Probabilités et Statistique

Cours de Licence 3 Miashs et Maths

année 2017/2018
(Version en cours de construction)

Faicel Chamroukhi

chamroukhi@unicaen.fr

http://math.unicaen.fr/ chamroukhi/

Table des matières

Chapitre 1

Introduction et fondements des probabilités

1.1 Introduction

La théorie des probabilités sert à modéliser des situations, phénomènes, etc, dont notre connaissance est imparfaite/imprécise. Cet aspect d'imprécision/d'incertitude, de manque d'information, est alors caractérisé par une composante *aléatoire*.

Par exemple, lors du jet d'un dé, les lois de Newton devraient en principe nous permettre de calculer la trajectoire exacte du dé, connaissant sa position et sa vitesse initiales, et d'en déduire sur quelle face il va tomber. En pratique, non seulement ce calcul est extrêmement difficile, mais le résultat dépend aussi de manière très sensible des conditions initiales (condition de l'expérience). Il est alors plus simple d'admettre que le dé peut tomber sur chacune de ses six faces avec la même probabilité de $1/6$ (si le dé est parfaitement symétrique - sinon, il peut être préférable d'associer des probabilités différentes aux différentes faces).

Le calcul des probabilités est donc la branche de l'analyse qui étude des phénomènes résultants d'*expérience aléatoires*

1.1.1 Expérience aléatoire

◇ **Définition 1.1.1.** *On appelle* **expérience (ou épreuve) aléatoire** *une expérience dont on connaît l'ensemble des résultats possibles mais dont on ne peut prédire le résultat effectif avec certitude, et qui, répétée plusieurs fois dans des conditions opératoires identiques, produit des résultats qui peuvent être différents.*

▷ **Exemple 1.1.1.** *Les exemples de telles expériences sont nombreux : par exemple le lancer d'un dé est une expérience aléatoire : le résultat est un entier compris entre 1 et 6 dont la valeur ne peut être connue avant le lancer et en répétant le lancer dans des conditions identiques on peut avoir un résultat différent.*

Le jeu pile ou face, sexe d'un enfant à naître, temps d'attente d'un client à la poste, à la durée de vie d'une particule radioactive, etc

◇ **Définition 1.1.2.** *L'ensemble des résultats possibles d'une expérience aléatoire est appelé univers. Il est généralement noté Ω et aussi connu sous le nom d'ensemble fondamental ou d'espace des possibles.*

▷ **Exemple 1.1.2.** *L'univers de l'expérience du lancer de dé est l'ensemble*

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

▷ **Exemple 1.1.3.** *Le lancer d'une pièce de monnaie est une expérience aléatoire dont l'univers est l'ensemble*

$$\Omega = \{\textit{pile}, \textit{face}\}$$

⚠ **Remarque 1.1.1.** *L'univers d'une expérience aléatoire étant un ensemble, sa description s'appuie donc sur la théorie des ensembles. On utilise en particulier la notion de produit cartésien, comme l'illustrent les exemples suivants.*

▷ **Exemple 1.1.4.** *On lance deux fois de suite un dé. L'espace des possibles associé à un lancer est $\mathcal{J} = \{1, 2, 3, 4, 5, 6\}$. Comme on réalise deux lancers, l'expérience produit un couple de résultats, c'est-à-dire un élément du produit cartésien $\mathcal{J} \times \mathcal{J} = \mathcal{J}^2$. L'univers de l'expérience est donc $\Omega = \{1, 2, 3, 4, 5, 6\}^2$*

Univers plus complexes

▷ **Exemple 1.1.5.** *Considérons une urne contenant 3 jetons numérotés de 1 à 3. On tire au hasard un premier jeton puis un second jeton (sans remise). Le résultat de l'expérience est donc une paire (ordonnée) d'éléments choisis dans $\mathcal{J} = \{1, 2, 3\}$. L'univers n'est pas cependant $\mathcal{J}^2$. En effet, si on tire d'abord le jeton 1, par exemple, celui n'est plus disponible dans l'urne. Les résultats de l'expérience sont donc des paires d'éléments distincts. L'univers est alors*

$$\begin{aligned} \Omega &= \{(a, b) \in \mathcal{J}^2 | a \neq b\}, \\ &= \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}. \end{aligned}$$

Cet ensemble ne peut pas se formuler plus simplement sous forme d'un produit cartésien, par exemple.

▷ **Exemple 1.1.6.** *On lance deux dés simultanément. Si on considère les deux dés comme* **indiscernables***, on ne doit pas faire de différence entre les paires $(1, 2)$ et $(2, 1)$: rien ne permet d'ordonner les dés. On peut représenter l'univers sous la forme suivante :*

$$U = \{\{a, b\} | a \in \{1, 2, 3, 4, 5, 6\} \textit{ et } b \in \{1, 2, 3, 4, 5, 6\}\}.$$

Dans cette définition, la notation $\{a, b\}$ désigne l'ensemble contenant les deux éléments a et b. Par nature, un ensemble n'est pas ordonné et de ce fait, $\{a, b\} = \{b, a\}$, ce qui semble bien adapté pour l'expérience considérée. Cependant, un ensemble ne contient qu'une seule fois chacun de ses éléments. De ce fait, l'ensemble $\{a, a\}$ correspond en réalité à l'ensemble $\{a\}$. Techniquement, ce détail ne change rien à la qualité de la modélisation : les seuls singletons de U correspondent aux lancers pour lesquels on obtient deux fois le même chiffre. Il est parfois cependant plus clair de supposer les objets étudiés comme **discernables** *et donc de considérer comme dans l'exemple 1.5 des n-uplets plutôt que des sous-ensembles de n éléments. En considérant les dés comme discernables, on obtient ici le même univers que dans l'exemple 1.1.4, c'est-à-dire :*

$$\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}.$$

1.1.2 Évènements aléatoires

◇ **Définition 1.1.3.** *Soit une expérience aléatoire* $\mathcal{E}$ *et son univers* Ω*. Un* **événement aléatoire** *correspond à une affirmation qui peut être vraie ou fausse suivant le résultat de l'expérience aléatoire. Un événement aléatoire est un sous-ensemble de tous les résultats possibles de l'expérience aléatoire, donc un sous-ensemble de l'univers* Ω*. On appelle* **évènement élémentaire** *les ensembles de résultats possibles réduits à un seul élément, c'est-à-dire les singletons de* $\mathcal{E}(\Omega)$*. Un évènement élémentaire est donc de la forme* $\{\omega\}$ *pour tout* $\omega \in \Omega$*.*

▷ **Exemple 1.1.7.** *On considère le lancer lancer d'un dé. Voici quelques exemples d'évènements :*
- *l'évènement* $A = \{1, 3, 5\}$ *correspond à l'obtention d'un chiffre impair ;*
- *l'évènement* $B = \{2, 4, 6\}$ *correspond à l'obtention d'un chiffre pair ;*
- *l'évènement* $C = \{4, 5, 6\}$ *correspond à l'obtention d'un chiffre supérieur ou égal à trois ; l'évènement* $D = \{4\}$ *correspond à l'obtention d'un quatre. C'est un évènement élémentaire.*

▷ **Exemple 1.1.8.** *Dans l'expérience des deux lancers successifs d'un dé :*
- *l'obtention de deux fois de suite la même valeur est l'évènement*

$$A = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\};$$

- *l'obtention d'un tirage dont le premier lancer est pair est l'évènement*

$$B = \{2, 4, 6\} \times \{1, 2, 3, 4, 5, 6\}.$$

Une formule logique relativement simple peut conduire à un évènement complexe, comme le montre l'exemple suivant.

▷ **Exemple 1.1.9.** *On considère l'expérience des deux lancers d'un dé et l'évènement "obtenir une somme de 6 en ajoutant le résultat des deux lancers". Une façon simple d'écrire l'évènement de façon ensembliste est la suivante*

$$A = \{(u, v) \in \{1, 2, 3, 4, 5, 6\}^2 | u + v = 6\}.$$

Bien que parfaitement correcte mathématiquement, cette formulation ne renseigne pas beaucoup sur l'évènement. Malheureusement, il n'existe pas d'autre formulation simple et on doit donc se contenter soit de la formule précédente, soit d'une description exhaustive de A (en extension)

$$A = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}.$$

L'expérience étant très simple, la description exhaustive n'est pas trop grande.

⚠ **Remarque 1.1.2.** *Dans le cas d'une expérience plus complexe, on peut rapidement arriver à des formulations ensemblistes assez lourdes.*

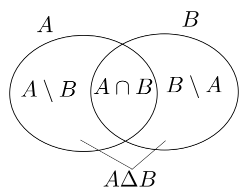
1.1.3 Vocabulaire probabiliste VS ensembliste

Vocabulaire probabiliste Bien que l'univers d'une expérience et les évènements associés soient décrits grâce aux concepts de la théorie des ensembles, la théorie des probabilités utilise un vocabulaire différent détaillé ci-dessous (ce vocabulaire est en partie partagé avec la logique mathématique).

Vocabulaire probabiliste	Vocabulaire ensembliste	Notation
évènement de l'univers Ω	sous-ensemble de Ω	$A \subset \Omega$
évènement **impossible**	ensemble vide	$\emptyset$
évènement **certain**	ensemble plein Ω	Ω
évènement contraire de A	complémentaire de A dans Ω	$\bar{A}$
A et B sont **incompatibles**	A et B sont disjoints	$A \cap B = \emptyset$
A **implique** B	A est inclus dans B	$A \subset B$
A **et** B	intersection de A et B	$A \cap B$
A **ou** B	union de A et B	$A \cup B$
A **mais pas** B	A privé de B : différence A moins B	$A \setminus B$
A **ou bien (ou exclusif)** B	différence symétrique de A et B	$A \Delta B$

où

- $A \setminus B = A \cap \bar{B}$ est la différence A moins B qui représente l'ensemble des éléments qui se trouvent dans A mais pas dans B
- $A \Delta B = (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$ est la différence symétrique de A et B qui représente l'ensemble des éléments qui se trouvent soit dans A soit dans B mais pas simultanément dans A et B



Pour bien utiliser les notations ensemblistes et le vocabulaire probabiliste, il faut conserver à l'esprit le sens d'un évènement : c'est un sous-ensemble de tous les résultats possibles pour une expérience. Chacun de ces résultats réalise l'évènement. En ce sens, A et B doit bien se traduire par l'intersection des ensembles A et B : en effet, pour réaliser à la fois A et B, il faut qu'un résultat de l'expérience appartienne aux deux ensembles, ce qui est exactement la définition de $A \cap B$. On raisonne de la même façon pour A ou B et pour $AimpliqueB$. La traduction des connecteurs logiques (et, ou, implique) en opérations ensemblistes facilite le passage d'une formule logique à une description ensembliste pour un évènement.

1.1.4 Tribu

Soit une expérience aléatoire et sont univers Ω, $\mathcal{A}$ un ensemble de sous-ensembles (parties) quelconque de Ω et $\mathcal{P}(\Omega)$ l'ensemble des sous-ensembles (parties) de Ω.

- Quand Ω est fini : on peut s'intéresser à tous les évènements de $\mathcal{P}(\Omega)$. $\mathcal{P}(\Omega)$ est alors une **tribu**
- Quand Ω est infini : on se restreint à une classe d'évènements plus petite de $\mathcal{P}(\Omega)$. Cette classe (ensemble de parties) est appelée tribu (qu'on notera $\mathcal{A}$). Une telle famille est appelée **tribu** si elle vérifie les propriétés suivantes :
  - (i) $\Omega \in \mathcal{A}$
  - (ii) si $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$ (stabilité par complémentarité)
  - (iii) si $A_1, A_2, \ldots$ est une suite finie ou infinie dénombrable d'éléments de $\mathcal{A}$, alors l'évènement $\bigcup_{i \geq 1} A_i \in \mathcal{A}$ (stabilité par union dénombrable)

⚠ **Remarque 1.1.3.** *: Attention ! une tribu sur Ω est un ensemble dont les éléments sont des parties de l'univers Ω.*
Proposition : Soit $\mathcal{A}$ une tribu sur l'univers Ω. Alors

— $\emptyset \in \mathcal{A}$

— si $A_1, A_2, \ldots$ est une suite finie ou infinie dénombrable d'éléments de $\mathcal{A}$, alors l'évènement $\bigcap_{i\geq 1} A_i \in \mathcal{A}$ (stabilité par intersection dénombrable)

Démonstration. Ces propriétés se déduisent des axiomes de la tribu qu'on vient de voir en passant au complémentaire. $\mathcal{A}$ est une tribu sur $\mathcal{A}$, on a donc

— $\Omega \in \mathcal{A} \Rightarrow \overline{\Omega} = \emptyset \in \mathcal{A}$ (axiome 2)

— On a $\overline{\left(\bigcap_{i\geq 1} A_i\right)} = \bigcup_{i\geq 1} \overline{A_i}$. Si $A_{i,i\geq 1} \in \mathcal{A}$ alors $\overline{A_{i,i\geq 1}} \in \mathcal{A}$ et donc de part l'axiome portant sur la stabilité par union dénombrable on a $\bigcup_{i\geq 1} \overline{A_i} \in \mathcal{A}$

□

▷ **Exemple 1.1.10.** *exemples de tribu*

(i) $\mathcal{A} = \{\emptyset, \Omega\}$ *(c'est la plus petite tribu)*

(ii) $\mathcal{A} = 2^{\Omega}$ *(c'est la plus grande tribu)*

(iii) si $A \subset \Omega$, $\mathcal{A} = \{\emptyset, A, \bar{A}, \Omega\}$

⚠ **Remarque 1.1.4.** *Lorsque l'univers Ω est dénombrable (en particulier fini), on prends toujours pour tribu $\mathcal{A} = 2^{\Omega}$: l'ensemble de toutes les parties de Ω (la plus grande tribu)*

1.1.5 Objectifs du calcul des probabilités

Après avoir défini une expérience, évènement, etc, maintenant nous allons voir comment on calcule "les chances" de voir un tel ou tel évènement se réaliser à la suite d'une expérience aléatoire. Ceci est fait par le calcul des *probabilités*

1.1.6 Probabilité

Soit une expérience aléatoire et son univers Ω. à laquelle on associe le couple $(\Omega, \mathcal{A})$ où $\mathcal{A}$ est une tribu de l'univers Ω.

⋄ **Définition 1.1.4.** *On appelle probabilité sur $\mathcal{A}$ (ou mesure de probabilité) une fonction $\mathbb{P}$ de $\mathcal{A}$ dans $[0, 1]$ telle que :*

1. $\mathbb{P}(\Omega) = 1$
2. *pour toute suite de sous-ensembles de Ω, $(A_i)_{i\geq 1}$ disjoints deux à deux (c'est-à-dire tels que pour tout $j \neq k$, $A_j \cap A_k = \emptyset$),*

$$\mathbb{P}\left(\bigcup_{i\geq 1} A_i\right) = \sum_{i\geq 1} \mathbb{P}(A_i)\cdot$$

Cette seconde propriété est la **sigma additivité** *de la mesure de probabilité.*

Il provient immédiatement de la définition 1.1.4,

(i) $\mathbb{P}(\emptyset) = 0$: en choisissant $A_1 = A_2 = \emptyset$, on sait que $0 \leq P(\emptyset) = \lim_{n\to\infty} \mathbb{P}(\emptyset)$ et par conséquent $\mathbb{P}(\emptyset) = 0$;

(ii) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ pour tous $A, B \in \mathcal{A}$ disjoints : en choisissant $A_1 = A, A_2 = B$ et $A_3 = A_4 = \ldots = \emptyset$, on a $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Voici quelques conséquences immédiates de la définition de $\mathbb{P}$. Pour tout $A, B \in \mathcal{A}$, nous avons

1. $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$
2. $A \subset B \Rightarrow \mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq P(A)$
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - P(A \cap B)$

▷ **Exemple 1.1.11.**

— *l'expérience du jeu plie ou face correspond) :* $\Omega = \{pile, face\}$*,* $\mathcal{A} = \{\emptyset, \{pile\}, \{face\}, \Omega\}$ *et* $\mathbb{P}(\emptyset = 0, \mathbb{P}(\{pile\}) = \mathbb{P}(\{face\}) = \frac{1}{2}, \mathbb{P}(\Omega) = 1$

—

—

1.1.7 Cardinaux et opérations ensemblistes

1.1.8 Cardinal d'un ensemble

◇ **Définition 1.1.5.** *Soit A un ensemble fini. On appelle le cardinal de A qui est noté $|A|$, $card(A)$ ou encore $\#A$. le nombre d'éléments de A. L'ensemble vide est de cardinal nul.*

Quand A est fini, on peut numéroter ces éléments et donc écrire

$$A = \{a_1, a_2, \ldots, a_n\}$$

si $n = |A|$.

1.1.9 Cardinaux et opérations ensemblistes

On peut parfois calculer le cardinal du résultat d'une opération ensembliste à partir des cardinaux des ensembles impliquées, comme dans les cas suivants :

Union disjointe soient A et B deux ensembles finis disjoints, c'est-à-dire tels que $A \cap B = \emptyset$. On a alors

$$|A \cup B| = |A| + |B|.$$

Union soient plus généralement deux ensembles finis, on a

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Ensemble des parties soit A un ensemble fini, on a

$$|\mathcal{P}(A)| = 2^{|A|}$$

Produit cartésien soient $A_1, \ldots, A_n$, n ensembles finis, on a

$$|A_1 \times A_2 \times \ldots \times A_n| = |A_1| \times |A_2| \times \ldots \times |A_n|$$

1.1.10 Équiprobabilité

◇ **Définition 1.1.6.** *Soit une expérience aléatoire et son univers* **fini** Ω. *Les évènements élémentaires* $\{\omega_i\}$ *sont dit équiprobables si*

$$\forall\omega \in \Omega,\ \mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}.$$

Dans ce cas on dit que $\mathbb{P}$ *est une probabilité uniforme sur* Ω *car elle associe à chaque évènement élémentaire la même probabilité.*

1.1.11 Espace de probabilité

◇ **Définition 1.1.7.** *Soit* Ω *l'univers d'une expérience aléatoire,* $\mathcal{A}$ *une tribu de* Ω *et* $\mathbb{P}$ *une mesure de probabilité définie sur* $\mathcal{A}$*, alors le triplet* $(\Omega, \mathcal{A}, \mathbb{P})$ *et appelé espace probabilisé ou espace de probabilité. la fonction* $\mathbb{P}$ *est appelée loi de probabilité ou mesure de probabilité sur* $\mathcal{A}$ *ou sur l'espace probabilisable formé par le couple* $(\Omega, \mathcal{A})$

résumé

Pour résumer, la description d'une expérience aléatoire $\mathcal{E}$ se formalise mathématiquement en définissant

1. l'ensemble fondamental Ω (l'univers) définissant l'ensemble des résultats possibles de $\mathcal{E}$, appelés événements élémentaires ;
2. un ensemble $\mathcal{A}$ de parties de Ω, appelées événements : une tribu.
3. une fonction $\mathbb{P} : \mathcal{A} \to [0, 1]$, appelée mesure ou distribution de probabilité, qui à tout événement A associe $\mathbb{P}(A)$, la probabilité de cet événement,

soit l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$

1.1.12 Probabilités conditionnelles

Nous rappelons dans ce paragraphe la notion de conditionnement par rapport à un évènement.

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité associé à une expérience aléatoire.

◇ **Définition 1.1.8.** *Soit un évènement* B *de probabilité* $\mathbb{P}(B) > 0$. *La* **probabilité conditionnelle** *d'un évènement* A **sachant** *l'evènnement* B *représente la probabilité que l'évènement* A *se réalise, sachant que l'évènement* B *s'est réalisé. Elle est noté* $\mathbb{P}(A|B)$ *et est donnée par*

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

Ceci explicite le fait que, il se peut que la réalisation de B, fournisse une information sur la réalisation possible de A. Cette information est représentée par la probabilité conditionnelle $\mathbb{P}(A|B)$. On a également

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

Le calcul de cette probabilité ($\mathbb{P}(A|B)$) s'effectue donc connaissant la probabilité $\mathbb{P}(B)$ et la probabilité $\mathbb{P}(B|A)$. C'est la règle de Bayes que nous allons le voir dans le paragraphe suivant.

1.1.13 Théorème de Bayes

1.1.13.1 Théorème de Bayes : cas de 2 évènements

Soient A et B deux évènements.

Théorème 1.1.1. *Le théorème de Bayes énonce des probabilités conditionnelles : étant donné deux évènements A et B, le théorème de Bayes permet de déterminer la probabilité $\mathbb{P}(A|B)$, si l'on connaît les probabilités $\mathbb{P}(B)$, $\mathbb{P}(B)$ et $\mathbb{P}(B|A)$. En partant de la définition de la probabilité conditionnelle 1.1.8, le théorème de Bayes est donné par :*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}. \tag{1.1}$$

⚠ **Remarque 1.1.5.** *Chaque terme du théorème de Bayes a une dénomination usuelle.*

- *$\mathbb{P}(A)$ est la probabilité a priori de A. Elle est "antérieure" au sens qu'elle précède toute information sur B. $\mathbb{P}(A)$ est aussi appelée la probabilité marginale de A*
- *de même $\mathbb{P}(B)$ est appelé la probabilité marginale ou a priori de B.*
- *$\mathbb{P}(A|B)$ est appelée la probabilité a posteriori de A sachant B. Elle est "postérieure", au sens qu'elle dépend directement de B.*
- *$\mathbb{P}(B|A)$, pour un B connu, est appelé la fonction de vraisemblance de A.*
- *$\frac{\mathbb{P}(B|A)}{\mathbb{P}(B)}$ représente "la connaissance" que B fournisse sur A.*

⚠ **Remarque 1.1.6.** *on a également*

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}.$$

1.1.13.2 Formule des probabilités totales

Soit un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Si $(B_i)_{i\in I}$ un ensemble d'évènements disjoints et dont l'union est l'univers (évènements formant une partition de l'univers), et si quel que soit $i \in I, \mathbb{P}(B_i) > 0$, alors, pour tout évènement A,

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Le théorème de Bayes énoncé précédemment peut donc s'écrire :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\bar{A})\mathbb{P}(\bar{A})}$$

du fait que l'application de la formule des probabilités totales sur les évènements A et $\bar{A}$ permet de calculer $\mathbb{P}(B)$ comme $\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\bar{A})\mathbb{P}(\bar{A})$. Idem pour l'autre forme du théorème de Bayes.
De façon générale, le théorème Bayes prend la forme suivante :

1.1.13.3 Théorème de Bayes : cas de n évènemts

Soient n évènements $A_1, A_2, \ldots, A_n$ deux à deux disjoints de l'univers et dont l'union est l'univers (formant une partition) avec $\sum_{i=1}^{n} \mathbb{P}(A_i) = 1$ et $\mathbb{P}(A_i) > 0\, \forall i$:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{P}(B|A_i)}$$

où on a appliqué la formule des probabilités totales pour le calcul de $\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$.

1.1.14 Indépendance d'événements

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité associé à une expérience aléatoire.

1.1.14.1 Cas de deux évènements

⋄ **Définition 1.1.9.** *Deux évènements A et B (dans $\mathcal{A}$) sont dits indépendants si*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\cdot$$

On note $A \perp B$. Ceci traduit le fait que la réalisation de B n'apporte pas d'information sur la réalisation possible de A.
A ne pas confondre avec le cas où A et B sont disjoints ç-à-d

$$A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

⚠ **Remarque 1.1.7.** *Si l'évènement A et indépendant de l'évènement B alors :*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)\cdot$$

et par équivalence

$$\mathbb{P}(B|A) = \mathbb{P}(B)\cdot$$

(si $\mathbb{P}(B) > 0$).

La définition 1.1.9 se généralise au cas de plusieurs évènements comme suit.

1.1.14.2 Cas de plusieurs évènements

Soient $A_1, A_2, \ldots, A_n$ n évènements de la tribu $\mathcal{A}$. Ces évènements sont dits indépendants si et seulement si, pour toute partie $I=\{i_1,\ldots,i_j\}\subset\{1,2,\ldots,n\}$, de j évènements parmi n $(j = 2, \ldots, n)$ on a

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap, \ldots, \cap A_{i_j}) = \mathbb{P}(A_{i_1}) \times \mathbb{P}(A_{i_2}) \times \ldots \times \mathbb{P}(A_{i_j})\cdot$$

En particulier, pour $j = n$ on a :

$$\mathbb{P}(A_1 \cap A_2 \cap, \ldots, \cap A_n) = \mathbb{P}(A_1) \times \mathbb{P}(A_2) \times \ldots \times \mathbb{P}(A_n)\cdot$$

⚠ **Remarque 1.1.8.** *On utilise souvent la notation du produit cartésien $\prod$ et l'on a :*

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

et pour $I = \{1, 2, \ldots, n\}$ on a

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}(A_i)\cdot$$

1.2 Lois de probabilités discrètes

Une loi de probabilité $\mathbb{P}$ est *concentrée* ou *est portée* sur un ensemble $A \in \mathcal{A}$ lorsque $\mathbb{P}(A) = 1$.

1.2.1 probabilités discrètes : définition

◇ **Définition 1.2.1.** *Une loi de Probabilité $\mathbb{P}$ est dite discrète si elle est concentrée sur un ensemble fini ou dénombrable d'évènements élémentaires.*

◇ **Définition 1.2.2.** *Un ensemble est dit dénombrable, ou infini dénombrable, lorsque ses éléments peuvent être listés sans omission ni répétition dans une suite indexée par les entiers. Ω est donc dit dénombrable si ces éléments peuvent être listés en une suite infini d'évènements élémentaires $\omega_1, \omega_2 \dots$*

1.2.2 Loi de probabilité discrète : définition

◇ **Définition 1.2.3.** *Loi de probabilité discrète*

Soit l'univers $\Omega = \{\omega_1, \omega_2, \dots\}$ un ensemble fini ou dénombrable d'évènements élémentaires ω_i. Soit A un évènement constitué d'évènements élémentaires deux à deux disjoints, on a donc (d'après la propriété de sigma-additivité (additivité des probas d'vènemts disjoints))

$$\mathbb{P}(A) = \sum_{w_i \in A} \mathbb{P}(w_i) = \sum_{w_i \in A} p_i.$$

La loi de probabilité $\mathbb{P}$ est donc entièrement définie par les probabilités élémentaires $\mathbb{P}(\omega_i)$ qu'on note p_i et vérifié les deux propriétés suivantes

Propriétés 1.2.1.

1. $\mathbb{P}(\omega_i) \in [0, 1]$
2. $\sum_{w_i \in \Omega} \mathbb{P}(\omega_i) = 1$ *(ç-à-d* $\sum_{i=1}^{\sharp\Omega} \mathbb{P}(\omega_i) = 1$*)*

Chapitre 2

Variables aléatoires discrètes

2.1 Introduction

Pour définir une variable aléatoire, seul l'espace probabilisable $(\Omega, \mathcal{A})$ suffit ($\mathbb{P}$ on la laisse de côté pour le moment)

2.2 Variables aléatoires

⋄ **Définition 2.2.1.** *Soit un espace probabilisable* $(\Omega, \mathcal{A})$*. Une* **Variable aléatoire** X *est une fonction numérique sur l'univers* Ω *qui pour tout événement élémentaire* ω *(résultat de l'univers), lui associe une valeur* $X(\omega)$ *:*

$$\begin{array}{rcl} X & : & \Omega \longrightarrow E \\ & & \omega \longrightarrow X(\omega) \end{array}$$

$X : \Omega \longrightarrow E$ *est une variable aléatoire (v.a.) si pour tout ensemble* $I \in E$*, l'image réciproque de* I *appartient à la tribu :*

$$\forall I \in E, X^{-1}(I) \in \mathcal{A}\cdot$$

Remarque : Lorsque Ω est dénombrable on prend $\mathcal{A} = 2^{\Omega}$

⋄ **Définition 2.2.2.** *Soit un espace probabilisable* $(\Omega, \mathcal{A})$ *et une v.a.* X*.* X *est une v.a. réèlle si*

$$X \quad : \quad \Omega \longrightarrow \mathbb{R}$$

⚠ **Remarque 2.2.1.** *Il est très pratique d'introduire la notation suivante*

$$\{\omega \in \Omega; X(\omega) \in I\} := \{X \in I\}, I \subset \mathbb{R}.$$

En particulier, nous noterons $\{\omega \in \Omega; X(\omega) \leq x\} := \{X \leq x\}$..

▷ **Exemple 2.2.1.** *Exemples de v.a. : Soit l'expérience qui consiste à jouer deux fois de suite pile ou face. L'univers est dans ce cas* $\Omega = \{pp, pf, fp, ff\}$ *(l'ordre des lancers étant pris en compe). Si l'on désigne par* X *la variable aléatoire représentant le nombre d'apparitions de pile dans cette expérience, on a alors*

$$X(\omega) = \begin{cases} 2 & si \quad \omega = pp \\ 1 & si \quad \omega \in \{pf, fp\} \\ 0 & si \quad \omega = ff \end{cases}$$

On en trouve plusieurs types de variables aléatoires.

2.2.1 Types de variables aléatoires

Soit X un variable aléatoire. X est dite une **variable quantitative** si elle prend des valeurs dans $\mathbb{R}$. Il s'agit d'une variable représentée par une quantité (une valeur) telle que l'âge, le poids et la taille, etc.

Une **variable qualitative** est quant à elle une variable représentée par une qualité (ou catégorie), telles que le sexe, encore l'état civil, le degré de satisfaction d'un service quelconque, etc. La variable qualitative est donc représenté par une modalité plutôt que d'une valeur.

Il existe deux types de variables de variables quantitatives : **variables discrètes** et **variables continues**.

2.2.2 Variables aléatoires discrètes

$\diamond$ **Définition 2.2.3.** *Une variable aléatoire est dite* **discrète** *si elle ne prend que des valeurs discontinues dans un ensemble fini ou dénombrable : c'est à dire si les valeurs x de X sont en nombre fini ou dénombrable ($X(\Omega)$ est fini ou dénombrable).*

Exemples : le résultat du jet d'un dé, le nombre d'enfants dans une famille, le nombre de personnes ayant une maladie rare dans une ville donnée, sont des variables aléatoires discrètes.

$\diamond$ **Définition 2.2.4.** *Une variable aléatoire est dite* **continue** *ou* **a densité** *si elle prend ses valeurs dans $\mathbb{R}$ ou une partie ou un ensemble de parties de $\mathbb{R}$.*

Exemple : la moyenne des étudiants, la taille, etc sont des variables aléatoires continues.

Pour les variable aléatoires qualitatives, il existe également deux types de variables : **variables nominales** et **variables ordinales**.

$\diamond$ **Définition 2.2.5.** *Une variable aléatoire qualitative* **nominale** *est une variable qui correspondent à des noms, il n'y a aucun ordre précis sur les modalités. Ce sont seulement des mots dans le désordre.*

Par exemple, la variable sexe est une variable qualitative nominale qui a deux modalités possibles : féminin ou masculin et dont l'ordre n'importe pas. Un autre exemple est celui de la catégorie socioprofessionnelle

2.3 Probabilité image

Soit un espace probabilisable $(\Omega, \mathcal{A})$ et une v.a. X sur $(\Omega, \mathcal{A})$, alors la fonction p définie par

$$\forall I \in \mathbb{R},\ p(I) = \mathbb{P}\left(X^{-1}(I)\right)$$

est une probabilité. p s'appelle **probabilité image** de $\mathbb{P}$. La probabilité image vérifie les propriétés de probabilités (vues dans 1.1.6)

⚠ **Remarque 2.3.1.** *La fonction $p(I)$ sera notée dans la suite $\mathbb{P}(I)$. Elle est très intéressante parce qu'elle permet le calcul des proababilités de la v.a. X sans continuer à se référer explicitement à l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$*

2.4 Fonction de répartition d'une v.a.

2.4.1 Fonction de répartition d'une v.a.

La fonction de répartition d'une variable aléatoire X, notée F_X, caractérise la loi de probabilité de cette variable. La fonction de répartition F_X associe à toute valeur de x (réelle ou discrète)

$$F_X(x) = \mathbb{P}(X \leq x) \tag{2.1}$$

Elle représente la probabilité que la variable aléatoire X prenne une valeur inférieure ou égale à x. Cette fonction est définition pour une v.a. discrète (et continue) et vérifie les propriétés suivantes :

Propriétés 2.4.1.

(i) $F_X(x) \in [0, 1]$

(ii) $F_X(x)$ *est croissante*

(iii) $\lim\limits_{x\to+\infty} F_X(x) = 1$ *et* $\lim\limits_{x\to-\infty} F_X(x) = 0$

(iv) $F_X(x)$ *est continue à droite :* $F_X(x) = F_X(x^+)$

⚠ **Remarque 2.4.1.** *Si $a < b$, la probabilité que X se trouve dans l'intervalle $]a, b]$ est donc*

$$\mathbb{P}(a < X \leq b) \;=\; F_X(b) - F_X(a)$$

2.4.2 Calcul des probabilités avec la fonction de répartition

Soit une v.a. X est sa fonction de répartition $F_X(x)$. On suppose que $F_X(x)$ est connue, on alors

Propriétés 2.4.2.

(i) $\mathbb{P}(X > x) \;= 1 - F_X(x)$

(ii) $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$

(iii) $\mathbb{P}(X < x)$ *est continue à gauche :* $\mathbb{P}(X < x) = F_X(x^-)$

(iv) $\mathbb{P}(X = x) = F_X(x^+) - F_X(x^-)$

⚠ **Remarque 2.4.2.** *Deux conséquences importantes de cette dernière propriété : v.a. continues et v.a. discrètes :*

- *Si F est continue en x, alors $\mathbb{P}(X = x) = 0$ (car $F_X(x^+) - F_X(x^-) = F_X(x)$. La probabilité attaché à un point d'une v.a. continue et nulle*
- *Si F n'est pas continue en x, alors $F_X(x^+) > F_X(x^-)$ (F étant croissante) et donc $\mathbb{P}(X = x) > 0$.*
  $\Rightarrow$ La probabilité attaché à un point d'une v.a. discrète est non nulle

⚠ **Remarque 2.4.3. Probabilités attachée à un intervalle d'une v.a. discrète :**
On a, la probabilité attachée à un intervalle d'une v.a. discrète est nulle :

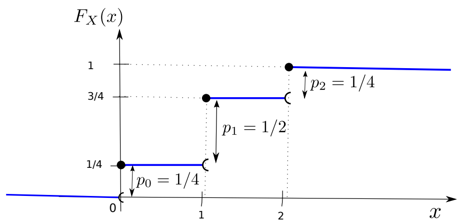
$$\mathbb{P}(x_i < X < x_{i+1}) = \mathbb{P}(X < x_{i+1}) - \mathbb{P}(X \leq x_i) = F_X(x_{i+1}^-) - F_X(x_i) = F_X(x_i) - F_X(x_i) = 0$$

(du fait que $\mathbb{P}(X < x_{i+1})$ est continue à gauche comme vu précédemment)

▷ **Exemple 2.4.1. Exemple d'une fonction de répartition :** *On reprend la v.a. de l'exemple 2.2.1. L'espace probabilisé dans ce cas est* $(\Omega, \mathcal{A}, \mathbb{P})$ *avec* $\Omega = \{pp, pf, fp, ff\}$, $\mathcal{A} = 2^{\Omega}$ *et* $\mathbb{P}(\{pp\}) = \mathbb{P}(\{pf\}) = \mathbb{P}(\{fp\}) = \mathbb{P}(\{ff\}) = 1/4$. *Et on a aussi* $p_0 = \mathbb{P}(X = 0) = 1/4$, $p_1 = \mathbb{P}(X = 1) = 1/2$ *et* $p_2 = \mathbb{P}(X = 2) = 1/4$. *La fonction de répartition de* X *est*

$$F_X(x) = \begin{cases} 0 & si \quad x \in]-\infty 0[\\ 1/4 & si \quad x \in [0, 1[\\ 3/4 & si \quad x \in [1, 2[\\ 1 & si \quad x \in [2, +\infty] \end{cases}$$

et son graphe est



2.5 Loi de probabilités

$\mathbb{P}(X = x)$ est appelée loi de probabilités de la v.a. X et vérifiés les propriétés d'une loi de probabilités vues en 1.2.1 et en particulier

Propriétés 2.5.1. *(i)* $\mathbb{P}(X = x) \in [0, 1]$

(ii) $\sum_{x \in \mathcal{X}} \mathbb{P}(X = x) = 1$

2.5.1 Relation entre la fonction de répartition et la loi de probabilités

Soit X une v.a. réèlle discrète de loi de probabilité $\mathbb{P}(X = x)$ et de fonction de répartition $F_X(x)$, on a

$$F_X(x) = \sum_{x_i \leq x} \mathbb{P}(X = x_i)$$

2.6 Espérance mathématique d'une variable aléatoire discrète

L'espérance mathématique d'une variable aléatoire réelle représente la valeur moyenne prise par cette variable aléatoire.

2.6.1 Définition de l'espérance

Soit X une variable aléatoire discrète prenant ses valeurs dans l'ensemble $\{x_1, \ldots, x_n\}$ avec des probabilités respectives $p_1, \ldots, p_n$ ç.à.d $\mathbb{P}(X = x_k) = p_k$ et $\sum_{k=1}^{n} p_k = 1$.

◇ **Définition 2.6.1.** *L'espérance de X est donnée par*

$$\mathbb{E}[X] = \sum_k x_k \mathbb{P}(X = x_k) = \sum_k p_k x_k \tag{2.2}$$

Pour que l'espérance existe (elle peut ne pas exister), il faut que la série $\sum_k x_k p_k$ soit convergente : $\mathbb{E}[|X|] = \sum_k p_k |x_k| < \infty$.

⚠ **Remarque 2.6.1.** *On peut remarquer que l'espérance d'une variable aléatoire binaire est égale à la sa probabilité de valoir 1, autrement dit si X est une v.a binaire (prenant ses valeurs dans $\{0, 1\}$, on a alors $\mathbb{E}[X] = \mathbb{P}(X = 1)$.*

2.6.2 Propriétés de l'espérance

(i) $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ (linéarité de l'espérance)

(ii) $|\mathbb{E}[XY] \ \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$ (inégalité de Cauchy-Schwartz)

(iii) Linéarité de l'espérance : Soit X une v.a réelle, deux fonctions $g, h : \mathbb{R} \to \mathbb{R}$ et deux réels a et b, alors

$$\mathbb{E}[a.g(X) + b.h(X)] = a.\mathbb{E}[g(X)] + b.\mathbb{E}[h(X)]$$

ainsi le cas particulier affine consiste en la propriété suivante :

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

$\Rightarrow$ On dit que l'espérance est **linéaire**

Conséquence : Soient n variables aléatoires $X_1, \ldots, X_n$. Alors

$$\mathbb{E}[\sum_i X_i] = \sum_i \mathbb{E}[X_i]$$

(iv) positivité de l'espérance : si X est une v.a. positive alors $\mathbb{E}[X] \geq 0$

(v) Croissance de l'espérance : Soient g et h deux fonctions numériques telles que $g(X) \leq h(X)$ alors $\mathbb{E}[g(X)] \leq \mathbb{E}[h(X)]$ (On a en effet $g(X) - h(X) \geq 0$ donc $\mathbb{E}[g(X) - h(X)] = \mathbb{E}[g(X)] - \mathbb{E}[h(X)] > 0$)

(vi) Soi a une constante, alors $\mathbb{E}(a) = a$

(vii) S'il existe une constante a telle que $\mathbb{P}(X \geq a) = 1$, alors $\mathbb{E}[X] \geq a$ (se démontre par l'inégalité de Markov)

(viii) S'il existe une constante b telle que $\mathbb{P}(X \leq b) = 1$, alors $\mathbb{E}[X] \leq b$ (de même)

(ix) S'il existe deux constantes a et b telles que $\mathbb{P}(a \leq X \leq b) = 1$, alors $a \leq \mathbb{E}[X] \leq b$

Théorème 2.6.1. *Soit une fonction φ avec $\sum_k p_k |\varphi(x_k)| < \infty$, alors,*

$$\mathbb{E}[\varphi(X) = \sum_k \mathbb{P}(X = x_k)\varphi(x_k) = \sum_k p_k \ \varphi(x_k)$$

Démonstration. En notant $Y = \varphi(X)$ nous avons

$$\begin{aligned}
\mathbb{E}[\varphi(X)] &= \mathbb{E}[Y] = \sum_{j=1}^{m} \mathbb{P}(Y = y_j) y_j \\
&= \sum_{j=1}^{m} \mathbb{P}(\varphi(X) = y_j) \varphi(x_j) \\
&= \sum_{j=1}^{m} \sum_{k=1|\varphi(X)=y_j}^{n} \mathbb{P}(X = x_k) \varphi(x_k) \\
&= \sum_{k=1}^{n} \mathbb{P}(X = x_k) \varphi(x_k) \\
&= \sum_{k} p_k \varphi(x_k)
\end{aligned}$$

□

2.7 Variance et écart type d'une variable aléatoire discrète

2.7.1 Définition

On appelle variance d'une variable aléatoire X la quantité définie par :

$$\mathrm{var}[X] = \mathbb{E}[(X - E[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \tag{2.3}$$

Ainsi, la variance notée σ_X^2 et sa racine carrée l'écart-type (noté σ_X) mesurent la dispersion de la variable aléatoire X autour de son espérance $\mathbb{E}[X]$.

Cas d'une v.a. discrète :

⚠ **Remarque 2.7.1.** *Comme on sait que* $\mathbb{E}[\varphi(X)] = \sum_x \mathbb{P}(X = x)\varphi(x)$*, la variance d'une v.a. discrète est donc aussi donnée par*

$$\begin{aligned}
\mathrm{var}[X] &= \sum_x (x - \mathbb{E}[X])^2 \mathbb{P}(X = x) \\
\mathrm{var}[X] &= \sum_x x^2 \mathbb{P}(X = x) - (\mathbb{E}[X])^2
\end{aligned}$$

En général, il est plus simple de calculer la variance avec la formule $\mathrm{var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

2.7.2 Propriétés de la variance

Comme la variance est une espérance, ses propriétés se démontrent donc à l'aide des propriétés de l'espérance.

Propriétés 2.7.1. *Soit X une v.a. réèlle.*

(i) $\mathrm{var}(X) \geq 0$

(ii) *Soient deux réels* a *et* b $Var(aX+b) = a^2 Var(X) \Rightarrow$ *On dit que la variance est* **quadratique**

(iii) $\mathrm{var}(X) = \mathrm{var}(-X)$

(iv) $\mathrm{var}(X+c) = \mathrm{var}(X)$ *pour tout réel* c

(v) $\mathrm{var}(c) = 0$ *pour tout réel* c

(vi) *Si* $\mathbb{P}(X=a) = 1$ *alors* $\mathrm{var}(X) = 0$

Démonstration. □

2.8 Inégalités Classiques

2.8.1 Inégalité de Markov

L'inégalité de Markov donne une borne supérieure de la probabilité qu'une variable aléatoire réelle positive soit supérieure ou égale à une constante positive. (Elle relie cette probabilité à l'espérance de la v.a.)

Théorème 2.8.1. *(***Markov***) Soit* X *une variable aléatoire réelle positive, à valeurs dans* $E \subset \mathbb{R}$*, définie sur un espace probabilisé* $(\Omega, \mathcal{A}, \mathbb{P})$*, alors*

$$\forall a > 0, \qquad \mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}[X]}{a} \cdot$$

Démonstration. On a :

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x \in E} x\, \mathbb{P}(X = x) \\ &= \sum_{x \in A} x\, \mathbb{P}(X = x) + \sum_{x \in \bar{A}} x\, \mathbb{P}(X = x) \end{aligned}$$

avec $A = \{x \in E | x \geq a\}$ et $\bar{A}$ le complémentaire de A dans E. Donc

$$\begin{aligned} \mathbb{E}[X] &\overset{x \geqslant 0}{\geq} \sum_{x \in A} x\, \mathbb{P}(X = x) \\ &\overset{x \geqslant a}{\geq} \sum_{x \in A} a\, \mathbb{P}(X = x) \\ &\geq a \sum_{x \in A} \mathbb{P}(X = x) = a\, \mathbb{P}(X \in A) = a\, \mathbb{P}(X \geq a). \end{aligned}$$

□

De cette inégalité, il en découle un corollaire, très connu également et qui donne une borne supérieure de la probabilité qu'une fonction positive d'une variable aléatoire réelle soit supérieure ou égale à une constante positive.

Corollaire 2.8.1. *Soit* f *une fonction croissante et positive ou nulle sur l'intervalle* I*. Soit* X *une variable aléatoire réelle définie sur un espace probabilisé* $(\Omega, \mathcal{A}, \mathbb{P})$*, et telle que* $\mathbb{P}(X \in I) = 1$*. Alors*

$$\forall a \in I, \text{ tel que } f(a) > 0, \qquad \mathbb{P}(f(X) \geqslant a) \leqslant \frac{\mathbb{E}[f(X)]}{a}.$$

Démonstration. même chose que pour la démonstration 2.8.1 : On a :

$$\begin{aligned}
\mathbb{E}[f(X)] &= \sum_{x\in E} f(x)\,\mathbb{P}(X=x) \\
&= \sum_{x\in A} f(x)\,\mathbb{P}(X=x) + \sum_{x\in \bar{A}} f(x)\,\mathbb{P}(X=x)
\end{aligned}$$

avec $A=\{x\in E|f(x)\geq a\}$ et $\bar{A}$ le complémentaire de A dans E. Donc

$$\begin{aligned}
\mathbb{E}[f(X)] &\geq \sum_{x\in A} f(x)\,\mathbb{P}(X=x) \\
&\geq \sum_{x\in A} a\,\mathbb{P}(X=x) \\
&\geq a\sum_{x\in A} \mathbb{P}(X=x) = a\,\mathbb{P}(X\in A) = a\,\mathbb{P}(f(X)\geq a).
\end{aligned}$$

□

2.8.2 Inégalité de Bienaymé-Tchebychev

L'inégalité de Bienaymé-Tchebychev permet de montrer qu'une variable aléatoire prendra presque sûrement une valeur proche de son espérance. (Elle relie cette probabilité avec la variance de cette variable aléatoire)

Théorème 2.8.2. *(**Tchebychev**) Soit X une variable aléatoire réelle à valeurs dans E définie sur un espace probabilisé $(\Omega,\mathcal{A},\mathbb{P})$, . Alors*

$$\forall\varepsilon>0, \qquad \mathbb{P}\big(|X-\mathbb{E}[X]|\geqslant\varepsilon\big)\leqslant\frac{\mathrm{var}(X)}{\varepsilon^2}.$$

Démonstration. On applique l'inégalité de Markov (corollaire 2.8.1) avec $f(X)=|X-\mathbb{E}[X]|^2$ et $a=\varepsilon^2$. □

2.8.3 Inégalité de Jensen

Soit f une fonction convexe sur un intervalle réel I et$X$ une variable aléatoire à valeurs dans I, dont les espérances $\mathbb{E}(X)$ et $\mathbb{E}[f(X)]<\infty$ existent ($\mathbb{E}(|X|)<\infty$ et $\mathbb{E}[|f(X)|]<\infty$). Alors,

$$f\left(\mathbb{E}\left(X\right)\right)\leq\mathbb{E}[f(X)].$$

2.9 Moments d'une v.a. et fonction génératrice des moments

2.9.1 Moments d'une v.a.

Pour l'instant on s'était intéressé à l'espérance et la variance d'une v.a. Comme on va le voir, ces deux statistiques correspondent à deux cas de ce que l'on appelle les moments. Plus généralement le moment d'ordre r ($r>0$) de la v.a X est

$$\mathbb{E}[X^r]=\sum_x x^r\mathbb{P}(X=x)$$

et respectivement de moment centré d'ordre r ($r > 0$) la quantité donnée par

$$\mathbb{E}[(X - \mathbb{E}[X])^r] = \sum_x (x - \mathbb{E}(X))^r \, \mathbb{P}(X = x).$$

On peut tout de suite remarquer que l'espérance mathématique d'une v.a. correspond au moment d'ordre 1 et que la variance correspond au moment centré d'ordre 2.

⚠ **Remarque 2.9.1.** *Notons que les puissances non entières ne sont définies que pour les nombres positifs* $x^r = \exp(r \ln(x)),\ x > 0.$

2.9.2 Fonction génératrice des moments

La fonction génératrice des moments d'une variable aléatoire X engendre les moments associés à la distribution de probabilités de la variable aléatoire X. Elle est définie par

$$M_X(t) = \mathbb{E}\left(e^{tX}\right), \quad t \in \mathbb{R}\cdot$$

Pour une v.a. discrète, la fonction génératrice des moments de X est donc donnée par

$$M_X(t) = \mathbb{E}\left(e^{tX}\right) = \sum_x e^{tx}\mathbb{P}(X = x), \quad t \in \mathbb{R}\cdot$$

Dans la suite, toutes les variables aléatoires considérées seront discrètes (i.e. l'ensemble $X(\Omega)$ des valeurs prises par X est fini ou dénombrable).

Chapitre 3

Couples de variables aléatoires discrètes

Soient $X : \Omega \to \mathcal{X}$ et $Y : \Omega \to \mathcal{Y}$ deux variables aléatoires discrètes définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Notons $p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$ la loi du couple (X, Y) et $p_X(x) = \mathbb{P}(X = x)$ la loi de X.

3.1 Loi jointe d'un couple de variables aléatoires discrètes

Soient X et Y deux variables discrètes telle que X prend ses valeurs dans l'ensemble $\mathcal{X}$ et $\mathcal{Y}$ prend ses valeurs dans l'ensemble $\mathcal{Y}$. La loi jointe du couple (X, Y) est définie par l'ensemble des probabilités :

$$\mathbb{P}(X = x, Y = y) \quad \text{avec } x \in \mathcal{X} \text{ et } y \in \mathcal{Y}.$$

⚠ **Remarque 3.1.1.** *Cette loi peut être représenté sous la forme d'un tableau dans le cas où les v.a prennent un nombre petit de valeurs*

Y/X	$\dots$	x	$\dots$	Somme des colonnes
$\vdots$				
y		$\mathbb{P}(X = x, Y = y)$		$\mathbb{P}(Y = y)$
$\vdots$				
Somme des lignes		$\mathbb{P}(X = x)$		

TABLE 3.1 – Loi jointe de couple de v.a discrètes

3.1.1 Loi marginale d'une v.a. discrète

La loi de chaque variable est donc donnée par la loi marginale comme suit :

$$\mathbb{P}(X = x) = \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y) \text{ et } \mathbb{P}(Y = y) = \sum_{x \in \mathcal{X}} \mathbb{P}(X = x, Y = y) \quad (3.1)$$

On peut remarquer que ces lois correspondent respectivement à la somme des lignes et la somme des colonnes dans le tableau précédent (3.1).

Les variables X et Y sont des variables aléatoires discrètes de lois respectives

$$P_X(x) = \sum_{y \in B} \mathbb{P}(X = x, Y = y) \tag{3.2}$$

et

$$P_X(y) = \sum_{x \in A} \mathbb{P}(X = x, Y = y) \tag{3.3}$$

3.2 Lois conditionnelles et théorème de Bayes

Soient X et Y deux variables aléatoires de loi respectives $p_X(x)$ $\mathbb{P}(X)$ et $p_Y(y)$ $\mathbb{P}(Y)$. Le comportement du couple (X, Y) est entièrement décrit par leur **distribution jointe** $p_{X,Y}(x, y) = \mathbb{P}_{XY}(X = x, Y = y)$. Il se peut que lors du tirage d'une réalisation de (X, Y), que la valeur observée (réalisation) x de X fournisse une information sur la valeur possible de Y. Cette information est représentée par la distribution de Y conditionnellement à $X = x$ (distribution de Y étant donné à $X = x$) soit $p_{Y|X}(y|x)$. Ceci est explicité par le théorème de suivant, appelé théorème de Bayes (ou aussi règle de Bayes ou formule de Bayes)

Théorème 3.2.1.

$$\mathbb{P}(Y = y, X = x) = \mathbb{P}(Y = y|X = x)\mathbb{P}(X = x)$$

et par symétrie on a également

$$\mathbb{P}(Y = y, X = x) = \mathbb{P}(X = x|Y = y)\mathbb{P}(Y = y) \ ,$$

et donc

$$\mathbb{P}(Y = y|X = x) = \frac{\mathbb{P}(X = x|Y = y)\mathbb{P}(Y = y)}{\mathbb{P}(X = x)}$$

3.2.1 Loi conditionnelle d'une v.a. discrète

Soient $X : \Omega \to E$ et $Y : \Omega \to F$ deux variables aléatoires discrètes définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Notons $p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$ la loi du couple (X, Y) et $p_X(x) = \mathbb{P}(X = x)$ la loi de X.

L'évènement $\{Y = y\} \forall y \in F$, conditionnellement a $\{X = x\}$ (avec $\mathbb{P}(X = x) > 0$, a une probabilité

$$\begin{aligned}\mathbb{P}(Y = y|X = x) &= \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)} \\ &= \frac{p_{X,Y}(x, y)}{p_X(x)}\end{aligned}$$

Les valeurs $\mathbb{P}(Y = y|X = x)$ $y \in F$ définissent une probabilité sur F et représentent la distribution de Y sachant $X = x$.

$\diamond$ **Définition 3.2.1. Loi Conditionnelle d'une v.a. discrète** *La loi conditionnelle de Y sachant $X = x$, notée généralement $p_{Y|X}(.|x)$ est définie par*

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

pour chaque $x \in E$ tel que $p_X(x) > 0$.

3.2.2 Espérance conditionnelle d'une v.a.

3.2.2.1 Conditionnement par rapport à un évènement

Soit $X : \Omega \to \mathcal{X}$ une v.a. réelle discrète et soit B un évènement de probabilité $\mathbb{P}(B) > 0$.

$\diamond$ **Définition 3.2.2.** *L'espérance conditionnelle de X sachant l'évènement B est définie par*

$$\mathbb{E}[X|B] = \sum_{x \in \mathcal{X}} x\, \mathbb{P}(X = x|B)$$

$\triangleright$ **Exemple 3.2.1.** *On lance un dé à 6 faces et on note par X la valeur du résultat à chaque lancé. Si B est l'évènement "X est supérieure ou égale à trois", alors l'espérance de X sachant B est*

$$\mathbb{E}[X|B] = \sum_{x=1}^{6} x\, \mathbb{P}(X = x|B)$$

Comme $\mathbb{P}(B) = 2/3$ (car $\mathbb{P}(B) = \mathbb{P}(\{3\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{5\}) + \mathbb{P}(\{6\}) = 4 \times 1/6 = 2/3$), on a alors, d'après le théorème de Bayes

$$\begin{aligned}
\mathbb{P}(X = x|B) &= \frac{\mathbb{P}(\{X = x\} \cap B)}{\mathbb{P}(B)} \\
&= \frac{3}{2}\mathbb{P}(\{X = x\} \cap B) \\
&= \frac{3}{2} \times \begin{cases} 0 & \textit{si } x \in \{1, 2\} \\ \frac{1}{6} & \textit{sinon } (i.e. \textit{ si } x \in \{3, 4, 5, 6\}) \end{cases} \\
&= \begin{cases} 0 & \textit{si } x \in \{1, 2\} \\ \frac{1}{4} & \textit{sinon } (i.e. \textit{ si } x \in \{3, 4, 5, 6\}) \end{cases}
\end{aligned}$$

et donc finalement

$$\mathbb{E}[X|B] = (1 + 2) \times 0 + (3 + 4 + 5 + 6) \times \frac{1}{4} = \frac{9}{2}.$$

Dans la suite, nous allons considérer le conditionnement par rapport à une v.a.

3.2.2.2 Conditionnement par rapport à une v.a.

$\diamond$ **Définition 3.2.3.** *L'espérance de Y pour la loi conditionnelle $\mathbb{P}(Y = .|X = x)$ s'appelle espérance conditionnelle de Y sachant $X = x$ et vaut :*

$$\begin{aligned}
\mathbb{E}[Y|X = x] &= \sum_{y \in \mathcal{Y}} y\, \mathbb{P}(Y = y|X = x) \\
&= \sum_{y \in \mathcal{Y}} y\, p_{Y|X}(y|x)
\end{aligned}$$

qui est une fonction de x.

Soit $\psi(x) = \mathbb{E}[Y|X = x]$. La variable aléatoire $\psi(X)$ est **l'espérance conditionnelle de Y sachant X** et se note $\psi(X) = \mathbb{E}[Y|X]$.

⚠ **Remarque 3.2.1.** *$\mathbb{E}[Y|X = x]$ est un réel alors que $\mathbb{E}[Y|X]$ est une v.a.*

▷ **Exemple 3.2.2.** *On lance un dé à 6 faces et on note par X la valeur du résultat obtenu. Ensuite on relance le dé $X = x$ fois de façon indépendante et on note Y la v.a. représentant le produit des valeurs obtenus lors des x lancers. Par exemple on peut calculer $\mathbb{E}[Y|X = 4]$; soit la valeur moyenne des résultats pour 4 lancers. On a $X = 4$ donc pour les 4 lancers indépendants*

$$\mathbb{E}[Y|X = 4] = \mathbb{E}[X_1 \times \ldots \times X_4] = \mathbb{E}[X_1] \times \ldots \times \mathbb{E}[X_4]$$

où $X_i, i = 1, ..., 4$ est la v.a. représentant le résultat du ième lancer. Comme,

$$\mathbb{E}[X_i] = \mu = \sum_{k=1}^{6} k\, p_k = \sum_{k=1}^{6} k\, \frac{1}{6} = \frac{7}{2}$$

(qui représente la valeur moyenne pour un lancer), on a donc

$$\mathbb{E}[Y|X = 4] = \mu \times \ldots \times \mu = \mu^4$$

qui est bien un réel. On voit bien que lorsqu'on connaît X on s'attend à avoir en moyenne pour Y la valeur μ^X et on a l'espérance conditionnelle de Y sachant X

$$\mathbb{E}[Y|X] = \mu^X$$

qui est une variable aléatoire.

3.3 Covariance de deux variables aléatoires

Quand il s'agit d'une variable aléatoire, on parle de variance comme dans les définitions (2.3)(2.4). Quand il s'agit d'un couple de deux variables aléatoires différentes, on parle de **covariance** de deux variables aléatoires.

On définit la covariance de deux variables aléatoires X et Y par le terme

$$\mathrm{Cov}(X, Y) \equiv \mathbb{E}[(X - \mathbb{E}[X])\,(Y - \mathbb{E}[Y])] \tag{3.4}$$

et on la propriété suivante

$$\mathrm{Cov}(X, Y) = \sigma_{XY} = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \tag{3.5}$$

⚠ **Remarque 3.3.1.** *Attention ! L'espérance du produit $\mathbb{E}[XY]$ se calcule à partir de la loi jointe du couple (X, Y). Ainsi, dans le cas discret, si X prend ses valeurs dans A et Y prends ses valeurs dans B, on a*

$$\mathbb{E}[XY] = \sum_{x\in A}\sum_{y\in B} xy\mathbb{P}(X = x, Y = y). \tag{3.6}$$

La covariance mesure la dépendance entre deux v.a. Si les deux v.a sont indépendantes, leur covariance est donc nulle.

3.3.1 Inégalité de Cauchy-Schwartz

L'inégalité de Cauchy-Schwartz permet de contrôler en espérance les fluctuations jointes de (X, Y) à l'aide des variances individuelles de X et Y,

Pour tout couple aléatoire discret (ou continu) (X, Y) nous avons

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2].$$

avec égalité si et seulement s'il existe $(a, b) \in \mathbb{R}^2$ dont l'un au moins est non nul tels que $\mathbb{P}(aX = bY) = 1$

3.4 Corrélation entre deux variables aléatoires

La corrélation entre deux variables aléatoires se mesure par le coefficient de corrélation linaire noté ρ.

◇ **Définition 3.4.1.** *Le coefficient de corrélation linéaire entre deux variables aléatoires X et Y est défini par :*

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}. \tag{3.7}$$

On a toujours $\rho_{X,Y} \in [-1, 1]$. Plus $|\rho_{X,Y}|$ est proche de 1 plus la corrélation entre les variables X et Y est forte.

⚠ **Remarque 3.4.1.** *Si X et Y sont indépendantes ou décorrélées, alors $Cov(X, Y) = 0$ et donc $\rho_{X,Y} = 0$. On a par conséquent, $Var(X + Y) = Var(X) + Var(Y)$.*

3.5 Indépendance de deux variables aléatoires

3.5.1 Indépendance et loi de probabilités

Soient X et Y deux variables aléatoires de lois respectives $p_X(x)$ et $p_Y(y)$

◇ **Définition 3.5.1.** *X et Y sont dites indépendantes si et seulement si leur loi jointe est égale au produit de lois marginales :*

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

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

Cela vient du fait que, d'une part le comportement du couple (X, Y) est entièrement décrit par leur distribution jointe $p_{XY}(x, y)$ et d'autre part, après le tirage d'une réalisation de (X, Y), la valeur observée (réalisation) x de X fournit une certaine quantité d'information sur la valeur de Y. Cette information est représentée par la distribution de Y conditionnellement à $X = x$ soit $p_{Y|X}(y|x)$. X et Y sont dites indépendantes si observer une réalisation x de X n'apporte aucune information sur la réalisation possible de Y étant donné x. Autrement dit, la distribution de Y conditionnellement à X ne dépend pas de x.

La loi jointe $p_{XY}(x, y)$ est :

$$p_{XY}(x, y) = p_{Y|X}(y|x)p_X(x) \tag{3.8}$$

par le théorème de Bayes. Ensuite la densité marginale de Y est obtenue en intégrant cette densité jointe par rapport à x :

$$p_Y(y) = \sum_{x\in\mathcal{X}} p_{XY}(x,y) = \sum_{x\in\mathcal{X}} p_{Y|X}(y|x)p_X(x) \tag{3.9}$$

Mais si la loi conditionnelle de Y ne dépend pas de x (cas de v.a. indépendantes), on peut la sortir de la somme et nous avons alors

$$p_Y(y) = p_{Y|X}(y|x)\sum_{x\in\mathcal{X}} p_X(x) = p_{Y|X}(y|x) \tag{3.10}$$

donc

$$p_{Y|X}(y|x) = p_Y(y) \tag{3.11}$$

et par conséquent

$$p_{XY}(x,y) = p_X(x)p_Y(y) \tag{3.12}$$

Plus généralement, deux variables aléatoires réèlles X et Y sont indépendantes si pour tout intervalle A et B de $\mathbb{R}$ on a

$$\mathbb{P}(X\in A, Y\in B) = \mathbb{P}(X\in A)\mathbb{P}(Y\in B).$$

La condition d'indépendance s'exprime en termes de distributions de probabilité, mais également en termes de fonction de répartition.

3.5.2 Indépendance et fonction de répartition

Soient X et Y deux variables aléatoires de fonctions de répartitions respectives $F_X(x)$ et $F_Y(y)$

◇ **Définition 3.5.2.** *X et Y sont dites indépendantes si et seulement si leur fonction de répartition jointe est égale au produit de fonctions de répartitions marginales :*

$$\begin{aligned} F_{X,Y}(x,y) &= \mathbb{P}(X\leq x, Y\leq y) \\ &= \mathbb{P}(X\leq x)\mathbb{P}(Y\leq y) \\ &= F_X(x)F_Y(y) \end{aligned}$$

3.5.3 Indépendance et espérance

X et Y sont indépendantes si et seulement si pour toutes fonctions $f(X)$ et $g(Y)$

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$$

si les espérances existent. Un cas particulier est bien celui où $f(X) = X$ et $g(Y) = Y$. On peut alors énoncer le théorème ainsi Si X et Y sont indépendantes, alors

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Une conséquence de ce théorème est que deux variables indépendantes sont décorrélées (leur covariance est nulle).

⚠ **Remarque 3.5.1.** *La réciproque est fausse : la décorrélation de deux variables n'implique pas indépendance, sauf dans le cas v.a à densité normale comme on le verra plus tard dans le chapitre dédié aux variables aléatoires gaussiennes.*

Chapitre 4

Vecteurs Aléatoires discrets

4.1 Vecteurs Aléatoires discrets

Dans cette partie, on généralise les définitions et réultats vus précédemment dans le cas d'une v.a. ou d'un couple de v.a., au cas d'un vecteur aléatoire.

$\diamond$ **Définition 4.1.1.** *Soit un espace probabilisable* $(\Omega, \mathcal{A})$ *et une fonction* $\mathbf{X} : (\Omega, \mathcal{A}) \to \mathbb{R}^n$. $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ *est un* **vecteur aléatoire** *réel de dimension* n *(ou une v.a. réelle de dimension* n*) si pour tout intervalle* $\boldsymbol{I}$ *de* $\mathbb{R}^n$*, l'image réciproque de l'intervalle appartient à la tribu :*

$$\forall \boldsymbol{I} \in \mathbb{R}^n, \boldsymbol{X}^{-1}(\boldsymbol{I}) \in \mathcal{A}\cdot$$

avec

$$\boldsymbol{X}^{-1}(\boldsymbol{I}) = \{\omega; \exists x_1 \in I_1, \ldots, \exists x_n \in I_n | X_1(\omega) = x_1, \ldots, X_n(\omega) = x_n\}$$

Remarque : $\boldsymbol{X}(\omega)$ *est le vecteur* $\boldsymbol{X}(\omega) = \big(X_1(\omega), \ldots, X_n(\omega)\big)^T$ *et* $\boldsymbol{I} = (I_1, \ldots, I_n)$.

Un vecteur aléatoire réel est donc un vecteur dont les composantes sont des variables aléatoires réelles.

4.1.1 Fonction de répartition

La fonction de répartition d'un vecteur aléatoire $\boldsymbol{X}$ décrit la loi de probabilité d'un vecteur aléatoire et est donnée par

$$F_{\boldsymbol{X}}(\mathbf{x}) = F_{\boldsymbol{X}}(x_1, \ldots, x_n) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) \tag{4.1}$$

Propriétés 4.1.1.

(i) $F_{\boldsymbol{X}}(x_1, \ldots, x_n) \in [0, 1]$

(ii) $F_{\boldsymbol{X}}(x_1, \ldots, x_n)$ *est croissante par rapport à chacune des variables* x_i

(iii) $\lim\limits_{\substack{x_i \to +\infty \\ i=1,\ldots,n}} F_{\boldsymbol{X}}(x_1, \ldots, x_n) = 1$ *et* $\lim\limits_{\substack{x_i \to -\infty \\ i=1,\ldots,n}} F_{\boldsymbol{X}}(x_1, \ldots, x_n) = 0$

(iv) $F_{\boldsymbol{X}}(x_1, \ldots, x_n)$ *est continue à droite :* $F_{\boldsymbol{X}}(x_1, \ldots, x_n) = F_{\boldsymbol{X}}(x_1^+, \ldots, x_n^+)$

4.1.2 Loi de probabilité

Lorsque $\boldsymbol{X}$ est un vecteur aléatoire discret, la loi de probabilité (fonction de masse de probabilité) se définit par

$$p_{\boldsymbol{X}}(\mathbf{x}) = \mathbb{P}(\boldsymbol{X} = \mathbf{x}) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n).$$

On a

$$\mathbb{P}(\boldsymbol{X} \in A) = \sum_{\mathbf{x} \in A} p_{\boldsymbol{X}}(\mathbf{x}),$$

et

$$\sum_{\mathbf{x} \in \mathcal{X}} p_{\boldsymbol{X}}(\mathbf{x}) = 1,$$

$\mathcal{X}$ étant l'ensemble de valeurs que prend $\boldsymbol{X}$.

Propriétés 4.1.2.

(i) $p(\mathbf{x}) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \in [0, 1]$

(ii)

$$\sum_{\mathbf{x}} p(\mathbf{x}) = \sum_{x_1} \sum_{x_2} \ldots \sum_{x_n} \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = 1$$

(iii)

$$F_{\boldsymbol{X}}(x_1, \ldots, x_n) = \sum_{x_{1_j} \leq x_1} \sum_{x_{2_j} \leq x_2} \ldots \sum_{x_{n_j} \leq x_n} \mathbb{P}(X_{1_j} = x_{1_j}, X_{2_j} = x_{2_j}, \ldots, X_{n_j} = x_{n_j})$$

4.2 Espérance d'un vecteur aléatoire

Soit $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ un vecteur aléatoire réel. Notons par $\mu_j = \mathbb{E}[X_j]$ l'espérance de chaque v.a composante X_j.

$\diamond$ **Définition 4.2.1.** *L'espérance du vecteur aléatoire* $\mathbf{X}$ *est donnée par le vecteur déterministe*

$$\mathbb{E}[\boldsymbol{X}] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n])^T = (\mu_1, \ldots, \mu_n)^T$$

4.3 Matrice de covariance

La matrice de variance-covariance (appelée aussi matrice de covariance) d'un vecteur aléatoire $\boldsymbol{X}$ est la matrice carrée parfois notée $\boldsymbol{\Sigma}$ dont le terme général est donné par :

$$\boldsymbol{\Sigma}_{i,j} = \text{cov}\,(X_i, X_j)$$

Elle est définie comme :

$$\boldsymbol{\Sigma}_{\boldsymbol{X}} = var(\boldsymbol{X}) = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^T\right], \tag{4.2}$$

et donc donnée aussi par

$$\boldsymbol{\Sigma}_{\boldsymbol{X}} = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \mathbb{E}[\boldsymbol{X}](\mathbb{E}[\boldsymbol{X}])^T, \tag{4.3}$$

$\mathbb{E}[\boldsymbol{X}]$ étant l'espérance mathématique de $\boldsymbol{X}$. En développant les termes on obtient la forme suivante :

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathrm{E}[(X_1-\mu_1)(X_1-\mu_1)] & \mathrm{E}[(X_1-\mu_1)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_1-\mu_1)(X_n-\mu_n)] \\ \mathrm{E}[(X_2-\mu_2)(X_1-\mu_1)] & \mathrm{E}[(X_2-\mu_2)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_2-\mu_2)(X_n-\mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n-\mu_n)(X_1-\mu_1)] & \mathrm{E}[(X_n-\mu_n)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_n-\mu_n)(X_n-\mu_n)] \end{bmatrix}$$

$$= \begin{pmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_1X_2) & \cdots & \mathrm{cov}(X_1X_p) \\ \mathrm{cov}(X_2X_1) & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_PX_1) & \cdots & \cdots & \mathrm{var}(X_p) \end{pmatrix} = \begin{pmatrix} \sigma^2_{X_1} & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_p} \\ \sigma_{X_2X_1} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_pX_1} & \cdots & \cdots & \sigma^2_{X_p} \end{pmatrix}$$

où les $\sigma^2_{X_j}, j = 1, \ldots, p$ sont les variances respectives des v.a X_j et les $\sigma^2_{X_i,X_j}, i \neq j$ sont les covariances respectives des couples de v.a (X_i, X_j) : $\sigma^2_{X_i,X_j} = cov(X_i, X_j)$.

La matrice de covariance possède les propriétés suivantes :

1. La matrice est **symétrique** car on a $cov(X,Y) = cov(Y,X)$
2. Elle est **semi- définie positive** (ses valeurs propres sont positives ou nulles).
3. Les éléments de sa diagonale ($\boldsymbol{\Sigma}_{i,i}$) représentent la variance de chaque variable, étant donné la propriété $cov(X,X) = var(X)$
4. Les éléments en dehors de la diagonale ($\boldsymbol{\Sigma}_{i,j}, i \neq j$) représentent la covariance entre les variables i et j.
5. etc

4.4 Matrice de corrélation

La matrice de corrélation (notée R) du vecteur aléatoire $\boldsymbol{X}$, définie de manière analogue à la matrice de covariance, est la matrice dont le terme général est le cœfficient de corrélation linéaire $\rho_{i,j}$ donné par l'équation (3.7). Chaque terme $\rho_{i,j}$ représente la corrélation entre le couple de variables (X_i, X_j) du vecteur $\boldsymbol{X}$ et pour rappel est donné par $\rho_{i,j} = \frac{Cov(X_i,X_j)}{\sqrt{Var(X_i)Var(X_j)}} = \frac{\sigma_{i,j}}{\sigma_i\sigma_j}$.

⚠ **Remarque 4.4.1.** *Toutes les valeurs de la matrice de corrélation sont donc comprises entre -1 et +1 et les termes de la diagonale sont égaux à 1.*

4.5 Indépendance de vecteurs aléatoires discrets

notion de v.a. i.i.d

Soit $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ une suite de variables aléatoires discrètes définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et soit $(S_1, S_2, ..., S_n)$ une suite d'ensembles finis ou dénombrables tels que, pour tout $i \leq n$, $\mathbb{P}(X_i \in S_i) = 1$.

4.5.1 Indépendance et loi de probabilité

Alors $(X_1, X_2, ..., X_n)$ est une suite de variables aléatoires indépendantes (mutuellement indépendantes) si et seulement si, pour tout $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \prod_{i=1}^{n} S_i$,

$$\mathbb{P}(\boldsymbol{X} = \mathbf{x}) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i).$$

La loi jointe est donc égale aux produit des lois marginales

4.5.2 Indépendance et fonction de répartition

$(X_1, \ldots, X_n)$ sont dites indépendantes si et seulement si leur fonction de répartition jointe est égale au produit des fonctions de répartitions marginales :

$$\begin{aligned} F_{\boldsymbol{X}}(x_1, \ldots, x_n) &= \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) \\ &= \mathbb{P}(X_1 \leq x_1) \times, \ldots, \times \mathbb{P}(X_n \leq x_n) \\ &= F_{X_1}(x_1) \times, \ldots, \times F_{X_n}(x_n) \\ &= \prod_{i=1}^{n} F_{X_i}(x_i) \end{aligned}$$

4.5.3 Indépendance et espérance

Si $(X_1, \ldots, X_n)$ sont des v.a. indépendantes alors l'espérance du produit des v.a est égale au produit des espérance :

$$\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathbb{E}[X_i].$$

4.5.4 Espérance de plusieurs variables aléatoires

Soient $(X_1, \ldots, X_n)$ n v.a. alors

Propriétés 4.5.1. *(i) on a*

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i]$$

(ii) Si $a_1, \ldots, n_n$ et b des réels alors

$$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i] + b.$$

4.5.5 Variance de la somme de variables aléatoires indépendantes

Soient $(X_1, \ldots, X_n)$ n v.a. Indépendantes alors

Propriétés 4.5.2. *(i) on a*

$$\operatorname{var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \operatorname{var}[X_i]$$

(ii) Si $a_1, \ldots, n_n$ et b des réels alors

$$\operatorname{var}\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i^2 \operatorname{var}[X_i].$$

4.6 Loi des grands nombres

4.6.1 Moyenne empirique

Soit $(X_1, \ldots, X_n)$ une suite de v.a. de distribution de probabilités $p_X(x)$ on appelle *moyenne empirique* (ou moyenne arithmétique) la quantité notée $\bar{X}$ et donnée par

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

4.6.1.1 Convergence en probabilité : Définition

On dit qu'une suite (X_n) de v.a. converge en probabilité vers une v.a. X si

$$\forall \epsilon > 0, \qquad \lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \epsilon\right) = 0. \tag{4.4}$$

Théorème 4.6.1 (Loi des grands nombres)**.** *Soit $X_1, \ldots, X_n$ un échantillon indépendant d'une suite de v.a. indépendantes et de même loi d'espérance $\mathbb{E}[X] = \mu$ et de variance* $\operatorname{var}(X) = \sigma^2$ *. Alors on a :*

$$\forall \epsilon > 0, \quad \lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mathbb{E}(X)\right| \geqslant \epsilon\right) = 0 \tag{4.5}$$

$\Rightarrow$ La moyenne empirique $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ converge en probabilité vers l'espérance $\mathbb{E}[X]$ quand $n \to +\infty$ (asymptotiquement)

⚠ **Remarque 4.6.1.** *Ici on parle donc de convergence en probabilité.*

$\Rightarrow$ La L.G.N nous dit que, pour tout réel ϵ strictement positif, la probabilité que la moyenne empirique $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ s'éloigne de l'espérance $\mathbb{E}[X] = \mu$ d'au moins ϵ tend vers 0 quand la taille de l'échantillon n tend vers l'infini.

4.7 Théorème central limite

La grande importance pratique associée à la distribution normale découle du théorème central limite présenté ci-dessous (thèorème 4.7.1).

Théorème 4.7.1. Le Théorème Central Limite*. Soit $X_1, X_2, \cdots, X_n$ une suite de variables aléatoires réelles mutuellement indépendantes et suivant la même densité f (variables i.i.d) d'espérance μ et d'écart-type σ. Soit la variable aléatoire somme*

$$S = X_1 + X_2 + \cdots + X_n$$

Alors la densité de probabilité de la somme S converges vers la loi normale $\mathbb{N}(n\mu, n\sigma^2)$ *quand n tend vers l'infini :*

$$\lim_{n\to\infty} f(s) = \mathcal{N}(n\mu, n\sigma^2)$$

Ceci montre donc l'importance que joue la loi normale pour approximer la densité de données issues de l'accumulation de plusieurs phénomènes notamment physiques.

$\diamond$ **Définition 4.7.1.** *On appelle loi ou densité normale (ou gaussienne) univariée de d'espérance μ et de variance σ^2 (où $\sigma \geq 0$) la loi de probabilité continue définie par la densité*
$f : \mathbb{R} \to \mathbb{R}^+$*, telle que pour tout $x \in \mathbb{R}$:*

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

et elle se note $\mathcal{N}(\mu, \sigma^2)$.

⚠ **Remarque 4.7.1.** *Ce théorème peut aussi se formuler ainsi. Soit la variable aléatoire moyenne*

$$\bar{X} = \frac{S}{n} = \frac{X_1 + \cdots + X_n}{n}\,,$$

et la variable aléatoire centrée réduite

$$Z \;=\; \frac{S - n\mu}{\sigma\sqrt{n}} \;=\; \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

On a donc la densité de Z converges vers la loi normale centrée réduite quand n tend vers l'infini.

Chapitre 5

Lois de probabilités discrètes usuelles

Dans ce chapitre, nous allons voir les loi de probabilités discrètes usuelles.

5.1 Loi de Bernoulli $\mathcal{B}(p)$

5.1.1 Définition

$\diamond$ **Définition 5.1.1. Loi de Bernoulli :** *La loi de Bernoulli est une distribution de probabilité discrète pour une v.a binaire X qui prend la valeur 1 avec la probabilité p (probabilité de succès) et 0 avec la probabilité $q = 1 - p$ (probabilité d'échec). Elle est donc caractérisé par le seul paramètre p et se définie ainsi*

$$\mathbb{P}(X = x) = \begin{cases} p & si\ x = 1, \\ 1 - p & si\ x = 0, \end{cases} \tag{5.1}$$

On peut la noter d'une manière équivalente $\mathbb{P}(X = x) = p^x(1-p)^{1-x}$, $x \in \{0, 1\}$ et elle se note $\mathcal{B}(p)$.
On a

1. $\mathbb{E}[X] = p$
2. $\text{var}[X] = p(1 - p)$.
3. Fonction génératrice des moments : $M_X(t) = pe^t + (1 - p)$

Démonstration. 1. Pour l'espérance, on a

$$\begin{aligned} \mathbb{E}[X] &= \sum_k x_k \mathbb{P}(X = x_k) \\ &= 0 \times \mathbb{P}(X = 0) + 1 \times \mathbb{P}(X = 1) \\ &= 1 \times \mathbb{P}(X = 1) \\ &= 1 \times p = p \end{aligned}$$

2. Pour la variance, on a

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_k x_k^2 \mathbb{P}(X = x_k)\\ &= 0^2 \times \mathbb{P}(X=0) + 1^2 \times \mathbb{P}(X=1)\\ &= 1 \times \mathbb{P}(X=1) = p\end{aligned}$$

donc

$$\begin{aligned}\mathrm{var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2\\ &= p - p^2\\ &= p(1-p)\end{aligned}$$

3. Pour la f.g. des moments on a

$$\begin{aligned}M_X(t) &= \mathbb{E}\left(e^{tX}\right) = \sum_x e^{tx}\mathbb{P}(X=x)\\ &= e^{t\times 0}(1-p) + e^{t\times 1}p\\ &= pe^t + (1-p)\end{aligned}$$

□

5.2 Loi Binomiale $\mathcal{B}(n,p)$

5.2.1 Définition

◇ **Définition 5.2.1.** *La loi Binomiale, notée $\mathcal{B}(n,p)$ se définissant ainsi*

$$\forall x \in \mathbb{N}, P(X=x) = C_n^x\, p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \tag{5.2}$$

peut être décrite comme la somme de n v.a. **indépendantes** *$X_1, \ldots, X_n$ de Bernoulli de paramètre p . Elle est donc caractérisée par deux paramètres (n,p) et est donnée par*

$$Y = \sum_{i=1}^n X_i \sim \mathcal{B}(n,p). \tag{5.3}$$

avec $X_i \overset{ind.}{\sim} \mathcal{B}(p)$

5.2.2 Espérance, variance, fonction génératrice des moments

Si Y est une v.a. de loi Binomiale $\mathcal{B}(n,p)$, alors

1. $\mathbb{E}[Y] = np$
2. $\mathrm{var}[Y] = np(1-p)$
3. Fonction génératrice des moments : $M_X(t) = (pe^t + (1-p))^n$

Démonstration. 1. On a $\sum_{i=1}^{n} X_i \sim \mathcal{B}(n,p)$ ou $X_i \overset{ind.}{\sim} \mathrm{Bern}(p)$, donc l'espérance d'une v.a Y de distribution $\mathcal{B}(n,p)$ est donnée par

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}[\sum_{i=1}^{n} X_i] \\ &= \sum_{i=1}^{n} \mathbb{E}[X_i] \\ &= \sum_{i=1}^{n} p \\ &= np\end{aligned}$$

2. de même pour la variance, on en trouve $np(1-p)$

$$\begin{aligned}\mathrm{var}[Y] &= \mathrm{var}[\sum_{i=1}^{n} X_i] \\ &= \sum_{i=1}^{n} \mathrm{var}[X_i] \\ &= \sum_{i=1}^{n} p(1-p) \\ &= np(1-p)\end{aligned}$$

3. Pour la f.g. des moments on a $Y = f(X_i) = \sum_i f(X_i)$, donc

$$\begin{aligned}M_Y(t) &= \mathbb{E}\left(e^{tY}\right) \\ &= \mathbb{E}\left(e^{t\sum_{i=1}^{n} X_i}\right) \\ &= \mathbb{E}\left(\prod_{i=1}^{n} e^{tX_i}\right) \\ &\overset{X_i \text{ i.i.d}}{=} \prod_{i=1}^{n} \mathbb{E}\left(e^{tX_i}\right) \\ &= \prod_{i=1}^{n} M_X(t) \\ &= (pe^t + (1-p))^n\end{aligned}$$

□

5.3 Loi Multinomiale $\mathcal{M}(n, p, K)$

La loi binomiale concerne le nombre de succès dans n épreuves de Bernoulli indépendantes donnant chacune un résultat binaire (comme dans le jeu de pile ou face). La loi multinomiale est une généralisation de celle-ci, applicable par exemple à n jets d'un dé à six faces. Contrairement à ces exemples simples, les différentes possibilités ne sont généralement pas équiprobables.

La fonction de probabilité de la variable aléatoire binomiale X ($X = x \in \{0, 1\}$ s'écrit

$$\mathbb{P}(X = x) = C_n^x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

peut se réécrire de manière symétrique en faisant intervenir deux variables X_1 et X_2 dont la somme est égale à n : ($x_2 = n - x_1$) et $p_1 = p; p_2 = 1 - p$:

$$\mathbb{P}(X_1 = x_1, X_2 = x_2) = \frac{n!}{x_1!(x_2)!} \frac{n!}{n_1! n_2!} p_1^{n_1} p_2^{n_2}$$

Généralisation

Dans le cas multinomial à K résultats possibles au lieu de 2, les variables deviennent $X_i \, , i = \{1, \ldots, K\}$ et correspondent aux probabilités $p_i \, , i = \{1, \ldots, K\}$ avec les contraintes

$$\sum_{k=1}^{K} x_k = n \quad \sum_{k=1}^{K} p_k = 1$$

La fonction de probabilité s'écrit alors, sous la condition portant sur la somme des variables :

$$\mathbb{P}(X = x) = \mathbb{P}(X_1 = x_1, \ldots X_K = x_k) = \frac{n!}{\prod_{k=1}^{K} x_k!} \times \prod_{k=1}^{K} p_k^{x_x} = \frac{n!}{x_1! \ldots x_k!} p_1^{x_1} \ldots p_K^{x_K}$$

5.3.1 Espérance, variance, fonction génératrice des moments

Chacune des variables reste une variable binomiale dont la moyenne et la variance sont

$$\mathrm{E}(X_k) = np_k$$

$$\mathrm{var}(X_k) = np_k(1 - p_k)$$

tandis que les covariances s'écrivent $\mathrm{cov}(X_k, X_l) = -np_k p_l$

5.3.2 Moments et fonction génératrice des moments

5.3.3 Définition

5.3.4 Espérance, variance, fonction génératrice des moments

5.4 Loi Géométrique

5.4.1 Définition

Considérons une expérience de Bernoulli dont la probabilité de succès est $p (0 < p < 1)$ et celle d'échec $q = 1 - p$. On renouvelle cette expérience de manière *indépendante* jusqu'au premier succès. On appelle X la variable aléatoire donnant le rang du premier succès. Les réalisations x de X sont donc les entiers naturels non nuls 1, 2, 3, ...

La loi de probabilité que $X = x$ est alors, pour $x \in \mathbb{N}^\star$

$$\mathbb{P}(X = x) = p(x) = pq^{x-1}. \tag{5.4}$$

On dit que X suit une loi géométrique de paramètre p et on note $X \sim \mathcal{G}(p)$.

⚠ **Remarque 5.4.1.** *On est donc dans les mêmes hypothèses que pour la loi binomiale, mais le nombre de répétitions n'est pas fixé à l'avance. On s'arrête au premier succès*

⚠ **Remarque 5.4.2.** *L'appellation géométrique vient du fait qu'en sommant toutes les probabilités, on obtient la somme des termes d'une suite géométrique. En effet :*

$$\begin{aligned}
\sum_{x=1}^{+\infty} \mathbb{P}(X=x) &= \sum_{x=1}^{+\infty} pq^{x-1} = \lim_{n\to\infty} \sum_{x=1}^{n} pq^{x-1} = p \lim_{n\to\infty} \sum_{x=1}^{n} q^{x-1} \\
&= p \lim_{n\to\infty} \frac{1-q^n}{1-q} \\
&\overset{\lim_{n\to\infty} q^n = 0 (0<q<1)}{=} \frac{p}{1-q} = 1.
\end{aligned}$$

5.4.2 Espérance, variance, fonction génératrice des moments

Soit X suit une loi géométrique de paramètre p, alors

1. *Espérance :* $\mathbb{E}[X] = \frac{1}{p}$
2. *Variance :* $\operatorname{var}[X] = \frac{q}{p^2}$
3. *Fonction génératrice des moments :* $M_X(t) = \mathbb{E}(e^{tX}) = \frac{pe^t}{1-q\,e^t}$

Démonstration.

1. Espérance

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{+\infty} \mathbb{P}(X=x) = \sum_{x=1}^{+\infty} xpq^{x-1} \\
&= p\sum_{x=1}^{+\infty} xq^{x-1} = p\frac{\mathrm{d}\left(\sum_{x=1}^{+\infty} q^x\right)}{\mathrm{d}\,q} \\
&= p\frac{\mathrm{d}\left(\frac{1}{1-q}\right)}{\mathrm{d}\,q} = p\frac{1}{(1-q)^2} = \frac{1}{p}
\end{aligned}$$

2. variance :

$$\begin{aligned}
\operatorname{var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X^2] - \frac{1}{p^2} \\
&= \sum_{x=1}^{+\infty} x^2 pq^{x-1} - \frac{1}{p^2} = p\sum_{x=1}^{+\infty} x^2 q^{x-1} - \frac{1}{p^2} \\
&= \frac{2p}{(1-q)^3} - \frac{1}{(1-q)^2} - \frac{1}{p^2} \\
&= \frac{q}{p^2}
\end{aligned}$$

car si on prend $f(x) = \sum_{k=1}^{+\infty} x^{k-1}$ on a

$$\begin{aligned}
f(x) &= \sum_{k=1}^{+\infty} x^{k-1} = \sum_{k=0}^{+\infty} x^k = \frac{1}{1-x} \\
f'(x) &= \sum_{k=1}^{+\infty} kx^{k-1} = \sum_{k=0}^{+\infty} (k+1)x^k = \frac{1}{(1-x)^2} \\
f''(x) &= \sum_{k=1}^{+\infty} k(k+1)x^{k-1} = \sum_{k=1}^{+\infty} k^2x^{k-1} + \sum_{k=1}^{+\infty} kx^{k-1} = \frac{2}{(1-x)^3}
\end{aligned}$$

En particulier

$$\sum_{k=1}^{+\infty} k^2x^{k-1} = \frac{2}{(1-x)^3} - \frac{1}{(1-x)^2}.$$

3. Pour la f.g. des moments on a $Y = f(X_i) = \sum_i f(X_i)$, donc

$$\begin{aligned}
M_X(t) &= \mathbb{E}\left(e^{tX}\right) = \sum_x e^{tx}\mathbb{P}(X = x) \\
&= \sum_x e^{tx}pq^{x-1} = p\sum_x e^{tx}q^{x-1} \\
&= \frac{p}{q}\sum_x e^{tx}q^x = \frac{p}{q}\sum_x (qe^t)^x \\
&= \frac{p}{q} \times \lim_{n\to\infty} qe^t \frac{1-(qe^t)^n}{1-qe^t} \\
&= \frac{pe^t}{1-qe^t}
\end{aligned}$$

□

5.5 Loi Uniforme discrète $\mathcal{U}(n)$

5.5.1 Définition

Une variable aléatoire X prenant ses valeurs dans $N^\star$ suit une loi uniforme discrète de paramètre n si :

Loi de probabilités :

$$P(X = x) = \frac{1}{n} \forall x \in \{1, 2, \ldots, n\}$$

5.5.2 Espérance, variance, fonction génératrice des moments

$$\mathbb{E}[X] = \frac{n+1}{2}$$

$$\mathrm{var}[X] = \frac{n^2-1}{12}$$

Fonction génératrice des moments

$$M_X(t) = \mathbb{E}(e^{tX}) = \frac{1}{n}\sum_{x=1}^{n} e^{tx}.$$

5.5.3 Cas particulier $\mathcal{U}_{[a,b]}$

Loi uniforme sur $[a, b]$
Loi de probabilités :

$$P(X = x) = \frac{1}{n} \forall x \in \{1, 2, \ldots, n\}$$

avec $n = \#[a, b]$

$$\mathbb{E}[X] = \frac{a+b}{2}$$

$$\mathrm{var}[X] = \frac{(b-a+1)^2 - 1}{12}$$

Fonction génératrice des moments

$$M_X(t) = \mathbb{E}(e^{tX}) = \frac{e^{at}}{n}\sum_{x=0}^{n-1} e^{tx}.$$

5.6 Loi de Poisson $\mathcal{P}(\lambda)$

5.6.1 Définition

$\diamond$ **Définition 5.6.1. Loi de poisson :** *La loi de Poisson est une loi de probabilité discrète qui décrit le comportement du nombre d'évènements se produisant dans un laps de temps fixé, si ces évènements se produisent avec une fréquence moyenne connue et indépendamment du temps écoulé depuis l'évènement précédent. Si le nombre moyen d'occurrences dans cet intervalle est* λ *(le paramètre de la loi), alors la probabilité qu'il existe exactement* x *occurrences (*$x \in \mathbb{N}$*) est*

$$\forall x \in \mathbb{N}, \quad p_X(x; \lambda) = \mathbb{P}(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}.$$

Elle se note généralement $\mathcal{P}(\lambda)$*.*

5.6.2 Espérance, variance, fonction génératrice des moments

Soit, X *une v.a. de loi de poisson de paramètre* λ*, alors*

1. $\mathbb{E}[X] = \lambda$
2. $\mathrm{var}[X] = \lambda$
3. *La fonction génératrice des moments d'une loi de Poisson est*

$$M_X(t) \equiv \mathbb{E}(e^{tX}) = \exp\left(\lambda(e^t - 1)\right).$$

Démonstration. 1. **Espérance**

Si X suit une loi de poisson de paramètre λ, soit $X \sim \mathcal{P}(\lambda)$.

Alors, on a par définition que $\mathbb{P}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$ et que :

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=1}^{+\infty} k\,\mathbb{P}(X=k) \\
&= \sum_{k=1}^{+\infty} k\,e^{-\lambda}\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} \\
&= \lambda\,e^{-\lambda}\sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda\,e^{-\lambda}\,e^{\lambda} = \lambda \qquad \text{car } e^{\lambda} = \sum_{k=0}^{+\infty}\frac{\lambda^k}{k!}
\end{aligned}$$

Dans la dernière ligne, on reconnaît le développement en série entière de e^x.

2. **Variance**

$$\begin{aligned}
V(X) &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \\
&= \sum_{k=1}^{+\infty} k^2\,\mathbb{P}(X=k) - \lambda^2 \\
&= \sum_{k=1}^{+\infty} k^2\,e^{-\lambda}\frac{\lambda^k}{k!} - \lambda^2 \\
&= \lambda\,e^{-\lambda}\sum_{k=1}^{+\infty} \frac{k\lambda^{k-1}}{(k-1)!} - \lambda^2 \\
&= \lambda\,e^{-\lambda}\sum_{k=1}^{+\infty} \frac{d}{d\lambda}\frac{\lambda^k}{(k-1)!} - \lambda^2 \\
&= \lambda\,e^{-\lambda}\frac{d}{d\lambda}\sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} - \lambda^2 \\
&= \lambda\,e^{-\lambda}\frac{d}{d\lambda}[\lambda\sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!}] - \lambda^2 \\
&= \lambda\,e^{-\lambda}\frac{d}{d\lambda}[\lambda\,e^{\lambda}] - \lambda^2 \qquad \text{car } \exp(x) = \sum_{n=0}^{+\infty}\frac{x^n}{n!} \\
&= \lambda\,e^{-\lambda}(\lambda+1)\,e^{\lambda} - \lambda^2 \\
&= \lambda\,(\lambda+1) - \lambda^2 \\
&= \lambda
\end{aligned}$$

3. **Fonction génératrice**
   La fonction génératrice d'une variable aléatoire X est définie par $G_X(t) = \mathbb{E}(t^X)$. Ainsi on obtient :

$$\begin{aligned}
\mathbb{E}(t^X) &= \sum_{k=0}^{\infty} t^k \mathbb{P}(X=k) \\
&= \sum_{k=0}^{\infty} t^k e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} t^k \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(t\lambda)^k}{k!} \\
&= e^{-\lambda} e^{t\lambda} && \text{car } \exp(x) = \sum_{n=0}^{+\infty} \frac{x^n}{n!} \\
&= e^{\lambda(t-1)}
\end{aligned}$$

4. **Fonction génératrice des moments**

$$\begin{aligned}
M_X(t) &= \sum_{k=0}^{+\infty} e^{tk} \mathbb{P}(X=k) \\
&= \sum_{k=0}^{+\infty} e^{tk} \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \\
&= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda \mathrm{e}^t)^k}{k!} \\
&= e^{-\lambda} e^{\lambda \mathrm{e}^t} && \text{car } \exp(x) = \sum_{n=0}^{+\infty} \frac{x^n}{n!} \\
&= e^{\lambda(e^t-1)}
\end{aligned}$$

□