand x parcourt A :

$$\begin{aligned} P(X \in A, Y \in B) &= \sum_{x \in A} P(X = x, Y \in B) \\ &= \sum_{x \in A} P(X = x)\, P(Y \in B) \\ &= \Big(\sum_{x \in A} P(X = x)\Big)\, P(Y \in B) \\ &= P(X \in A)\, P(Y \in B) \end{aligned}$$

Si X et Y sont indépendantes, alors pour toute fonction f et g de $\mathbb{R}$ dans $\mathbb{R}$, $f(X)$ et $g(Y)$ sont indépendantes. En effet, Pour toute valeur u et v que peuvent prendre $f(X)$ et $g(Y)$, on a :

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

$= \mathrm{P}(f(\mathrm{X}) = u)\ \mathrm{P}(g(\mathrm{Y}) = v)$

EXEMPLE :
Pour (n, m) éléments de $\mathbb{N}^2$, soit $\mathrm{P}(\mathrm{X} = n, \mathrm{Y} = m) = \dfrac{\mathrm{C}}{2^{n+m+1}}$. Déterminons C pour que l'on ait une loi de probabilité et voyons si X et Y sont indépendantes.

La loi de X est $\mathrm{P}(\mathrm{X} = n) = \sum\limits_{m=0}^{\infty} \dfrac{\mathrm{C}}{2^{n+m+1}} = \dfrac{\mathrm{C}}{2^n}$ et l'on a $\sum\limits_{n=0}^{\infty} \mathrm{P}(\mathrm{X} = n) = 1$ si et seulement si $\mathrm{C} = \dfrac{1}{2}$.

De même, la loi de Y est $\mathrm{P}(\mathrm{Y} = m) = \sum\limits_{n=0}^{\infty} \dfrac{\mathrm{C}}{2^{n+m+1}} = \dfrac{\mathrm{C}}{2^m}$.

On a bien $\mathrm{P}(\mathrm{X} = n, \mathrm{Y} = m) = \dfrac{1}{2^{n+m+2}} = \dfrac{1}{2^{n+1}}\dfrac{1}{2^{m+1}} = \mathrm{P}(\mathrm{X} = n)\ \mathrm{P}(\mathrm{Y} = m)$

Si les variables aléatoires ne sont pas indépendantes, on aura souvent recours à des probabilités conditionnelles. En vertu de la relation $\mathrm{P}(\mathrm{X} = x, \mathrm{Y} = y) = \mathrm{P}(\mathrm{Y} = y \mid \mathrm{X} = x)\ \mathrm{P}(\mathrm{X} = x)$, on peut reconstituer la loi du couple si on connaît la loi de X et toutes les valeurs de $\mathrm{P}(\mathrm{Y} = y \mid \mathrm{X} = x)$, i.e. la loi conditionnelle de Y relativement à X.

EXEMPLE :
Soit X de loi de Poisson de paramètre λ et Y de loi conditionnelle sachant que $\mathrm{X} = n$ de loi binomiale $\mathcal{B}(n, p)$ et $\mathrm{Z} = \mathrm{X} - \mathrm{Y}$. Quelles sont les lois de Y et Z ? Voici quelques cas où l'on rencontre cette situation :
❑ X est le nombre de clients d'un grand magasin durant une journée donnée. λ est le nombre moyen de clients dans le magasin. Y est le nombre de clients se faisant voler leur portefeuille, p la probabilité de se faire voler son portefeuille. (Exemple original tiré des annales de BCPST).
❑ X est le nombre de voitures arrivant à une station service durant une journée donnée, λ est le nombre moyen de voitures, Y est le nombre de clients qui achètent du diesel, Z ceux qui achètent de l'essence, p la proportion de voitures qui achète du diesel.
❑ Z est le nombre de participants à un congrès, λ le nombre moyen de participants, Y le nombre de personnes qui s'inscrivent au repas, Z le nombre de ceux qui mangent par leur propre moyen, p la proportion de personnes qui prend un repas.

Loi conjointe de (X, Y):

$$\begin{aligned}\mathrm{P}(\mathrm{X} = n, \mathrm{Y} = k) &= \mathrm{P}(\mathrm{X} = n) \times \mathrm{P}(\mathrm{Y} = k \mid \mathrm{X} = n) \\ &= \frac{\lambda^n}{n!} e^{-\lambda} \binom{n}{k} p^k (1-p)^{n-k}\end{aligned}$$

Cette quantité est nulle si $k > n$, et sinon se simplifie en $\dfrac{\lambda^n}{k!(n-k)!} e^{-\lambda} p^k (1-p)^{n-k}$. Donc :

$$\mathrm{P}(\mathrm{Y} = k) = \sum_{n=k}^{\infty} \mathrm{P}(\mathrm{X} = n, \mathrm{Y} = k) = \sum_{n=k}^{\infty} \frac{\lambda^n}{k!(n-k)!} e^{-\lambda} p^k (1-p)^{n-k}$$

$$= e^{-\lambda} \frac{p^k}{k!} \sum_{n=k}^{\infty} \frac{\lambda^n}{(n-k)!} (1-p)^{n-k}$$

$$= e^{-\lambda} \frac{(\lambda p)^k}{k!} \sum_{n=k}^{\infty} \frac{\lambda^{n-k}}{(n-k)!} (1-p)^{n-k}$$

On reconnaît dans la somme le développement en série de $\exp(\lambda(1-p))$. Donc :

$$P(Y = k) = e^{-\lambda} \frac{(\lambda p)^k}{k!} e^{\lambda(1-p)} = \frac{(\lambda p)^k}{k!} e^{-\lambda p}$$

qui n'est autre qu'une loi de Poisson de paramètre λp. En particulier, $\mathbf{E}(Y) = p\lambda$, ce qui, finalement est bien naturel, puisque $\mathbf{E}(X) = \lambda$, et que dans la population X, une proportion p en moyenne forme la population X.

Loi conjointe de (Y, Z) :

$$P(Y = k, Z = n) = P(X = n + k, Y = k) = \frac{\lambda^{n+k}}{(n+k)!} e^{-\lambda} \binom{n+k}{k} p^k (1-p)^n$$
$$= \frac{\lambda^{n+k}}{n!\,k!} e^{-\lambda} p^k (1-p)^n = \frac{(\lambda p)^k}{k!} e^{-\lambda p} \frac{(\lambda(1-p))^n}{n!} e^{-\lambda(1-p)}$$

En sommant sur k, on obtient $P(Z = n) = \frac{(\lambda(1-p))^n}{n!} e^{-\lambda(1-p)}$, loi de Poisson de paramètre $\lambda(1-p)$. De plus, on constate que Y et Z sont indépendants.

Si l'on dispose de n variables aléatoires $X_1, ..., X_n$, on dira qu'elles sont mutuellement indépendantes si et seulement si pour tout $(x_1, ..., x_n)$, les événements $X = x_1, ..., X = x_n$ sont indépendants. Il ne suffit pas que les variables aléatoires soient deux à deux indépendantes pour l'être mutuellement, pas plus qu'il ne suffit pas que des événements soient deux à deux indépendants pour l'être mutuellement.

EXEMPLE :
On lance n dés. Pour $1 \le i \le n$, X_i est le tirage du i-ème dè. Notons Y_n la parité de la somme des n dès ($Y_n = X_1 + ... + X_n \bmod 2 = 0$ si la somme est paire, 1 si la somme est impaire). $(X_1, ..., X_n)$ sont indépendants, mais $(X_1, ..., X_n, Y_n)$ ne le sont pas puisque la valeur de Y_n est imposée par celle de $X_1, ..., X_n$. Cependant, pour tout $i \le n$, (X_i, Y_n) sont indépendants. Montrons-le pour $i = n$. On a :

$$P(Y_n = 0, X_n = k) = P(Y_{n-1} = k \bmod 2, X_n = k)$$

où $Y_{n-1} = X_1 + ... + X_{n-1} \bmod 2$, indépendant de X_n

$$= P(Y_{n-1} = k \bmod 2)\, P(X_n = k)$$

Mais, pour tout k, $P(Y_{n-1} = k \bmod 2) = \frac{1}{2} = P(Y_n = 0)$ (car pour chaque dé, il y a autant de chance de tirer un nombre pair qu'un nombre impair, et il en sera de même pour la somme). Donc :

$$P(Y_n = 0, X_n = k) = P(Y_n = 0)\, P(X_n = k)$$

PROPOSITION

Soit X *de loi de Poisson* $\mathcal{P}(\lambda)$ *et* Y $\mathcal{P}(\mu)$ *deux variables aléatoires indépendantes. Alors* X + Y *suit une loi de Poisson* $\mathcal{P}(\lambda + \mu)$.

Ce résultat s'interprète aussi concrètement en réunissant deux files d'attente en une seule, la première ayant en moyenne λ clients par unité de temps et la seconde μ clients. La réunion aura en moyenne $\lambda + \mu$ clients.

<u>Démonstration</u> :

$$P(X + Y = n) = \sum_{k=0}^{n} P(X = k)\, P(Y = n - k)$$

$$= \sum_{k=0}^{n} \frac{\lambda^k}{k!} e^{-\lambda} \frac{\mu^{n-k}}{(n-k)!} e^{-\mu}$$

$$= \frac{1}{n!} e^{-(\lambda+\mu)} \sum_{k=0}^{n} \binom{n}{k} \lambda^k \mu^{n-k}$$

On reconnaît le développement du binôme de Newton de $(\lambda + \mu)^n$. Donc :

$$P(X + Y) = \frac{(\lambda + \mu)^n}{n!} e^{-(\lambda+\mu)}$$

ce qui est bien la loi cherchée.

On prouve de même, par récurrence, que la somme de n variables aléatoires indépendantes suivant des loi de Poisson de paramètre $\lambda_1, ..., \lambda_n$, suit une loi de Poisson de paramètre $\lambda_1 + ... + \lambda_n$

6- Espérance d'une variable aléatoire

L'espérance d'une variable aléatoire se calcule comme dans le cas fini, mais le nombre de valeurs prises par la variable peut être infini dénombrable. Si X est à valeurs dans $\{x_n, n \in \mathbb{N}\}$, on dit que X est d'espérance finie si la série x_i P(X = x_i) est **absolument** convergente et on pose :

DEFINITION DE L'ESPERANCE

$$\mathbf{E}(X) = \sum_{n=0}^{\infty} x_n\, P(X = x_n)$$

Pour le calcul de $\mathbf{E}(f(X))$, on dispose du théorème du transfert suivant (admis) :
Si X *est une variable aléatoire discrète et f une fonction définie sur* $X(\Omega) = \{x_n, n \in \mathbb{N}\}$, *alors* $f(X)$ *est d'espérance finie si et seulement si* $\sum P(X = x_n) f(x_n)$ *est absolument convergente et on a*

$\mathbf{E}(f(X)) = \sum_{n=0}^{\infty} P(X = x_n) f(x_n)$.

Une démonstration dans le cas où Ω est fini est donnée dans le cours de probabilité de première année. La régle précédente s'applique également aux couples de variables aléatoires discrètes. On exige des séries d'être absolument convergente de façon à ce que l'ordre dans lequel on somme les termes n'intervienne pas. Il suffit pour cela que X (respectivement $f \circ X$) soit bornée. Dans le cas contraire, il est tout à fait possible que la série diverge ou ait une valeur qui dépende de l'ordre des termes, auquel cas l'espérance n'est pas définie.

On a également la propriété suivante :
Si Y *est d'espérance finie et si* $|X| \leq Y$, *alors* X *est d'espérance finie.*

Notons $\{x_n, n \in \mathbb{N}\}$ les valeurs prises par X et $\{y_i, i \in I\}$ celles prises par Y, I étant un ensemble d'indices fini ou dénombrable. Pour tout entier n, on a :

$$|x_n|\, P(X = x_n) = \sum_{i \in I} |x_n|\, P(X = x_n, Y = y_i)$$

$$\leq \sum_{i\in I} y_i \, P(X = x_n, Y = y_i)$$

En effet, pour tout i, ou bien $P(X = x_n, Y = y_i) = 0$ et le terme correspondant n'a aucune contribution dans la somme, ou bien $P(X = x_n, Y = y_i) > 0$, mais comme $|X| > Y$ est impossible et est de probabilité nulle, on a nécessairement $|x_n| \leq y_i$.
On prend ensuite une somme partielle. Pour tout N :

$$\sum_{n=0}^{N} |x_n| \, P(X = x_n) \leq \sum_{n=0}^{N} \sum_{i\in I} y_i \, P(X = x_n, Y = y_i) = \sum_{i\in I} \sum_{n=0}^{N} y_i \, P(X = x_n, Y = y_i)$$

$$= \sum_{i\in I} y_i \, P(X \in \{x_0, ..., x_N\}, Y = y_i\}$$

$$\leq \sum_{i\in I} y_i \, P(Y = y_i\} = E(Y)$$

Les sommes partielles $\sum_{n=0}^{N} |x_n| \, P(X = x_n)$ formant une suite croissante majorée, elles convergent et X a une espérance finie.

PROPRIETES
Soient X *et* Y *deux variables aléatoires ayant une espérance finie, on a* :

i) $\mathbf{E}(\lambda X) = \lambda \mathbf{E}(X)$
ii) $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$
iii) $X \geq 0 \Rightarrow \mathbf{E}(X) \geq 0$
iv) $X \leq Y \Rightarrow \mathbf{E}(X) \leq \mathbf{E}(Y)$
v) *Si* X *et* Y *sont* ***indépendantes****, alors* $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$
vi) *Si* X *est à valeurs dans* $\mathbb{N}$, $\mathbf{E}(X)$ *est défini si et seulement si* $\sum P(X \geq k)$ *converge, et dans ce cas,* $\mathbf{E}(X) = \sum_{k=1}^{\infty} P(X \geq k)$

Démonstration :
On glisse sur d'éventuelles difficultés portant sur les séries doubles, qui ne sont pas l'objet du programme. On admet que le maniement de ces séries doubles est valide dans le cas de séries absolument convergentes.

i) En prenant $f(X) = \lambda X$, on a $\mathbf{E}(\lambda X) = \sum_{n=0}^{\infty} \lambda x_n P(X = x_n) = \lambda \sum_{n=0}^{\infty} x_n P(X = x_n) = \lambda \mathbf{E}(X)$

ii) En prenant $f(X, Y) = X + Y$, on a :

$$\mathbf{E}(X + Y) = \sum_{n,m} (x_n + y_m) \, P(X = x_n, Y = y_m)$$

$$= \sum_{n,m} x_n \, P(X = x_n, Y = y_m) + \sum_{n,m} y_m \, P(X = x_n, Y = y_m)$$

$$= \sum_{n=0}^{\infty} x_n \sum_{m=0}^{\infty} \mathrm{P}(\mathrm{X} = x_n, \mathrm{Y} = y_m) + \sum_{m=0}^{\infty} y_m \sum_{n=0}^{\infty} \mathrm{P}(\mathrm{X} = x_n, \mathrm{Y} = y_m)$$

$$= \sum_{n=0}^{\infty} x_n \, \mathrm{P}(\mathrm{X} = x_n) + \sum_{m=0}^{\infty} y_m \, \mathrm{P}(\mathrm{Y} = y_m)$$

$$= \mathbf{E}(\mathrm{X}) + \mathbf{E}(\mathrm{Y})$$

iii) évident

iv) Si $\mathrm{X} \leq \mathrm{Y}$, alors $0 \leq \mathbf{E}(\mathrm{Y} - \mathrm{X}) = \mathbf{E}(\mathrm{Y}) - \mathbf{E}(\mathrm{X})$

v) En prenant $f(\mathrm{X}, \mathrm{Y}) = \mathrm{XY}$, on a (en admettant comme valide le calcul suivant effectué sur des séries doubles) :

$$\mathbf{E}(\mathrm{XY}) = \sum_{n,m} x_n y_m \, \mathrm{P}(\mathrm{X} = x_n, \mathrm{Y} = y_m)$$

$$= \sum_{n=0}^{\infty} x_n \, \mathrm{P}(\mathrm{X} = x_n) \times \sum_{m=0}^{\infty} y_m \, \mathrm{P}(\mathrm{Y} = y_m)$$

$$= \mathbf{E}(\mathrm{X})\mathbf{E}(\mathrm{Y})$$

vi) Considérons une somme partielle de $\mathbf{E}(\mathrm{X}) = \sum_{n=0}^{\infty} n \, \mathrm{P}(\mathrm{X} = n)$ et transformons-la :

$$\sum_{k=0}^{n} k \, \mathrm{P}(\mathrm{X} = k) = \sum_{k=0}^{n} k \, (\mathrm{P}(\mathrm{X} \geq k) - \mathrm{P}(\mathrm{X} \geq k + 1))$$

$$= \sum_{k=0}^{n} k \, \mathrm{P}(\mathrm{X} \geq k) - \sum_{k=0}^{n} k \, \mathrm{P}(\mathrm{X} \geq k + 1)$$

$$= \sum_{k=1}^{n} k \, \mathrm{P}(\mathrm{X} \geq k) - \sum_{k=1}^{n+1} (k - 1) \, \mathrm{P}(\mathrm{X} \geq k) \qquad \text{en changeant d'indice}$$

$$= \sum_{k=1}^{n} (k - (k - 1)) \, \mathrm{P}(\mathrm{X} \geq k) - n \, \mathrm{P}(\mathrm{X} \geq n + 1)$$

$$= \sum_{k=1}^{n} \mathrm{P}(\mathrm{X} \geq k) - n \, \mathrm{P}(\mathrm{X} \geq n + 1)$$

Supposons que $\mathbf{E}(\mathrm{X})$ soit défini, i.e. que $\sum n \, \mathrm{P}(\mathrm{X} = n)$ converge, alors :

$$n \, \mathrm{P}(\mathrm{X} \geq n + 1) = n \sum_{k=n+1}^{\infty} \mathrm{P}(\mathrm{X} = k) = \sum_{k=n+1}^{\infty} n \, \mathrm{P}(\mathrm{X} = k)$$

$$\leq \sum_{k=n+1}^{\infty} k \, \mathrm{P}(\mathrm{X} = k) \qquad \text{car } n \geq k$$

$$\leq \mathbf{E}(X) - \sum_{k=0}^{n} k\,P(X = k) \qquad \text{quantité qui tend vers 0 quand } n \text{ tend vers } +\infty$$

puisqu'on reconnaît le reste de la série définissant **E**(X)

Donc $\lim\limits_{n\to\infty} n\,P(X \geq n + 1) = 0$

Par conséquent, en passant à la limite dans l'égalité $\sum\limits_{k=1}^{n} P(X \geq k) = \sum\limits_{k=0}^{n} k\,P(X = k) + n\,P(X \geq n + 1)$, on obtient :

$$\sum_{k=1}^{\infty} P(X \geq k) = \mathbf{E}(X)$$

Réciproquement, si $\sum P(X \geq k)$ converge, alors **E**(X) existe car la suite des sommes partielles $\sum\limits_{k=0}^{n} k\,P(X = k)$ est croissante avec n et majorée par $\sum\limits_{k=1}^{\infty} P(X \geq k)$. En effet :

$$\sum_{k=0}^{n} k\,P(X = k) = \sum_{k=1}^{n} P(X \geq k) - n\,P(X \geq n + 1)$$

$$\leq \sum_{k=1}^{n} P(X \geq k)$$

$$\leq \sum_{k=1}^{\infty} P(X \geq k)$$

On est donc ramené au cas précédent, donnant l'égalité entre **E**(X) et $\sum\limits_{k=1}^{\infty} P(X \geq k)$

Une démonstration plus simple, mais utilisant les séries doubles (qui sont hors programme en PC/PSI) peut être menée comme suit :

$$\mathbf{E}(X) = \sum_{n=1}^{\infty} n\,P(X = n)$$

$$= \sum_{n=1}^{\infty} \sum_{k=1}^{n} P(X = n) = \sum_{1 \leq k \leq n} P(X = n)$$

$$= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} P(X = n)$$

$$= \sum_{k=1}^{\infty} P(X \geq k)$$

Enfin, remarquons que la relation $\sum_{k=1}^{n} P(X \geq k) = \sum_{k=0}^{n} k\ P(X = k) + n\ P(X \geq n+1)$ peut se visualiser sous la disposition suivante : pour tout n, disposons la somme partielle $\sum_{k=1}^{n} P(X \geq k)$ en colonne :

$$\begin{array}{lllll}
\ \ P(X \geq 1) & & & & \\
+\ P(X \geq 2) & & & & \\
+\ P(X \geq 3) & & & & \\
+\ \ldots & & & & \\
+\ P(X \geq n) & & & & \\
=\ P(X = 1) + & P(X = 2) + & P(X = 3) + \ldots + & P(X = n) + P(X \geq n+1) \\
+ & P(X = 2) + & P(X = 3) + \ldots + & P(X = n) + P(X \geq n+1) \\
+ & & P(X = 3) + \ldots + & P(X = n) + P(X \geq n+1) \\
+\ \ldots & & & \\
+ & & & P(X = n) + P(X \geq n+1)
\end{array}$$

On voit donc que :

$$\sum_{k=1}^{n} P(X \geq k) = P(X = 1) + 2P(X = 2) + 3P(X = 3) + \ldots + nP(X = n) + nP(X \geq n+1)$$

$$= \sum_{k=1}^{n} kP(X = k) + nP(X \geq n+1)$$

On peut aussi prouver cette relation par récurrence.

Les espérances des lois usuelles sont les suivantes :

❑ *Loi de Poisson* :

$$\mathbf{E}(X) = \sum_{k=0}^{\infty} k\ P(X = k) = \sum_{k=1}^{\infty} \frac{k}{k!} \lambda^k e^{-\lambda} = \sum_{k=1}^{\infty} \frac{1}{(k-1)!} \lambda^k e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{1}{(k-1)!} \lambda^{k-1}$$

On reconnaît dans la série le développement de e^{λ}. Donc :

$$\mathbf{E}(X) = \lambda$$

❑ *Loi géométrique* :

$$\mathbf{E}(X) = \sum_{k=1}^{\infty} k\ P(X = k) = \sum_{k=1}^{\infty} k\ (1-p)^{k-1}\ p$$

On reconnaît une série du type $\sum kx^{k-1}$ de somme $\frac{1}{(1-x)^2}$. D'où :

$$\mathbf{E}(X) = \frac{1}{p}$$

EXEMPLE 1 :
Un joueur joue au loto une grille (5 numéros sur 49) par semaine jusqu'à ce qu'il gagne le gros lot. Quelle est la durée moyenne de jeu ? Le nombre de semaines est une variable aléatoire de loi

géométrique $p = \dfrac{1}{\binom{49}{5}} = \dfrac{1}{1906884}$. Son espérance est donc de 1906884 semaines, soit plus de 35 000 ans.

7- Variance et covariance

Si X^2 est d'espérance finie, alors il en est de même de X. En effet, $\sum x_n^2\ P(X = x_n)$ est une série convergente, ainsi que $\sum P(X = x_n)$. Or $|x_n|\ P(X = x_n) \leq \dfrac{x_n^2\ P(X = x_n) + P(X = x_n)}{2}$ car cette inégalité se déduit de $|x_n| \leq \dfrac{x_n^2 + 1}{2}$ ou de $x_n^2 - 2|x_n| + 1 \geq 0$ ou de $(|x_n| - 1)^2 \geq 0$ qui est vraie.

Donc $\sum |x_n|\ P(X = x_n)$ converge.

On définit la variance de X comme suit :

DEFINITION DE LA VARIANCE

$$V(X) = \mathbf{E}((X - \mathbf{E}(X))^2) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$$

Son écart-type est défini par :

$$\sigma(X) = \sqrt{V(X)}$$

On remarquera que $V(aX + b) = V(aX) = a^2V(X)$

Les variances des lois usuelles sont les suivantes :

❑ *Loi de Poisson* :

De même, calculons $\mathbf{E}(X(X - 1))$, plus facile à calculer que $\mathbf{E}(X^2)$

$$\mathbf{E}(X(X - 1)) = \sum_{k=0}^{\infty} k(k - 1)\ P(X = k) = \sum_{k=2}^{\infty} \frac{k(k - 1)}{k!} \lambda^k e^{-\lambda}$$

$$= \sum_{k=2}^{\infty} \frac{1}{(k - 2)!} \lambda^k e^{-\lambda} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{1}{(k - 2)!} \lambda^{k-2}$$

$$= \lambda^2$$

Donc $\mathbf{E}(X^2) = \mathbf{E}(X(X - 1)) + \mathbf{E}(X) = \lambda^2 + \lambda$

Donc $V(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \lambda$

Ces valeurs sont à rapprocher de celles de la loi binomiale :

$$\mathbf{E}(X) = np$$
$$V(X) = np(1 - p)$$

En effet, lorsque n tend vers l'infini, p vers 0 de sorte que np tende vers λ, la loi binomiale converge vers la loi de Poisson. On remarque qu'il y a également convergence de l'espérance de la variance :

$$np \to \lambda$$
$$np(1 - p) \to \lambda$$

Une autre façon de calculer les espérances et les variances est donnée dans le paragraphe sur les fonctions génératrices

❑ *Loi géométrique* :

Calculons également $\mathbf{E}(X(X-1))$, plus facile à calculer que $\mathbf{E}(X^2)$:

$$\mathbf{E}(X(X-1)) = \sum_{k=1}^{\infty} k(k-1)\, P(X=k) = \sum_{k=1}^{\infty} k(k-1)\,(1-p)^{k-1}\, p$$

$$= p(1-p) \sum_{k=2}^{\infty} k(k-1)\,(1-p)^{k-2}$$

On reconnaît une série du type $\sum k(k-1)\, x^{k-2}$ de somme $\dfrac{2}{(1-x)^3}$. D'où :

$$\mathbf{E}(X(X-1)) = \frac{2(1-p)}{p^2}$$

$$\Rightarrow \qquad \mathbf{E}(X^2) = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}$$

Donc $\quad V(X) = \dfrac{1-p}{p^2}$

EXEMPLE, issu de la mécanique quantique :
On suppose qu'une particule possède une énergie quantifié $nh\nu$, $n \geq 0$, et la probabilité que son énergie E soit $nh\nu$ est proportionnelle à $\exp(-\dfrac{nh\nu}{kT})$ où h est la constante de Planck, k la constante de Boltzmann, ν une fréquence associée à la particule et T la température (en K).

$$\exists\, C,\ \forall n \in \mathbb{N},\ P(E = nh\nu) = C \exp(-\frac{nh\nu}{kT})$$

La variable aléatoire $X = \dfrac{E}{h\nu}$ suit donc d'une loi géomètrique de paramètre $p = 1 - \exp(-\dfrac{h\nu}{kT})$. La variable n décrit $\mathbb{N}$ et pas $\mathbb{N}^*$, ce qui entraîne un décalage d'une unité par rapport à la situation décrite précédemment. La constante de proportionalité C vérifie :

$$C = \frac{1}{1-p} = \frac{1}{1-\exp(-\frac{h\nu}{kT})} = \frac{\exp(\frac{h\nu}{kT})}{\exp(\frac{h\nu}{kT}) - 1}.$$

L'énergie moyenne (espérance de la loi de probabilité) est ici $h\nu\, \dfrac{1-p}{p} = \dfrac{h\nu}{\exp(\frac{h\nu}{kT}) - 1}$. On pourra vérifier que la fonction $\nu \to \dfrac{h\nu}{\exp(\frac{h\nu}{kT}) - 1}$ est strictement décroissante et tend vers kT quand ν tend vers 0. Ainsi, l'énergie moyenne vaut environ kT si $h\nu$ est petit devant kT (cas où T est élevé ou ν faible). Si, au contraire, $h\nu$ est élevé devant kT, l'énergie moyenne est de l'ordre de $h\nu \exp(-\dfrac{h\nu}{kT})$ petit devant kT (dans ce dernier cas, la probabilité que l'énergie soit strictement positive est faible).

La médiane m est la valeur telle que $P(X \geq m) = \dfrac{1}{2}$ ce qui correspond à :

$$(1-p)^m = \frac{1}{2} \Leftrightarrow m \ln(1-p) = -\ln(2) \Leftrightarrow m = \frac{kT}{h\nu} \ln(2)$$

L'énergie médiane est donc kT ln(2), inférieure à l'énergie moyenne maximale kT, mais du même ordre. Si la particule sert à effectuer une mesure, déclenchant un système de détection dès qu'elle reçoit une dose d'énergie ΔE significative, il convient que ΔE > kT ln(2) de façon à être distinguée du bruit de fond thermique. Cette mesure se traduit par un accroissement d'entropie $\Delta S = \dfrac{\Delta E}{T} > k \ln(2)$. k est l'ordre de grandeur de l'accroissement minimal d'entropie nécessaire pour effectuer une mesure ou acquérir une information. C'est une quantité d'information minimale.

La variance de la loi géométrique étant $V(X) = \dfrac{1-p}{p^2} = \dfrac{\exp(-\frac{h\nu}{kT})}{(1-\exp(-\frac{h\nu}{kT}))^2} = \dfrac{\exp(\frac{h\nu}{kT})}{(\exp(\frac{h\nu}{kT})-1)^2}$, son écart-type vaut $\sigma(X) = \sqrt{V(X)} = \dfrac{\exp(\frac{h\nu}{2kT})}{\exp(\frac{h\nu}{kT})-1}$ et l'écart-type de l'énergie est $h\nu \dfrac{\exp(\frac{h\nu}{2kT})}{\exp(\frac{h\nu}{kT})-1}$. Cet écart-type mesure la dispersion de l'énergie due autour de l'énergie moyenne. Si $h\nu$ est petit devant kT, elle est de l'ordre de kT. On retrouve ainsi le fait qu'une variation d'énergie ΔE sera d'autant plus significative qu'elle sera supérieure à kT. Si $h\nu$ est grand devant kT, l'écart-type est de l'ordre de $h\nu \exp(-\frac{h\nu}{2kT})$ bien plus grand que l'énergie moyenne $h\nu \exp(-\frac{h\nu}{kT})$. A haute fréquence, les fluctuations relatives sont plus importantes qu'à basse fréquence.

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

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

Toutes les quantités intervenant ci-dessus sont définies si X et Y sont de variance finie. En effet, on sait déjà que, dans ce cas, X et Y sont d'espérance finie. Reste à montrer que (X – **E**(X))(Y – **E**(Y)) est d'espérance finie. Posons X' = X – **E**(X) et Y' = Y – **E**(Y). Si x_i' sont les valeurs prises par X' et y_j' celles prises par Y', on a :

$$|x_i'\, y_j'|\ P(X' = x_i', Y' = y_j') \leq \frac{x_i'^2 + y_j'^2}{2} P(X' = x_i', Y' = y_j')$$

Mais $\displaystyle\sum_{i,j} \frac{x_i'^2 + y_j'^2}{2} P(X' = x_i', Y' = y_j')$ converge. Elle vaut $\dfrac{\mathbf{E}(X'^2) + \mathbf{E}(Y'^2)}{2} = \dfrac{V(X') + V(Y')}{2}$

On pose alors la covariance de X et Y comme suit :

DEFINITION DE LA COVARIANCE

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

D'où :

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

(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

Variances et covariances vérifient les propriétés suivantes :

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

$$\begin{aligned}\text{cov}(X,Y) &= \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y)))\\ &= \mathbf{E}(XY - X\mathbf{E}(Y) - Y\mathbf{E}(X) + \mathbf{E}(X)\mathbf{E}(Y))\\ &= \mathbf{E}(XY) - \mathbf{E}(X\mathbf{E}(Y)) - \mathbf{E}(Y\mathbf{E}(X)) + \mathbf{E}(\mathbf{E}(X)\mathbf{E}(Y))\\ &= \mathbf{E}(XY) - \mathbf{E}(Y)\mathbf{E}(X) - \mathbf{E}(X)\mathbf{E}(Y) + \mathbf{E}(X)\mathbf{E}(Y)\\ &= \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)\end{aligned}$$

iii) La vérification est laissée au lecteur
iv) résulte du fait qu'une forme bilinéaire symétrique positive vérifie l'inégalité de Cauchy-Schwarz.
v) Si X et Y sont indépendantes, alors $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$. 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 $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$ 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 l si $a > 0$ et – l si $a < 0$.

INEGALITE DE BIENAYME-TCHEBYCHEV

Soit X *une variable aléatoire telle que* X^2 *a une espérance finie*. *Soit* $\mu = \mathbf{E}(X)$ *et* $\sigma^2 = V(X)$. *Alors, pour tout* $\varepsilon > 0$:

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

<u>Démonstration</u> :

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

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

D'où le résultat.

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 :

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

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

EXEMPLE :
Soit X une variable aléatoire telle que $V(X) = 0$. Alors X est presque sûrement constante. En effet, l'inégalité de Bienaymé-Tchebychev donne, pour tout entier n :

$$P(|X - \mathbf{E}(X)| \geq \frac{1}{n}) \leq V(X)n^2 = 0$$

Les événements $|X - \mathbf{E}(X)| \geq \frac{1}{n}$ sont croissants, de réunion $|X - \mathbf{E}(X)| > 0$. On a donc :

$$P(|X - \mathbf{E}(X)| > 0) = \lim_{n\to\infty} P(|X - \mathbf{E}(X)| \geq \frac{1}{n}) = 0$$

Donc $P(|X - \mathbf{E}(X)| = 0) = 1$ et X est presque sûrement égale à la constante $\mathbf{E}(X)$.

LOI FAIBLE DES GRANDS NOMBRES

Soit (X_n) *une suite de variables aléatoires indépendantes de même loi, d'espérance* μ *et d'écart-type* σ. *Posons* $S_n = X_1 + ... + X_n$. *Alors, pour tout* $\varepsilon > 0$:

$$\lim_{n\to\infty} P(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon) = 0$$

<u>Démonstration</u> :
Appliquons l'inégalité de Bienaymé-Tchebychev à $Z_n = \frac{X_1 + ... + X_n}{n}$. On a $\mathbf{E}(Z_n) = \mu$ et :

$$V(Z_n) = \frac{1}{n^2} V(X_1 + ... + X_n) = \frac{1}{n^2} \times (V(X_1) + ... + V(X_n)) \text{ car les } X_i \text{ sont indépendantes}$$
$$= \frac{\sigma^2}{n}$$

Donc :

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

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.

Donnons également un exemple montrant que l'inégalité de Bienaymé-Tchebychev peut également s'appliquer à des domaines a priori sans lien avec les probabilités.

EXEMPLE :

Prouver que $\sum_{k=1}^{\infty} \frac{x^k}{k!\sqrt{k}} \sim \frac{e^x}{\sqrt{x}}$ quand x tend vers $+\infty$.

C'est équivalent à montrer que, si on pose $S(x) = e^{-x} \sum_{k=1}^{\infty} \frac{\sqrt{x}}{\sqrt{k}} \frac{x^k}{k!}$, on a $\lim_{x \to +\infty} S(x) = 1$. Considérons une variable aléatoire X suivant une loi de Poisson de paramètre x. On constate que :

$$S(x) = \sum_{k=1}^{\infty} \frac{\sqrt{x}}{\sqrt{k}} P(X = k)$$

donc
$$S(x) - 1 = \sum_{k=1}^{\infty} \frac{\sqrt{x}}{\sqrt{k}} P(X = k) - 1$$

$$= \sum_{k=1}^{\infty} \frac{\sqrt{x}}{\sqrt{k}} P(X = k) - \sum_{k=0}^{\infty} P(x = k)$$

$$= \sum_{k=1}^{\infty} (\frac{\sqrt{x}}{\sqrt{k}} - 1)\ P(X = k) - P(X = 0)$$

$$= \sum_{k=1}^{\infty} (\frac{\sqrt{x}}{\sqrt{k}} - 1)\ P(X = k) - e^{-x}$$

Comme $\lim_{x \to +\infty} e^{-x} = 0$, il suffit de montrer que $\lim_{x \to \infty} \sum_{k=1}^{\infty} (\frac{\sqrt{x}}{\sqrt{k}} - 1)\ P(X = k) = 0$. Pour cela, écrivons l'inégalité de Bienaymé-Tchebychev. Pour tout a strictement positif, on a :

$$P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}$$

donc $P(|X - x| \geq a) \leq \frac{x}{a^2}$ puisque $\mathbf{E}(X) = x$ et $V(x) = x$.

Dans la suite, on prendra $a = x^{7/8}$, un peu inférieur à x, tout en étant négligeable devant x quand x tend vers $+\infty$. En notant $m = \lfloor x - a \rfloor$ et $n = \lfloor x + a \rfloor$, on coupe la valeur absolue de la somme considérée en trois morceaux dont on va montrer que chacun tend vers 0 :

$$\left| \sum_{k=1}^{\infty} (\frac{\sqrt{x}}{\sqrt{k}} - 1)\ P(X = k) \right| \leq \sum_{k=1}^{m} \left| \frac{\sqrt{x}}{\sqrt{k}} - 1 \right| P(X = k) + \sum_{k=m+1}^{n} \left| \frac{\sqrt{x}}{\sqrt{k}} - 1 \right| P(X = k) + \sum_{k=n+1}^{\infty} \left| \frac{\sqrt{x}}{\sqrt{k}} - 1 \right| P(X = k)$$

<u>Premier terme</u> : Pour $k \leq m \leq x - a \leq x$, on a $\left| \frac{\sqrt{x}}{\sqrt{k}} - 1 \right| = \frac{\sqrt{x}}{\sqrt{k}} - 1 \leq \sqrt{x} - 1 \leq \sqrt{x}$. Donc :

$$\sum_{k=1}^{m} \left| \frac{\sqrt{x}}{\sqrt{k}} - 1 \right| P(X = k) \leq \sum_{k=1}^{m} \sqrt{x}\ P(X = k)$$

$\leq \sqrt{x}\ P(X \leq m)$

$\leq \sqrt{x}\ P(X \leq x - a)$ car $\{X \leq m\} \subset \{X \leq x - a\}$

$\leq \sqrt{x}\ P(|X - x| \geq a)$ car $\{X \leq x - a\} \subset \{|X - x| \geq a\}$

$$\leq \frac{x\sqrt{x}}{a^2} \qquad \text{d'après l'inégalité de Bienaymé-Tchebychev}$$
$$\leq \frac{1}{x^{1/4}} \qquad \to 0 \text{ quand } x \text{ tend vers } +\infty$$

Troisième terme : de même, pour $k \geq n + 1 \geq x + a \geq x$, on a $\left|\frac{\sqrt{x}}{\sqrt{k}} - 1\right| = 1 - \frac{\sqrt{x}}{\sqrt{k}} \leq 1$. Donc :

$$\sum_{k=n+1}^{\infty} \left|\frac{\sqrt{x}}{\sqrt{k}} - 1\right| \mathrm{P}(\mathrm{X} = k) \leq \sum_{k=n+1}^{\infty} \mathrm{P}(\mathrm{X} = k)$$
$$\leq \mathrm{P}(\mathrm{X} \geq n + 1)$$
$$\leq \mathrm{P}(\mathrm{X} \geq x + a)$$
$$\leq \mathrm{P}(|\mathrm{X} - x| \geq a)$$
$$\leq \frac{x}{a^2}$$
$$\leq \frac{1}{x^{3/4}} \qquad \to 0 \text{ quand } x \text{ tend vers } +\infty$$

Deuxième terme :

$$\sum_{k=m+1}^{n} \left|\frac{\sqrt{x}}{\sqrt{k}} - 1\right| \mathrm{P}(\mathrm{X} = k) \leq \underset{k \in [\![m,n+1]\!]}{\mathrm{Sup}} \left|\frac{\sqrt{x}}{\sqrt{k}} - 1\right| \sum_{k=m+1}^{n} \mathrm{P}(\mathrm{X} = k)$$
$$\leq \underset{k \in [\![m,n+1]\!]}{\mathrm{Sup}} \left|\frac{\sqrt{x}}{\sqrt{k}} - 1\right|$$

Comme la suite $k \to \frac{\sqrt{x}}{\sqrt{k}} - 1$ est décroissante, positive pour $k = m$ puis négative pour $k = n + 1$, la suite $k \to \left|\frac{\sqrt{x}}{\sqrt{k}} - 1\right|$ décroît, passe par un minimum, puis croît lorsque k varie de m à $n + 1$, de sorte que sa borne supérieure n'est autre que Max $\{\frac{\sqrt{x}}{\sqrt{m}} - 1, 1 - \frac{\sqrt{x}}{\sqrt{n+1}}\}$. Or :

$$0 \leq \frac{\sqrt{x}}{\sqrt{m}} - 1 \leq \frac{\sqrt{x}}{\sqrt{x-a-1}} - 1 = \frac{1}{\sqrt{x-a-1}} (\sqrt{x} - \sqrt{x-a-1})$$
$$\leq \frac{1}{\sqrt{x-a-1}} \frac{a+1}{\sqrt{x} + \sqrt{x-a-1}} \sim \frac{a}{2x} = \frac{1}{2x^{1/4}} \to 0$$

et de même pour $1 - \frac{\sqrt{x}}{\sqrt{n+1}}$. Donc le Max des deux termes tend aussi vers 0.

III : Fonctions génératrices

1- Définition

Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On cherche à définir une fonction dont la seule connaissance suffise à retrouver la loi. La fonction la plus naturelle est la série entière sur $\mathbb{C}$ définie par :

$$G_X(z) = \sum_{n=0}^{\infty} P(X = n)\, z^n$$

Cette fonction s'appelle fonction génératrice de X. Elle n'est autre que $\mathbf{E}(z^X)$, en vertu de la formule générale $\mathbf{E}(f(X)) = \sum_{n=0}^{\infty} f(n)\, P(X = n)$. On a donc :

$$G_X(z) = \mathbf{E}(z^X)$$

PROPRIETES DE LA FONCTION GENERATRICE

i) *Le rayon* R *de convergence de la série est supérieur ou égal à* 1

ii) G_X *est dérivable en* 1 *si et seulement si* X *admet une espérance, et dans ce cas :*

$$\mathbf{E}(X) = G_X'(l)$$

iii) G_X *est deux fois dérivable en* 1 *si et seulement si* X *admet une variance, et dans ce cas :*

$$V(X) = G_X''(l) + G_X'(l) - G_X'(l)^2$$

Démonstration :

i) Pour $z = 1$, on a $G_X(l) = \sum_{n=0}^{\infty} P(X = n) = 1$. Puisque la série entière définissant G_X converge en 1, le rayon de convergence R est supérieur ou égal à 1.

ii) Si R > 1, G_X est indéfiniment dérivable sur]– R, R[, et l'on a, pour tout t élément de]–R, R[(et donc en 1) :

$$G_X'(t) = \sum_{n=1}^{\infty} P(X = n)\, nt^{n-1} = \mathbf{E}(X\, t^{X-1}) \qquad \text{donc} \quad G_X'(l) = \mathbf{E}(X)$$

Le cas R = 1 est plus délicat :

Si X admet une espérance, le calcul précédent reste valide car $\sum n\, P(X = n)$ converge, donc la série de fonctions $\sum n\, P(X = n)\, t^{n-1}$ converge normalement sur [0, 1], donc G_X est C^1 sur [0, 1] et sa dérivée se calcule terme à terme.

Réciproquement, supposons que $G_X'(1)$ est défini. G_X étant une série entière de rayon 1, on sait que, pour t élément de]–1, 1[, $G_X'(t) = \sum_{n=1}^{\infty} n\, P(X = n)\, t^{n-1}$ mais on ignore a priori si cette relation reste vraie en 1. On note cependant que G_X est continue sur [0, 1], et sur [0, 1[, G_X' est croissante, donc admet une limite en 1, finie ou non. Mais si une fonction est C^1 sur un intervalle $[a, b[$, continue sur $[a, b]$ et si sa dérivée admet une limite en b, ou bien cette limite est infinie et la fonction n'est pas dérivable en b, ou bien cette limite est finie en b et la fonction est C^1 sur $[a, b]$. Comme on a supposé que G_X est dérivable en 1, c'est le second cas qui s'applique et on a $\lim_{t \to 1,\, t<1} G_X'(t) = G_X'(1)$ et en particulier, $G_X'(t) \leq G_X'(1)$ car G_X' étant croissante, elle est inférieure à sa limite. On a alors, pour tout N, et tout t élément de [0, 1[:

$$\sum_{n=1}^{N} n\, P(X = n)\, t^{n-1} \leq \sum_{n=1}^{\infty} n\, P(X = n)\, t^{n-1} = G_X'(t) \leq G_X'(1)$$

donc, en passant à la limite quand t tend vers 1 :

$$\forall \text{ N}, \sum_{n=1}^{\text{N}} n \text{ P}(\text{X} = n) \leq \text{G}_\text{X}'(1)$$

La suite des sommes partielles $\sum_{n=1}^{\text{N}} n \text{ P}(\text{X} = n)$ est croissante majorée, donc converge. Il en résulte que la série $\sum n \text{ P}(\text{X} = n)$ converge et donc que $\mathbf{E}(\text{X})$ existe. Mais nous avons vu précédemment que, dans ce cas $\mathbf{E}(\text{X}) = \text{G}_\text{X}'(1)$.

iii) On se bornera à vérifier les formules indiquées lorsque R > 1 en supposant que le calcul reste valide si R = 1 :

$$\text{G}_\text{X}''(z) = \sum_{n=2}^{\infty} \text{P}(\text{X} = n)\, n(n-1)z^{n-2} = \mathbf{E}(\text{X}(\text{X}-1)\, z^{\text{X}-2}) \qquad \text{donc } \text{G}_\text{X}''(1) = \mathbf{E}(\text{X}(\text{X}-1))$$

et $\quad \mathbf{E}(\text{X}^2) = \text{G}_\text{X}''(1) + \mathbf{E}(\text{X}) = \text{G}_\text{X}''(1) + \text{G}_\text{X}'(1)$

Donc $\quad \text{V}(\text{X}) = \text{G}_\text{X}''(1) + \text{G}_\text{X}'(1) - \text{G}_\text{X}'(1)^2$

2- Lois usuelles

Calculons la fonction génératrice des lois usuelles, et retrouvons les expressions des espérances et variances de ces lois.

❑ *Loi de Bernoulli* :

$\text{G}_\text{X}(z) = 1 - p + pz$

$\text{G}_\text{X}'(z) = p \qquad$ donc $\quad \mathbf{E}(\text{X}) = p$

$\text{G}_\text{X}''(z) = 0 \qquad$ donc $\quad \text{V}(\text{X}) = p(1-p)$

❑ *Loi binomiale* $\mathcal{B}(n, p)$:

$$\text{G}_\text{X}(z) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} z^k = (pz + 1 - p)^n$$

$\text{G}_\text{X}'(z) = np(pz + 1 - p)^{n-1} \qquad$ donc $\quad \mathbf{E}(\text{X}) = np$

$\text{G}_\text{X}''(z) = n(n-1)p^2(pz + 1 - p)^{n-2} \qquad$ donc $\quad \text{V}(\text{X}) = n(n-1)p^2 + np - n^2p^2 = np(1-p)$

On notera la facilité avec laquelle on peut calculer la variance de la loi binomiale.

❑ *Loi de Poisson* $\mathcal{P}(\lambda)$:

$$\text{G}_\text{X}(z) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} z^n = \exp(\lambda z - \lambda)$$

$\text{G}_\text{X}'(z) = \lambda\exp(\lambda z - \lambda) \qquad$ donc $\quad \mathbf{E}(\text{X}) = \lambda$

$\text{G}_\text{X}''(z) = \lambda^2\exp(\lambda z - \lambda) \qquad$ donc $\quad \text{V}(\text{X}) = \lambda^2 + \lambda - \lambda^2 = \lambda$

❑ *Loi géométrique sur* $\mathbb{N}^*$:

$$\text{G}_\text{X}(z) = \sum_{n=1}^{\infty} (1-p)^{n-1} p\, z^n = \frac{pz}{1-(1-p)z}$$

Cette série n'est définie que pour $|z| < \dfrac{1}{1-p}$.

$$G_X'(z) = \frac{p}{(1-(1-p)z)^2} \qquad \text{donc} \quad \mathbf{E}(X) = \frac{1}{p}$$

$$G_X''(z) = \frac{2p(1-p)}{(1-(1-p)z)^3} \qquad \text{donc} \quad V(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

3- Fonction génératrice d'une somme de variables aléatoires indépendantes

Soit X et Y deux variables aléatoires indépendantes de fonctions génératrices G_X et G_Y. Quelle est la fonction génératrices de X + Y ?

$$G_{X+Y}(z) = \sum_{n=0}^{\infty} P(X + Y = n)\, z^n$$

$$= \sum_{n=0}^{\infty} \sum_{k=0}^{n} P(X = k, Y = n - k)\, z^n$$

$$= \sum_{n=0}^{\infty} \sum_{k=0}^{n} P(X = k)\, P(Y = n - k)\, z^n \qquad \text{car les variables sont indépendantes}$$

On reconnaît alors l'expression d'une série produit :

$$G_{X+Y}(z) = \sum_{n=0}^{\infty} P(X = n)\, z^n \times \sum_{n=0}^{\infty} P(Y = n)\, z^n$$

Ainsi :

$$G_{X+Y} = G_X \times G_Y$$

Une autre démonstration consiste à dire :

$$G_{X+Y}(z) = \mathbf{E}(z^{X+Y}) = \mathbf{E}(z^X z^Y) = \mathbf{E}(z^X)\, \mathbf{E}(z^Y) = G_X(z) \times G_Y(z)$$

Cette démonstration utilise les résultats suivants :

i) si X et Y sont indépendantes, alors une fonction de X (à savoir z^X) est indépendante d'une fonction de Y (à savoir z^Y).

ii) L'espérance du produit de deux variables aléatoires indépendantes est égal au produit des espérances de chacune de ces variables.

Le résultat prouvé se généralise à la somme de *n* variables aléatoires indépendantes. La fonction génératrice de la somme est le produit des *n* fonctions génératrices.

EXEMPLES D'UTILISATION :

❑ La loi binomiale $\mathcal{B}(n, p)$ est la loi de la somme de *n* variables aléatoires indépendantes de Bernoulli de paramètre *p*. On retrouve effectivement :

$$G_{\text{binomiale}}(z) = (1 - p + pz)^n = G_{\text{Bernoulli}}(z)^n$$

❑ La somme de deux variables aléatoires indépendantes de lois binomiales $\mathcal{B}(n, p)$ et $\mathcal{B}(m, p)$ suit une loi binomiale $\mathcal{B}(n + m, p)$. Et en effet :

$$(1 - p + pz)^n\ (1 - p + pz)^m = (1 - p + pz)^{n+m}$$

❑ La somme de *n* variables aléatoires indépendantes de loi de Poisson de paramètres $\lambda_1, ..., \lambda_n$ suit une loi de Poisson de paramètre la somme des λ_i. En effet :

$$\exp(\lambda_1(z-1))...\exp(\lambda_n(z-1)) = \exp((\lambda_1 + ... + \lambda_n)(z-1))$$

Une utilisation des séries génératrices et de la diagonalisation des matrices est donnée dans l'annexe qui suit.

Annexe I : Chaînes de Markov

1- Matrice de transition d'un système

Considérons un système pouvant se trouver dans divers états possibles, numérotés de 1 à n. A intervalle de temps régulier, le système change d'état, la probabilité qu'il passe de l'état i à l'état j étant p_{ij}. Il est possible également qu'on puisse rester dans le même état i avec une certaine probabilité, ce qui signifie que p_{ii} est non nul. Parfois, parmi les n états, il existe un état (ou plusieurs) final, par exemple l'état n. Si on atteint cet état, on y reste définitivement, ce qu'on peut traduire par :

$$p_{nn} = 1 \qquad \text{et} \qquad p_{ni} = 0 \text{ si } i \neq n$$

Notons M la matrice de $\mathcal{M}_n(\mathbb{R})$ dont le terme (i, j) vaut p_{ij}. M est la matrice de transition du système. On remarque que la somme des coefficients de chaque ligne vaut 1, égale à la probabilité qu'on passe de l'état i à un autre état (éventuellement le même).

A l'instant k, on note $X(k)$ la variable aléatoire précisant dans quel état se trouve le système. Soit $L(k)$ le vecteur-ligne dont la i-ème composante est égale à $P(X(k) = i)$, probabilité que le système se trouve à l'état i. $L(k)$ représente la loi de $X(k)$. A l'instant initial, le système de trouve dans l'état i suivant une certaine loi de probabilité, définissant le vecteur $L(0)$. Par exemple, si le système se trouve de manière certaine à l'état 1, alors $L(0) = (1\ 0\ \dots\ 0)$. S'il se trouve à l'instant initial dans l'état 1 ou 2 avec équiprobabilité, alors $L(0) = (\frac{1}{2}\ \ \frac{1}{2}\ \ 0\ \ \dots\ \ 0)$.

Considérons un instant k particulier. L'évènement $X(k) = i$ consiste à se trouver dans l'état i à cet instant. L'évènement $X(k + 1) = j$, consistant à se trouver dans l'état j à l'instant $k + 1$. La probabilité conditionnelle $P(X(k + 1) = j \mid X(k)=i)$ n'est autre que p_{ij}. La loi des probabilité totale énonce que :

$$P(X(k + 1) = j) = \sum_{j=1}^{n} P(X(k + 1) = j \mid X(k) = i)\ P(X(k) = i)$$

Or $P(X(k) = i)$ n'est autre que la i-ème composante de $L(k)$, qu'on peut noter $L_i(k)$, et $P(X(k + 1) = j)$ est la j-ème composante de $L(k + 1)$, que l'on peut noter $L_j(k + 1)$. On a donc :

$$\forall\ j,\ L_j(k + 1) = \sum_{j=1}^{n} p_{ij}\ L_i(k)$$

ce qui exprime exactement le fait que $L(k + 1) = L(k)M$. On notera bien qu'il s'agit dans le membre de droite du produit de la ligne $L(k)$ par la matrice M. Si on souhaitait manipuler les vecteurs donnant la loi de $X(k)$ en colonne, il conviendrait de prendre ${}^{t}M$.

Par récurrence, on en déduit que $L(k) = L(0)\ M^k$. On détermine explicitement $L(k)$ en parvenant à calculer M^k.

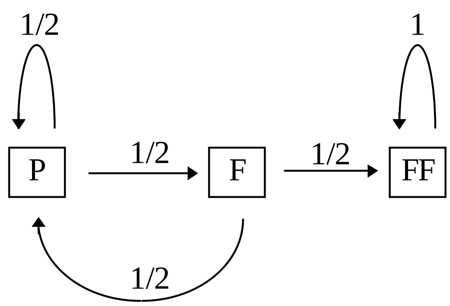
Le schéma précédent caractérise les chaînes de Markov. Dans ce type de processus, ce qui se passe en partant de l'état i est indépendant de la façon dont on est arrivé à l'état i : le futur est indépendant du passé. On peut résumer l'étude précédente de la façon suivante :

PROPOSITION

Soit une chaîne de Markov $(X(k))_{k\in\mathbb{N}}$, *où* $X(k)$ *désigne la variable aléatoire donnant l'état d'un système fini au bout du temps k. On se donne la matrice de transition* M *du système dont le terme général vaut* $p_{ij} = P(X(k+1) = j \mid X(k) = i)$ *et ne dépend pas de k. Soit* $L(k)$ *le vecteur-ligne de composantes* $P(X(k) = i)$. *Alors* $L(k) = L(0)\, M^k$ *ce qui permet de déterminer la loi de* $X(k)$ *à partir de celle de* $X(0)$.

EXEMPLE :
On lance une pièce jusqu'à obtenir deux Faces (FF). Il est clair que, si on tire Pile (P) avant d'avoir tiré les deux Faces, tout est à recommencer. On dispose donc des trois états suivants :

état 1 P : état initial ou état après avoir tiré un Pile
état 2 F : état suivant l'état 1 après tirage d'un Face
état 3 FF : état final, après tirage successif de deux Faces



Pour une pièce équilibrée, la matrice de transition est $M = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{pmatrix}$. Supposons que l'on parte de $L(0) = (1 \ \ 0 \ \ 0)$ (probabilité certaine d'être dans l'état initial). Les lois $L(k)$ de $X(k)$ valent successivement, en multipliant itérativement par M :

$$L(0) = (1 \quad 0 \quad 0)$$
$$L(1) = (\tfrac{1}{2} \quad \tfrac{1}{2} \quad 0)$$
$$L(2) = (\tfrac{1}{2} \quad \tfrac{1}{4} \quad \tfrac{1}{4})$$
$$L(3) = (\tfrac{3}{8} \quad \tfrac{1}{4} \quad \tfrac{3}{8})$$
$$L(4) = (\tfrac{5}{16} \quad \tfrac{3}{16} \quad \tfrac{1}{2})$$
$$L(5) = (\tfrac{1}{4} \quad \tfrac{5}{32} \quad \tfrac{19}{32})$$
$$L(6) = (\tfrac{13}{64} \quad \tfrac{1}{8} \quad \tfrac{43}{64})$$
...

La somme des termes de chaque ligne doit donner 1. On peut déterminer $L(k)$ en calculant M^k en la diagonalisant. Le polynôme caractéristique de M vaut :

$$\chi_M(X) = X^3 - \frac{3X^2}{2} + \frac{X}{4} + \frac{1}{4} = (X-1)(X^2 - \frac{X}{2} - \frac{1}{4})$$

de racines 1, $\varphi = \dfrac{1+\sqrt{5}}{4}$ et $\psi = \dfrac{1-\sqrt{5}}{4}$. M est diagonalisable de matrice diagonale $D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \varphi & 0 \\ 0 & 0 & \psi \end{pmatrix}$.

Une matrice de passage est donnée par $Q = \begin{pmatrix} 1 & 2\varphi & 2\psi \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}$. $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ est toujours vecteur propre d'une telle matrice, associée à la valeur propre 1 du fait que la somme des lignes de M vaut 1. Q^{-1} vaut :

$$Q^{-1} = \frac{1}{2\varphi - 2\psi} \begin{pmatrix} 0 & 0 & 2\varphi-2\psi \\ 1 & -2\psi & -2\varphi \\ -1 & 2\varphi & 2\psi \end{pmatrix}$$

On a $M = QDQ^{-1}$ et $M^k = QD^kQ^{-1}$.

Donc : $L(k) = L(0)Q\,D^k\,Q^{-1}$

$$= (1\ 0\ 0) \begin{pmatrix} 1 & 2\varphi & 2\psi \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \varphi^k & 0 \\ 0 & 0 & \psi^k \end{pmatrix} \frac{1}{2\varphi - 2\psi} \begin{pmatrix} 0 & 0 & 2\varphi-2\psi \\ 1 & -2\psi & -2\varphi \\ -1 & 2\varphi & 2\psi \end{pmatrix}$$

$$= (1\ 2\varphi\ 2\psi) \begin{pmatrix} 1 & 0 & 0 \\ 0 & \varphi^k & 0 \\ 0 & 0 & \psi^k \end{pmatrix} \frac{1}{2\varphi - 2\psi} \begin{pmatrix} 0 & 0 & 2\varphi-2\psi \\ 1 & -2\psi & -2\varphi \\ -1 & 2\varphi & 2\psi \end{pmatrix}$$

$$= (1\ 2\varphi^{k+1}\ 2\psi^{k+1}) \frac{1}{2\varphi - 2\psi} \begin{pmatrix} 0 & 0 & 2\varphi-2\psi \\ 1 & -2\psi & -2\varphi \\ -1 & 2\varphi & 2\psi \end{pmatrix}$$

$$= \left(\frac{\varphi^{k+1} - \psi^{k+1}}{\varphi - \psi} \quad \frac{2(\varphi\psi^{k+1} - \varphi^{k+1}\psi)}{\varphi - \psi} \quad 1 + \frac{2(\psi^{k+2} - \varphi^{k+2})}{\varphi - \psi}\right)$$

Ces trois nombres donnent la probabilité de se trouver à l'instant k respectivement dans les états 1, 2 et 3.

Que se passe-t-il quand k tend vers l'infini ? Comme $|\varphi| < 1$ et $|\psi| < 1$, φ^k et ψ^k tendent vers 0 et $L(k)$ tend vers $(0\ \ 0\ \ 1)$ et donc la probabilité de se trouver dans l'état final FF tend vers 1.

Soit T la variable aléatoire donnant l'instant k auquel on atteint l'état FF. On a :

$$P(T \le k) = P(X(k) = 3) = 1 + \frac{2(\psi^{k+2} - \varphi^{k+2})}{\varphi - \psi}$$

$$P(T \text{ fini}) = P(\cup \{T \le k\}) = \lim_{k\to\infty} P(T \le k) = 1$$

Ainsi, on est certain d'atteindre l'état final en un temps fini. Boucler indéfiniment dans le diagramme est un évènement de probabilité nulle. T s'appelle le temps d'arrêt du processus. Cherchons son espérance :

$$\mathbf{E}(T) = \sum_{k=0}^{\infty} k\,P(T = k) = \sum_{k=0}^{\infty} P(T > k) = \sum_{k=0}^{\infty} (1 - P(T \le k))$$

$$= \sum_{k=0}^{\infty} \frac{2(\varphi^{k+2} - \psi^{k+2})}{\varphi - \psi} = \frac{2}{\varphi - \psi}\left(\frac{\varphi^2}{1-\varphi} - \frac{\psi^2}{1-\psi}\right)$$

$$= \frac{2}{\varphi - \psi} \frac{\varphi^2 - \psi^2 - \varphi^2\psi + \varphi\psi^2}{1 - \varphi - \psi + \varphi\psi} = 2\,\frac{\varphi + \psi - \varphi\psi}{1 - \varphi - \psi + \varphi\psi} \qquad \text{or } \varphi + \psi = \frac{1}{2} \text{ et } \varphi\psi = -\frac{1}{4}$$

$$= 6$$

2- Temps d'arrêt

Soit, comme précédemment, un système de Markov doté d'un état final n. Soit T_i le temps d'arrêt pour passer de l'état i à l'état final n. On a alors, pour tout $i < n$ et tout $k > 0$:

$$T_i = 1 + \sum_{i=1}^{n} X_{ij}T_j \qquad \text{où } X_{ij} = 1 \text{ si on passe de l'état } i \text{ à l'état } j \text{ au premier saut et } = 0 \text{ sinon}$$

mais $T_n = 0$

donc $T_i = 1 + \sum_{i=1}^{n-1} X_{ij}T_j$

Si $k = 1$, $P(T_i = k) = P(T_i = 1) = p_{in}$. Si $k \geq 2$, on a :

$$P(T_i = k) = P(\sum_{i=1}^{n-1} X_{ij}T_j = k - 1) = P(\bigcup_{i=0}^{n-1} \{X_{ij} = 1 \cap T_j = k - 1\})$$

$$= \sum_{i=1}^{n-1} P(X_{ij} = 1 \cap T_j = k - 1) = \sum_{i=1}^{n-1} P(X_{ij} = 1)\, P(T_j = k-1)$$

$$= \sum_{i=1}^{n-1} p_{ij}\, P(T_j = k-1)$$

Si l'on note $T(k)$ le vecteur-colonne de composantes $P(T_i = k)$, $1 \leq i \leq n - 1$, on obtient la relation $T(k) = N\, T(k - 1)$, où N est la matrice $(n - 1) \times (n - 1)$ dont le terme général est p_{ij}, $1 \leq i \leq n - 1$. Il s'agit d'une sous-matrice de la matrice M déjà rencontrée.

On aura donc $T(k) = N^{k-1}\, T(1)$, avec $T(1) = \begin{pmatrix} p_{1n} \\ p_{2n} \\ \dots \\ p_{n-1,n} \end{pmatrix}$ donc $T(k) = N^{k-1}\, T(1)$. Le calcul de la loi des T_i peut donc se faire également en diagonalisant la matrice N.

EXEMPLE :

On reprend la matrice de transition de l'exemple précédent, on a $N = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 0 \end{pmatrix}$ et $T(1) = \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}$. On trouve que les $T(k)$ successifs valent :

$$\underset{T(1)}{\begin{pmatrix} 0 \\ 1/2 \end{pmatrix}} \quad \underset{T(2)}{\begin{pmatrix} 1/4 \\ 0 \end{pmatrix}} \quad \underset{T(3)}{\begin{pmatrix} 1/8 \\ 1/8 \end{pmatrix}} \quad \underset{T(4)}{\begin{pmatrix} 1/8 \\ 1/16 \end{pmatrix}} \quad \underset{T(5)}{\begin{pmatrix} 3/32 \\ 1/16 \end{pmatrix}} \quad \underset{T(6)}{\begin{pmatrix} 5/64 \\ 3/64 \end{pmatrix}} \quad \underset{\dots}{\dots}$$

La loi de T_1 se trouve dans la première composante de chaque colonne, la loi de T_2 dans la deuxième composante. On retrouve comme polynôme caractéristique de N le polynôme $X^2 - \frac{X}{2} - \frac{1}{4}$ de racine φ et ψ. N est diagonalisable de matrice diagonalisable $\Delta = \begin{pmatrix} \varphi & 0 \\ 0 & \psi \end{pmatrix}$ et de matrice de passage $R = \begin{pmatrix} 2\varphi & 2\psi \\ 1 & 1 \end{pmatrix}$ et d'inverse $R^{-1} = \frac{1}{2\varphi - 2\psi} \begin{pmatrix} 1 & -2\psi \\ -1 & 2\varphi \end{pmatrix}$. On a $N = R\Delta R^{-1}$ donc $N^{k-1} = R\, \Delta^{k-1}\, R^{-1}$ donc :

$$T(k) = R\, \Delta^{k-1}\, R^{-1}\, T(1)$$

$$= \begin{pmatrix} 2\varphi & 2\psi \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \varphi^{k-1} & 0 \\ 0 & \psi^{k-1} \end{pmatrix} \frac{1}{2\varphi - 2\psi} \begin{pmatrix} 1 & -2\psi \\ -1 & 2\varphi \end{pmatrix} \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}$$

$$= \frac{1}{2\varphi - 2\psi} \begin{pmatrix} 2\varphi & 2\psi \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \varphi^{k-1} & 0 \\ 0 & \psi^{k-1} \end{pmatrix} \begin{pmatrix} -\psi \\ \varphi \end{pmatrix}$$
$$= \frac{1}{2\varphi - 2\psi} \begin{pmatrix} 2\varphi & 2\psi \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -\psi\varphi^{k-1} \\ \varphi\psi^{k-1} \end{pmatrix}$$
$$= \frac{1}{2\varphi - 2\psi} \begin{pmatrix} 2\varphi\psi^k - 2\psi\varphi^k \\ \varphi\psi^{k-1} - \psi\varphi^{k-1} \end{pmatrix}$$

Ainsi :

$$P(T_1 = k) = \frac{\varphi\psi^k - \psi\varphi^k}{\varphi - \psi}$$
$$P(T_2 = k) = \frac{\varphi\psi^{k-1} - \psi\varphi^{k-1}}{2\varphi - 2\psi}$$

On en déduit que :

$$\mathbf{E}(T_1) = \sum_{k=0}^{\infty} k \frac{\varphi\psi^k - \psi\varphi^k}{\varphi - \psi} = \frac{1}{\varphi - \psi} \left(\frac{\varphi\psi}{(1-\psi)^2} - \frac{\varphi\psi}{(1-\varphi)^2}\right) = ... = 6$$

$$\mathbf{E}(T_2) = \sum_{k=0}^{\infty} k \frac{\varphi\psi^{k-1} - \psi\varphi^{k-1}}{2\varphi - 2\psi} = \frac{1}{2(\varphi - \psi)} \left(\frac{\varphi}{(1-\psi)^2} - \frac{\psi}{(1-\varphi)^2}\right) = ... = 4$$

On peut aussi utiliser les fonctions génératrices. Soit $g_i(z) = \sum_{k=0}^{\infty} P(T_i = k)\, z^k$. n étant l'état final, on a $g_n(z) = 1$. Par ailleurs, les relations $P(T_i = 0) = 0$, $P(T_i = 1) = p_{in}$ et $P(T_i = k) = \sum_{i=1}^{n-1} p_{ij}\, P(T_j = k - 1)$ pour $k \geq 2$ donne :

$$g_i(z) = p_{in}\, z + \sum_{i=1}^{n-1} p_{ij} \sum_{k=2}^{\infty} P(T_j = k - 1)\, z^k$$

$$= p_{in}\, z + \sum_{i=1}^{n-1} p_{ij} \sum_{k=1}^{\infty} P(T_j = k)\, z^{k+1}$$

$$= p_{in}\, z + \sum_{i=1}^{n-1} p_{ij}\, z g_j(z)$$

$$\Rightarrow \qquad g_i(z) = \sum_{i=1}^{n} p_{ij}\, z g_j(z)$$

On obtient ainsi un système d'inconnues les g_i, qui, une fois résolu, permettent, en théorie, d'en tirer les lois des T_i, ainsi que leur espérance et variance.

EXEMPLE :
Toujours avec l'exemple précédent, on a :

$$g_1 = \frac{1}{2} z g_1 + \frac{1}{2} z g_2$$

$$g_2 = \frac{1}{2} z g_1 + \frac{1}{2} z$$

d'où, en résolvant le système :

$$g_1 = \frac{z^2}{4 - 2z - z^2} = \frac{z^2}{4} + \frac{z^3}{8} + \frac{z^4}{8} + \frac{3z^5}{32} + \frac{5z^6}{64} + \ldots$$

$$g_2 = \frac{2z - z^2}{4 - 2z - z^2} = \frac{z}{2} + \frac{z^3}{8} + \frac{z^4}{16} + \frac{z^5}{16} + \frac{3z^6}{64} + \ldots$$

Le calcul du terme du général du développement en série entière demanderait de factoriser le dénominateur, de réduire la fraction obtenue en éléments simples et enfin d'en développer en série chacun des termes obtenus. Les fonctions génératrices sont plus efficaces pour calculer espérance et variance des temps d'arrêts. On trouve ainsi que :

$$\mathbf{E}(T_1) = g_1'(z) = 6 \qquad V(T_1) = g_1''(z) + g_1'(z) - g_1'(z)^2 = 22$$
$$\mathbf{E}(T_2) = g_2'(z) = 4 \qquad V(T_2) = g_2''(z) + g_2'(z) - g_2'(z)^2 = 20$$

On peut directement établir un système vérifié par les espérances en dérivant les relations $g_i(z) = \sum_{i=1}^{n} p_{ij}\, z g_j(z)$, ce qui donne :

$$g_i'(z) = \sum_{i=1}^{n} p_{ij}\, (g_j(z) + z g_j'(z)) \qquad \text{qu'on évalue en } z = 1$$

$$\Rightarrow \qquad \mathbf{E}(T_i) = \sum_{i=1}^{n} p_{ij}\, (1 + \mathbf{E}(T_j)) = 1 + \sum_{i=1}^{n} p_{ij}\, \mathbf{E}(T_j)$$

En moyenne, le temps d'arrêt de T_i est égal à 1 (correspondant à un saut de transition) + la moyenne des temps d'arrêt des états suivants, pondérés par la probabilité de passer de l'état i à l'état j.

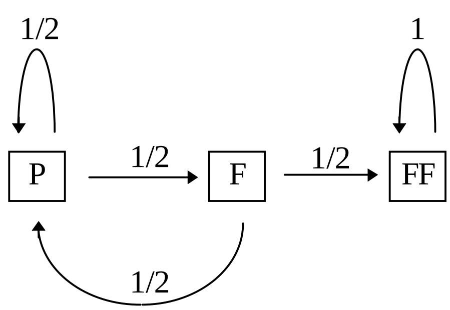
EXEMPLES :

Retrouvons les valeurs déjà trouvées pour atteindre FF : l'espérance du temps d'arrêt vaut 6.

Soient E_P = espérance du temps d'arrêt P $\xrightarrow{*}$ FF

E_F = espérance du temps d'arrêt F $\xrightarrow{*}$ FF

où $\xrightarrow{*}$ désigne un nombre indéterminé de changement d'états



$$\begin{cases} E_P = \dfrac{E_P + 1}{2} + \dfrac{E_F + 1}{2} \\[2ex] E_F = \dfrac{E_P + 1}{2} + \dfrac{1}{2} = \dfrac{E_P}{2} + 1 \end{cases} \qquad \text{donc } E_P = 6 \text{ et } E_F = 4.$$

3- Compétition entre deux états finaux

Considérons un système possédant deux états finaux. On cherche la probabilité d'atteindre le premier état plutôt que le second.

❑ On considère une pièce ayant une probabilité p de tomber sur P et $q = 1 - p$ de tomber sur F. Sur l'axe réel, on part de 0. A chaque P, on fait un saut vers la droite et à chaque F, on fait un saut vers la gauche. Soit n un entier donné. On arrête les lancers quand on a atteint n ou $-n$. On cherche la probabilité d'atteindre n et celle d'atteindre $-n$.

Soit $P(k)$ la probabilité d'atteindre n en partant de k. On a :

$$P(n) = 1$$
$$P(-n) = 0$$
$$\forall k \in \{-(n-1), -(n-2), \ldots, n-2, n-1\},\ P(k) = pP(k+1) + qP(k-1)$$

On obtient donc une suite récurrente linéaire d'ordre 2 dont l'équation caractérisque d'inconnue r est :

$$pr^2 - r + q = 0$$

et dont les racines sont $r = 1$ et $r = \frac{q}{p}$. Dans le cas où $q \neq p$, il existe λ et μ tel que, pour tout k :

$$P(k) = \lambda + \mu\left(\frac{q}{p}\right)^k$$

On trouve λ et μ avec les conditions :

$$\begin{cases} \lambda + \mu\left(\frac{q}{p}\right)^n = 1 \\ \lambda + \mu\left(\frac{q}{p}\right)^{-n} = 0 \end{cases}$$

donc $\mu = \dfrac{1}{\left(\frac{q}{p}\right)^n - \left(\frac{q}{p}\right)^{-n}} = \dfrac{\left(\frac{q}{p}\right)^n}{\left(\frac{q}{p}\right)^{2n} - 1} = \dfrac{p^n q^n}{q^{2n} - p^{2n}}$ et $\lambda = -\dfrac{p^{2n}}{q^{2n} - p^{2n}}$

La valeur de $P(0)$ est $\lambda + \mu = \dfrac{p^n}{q^n + p^n}$.

EXEMPLE :

D'après Huygens[2] : "Ayant pris chacun 12 jetons, A et B jouent avec trois dés à cette condition qu'à chaque coup de 11 points, A doit donner un jeton à B et que B en doit donner un à A à chaque coup de 14 points, et que celui là gagnera qui sera le premier en possession de tous les jetons. On trouve dans ce cas que la chance de A est à celle de B comme 244 140 625 est à 282 429 536 481".

La probabilité d'obtenir 11 avec trois dés est $\frac{27}{6^3}$ et celle d'obtenir 14 est de $\frac{15}{6^3}$. Comme les autres totaux n'interviennent pas, on peut considérer les probabilités conditionnelles sachant qu'on tire 11 ou 14, celles-ci valant respectivement $\frac{27}{27+15} = \frac{9}{14} = p$ et $\frac{15}{27+15} = \frac{5}{14} = q$.

B gagne avec la probabilité $\dfrac{p^{12}}{p^{12} + q^{12}} = \dfrac{9^{12}}{9^{12} + 5^{12}}$, de l'ordre de $\dfrac{999}{1000}$.

[2] *De ratiociniis in ludo aleae* (1657), cinquième problème, http://gallica.bnf.fr/ark:/12148/bpt6k77862v/f95.image

Annexe II : Ensembles dénombrables et non dénombrables

a) On pourrait penser qu'il n'y a que deux types d'ensembles, les ensembles finis et les ensembles infinis, ces derniers étant tous de même nature. Cette vision a été mise en défaut par Georg Cantor (1845-1918). Ses travaux sont à la base de la théorie des ensembles au XX$^{\text{ème}}$ siècle. Il définit plusieurs types d'infinis.

Un ensemble infini est en bijection avec l'une de ses parties strictes. Par exemple, $\mathbb{N}$ est en bijection avec $\mathbb{N}^*$, au moyen de la bijection suivante :

$$\begin{aligned} \mathbb{N} &\to \mathbb{N}^* \\ n &\to n+1 \end{aligned}$$

Soient plusieurs ensembles infinis, par exemple $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ et $\mathbb{R}$. Sont–ils en bijection les uns avec les autres ? On prouvera que $\mathbb{N}$, $\mathbb{Z}$ et $\mathbb{Q}$ sont effectivement en bijection, mais ce n'est pas le cas de $\mathbb{R}$. Les premiers sont dits dénombrables.

Galilée a bien remarqué que les termes "autant d'éléments", "moins d'éléments" ou "plus d'éléments" ne peuvent s'appliquer sans paradoxe aux ensembles infinis. Le terme *bijection* n'était pas encore inventé, mais Galilée a mis en évidence une bijection entre $\mathbb{N}$ et une partie stricte de $\mathbb{N}$:

$$\begin{matrix} 1 & 2 & 3 & 4 & \dots & n & \dots \\ 1 & 4 & 9 & 16 & \dots & n^2 & \dots \end{matrix}$$

b) Deux ensembles en bijection sont dits équipotents. S'ils sont finis, cela signifie simplement qu'ils ont le même nombre d'éléments. Soit E un ensemble quelconque, et $\mathcal{P}(E)$ l'ensemble de ses parties. Alors E et $\mathcal{P}(E)$ ne sont pas équipotents. Cela est évident si E est fini, à n éléments, puisqu'alors $\mathcal{P}(E)$ possède 2^n éléments, et pour tout n, $2^n > n$. Mais cette propriété reste vraie si E est infini. Il faut prouver qu'il ne peut exister de bijection f entre E et $\mathcal{P}(E)$. Raisonnons par l'absurde et supposons l'existence d'une telle bijection f :

$$\begin{aligned} f : E &\to \mathcal{P}(E) \\ x &\to f(x) \end{aligned}$$

A tout élément x de E, f associe $f(x)$, élément de $\mathcal{P}(E)$, autrement dit, $f(x)$ est une partie de E. Considérons maintenant la partie A de E définie de la façon suivante :

$$A = \{x \in E \mid x \notin f(x)\}.$$

Par définition de A, on a l'équivalence : $x \in A \Leftrightarrow x \notin f(x)$. Puisque f est supposée être une bijection de E sur $\mathcal{P}(E)$, et que A étant une partie de E est un élément de $\mathcal{P}(E)$, A possède un antécédent unique par f, a. On a donc $f(a) = A$. On se pose alors la question suivante : a-t-on $a \in f(a)$? Or :

$$\begin{aligned} a \in f(a) &\Leftrightarrow a \in A \text{ car } f(a) = A \\ &\Leftrightarrow a \notin f(a) \text{ par définition de l'appartenance à A} \end{aligned}$$

Ainsi la proposition $a \in f(a)$ est équivalente à sa négation. La contradiction ne peut être levée qu'en rejetant l'hypothèse de l'existence de f.

Cette démonstration assure l'existence d'ensembles non dénombrables, c'est-à-dire qui ne sont pas en bijection avec $\mathbb{N}$, par exemple $\mathcal{P}(\mathbb{N})$. On conçoit même une hiérarchie infinie d'espaces $\mathbb{N}$, $\mathcal{P}(\mathbb{N})$, $\mathcal{P}(\mathcal{P}(\mathbb{N}))$, ...

c) $\mathbb{N}$ est le plus petit ensemble infini. Si E est un ensemble quelconque, alors ou bien E est fini, ou bien il est dénombrable (en bijection avec $\mathbb{N}$), ou bien il existe une injection de $\mathbb{N}$ dans E mais pas de bijection (exemples: E = $\mathcal{P}(\mathbb{N})$ ou E = $\mathbb{R}$). Un ensemble dénombrable, étant en bijection avec $\mathbb{N}$, peut s'écrire sous la forme $\{x_n \mid n \in \mathbb{N}\}$; la bijection est l'application $f : \mathbb{N} \to E$, $n \to x_n$. Un ensemble dénombrable se reconnaît à ce qu'on peut énumérer ses éléments. On a vu que $\mathbb{Z}$ et $\mathbb{Q}$ étaient dénombrables.

d) $\mathbb{R}$ n'est pas dénombrable. S'il l'était, il en serait de même de [0,1[. Considérons alors une énumération $(x_n)_{n\in\mathbf{N}^*}$ de [0,1[, obtenue au moyen d'une bijection $f : \mathbb{N}^* \to [0,1[$, $n \to x_n$, et considérons le développement décimal des x_n.

$x_1 = 0,a_{11}a_{12}a_{13}...a_{1p}...$
$x_2 = 0,a_{21}a_{22}a_{23}...a_{2p}...$
...
$x_n = 0,a_{n1}a_{n2}a_{n3}...a_{np}...$
...

a_{np} est le $p^{\text{ème}}$ chiffre de la décomposition décimale de x_n. C'est un élément de {0,1,...,9}.
Considérons maintenant l'élément y de]0,1[défini de la façon suivante :

$y = 0,b_1b_2b_3...b_p...$

où $b_p = 0$ si $a_{pp} \neq 0$ et $b_p = 1$ si $a_{pp} = 0$.
On obtient le développement décimal d'un réel distinct de tous les x_n. En effet, le n^{eme} chiffre de x_n et y sont différents ($\forall\ n$, $b_n \neq a_{nn}$). Par ailleurs, il est évident que y appartient à [0,1[. Cela est contradictoire avec le fait que f soit bijective, puisqu'alors, tout élément de [0,1[serait de la forme d'un des x_n. Cette démonstration est connue sous le nom de diagonalisation de Cantor.

On peut prouver que $\mathbb{R}$ est équipotent à $\mathcal{P}(\mathbb{N})$, et que les trois ensembles suivants sont équipotents : $\mathcal{P}(\mathbf{R})$, $\mathcal{P}(\mathcal{P}(\mathbb{N}))$ et $C^0(\mathbf{R})$ ensemble des fonctions continues sur $\mathbb{R}$.

e) Signalons également une question étonnante. Peut–on trouver un ensemble E compris entre $\mathbb{N}$ et $\mathbb{R}$, mais qui ne soit équipotent ni à $\mathbb{N}$, ni à $\mathbb{R}$? On aurait seulement des injections de $\mathbb{N}$ dans E et de E dans $\mathbb{R}$. Rappelons que $\mathbb{Q}$ ne répond pas à la question puisqu'il est en bijection avec $\mathbb{N}$. On peut prouver qu'il était *impossible* de répondre à cette question. Cela ne signifie pas qu'on n'ait pas encore trouvé si cette propriété était vraie ou fausse, mais bel et bien qu'on ne peut ni prouver qu'elle est vraie, ni prouver qu'elle est fausse. Elle est dite indécidable. Elle ne découle pas des axiomes de la théorie des ensembles, pas plus que sa négation. Cela signifie également qu'on peut prendre comme axiome supplémentaire l'existence d'un tel ensemble E sans apporter de contradiction à l'édifice des Mathématiques, ou au contraire, de prendre comme axiome la non-existence de E. Dans ce dernier cas, on adopte ce qu'on appelle l'hypothèse du continu. L'un ou l'autre choix conduit donc à deux théories mathématiques différentes.

Ces considérations n'ont aucune importance en ce qui nous concerne, car nous n'utiliserons jamais cette propriété, ni sa négation !

f) Donnons enfin une conséquence curieuse de ce qui précède en informatique. On peut montrer que l'ensemble de tous les algorithmes possibles est dénombrable, alors que l'ensemble des fonctions de

$\mathbb{N}$ dans $\mathbb{N}$ est équipotent à $\mathbb{R}$. Il y a donc des fonctions de $\mathbb{N}$ dans $\mathbb{N}$ qui ne sont calculables par aucun ordinateur. Aucun algorithme ne permet de les calculer. De telles fonctions ont été explicitement définies.

◆