Probabilités - Résumé

1. Cadre théorique de la théorie des probabilités

1.1. **L'axiomatique de Kolmogorov.** Cette axiomatique donne le cadre dans lequel s'inscrit la théorie des probabilités depuis la parution des *Grundbegriffe der Wahrscheinlichkeitsrechnung* (principes de base du calcul des probabilités) en 1933. Ce cadre s'inscrit plus largement dans la *théorie de la mesure* dont est issue la théorie générale de l'intégrale de Lebesgue (1904). À partir du niveau L3, il est fréquent qu'un cours unique traite des théories de l'intégration et des probabilités dans le cadre de celle de la mesure. Cette théorie est hors programme en CPGE.

On n'en conserve ici que quelques définitions, ce qui a pour conséquence que, si Ω est infini non dénombrable, on ne décrit pas l'univers Ω (sauf dans dans le cas d'école du pile ou face), ni la tribu $\mathcal{A}$ des événements (même dans le cadre du pile ou face). Si Ω est au plus dénombrable, on prend $\mathcal{A} = \mathscr{P}(\Omega)$. De plus, dans le cadre du programme, on se restreint aux variables aléatoires discrètes, c'est-dire à valeurs dans un ensemble au plus dénombrable (cette restriction exclut des lois parmi les plus courantes : lois exponentielle, uniforme sur un intervalle et normales).

1.2. **Définitions.**

Définition 1. *Si Ω est un ensemble non vide, une partie $\mathcal{A} \subset \mathscr{P}(\Omega)$ est une* tribu *(ou σ-*algèbre*) sur Ω si :*

(i) $\Omega \in \mathcal{A}$ *;*

(ii) $\forall A \in \mathcal{A} \colon \overline{A} \in \mathcal{A}$ *;*

(iii) $\forall (A_n)_n \in \mathcal{A}^{\mathbb{N}} \colon \bigcup\limits_{n \geqslant 0} A_n \in \mathcal{A}$.

Un espace probabilisable est un couple $(\Omega, \mathcal{A})$, où Ω est un ensemble non vide, $\mathcal{A}$ une tribu sur Ω.

Définition 2. *Si Ω est un ensemble non vide et $\mathcal{A}$ une tribu sur Ω, on appelle (mesure de) probabilité sur $(\Omega, \mathcal{A})$ une application $\mathbf{P} \colon \mathcal{A} \to [0,1]$ telle que :*

(i) $\mathbf{P}(\Omega) = 1$ *;*

(ii) Pour toute suite d'événements $(A_n)_n$ deux à deux incompatibles (i.e. $A_i \cap A_j = \varnothing$), la série de terme général $\mathbf{P}(A_n)$ converge et $\mathbf{P}\left(\biguplus\limits_{n=0}^{\infty} A_n\right) = \sum\limits_{n=0}^{\infty} \mathbf{P}(A_n)$ *(*additivité dénombrable*).*

Un espace probabilisé est un triplet $(\Omega, \mathcal{A}, \mathbf{P})$, où Ω est un ensemble non vide, $\mathcal{A}$ une tribu sur Ω et $\mathbf{P}$ une probabilité sur $(\Omega, \mathcal{A})$.

Notations : $\overline{A} = \Omega \setminus A$ est l'événement contraire de A. Le signe $\biguplus$ désigne une réunion disjointe et exprime l'incompatibilité ; on rencontre aussi $\bigsqcup$. On peut aussi utiliser $\bigcap\!\!\!\downarrow A_n$ pour une intersection décroissante (*i.e.* $A_n \supset A_{n+1}$) et $\bigcup\!\!\!\uparrow A_n$ pour une réunion croissante (*i.e.* $A_n \subset A_{n+1}$). Notons qu'une intersection croissante ou une union décroissante sont sans intérêt car elles sont égales à A_0.

1.3. **Opérations sur les événements et premières propriétés.** Comme en algèbre linéaire (définitions d'un espace vectoriel et d'une application linéaire), si les définitions sont à connaître, il est encore plus important de savoir quelles sont les opérations permises, de connaître leur signification et leurs propriétés.

Opérations sur les événements. Passage au complémentaire, réunion et intersection finies ou dénombrables dans un premier temps. Dans un deuxième temps, toute combinaison finie de ces opérations. Par exemple, $A \setminus B = A \cap \overline{B}$ et la *différence symétrique* $A \bigtriangleup B = (A \cup B) \setminus (A \cap B) = (A \setminus B) \uplus (B \setminus A)$. On rappelle les lois de de Morgan :

$$\overline{\bigcup_{i \in I} A_i} = \bigcap_{i \in I} \overline{A_i} \quad \& \quad \overline{\bigcap_{i \in I} A_i} = \bigcup_{i \in I} \overline{A_i}.$$

Il est bon de savoir décoder quelques cas particuliers usuels, comme

$$A_* = \bigcup_{p \geqslant 0} \bigcap_{n \geqslant p} A_n \quad \text{et} \quad A^* = \bigcap_{p \geqslant 0} \bigcup_{n \geqslant p} A_n = A \text{ i.s.}$$

A_* est l'événement « tous les A_n à partir d'un certain rang » et A^* est l'événement « A_n pour une infinité de valeurs de n » ou, ce qui est équivalent, « A_n pour des valeurs de n arbitrairement grandes ». On l'appelle A *infiniment souvent*, d'où la notation A i.s.

Proposition 1. *Propriétés du calcul sur les événements :*

(i) $\mathbf{P}(\varnothing) = 0$ *;*

(ii) $\mathbf{P}(A_1 \uplus A_1 \uplus \cdots \uplus A_n) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \cdots + \mathbf{P}(A_n)$ *(*additivité finie*) ;*

(iii) $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$ *;*

(iv) $A \subset B \Longrightarrow \mathbf{P}(A) \leqslant \mathbf{P}(B)$ *(*croissance*) ;*

(v) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$ *;*

(vi) $\mathbf{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leqslant i_1 < i_1 < \cdots < i_k \leqslant n} \mathbf{P}\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right)$ *(*formule du crible*, HP) ;*

(vii) $\mathbf{P}\left(\bigcap\!\!\!\downarrow A_n\right) = \lim_{n \to \infty} \mathbf{P}(A_n)$ *(*continuité décroissante*) ;*

(viii) $\mathbf{P}\left(\bigcup\!\!\!\uparrow A_n\right) = \lim_{n \to \infty} \mathbf{P}(A_n)$ *(*continuité croissante*) ;*

(ix) $\mathbf{P}\left(\bigcup_{n=0}^{N} A_n\right) \leqslant \sum_{n=0}^{N} \mathbf{P}(A_n)$ *(*sous-additivité finie*) ;*

(x) $\mathbf{P}\left(\bigcup_{n=0}^{\infty} A_n\right) \leqslant \lim_{N \to +\infty} \sum_{n=0}^{N} \mathbf{P}(A_n) = \sum_{n=0}^{\infty} \mathbf{P}(A_n)$ *(sous-additivité* dénombrable*) ;*

(xi) $\left|\mathbf{P}(B) - \mathbf{P}(A)\right| \leqslant \max\left(\mathbf{P}(A \cap \overline{B}), \mathbf{P}(B \cap \overline{A})\right) \leqslant \mathbf{P}(A \bigtriangleup B)$*.*

L'inégalité *(x)* s'écrit ainsi plus simplement $\mathbf{P}\left(\bigcup_{n=0}^{\infty} A_n\right) \leqslant \sum_{n=0}^{\infty} \mathbf{P}(A_n)$, étant entendu qu'en cas de divergence de la série, on peut lui attribuer la valeur $+\infty$, ce qui, pour majorer une probabilité, est de toute façon peu informatif.

Astuce utile. Toute réunion au plus dénombrable peut s'écrire de manière croissante et de manière disjointe et toute intersection peut s'écrire de manière décroissante. Ainsi,

$$\bigcup_{n \geqslant 0} A_n = \biguplus_{n \geqslant 0} B_n = \bigcup\!\!\!\uparrow_{n \geqslant 0} C_n, \text{ avec } B_0 = A_0,\ B_n \underset{(n \geqslant 1)}{=} A_n \setminus (A_0 \cup A_1 \cup \cdots \cup A_{n-1}) \text{ et } C_n = A_0 \cup A_1 \cup \cdots \cup A_n;$$

$$\bigcap_{n=0}^{\infty} A_n = \bigcap\!\!\!\downarrow_{n \geqslant 0} B_n, \text{ où } B_n = A_0 \cap A_1 \cap \cdots \cap A_n.$$

Ainsi, pour $C_n = B_0 \cup B_1 \cup \cdots \cup B_n$,

$$\biguplus_{n=0}^{N} B_n = \bigcup\!\!\!\uparrow_{n=0}^{N} C_n \quad \therefore \quad \mathbf{P}\left(\biguplus_{n=0}^{N} B_n\right) = \sum_{n=0}^{N} \mathbf{P}(B_n) = \mathbf{P}(C_N) \quad \therefore$$

$$\mathbf{P}\left(\biguplus_{n \geqslant 0} B_n\right) = \lim_{N \to \infty} \sum_{n=0}^{N} \mathbf{P}(B_n) = \lim_{N \to \infty} \mathbf{P}(C_n).$$

Cela montre la convergence de la série de terme général $\mathbf{P}(B_n)$ et le fait que, si $(B_n)_n$ est une suite d'événements deux à deux incompatibles, alors $\lim \mathbf{P}(B_n) = 0$.

En général, la probabilité d'une intersection ou d'une réunion n'est pas la limite des probabilités des événements considérés. Toutefois, si $\lim \mathbf{P}(A_n) = 0$, alors $\mathbf{P}\left(\bigcap_{n=0}^{\infty} A_n\right) = 0$ car, pour tout $p \in \mathbb{N}$, $\bigcap_{n=0}^{\infty} A_n \subset A_p$, d'où $\mathbf{P}\left(\bigcap_{n=0}^{\infty} A_n\right) \leqslant \lim_{p \to \infty} \mathbf{P}(A_p) = 0$. De même, si $\lim \mathbf{P}(A_n) = 1$, alors $\mathbf{P}\left(\bigcup_{n=0}^{\infty} A_n\right) = 1$. Cela n'est pas à savoir par cœur, mais l'on peut à l'occasion refaire le raisonnement pour se passer du passage à une suite monotone.

1.4. **Axiomatique des variables aléatoires.**

Définition 3. *Une variable aléatoire discrète (v.a.d.) est une fonction définie sur un espace probabilisé* $(\Omega, \mathcal{A}, \mathbf{P})$ *et à valeurs dans un ensemble* E *telle que* $X(\Omega)$ *soit fini ou dénombrable et telle que* $(X = x) = X^{-1}(\{x\}) \in \mathcal{A}$ *pour tout* $x \in X(\Omega)$.

Théorème et définition 1. *Si* X *est une v.a.d., on définit un espace probabilisé* $\Big(X(\Omega), \mathscr{P}\big(X(\Omega)\big), P_X\Big)$ *par*

$$\forall A \in \mathscr{P}\big(X(\Omega)\big) \colon\ \mathbf{P}_X(A) = \mathbf{P}(X \in A) = \mathbf{P}\left(X^{-1}(A)\right) = \sum_{a \in A} \mathbf{P}(X = a).$$

La mesure de probabilité $\mathbf{P}_X$ *est appelée la* loi de X.

Tout cela est plutôt abstrait... Retenir avant tout que la loi de X est entièrement donnée et décrite par les valeurs $\mathbf{P}(X = x)$ pour $x \in X(\Omega)$, un peu comme une application linéaire est entièrement décrite par l'image d'une base. Cela est spécifique aux v.a. discrètes. Comme $X(\Omega)$ est au plus dénombrable, on peut le décrire en extension sous la forme $X(\Omega) = \{x_0, x_1, \ldots, x_n\}$ ou $X(\Omega) = \{x_0, x_1, \ldots, x_n, \ldots\}$. De plus, pour $A \subset X(\Omega)$, $\mathbf{P}(A) = \sum_{a \in A} \mathbf{P}(X = a)$ ne dépend pas de la numérotation de A.

Quand un énoncé demande de déterminer la loi d'une v.a. X, il est attendu, dans l'ordre, de déterminer $X(\Omega)$, de calculer $\mathbf{P}(X = x)$ pour tout $x \in X(\Omega)$ et, enfin, de reconnaître la loi si X suit une loi au programme. Il est parfois possible de court-circuiter les calculs en reconnaissant directement une loi usuelle.

Exemple 1. La plupart des v.a. rencontrées dans les énoncés sont à valeurs dans $\mathbb{N}$. Si $A = (X \text{ est paire})$, alors,

$$\mathbf{P}(A) = \mathbf{P}(X \in 2\mathbb{N}) = \sum_{n \in 2\mathbb{N}} \mathbf{P}(X = n) = \sum_{k=0}^{\infty} \mathbf{P}(X = 2k).$$

Remarque 1. On se contente, sauf exception, des $\big\{x \in X(\Omega)\,;\, \mathbf{P}(X = x) > 0\big\}$ à la place de $X(\Omega)$ en écartant donc les valeurs prises par X avec une probabilité nulle. Un cas typique est l'indice d'apparition du premier *face* lors d'une série de lancers de *pile/face*, cas fondateur de la loi géométrique. On peut définir une valeur bidon relative à l'événement presque impossible « *face* n'est jamais apparu », comme $X = 0$, ou $X = +\infty$ (valeur plus naturelle, mais qui n'est pas entière) ; cela n'a pas d'importance et l'on peut considérer sans inconvénient que $X(\Omega) = \mathbb{N}^*$. Vérifier que X est à valeurs entières, c'est donc vérifier que $\mathbf{P}(X \in \mathbb{N}) = 1$. En L3 et au-delà, on travaille dans des espaces quotients construits pour négliger rigoureusement les événements presque impossibles ; cela n'est pas du tout au programme, mais on peut garder l'idée que l'on ne se laisse pas compliquer la vie par ce genre d'événements.

Propriétés. L'ensemble des variables aléatoires réelles forme un espace vectoriel. Plus généralement, si $(X_i)_{1 \leqslant i \leqslant n}$ est une suite finie de variables aléatoires et si $f(X_1, X_2, \ldots, X_n)$ est bien définie, alors c'est une variable aléatoire. À titre d'exemple, montrons qu'une somme de deux v.a. à valeurs dans $\mathbb{N}$ est une v.a. Comme $\mathbb{N}$ est stable par addition, $X_1 + X_2$ est bien à valeurs dans $\mathbb{N}$. Pour vérifier que $X_1 + X_2$ est une v.a., il faut donc montrer que, pour tout $n \in \mathbb{N}$, $(X_1 + X_2 = n)$ est un événement, ce que prouve la décomposition $(X_1 + X_2 = n) = \bigcup_{k=0}^{n} (X_1 = k) \cap (X_2 = n - k)$, puisque, par hypothèse, $(X_1 = k)$ et $(X_2 = n - k)$ en sont.

Définition 4. *Si* X *est une v.a.d. réelle, c'est-à-dire prenant ses valeurs dans une partie au plus dénombrable de* $\mathbb{R}$, *on appelle* fonction de répartition *de* X *la fonction* $F_X \colon \mathbb{R} \to [0, 1]$ *définie par* $F_X(x) = \mathbf{P}(X \leqslant x)$.

Propriétés. F_X est croissante, continue à droite, $\lim_{x \to -\infty} F_X(x) = 0$ et $\lim_{x \to +\infty} F_X(x) = 1$: pour $x \leqslant y$, deux réels, on a $(X \leqslant x) \subset (X \leqslant y)$, d'où, par croissance des mesures de probabilité, $F_X(x) \leqslant F_X(y)$. Cela entraîne l'existence de $\lim_{x \to \pm\infty} F_X(x)$ et l'on a

$$\lim_{x \to +\infty} F_X(x) = \lim_{\substack{n \to \infty \\ n \in \mathbb{N}}} F_X(n) = \lim_{n \to +\infty} \mathbf{P}(X \leqslant n) = \mathbf{P}\left(\biguplus_{n \geqslant 0} (X \leqslant n)\right) = \mathbf{P}(X \in \mathbb{R}) = 1.$$

La limite quand x tend vers $-\infty$ se calcule de manière similaire. Pour $h > 0$,

$$F_X(x + h) - F_X(x) = \mathbf{P}(x < X \leqslant x + h) \qquad \& \qquad F_X(x) - F - X(x - h) = \mathbf{P}(x - h < X \leqslant x).$$

En prenant $x = n^{-1}$, le théorème de continuité décroissante montre que F_X est continue à droite en tout point et continue à gauche en x si, et seulement si, $\mathbf{P}(X = x) = 0$. Autrement dit, les points de discontinuité de F_X sont les éléments de $X(\Omega)$.

2. Probabilités conditionnelles et indépendance

2.1. Définition.

Théorème et définition 2. *Soient* $(\Omega, \mathcal{A}, \mathbf{P})$ *un espace probabilisé et* $B \in \mathcal{A}$ *un événement tel que* $\mathbf{P}(B) > 0$. *Alors, on définit un nouvel espace probabilisé* $(\Omega, \mathcal{A}, \mathbf{P}_B)$ *par* $\mathbf{P}_B(A) = \dfrac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$.

L'autre notation usuelle pour les probabilités conditionnelles est $\mathbf{P}(A \,|\, B) = \mathbf{P}_B(A)$. La première se lit « probabilité de A sachant B ». Il faut bien comprendre que le conditionnement est une modification de la **mesure de la probabilité** et non des événements eux-mêmes. En particulier, $A \,|\, B$ n'est pas un événement. On peut comprendre le calcul des probabilités comme une évaluation des chances qu'adviennent les issues possibles d'une expérience en présence d'une information partielle. Si l'on rajoute de l'information, la probabilité change. Or calculer une probabilité, c'est intrinsèquement le faire d'un certain point de vue, eu égard aux connaissances à disposition.

Définition 5. *Deux événements* A *et* B *sont indépendants si* $\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B)$. *On note alors* $A \perp\!\!\!\perp B$.

En particulier, si B n'est pas presque impossible, A et B sont indépendants si, et seulement si, $\mathbf{P}_B(A) = \mathbf{P}(A)$. Par symétrie, on a la même équivalence en intervertissant A et B, soit $\mathbf{P}_A(B) = \mathbf{P}(B)$ si $\mathbf{P}(A) \neq 0$. L'interprétation est naturelle : la connaissance de B n'influe pas sur la probabilité de A. Le calcul et l'intuition annoncent de même que A et B sont indépendants ssi $\overline{A}$ et B le sont, ssi $\overline{A}$ et $\overline{B}$ le sont.

Des conditionnements itérés reviennent à conditionner par une intersection. On vérifie facilement que si $B \cap C$ n'est pas presque impossible, alors

$$\mathbf{P}_B(A \,|\, C) = \mathbf{P}_C(A \,|\, B) = \mathbf{P}(A \,|\, B \cap C).$$

Définition 6. *Les événements d'une famille* $(A_i)_{i \in I}$ *sont* indépendants *(on dit aussi « mutuellement indépendants ») si, pour toute partie* $J \subset I$ *finie, on a*

$$\mathbf{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbf{P}(A_j).$$

Il est évident que si des événements sont indépendants, ils le sont deux à deux. La réciproque est fausse. La définition avec les parties finies vaut aussi si I est fini. Ainsi, pour que les événements $A_1, A_2, ..., A_n$ soient indépendants, il ne suffit pas que $\mathbf{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbf{P}(A_1)\,\mathbf{P}(A_2) \cdots \mathbf{P}(A_n)$. En revanche, $A_1, A_2, ..., A_n$ sont indépendants si, et seulement si,

$$\forall (B_1, \ldots, B_n) \in \prod_{i=1}^{n} \left\{A_i, \overline{A_i}\right\} \colon\ \mathbf{P}(B_1 \cap B_2 \cap \cdots \cap B_n) = \mathbf{P}(B_1)\,\mathbf{P}(B_2) \cdots \mathbf{P}(B_n). \tag{1}$$

2.2. Couples de variables aléatoires.

Définition 7. *Si* X *et* Y *sont deux variables aléatoires discrètes définies sur un même espace probabilisé* $(\Omega, \mathcal{A}, \mathbf{P})$, *le couple* $(X, Y) \colon \omega \mapsto (X(\omega), Y(\omega))$ *définit une v.a.d. à valeurs dans l'ensemble au plus dénombrable* $X(\Omega) \times Y(\Omega)$. *Les lois de* X *et de* Y *sont dites* lois marginales *du couple et celle du couple est dite* loi conjointe. *Les lois de* X *sachant* $(Y = y)$ *ou de* X *sachant* $(X = x)$ *sont dites* lois conditionnelles.

Notation : on note généralement $\mathbf{P}(X = x, Y = y)$ pour $\mathbf{P}\big((X = x) \cap (Y = y)\big)$.
La loi conjointe détient l'ensemble de l'information relative au couple. On peut en déduire les lois marginales et conditionnelles :

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

$$\mathbf{P}_{(X=x)}(Y = y) = \frac{\mathbf{P}(X = x, Y = y)}{\mathbf{P}(X = x)} \quad \& \quad \mathbf{P}_{(Y=y)}(X = x) = \frac{\mathbf{P}(X = x, Y = y)}{\mathbf{P}(Y = y)},$$

la donnée des lois conditionnelles étant bien sûr valables si les événements par rapport auxquels on conditionne — ici, $(X = x)$ et $(Y = y)$ — ne sont pas négligeables, ce que l'on a supposé par convention. On peut, plus généralement, définir la loi conditionnelle de X sachant tout événement A de probabilité non nulle par $x \mapsto \mathbf{P}_A(X = x)$. Si (X, Y) est un couple de v.a.d., on peut retrouver la loi de conjointe du couple à partir de la loi marginale de X et des lois conditionnelles de Y sachant $(X = x)$.

Remarque 2. Il existe des théorèmes qui garantissent l'existence d'un espace probabilisé approprié permettant de considérer des suites de variables aléatoires vérifiant telle ou telle hypothèse (suite de v.a.i.i.d, chaîne de Markov, etc). Ces théorèmes sont entièrement hors programme, jusque dans leur énoncé. On ne se posera donc pas, en pratique (sauf si la question est posée, évidemment), la question de savoir ce qu'est Ω, ni $\mathcal{A}$, ni même s'ils existent. Les variables aléatoires sont des fonctions, mais on ne raisonne pas du tout sur elles comme sur des fonctions de la variable réelle en analyse.

2.3. **Variables aléatoires indépendantes.**

Définition 8. *1. Deux v.a.d. X et Y définies sur un même espace probabilisé sont indépendantes si*

$$\forall (x,y) \in X(\Omega) \times Y(\Omega) \colon\ \mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x)\,\mathbf{P}(Y = y),$$

i.e., si les événements $(X = x)$ et $(Y = y)$ sont indépendants pour tout couple de valeurs prises par X et par Y. De manière équivalente, pour tout $x \in X(\Omega)$, les v.a. Y et $Y_{|(X=x)}$ ont même loi. On note alors $X \perp\!\!\!\perp Y$.

2. Les v.a. d'une famille finie $(X_1, X_2, \ldots, X_n)$ définies sur un même espace probabilisé sont indépendantes si

$$\forall (x_i)_{1 \leqslant i \leqslant n} \in \prod_{i=1}^{n} X_i(\Omega) \colon\ \mathbf{P}\left(\bigcap_{i=1}^{n} (X_i = x_i)\right) = \prod_{i=1}^{n} \mathbf{P}(X_i = x_i).$$

3. Les v.a. d'une famille quelconque de v.a.d. définies sur un même espace probabilisé sont indépendantes si toute sous-famille finie est constituée de v.a. indépendantes.

Proposition 2. *Soit une famille finie $(X_1, X_2, \ldots, X_n)$ de v.a.d. définies sur un même espace probabilisé. Les propositions suivantes sont équivalentes :*

(i) Les v.a. $X_1, X_2, \ldots, X_n$ sont indépendantes.

(ii) $\forall (A_i)_{1 \leqslant i \leqslant n} \in \prod_{i=1}^{n} \mathscr{P}\big(X_i(\Omega)\big) \colon\ \mathbf{P}\left(\bigcap_{i=1}^{n} (X_i \in A_i)\right) = \prod_{i=1}^{n} \mathbf{P}(X_i \in A_i).$

Remarque 3. — La définition n'est pas exactement analogue à celle des familles finies d'événements indépendants, pour lesquelles on considère des sous-familles quelconques, même dans le cas d'une famille finie, alors que ce n'est pas nécessaire pour les v.a.d. Cela est dû au fait que l'on peut expulser des i du produit en prenant $A_i = X_i(\Omega)$.
— Quand deux v.a. sont indépendantes, la loi conjointe est donnée par le produit des lois marginales et les lois conditionnelles se confondent avec les lois marginales (c'est une reformulation de la définition).

Proposition 3 (Lemme des coalitions)**.** *Si les v.a. $(X_i)_{i \in \mathbb{N}}$ sont indépendantes dans leur ensemble, et si J et K sont deux parties finies de I telles que $J \cap K = \varnothing$, si f et g sont des fonctions donnant un sens à $U = f(X_j, j \in J)$ et $V = g(X_k, k \in K)$, alors U et V sont indépendantes.*

À nouveau, la formulation est abstraite, mais facile à comprendre : si X_1, X_2, X_3, X_4, X_5 sont mutuellement indépendantes, alors $U = X_1 X_3$ et $V = X_2 + X_4 - 2X_5$ sont indépendantes. C'est évident intuitivement. Noter que le lemme des coalitions est déjà intéressant quand J et K sont des singletons : si X et Y sont indépendantes, alors $f(X)$ et $g(Y)$ le sont aussi.

Exemple 2. Soient X et Y indépendantes suivant une loi uniforme $\mathcal{U}\big(\{-1, 1\}\big)$ et $Z = XY$. Alors, X, Y, Z ne sont pas indépendantes dans leur ensemble (considérer $(X = 1, Y = 1, Z = -1)$), mais indépendantes deux à deux car $\mathbf{P}(X = 1, Z = 1) = \mathbf{P}(X = 1, Y = 1) = \mathbf{P}(X = 1)\,\mathbf{P}(Z = 1) = 1/4$. En effet, on a ainsi montré que $(X = 1)$ et $(Z = 1)$ étaient indépendants, donc, par complémentation, il en va de même de $(X = 1)$ et $(Z = -1)$, de $(X = -1)$ et $(Z = 1)$, et de $(X = -1)$ et $(Z = -1)$. Ainsi, X et Z sont indépendantes, donc Y et Z également par symétrie.

Exemple 3 (à connaître)**.** Soient $(X_i)_{1 \leqslant i \leqslant n}$ une famille de v.a.i.i.d. à valeurs dans $\mathbb{N}$, $U = \min(X_1, X_2, \ldots, X_n)$ et $V = \max(X_1, X_2, \ldots, X_n)$. Déterminer la loi de U et celle de V en fonction de celle de X_1.

Les événements les plus simples associés sont $(U \geqslant k) = \bigcap_{i=1}^{n} (X_i \geqslant k)$ et $(V \leqslant k) = \bigcap_{i=1}^{n} (X_i \leqslant k)$, d'où les probabilités $\mathbf{P}(U \geqslant k) = \mathbf{P}(X_1 \geqslant k)^n$ et $\mathbf{P}(V \leqslant k) = \mathbf{P}(X_1 \leqslant k)^n$, dont on déduit facilement

$$\mathbf{P}(U = k) = \mathbf{P}(X_1 \geqslant k)^n - \mathbf{P}(X_1 \geqslant k+1)^n \text{ et } \mathbf{P}(V = k) = \mathbf{P}(X_1 \leqslant k)^n - \mathbf{P}(X_1 \leqslant k-1)^n.$$

3. Compléments sur les événements

3.1. **Presque.** Dans le cadre des univers finis, un événement de probabilité nulle est impossible. Il n'en va pas de même quand l'univers est infini.

Définition 9. *Un événement est dit* presque sûr *s'il est de probabilité 1 et* négligable *s'il est de probabilité nulle (on dit aussi* presque impossible*).*

Propriétés Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace probabilisé. Les événements ci-dessous appartiennent tous à $\mathcal{A}$.

— On suppose que $A \subset B$. Si A est presque sûr, alors B est presque sûr. Si B est presque impossible, alors A est presque impossible ;

— si A est presque sûr, alors, pour tout événement B, $\mathbf{P}(B) = \mathbf{P}(B \cap A)$ ou, ce qui revient au même, $\mathbf{P}(B \mid A) = \mathbf{P}(B)$;

— une réunion au plus dénombrable d'événements presque impossible est presque impossible. Une intersection au plus dénombrable d'événements presque sûrs est presque sûre.

Exemple 4. On joue à pile ou face sans s'arrêter. Si A_n est l'événement « les n premiers lancers ont amené pile » et A est « on n'a jamais fait face », alors $A = \bigcap\!\!\downarrow A_n$, donc $\mathbf{P}(A) = \lim \mathbf{P}(A_n) = \lim \left(\frac{1}{2}\right)^n = 0$. Ainsi, A est presque impossible.

Exemple 5. On reprend le jeu du pile ou face. Soit l'événement « on n'a jamais fait p fois pile consécutivement ». On note A cet événement. On note F_n l'événement « le n^{e} tirage a donné face » et, pour $m \in \mathbb{N}$, $\Phi_m = \bigcup_{k=pm+1}^{p(m+1)} F_k$. Ainsi, Φ_m exprime que les p tirages consécutifs d'indices $pm+1$ à $pm+p$ n'ont pas tous donné pile et l'on a

$$A \subset \bigcap_{m \geqslant 0} \Phi_m = \bigcap\!\!\downarrow_{n \geqslant 0} \left(\Phi_0 \cap \Phi_1 \cap \cdots \cap \Phi_n\right) \qquad \therefore$$

$$\mathbf{P}(A) \leqslant \mathbf{P}\left(\bigcap_{m \geqslant 0} \Phi_m\right) = \lim_{n \to \infty} \mathbf{P}\left(\Phi_0 \cap \Phi_1 \cap \cdots \cap \Phi_n\right) = \lim_{n \to \infty} \left(1 - \frac{1}{2^p}\right)^{n+1} \xrightarrow[m \to \infty]{} 0.$$

Le fait d'avoir introduit les événements Φ_m permet de travailler sur des événements indépendants car liés à des tirages distincts. Plus précisément, les événements Φ_m sont indépendants et de même probabilité, obtenue en passant à l'événement complémentaire, soit

$$\overline{\Phi_m} = \bigcap_{k=pm+1}^{p(m+1)} \overline{F_k} \qquad \therefore \qquad \mathbf{P}(\Phi_m) = 1 - \left(\frac{1}{2}\right)^p.$$

Exemple 6. On reste dans le pile ou face. Soient F_n l'événement « le n-ème lancer a amené face » et B l'événement « la suite de lancers a vu des séries arbitrairement longues de pile et de face alternés consécutivement ». Alors, $B = \bigcap\!\!\downarrow B_\ell$, avec B_ℓ l'événement « la suite de lancers a vu au moins une série de 2ℓ lancers consécutifs ayant alterné pile et face et commençant par face » (une telle série commençant par pile donne une série de longueur $2(\ell-1)$ commençant par face). On peut écrire

$$B_\ell = \bigcup_{n \geqslant 1} \left[\left(\bigcap_{i=0}^{\ell-1} F_{n+2i}\right) \bigcap \left(\bigcap_{i=0}^{\ell-1} \overline{F_{n+2i+1}}\right)\right] = \bigcup_{n \geqslant 1} C_{n,\ell}.$$

L'indépendance des lancers donne $\mathbf{P}(C_{n,\ell}) = 2^{-2\ell}$. Par ailleurs,

$$\overline{B_\ell} = \bigcap_{n \geqslant 1} \overline{C_{n,\ell}} \subset \bigcap_{k=1}^{K} \overline{C_{2\ell k,\ell}}.$$

Or, les événements $(C_{2\ell k,\ell})_k$ sont indépendants dans leur ensemble (ce que ne sont pas les $C_{n,\ell}$) car ils sont construits à partir d'expériences qui le sont. Il s'ensuit $\mathbf{P}(\overline{B_\ell}) \leqslant \left(1 - 2^{-2\ell}\right)^K$ pour tout K, donc $\mathbf{P}(\overline{B_\ell}) = 0$. Finalement, B_ℓ est presque sûr pour tout ℓ, donc $\mathbf{P}(B) = \lim \mathbf{P}(B_\ell) = 1$.

4. Quelques modélisations d'univers

4.1. Modélisation d'univers finis liés à la physique statistique.

Exemple 7 (Modèle de Maxwell-Boltzmann)**.** Ce modèle est utilisé pour étudier la répartition d'un système de particules selon différents niveaux d'énergie. Il repose sur un modèle d'urnes : on veut répartir r boules discernables dans n urnes. Les boules sont réparties de manière indépendante et équiprobable dans chaque urne. On peut prendre comme univers $\Omega = [\![1,n]\!]^{[\![1,r]\!]}$, ensemble des applications de $[\![1,r]\!]$ dans $[\![1,n]\!]$ (à chaque boule est associée une urne), muni de l'équiprobabilité. Chaque répartition est donc individuellement de probabilité $\dfrac{1}{n^r}$. Notons A l'événement « Chaque urne contient au plus une boule », et X la variable aléatoire comptant le nombre de boules présentes dans la première urne. Alors

$$\mathbf{P}(A) = \frac{\text{nombre d'applications injectives de } [\![1,r]\!] \text{ dans } [\![1,n]\!]}{\#\Omega} = \frac{n!}{(n-r)!\, n^r} \quad \text{et} \quad X \sim \mathcal{B}\left(r, \frac{1}{n}\right).$$

Exemple 8 (Modèle de Fermi-Dirac)**.** Ce modèle est utilisé pour étudier la répartition de fermions indiscernables sur les états d'énergie d'un système à l'équilibre thermodynamique. Il repose sur la répartition de r boules indiscernables dans n urnes, étant entendu qu'une urne ne peut pas contenir plus d'une boule. On peut prendre comme univers $\Omega = \mathscr{P}_r\big([\![1,n]\!]\big)$, ensemble des parties de $[\![1,n]\!]$ à r éléments, muni de l'équiprobabilité. Chaque répartition est donc individuellement de probabilité $\dbinom{n}{r}^{-1}$.

En reprenant la v.a. X de l'exemple précédent, X suit une loi de Bernoulli de paramètre $\dbinom{n-1}{r-1}\dbinom{n}{r}^{-1} = \dfrac{r}{n}$.

Exemple 9 (Modèle de Bose-Einstein)**.** Ce modèle s'intéresse aux bosons. Il repose sur la répartition de r boules indiscernables dans n urnes (comme pour le modèle de Maxwell-Boltzmann, mais avec des boules indiscernables). On peut cette fois prendre comme univers l'ensemble

$$\Omega_{n,r} = \big\{(\alpha_1, \alpha_2, \dots, \alpha_n) \in \mathbb{N}^n\,;\ \alpha_1 + \alpha_2 + \cdots + \alpha_n = r\big\}.$$

muni de l'équiprobabilité. Le calcul de $\#\Omega_{n,r}$ peut se faire à partir des suites croissantes $(\alpha_1, \alpha_1+\alpha_2, \dots, \alpha_1+\alpha_2+\cdots+\alpha_n)$, lequel se ramène aux suites strictement croissantes. Une alternative consiste à représenter en lignes les boules et les urnes avec des traits de séparation. Par exemple,

$$\circ\circ\circ\circ \,|\, \circ \,|\, \circ\circ |||\, \circ \,| \in \Omega_{7,8} \text{ correspond à } 4+1+2+0+0+1+0.$$

Il y a $n-1$ symboles $|$ représentant les cloisons entre les urnes et r symboles $\circ$ correspondant aux boules. En tout, il y a donc $r+n-1$ symboles parmi lesquels il faut placer $r\circ$ (ou $n-1\,|$), soit $\#\Omega_{n,r} = \dbinom{r+n-1}{r} = \dbinom{r+n-1}{n-1}$. La v.a. X comptant le nombre de boules se trouvant dans la première urne est à nouveau à valeurs dans $[\![0,r]\!]$ et l'événement $(X=k)$ correspond à une représentation de préfixe $\circ^k|$, soit

$$\mathbf{P}(X=k) = \frac{\#\Omega_{n-1,r-k}}{\#\Omega_{n,r}} = \binom{r+n-k-2}{r-k}\binom{r+n-1}{r}^{-1}.$$

L'événement A correspond à $\alpha_i \in \{0,1\}$ et revient à choisir les r urnes ayant recueilli une boule, d'où $\mathbf{P}(A) = \dbinom{n}{r}\dbinom{r+n-1}{r}^{-1}$.

Si l'on s'intéresse en variante à l'événement B :« chaque urne contient *au moins* une boule », il est pratique de revenir aux suites $(\alpha_1, \alpha_1+\alpha_2, \dots, \alpha_1+\alpha_2+\cdots+\alpha_{n-1})$, lesquelles décrivent maintenant les suites $(u_m)_{1\leqslant m\leqslant n-1}$ strictement croissantes (donc les parties) de $[\![1,r-1]\!]$ à $n-1$ éléments, d'où $\mathbf{P}(B) = \dbinom{r-1}{n-1}\dbinom{r+n-1}{r}^{-1}$ (dans le cas du modèle de Maxwell-Boltzmann, cela aurait conduit à dénombrer les applications surjectives, ce qui est un exercice classique, mais nécessitant des indications). Enfin, si Y désigne le nombre d'urnes restées vides, $Y(\Omega_{n,r}) = [\![\max(n-r,0), n-1]\!]$ et $\mathbf{P}(Y=k) = \dbinom{n}{k}\dbinom{r-1}{n-k-1}\dbinom{r+n-1}{r}^{-1}$, $\dbinom{n}{k}$ pour le choix des k urnes vides et $\dbinom{r-1}{n-k-1}$ correspondant au dénombrement ayant conduit au calcul de $\mathbf{P}(B)$ avec $n-k$ urnes et r boules.

4.2. Univers du pile ou face infini.

Posons $\Omega = \{P,F\}^{\mathbb{N}}$. Les éléments de Ω sont les suites infinies de lancers $(P,P,F,P,F,\dots)$. Un cylindre est une partie de Ω prescrivant le résultat d'un nombre fini de lancers initiaux. Par exemple, $[PP]$ est l'ensemble des suites infinies de tirages ayant commencé par deux fois pile. Un tel cylindre est de longueur 2. On peut aussi définir des cylindres généralisés, qui sont en fait des réunions finies de cylindres du type $[P*FF] = [PPFF] \cup [PFFF]$, ensemble des suites infinies de tirage dont le premier lancer est P, les troisième et quatrième F.

On peut montrer (la preuve est difficile et totalement hors programme) que si l'on prend pour $\mathcal{A}$ la plus petite tribu contenant tous les cylindres (son existence est facile à prouver car c'est l'intersection de toutes les tribus contenant les cylindres, cet ensemble étant non vide puisque $\mathscr{P}(\Omega)$ est une tribu — il est plus délicat de voir que cette tribu n'est pas $\mathscr{P}(\Omega)$), il existe une unique probabilité $\mathbf{P}$ sur $(\Omega, \mathcal{A})$ telle que, si C est un cylindre de longueur n, alors $\mathbf{P}(C) = 2^{-n}$. Un cylindre généralisé au sens ci-dessus est de probabilité 2^{-m} où m est sa longueur, compte non tenu des étoiles. On peut modéliser tout ce qui a trait aux séries de lancers de pile ou face en prenant cet espace probabilisé. La construction s'adapte à des lancers de dés ou à toute répétition indépendante d'expériences à valeurs discrètes.

5. Formules liées au conditionnement

Définition 10. *Soient* $(\Omega, \mathcal{A}, \mathbf{P})$ *un espace probabilisé,* I *un ensemble au plus dénombrable et* $(A_i)_{i\in I}$ *une famille d'événements.*

1. On dit que $(A_i)_{i\in I}$ *est un système complet d'événements si* $\Omega = \biguplus_{i\in I} A_i$.

2. On dit que $(A_i)_{i\in I}$ *est un système quasi-complet d'événements si* $\biguplus_{i\in I} A_i$ *est presque sûr.*

Exemple 10. Dans une suite infinie de « pile ou face », si A_n est l'événement « le $n$ème lancer a amené le premier face », $(A_n)_{n\geqslant 1}$ est un système quasi-complet d'événements.

5.1. Formule des probabilités composées.

Proposition 4. *Pour une famille finie* $(A_k)_{1\leqslant k\leqslant n}$ *d'événements d'un espace probabilisé* $(\Omega, \mathcal{A}, \mathbf{P})$ *et si* $A_1\cap\cdots\cap A_{n-1}$ *n'est pas presque impossible, on a*

$$\mathbf{P}(A_1 \cap \cdots \cap A_n) = \mathbf{P}(A_1)\,\mathbf{P}(A_2 \mid A_1)\,\mathbf{P}(A_3 \mid A_1 \cap A_2)\cdots\mathbf{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}).$$

La formule des probabilités composées est utile pour estimer la probabilité d'un événement dépendant d'une suite antérieure d'expériences ; il s'applique notamment aux tirages sans remise.

5.2. Formule des probabilités totales.

Proposition 5. *Soient* I *un ensemble au plus dénombrable et* $(A_i)_{i\in I}$ *un système complet ou quasi-complet d'événements d'un espace probabilisé* $(\Omega, \mathcal{A}, \mathbf{P})$*. Alors, pour tout* $A \in \mathcal{A}$ *:*

$$\mathbf{P}(A) = \sum_{i\in I} \mathbf{P}(A \cap A_i) = \sum_{\substack{i\in I \\ \mathbf{P}(A_i)\neq 0}} \mathbf{P}(A \mid A_i)\,\mathbf{P}(A_i) = \sum_{i\in I} \mathbf{P}(A \mid A_i)\,\mathbf{P}(A_i).$$

Les deux premières égalités sont vraies même si certains A_i sont négligeables. La troisième est un abus de langage toléré et pratique, les éventuelles probabilités non définies $\mathbf{P}(A \mid A_i)$ étant multipliée par 0.

La formule des probabilités totales s'applique notamment au résultat d'une deuxième expérience dont les conditions dépendent du résultat d'une première expérience aléatoire, dont les issues sont prises comme système complet ou quasi-complet.

5.3. Formule de Bayes.

Proposition 6. *Soient* I *un ensemble au plus dénombrable et* $(A_i)_{i\in I}$ *un système complet ou quasi-complet d'événements d'un espace probabilisé* $(\Omega, \mathcal{A}, \mathbf{P})$*. Alors, pour tout* $A \in \mathcal{A}$ *et* $i_0 \in I$ *:*

$$\mathbf{P}(A_{i_0} \mid A) = \frac{\mathbf{P}(A_{i_0} \cap A)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A \mid A_{i_0})\,\mathbf{P}(A_{i_0})}{\sum_{i\in I} \mathbf{P}(A \mid A_i)\,\mathbf{P}(A_i)}.$$

La formule de Bayes se place dans la même configuration que la formule des probabilités totales, à ceci près que l'on cherche à retrouver le résultat de la première expérience en observant le résultat de la deuxième. Cette formule s'appelle ainsi aussi *formule de probabilités des causes*.

Exemple 11. Lors d'une épidémie, 35% des animaux d'un élevage sont atteints par la maladie. Il existe un test de dépistage avec les caractéristiques suivantes : la probabilité qu'un animal malade montre une réaction positive est 0,9 ; la probabilité qu'un animal sain montre une réaction négative est 0,8. On prend un animal au hasard. On note M l'événement qu'il soit malade, $\overline{M}$, qu'il soit sain, N, qu'il soit négatif au test et $\overline{N}$ qu'il soit positif. Les données se traduisent par $\mathbf{P}(M) = \frac{35}{100} = \frac{7}{20}$, donc $\mathbf{P}(\overline{M}) = \frac{13}{20}$, puis $\mathbf{P}(\overline{N} \mid M) = \frac{9}{10}$ et $\mathbf{P}(N \mid \overline{M}) = \frac{8}{10}$ (il serait maladroit ici de simplifier la fraction).

On cherche dans un premier temps les probabilités qu'un animal soit malade (resp. sain) en fonction du résultat du test. La formule de Bayes donne

$$\mathbf{P}(M\,|\,\overline{N}) = \frac{\mathbf{P}(\overline{N}\,|\,M)\,\mathbf{P}(M)}{\mathbf{P}(\overline{N}\,|\,M)\,\mathbf{P}(M) + \mathbf{P}(\overline{N}\,|\,\overline{M})\,\mathbf{P}(\overline{M})} = \frac{\frac{9}{10}\times\frac{7}{20}}{\frac{9}{10}\times\frac{7}{20}+\frac{2}{10}\times\frac{13}{20}} = \frac{63}{89}.$$

On suppose maintenant que si l'on répète plusieurs fois le test sur le même animal, les résultats des tests sont indépendants. On cherche la probabilité qu'un animal testé deux fois comme sain soit sain. Avec des notations transparentes, on cherche la probabilité $\mathbf{P}(\overline{M}\,|\,NN)$. L'hypothèse d'indépendance donne $\mathbf{P}(NN\,|\,\overline{M}) = \mathbf{P}(N\,|\,\overline{M})^2 = \dfrac{64}{100}$ et $\mathbf{P}(NN\,|\,M) = \mathbf{P}(N\,|\,M)^2 = \dfrac{1}{100}$. On utilise à nouveau la formule de Bayes.

$$\mathbf{P}(\overline{M}\,|\,NN) = \frac{\mathbf{P}(NN\,|\,\overline{M})\,\mathbf{P}(\overline{M})}{\mathbf{P}(NN\,|\,M)\,\mathbf{P}(M)} = \frac{\frac{64}{100}\times\frac{13}{20}}{\frac{64}{100}\times\frac{13}{20}+\frac{1}{100}\times\frac{7}{20}} = \frac{832}{839}.$$

Cherchons maintenant la probabilité qu'un animal testé une fois comme sain, puis une fois comme malade soit malade. On commence par calculer

$$\mathbf{P}(N\overline{N}\cup\overline{N}N\,|\,M) = 1 - \mathbf{P}(NN\,|\,M) - \mathbf{P}(\overline{N}\overline{N}\,|\,M) = 1 - \frac{1}{100} - \frac{81}{100} = \frac{18}{100}.$$
$$\mathbf{P}(N\overline{N}\cup\overline{N}N\,|\,\overline{M}) = 1 - \mathbf{P}(NN\,|\,\overline{M}) - \mathbf{P}(\overline{N}\overline{N}\,|\,\overline{M}) = 1 - \frac{64}{100} - \frac{4}{100} = \frac{32}{100}.$$

On peut alors une nouvelle fois utiliser la formule de Bayes.

$$\mathbf{P}(M\,|\,N\overline{N}\cup\overline{N}N) = \frac{\mathbf{P}(N\overline{N}\cup\overline{N}N\,|\,M)\,\mathbf{P}(M)}{\mathbf{P}(N\overline{N}\cup\overline{N}N\,|\,M)\,\mathbf{P}(M) + \mathbf{P}(N\overline{N}\cup\overline{N}N\,|\,\overline{M})\,\mathbf{P}(\overline{M})} = \frac{\frac{18}{100}\times\frac{7}{20}}{\frac{18}{100}\times\frac{7}{20}+\frac{32}{100}\times\frac{13}{20}} = \frac{63}{271}.$$

On peut aussi utlliser la formule des probabilités totales avec le système complet d'événements $(M\overline{M})$ pour calculer $\mathbf{P}(NN) = \dfrac{839}{2000}$ et $\mathbf{P}(\overline{N}\overline{N}) = \dfrac{619}{2000}$, d'où $\mathbf{P}(N\overline{N}\cup\overline{N}N) = \dfrac{542}{2000}$, ce qui redonne le dénominateur précédent.

6. Dictionnaire

L'axiomatique de Kolmogorov utilise la théorie des ensembles. L'ensemble Ω modélise l'univers des issues possibles d'une expérience. Les événements sont des parties de Ω, donc des ensembles, et $\mathbf{P}(A)$ une mesure de ces ensembles.

Ω	événement certain	$\varnothing$	événement impossible
$\mathbf{P}(A) = 1$	événement presque sûr	$\mathbf{P}(A) = 0$	événement négligeable (presque impossible)
$\{\omega\}$ $(\omega \in \Omega)$	événement élémentaire	$\overline{A}$	événement contraire
$A \cup B$	A ou B	$A \cap B$	A et B
$A \bigtriangleup B$	A ou (exlusif) B	$A \cap B = \varnothing$	événements incompatibles
$A \subset B$	A implique B	$\biguplus A_n$	réunion d'événements 2 à 2 incompatibles
$\biguplus A_n = \Omega$	système complet	$\mathbf{P}\left(\biguplus A_n\right) = 1$	système quasi-complet
$\bigcup_{n\geqslant 0}\bigcap_{k\geqslant n} A_k$	A_n à.p.c.r.	$\bigcap_{n\geqslant 0}\bigcup_{k\geqslant n} A_k$	A_n pour une infinité de n (A_n *i.s.*)

Remarque 4. Les probabilités (*a priori*) faciles à calculer sont :
— celle d'une réunion au plus dénombrable d'événements deux à deux incompatibles ;
— celle d'une intersection finie d'événements indépendants dans leur ensemble ;
— celle de l'intersection d'une suite décroissante ou de la réunion d'une suite croissante d'événements ;
— celles données par l'une des trois formules des probabilités conditionnelles (probabilités composées, probabilités totales, formule de Bayes).

7. Moments

Définition 11. *Les* moments *d'une v.a.r. sont les* $\mathbf{E}(X^k)$ *et les* moments centrés *sont les* $\mathbf{E}\left((X-\mathbf{E}(X))^k\right)$, $k \in \mathbb{N}^*$.

Pour donner un sens à cela, il faut commencer par définir le moment d'ordre 1, c'est-à-dire l'*espérance*. On pourra noter que le moment centré d'ordre 1 est nul et n'a donc pas d'intérêt.

7.1. **Espérance.**

Définition 12. *Si X est une v.a.d. à valeurs réelles ou complexes, on dit que X admet une espérance si la famille $(\mathbf{P}(X=x)_{x\in X(\Omega)}$ est sommable. Si tel est le cas, la somme de la série est l'espérance de X et est notée* $\mathbf{E}(X)$.

Dans le cas des variables aléatoires positives, on peut indifféremment utiliser *d'espérance finie* et *admettant une espérance*. De fait, une famille positive non sommable peut sans inconvénient se voir affecter la valeur $+\infty$.

Remarque 5. On rappelle que la *sommabilité* signifie, par définition, la convergence absolue de $\sum x_n \mathbf{P}(X=x_n)$ d'une énumération $X(\Omega) = \{x_0, x_1, \ldots\}$ et que tant la sommabilité que la somme sont indépendantes de cette énumération.

La raison de l'hypothèse *à valeurs réelles* est qu'une espérance est une moyenne. Si l'on modélise un jeu de cartes, on peut définir une v.a. à valeurs dans $\{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$. Une moyenne n'a dans ce contexte aucun sens ; quelle serait la moyenne de $\heartsuit$ et de $\spadesuit$? Cette hypothèse de valeurs réelles peut être naturellement amoindrie car il suffit précisément que l'on puisse faire une moyenne, donc additionner et multiplier par un scalaire... Une structure adéquate est bien sûr celle d'espace vectoriel. En particulier, il n'est pas compliqué de parler d'espérance d'une variable aléatoire à valeurs dans $\mathbb{R}^n$. On se ramène au cas réel en raisonnant coordonnée par coordonnée.

Propriétés.

(i) X admet une espérance si, et seulement si, $|X|$ admet une espérance.

(ii) Si $(a,b) \in \mathbb{R}^2$ et si X admet une espérance, alors $aX+b$ aussi et $\mathbf{E}(aX+b) = a\,\mathbf{E}(X)+b$.

(iii) Si $X \geqslant 0$ et si X admet une espérance, alors $\mathbf{E}(X) \geqslant 0$.

(iv) Si $X \geqslant 0$, si X admet une espérance et si $\mathbf{E}(X)=0$, alors $X=0$ p.s.

(v) *Théorème de comparaison.* Si $0 \leqslant |X| \leqslant Y$ et si Y admet une espérance, alors X admet une espérance et
$$|\,\mathbf{E}(X)| \leqslant \mathbf{E}(|X|) \leqslant \mathbf{E}(Y);$$
en particulier, l'espérance est croissante et toute v.a. réelle bornée admet une espérance.

(vi) *Théorème de transfert.* Si $\varphi\colon \mathbb{R} \to \mathbb{R}$ alors $\mathbf{E}\left(\varphi(X)\right) = \displaystyle\sum_{x\in X(\Omega)} \varphi(x)\,\mathbf{P}(X=x)$, où il est entendu que l'existence de $\mathbf{E}\left(\varphi(X)\right)$ est équivalente à la convergence absolue de la série $\displaystyle\sum_{x\in X(\Omega)} \varphi(x)\,\mathbf{P}(X=x)$ et qu'en cas de convergence absolue, ils sont égaux.

(vii) L'ensemble des v.a.d. définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et admettant une espérance est un espace vectoriel, sur lequel l'espérance définit une application linéaire.

(viii) *Inégalité de Markov.* Si $X \geqslant 0$ admet une espérance, alors, pour tout réel $a>0$, on a $\mathbf{P}(X \geqslant a) \leqslant \dfrac{\mathbf{E}(X)}{a}$.

Preuve : l'inégalité $X \geqslant a\,\mathbb{1}_{(X\geqslant a)}$ donne $\mathbf{E}(X) \geqslant a\,\mathbf{P}(X\geqslant a)$ par croissance de l'espérance.

Remarque 6. Toutes ces propriétés sont faciles à retenir car ce sont des propriétés que l'on a déjà vues dans le cadre de l'intégration (sauf le théorème de transfert et l'inégalité de Markov). Ce n'est pas un hasard, la théorie générale de l'intégrale englobant les probabilités. Même si cela ne se voit pas à bac + 2, une espérance *est* une intégrale. On notera qu'on n'a pas mentionné la propriété la plus importante : la linéarité de l'espérance. Elle s'énonce en effet dans le cadre des couples.

Exemple 12. Un joueur joue contre contre la banque du casino à pile ou face. Contre une mise initiale m, il remporte 1 si face sort du premier coup, 2 si face sort seulement au deuxième et 2^{n-1} si le premier face est obtenu au n-ème coup. Que doit valoir m pour que le jeu soit *équitable*, c'est-à-dire pour que l'espérance de gain soit nulle de chaque côté ?

Le gain G du joueur est $g-m$, où g est ce que lui verse le casino. En notant X la date d'apparition du premier face, g vaut 2^{n-1} avec probabilité 2^{-n}, ce qui donne $m = \mathbf{E}(G) = \displaystyle\sum_{n=1}^{\infty} 2^{n-1} \times 2^{-n} = \sum_{n=1}^{\infty} 2^{-1} = +\infty$. La variable g n'admet pas d'espérance et la mise devrait être infinie.

Il s'agit du *Paradoxe de Saint-Pétersbourg*. C'est un paradoxe en ce que toute mise initiale finie donne une espérance de gain infinie, mais qu'il ne viendrait à l'idée de personne de miser à un tel jeu, contrairement à ce que « conseille » la théorie des probabilités. En fait, il est intéressant de jouer à un jeu d'espérance strictement positive car la loi des grands nombres montre que la probabilité de gagner tend vers 1 quand le nombre de répétitions du jeu tend vers l'infini.

Toutefois, le fait de disposer d'une fortune finie au départ, ce qui est le cas de la plupart des gens, interdit d'une part de miser des millions de fois une certaine somme, et l'esprit se refuse d'autre part à la perspective de gagner peu avec une probabilité énorme et de tout perdre avec une probabilité infime. Les jeux comme Euromillions proposent au contraire de perdre une mise consentie et maîtrisée avec une probabilité élevée, de gagner un peu de temps en temps, et de rêver de gagner une fortune. Une variante de ce paradoxe est la martingale bien connue consistant à doubler sa mise à la roulette à chaque fois que l'on perd, jusqu'à ce que l'on gagne... ou que l'on soit ruiné.

7.2. **Moments d'ordre 2.**

7.2.1. *Variance.*

Définition 13. *Le moment centré d'ordre 2 s'appelle la* variance, *soit* $\mathbf{V}(X) = \mathbf{E}\left((X - \mathbf{E}(X))^2\right)$. *On appelle* écart-type *de X la grandeur* $\sigma(X) = \sqrt{\mathbf{V}(X)}$. *Une variable aléatoire est dite* centrée *si elle est d'espérance nulle et* réduite *si son écart-type vaut 1.*

Proposition 7. *Si X est une v.a.d., l'existence de $\mathbf{E}(X^2)$ entraîne celle de $\mathbf{E}(X)$ et est équivalent à celle de $\mathbf{V}(X)$.*

Pour voir que l'existence de $\mathbf{E}(X^2)$ entraîne celle de $\mathbf{E}(X)$, on peut développer $(|X| - 1)^2 \geqslant 0$, ou utiliser une majoration plus probabiliste dans l'esprit, soit

$$|X| = \left(\mathbb{1}_{|X|\leqslant 1} + \mathbb{1}_{|X|>1}\right)|X| \leqslant 1 + \mathbb{1}_{|X|>1}|X| \leqslant 1 + X^2.$$

La deuxième majoration est plus faible que la première d'un facteur 2, mais elle se généralise bien à des puissances supérieures.

Propriétés.

(i) $\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$ (formule de Koenig) ;

(ii) $\mathbf{V}(aX + b) = a^2\,\mathbf{V}(X)$, donc $\sigma(aX + b) = |a|\sigma(X)$;

(iii) $\mathbf{V}(X) = 0$ si, et seulement si, X est presque sûrement constante ;

(iv) *Inégalité de Bienaymé-Tchebychev.* Si X admet un moment d'ordre 2, alors, pour tout $a > 0$, on a

$$\mathbf{P}\left(|X - \mathbf{E}(X)| \geqslant a\right) \leqslant \frac{\mathbf{V}(X)}{a^2}.$$

Démonstration. Montrons l'inégalité de Bienaymé-Tchebychev. En utilisant la croissance de la fonction $x \longmapsto x^2$ sur $\mathbb{R}_+$ (astuce utilisée à l'occasion avec d'autres fonctions croissantes),

$$\mathbf{P}\left(|X - \mathbf{E}(X)| \geqslant a\right) = \mathbf{P}\left((X - \mathbf{E}(X))^2 \geqslant a^2\right) \underset{\text{(Markov)}}{\leqslant} \frac{\mathbf{E}\left((X - \mathbf{E}(X)^2\right)}{a^2} = \frac{\mathbf{V}(X)}{a^2}.$$

□

7.2.2. *Couples.*

Proposition 8. ***1.*** *Si X et Y admettent un moment d'ordre 2, alors XY admet une espérance et*

$$\mathbf{E}(XY) = \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xy\,\mathbf{P}(X = x, Y = y).$$

2. *L'application $(X, Y) \mapsto \mathbf{E}(XY)$ définit sur l'espace vectoriel des v.a.d. admettant une espérance une application bilinéaire symétrique positive. C'est « presque » un produit scalaire en ce sens que $\mathbf{E}(X^2) = 0$ n'équivaut pas à $X = 0$, mais à $X = 0$ p.s.*
3. *L'inégalité de Cauchy-Schwarz montre que si X et Y admettent un moment d'ordre 2, alors XY admet une espérance et $\mathbf{E}(XY)^2 \leqslant \mathbf{E}(X^2)\,\mathbf{E}(Y^2)$. Dans le cas de v.a.i., on a un résultat plus fort :*
3. *Si $X \perp\!\!\!\perp Y$ admettent une espérance, alors XY admet une espérance et $\mathbf{E}(XY) = \mathbf{E}(X)\,\mathbf{E}(Y)$.*

Définition 14. *Si X et Y admettent un moment d'ordre 2, on définit la covariance du couple (X,Y) par*

$$\mathbf{cov}(X,Y) = \mathbf{E}\Big((X-\mathbf{E}(X))(Y-\mathbf{E}(Y)\Big) = \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y).$$

Si $\mathbf{cov}(X,Y)=0$, on dit que X et Y sont non corrélées. *C'est un peu plus faible que l'indépendance en vertu de la proposition 8 et d'un contrexemple vu en exercice.*

Propriétés.

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

(ii) L'inégalité de Cauchy-Schwarz donne $|\,\mathbf{cov}(X,Y)| \leqslant \sigma(X)\sigma(Y)$.

(iii) Les v.a.d. admettant un moment d'ordre 2 forment un espace vectoriel.

(iv) *Théorème de transfert.* Si X et Y sont des v.a.d., et $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$, alors

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

au sens où l'existence de $\mathbf{E}(\varphi(X,Y))$ est équivalente à la convergence absolue de la série

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

et qu'en cas d'existence, il y a égalité numérique.

(v) Corrélativement, si X et Y admettent une espérance et si $(\lambda,\mu)\in\mathbb{R}^2$, alors $\lambda X + \mu Y$ admet une espérance et l'on a $\mathbf{E}(\lambda X + \mu Y) = \lambda\,\mathbf{E}(X) + \mu\,\mathbf{E}(Y)$.

(vi) Si $(X_1, X_2, \ldots, X_n)$ admettent un moment d'ordre 2, alors

$$\mathbf{V}(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} \mathbf{V}(X_i) + 2 \sum_{1\leqslant i<j\leqslant n} \mathbf{cov}(X_i, X_j).$$

(vii) Si $(X_1, X_2, \ldots, X_n)$ sont deux à deux indépendantes et admettent une variance, alors

$$\mathbf{V}(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} \mathbf{V}(X_i).$$

8. Séries génératrices

Cette section introduit un outil spécifique aux variables aléatoires à valeurs dans $\mathbb{N}$ (et donc discrètes).

Définition 15. *Si X est une v.a. à valeurs dans $\mathbb{N}$, on définit sa série (ou fonction) génératrice par*

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

L'expression comme série entière est donnée par le théorème de transfert. La série génératrice est une série entière de rayon de convergence $R \geqslant 1$ et vérifie $G_X(1) = 1$. Par unicité du développement en série entière, elle caractérise entièrement la loi : $G_X = G_Y$ si, et seulement si, X et Y suivent la même loi. On peut dériver G_X terme à terme. Si $R > 1$, le théorème de transfert donne $G_X^{(p)}(1) = \mathbf{E}\left(X(X-1)\cdots(X-p+1)\right)$ pour tout $p \geqslant 1$ et, en particulier, pour $p=1$, $G'_X(1) = \mathbf{E}(X)$. Ainsi, les moments $\mathbf{E}(X^p)$ existent pour tout p. Si le rayon de convergence vaut 1, c'est plus subtil, comme le montre le théorème suivant.

Théorème 1. *Soit X, une variable aléatoire à valeurs dans $\mathbb{N}$.*
1. La v.a. X admet une espérance si, et seulement si, G_X est dérivable à gauche en 1. On a alors $\mathbf{E}(X) = G'_X(1)$.

2. La v.a. X admet une variance si, et seulement si, G_X est deux fois dérivable à gauche en 1. On a alors la relation $G''_X(1) = \mathbf{E}(X(X-1))$, d'où $\mathbf{V}(X) = G''_X(1) + G'_X(1) - G'_X(1)^2$.

3. La v.a. X admet une espérance si, et seulement si, la série $\sum \mathbf{P}(X>n)$ converge. On a alors

$$\mathbf{E}(X) = \sum_{n=0}^{\infty} \mathbf{P}(X>n) = \sum_{n=1}^{\infty} \mathbf{P}(X\geqslant n).$$

Ce théorème est important, même s'il ne faut pas forcément se précipiter sur les séries génératrices pour calculer une espérance ou une variance. Sa démonstration est non exigible. Les trois propriétés sont des conséquences assez immédiates de la version suivante du théorème de la double limite, qui se démontre lui-même bien plus facilement que celui admis dans le chapitre sur le convergence uniforme et qu'il faut prendre comme un exercice instructif, et non comme un résultat (de plus...) à apprendre sur la convergence des séries de fonctions :

Lemme 1. *Soit une suite de fonctions continues et croissantes* $f_n\colon [a,b] \to \mathbb{R}_+$ *avec* $-\infty < a < b < +\infty$. *On suppose que la série* $\sum_{n=0}^{\infty} f_n(x)$ *converge simplement sur* $[a,b[$ *et l'on note* f *sa somme. Alors, les limites étant dans* $[0,+\infty]$,

$$\lim_{\substack{x\to b\\ x<b}} f(x) = \lim_{N\to\infty} \sum_{n=0}^{N} f_n(b),$$

Les séries génératrices ne servent pas qu'à calculer les moments. Elles permettent aussi d'utiliser simplement l'indépendance.

Proposition 9. *Soient* X *et* Y *des v.a. indépendantes à valeurs dans* $\mathbb{N}$. *Alors,* $G_{X+Y} = G_X \times G_Y$. *Plus généralement, si* $X_1, X_2, \dots, X_n$ *est une famille de v.a. indépendantes, alors* $G_{X_1+X_2+\cdots+X_n} = \prod_{k=1}^{n} G_{X_i}$.

Démonstration. Supposons que X et Y sont indépendantes. Le produit de Cauchy donne

$$\mathbf{P}(X+Y=n) = \mathbf{P}\left(\biguplus_{k=0}^{n} (X=k, Y=n-k)\right) \underset{\text{(incomp.)}}{=} \sum_{k=0}^{n} \mathbf{P}(X=k,\, Y=n-k) \underset{\text{(indép.)}}{=} \sum_{k=0}^{n} \mathbf{P}(X=k)\,\mathbf{P}(Y=n-k)$$

$$\therefore \quad \begin{aligned} G_X(x) \times G_Y(x) &= \sum_{n=0}^{\infty} \mathbf{P}(X=n)x^n \times \sum_{n=0}^{\infty} \mathbf{P}(Y=n)x^n \\ &= \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} \mathbf{P}(X=k)\,\mathbf{P}(Y=n-k) \right) x^n = \sum_{n=0}^{\infty} \mathbf{P}(X+Y=n)\, x^n. \end{aligned}$$

Le cas général suit par une récurrence immédiate sur n. □

9. Lois usuelles

9.1. **Lois certaine et quasi-certaine.** C'est le cas $X(\Omega) = \{a\}$, soit $X = a$ pour la loi certaine et $X = a$ p.s. pour la loi quasi-certaine. On a évidemment $\mathbf{E}(X) = a$ et la loi quasi-certaine caractérise les v.a. de variance nulle.

9.2. **Loi uniforme.** C'est une loi d'équiprobabilité sur $X(\Omega)$, qui ne peut être dénombrable et est donc fini (mais il existe des lois uniformes sur des domaines non dénombrables, comme des intervalles de $\mathbb{R}$). Si $X \sim \mathcal{U}(E)$, alors $\mathbf{E}(X) = \dfrac{1}{\#E} \sum_{x\in E} x$. Si $E = [\![1, n]\!]$, on a $G_X(t) = \dfrac{t(1-t^n)}{n(1-t)}$.

9.3. **Loi de Bernoulli.** Une v.a. suit une loi de Bernoulli si $X(\Omega) = \{0,1\}$. Le paramètre d'une v.a. suivant une loi de Bernouilli est $\mathbf{P}(X=1) = \mathbf{E}(X)$. On a aussi $G_X(x) = q + px$, avec $q = 1-p$ et $\mathbf{V}(X) = pq$. Une v.a. de Bernoulli est une fonction indicatrice car $X = \mathbb{1}_{(X=1)}$ (X vaut 1 quand elle vaut 1 et 0 sinon).

Les fonctions indicatrices sont des outils utiles (on l'a vu par exemple pour prouver l'inégalité de Markov). En voici quelques propriétés :

- A et B sont indépendants si, et seulement si $\mathbb{1}_A$ et $\mathbb{1}_B$ le sont ;
- $\mathbb{1}_{\overline{A}} = \mathbb{1} - \mathbb{1}_A$;
- $\mathbb{1}_{A\cap B} = \mathbb{1}_A \times \mathbb{1}_B$;
- $\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \mathbb{1}_B$.

On peut facilement obtenir la formule du crible (c'est un exercice, la formule n'est pas exigible) en utilisant les fonctions indicatrices :

$$\mathbb{1}_{A_1\cup A_2\cup\cdots\cup A_n} = 1 - \mathbb{1}_{\overline{A_1}\cap\overline{A_2}\cap\cdots\cap\overline{A_n}} = 1 - \prod_{i=1}^{n} (1 - \mathbb{1}_{A_i}) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n} \mathbb{1}_{A_{i_1}\cap\cdots\cap A_{i_k}} \quad \therefore$$

$$\mathbf{P}(A_1\cup A_2\cup\cdots\cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n} \mathbf{P}(A_{i_1}\cap\cdots\cap A_{i_k})$$

en prenant l'espérance. La proposition suivante est une conséquence immédiate de la caractérisation de l'indépendance des événements (équation 1).

Proposition 10. *Soient $(\Omega, \mathcal{A}, \mathbf{P})$ un espace probabilisé et $(A_i)_{i \in I}$ une famille d'événements de cet espace. Alors, les événements A_i sont indépendants si, et seulement si, la famille des indicatrices $\mathbb{1}_{A_i}$ le sont.*

9.4. **Loi binomiale.** Si $(X_i)_{1 \leqslant i \leqslant n}$ est une suite de v.a.i.i.d. telle que $X_1 \sim \mathcal{B}(p)$, alors $X = \sum_{i=1}^{n} X_i \sim \mathcal{B}(n,p)$, loi binomiale. On a $X(\Omega) = [\![0,n]\!]$ et $\mathbf{P}(X=k) = \binom{n}{k} p^k q^{n-k}$ pour tout $k \in [\![0,n]\!]$. De plus, $\mathbf{E}(X) = np$, $\mathbf{V}(X) = npq$ et $G_X(x) = (q+px)^n$. La loi binomiale $\mathcal{B}(n,p)$ modélise le nombre de succès lors de n expériences indépendantes et similaires, de probabilité de succès individuel p. En particulier, si $X_1 \sim \mathcal{B}(n_1,p)$ et $X_2 \sim \mathcal{B}(n_2,p)$ sont indépendantes, alors $X_1 + X_2 \sim \mathcal{B}(n_1+n_2,p)$. C'est aussi immédiat en utilisant les séries génératrices.

Exemple 13. Exercice classique : conditionnement de la loi binomiale par elle-même. Si $X \sim \mathscr{B}(n,p)$ et si, pour tout $k \in [\![0,n]\!]$, $Y_{|(X=k)} \sim \mathscr{B}(k,p')$, alors $Y \sim \mathscr{B}(n,pp')$.

9.5. **Loi de Poisson.** C'est la loi dite des *événements rares*. Elle approche la loi binomiale quand la probabilité est faible et le nombre de répétitions grand. Plus précisément :

Proposition 11. *Si $\lambda > 0$ et si $X_n \sim \mathcal{B}\left(n, \frac{\lambda}{n}\right)$, alors, pour tout $k \in \mathbb{N}$, $\lim_{n\to\infty} \mathbf{P}(X_n = k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}$.*

Définition 16. *On dit que X suit une loi de Poisson de paramètre λ et l'on note $X \sim \mathscr{P}(\lambda)$ si $X(\Omega) = \mathbb{N}$ et si, pour tout $k \in \mathbb{N}$, $\mathbf{P}(X=k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}$.*

Pour $X \sim \mathscr{P}(\lambda)$, on calcule $G_X(x) = \mathrm{e}^{\lambda(x-1)}$, $\mathbf{V}(X) = \mathbf{E}(X) = \lambda$. Si $X \sim \mathscr{P}(\lambda)$ et $Y \sim \mathscr{P}(\mu)$ sont indépendantes, alors $X + Y \sim \mathscr{P}(\lambda+\mu)$. La démonstration est immédiate si l'on utilise les séries génératrices.

Exemple 14. Exercice classique : conditionnement d'une loi de Poisson par une loi binomiale. Si $N \sim \mathscr{P}(\lambda)$ et si, pour tout $n \in \mathbb{N}$, $X_{|(N=n)} \sim \mathscr{B}(n,p)$, alors $X \sim \mathscr{P}(\lambda p)$ et $X \perp\!\!\!\perp N - X$.

9.6. **Loi géométrique.** La loi géométrique modélise le temps d'attente du premier succès lors d'une suite d'expériences indépendantes et de même probabilité individuelle de succès. On note $X \sim \mathscr{G}(p)$ si $X(\Omega) = \mathbb{N}^*$ et si, pour tout $k \in \mathbb{N}^*$, $\mathbf{P}(X=k) = q^{k-1}p$ ($k-1$ échecs suivis d'un succès). La série génératrice de la loi géométrique de paramètre p est $G_X(x) = \frac{px}{1-qx}$, on a $\mathbf{E}(X) = \frac{1}{p}$ et $\mathbf{V}(X) = \frac{q}{p^2}$. Plutôt que la valeur de $\mathbf{P}(X=k)$, il est souvent pratique d'utiliser $\mathbf{P}(X>k) = q^k$.

Remarque 7. Si $X(\Omega) = \mathbb{N}^*$, X suit une loi géométrique si, et seulement si, X est *sans mémoire* au sens suivant :

$$\forall (n,k) \in \mathbb{N}^2 \colon\ \mathbf{P}(X > n+k \,|\, X > n) = \mathbf{P}(X > k).$$

On a alors $q = \mathbf{P}(X > 1)$. La condition est en effet nécessaire : par définition de la probabilité conditionnelle,

$$\mathbf{P}(X > n+k \,|\, X > n) = \frac{\mathbf{P}\big((X > n+k) \cap (X > n)\big)}{\mathbf{P}(X > n)} = \frac{\mathbf{P}(X > n+k)}{\mathbf{P}(X > n)} = \frac{q^{n+k}}{q^n} = q^k = \mathbf{P}(X > k).$$

Pour la réciproque, prenons $k = 1$ et posons $q = \mathbf{P}(X > 1)$. Alors, la relation $\mathbf{P}(X > n+1 \,|\, X > n) = \frac{\mathbf{P}(X > n+1)}{\mathbf{P}(X > n)} = q$ montre que la suite $\big(\mathbf{P}(X > n)\big)_n$ est géométrique de raison q et l'on a donc $\mathbf{P}(X > n) = q^n$, soit $X \sim \mathscr{G}(p)$, puisque

$$\mathbf{P}(X = n) = \mathbf{P}(X > n-1) - \mathbf{P}(X > n) = q^{n-1} - q^n = q^{n-1}p.$$

10. Théorèmes limites

10.1. Loi faible des grands nombres.

Théorème 2. *Soit une suite* $(X_n)_n$ *de v.a.i.i.d. possédant un moment d'ordre 2. On note* $M_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ *la moyenne des n premières v.a. X_k. Alors :*

$$\forall \varepsilon > 0 \colon \lim_{n\to\infty} \mathbf{P}\Big(|M_n - \mathbf{E}(X_1)| \geqslant \varepsilon\Big) = 0.$$

L'inégalité de Bienaymé-Tchebychev donne une *one-line proof* de la loi faible et la majoration explicite

$$\mathbf{P}\Big(|M_n - \mathbf{E}(X_1)| \geqslant \varepsilon\Big) \leqslant \frac{\mathbf{V}(X)}{n\varepsilon^2}.$$

La loi des grands nombres jette un pont entre les probabilités et les statistiques et répond notamment à la question « comment estimer empiriquement p à partir d'un (gros) échantillon ? » Elle est valide sous l'hypothèse plus faible de l'existence de $\mathbf{E}(X)$, mais ce raffinement est hors programme.

11. Convergences (HP)

En probabilités, comme en analyse, il existe de nombreux modes de convergence. On en a vu quelques uns en mode *Monsieur Jourdain*, qui sont dans l'adhérence du programme et qui peuvent faire l'objet de problèmes de concours. Réservé à la préparation X-ENS pour ceux ayant assimilé tout le reste.

Définition 17. *Soient X et $X_1, X_2, \ldots,$ des variables aléatoires définies sur un même espace probabilisé* $(\Omega, \mathcal{A}, \mathbf{P})$.

1. *On dit que $(X_n)_n$ converge en loi vers X si, pour tout point de continuité $x \in \mathbb{R}$ de la fonction de répartition F_X,* $\lim F_{X_n}(x) = F_X(x)$. *Si les v.a. sont à valeurs dans $\mathbb{N}$, cela équivaut à* $\lim \mathbf{P}(X_n = k) = \mathbf{P}(X = k)$ *pour tout $k \in \mathbb{N}$. On note* $X_n \xrightarrow[n\to\infty]{\mathcal{L}} X$.

2. *On dit que $(X_n)_n$ converge en probabilité vers X si, pour tout $\varepsilon > 0$,* $\lim \mathbf{P}(|X_n - X| \geqslant \varepsilon) = 0$. *On note* $X_n \xrightarrow[n\to\infty]{\mathrm{P}} X$.

3. *On dit que $(X_n)_n$ converge presque sûrement vers X si l'événement* $\{\omega \in \Omega\,;\ \lim X_n(\omega) = X(\omega)\}$ *est presque sûr. On note* $X_n \xrightarrow[n\to\infty]{\text{p.s.}} X$, *ou* $X_n \xrightarrow[n\to\infty]{} X$ *p.s.*

Propriétés. — La convergence de $X_n \sim \mathcal{B}(n, \lambda/n)$ vers $X \sim \mathscr{P}(\lambda)$ est une convergence en loi.

— La convergence presque sûre entraîne la convergence en probabilité. C'est un bon exercice formel, mais très abstrait : on suppose que $X_n \xrightarrow[n\to\infty]{\text{p.s.}} X$. Notons $A = \{\omega \in \Omega\,;\ \lim X_n(\omega) = X(\omega)\}$. En écrivant la définition de la limite de manière dénombrable, c'est-à-dire en y substituant $1/m$ à ε, il vient

$$\omega \in A \iff \forall m \in \mathbb{N},\ \exists n \in \mathbb{N},\ \forall k \geqslant n \colon |X_k(\omega) - X(\omega)| \leqslant \frac{1}{m} \qquad \therefore$$

$$A = \bigcap_{m\geqslant 1}\bigcup_{n\geqslant 1}\bigcap_{k\geqslant n}\left(|X_k - X| < \frac{1}{m}\right) \qquad \& \qquad \overline{A} = \bigcup_{m\geqslant 1}\bigcap_{n\geqslant 1}\bigcup_{k\geqslant n}\left(|X_k - X| \geqslant \frac{1}{m}\right).$$

Par hypothèse, $\mathbf{P}(\overline{A}) = 0$, donc tous les événements dont $\overline{A}$ est l'union sont presque impossibles. Ainsi,

$$\forall m \in \mathbb{N}^* \colon \mathbf{P}\left(\bigcap_{n\geqslant 1}^{\downarrow}\bigcup_{k\geqslant n}\left(|X_k - X| \geqslant \frac{1}{m}\right)\right) = 0 \quad \therefore \quad \lim_{n\to\infty}\mathbf{P}\left(\bigcup_{k\geqslant n}\left(|X_k - X| \geqslant \frac{1}{m}\right)\right) = 0.$$

alors, l'inclusion $\left(|X_n - X| \geqslant \dfrac{1}{m}\right) \subset \displaystyle\bigcup_{k\geqslant n}\left(|X_k - X| \geqslant \frac{1}{m}\right)$ et le théorème des gendarmes assurent que l'on peut passer à la limite dans $\displaystyle\lim_{n\to\infty}\mathbf{P}\left(|X_n - X| \geqslant \frac{1}{m}\right) = 0$, d'où la convergence en probabilité en prenant $m = \lceil 1/\varepsilon \rceil$.

— La convergence en probabilité entraîne la convergence en loi : dans le cas où toutes les v.a. prennent leurs valeurs dans $\mathbb{N}$, on a, pour tout $(k,n) \in \mathbb{N}^2$, $(X_n = k)\Delta(X = k) \subset \big(|X - X_n| \geqslant 1\big)$, d'où

$$|\,\mathbf{P}(X = k) - \mathbf{P}(X_n = k)| \leqslant \mathbf{P}\big((X_n = k)\Delta(X = k)\big) \leqslant \mathbf{P}\big(|X - X_n| \geqslant 1\big) \xrightarrow[n\to\infty]{} 0.$$

— La loi faible des grands nombres exprime la convergence en probabilité de $(M_n)_n$ vers la v.a. quasi-certaine $\mathbf{E}(X_1)$.

— La loi forte des grands nombres (s'il y a une *loi faible*, c'est qu'il existe aussi une *loi forte*...) dit que, sous l'hypothèse de l'existence de $\mathbf{E}(X_1)$, la suite $(M_n)_n$ converge presque sûrement vers la v.a. quasi-certaine $\mathbf{E}(X_1)$. Elle est démontrée dans le cas centré dans le problème CCINP, PSI, 2018.

L'erreur commise dans la loi des grands nombres est précisée par le *théorème central limite* : c'est une erreur en $n^{-1/2}$ quantifiée par la loi de Gauß. Plus précisément,

$$\frac{M_n - \mathbf{E}(M_n)}{\sigma(M_n)} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1) \quad \textit{i.e.} \quad \forall x \in \mathbb{R} \colon \lim_{n\to\infty} \mathbf{P}\left(\left|M_n - \mathbf{E}(X_1)\right| \leqslant \frac{x}{\sqrt{n}}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\,\mathrm{d}t.$$