

*Aix-Marseille Université
Institut Universitaire de Technologie
Département Réseaux & Télécommunications*

INTRODUCTION AU CALCUL DES PROBABILITÉS

1. *Espaces probabilisés*
2. *Variables aléatoires discrètes*
3. *Variables aléatoires réelles*

Table des matières

1	Espaces probabilisés	5
1.1	Notions fondamentales	5
1.1.1	Expérience aléatoire	5
1.1.2	Événement (d'une expérience aléatoire)	5
1.2	Probabilité	7
1.2.1	La loi empirique des grands nombres	7
1.2.2	Définition et premières propriétés	8
1.3	Probabilité uniforme et dénombrement	9
1.3.1	Etude d'un ensemble fini d'éventualités équiprobables	9
1.3.2	Dénombrement	10
1.4	Conditionnement et indépendance	12
1.4.1	Probabilités conditionnelles	12
1.4.2	Définition et propriétés	12
1.4.3	Événements indépendants	13
1.5	Un exemple récapitulatif	14
1.5.1	Description de l'expérience et du problème	14
1.5.2	Solution	15
2	Variables aléatoires discrètes	17
2.1	Loi de probabilité	17
2.1.1	Définition	17
2.1.2	Transformation de v.a.	18
2.1.3	Loi de probabilité conditionnelle	19
2.1.4	V.a. indépendantes	21
2.2	Grandeurs caractéristiques d'une v.a. réelle discrète	21
2.2.1	Espérance	22
2.2.2	Variance, écart-type	25
2.3	Quelques modèles aléatoires discrets	27
2.3.1	Lois géométriques	27
2.3.2	Lois de Bernoulli	27
2.3.3	Lois binomiales	27
2.3.4	Lois multinomiales et test d'adéquation du χ^2	29
2.3.5	Lois de Poisson	30

3	Variables aléatoires réelles	33
3.1	<i>Densités de probabilité</i>	33
3.1.1	<i>Définition et propriétés immédiates</i>	33
3.1.2	<i>Exemples de densités de probabilité</i>	34
3.2	<i>Outils permettant l'étude des v.a.r.</i>	36
3.2.1	<i>Fonction de répartition de X</i>	36
3.2.2	<i>Valeurs caractéristiques d'une v.a.r.</i>	37
3.2.3	<i>Fonction caractéristique d'une v.a.r. X.</i>	40
3.3	<i>Indépendance et convolution</i>	41
3.3.1	<i>Vecteurs aléatoires réels</i>	41
3.3.2	<i>Indépendance</i>	42
3.3.3	<i>Convolution</i>	42

Chapitre 1

Espaces probabilisés

1.1 Notions fondamentales

1.1.1 Expérience aléatoire

Cette première notion de la théorie des probabilités s'est imposée au 17^{ème} siècle dans l'étude des jeux de hasard (jeux de dés, de cartes, etc...). Une expérience aléatoire se décrit mathématiquement par la donnée de l'ensemble des résultats possibles de l'expérience en question. Il est de tradition de noter ω un tel résultat (parfois appelé éventualité dans la suite) et de désigner par Ω l'espace de tous ces résultats possibles.

Exemples

1. On lance un dé et l'on s'intéresse au chiffre lu sur sa face supérieure : $\Omega = \{1, 2, 3, 4, 5, 6\}$.
2. On lance deux dés et l'on s'intéresse à leurs faces supérieures :

$$\Omega = \{\omega = (i, j), 1 \leq i \leq 6, 1 \leq j \leq 6\}.$$

3. On lance deux dés et l'on s'intéresse à la somme des deux faces supérieures : $\Omega = \{2, 3, \dots, 12\}$.
4. A l'instant $t = 0$ on met en fonctionnement un appareil et l'on s'intéresse au temps de bon fonctionnement : $\Omega = \mathbb{R}_+$.
5. On observe les points d'impacts de neutrinos sur une cible plane : $\Omega = \mathbb{R}^2$.
6. On observe pendant un intervalle de temps $[t_1, t_2]$ un bruit électronique continu : $\Omega = C^0([t_1, t_2])$

1.1.2 Evénement (d'une expérience aléatoire)

Définition

Un événement est un ensemble de résultats ω d'une expérience aléatoire (caractérisée comme on vient de le voir par la donnée d'un ensemble Ω) possédant une même propriété. On peut ainsi faire correspondre un sous-ensemble (ou une partie) de Ω à chaque événement :

$\text{Ensemble des événements} \xleftrightarrow{\text{en bijection}} \underbrace{\text{Ensemble des parties de } \Omega}_{P(\Omega)}$
--

Si $A \subset \Omega$, on dira que l'événement A se réalise dans l'expérience si l'éventualité ω qui se produit au cours de cette expérience est telle que $\omega \in A$.

Exemple. On reprend l'exemple 1 du paragraphe précédent et l'on considère l'événement :

"le chiffre lu est pair".

Le sous-ensemble de Ω associé à cet événement est $A = \{2, 4, 6\}$.

Si le résultat ω de l'expérience est 2, 4 ou 6, l'événement A est réalisé.

Opérations logiques sur les événements aléatoires

- A tout événement A est associé son contraire ("non A " en logique) noté A^c , qui est réalisé lorsque A ne l'est pas. Dans la représentation des événements par des parties de Ω , l'ensemble correspondant à A^c est donc le complémentaire (dans Ω) de (l'ensemble représentatif de) A .
- Pour tout couple d'événements A_1 et A_2 , l'événement " A_1 et A_2 " est par définition celui qui se réalise si les événements A_1 et A_2 sont réalisés en même temps. Dans l'espace Ω , l'événement " A_1 et A_2 " est représenté par les ω réalisant à la fois A_1 et A_2 c'est-à-dire par l'intersection (des ensembles représentatifs) de A_1 et de A_2 : $A_1 \cap A_2$.
- L'événement impossible (qui ne se réalise jamais) sera noté \emptyset comme l'ensemble vide qui le représente dans Ω . L'équation $A_1 \cap A_2 = \emptyset$, qui exprime que les parties (représentatives de) A_1 et A_2 sont disjointes, signifie alors que les événements A_1 et A_2 sont incompatibles (ils ne se réalisent jamais simultanément au cours d'une même expérience).
- Pour tout couple d'événements A_1 et A_2 , l'événement " A_1 ou A_2 " est celui qui par définition est réalisé si l'un au moins des événements A_1 ou A_2 est réalisé; il s'agit donc ici du "ou" non exclusif. Dans l'espace Ω , l'événement " A_1 ou A_2 " est représenté par la réunion (des ensembles représentatifs) de A_1 et de A_2 : $A_1 \cup A_2$. Lorsque A_1 et A_2 sont incompatibles, et seulement dans ce cas, on notera $A_1 + A_2$ à la place de $A_1 \cup A_2$.
- L'événement certain (qui se réalise toujours) sera noté Ω puisqu'il est réalisé quel que soit le résultat ω de l'expérience aléatoire. L'équation $A_1 \cup A_2 = \Omega$ signifie donc que quel que soit le résultat de l'expérience aléatoire, l'un (au moins) des événements A_1 , A_2 est réalisé.
- L'événement A implique l'événement B si l'événement A ne peut être réalisé sans que B ne le soit aussi. Il revient au même de dire que dans l'espace Ω , tout ω appartenant à A appartient aussi à B , donc que A est contenu dans B . La relation logique d'implication sera donc notée $A \subset B$.

Système exhaustif d'événements

Evidemment, plus de deux événements peuvent être combinés par les opérations précédentes. Ainsi, si $(A_n)_{n \in I}$ est une suite (finie ou infinie mais dénombrable) d'événements, la réunion $\bigcup_{n \in I} A_n$ désignera l'événement

" A_1 ou A_2 ou ...". On conviendra de noter $\sum_{n \in I} A_n$ cette réunion lorsque les événements A_n , $n \in I$, sont deux à deux incompatibles. Ainsi, les relations

$$\left\{ \begin{array}{l} \forall (n, m) \in I^2, n \neq m \implies A_n \cap A_m = \emptyset; \\ \sum_{n \in I} A_n = \Omega, \end{array} \right.$$

qui expriment que les ensembles A_n , $n \in I$, forment une partition de Ω , signifient qu'il est certain qu'un et un seul événement A_n , parmi tous les $n \in I$, sera réalisé. Un tel système d'événements est appelé système exhaustif d'événements.

Exemple. D'une façon générale, pour tout événement A , le système formé des événements A et A^c est exhaustif. Ainsi, dans le cas du lancer d'un dé, les deux événements suivants :

- A = "le chiffre lu est pair" ;
- B = "le chiffre lu est impair",

forment un système exhaustif puisque $B = A^c$.

Résumé

Terminologie probabiliste	Terminologie ensembliste	Notations
Événement	Partie de Ω	A, B, \dots
Événement certain	Espace entier Ω	Ω
Événement impossible	Partie vide	\emptyset
Événement contraire	Partie complémentaire	A^c
Et	Intersection	\cap
Événements incompatibles	Parties disjointes	$A_1 \cap A_2 = \emptyset$
Ou (non exclusif)	Réunion	\cup
Ou (pour des événements 2 à 2 incompatibles seulement)	Somme (pour des parties 2 à 2 disjointes seulement)	$+, \Sigma$
Implication	Inclusion	$A \subset B$
Système exhaustif	Partition de Ω	$\sum_n A_n = \Omega$

1.2 Probabilité

1.2.1 La loi empirique des grands nombres

Faute de pouvoir prédire qu'une expérience donne lieu à un résultat déterminé on veut pouvoir associer à chaque événement un nombre représentant la probabilité qu'il se réalise. Cette probabilité est l'expression d'une propriété de permanence lors de répétitions de l'expérience car une propriété essentielle d'une expérience aléatoire est de pouvoir être répétée indéfiniment (dans des conditions "identiques et indépendantes").

La fréquence de réalisation d'un événement A au cours de N répétitions de l'expérience à laquelle il est lié s'écrit :

$$F_N(A) = \frac{N_A}{N} = \frac{\text{Nombre de réalisations de l'événement } A \text{ au cours de } N \text{ expériences}}{N}.$$

On vérifie expérimentalement que cette fréquence $F_N(A)$ fluctue de moins en moins lorsque N augmente. C'est ce résultat expérimental, connu sous le nom de loi empirique des grands nombres, qui permet de définir mathématiquement la notion idéalisée de fréquence de l'événement A :

$$F(A) = \lim_{N \rightarrow +\infty} F_N(A) = \lim_{N \rightarrow +\infty} \frac{N_A}{N}.$$

On vérifie facilement que

$$\left\{ \begin{array}{l} (i) \text{ Pour tout événement } A, F(A) \geq 0 \text{ et } F(\Omega) = 1; \\ (ii) F\left(\sum_{i \in I} A_i\right) = \sum_{i \in I} F(A_i) \text{ pour des événements } A_i, i \in I \text{ ensemble dénombrable, deux à deux incompatibles.} \end{array} \right.$$

On retiendra ces deux propriétés pour définir axiomatiquement les probabilités (sans faire reposer cette définition sur la loi empirique des grands nombres). La théorie mathématique rigoureuse développée à partir de cette définition s'est révélée à même de rendre compte des phénomènes aléatoires.

1.2.2 Définition et premières propriétés

Pour compléter la description d'une expérience aléatoire Ω , il nous reste à associer à tout événement $A \in P(\Omega)$ sa probabilité $P(A)$ suivant la définition suivante que justifie le paragraphe précédent.

Définition et application au cas discret

Définition. On appelle probabilité sur Ω toute application $P : P(\Omega) \rightarrow [0, 1]$ telle que :

$$\left\{ \begin{array}{l} (i) P(\Omega) = 1 \\ (ii) \text{ Pour toute famille dénombrable } (A_i)_{i \in I} \text{ de } A_i \in P(\Omega) \text{ deux à deux disjointes, on a :} \\ \qquad P\left(\sum_{i \in I} A_i\right) = \sum_{i \in I} P(A_i) \quad (\text{axiome de } \sigma\text{-additivité}) \end{array} \right.$$

L'ensemble Ω muni de la probabilité P , noté (Ω, P) , est appelé espace probabilisé. C'est l'espace fondamental associé à l'expérience aléatoire étudiée.

Exemple : le cas discret. On considère un ensemble Ω dénombrable muni d'une probabilité P . Pour tout $\omega \in \Omega$, posons $p_\omega = P(\{\omega\})$. Alors, pour n'importe quel événement $A \subset \Omega$, l'égalité évidente $A = \sum_{\omega \in A} \{\omega\}$ et l'axiome de σ -additivité impliquent :

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} p_\omega.$$

Ainsi, toute probabilité P pour un espace Ω dénombrable est entièrement déterminée par la donnée de la suite $(p_\omega)_{\omega \in \Omega}$ des probabilités des événements élémentaires (c'est-à-dire réduits à une éventualité). Nous voyons ainsi qu'il existe autant de probabilités sur Ω que de suites $(p_\omega)_{\omega \in \Omega} \in [0, 1]^\Omega$ satisfaisant la condition

$$\sum_{\omega \in \Omega} p_\omega = P(\Omega) = 1.$$

Dans le cas particulier où $\Omega = \mathbb{N}$ par exemple, on peut choisir

$$\forall n \in \Omega, P_\lambda(\{n\}) = \frac{e^{-\lambda} \lambda^n}{n!},$$

où λ est un réel strictement positif fixé. Cette loi de probabilité classique, appelée loi de Poisson sera étudiée au chapitre suivant.

Propriétés

Ce sont des conséquences simples de la σ -additivité de P .

PROPOSITION 1

1. $\forall A \subset \Omega, P(A^c) = 1 - P(A)$
2. $P(\emptyset) = 0$
3. $\forall A, B \subset \Omega, P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Démonstration

1. Comme $A + A^c = \Omega$, la σ -additivité de P implique $P(\Omega) = P(A) + P(A^c) = 1$.
2. On applique l'égalité précédente à $A = \emptyset$ (ce qui implique $A^c = \Omega$).
3. Les événements A et $B - A = \{\omega \in B, \omega \notin A\}$ sont incompatibles et $A \cup B = A + (B - A)$. Par application de la σ -additivité de P , il vient ainsi :

$$P(A \cup B) = P(A) + \underbrace{P(B - A)}_{P(B) - P(A \cap B) \text{ car } B = (B - A) + (A \cap B)} \quad \blacksquare$$

1.3 Probabilité uniforme et dénombrement

1.3.1 Etude d'un ensemble fini d'éventualités équiprobables

On considère ici un espace probabilisé (Ω, P_u) , où l'espace Ω est fini (c'est donc un cas particulier du cas " Ω dénombrable" envisagé plus haut).

On suppose de plus que toutes les éventualités sont équiprobables, c'est-à-dire :

$$\boxed{\forall \omega, \omega' \in \Omega, P_u(\{\omega\}) = P_u(\{\omega'\})}$$

Comme $\sum_{\omega \in \Omega} \{\omega\} = \Omega$ et que $P_u(\Omega) = 1$, la σ -additivité de P_u permet d'écrire :

$$P_u\left(\sum_{\omega \in \Omega} \{\omega\}\right) = \sum_{\omega \in \Omega} P_u(\{\omega\}) = 1.$$

Pour chaque $\omega_0 \in \Omega$, on a donc

$$\sum_{\omega \in \Omega} P_u(\{\omega\}) = \text{Card}(\Omega) \times P_u(\{\omega_0\}),$$

soit finalement :

$$\boxed{P_u(\{\omega_0\}) = \frac{1}{\text{Card}(\Omega)}}$$

Cette expression justifie que cette probabilité soit appelée probabilité uniforme (d'où le u en indice). Cette probabilité uniforme régit la plupart des modèles finis de jeux de hasard et d'ailleurs, lorsque dans le langage courant on dit que l'on "choisit au hasard" un élément d'un ensemble fini Ω , on sous-entend le plus souvent que les divers choix possibles sont équiprobables et donc que l'espace Ω est muni de la probabilité uniforme.

Plus généralement, le calcul de la probabilité uniforme d'un événement $A \in \Omega$ se déduit facilement de l'expression précédente, puisque :

$$P_u(A) = \sum_{\omega \in A} P_u(\{\omega\}) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}$$

Nous voyons ainsi que le calcul de la probabilité uniforme d'un événement A se ramène à celui du cardinal du sous-ensemble de Ω associé à A , ce qui revient finalement à dénombrer les éléments de cet ensemble. Lorsqu'il peut être effectué, ce calcul est le plus souvent de nature combinatoire, c'est pourquoi il utilise généralement les notions classiques (que vous connaissez déjà normalement) de combinatoire qui sont rappelées dans le paragraphe suivant.

1.3.2 Dénombrément

Soient $E = \{x_1, x_2, \dots, x_n\}$ un ensemble fini à n ($n \geq 1$) éléments et p un entier de $\{1, 2, \dots, n\}$.

p -arrangements sans répétition de n éléments

Définition. On appelle p -arrangements sans répétition de n éléments de E toute liste ordonnée du type

$$(x_{i_1}, x_{i_2}, \dots, x_{i_p}),$$

où les indices i_1, i_2, \dots, i_p sont deux à deux distincts (ce qui implique que les éléments d'une telle liste sont également deux à deux distincts).

Ainsi, deux p -arrangements diffèrent :

- soit par la nature de leurs éléments ;
- soit par l'ordre dans lequel sont rangés ces éléments.

Il est de tradition de noter A_n^p le nombre (ou cardinal) de tous ces p -arrangements sans répétition de n éléments :

$$A_n^p = A_n^{p-1} \times [n - (p - 1)].$$

D'où, successivement :

$$\begin{cases} A_n^{p-1} = A_n^{p-2} \times [n - (p - 2)] \\ A_n^{p-2} = A_n^{p-3} \times [n - (p - 3)] \\ \vdots \\ A_n^2 = A_n^1 \times (n - 1) \\ A_n^1 = n \end{cases}$$

En multipliant membre à membre ces égalités, on obtient finalement :

$$A_n^p = n(n - 1)(n - 2) \dots (n - p + 1)$$

Si l'on introduit la quantité factorielle n , classiquement notée $n!$ et définie pour tout entier $n \geq 1$ par

$$n! = n \times (n - 1) \times (n - 2) \times \dots \times 3 \times 2 \times 1,$$

avec la convention $0! = 1$, l'égalité précédente s'écrit donc simplement :

$$A_n^p = \frac{n!}{(n - p)!}$$

Exemple. Répartir 3 ordinateurs différents entre 6 étudiants revient à choisir 3 personnes parmi les 6 susceptibles d'obtenir une machine. Comme les 3 machines sont différentes, on doit tenir compte de l'ordre dans lequel ces 3 étudiants sont choisis (le premier choisi obtient la machine n°1, le second la machine n°2, etc...). C'est pourquoi, le nombre de possibilités de répartir ces 3 machines différentes entre les 6 étudiants est $A_6^3 = 120$.

Cas particulier : les permutations. Si $p = n$, A_n^p désigne alors le nombre de permutations (des n éléments) de E .

Deux permutations ne diffèrent donc que par l'ordre dans lequel les n objets sont rangés.

D'après la formule précédente, le nombre de permutations de E est :

$$p_n = n!$$

Ainsi, dans le cas où $n = 3$, les différentes permutations de $\{x_1, x_2, x_3\}$ sont, outre $\{x_1, x_2, x_3\}$ lui-même, $\{x_1, x_3, x_2\}$, $\{x_2, x_1, x_3\}$, $\{x_2, x_3, x_1\}$, $\{x_3, x_1, x_2\}$ et $\{x_3, x_2, x_1\}$. Ceci montre qu'il y a bien $3! = 6$ possibilités de répartir 3 machines différentes entre 3 étudiants.

Combinaisons de n éléments pris p à p

Définition. Toute sous partie $F = \{x_{i_1}, x_{i_2}, \dots, x_{i_p}\}$ de E contenant p éléments, dans laquelle ces éléments sont distincts et énoncés dans un ordre quelconque, est appelée combinaison de n éléments pris p à p .

On note généralement C_n^p leur nombre.

Les éléments de chacune de ces combinaisons pouvant être ordonnés de $p!$ façons différentes, on peut écrire

$$p!C_n^p = A_n^p,$$

ce qui implique finalement :

$$C_n^p = \frac{n!}{p!(n-p)!}$$

Lorsque $p = 0$, il n'y a qu'un seul sous-ensemble de E ne contenant aucun élément (c'est l'ensemble vide), c'est pourquoi on convient de poser :

$$C_n^0 = 1$$

Exemple. Répartir 3 ordinateurs identiques entre 6 étudiants revient à nouveau à choisir 3 personnes parmi les 6 susceptibles d'obtenir une machine. Mais comme les 3 machines sont, cette fois, identiques, l'ordre dans lequel ces 3 étudiants sont choisis n'importe pas. C'est pourquoi, le nombre de possibilités de répartir ces 3 machines identiques entre les 6 étudiants est $C_6^3 = 20$.

Propriétés des C_n^p . Les propriétés essentielles à retenir sont les suivantes :

$$\begin{array}{ll} C_n^p \in \mathbb{N} & (p \in \{0, 1, \dots, n\}) \\ C_n^p = C_n^{n-p} & (p \in \{0, 1, \dots, n\}) \\ C_n^p = C_{n-1}^p + C_{n-1}^{p-1} & (p \in \{1, 2, \dots, n-1\}) \quad (\text{formule de Pascal}) \end{array}$$

La formule de Pascal est à la base du calcul du triangle de Pascal.

Une application : la formule du binôme de Newton. Quels que soient les réels a, b et l'entier $n \in \mathbb{N}^*$, on démontre (par récurrence sur n) que

$$(a+b)^n = \underbrace{a^n + C_n^1 a^{n-1} b + C_n^2 a^{n-2} b^2 + \dots + C_n^{n-1} a b^{n-1} + b^n}_{\sum_{p=0}^n C_n^p a^{n-p} b^p}$$

1.4 Conditionnement et indépendance

1.4.1 Probabilités conditionnelles

Exemple introductif

On considère une famille ayant deux enfants. Chaque enfant est désigné par F lorsque c'est une fille et par G si c'est un garçon. L'espace Ω dans lequel on se place est donc :

$$\Omega = \{(F, F), \underbrace{(F, G)}_{\text{la fille est née avant le garçon}}, (G, F), (G, G)\}.$$

On suppose qu'il y a équiprobabilité entre chacune des 4 éventualités composant Ω .

Définissons les deux événements suivants :

- $A =$ "La famille a deux garçons" : $P(A) = \frac{1}{4}$;
- $B =$ "La famille a au moins un garçon" : $P(B) = \frac{3}{4}$.

On suppose que B est réalisé.

Les éventualités qui réalisent A sont donc dans $A \cap B$, et :

$$\text{La probabilité de } A \text{ (lorsque } B \text{ est réalisé)} = \frac{\text{nombre d'éventualités de } A \cap B}{\text{nombre d'éventualités de } B} = \frac{1}{3}.$$

Et l'on constate sur cet exemple que $\frac{P(A \cap B)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}$.

1.4.2 Définition et propriétés

Définition. Soit (Ω, P) un espace probabilisé. Soient A et B deux événements tels que $P(B) \neq 0$.

On appelle probabilité conditionnelle de A sachant B la quantité

$$P(A/B) = \frac{P(A \cap B)}{P(B)}.$$

On remarque que $A \mapsto P(A/B)$ est une nouvelle probabilité sur Ω (notée $P(\cdot/B)$) qui vaut 0 sur tous les événements incompatibles avec B .

Exemple. Considérons la probabilité uniforme P_u définie sur un espace fini Ω par

$$P_u(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}.$$

Si l'événement B n'est pas impossible ($B \neq \emptyset$), sa probabilité $P_u(B)$ est strictement positive et pour tout événement A

$$P_u(A/B) = \frac{P_u(A \cap B)}{P_u(B)} = \frac{\text{Card}(A \cap B)}{\text{Card}(B)}.$$

Si $A \subset B$, cette formule se réduit à

$$P_u(A/B) = \frac{\text{Card}(A)}{\text{Card}(B)},$$

et montre que pour ces événements A la probabilité conditionnelle $P_u(. / B)$ se réduit à la probabilité uniforme sur B .

Ce résultat, que nous avons déduit de la définition est bien conforme à l'intuition.

THÉORÈME 1 (Formule de Bayes) – Soit $(B_n)_{n \in I}$ un système exhaustif d'événements tels que

$$\forall n \in I, P(B_n) \neq 0.$$

Alors :

$$\forall A \subset \Omega, P(A) = \sum_{n \in I} P(A/B_n) \times P(B_n).$$

Démonstration. Comme les B_n , $n \in I$, sont deux à deux incompatibles, il en est de même des $A \cap B_n$, $n \in I$. Ainsi, $A = \sum_{n \in I} (A \cap B_n)$ et

$$P(A) = \sum_{n \in I} \underbrace{P(A \cap B_n)}_{P(A/B_n) \times P(B_n)} . \blacksquare$$

1.4.3 Événements indépendants

Définition

Comme on vient de le voir, $P(A/B) \neq P(A)$ en général (lorsque A "dépend" de B). En fait, on a l'équivalence :

$$P(A/B) = P(A) \iff P(A \cap B) = P(A) \times P(B).$$

Définition. Dans un espace probabilisé (Ω, P) , deux événements A et B sont dits indépendants si

$$\boxed{P(A \cap B) = P(A) \times P(B)}$$

Remarques. 1) Un événement A est toujours indépendant de \emptyset et de Ω .

2) Si deux événements A et B sont indépendants, alors il en est de même de A^c et B , de A et B^c et de A^c et B^c .

Exemple et généralisation

Exemple. L'expérience consiste à lancer deux dés :

$$\Omega = \{\omega = (i, j), 1 \leq i \leq 6, 1 \leq j \leq 6\}.$$

On s'intéresse aux deux événements suivants :

- $A =$ "le premier dé donne un chiffre pair" : $P(A) = \frac{1}{2}$;
- $B =$ "le deuxième dé donne un chiffre pair" : $P(B) = \frac{1}{2}$.

On voit facilement que $P(A \cap B) = \frac{1}{4}$. Par suite, $P(A \cap B) = P(A) \times P(B)$ donc A et B sont indépendants. Considérons maintenant l'événement :

- $C =$ "la somme des deux chiffres lus est paire" : $P(C) = \frac{1}{2}$.

On a $P(A \cap C) = \frac{1}{4} = P(A) \times P(C)$ donc A et C sont indépendants.

Par raisons de symétrie, B et C sont indépendants, donc A, B et C sont deux à deux indépendants.

Et pourtant,

$$P(A \cap B \cap C) \neq P(A) \times P(B) \times P(C).$$

En effet, $A \cap B \subset C$ donc $A \cap B \cap C = A \cap B$ et $P(A \cap B \cap C) = \frac{1}{4} \neq P(A) \times P(B) \times P(C)$.

Cet exemple montre que trois événements deux à deux indépendants ne sont pas forcément "globalement" indépendants lorsqu'ils sont pris "dans leur ensemble". Cela conduit à généraliser la définition précédente au cas de plus de deux événements comme dans la définition suivante.

Définition. Dans un espace probabilisé (Ω, P) , n ($n \geq 3$) événements A_1, A_2, \dots, A_n sont dits indépendants dans leur ensemble, si :

$$\forall k, 1 \leq k \leq n, P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \prod_{j=1}^k P(A_{i_j}).$$

1.5 Un exemple récapitulatif

1.5.1 Description de l'expérience et du problème

On dispose de deux dés supposés parfaits (disons le dé a et le dé b).

On jette ces deux dés et on lit le nombre apparaissant sur la face supérieure du dé a puis celui apparaissant sur la face supérieure du dé b .

L'entier n appartenant à $\{1, 2, 3, 4, 5, 6\}$, on va essayer de répondre aux questions suivantes :

1. Quelle est la probabilité que le premier chiffre lu soit n ? Et celle que le deuxième chiffre lu soit n ?
2. Quelle est la probabilité d'obtenir un double ?
3. Quelle est la probabilité d'obtenir au moins un 6 ?
4. Quelle est la probabilité d'obtenir un double ou au moins un 6 ?

1.5.2 Solution

On va voir comment le formalisme mathématique introduit dans ce chapitre permet de résoudre ce problème. Avant toute chose, il faut définir l'espace Ω des observations. Comme on l'a déjà vu dans l'exemple 2 du §1.1 :

$$\Omega = \{\omega_{i,j} = (i, j), 1 \leq i \leq 6, 1 \leq j \leq 6\}.$$

Par suite, $\text{Card}(\Omega) = 6 \times 6 = 36$. Ensuite, comme les deux dés sont supposés parfaits, toutes les éventualités $\omega_{i,j}$, pour i et j dans $\{1, 2, 3, 4, 5, 6\}$ sont équiprobables, donc :

$$P(\omega_{i,j}) = \frac{1}{\text{Card}(\Omega)} = \frac{1}{36}.$$

Ceci montre que le cadre mathématique adapté à la résolution de ce problème est l'espace probabilisé (Ω, P) où P désigne la probabilité uniforme.

1. Quelle est la probabilité que le premier chiffre lu soit n ? Et celle que le deuxième chiffre lu soit n ?
Soit A_n l'événement : "le premier chiffre lu est n ". Il est clair que $A_n = \{\omega_{n,j}, 1 \leq j \leq 6\}$. Par suite, $\text{Card}(A_n) = 6$, donc, comme P est la probabilité uniforme :

$$P(A_n) = \frac{\text{Card}(A_n)}{\text{Card}(\Omega)} = \frac{1}{6}.$$

De même, en posant $B_n =$ "le deuxième chiffre lu est n " et en appliquant un raisonnement totalement similaire, on obtiendrait

$$P(B_n) = \frac{1}{6}.$$

En fait, on peut procéder un peu différemment pour calculer $P(B_n)$ (bien que ce soit un peu artificiel, mais c'est pour utiliser les notions introduites plus haut).

En effet, $(A_i)_{1 \leq i \leq 6}$ étant un système exhaustif d'événements tels que $P(A_i) = \frac{1}{6} \neq 0$ pour tout $i \in \{1, 2, 3, 4, 5, 6\}$, la formule de Bayes s'applique :

$$P(B_n) = \sum_{i=1}^6 P(B_n/A_i) \times \underbrace{P(A_i)}_{\frac{1}{6}}.$$

Or, $P(B_n/A_i) = \frac{1}{6}$ (car $B_n \cap A_i = \{\omega_{i,n}\}$ et il résulte de l'exemple de la section 1.4.2 que

$$P_u(B_n/A_i) = \frac{\text{Card}(B_n \cap A_i)}{\text{Card}(A_i)} \text{ et donc que } P(B_n) = \sum_{i=1}^6 \frac{1}{6} \times \frac{1}{6} = 6 \times \left(\frac{1}{6}\right)^2 = \frac{1}{6}.$$

Remarque. Ces résultats confirment l'impression (totalement intuitive) que les événements A_n et B_m , pour $1 \leq n, m \leq 6$ fixés, sont indépendants. En effet, $A_n \cap B_m = \{\omega_{n,m}\}$ donc

$$P(A_n \cap B_m) = \frac{1}{36} = P(A_n) \times P(B_m).$$

2. Quelle est la probabilité d'obtenir un double ?

Soit $C = \text{"les deux chiffres lus sont égaux"}$. On a en fait $C = \sum_{n=1}^6 (A_n \cap B_n)$.

En effet, les événements de la suite $(A_n \cap B_n)_{1 \leq n \leq 6}$ sont deux à deux incompatibles puisque $A_n \cap B_n = \{\omega_{n,n}\}$. On en déduit donc :

$$P(C) = \sum_{n=1}^6 \underbrace{P(\{\omega_{n,n}\})}_{\frac{1}{36}} = \frac{1}{6}.$$

3. Quelle est la probabilité d'obtenir au moins un 6 ?

Soit $D = \text{"l'un au moins des chiffres lus est 6"}$. Comme $D = A_6 \cup B_6$, on a :

$$P(D) = \underbrace{P(A_6)}_{\frac{1}{6}} + \underbrace{P(B_6)}_{\frac{1}{6}} - \underbrace{P(A_6 \cap B_6)}_{P(\{\omega_{6,6}\}) = \frac{1}{36}} = \frac{11}{36}.$$

4. Quelle est la probabilité d'obtenir un double ou au moins un 6 ?

Posons $E = \text{"les chiffres lus forment un double ou contiennent au moins un 6"}$. Comme $E = C \cup D$, on a cette fois

$$P(E) = \underbrace{P(C)}_{\frac{1}{6}} + \underbrace{P(D)}_{\frac{11}{36}} - P(C \cap D).$$

Or, $C \cap D = \{\omega_{6,6}\}$, donc $P(C \cap D) = \frac{1}{36}$ et $P(E) = \frac{4}{9}$.

Chapitre 2

Variables aléatoires discrètes

Une variable aléatoire (v.a. en abrégé) est définie par référence à une expérience aléatoire comme une variable X dont la valeur dépend du résultat ω de cette expérience. A proprement parler une v.a. est donc une fonction définie sur l'espace Ω associé à l'expérience aléatoire.

Dans l'exemple du jet de deux dés, la somme S des chiffres lus sur la face supérieure de chacun des dés est une v.a. à valeurs entières positives égale à la fonction de $\omega \in \Omega = \{1, 2, \dots, 6\}^2$ définie par

$$S(\omega) = n_1 + n_2 \text{ si } \omega = (n_1, n_2).$$

On ne considère dans ce chapitre que des v.a. discrètes, c'est-à-dire des v.a. dont l'ensemble des valeurs, généralement noté E , est dénombrable :

$$X : \Omega \longrightarrow E.$$

Naturellement si $E \subset \mathbb{Z}$, la v.a. sera dite entière.

Si $E \subset \mathbb{Z}^d$ ($d \geq 2$) nous dirons que X est un vecteur aléatoire entier et l'on désignera ses coordonnées par X_1, X_2, \dots, X_d (chacune de ses coordonnées est une v.a. entière).

2.1 Loi de probabilité

Soient (Ω, P) un espace probabilisé et $X : \Omega \rightarrow E$ une v.a. discrète.

2.1.1 Définition

Pour tout $x \in E$, $\{X = x\} \stackrel{\text{def}}{=} \{\omega \in \Omega, X(\omega) = x\}$ est une partie de Ω , donc un événement. On peut donc calculer $P(\{X = x\})$ qui sera noté plus simplement $P(X = x)$ ou bien $p_X(x)$ dans la suite.

Définition. La famille des nombres $\{P(X = x), x \in E\}$ est appelée loi de probabilité de X .

L'intérêt de cette loi de probabilité est de permettre de calculer directement, sans passer par Ω , la probabilité de tout événement "ne dépendant que de X ". Un tel événement correspondant à une partie F de E est en effet de la forme

$$A = \{X \in F\} \stackrel{\text{def}}{=} \{\omega \in \Omega, X(\omega) \in F\} = \sum_{x \in F} \{X = x\}$$

puisque les événements $\{X = x\}$, $x \in F$, sont deux à deux disjoints.

Comme E est dénombrable, F l'est aussi et la σ -additivité de P montre alors que la probabilité de A vaut

$$P(A) = P(X \in F) = \sum_{x \in F} \underbrace{P(X = x)}_{p_X(x)}.$$

Exemple. Pour la somme S des chiffres lus en jetant deux dés supposés parfaits, le dénombrement des cas possibles donne, compte tenu du fait que $\text{Card}(\Omega) = 6^2 = 36$:

x	2	3	4	5	6	7	8	9	10	11	12
$p_S(x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Par suite,

$$P(S \leq 5) = P(S \in \{2, 3, 4, 5\}) = p_S(2) + p_S(3) + p_S(4) + p_S(5) = \frac{10}{36}$$

sans qu'il soit nécessaire de dénombrer les cas favorables à $S \leq 5$.

2.1.2 Transformation de v.a.

Dans l'étude de modèles probabilistes, on est fréquemment amené à transformer des v.a. en d'autres v.a. et donc à appliquer le résultat suivant :

PROPOSITION 1 – Si $X : \Omega \rightarrow E$ est une v.a. discrète sur un espace probabilisé (Ω, P) et si $f : E \rightarrow E'$ est une application arbitraire de l'espace dénombrable E dans un second espace dénombrable E' , alors

$$Y = f \circ X,$$

également notée $Y = f(X)$, est une v.a. discrète dont la loi de probabilité $\{p_Y(y), y \in E'\}$ se déduit de celle $\{p_X(x), x \in E\}$ de X par la formule

$$p_Y(y) = \sum_{\{x \in E, f(x)=y\}} p_X(x)$$

Démonstration. Comme pour tout $y \in E'$

$$\{Y = y\} = \sum_{\{x \in E, f(x)=y\}} \{X = x\},$$

et que la somme du second membre est dénombrable (puisque E l'est) il est clair que Y est une v.a. discrète et que sa loi est donnée par la formule précédente. ■

Attention : la deuxième partie de la proposition 1 ne se généralise pas telle quelle à des fonctions de plusieurs variables !

En effet, si $X_j : \Omega \rightarrow E_j$ ($1 \leq j \leq k$) sont k v.a. discrètes ($k \geq 2$) définies sur le même espace probabilisé (Ω, P) , toute fonction $f : E_1 \times E_2 \times \dots \times E_k \rightarrow E$ où E est un ensemble dénombrable donne encore une v.a. discrète par composition, soit $Y = f(X_1, X_2, \dots, X_k)$ puisque

$$\{Y = y\} = \sum_{\{(x_1, x_2, \dots, x_k) \in E_1 \times E_2 \times \dots \times E_k, f(x_1, x_2, \dots, x_k) = y\}} \{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \dots \cap \{X_k = x_k\}$$

est un événement pour tout $y \in E$.

Mais les lois de probabilité de chacune des variables X_j ($1 \leq j \leq k$), c'est-à-dire la suite des nombres $\{P(X_j = x), x \in E_j\}$ pour $j = 1, 2, \dots, k$, ne suffisent pas en général à déterminer la loi de Y (c'est-à-dire les $P(Y = y)$ pour tout $y \in E$). En effet, pour calculer cette loi, il faut connaître les probabilités $P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k)$ c'est-à-dire la loi du vecteur aléatoire (X_1, X_2, \dots, X_k) et considérer donc que Y est une fonction d'une seule variable, la variable vectorielle (X_1, X_2, \dots, X_k) , comme c'est le cas dans l'exemple suivant.

Exemple. Lors d'un tirage de cartes d'une même couleur (il s'agit de 13 cartes distinctes : as, roi, dame, valet, 10, 9, 8, ..., 2) un joueur reçoit successivement 2 cartes de hauteurs respectives X_1 et X_2 ($X = 13$ pour un roi, $X = 1$ pour un as, etc...) et s'intéresse à la plus haute des deux cartes, donc à $Y = \max(X_1, X_2)$. Evidemment, X_1 et X_2 ont des lois uniformes :

$$\forall x \in \{1, 2, \dots, 13\}, P(X_i = x) = \frac{1}{13} \quad (i = 1, 2),$$

et la loi du vecteur aléatoire (X_1, X_2) est donnée par

$$\forall (x_1, x_2) \in \{1, 2, \dots, 13\}^2, P(\underbrace{X_1 = x_1, X_2 = x_2}_{\{X_1=x_1\} \cap \{X_2=x_2\}}) = \begin{cases} \frac{1}{13 \times 12} & \text{si } x_1 \neq x_2; \\ 0 & \text{si } x_1 = x_2, \end{cases}$$

si le tirage est effectué sans remise de la première carte (on aurait simplement $P(X_1 = x_1, X_2 = x_2) = \frac{1}{13^2}$ pour un tirage avec remise).

Ainsi,

$$\begin{aligned} P(Y = 13) &= \sum_{\{(x_1, x_2) \in \{1, 2, \dots, 13\}^2, \max(x_1, x_2) = 13\}} P(X_1 = x_1, X_2 = x_2) \\ &= \sum_{\{(x_1, x_2) \in \{1, 2, \dots, 13\}^2, x_1 \neq x_2, \max(x_1, x_2) = 13\}} P(X_1 = x_1, X_2 = x_2) \\ &= 2 \times 12 \times \frac{1}{13 \times 12} = \frac{2}{13}. \end{aligned}$$

Remarque : formule des lois marginales. Si (X, Y) est un vecteur aléatoire discret de coordonnées X et Y (ce sont donc deux v.a. discrètes) dont la loi de probabilité est notée

$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

alors la loi de X est donnée par la somme

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

2.1.3 Loi de probabilité conditionnelle

Appliquons aux v.a. discrètes la définition des probabilités conditionnelles. Si B est un événement de probabilité $P(B) > 0$ ($X : \Omega \rightarrow E$ désignant toujours une v.a. discrète) nous appellerons loi de probabilité de X conditionnelle en B la loi de probabilité q sur E définie par

$$q(x) = P(X = x/B) = \frac{P(\{X = x\} \cap B)}{P(B)}.$$

Cette notion est surtout importante pour des événements B associés à une seconde variable Y comme dans la définition suivante.

Définition. Si $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$ sont deux v.a. discrètes définies sur le même espace probabilisé (Ω, P) , la loi conditionnelle de X en Y est définie comme la famille de lois de probabilité $p_{X/Y}(\cdot/y)$ sur E indexée par les valeurs $y \in F$ de Y telles que $P(Y = y) > 0$, qui est donnée par :

$$p_{X/Y}(x/y) = P(X = x/Y = y) \quad (x \in E)$$

Exemple : lois géométriques des durées de vie. Soit T une v.a. entière strictement positive représentant une durée (exprimée en secondes, par exemple) de vie (ou de bon fonctionnement) d'un objet électronique, ayant débuté à l'instant 0. Plaçons nous ensuite à un instant $t > 0$ et supposons que nous observons à cet instant que l'objet en question fonctionne toujours correctement. Conditionnellement en cette information $\{T > t\}$, la durée de bon fonctionnement résiduelle, $T - t$, possède une loi notée q_t sur \mathbb{N}^* que l'on peut calculer à partir de la loi initiale p_T :

$$\forall n \geq 1, q_t(n) = P(T - t = n/T > t) = \frac{P(T = t + n)}{P(T > t)} = \frac{p_T(t + n)}{\sum_{m>t} p_T(m)}.$$

Bien entendu pour $t = 0$, q_t est égale à la loi de probabilité de T .

Si l'on suppose que l'objet ne s'abîme pas en fonctionnant (hypothèse courante lorsqu'il s'agit d'objets électroniques), il est naturel de supposer

$$q_t(n) = p_T(n).$$

Par suite, nous avons :

$$\forall t \in \mathbb{N}, \forall t \in \mathbb{N}^*, p_T(t + n) = p_T(n) \times \left[\sum_{m>t} p_T(m) \right].$$

Par sommation de cette égalité sur $n > s$ pour $s \in \mathbb{N}^*$, on obtient

$$\sum_{n>s} p_T(t + n) = \left[\sum_{n>s} p_T(n) \right] \times \left[\sum_{m>t} p_T(m) \right],$$

soit, en posant $S(k) = \sum_{m>k} p_T(m)$:

$$\forall (t, s) \in (\mathbb{N}^*)^2, S(t + s) = S(s) \times S(t).$$

Par suite, $S(s) = a^s$ avec $a > 0$ (car $S(\cdot) > 0$) et $a < 1$ (car $S(s) \xrightarrow{s \rightarrow \infty} 0$).

Enfin, puisque $p_T(n) = S(n) - S(n - 1)$, la loi de T sous l'hypothèse précédente est nécessairement l'une des lois géométriques notées $G(a)$,

$$p_T(n) = (1 - a)a^{n-1} \quad (n \in \mathbb{N}^*)$$

dont le paramètre a est un réel de $]0, 1[$.

Pratiquement, on vérifie que ces lois géométriques décrivent bien le temps de bon fonctionnement des appareils

électroniques.

La formule de définition des probabilités conditionnelles est souvent utilisée sous sa forme multiplicative :

$$P(A \cap B) = P(A)P(B/A) \text{ si } P(A) \neq 0.$$

En prenant $A = \{X = x\}$ dans cette égalité puis en sommant sur x , on démontre le célèbre théorème de Bayes.

THÉORÈME (de Bayes) – Pour tout événement B et toute v.a. discrète $X : \Omega \rightarrow E$ telle que $P(X = x) \neq 0$ pour chaque $x \in E$, on a

$$P(B) = \sum_{x \in E} P(X = x)P(B/X = x).$$

2.1.4 V.a. indépendantes

Définition. Deux v.a. discrètes $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$ définies sur le même espace probabilisé (Ω, P) sont dites indépendantes si

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

pour tout $x \in E$ et tout $y \in F$. Cette égalité s'écrit encore

$$\boxed{p_{X,Y}(x, y) = p_X(x)p_Y(y)}$$

Remarque. Si X et Y sont indépendantes alors

$$P(X = x/Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = P(X = x)$$

pour tout $x \in E$ et tout $y \in F$ tel que $P(Y = y) \neq 0$.

Ceci traduit le fait que la connaissance de la valeur y prise par l'une des deux variables (la variable Y dans le cas envisagé ici), n'apporte aucune information sur la probabilité que X soit égal à x .

Généralisation. La définition précédente se généralise aisément au cas de n ($n \geq 2$) v.a. discrètes $X_m : \Omega \rightarrow E_m$ définies sur le même espace probabilisé (Ω, P) . Ces n v.a. seront dites indépendantes si

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_1) \times P(X_2 = x_2) \times \dots \times P(X_n = x_n).$$

2.2 Grandeurs caractéristiques d'une v.a. réelle discrète

Soit (Ω, P) un espace probabilisé.

Une v.a. discrète $X : \Omega \rightarrow E$ est dite réelle si ses valeurs possibles sont réelles, c'est-à-dire si $E \subset \mathbb{R}$. Toute v.a. entière $X : \Omega \rightarrow \mathbb{Z}$ par exemple est de ce type.

2.2.1 Espérance

Définition et première interprétation

Définition. L'espérance d'une v.a. réelle discrète $X : \Omega \rightarrow E$ ($E \subset \mathbb{R}$), notée $E(X)$, est définie à partir de sa loi de probabilité par :

$$E(X) = \sum_{x \in E} xp_X(x)$$

pourvu que la série du second membre soit absolument convergente.

Exemple. Soit X une v.a. suivant une loi géométrique de paramètre $a \in]0, 1[$, ce que l'on note $X \sim G(a)$. Alors, pour tout $n \in \mathbb{N}^*$, $P(X = n) = (1 - a)a^{n-1}$ donc l'espérance de X s'écrit :

$$E(X) = \sum_{n \in \mathbb{N}^*} n(1 - a)a^{n-1} = (1 - a) \sum_{n \geq 1} \underbrace{na^{n-1}}_{(a^n)'}$$

Or le rayon de convergence de la série entière (de la variable a) $\{a^n\}_{n \geq 0}$ est égal à 1, donc

$$\sum_{n \geq 1} na^{n-1} = \left(\sum_{n \geq 0} a^n \right)' = \left(\frac{1}{1 - a} \right)' = \frac{1}{(1 - a)^2}$$

et

$$E(X) = \frac{1}{1 - a}.$$

Interprétation. Comme $\sum_{x \in E} p_X(x) = 1$, la définition précédente s'écrit aussi

$$E(X) = \frac{\sum_{x \in E} xp_X(x)}{\sum_{x \in E} p_X(x)},$$

ce qui montre que l'espérance de X s'interprète comme l'isobarycentre (ou le centre de gravité) de la répartition de masses discrètes $p_X(x)$ aux points $x \in E$ de \mathbb{R} .

C'est pourquoi, la v.a. X est dite centrée lorsque $E(X) = 0$.

En particulier la v.a. $Y = X - E(X)$ est toujours centrée.

Remarque. Soient $A \subset \Omega$ et la fonction indicatrice de (support) A , notée $\mathbf{1}_A$, et définie par :

$$\mathbf{1}_A : \Omega \longrightarrow \{0, 1\}$$

$$\omega \longmapsto \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}$$

Comme $\mathbf{1}_A$ est une v.a. discrète réelle de Ω dans $E = \{0, 1\}$, nous pouvons calculer son espérance :

$$E(\mathbf{1}_A) = 0 \times P(\mathbf{1}_A = 0) + 1 \times P(\mathbf{1}_A = 1) = P(\mathbf{1}_A = 1) = P(A).$$

Autrement dit, la probabilité de A est l'espérance de la v.a. $\mathbf{1}_A$.

Autre interprétation de l'espérance

Considérons une v.a. réelle discrète $X : \Omega \rightarrow \{x_1, x_2, \dots, x_k\}$.

Afin d'estimer la loi de probabilité de X , on s'intéresse à la suite X_1, X_2, \dots, X_N des v.a. obtenues en répétant N ($N \geq 1$) fois de suite l'expérience aléatoire à laquelle X est liée.

D'après la loi empirique des grands nombres

$$P(X = x_i) = \lim_{N \rightarrow +\infty} \left(\frac{1}{N} \sum_{n=1}^N \mathbf{1}_{\{X_n = x_i\}} \right),$$

pour tout x_i dans E ($1 \leq i \leq k$). En d'autres termes, lorsque N est grand :

$$P(X = x_i) \approx \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{\{X_n = x_i\}}.$$

Or, pour tout $n \in \{1, 2, \dots, N\}$, $X_n = \sum_{i=1}^k x_i \mathbf{1}_{\{X_n = x_i\}}$ donc

$$\frac{1}{N} \sum_{n=1}^N X_n = \frac{1}{N} \sum_{n=1}^N \left(\sum_{i=1}^k x_i \mathbf{1}_{\{X_n = x_i\}} \right) = \sum_{i=1}^k x_i \underbrace{\left[\frac{1}{N} \sum_{n=1}^N \mathbf{1}_{\{X_n = x_i\}} \right]}_{\approx P(X=x_i)} \approx E(X).$$

Ainsi, d'après la loi empirique des grands nombres, la moyenne des N observations X_1, X_2, \dots, X_N estime l'espérance $E(X)$.

Ceci explique que $E(X)$ soit également appelée moyenne de X .

Exemple. Dans le cas de la somme S des chiffres lus sur la face supérieure de deux dés,

$$\begin{aligned} E(S) = \sum_{k=2}^{12} k p_S(k) &= \frac{1}{36} [2 \times 1 + 3 \times 2 + 4 \times 3 + 5 \times 4 + 6 \times 5 + 7 \times 6 \\ &\quad + 8 \times 5 + 9 \times 4 + 10 \times 3 + 11 \times 2 + 12 \times 1] = 7 \end{aligned}$$

donc la valeur moyenne de N observations de ce type est proche de 7 dès que N est assez grand.

Propriétés utiles

Expression de $E(f(X))$. Si $X : \Omega \rightarrow E$ désigne toujours une v.a. réelle discrète sur un espace probabilisé (Ω, P) et que $f : E \rightarrow E'$ est une application arbitraire de l'espace dénombrable E dans un second espace dénombrable E' , l'espérance de $Y = f(X)$ (dont l'ensemble des valeurs possibles est $f(E)$) s'écrit par définition :

$$E(Y) = \sum_{y \in f(E)} y P(Y = y).$$

Or, $\{Y = y\} = \sum_{\{x \in E, f(x)=y\}} \{X = x\}$, donc

$$E(Y) = \sum_{y \in f(E)} y \left(\sum_{\{x \in E, f(x)=y\}} P(X = x) \right) = \underbrace{\sum_{y \in f(E)} \sum_{\{x \in E, f(x)=y\}} f(x)P(X = x)}_{\sum_{x \in E}}$$

soit

$$E(f(X)) = \sum_{x \in E} f(x)P(X = x)$$

Ce résultat est totalement intuitif compte tenu de l'interprétation précédente de l'espérance.

Application. Reprenons l'exemple d'une v.a. $X \rightsquigarrow G(a)$ ($a \in]0, 1[$). D'après la formule précédente,

$$E(X^2) = \sum_{n \geq 1} n^2 \underbrace{P(X = n)}_{(1-a)a^{n-1}} = (1-a) \sum_{n \geq 1} \underbrace{n^2}_{n(n-1)+n} a^{n-1} = (1-a) \left[a \underbrace{\sum_{n \geq 2} n(n-1)a^{n-2}}_{\left(\frac{1}{1-a}\right)''} + \underbrace{\sum_{n \geq 1} na^{n-1}}_{\left(\frac{1}{1-a}\right)'} \right].$$

Par suite,

$$E(X^2) = (1-a) \left[\frac{2a}{(1-a)^3} + \frac{1}{(1-a)^2} \right] = \frac{1+a}{(1-a)^2}.$$

Linéarité. En s'appuyant une nouvelle fois sur l'interprétation précédente de l'espérance, on obtient :

PROPOSITION 2 – Si $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow E'$ sont deux v.a. réelles discrètes définies sur le même espace probabilisé (Ω, P) , alors

$$E(X + Y) = E(X) + E(Y)$$

Conséquence. $E(X - E(X)) = E(X) - E(E(X)) = E(X) - E(X) = 0$.

Cas de deux variables indépendantes. Soient $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$ deux v.a. discrètes définies sur le même espace probabilisé (Ω, P) .

PROPOSITION 3 – Quelles que soient les fonctions réelles $f : E \rightarrow \mathbb{R}$ et $g : F \rightarrow \mathbb{R}$, la formule multiplicative

$$E(f(X)g(Y)) = E(f(X))E(g(Y))$$

est valable si X et Y sont indépendantes.

Démonstration. Par définition de l'espérance de la v.a. $f(X)g(Y)$

$$E(f(X)g(Y)) = \sum_{(x,y) \in E \times F} f(x)g(y) \underbrace{p_{X,Y}(x,y)}_{p_X(x) \times p_Y(y)} = \left(\sum_{x \in E} f(x)p_X(x) \right) \times \left(\sum_{y \in F} g(y)p_Y(y) \right),$$

ce qui démontre bien le résultat. ■

2.2.2 Variance, écart-type

Définition. On appelle *variance* d'une v.a. réelle discrète $X : \Omega \rightarrow E$ ($E \subset \mathbb{R}$) et on note $\text{Var}(X)$ la quantité :

$$\text{Var}(X) = E\left([X - E(X)]^2\right)$$

Dans la pratique, on utilise plutôt l'écart-type de X , noté $\sigma(X)$, et défini par

$$\sigma(X) = \sqrt{\text{Var}(X)}$$

Cet écart-type sert à mesurer l'ampleur de l'étalement de la loi de probabilité de X autour de son centre de gravité $E(X)$, comme le montre l'inégalité suivante, dite de Bienaymé-Chebichev :

$$\forall \alpha > 0, P(|X - E(X)| \geq \alpha) \leq \frac{\sigma(X)^2}{\alpha^2}$$

Ainsi, la probabilité que X s'écarte de $E(X)$ d'une quantité supérieure ou égale à $\alpha = k\sigma(X)$ est inférieure à $\frac{1}{k^2}$ (par exemple, si $k = 10$ cette probabilité est inférieure à 0.01).

Démonstration de l'inégalité de Bienaymé-Chebichev. On part de l'expression de la variance,

$$\sigma(X)^2 = \sum_{x \in E} [x - E(X)]^2 P(X = x),$$

en remarquant que $E = \{x \in E, |x - E(X)| \leq \alpha\} + \{x \in E, |x - E(X)| > \alpha\}$, de sorte que :

$$\begin{aligned} \sigma(X)^2 &= \sum_{\{x \in E, |x - E(X)| > \alpha\}} \underbrace{[x - E(X)]^2}_{> \alpha^2} P(X = x) + \underbrace{\sum_{\{x \in E, |x - E(X)| \leq \alpha\}} [x - E(X)]^2 P(X = x)}_{\geq 0} \\ &\geq \alpha^2 \underbrace{\sum_{\{x \in E, |x - E(X)| > \alpha\}} P(X = x)}_{P(|X - E(X)| > \alpha)} \end{aligned}$$

ce qui démontre bien le résultat annoncé. ■

Lorsque $\sigma(X) = 1$, la v.a. X est dite *réduite*.

Ainsi, si $\sigma(X) \neq 0$, $\frac{X}{\sigma(X)}$ est toujours une v.a. réduite, car on sait que $\text{Var}(\lambda X) = \lambda^2 \text{Var}(X)$ pour tout réel λ ,

ce qui entraîne $\text{Var}\left(\frac{X}{\sigma(X)}\right) = \frac{1}{\sigma^2(X)} \text{Var}(X) = 1$.

De plus, comme $X - E(X)$ est centrée, $\sigma(X - E(X)) = \sigma(X)$, donc $\frac{X - E(X)}{\sigma(X)}$ est centrée réduite, puisque :

$$\sigma^2(X - E(X)) = E\left(\left[X - E(X) - \underbrace{E(X - E(X))}_0\right]^2\right) = E\left([X - E(X)]^2\right) = \sigma^2(X).$$

Autre expression de la variance. Elle s'obtient directement à partir du calcul suivant :

$$\text{Var}(X) = \sum_{x \in E} \underbrace{[x - E(X)]^2}_{x^2 - 2xE(X) + E(X)^2} p_X(x) = \underbrace{\sum_{x \in E} x^2 p_X(x)}_{E(X^2)} - 2E(X) \underbrace{\sum_{x \in E} x p_X(x)}_{E(X)} + E(X)^2 \underbrace{\sum_{x \in E} p_X(x)}_1,$$

soit

$$\boxed{\text{Var}(X) = E(X^2) - E(X)^2}$$

Exemple. Si l'on reconsidère l'exemple d'une v.a. $X \sim G(a)$ ($a \in]0, 1[$), on a :

$$\text{Var}(X) = \underbrace{E(X^2)}_{\frac{1+a}{(1-a)^2}} - \underbrace{E(X)^2}_{\frac{1}{(1-a)^2}} = \frac{a}{(1-a)^2}.$$

Cas de la somme de v.a. indépendantes. Soient (X_1, X_2, \dots, X_n) une suite de n ($n \geq 2$) v.a. réelles discrètes $X_m : \Omega \rightarrow E_m$, définies sur le même espace probabilisé (Ω, P) .

PROPOSITION 4 - Si X_1, X_2, \dots, X_n sont indépendantes alors

$$\boxed{\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)}$$

Démonstration. La linéarité de l'espérance donne (sans utiliser l'hypothèse d'indépendance) :

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n),$$

et donc

$$X_1 + X_2 + \dots + X_n - E(X_1 + X_2 + \dots + X_n) = [X_1 - E(X_1)] + [X_2 - E(X_2)] + \dots + [X_n - E(X_n)].$$

En prenant l'espérance des deux membres de cette égalité après les avoir élevés au carré, on obtient par linéarité :

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \sum_{1 \leq i, j \leq n} E([X_i - E(X_i)][X_j - E(X_j)]).$$

Mais lorsque $i \neq j$ les variables X_i et X_j sont indépendantes et la proposition 3 impose :

$$E([X_i - E(X_i)][X_j - E(X_j)]) = E([X_i - E(X_i)]) \times E([X_j - E(X_j)]) = 0 \times 0 = 0.$$

Comme $E([X_i - E(X_i)][X_j - E(X_j)]) = \text{Var}(X_i)$ lorsque $i = j$, le résultat s'en déduit. ■

2.3 Quelques modèles aléatoires discrets

2.3.1 Loïs géométriques

On a déjà défini dans ce chapitre la loi géométrique de paramètre $a \in]0, 1[$, notée $G(a)$. Si $X \sim G(a)$, on a vu que

$$P(X = n) = (1 - a)a^{n-1} \quad (n \in \mathbb{N}^*)$$

Espérance et variance. Les calculs effectués précédemment ont permis d'obtenir les expressions de l'espérance et de la variance de X :

$$E(X) = \frac{1}{1 - a}, \quad \text{Var}(X) = \frac{a}{(1 - a)^2}$$

2.3.2 Loïs de Bernoulli

Considérons un jeu de pile ou face. Le résultat X de ce jeu est symbolisé par 0 pour pile et 1 pour face. Pour représenter ce jeu, on peut poser $\Omega = \{\text{pile}, \text{face}\}$ de sorte que $X : \Omega \rightarrow \{0, 1\}$ est une v.a. discrète. On suppose que

$$P(X = 0) = 1 - p \quad \text{et} \quad P(X = 1) = p$$

pour un réel $p \in]0, 1[$ fixé.

On dit alors que X suit une loi de Bernoulli de paramètre p , et on note $X \sim B(p)$.

Lorsque $p = \frac{1}{2}$, on retrouve en fait la loi uniforme.

Espérance et variance. On a $E(X) = 1 \times p + 0 \times (1 - p)$ soit

$$E(X) = p$$

De même, $\text{Var}(X) = \underbrace{1^2 \times p + 0^2 \times (1 - p)}_{E(X^2)} - p^2$ soit

$$\text{Var}(X) = p(1 - p)$$

2.3.3 Loïs binomiales

On considère maintenant un jeu de pile ou face de durée $n \geq 2$. Les résultats X_m ($1 \leq m \leq n$) de ce jeu sont représentés par des 0 (pour pile) et des 1 (pour face). Chaque X_m , $1 \leq m \leq n$, est une v.a. suivant une loi de Bernoulli de paramètre $p \in]0, 1[$: $X_m \sim B(p)$.

Si l'on suppose que toutes ces v.a. sont indépendantes, on a

$$P[(X_1, X_2, \dots, X_n) = (x_1, x_2, \dots, x_n)] = \underbrace{P(X_1 = x_1) \times P(X_2 = x_2) \times \dots \times P(X_n = x_n)}_{\prod_{i=1}^n P(X_i = x_i)}$$

pour tout $x_i \in \{0, 1\}$ ($1 \leq i \leq n$).

Ainsi, en notant $|x|$ le nombre de coordonnées x_i du vecteur $x = (x_1, x_2, \dots, x_n)$ qui sont égales à 1, il vient immédiatement :

$$P[(X_1, X_2, \dots, X_n) = (x_1, x_2, \dots, x_n)] = p^{|x|}(1-p)^{1-|x|}.$$

En dehors des X_m , la première v.a. intéressante dans ce modèle est le nombre de X_m égaux à 1 (le nombre de "faces"), soit :

$$N_n = \sum_{m=1}^n \mathbf{1}_{\{X_m=1\}} = \sum_{m=1}^n X_m.$$

La loi de cette v.a. entière qui prend ses valeurs dans $\{0, 1, \dots, n\}$ se calcule aisément. En effet, pour tout entier k de $\{0, 1, \dots, n\}$:

$$\begin{aligned} P(N_n = k) &= \sum_{\{x=(x_1, x_2, \dots, x_n), |x|=k\}} P[(X_1, X_2, \dots, X_n) = (x_1, x_2, \dots, x_n)] \\ &= \text{Card}(\{x = (x_1, x_2, \dots, x_n), |x| = k\}) \times p^{|x|}(1-p)^{1-|x|}. \end{aligned}$$

Or $\text{Card}(\{x = (x_1, x_2, \dots, x_n), |x| = k\}) = C_n^k$ donc N_n suit la loi binomiale de paramètres n, p que l'on notera $B(n, p)$ et qui vaut :

$$\boxed{B_{n,p}(k) = C_n^k p^k (1-p)^{n-k} \quad (0 \leq k \leq n)}$$

soit

$$P(N_n = k) = B_{n,p}(k).$$

Espérance et variance. Comme $N_n = \sum_{m=1}^n X_m$, la linéarité de l'espérance impose :

$$E(N_n) = \sum_{m=1}^n E(X_m),$$

soit, puisque $E(X_m) = p$ pour tout $m \in \{1, 2, \dots, n\}$:

$$\boxed{E(N_n) = np}$$

De même, comme les X_m ($1 \leq m \leq n$) sont indépendantes,

$$\text{Var}(N_n) = \sum_{m=1}^n \text{Var}(X_m).$$

Or, $\text{Var}(X_m) = \underbrace{(1-p)^2 P(X_m=1) + (0-p)^2 P(X_m=0)}_{E([X_m - E(X_m)]^2) = E([X_m - p]^2)} = p(1-p)^2 + p^2(1-p) = p(1-p)$, donc

$$\boxed{\text{Var}(N_n) = np(1-p)}$$

Comportement asymptotique de N_n : lois des grands nombres On admettra les deux parties du résultat suivant, appelé lois des grands nombres, qui confirme rigoureusement la loi empirique des grands nombres.

PROPOSITION 5 (Lois des grands nombres)

(Loi faible). Lorsque $n \rightarrow +\infty$ la fréquence $\frac{N_n}{n}$ converge en probabilité vers la probabilité p en ce sens que pour tout $\varepsilon > 0$ fixé :

$$P \left[\left| \frac{N_n}{n} - p \right| > \varepsilon \right] \xrightarrow{n \rightarrow +\infty} 0.$$

(Loi forte). De plus :

$$P \left[\lim_{n \rightarrow +\infty} \frac{N_n}{n} = p \right] = 1.$$

2.3.4 Lois multinomiales et test d'adéquation du χ^2

Loi multinomiale

Généralisons le modèle précédent en permettant aux v.a. discrètes X_m de prendre un nombre fini d de valeurs distinctes plutôt que les seules valeurs 0 et 1, et notons E l'espace fini de cardinal d de ces valeurs possibles x . D'autre part, nous garderons l'hypothèse d'indépendance de la suite X_1, X_2, \dots, X_n ainsi que celle dite d'équidistribution qui suppose que les X_m ($1 \leq m \leq n$) suivent la même loi $\{p(x), x \in E\}$ sur E .

Pour compter les nombres de fois qu'une valeur possible x est prise par les X_m , on introduit les v.a. entières positives

$$N_{n,x} = \sum_{m=1}^n \mathbf{1}_{\{X_m=x\}},$$

qui vérifient manifestement l'égalité $\sum_{x \in E} N_{n,x} = n$.

On admettra que le vecteur aléatoire d – dimensionnel à coordonnées entières positives

$$(N_{n,x}, x \in E)$$

suit alors la loi multinomiale de paramètres n et $\{p(x), x \in E\}$ suivante

$$P[N_{n,x} = n_x] = n! \prod_{x \in E} \frac{p(x)^{n_x}}{n_x!} \text{ où } n_x \geq 0 \text{ (} x \in E \text{) et } \sum_{x \in E} n_x = n$$

Test d'adéquation du χ^2

Un problème important dans les applications est de tester si une v.a. X dans un espace fini E suit une loi donnée $\{p(x), x \in E\}$ sur cet espace (par exemple si un digit aléatoire suit la loi uniforme sur $\{0, 1, \dots, 9\}$). Pour procéder à un tel test, on se donne un échantillon de taille n de la v.a., c'est-à-dire la suite

$$X_1, X_2, \dots, X_n$$

de n v.a. indépendantes et de même loi sur E , obtenue en observant la v.a. X dans n répétitions indépendantes et similaires de l'expérience à laquelle X est liée. La loi commune de ces X_i est

- soit la loi $\{p(x), x \in E\}$ si "l'hypothèse est vraie";
- soit une autre loi $\{q(x), x \in E\}$ quelconque si "l'alternative est vraie".

Le vecteur aléatoire

$$\left(\frac{N_{n,x}}{n}, x \in E \right)$$

formé par la fréquence des x ($x \in E$) dans l'échantillon sera alors, d'après la loi des grands nombres, proche du vecteur $p = (p(x), x \in E)$ dans le cas où l'hypothèse est vraie et dans ce cas seulement.

Pour mesurer l'écart entre ces deux vecteurs, le statisticien K. Pearson a introduit la v.a. scalaire

$$Z_n = \sum_{x \in E} \frac{[N_{n,x} - np(x)]^2}{np(x)}$$

dont les valeurs probables sont très différentes selon que l'hypothèse est vraie ou non, lorsque n est grand. En effet, on montre que

- $E(Z_n) = d - 1$ si l'hypothèse est vraie;
- $E(Z_n) \geq n \sum_{x \in E} \frac{[q(x) - p(x)]^2}{p(x)}$ si l'alternative est vraie.

Par conséquent, lorsque n est assez grand, $E(Z_n) \gg d - 1$ si l'alternative est vraie (et si la loi p est suffisamment différente de la loi q).

2.3.5 Lois de Poisson

La loi de Poisson P_λ de paramètre λ (λ réel > 0) est la probabilité sur \mathbb{N} définie par la formule

$$P_\lambda(n) = e^{-\lambda} \frac{\lambda^n}{n!} \quad (n \in \mathbb{N})$$

Le facteur constant $e^{-\lambda}$ assure que $\sum_{n \in \mathbb{N}} P_\lambda(n) = 1$ comme il le faut.

Espérance et variance. Si X suit une loi de Poisson de paramètre λ , son espérance est égale à son paramètre. En effet,

$$E(X) = \sum_{n \in \mathbb{N}} n P_\lambda(n) = \sum_{n \geq 1} n e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} \lambda \underbrace{\sum_{n \geq 1} \frac{\lambda^{n-1}}{(n-1)!}}_{e^\lambda},$$

ce qui donne bien

$$E(X) = \lambda$$

Ensuite,

$$\text{Var}(X) = \underbrace{\sum_{n \in \mathbb{N}} n^2 P_\lambda(n)}_{E(X^2)} - \lambda^2,$$

et

$$\sum_{n \in \mathbb{N}} n^2 P_\lambda(n) = e^{-\lambda} \sum_{n \in \mathbb{N}} n^2 \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n \geq 1} \underbrace{n}_{(n-1)+1} \frac{\lambda^n}{(n-1)!} = e^{-\lambda} \left[\lambda^2 \underbrace{\sum_{n \geq 2} \frac{\lambda^{n-2}}{(n-2)!}}_{e^\lambda} + \lambda \underbrace{\sum_{n \geq 1} \frac{\lambda^{n-1}}{(n-1)!}}_{e^\lambda} \right],$$

soit

$$\boxed{\text{Var}(X) = \lambda}$$

Chapitre 3

Variables aléatoires réelles

On reprend ici la notion de variable aléatoire introduite au chapitre précédent pour l'étendre au cas où ces v.a. prennent leurs valeurs dans toute la droite réelle \mathbb{R} (qui est un espace non dénombrable). Nous dirons donc qu'une application

$$X : \Omega \rightarrow \mathbb{R}$$

définie sur un espace probabilisé (Ω, P) est une variable aléatoire réelle (v.a.r. en abrégé).

Pour tous réels $a < b$,

$$\{a < X < b\} \stackrel{\text{def}}{=} \{\omega \in \Omega, X(\omega) \in]a, b[\}$$

est alors un événement dont on peut calculer la probabilité

$$P(a < X < b) \stackrel{\text{def}}{=} P(\{a < X < b\}).$$

En particulier, une v.a. entière $X : \Omega \rightarrow \mathbb{Z}$ est une v.a.r. au sens précédent et

$$\{a < X < b\} = \sum_{\{n \in \mathbb{Z}, a < n < b\}} \{X = n\}, \quad (a < b)$$

de sorte que

$$P(a < X < b) = \sum_{\{n \in \mathbb{Z}, a < n < b\}} p_X(n) \quad (D)$$

si $\{p_X(n) = P(X = n), n \in \mathbb{Z}\}$ désigne la loi de probabilité de X .

3.1 Densités de probabilité

3.1.1 Définition et propriétés immédiates

Définition. Nous dirons que la v.a.r. X possède la densité de probabilité $x \mapsto p_X(x)$ sur \mathbb{R} si

$$P(a < X < b) = \int_a^b p_X(x) dx \quad (C)$$

quels que soient les réels $a < b$.

Remarque. Si X possède une densité de probabilité p_X , on a forcément

$$P(X = c) = 0$$

pour tout $c \in \mathbb{R}$, donc les événements $\{X = c\}$ ne sont pas intéressants pour une telle variable.

De ce fait, l'égalité $\{a \leq X < b\} = \{X = a\} + \{a < X < b\}$ implique $P(a \leq X < b) = P(a < X < b)$ et l'on a plus généralement

$$P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b)$$

pour tous $a < b$ de \mathbb{R} .

Propriétés immédiates. Comme $P(a < X < b) \geq 0$ pour tous réels $a < b$ et que $P(X \in \mathbb{R}) = 1$, la fonction p_X doit nécessairement être positive, d'intégrale égale à 1 sur \mathbb{R} :

$$\boxed{\begin{cases} \forall x \in \mathbb{R}, p_X(x) \geq 0 \\ \int_{\mathbb{R}} p_X(x) dx = \int_{-\infty}^{+\infty} p_X(x) dx = 1 \end{cases}}$$

Remarque. Une v.a. entière ne possède pas de densité de probabilité (à moins d'introduire les pseudo-fonctions de Dirac, ce que nous ne ferons pas ici). C'est pourquoi, toute v.a.r. X admettant une densité sera dite continue (par opposition aux v.a.r. entières, qui sont discrètes).

Néanmoins on peut remarquer que le maniement de sommes telles que (D) est analogue à celui des intégrales de (C) et d'ailleurs, on adopte généralement comme ici la même notation pour les lois discrètes $\{p_X(n), n \in \mathbb{Z}\}$ et les densités de probabilité $\{p_X(x), x \in \mathbb{R}\}$.

Transformation de v.a.r. Pour toute fonction $f : \mathbb{R} \mapsto \mathbb{R}$, $f(X)$ est une v.a.r.

De plus, si X est continue de densité de probabilité p_X , on a alors pour tous réels $a < b$:

$$\boxed{P(a < f(X) < b) = \int_{\{x \in \mathbb{R}, a < f(x) < b\}} p_X(x) dx = \int_{f^{-1}(]a,b[)} p_X(x) dx}$$

3.1.2 Exemples de densités de probabilité

Densité uniforme

La densité uniforme sur un intervalle réel fini $[u, v]$ est constante sur cet intervalle et nulle en dehors. Elle est donc définie par

$$u(x) = \frac{1}{v-u} \mathbf{1}_{[u,v]}(x) = \begin{cases} \frac{1}{v-u} & \text{si } u \leq x \leq v; \\ 0 & \text{si } x \in \mathbb{R} \setminus [u, v]. \end{cases}$$

Pour toute v.a.r. X possédant la densité uniforme sur $[u, v]$ on note $X \rightsquigarrow \mathcal{U}_{[u,v]}$.

Densité gaussienne

La fonction de Gauss $x \mapsto e^{-\frac{x^2}{2}}$ n'a pas de primitive explicite mais on montre que

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

Par conséquent

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

est une densité de probabilité appelée densité gaussienne.

Densités normales

La densité normale de paramètres m et σ^2 ($m \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+^*$) est la fonction

$$n_{m,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Remarque. La densité gaussienne est en fait la densité normale de paramètres $m = 0$ et $\sigma = 1$ puisque $g = n_{0,1}$.

Lorsqu'une v.a.r. X admet la densité normale de paramètres m et σ^2 , on note $X \sim N(m, \sigma^2)$.

Densités exponentielles sur \mathbb{R}_+

Pour tout réel $\theta > 0$, la fonction positive

$$e_\theta(x) = \theta e^{-\theta x} \mathbf{1}_{\mathbb{R}_+}(x)$$

est une densité de probabilité sur \mathbb{R} portée par \mathbb{R}_+ , appelée densité exponentielle sur \mathbb{R}_+ de paramètre θ . Ces densités jouent dans les applications le même rôle que les lois géométriques des v.a. discrètes. En effet, les densités exponentielles sont typiquement des densités de durées de vie ou de délais d'attente lorsque ces durées de vie et délais d'attente sont mesurés de manière continue plutôt que discrète.

Densités gamma

On appelle fonction d'Euler de première espèce, l'application définie pour tout $a \in \mathbb{R}_+$ par

$$\Gamma(a) = \int_0^{+\infty} e^{-x} x^{a-1} dx.$$

Si $a = n \in \mathbb{N}^*$ par exemple, on obtient (après intégration par parties) que $\Gamma(n) = (n-1)!$ et on montre (par une méthode hors programme) que $\Gamma(\frac{1}{2}) = \frac{1}{\sqrt{\pi}}$.

La densité gamma de paramètres a et p ($a \in \mathbb{R}_+^*$ et $p \in \mathbb{R}_+^*$) est alors définie par

$$\gamma_{a,p}(x) = \frac{1}{p^a \Gamma(a)} e^{-\frac{x}{p}} x^{a-1} \mathbf{1}_{\mathbb{R}_+}(x).$$

Remarque. La densité exponentielle de paramètre θ ($\theta > 0$) est en fait la densité gamma de paramètres $a = 1$ et $p = \frac{1}{\theta}$ puisque $e_\theta = \gamma_{1, \frac{1}{\theta}}$.

Lorsqu'une v.a.r. X admet la densité gamma de paramètres a et p , on note $X \rightsquigarrow \gamma(a, p)$.

Densité de Cauchy

La fonction

$$c(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

est une densité de probabilité appelée densité de Cauchy.

3.2 Outils permettant l'étude des v.a.r.

Soit $X : \Omega \mapsto \mathbb{R}$ une v.a.r. définie sur un espace probabilisé (Ω, P) .

3.2.1 Fonction de répartition de X

Définition. On appelle fonction de répartition de X l'application :

$$\begin{aligned} F_X : \mathbb{R} &\rightarrow [0, 1] \\ x &\mapsto P(X \leq x). \end{aligned}$$

Propriétés immédiates. Pour tous les x et x' de E tels que $x \leq x'$,

$$\{X \leq x\} \subset \{X \leq x'\} \implies F_X(x) \leq F_X(x'),$$

donc

$$\boxed{F_X \text{ est croissante}}$$

De plus, on déduit directement de la définition de F_X que

$$\boxed{\lim_{x \rightarrow -\infty} F_X(x) = 0, \quad \lim_{x \rightarrow +\infty} F_X(x) = 1}$$

Enfin, F_X est continue à droite et on vérifie par ailleurs que

$$F_X(x_0) - \lim_{x \rightarrow x_0^-} F_X(x) = P(X = x_0).$$

En particulier, si X est discrète et ne prend qu'un nombre fini de valeurs réelles x_1, x_2, \dots, x_n , alors F_X est une fonction en

Par contre, si X est continue, comme $P(X = x_0) = 0$, l'application F_X est continue à gauche en tout point x_0 et donc que F_X est continue sur \mathbb{R} . En fait, F_X est même dérivable sur \mathbb{R} , car X admettant la densité p_X , on a pour tout $x \in \mathbb{R}$:

$$\boxed{F_X(x) = \int_{-\infty}^x p_X(x) dx \implies p_X(x) = F_X'(x)}$$

Exemples.

Densité uniforme. Si $X \rightsquigarrow \mathcal{U}_{[a,b]}$ ($a < b$) alors

$$\forall x \in \mathbb{R}, F_X(x) = \int_{-\infty}^x \frac{1}{b-a} \mathbf{1}_{[a,b]}(t) dt = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a < x < b \\ 1 & \text{si } b \leq x \end{cases}$$

Densité de Cauchy. Si X possède la densité de Cauchy alors

$$\forall x \in \mathbb{R}, F_X(x) = \int_{-\infty}^x \frac{1}{\pi} \frac{dt}{1+t^2} = \frac{1}{\pi} \lim_{a \rightarrow -\infty} ([\arctan t]_a^x) = \frac{\arctan x}{\pi} + \frac{1}{2}.$$

Fonction de répartition de la transformée d'une v.a.r. continue. Evidemment, X étant une v.a.r. continue de densité p_X sur \mathbb{R} , on a pour toute fonction $f : \mathbb{R} \mapsto \mathbb{R}$:

$$\forall y \in \mathbb{R}, F_{f(X)}(y) = P(f(X) \leq y) = \int_{\{x, f(x) \leq y\}} p_X(x) dx = \int_{f^{-1}(\cdot]^{-\infty, y])} p_X(x) dx.$$

Application : $X \rightsquigarrow N(0, 1) \implies X^2 \rightsquigarrow \gamma(\frac{1}{2}, 2)$ En effet, Y désignant la v.a.r. à valeurs positives X^2 , on a pour tout $y \in \mathbb{R}_+$:

$$F_Y(y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

En dérivant cette égalité par rapport à $y \in \mathbb{R}_+$ on obtient alors

$$p_Y(y) = \frac{1}{2\sqrt{y}} p_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} p_X(-\sqrt{y}) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}.$$

De plus, $F_Y(y) = 0$ et donc $p_Y(y) = 0$ si $y < 0$, d'où

$$\forall y \in \mathbb{R}, p_Y(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}} \mathbf{1}_{\mathbb{R}_+}(y) = \gamma_{\frac{1}{2}, 2}(y),$$

soit $Y \rightsquigarrow \gamma(\frac{1}{2}, 2)$.

3.2.2 Valeurs caractéristiques d'une v.a.r.

Espérance

Définition. L'espérance d'une v.a.r. continue X admettant une densité de probabilité p_X sur \mathbb{R} est définie par

$$E(X) = \int_{\mathbb{R}} x p_X(x) dx$$

du moins si cette intégrale converge absolument.

Comme pour une v.a. entière, l'espérance de la v.a.r. continue X est le centre de gravité de la densité $p_X(x)$ (on parle également de moyenne de X). Ceci explique que la terminologie des v.a. entières soit conservée pour une v.a.r. continue X , qui sera dite centrée si

$$E(X) = 0.$$

Espérance de $f(X)$. Plus généralement, pour toute fonction $f : \mathbb{R} \mapsto \mathbb{R}$, l'espérance de la v.a.r. $f(X)$ est

$$\boxed{E(f(X)) = \int_{\mathbb{R}} f(x)p_X(x)dx}$$

Deux propriétés élémentaires. Compte tenu des propriétés de l'intégrale, elles s'écrivent

$$\boxed{E(aX + b) = aE(X) + b}$$

et, X_1, X_2, \dots, X_n désignant n ($n \geq 1$) v.a.r. continues :

$$\boxed{E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)}$$

On déduit notamment de la première de ces deux propriétés que $X - E(X)$ est centrée pour toute v.a.r. continue X .

Exemples.

Densité uniforme sur $[u, v]$. Si $X \rightsquigarrow \mathcal{U}_{[u,v]}$ ($u < v$), alors

$$E(X) = \int_{\mathbb{R}} x \times \frac{1}{v-u} \mathbf{1}_{[u,v]}(x) dx = \frac{1}{v-u} \underbrace{\int_u^v x dx}_{\left[\frac{x^2}{2} \right]_u^v} = \frac{u+v}{2}.$$

Densité gaussienne. Si $X \rightsquigarrow N(0, 1)$ on a :

$$E(X) = \int_{\mathbb{R}} x \times \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \left(e^{-\frac{x^2}{2}} \right)' dx = \lim_{a \rightarrow -\infty} \left\{ \lim_{b \rightarrow +\infty} \left(\frac{1}{\sqrt{2\pi}} \left[e^{-\frac{x^2}{2}} \right]_a^b = 0 \right) \right\} = 0.$$

Densités normales. Si $X \rightsquigarrow N(m, \sigma^2)$ ($m \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+^*$), on a :

$$\begin{aligned} E(X) &= \int_{\mathbb{R}} x \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx \\ &\stackrel{y=\frac{x-m}{\sigma}}{=} \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} (\sigma y + m) e^{-\frac{y^2}{2}} (\sigma dy) \\ &= \underbrace{\sigma \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} y e^{-\frac{y^2}{2}} dy}_{=0} + m \underbrace{\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{y^2}{2}} dy}_{=1} = m \end{aligned}$$

Densités exponentielles. Toute v.a.r. $X \rightsquigarrow \gamma(1, \frac{1}{\theta})$ ($\theta > 0$) a pour espérance

$$E(X) = \int_{\mathbb{R}} x \times \theta e^{-\theta x} \mathbf{1}_{\mathbb{R}_+}(x) dx = \int_{\mathbb{R}_+} x \times \underbrace{(\theta e^{-\theta x})}_{-(e^{-\theta x})'} dx = \lim_{b \rightarrow +\infty} \left(\left[-x(e^{-\theta x}) \right]_0^b \right) + \underbrace{\int_{\mathbb{R}_+} e^{-\theta x} dx}_{\frac{1}{\theta}} = \frac{1}{\theta}.$$

Densités gamma. Si $X \rightsquigarrow \gamma(a, p)$ ($a \in \mathbb{R}_+^*$ et $p \in \mathbb{R}_+^*$), alors

$$\begin{aligned} E(X) &= \int_{\mathbb{R}} x \times \frac{1}{p^a \Gamma(a)} e^{-\frac{x}{p}} x^{a-1} \mathbf{1}_{\mathbb{R}_+}(x) dx \\ &= \frac{1}{p^a \Gamma(a)} \int_{\mathbb{R}_+} x^a e^{-\frac{x}{p}} dx \\ &\stackrel{y=\frac{x}{p}}{=} \frac{1}{p^a \Gamma(a)} \int_{\mathbb{R}_+} (py)^a e^{-y} (p dy) \\ &= \frac{p^{a+1}}{p^a \Gamma(a)} \underbrace{\int_{\mathbb{R}_+} y^a e^{-y} dy}_{\Gamma(a+1)} = p \frac{\Gamma(a+1)}{\Gamma(a)}. \end{aligned}$$

Densité de Cauchy. Une v.a.r. admettant une densité de Cauchy n'a pas d'espérance car la fonction

$$x \mapsto \frac{x}{1+x^2}$$

qui admet $x \mapsto \frac{1}{2} \ln(1+x^2)$ pour primitive sur \mathbb{R} , n'est pas intégrable sur \mathbb{R} en entier.

Remarque de symétrie. L'espérance d'une v.a.r. X admettant une densité p_X symétrique autour d'un point a (c'est-à-dire vérifiant $p_X(x) = p_X(2a-x)$), si elle existe, est nécessairement égale à a . En effet, on a dans ce cas

$$E(X) = \int_{-\infty}^{+\infty} x p_X(x) dx = \int_{-\infty}^{+\infty} x p_X(2a-x) dx,$$

soit, en effectuant le changement de variable $y = 2a - x$ dans cette dernière intégrale,

$$E(X) = \int_{+\infty}^{-\infty} (2a-y) p_X(y) (-dy) = 2a \underbrace{\int_{-\infty}^{+\infty} p_X(y) dy}_1 - \underbrace{\int_{-\infty}^{+\infty} y p_X(y) dy}_{E(X)} = 2a - E(X),$$

ce qui donne bien $E(X) = a$.

Cette remarque permet ainsi de calculer directement l'espérance de la loi normale $N(m, \sigma^2)$.

Mais il faut bien prendre garde au fait que ce résultat ne s'applique que si l'espérance est effectivement définie. Ainsi, bien que la densité de Cauchy soit une fonction paire, on a vu qu'une v.a.r. admettant une telle densité ne possède pas d'espérance.

Moments

Définition. Le moment d'ordre k ($k \in \mathbb{N}$) d'une v.a.r. X possédant la densité de probabilité p_X sur \mathbb{R} est

$$m_k(X) = E(X^k) = \int_{\mathbb{R}} x^k f_X(x) dx$$

si toutefois cette intégrale est absolument convergente.

De même, on appelle moment centré d'ordre k de X la quantité $E([X - E(X)]^k)$, si elle est définie.

Exemple : calcul de $m_2(X)$ lorsque $X \sim N(0, 1)$. On a vu que $Y = X^2 \sim \gamma(\frac{1}{2}, 2)$ donc

$$m_2(X) = E(Y) = 2 \frac{\Gamma(3/2)}{\Gamma(1/2)}.$$

Cas particulier. La variance de X est le moment centré d'ordre 2 de X :

$$\text{Var}(X) = \sigma^2(X) = E\left([X - E(X)]^2\right) = E(X^2) - E(X)^2.$$

Exemple : calcul de $\text{Var}(X)$ lorsque $X \sim N(0, 1)$. D'après ce que l'on vient de voir :

$$\text{Var}(X) = 2 \frac{\Gamma(3/2)}{\Gamma(1/2)} - 0^2 = m_2(X).$$

Enfin, l'inégalité de Bienaymé-Chebichev, démontrée au chapitre précédent pour une v.a. discrète, s'étend sans difficulté (il suffit de remplacer les sommes par des intégrales dans la démonstration) au cas d'une v.a.r. continue X :

$$\forall \alpha > 0, P(|X - E(X)| \geq \alpha) \leq \frac{\sigma(X)^2}{\alpha^2}$$

3.2.3 Fonction caractéristique d'une v.a.r. X .

Définition. La fonction caractéristique d'une v.a.r. X , notée $t \mapsto \varphi_X(t)$ est la "transformée de Fourier" de $x \mapsto F_X(x)$, c'est-à-dire l'espérance de la v.a.r. e^{itX} pour tout réel t :

$$\forall t \in \mathbb{R}, \varphi_X(t) = E(e^{itX}).$$

Lorsque X est continue de densité p_X sur \mathbb{R} on a alors

$$\forall t \in \mathbb{R}, \varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} p_X(x) dx$$

Exemple. Soit $X \sim \gamma(a, p)$, ($a \in \mathbb{R}_+^*$ et $p \in \mathbb{R}_+^*$). Alors, pour tout $t \in \mathbb{R}$,

$$\begin{aligned} \varphi_X(t) &= \int_{\mathbb{R}} e^{itx} \frac{1}{p^a \Gamma(a)} e^{-\frac{x}{p}} x^{a-1} \mathbf{1}_{\mathbb{R}_+}(x) dx \\ &= \frac{1}{p^a \Gamma(a)} \int_{\mathbb{R}_+} e^{-(\frac{1}{p} - it)x} x^{a-1} dx \\ &\stackrel{py=(1-itp)x}{=} \frac{1}{p^a \Gamma(a)} \int_{\mathbb{R}_+} e^{-y} \left(\frac{py}{1-itp}\right)^{a-1} \times \frac{p}{1-itp} dy \\ &= \frac{1}{p^a \Gamma(a)} \left(\frac{p}{1-itp}\right)^a \underbrace{\int_{\mathbb{R}_+} e^{-y} y^{a-1} dy}_{\Gamma(a)} = (1-itp)^{-a}. \end{aligned}$$

Formule d'inversion. Elle permet généralement de déterminer la densité d'une v.a.r. continue X à partir de sa fonction caractéristique φ_X . En effet, si $t \mapsto \varphi_X(t)$ est absolument intégrable sur \mathbb{R} , c'est-à-dire si

$$\int_{\mathbb{R}} |\varphi_X(t)| dt < +\infty,$$

alors X admet une densité continue p_X donnée par

$$\forall x \in \mathbb{R}, p_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(t) e^{-itx} dt$$

Exemple. Pour tout $p > 0$ et tout réel $a > 1$, la fonction

$$t \mapsto (1 - itp)^{-a}$$

est absolument intégrable sur \mathbb{R} , donc la formule d'inversion permet d'écrire :

$$\forall x \in \mathbb{R}, \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-itx}}{(1 - itp)^a} dt = \gamma_{a,p}(x).$$

3.3 Indépendance et convolution

3.3.1 Vecteurs aléatoires réels

Définition. Nous dirons qu'une application

$$X : \Omega \rightarrow \mathbb{R}^d \quad (d \geq 2)$$

définie sur l'espace probabilisé (Ω, P) est un vecteur aléatoire réel.

Chacune des coordonnées X_i , $1 \leq i \leq d$, de X est alors une v.a.r.

Loi de probabilité Pour simplifier, on considère un vecteur aléatoire $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ (la généralisation à $d > 2$ se faisant sans difficulté), dont les coordonnées sont les v.a.r. X et Y .

Pour tout $(x, y) \in \mathbb{R}^2$, $\{X \leq x\} \cap \{Y \leq y\}$ est un événement dont on peut calculer la probabilité :

$$P(X \leq x, Y \leq y) \stackrel{\text{def}}{=} P(\{X \leq x\} \cap \{Y \leq y\}),$$

qui définit la fonction de répartition du couple (X, Y) :

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

S'il existe une fonction $p_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ telle que

$$\forall (x, y) \in \mathbb{R}^2, F_{X,Y}(x, y) = \int_{u=-\infty}^{u=x} \int_{v=-\infty}^{v=y} p_{X,Y}(u, v) dudv,$$

le vecteur (X, Y) est dit de type continu.

La fonction $p_{X,Y}$ s'appelle alors la densité du couple (X, Y) sur \mathbb{R}^2 .

Evidemment, une telle densité $p_{X,Y}$ est à valeurs positives et son intégrale est égale à 1 sur \mathbb{R}^2 :

$$\left\{ \begin{array}{l} \forall (x, y) \in \mathbb{R}^2, p_{X,Y}(x, y) \geq 0 \\ \int_{u \in \mathbb{R}} \int_{v \in \mathbb{R}} p_{X,Y}(u, v) dudv = 1 \end{array} \right.$$

Exemple. La fonction de deux variables

$$p_{X,Y}(x, y) = \frac{1}{\pi} e^{-(x^2+y^2)/2} \mathbf{1}_{\{xy>0\}}$$

est par exemple une densité sur \mathbb{R}^2 .

Application : calcul de la loi de $X + Y$ à partir de $p_{X,Y}$. Pour tout $z \in \mathbb{R}$,

$$F_{X+Y}(z) = P(X + Y \leq z) = \int \int_{\{(x,y) \in \mathbb{R}^2, x+y \leq z\}} p_{X,Y}(x, y) dx dy = \int \int_{x+y \leq z} p_{X,Y}(x, y) dx dy.$$

3.3.2 Indépendance

Définition. Nous dirons que deux v.a.r. continues X et Y définies sur le même espace probabilisé (Ω, P) sont indépendantes si elles vérifient

$$\forall (x, y) \in \mathbb{R}^2, F_{X,Y}(x, y) = F_X(x) \times F_Y(y)$$

Maintenant, si les v.a.r. X et Y sont continues et admettent pour densités respectives p_X et p_Y , cette condition s'écrit également

$$\forall (x, y) \in \mathbb{R}^2, p_{X,Y}(x, y) = p_X(x) \times p_Y(y)$$

Exemple. Supposons que $X \rightsquigarrow N(0, 1)$ et $Y \rightsquigarrow N(0, 1)$, de sorte que $p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ et $p_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$. Si la densité du couple (X, Y) est

$$p_{X,Y}(x, y) = \frac{1}{\pi} e^{-(x^2+y^2)/2} \mathbf{1}_{\{xy>0\}},$$

on voit bien que $p_{X,Y}(x, y) \neq p_X(x)p_Y(y)$ et donc que X et Y ne sont donc pas indépendantes.

PROPOSITION 1 – Si X et Y sont deux v.a.r. indépendantes, alors $f(X)$ et $g(Y)$ sont également indépendantes pour toutes fonctions f et g de \mathbb{R} dans \mathbb{R} .

Ensuite, comme dans le cas des variables discrètes, on démontre la proposition suivante.

PROPOSITION 2 – Si X et Y sont deux v.a.r. indépendantes :

- $E(XY) = E(X)E(Y)$
- $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$

3.3.3 Convolution

PROPOSITION 3 – Si X et Y sont deux v.a.r. indépendantes de densités respectives p_X et p_Y , alors $X + Y$ a pour densité

$$p_{X+Y}(u) = (p_X * p_Y)(u) = \int_{-\infty}^{+\infty} p_X(u - v)p_Y(v)dv.$$

Démonstration. Pour tout $z \in \mathbb{R}$, on a, en posant $Z = X + Y$:

$$F_Z(z) = P(Z \leq z) = \int \int_{x+y \leq z} \underbrace{p_{X,Y}(x,y)}_{p_X(x)p_Y(y)} dx dy.$$

Effectuons le changement de variable

$$\begin{cases} x + y = u \\ y = v \end{cases} \iff \begin{cases} x = u - v \\ y = v, \end{cases}$$

de sorte que

$$F_Z(z) = \int_{-\infty}^z \underbrace{\left(\int_{-\infty}^{+\infty} p_X(u-v)p_Y(v) dv \right)}_{p_{X+Y}(u)} du. \blacksquare$$

Exemple. Soient $X \rightsquigarrow U_{[0,1]}$ et $Y \rightsquigarrow U_{[0,1]}$, deux v.a.r. supposées indépendantes. La densité de la v.a.r. $X + Y$ est donc

$$\forall u \in \mathbb{R}, h(u) = \int_{-\infty}^{+\infty} \mathbf{1}_{[0,1]}(u-v) \times \mathbf{1}_{[0,1]}(v) dv.$$

Or la fonction intégrée est non nulle si et seulement si

$$\begin{cases} 0 \leq u - v \leq 1 \\ \text{et} \\ 0 \leq v \leq 1 \end{cases} \implies 0 \leq u \leq 2.$$

Choisissons donc $u \in [0, 2]$.

$$\text{— si } 0 \leq u \leq 1, h(u) = \int_0^u dv = u$$

$$\text{— si } 1 \leq u \leq 2, h(u) = \int_{u-1}^1 dv = 1 - (u - 1) = 2 - u.$$

Enfin, comme on l'a déjà mentionné plus haut, $h(u) = 0$ si $u < 0$ ou $u > 2$.

Caractérisation. Si X et Y deux v.a.r. continues, on a l'équivalence :

$$\boxed{X \text{ et } Y \text{ sont indépendantes} \iff \varphi_{X+Y} = \varphi_X \times \varphi_Y}$$

Démontrons l'implication " \implies ". D'après la proposition 1, les deux v.a.r. e^{itX} et e^{itY} sont indépendantes pour tout réel t . Il résulte alors de la proposition 2 :

$$\forall t \in \mathbb{R}, \varphi_{X+Y}(t) = \mathbb{E} \left(e^{it(X+Y)} \right) = \mathbb{E} \left(e^{itX} \times e^{itY} \right) = \underbrace{\mathbb{E} \left(e^{itX} \right)}_{\varphi_X(t)} \underbrace{\mathbb{E} \left(e^{itY} \right)}_{\varphi_Y(t)},$$

ce qui démontre la première implication de cette équivalence.

La réciproque, est alors une conséquence directe de la proposition 3 et de la formule d'inversion.

Exemple. Soient $X \rightsquigarrow \gamma(a, p)$ et $Y \rightsquigarrow \gamma(b, p)$ (a, b et p appartenant à \mathbb{R}_+^*) deux v.a.r. indépendantes. Pour tout $t \in \mathbb{R}$, on a donc

$$\varphi_{X+Y}(t) = \underbrace{\varphi_X(t)}_{(1-itp)^{-a}} \times \underbrace{\varphi_Y(t)}_{(1-itp)^{-b}} = (1-itp)^{-(a+b)},$$

donc la formule d'inversion montre que $X + Y \rightsquigarrow \gamma(a + b, p)$.

Conséquence. Soient $X_i \rightsquigarrow N(0, 1)$, $1 \leq i \leq n$ ($n \in \mathbb{N}^*$), n v.a.r. indépendantes. D'après la proposition 1, les variables X_i^2 sont indépendantes et comme (on a déjà vu que) $X_i^2 \rightsquigarrow \gamma(\frac{1}{2}, 2)$ pour tout $i \in \{1, 2, \dots, n\}$, l'exemple précédent implique

$$X_1^2 + X_2^2 + \dots + X_n^2 \rightsquigarrow \gamma(\frac{n}{2}, 2).$$

On dit alors que $X_1^2 + X_2^2 + \dots + X_n^2$ suit la loi du chi-deux à n degrés de liberté.