Rappels de Probabilités et Statistiques

Renaud Leplaideur

Université de Nouvelle Calédonie
Master MEEF

Table des matières

Chapitre 1

Probabilités, Espaces probabilisés.

1.1 Probabilités.

1.1.1 Événements, tribus, espace probabilisable.

La première étape lorsqu'on veut étudier mathématiquement le hasard c'est la modélisation. Il s'agit de préciser l'univers des possibles, la tribu (ensemble des événements) et la probabilité mise sur cette tribu.

Définition 1.1.1. *On appelle épreuve une expérience aléatoire dont le résultat est imprévisible.*
On appelle éventualité la résultat d'une épreuve.
On appelle univers associé à une épreuve l'ensemble des résultats possibles de cette épreuve.

Exemples
On considère l'épreuve "jeter une pièce". Une éventualité sera pile ou face. L'univers des possibles sera l'ensemble {pile,face}.
On lance deux dés discernables, Rouge et Noir (R et N).

Si on observe les chiffres marqués sur les deux dés, une épreuve sera le lancer des deux dés, une éventua lité sera les deux chiffres observés, et l'univers des possibles sera $\Omega = \{(i,j), 1 \leq i,j \leq 6\}$.

Si on veut observer quel dé fait apparaitre le chiffre le plus grand, une épreuve sera le lancer des deux dés, une éventualité sera R ou N, et l'univers des possibles sera $\Omega = \{N, R, \{N, R\}$ (si égalité)$\}$.

Si on observe la distance qui sépare les deux dés, une épreuve sera le lancer des deux dés, une éventualité sera une distance et l'univers des possibles sera $\Omega = \mathbb{R}_+$.

Si Ω est un ensemble et si $A \subset \Omega$ est une partie de Ω on notera $\overline{A}$ son complémentaire ($\overline{A} = \Omega \setminus A$).

Définition 1.1.2. *Soit Ω un ensemble non vide. Une partie $\mathcal{B}$ est une tribu de Ω si et seulement si*

1. $\Omega \in \mathcal{B}$
2. *Si A est dans $\mathcal{B}$ alors $\overline{A}$ est aussi dans $\mathcal{B}$*
3. *Si A et B sont dans $\mathcal{B}$, alors $A \bigcap B$ est encore dans $\mathcal{B}$*
4. *Si (A_n) est une famille dénombrable d'éléments de $\mathcal{B}$ alors $\bigcap_n A_n$ est encore dans $\mathcal{B}$*

On remarque qui si $\mathcal{B}$ est une tribu de Ω, alors $\emptyset$ est dans $\mathcal{B}$. On montre aussi en passant au complémentaire qu'une tribu est stable par réunion finie ou dénombrable.

Exemple

1. $\{\Omega, \emptyset\}$ est une tribu de Ω. c'est même la plus petite tribu possible.
2. L'ensemble $\mathcal{P}(\Omega)$ de toutes les parties de Ω est aussi une tribu de Ω. Cette tribu est généralement trop grosse.
3. Si A est un sous-ensemble de Ω qui n'est ni Ω, ni $\emptyset$, alors $\mathcal{B} = \{\emptyset, A, \overline{A}, \Omega\}$ est une tribu de Ω.

Définition 1.1.3. *Si Ω est un ensemble non vide et si $\mathcal{B}$ est une tribu de Ω, les éléments de la tribu $\mathcal{B}$ s'appellent des événements. Le couple $(\Omega, \mathcal{B})$ s'appelle un espace probabilisable.*

Parmi les événements certains sont plus remarquables que d'autres :

1. L'événement élémentaire $\{\omega\}$, où $\omega \in \Omega$. Attention un singleton n'est pas toujours un événement. Cela dépend de la tribu.
2. l'événement certain Ω.
3. l'événement impossible $\emptyset$.

Si A est un événement, on dit qu'il est réalisé, si à la fin de l'épreuve, le résultat est dans l'événement considérer. Par exemple, si on lance un dé à six faces, l'événement obtenir un chiffre pair ($= \{2, 4, 6\}$) sera réalisé si le résultat du lancer est soit 2, soit 4 soit 6. Si A est un événement, $\overline{A}$ s'appelle l'événement contraire de A ; $\overline{A}$ est en effet réalisé si et seulement si A n'est pas réalisé. Si B est un autre événement, A et B sont dits incompatibles si et seulement si $A \cap B = \emptyset$.

Rappels sur les ensembles

On rappelle ici quelques généralités sur la théorie des ensembles.

1. $\overline{\overline{A}} = A$, $\overline{\emptyset} = \Omega$, $\overline{A} \cup A = \Omega$.
2. (Loi de Morgan) $\overline{A \cup B} = \overline{A} \cap \overline{B}$ et $\overline{A \cap B} = \overline{A} \cup \overline{B}$. On obtient alors $\overline{\Omega} = \emptyset$ et $A \cap \overline{A} = \emptyset$.

3. $A \cup A = A \cap A = A$, $A \cap B = B \cap A$ et $A \cup B = B \cup A$.
4. associativités : $A \cup (B \cup C) = (A \cup B) \cup C$, et idem avec $\cap$.
5. $\emptyset \cup A = A$, $\emptyset \cap A = \emptyset$. On trouve alors $\Omega \cup A = \Omega$ et $\Omega \cap A = A$.
6. Double distributivité

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

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

1.1.2 Espace probabilisé.

Nous pouvons maintenant donner la définition axiomatique d'une probabilité :

Définition 1.1.4. *Une probabilité sur un espace Ω et une application ensembliste P définie sur une tribu $\mathcal{B}$, à valeur dans $[0,1]$ et telle que*

1. $P(\Omega) = 1$,
2. $P(A \bigsqcup B) = P(A) + P(B)$,
3. *P vérifie la propriété de limite séquentielle : Si $(A_n)_n$ est une suite d'éléments de la tribu et si $\bigcap_n A_n = \emptyset$, alors*

$$\lim_{N\to+\infty} P(\bigcap_{n\leq N} A_n) = 0.$$

Le triplet $(\Omega, \mathcal{B}, P)$ s'appelle un espace probabilisé.

Exemple
On lance un dé non pipé jusqu'à l'apparition du 6. Pour $n \geq 1$ on note
$\omega_n =$"6 apparait pour la première fois au n-ième lancer.",
et $\omega_0 =$"6 n'apparaît jamais". L'espace Ω est l'ensemble des ω_n (avec n dans $\mathbb{N}$) et la tribu qu'on met dessus est l'ensemble des parties de Ω, $\mathcal{P}(\Omega)$.
Posons alors $p_n = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6}$ et $p_0 = 0$. Alors $\sum_{n=0}^{+\infty} p_n = 1$ donc (p_n) permet de définir une probabilité sur $\Omega \stackrel{def}{=} \{\omega_n\}$.
On définit ensuite l'événement :
$A =$"le rang d'apparition du premier 6 est pair".
Alors $P(A) = \displaystyle\sum_{n=0, n\ pair} p_n = \frac{5}{11} < \frac{1}{2}$.

Théorème 1.1.1. *Soient Ω un univers et P une probabilité sur $\mathcal{B}$. On considère aussi $A_1, \ldots, A_n$ des événements . Alors :*
(1) $P(\overline{A}) = 1 - P(A)$. en particulier $P(\emptyset) = 0$.
(2) Si A et B sont deux événements tels que $A \subseteq B$ alors $P(A) \leq P(B)$.
(3) Si les A_i sont 2 à 2 disjoints alors $P(A_1 \cup \ldots \cup A_n) = P(A_1) + \ldots + P(A_n)$.
(4) Si A et B sont deux événements , on a $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
(5) $P(A_1 \cup \ldots \cup A_n) \leq P(A_1) + \ldots + P(A_n)$.

Démonstration. (1). On remarque que $\Omega = A \bigsqcup \overline{A}$. Ceci est valable aussi pour $A = \emptyset$.
(2) On aura donc $B = B \setminus A \bigsqcup A$.
(3) On le fait par récurrence.
(4) On a $A \cup B = A \setminus B \bigsqcup B \setminus A \bigsqcup A \cap B$. De plus $B = B \setminus A \bigsqcup A \cap B$ et $A = A \setminus B \bigsqcup A \cap B$.
(5) Le terme de droite compte plusieurs fois les élements qui sont dans 2 A_i au moins. □

On peut remplacer dans ce théorèmes les familles finies par des familles infinies dénombrables.

1.1.3 Nécessité de la propriété de limite séquentielle.

On joue à “pile ou face” et on regarde l'apparition du premier pile. On note alors
$w_0 =$“on n'obtient que des faces”,
$w_n =$“on obtient le premier pile au n-ième lancer”.
Posons
$$B^N = \bigcup_{n \geq N} w_n.$$
B^N est l'événement “on tire le premier pile après au moins N lancers”. On écrit ensuite
$$B^\infty \stackrel{def}{=} \bigcap_{N=1}^{+\infty} B^N.$$
Si B^∞ est un événement, c'est l'événement , “on tire pile à l'infinième lancer”... ce qui n'est pas très clair, mais qui peut se comprendre par “on ne tire jamais pile”. Supposons en outre qu'on ait choisi $p(\{w_0\}) \stackrel{def}{=} p_0 > 0$ et $p_n = \frac{1-p_0}{2^n}$. On aura bien $\sum_n p_n = 1$. Calculons maintenant la probabilité de B^∞.
D'un côté, on a $B^\infty = \{w_0\}$, donc $P(B^\infty) = p_0$. D'un autre côté, on a $P(B^\infty) \leq P(B^N) = \frac{1-p_0}{2^{N-1}}$ et donc on aura $P(B^\infty) = 0$. ce qui montre que B^∞ ne peut pas être $\{w_0\}$.
Ceci explique que la condition de stabilité par intersection dénombrable dans la définition d'une tribu et la propriété de limite séquentielle requise dans la définition d'une probabilité.

1.2 Conditionnement et indépendance

1.2.1 Probabilités conditionnelles

Ici on se restreint au cas d'un ensemble Ω fini.

Définition 1.2.1. *Soient Ω un univers et P une probabilité sur $\mathcal{P}(\Omega)$ de distributon (p_n). Si B est un événement tel que $P(B) \neq 0$, on appelle probabilité de A sachant B le réel*

$$P(A|B) = P_B(A) \stackrel{def}{=} \frac{P(A \cap B)}{P(B)},$$

où A est un événement.

Comme $P(B) = \sum_{n,\omega_n \in B} p_n$, il existe des n tels que $\omega_n \in B$ et $p_n \neq 0$. On pose alors

$$q_n = \begin{cases} 0 \text{ si } \omega_n \notin B \\ \dfrac{p_n}{P(B)} \text{ sinon.} \end{cases}$$

On constate alors que P_B n'est rien d'autre que la probabilité sur $\mathcal{P}(\Omega)$ associée àla distribution (q_n).
Le théorème 1.1.1 est donc valable avec P_B. En particulier $P(\overline{A}|B) = 1 - P(A|B)$.
On se souviendra aussi que la formule

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

est souvent très utile.

Exemple

Une famille a 2 enfants. L'univers des possibles est donc $\Omega = \{FF, FG, GF, GG\}$ en tenant compte de l'ordre de naissance.

1. Probabilité d'avoir 2 garçons sachant que l'ainé est un garçon ?
   On pose $A =$"les 2 enfants sont des G" et $B =$"l'ainé est un G". Clairement $P(B) = \frac{1}{2}$ et donc
   $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}.$$
2. Probabilité d'avoir 2 garçons sachant que l'un est un garçon ?
   On pose $C =$"il y a au moins un G" ; $P(C) = \frac{3}{4}$ et donc
   $$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

1.2.2 Formule des probabilités totales

Définition 1.2.2. *Soient $A_1, \ldots, A_n$ des événements. On dit que $\{A_1, \ldots, A_n\}$ constitue un système complet d'événements si*
pour tout $1 \leq i \leq n$, $A_i \neq \emptyset$;
pour tout $1 \leq i, j \leq n$, $A_i \cap A_j = \emptyset$ dès que $i \neq j$;
$\Omega = \bigcup_{i=1}^{n} A_i \stackrel{def}{=} \bigsqcup_{i=1}^{n} A_i$.

Par exemple si A est un événement non vide, alors $\{A, \overline{A}\}$ est un système complet d'événements.

Théorème 1.2.1 (Formule des probabilités totales)**.** *Si* $\{A_1, \ldots, A_n\}$ *est un système complet d'événements, et si pour tout* i*,* $P(A_i) \neq 0$*, alors pour tout* $B \subset \Omega$

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

Démonstration. On a $\Omega = \bigsqcup_{i=1}^{n} A_i$ et donc pour tout $B \subset \Omega$, $B = \bigsqcup_{i=1}^{n} (B \cap A_i)$. D'après la théorème 1.1.1 on doit avoir

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i).$$

Or chaque terme du membre de droite vaut exactement $P(B|A_i)P(A_i)$. □

Exemple et application

Si A est tel que $0 < P(A) < 1$, alors $P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A})$.

On prend un individu I au hasard dans une population qui contient une proportion $p \in]0, 1[$ de tricheurs. On fait tirer à I une carte au hasard dans un jeu de 52 cartes. On suppose que si I est un tricheur il est sûr de piocher un as. quelle est la probabilité que I tire un as ?

On pose $A =$"I est un tricheur" et $B =$"I trouve un as". Alors

$$\begin{aligned} P(B) &= P(B|A)P(A) + P(B|\overline{A})P(\overline{A}) \\ &= p \times 1 + (1-p) \times \frac{4}{52} = \frac{1+12p}{13}. \end{aligned}$$

1.2.3 Formule de Bayes

Le but de la formule de Bayes est d'exprimer les valeurs de $P(A_i|B)$ en fonction des $P(B|A_i)$. On verra un exemple courant d'application.

Théorème 1.2.2. *Soient* $\{A_1, \ldots, A_n\}$ *un système complet d'événements, tel que pour tout* i*,* $P(A_i) \neq 0$ *et* B *un événement de probabilité non nulle. Alors pour tout* i

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \ldots + P(B|A_n)P(A_n)}.$$

Démonstration. On sait que $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$, et pour tout i, $P(B|A_i)P(A_i) = P(B \cap A_i) = P(A_i|B)P(B)$. Ceci donne donc

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)},$$

et il suffit de remplacer $P(B)$ par sa valeur en fonction des $P(B|A_i)$. □

Exemple d'application

Au cours d'un sondage on "trouve" que 32% des gens se réclament de la majorité (événement M), 42% de l'opposition (événement O), et 26% sont indécises (événement I). De plus 85% des personnes qui se réclament de la majorité sont satisfaites de l'action du gouvernement, contre 25% des personnes de l'opposition et 60% des indécises. Un individu pris au hasard approuve l'action du gouvernement, quelle est la probabilité qu'il se réclame de la majorité, de l'opposition ou qu'il soit indécis ?
On note S l'événement "être satisfait" et on remarque que $\{M, O, I\}$ est un sytème complet. On a alors

$$P(S) = P(S|M)P(M) + P(S|O)P(O) + P(S|I)P(I) = 0,533.$$

La formule de Bayes donne

$$P(M|S) = \frac{P(S|M)P(M)}{P(S)} = 0,510$$

$P(O|S) = 0,197$ et $P(I|S) = 0,293$.

1.2.4 Événements indépendants

Définition 1.2.3. *On dit que deux événements A et B sont indépendants pour le probabilité P si et seulement si $P(A \cap B) = P(A)P(B)$.*

Supposons que $P(B) \neq 0$. alors A et B sont indépendants si et seulement si $P(A|B) = P(A)$. Cela correspond bien àla notion intuitive d'événements indépendants : l'événement A se produit de manière indépendante de l'événement B ce qui signifie que si l'on fait une multitude d'expériences, la fréquence de réalisation de A sera la même (en moyenne) que la la fréquence de réalisation de A lorsque B se réalise.
On remarquera aussi que sauf si $P(A) = 0$ ou 1, A et $\overline{A}$ ne sont pas indépendants.

Exemple

On considère les événements
$A =$"la famille Toto a des enfants de sexes différents".
$B =$"la famille Toto a au plus une fille".

1. Supposons que la famille Toto ait 2 enfants. L'univers des possibles est $\Omega = \{FF, FG, GF, GG\}$. Alors $P(A) = \frac{1}{2}$, $P(B) = \frac{3}{4}$ et $P(A \cap B) = \frac{1}{2}$. Donc A et B ne sont pas indépendants.
2. Si la famille Toto a 3 enfants, on a $\Omega = \{FFF, FFG, FGF, FGG, GFF, GFG, GGF, GGG\}$, et donc $P(A) = \frac{6}{8} = \frac{3}{4}$, $P(B) = \frac{4}{8} = \frac{1}{2}$ et $P(A \cap B) = \frac{3}{8}$. Ainsi A et B sont indépendants.

Proposition 1.2.4. *Si A et B sont indépendants, alors A et $\overline{B}$ le sont aussi.*

Démonstration. On a $P(A) = P(A \cap B) + P(A \cap \overline{B})$ et comme $P(A \cap B) = P(A)P(B)$, cela donne
$$P(A \cap \overline{B}) = P(A)(1 - P(B)) = P(A)P(\overline{B}).$$
□

comme parfois nous allons rencontrer des systèmes complets avec plus de 2 événements, il faut étendre la définition d'événements indépendants.

Définition 1.2.5. *Soient $A_1, \ldots, A_n$ des événements. Ils sont dits indépendants dans leur ensemble si et seulement si pour tout $1 \leq i_1 < i_2 < \ldots < i_r \leq n$ on a*
$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_r}) = \prod_{j=1}^{r} P(A_{i_j}).$$

Attention il faut rapprocher cela de la notion d'indépendance de n-vecteurs dans un Espace vectoriel : tout comme 3 vecteurs peuvent être 2 à 2 indépendants mais pas indépendants dans leur ensemble, 3 événements peuvent être 2 à 2 indépendants mais pas indépendants dans leur ensemble.

Exemple

On considères 2 dés discernables, $A =$"le chiffre du premier dé est pair", $B =$"le chiffre du deuxième dé est pair" et $C =$"la somme des deux chiffres est paire".
On a $P(A) = P(B) = P(C) = \frac{1}{2}$, et $P(A \cap B) = P(B \cap C) = P(C \cap A) = \frac{9}{36}$ (en fait il faut observer que $A \cap B = A \cap C = B \cap C$), mais, en revanche
$$P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C).$$

1.2.5 Un complément pour les univers infinis

Lorsque l'univers est infini, on peut rencontrer des suites infinies d'événements 2 à 2 disjoints, et donc des systèmes complets de cardinal infini. Il se pose alors des problèmes de limites. Le but de cette partie est de montrer qu'on peut raisonner comme dans le cas fini. On suppose donc dans cette partie seulement que Ω est infini (dénombrable). On posera $\Omega = \{\omega_n\}_{n \in \mathbb{N}}$

Proposition 1.2.6 (Union disjointe)**.** *Soient P une probabilité sur $\mathcal{P}(\Omega)$ de distribution (p_n) et (A_k) une suite d'événements deux àdeux disjoints. Alors*

$$P(\bigcup_{k\in\mathbb{N}} A_k) = \sum_{k=0}^{+\infty} P(A_k).$$

Attention, sauf si l'ensemble des événements vaut $\mathcal{P}(\Omega)$, il n'est pas clair qu'une union infinie d'événements soit encore un événement. Pour cela il faut demander que l'ensemble des événements soit une *tribu*, c'est àdire que c'est une algèbre stable par réunion infinie, et donc aussi par intersection infinie.

Démonstration. Notons $A \stackrel{def}{=} \bigcup_{k\in\mathbb{N}} A_k$. Si K est fixé, on a $\displaystyle\bigcup_{k=0}^{K} A_k \subset A$, et donc pour tout K, $P(\bigcup_{k=0}^{K} A_k) \leq P(A)$. De plus la suite $(P(\bigcup_{k=0}^{K} A_k))_{K\in\mathbb{N}}$ est une suite croissante et majorée (par 1), ce qui montre qu'elle converge vers une limite qu'on note $\displaystyle\sum_{k=0}^{+\infty} P(A_k)$. Ainsi on trouve

$$\sum_{k=0}^{+\infty} P(A_k) \leq P(A).$$

Réciproquement, fixons N, et notons $k(N)$ le plus grand entier k tel qu'il existe $n \leq N$ avec $\omega_n \in A_k$. Un tel entier $k(N)$ doit forcément exister car $\{0, \ldots, N\}$ est fini, et il y a une infinité de A_j. Si B est un événement quelconque,on rappelle qu'on avait noté

$$P_N(B) = \sum_{n\leq N, \omega_n \in B} p_n,$$

et on avait vu que $(P_N(B))_{N\in\mathbb{N}}$ est une suite croissante qui converge vers $P(B)$. On a alors

$$P_N(A) = \sum_{k\leq k(N)} P_N(A_k) \leq \sum_{k\leq k(N)} P(A_k) \leq 1.$$

En passant à la limite on trouve l'autre inégalité, et donc la proposition est prouvée. $\square$

Proposition 1.2.7. *Soit $\{A_k\}_{k\in\mathbb{N}}$ un système complet d'événements de Ω, tous de probabilité non nulle. Alors pour tout événement A :*

$$P(A) = \sum_{k=0}^{+\infty} P(A|A_k)P(A_k).$$

Démonstration. On utilise la proposition précédente pour écrire :

$$P(A) = P(\bigsqcup_{k\in\mathbb{N}}(A \cap A_k)) = \sum_{k=0}^{+\infty} P(A \cap A_k),$$

et par définition chaque terme de la somme de droite vaut $P(A|A_k)P(A_k)$. $\square$

Exemple

On joue à“pile ou face” et on regarde l'apparition du premier pile. On note alors
$w_0 =$“on n'obtient que des faces”,
$w_n =$“on obtient le premier pile au n-ième lancer”.
l'intuition nous pousse àposer $p_n \stackrel{def}{=} p(\{w_n\}) = \frac{1}{2^n}$, et pour avoir $\sum_n p - n = 1$, on est amené à poser $p_0 = 0$. On a bien $\displaystyle\sum_n p_n = 1$. Supposons maintenant qu'on se donne en plus une urne qui contient une boule blanche. Si on tire face on rajoute une boule noire dans l'urne, et si on tire pile on tire au hasard ue boule dans l'urne. On s'arrête lorsqu'on tire la boule blanche.
La probabilité de tirer la boule blanche est donc (événement T) $P(T) = \sum_{n=1}^{+\infty} \frac{1}{n}\frac{1}{2^n} = \ln 2$.

1.3 Rappels de dénombrements

1.3.1 Listes

Définition 1.3.1. *Soit E un ensemble àn éléments. On appelle liste de p éléments de E tout p-uplet de E, ie tout élément de $E^p = E \times E \times \ldots \times E$.*

La définition de liste n'interdit en aucune manière qu'un élément figure plusieurs fois dans la liste.

Proposition 1.3.2. *Soit E un ensemble àn éléments. Il y a n^p listes de p éléments.*

Démonstration. On a n choix pour le premier élément, n choix pour le deuxième, ... et n choix pour le dernier. $\square$

Exemples

1. On jete p dés discernables (couleurs différentes), et on note le résultat de chaque dé. Au final on obtient un liste de p éléments àvaleur dans $\mathbb{N}_6 = \{1, 2, 3, 4, 5, 6\}$. Il y a 6^p possibiliés.
2. On jette p fois de suite une pièce de monnaie. Le résultat est une de p éléments à valeur dans $\{$Pile,Face$\}$. Il y a 2^p possibilités.

1.3.2 Arrangements

Définition 1.3.3. *Soit E un ensemble àn éléments. On appelle arrangement de p éléments pris dans E une liste de p éléments tous 2 à2 distincts.*
On notera A_n^p le nombre d'arrangements de p éléments pris dans un ensemble àn éléments.

On remarque qu'il est nécessaire d'avoir $p \leq n$ pour avoir un arrangement de p éléments pris dans E.

Exemple

Un tiercé dans l'ordre de n chevaux est un arrangement de 3 de ces n chevaux.

Proposition 1.3.4. *Si $1 \leq p \leq n$, alors $A_n^p = \dfrac{n!}{(n-p)!}$.*
Si $p > n$, alors $A_n^p = 0$.

Démonstration. Il est clair que si $p > n$, alors $A_n^p = 0$. supposons donc que $1 \leq p \leq n$.

$$\left.\begin{array}{l} \text{Il y a } n \text{ choix pour le premier élément} \\ \text{Il y a } n-1 \text{ choix pour le deuxième élément} \\ \vdots \\ \text{Il y a } n-p+1 \text{ choix pour le } p\text{-ième élément} \end{array}\right\} \begin{array}{c} \text{en tout il y a} \\ n(n-1)\dots(n-p+1) \\ \text{possibilités.} \end{array}$$

□

Par convention on posera $A_n^0 = 1$.

Exemple

Dans une course de n chevaux, il y a $A_n^3 = n(n-1)(n-2)$ tiercés possibles.

1.3.3 Permutations

Définition 1.3.5. *Une permutation d'un ensemble E àn éléments est un arrangement de n éléments. C'est donc une liste de tous les éléments de E, chaque élément apparaissant exactement une fois dans la liste.*

Exemple

Dans une course de n chevaux une permutation est un classement entier après la course.

Proposition 1.3.6. *Il y a exactement $n!$ permutations dans un ensemble E àn éléments.*

Démonstration. Il suffit de calculer A_n^n. □

1.3.4 Combinaisons

Définition 1.3.7. *Une combinaison de p éléments d'un ensemble E àn éléments est une partie de E àp éléments.*
On notera $\binom{n}{p}$ le nombre de combinaisons p éléments d'un ensemble à$n$éléments.

On remarque que, comme pour les arrangements, il est nécessaire d'avoir $p \leq n$ pour pouvoir définir une combinaison.

Exemples

1. 5 cartes prises parmi un jeu de 32 cartes.
2. un tiercé dans l'ordre ou le désordre est une combinaison de 3 éléments parmi n.

Proposition 1.3.8. *Si $p \leq n$, on a $\binom{n}{p} = \frac{A_n^p}{p!} = \frac{n(n-1)\dots(n-p+1)}{p!} = \frac{n!}{p!(n-p)!}$, avec la convention $0! = 1$.*

Démonstration. Un arrangement de p éléments parmi n consiste à d'abord choisir p éléments parmi n, puis à ranger ces p élèments dans un certain ordre.
Il y a $\binom{n}{p}$ manière de choisir les p éléments parmi n, puis ces éléments étant fixés, il y a $p!$ manières de les classer (permutation). On trouve donc

$$A_n^p = \binom{n}{p} p!.$$

Les autres égalités s'obtiennent par le calcul. □

On remarquera que $\binom{n}{0} = \binom{n}{n} = 1$, $\binom{n}{1} = \binom{n}{n-1} = n$, $\binom{n}{2} = \binom{n}{n-2} = \frac{n(n-1)}{2}$.

Propriétés des coefficients binômiaux

Nous allons donner une série de propriété des $\binom{n}{p}$.

Proposition 1.3.9. *Pour tout p dans $[0, n]$, $\binom{n}{p} = \binom{n}{n-p}$.*
Pour tout p dans $[0, n-1]$, $\binom{n}{p} = \binom{n-1}{p} + \binom{n-1}{p-1}$.

Démonstration. La première égalité résulte de la proposition 1.3.8. Pour la seconde, donnons nous un ensemble E àn éléments, et x un des éléments de E.
Une partie de p éléments de E contient x ou ne contient pas x. De plus il y a $\binom{n-1}{p}$ parties de E àp éléments qui ne contiennent pas x et $\binom{n-1}{p-1}$ parties de E àp éléments qui contiennent x. Ainsi

$$\binom{n}{p} = \binom{n-1}{p} + \binom{n-1}{p-1}.$$

□

Proposition 1.3.10 (Formule de Newton)**.** *Pour tout* (a, b) *dans* $\mathbb{C}$, $(a+b)^n = \sum_{p=0}^{n} \binom{n}{p} a^p b^{n-p}$.

Démonstration. On commence par remarque que $(1+X)^n$ est un polynôme de degré n, et donc il s'écrit $\sum_{p=0}^{n} \alpha_p X^p$.
Si on met b^n en facteur dans $(a+b)^n$, on voit donc que $(a+b)^n$ s'écrit sous la forme $\sum_{p=0}^{n} \alpha_p a^p b^{n-p}$. Reste àcalculer le coefficient α_p. Or α_p correspont aunombre de manières d'obtenir X^p lorsqu'on développe $(1+X)^n = (1+X)\times(1+X)\times\ldots\times(1+X)$. Cela revient donc àchoisir p éléments parmi n (ceux qui vont donner un X), et donc il y a $\binom{n}{p}$ possibilité, chaque possibilité ayant une contribution de 1 (les autres ne sont que des 1), et donc $\alpha_p = \binom{n}{p}$.
On peut aussi obtenir une preuve par récurrence, mais qui est moins élégante. □

Corollaire 1.3.11. *Si* $\mathrm{Card}\, E = n$ *alors le nombre de parties de* E*, ie* $\mathrm{Card}\,\mathcal{P}(E)$*, est* 2^n*.*

Démonstration. Il y a exactement $\binom{n}{p}$ parties de E àp éléments. Ainsi il y a $\sum_{p=0}^{n} \binom{n}{p}$ parties de E, soit 2^n par la proposition 1.3.10 □

De la proposition 1.3.10 découle aussi le corollaire suivant :

Corollaire 1.3.12. *On a la relation* $\sum_{p=0}^{n} \binom{n}{p} (-1)^p = 0$.

On peut aussi montrer par récurrence en utilisant la proposition 1.3.9 la **formule de Leibnitz** :

Proposition 1.3.13. *Soient* f *et* g *deux fonctions de* $\mathbb{R}$ *dans* $\mathbb{R}$ *qui sont* n *fois dérivables. Alors*

$$(fg)^{(n)} = \sum_{p=0}^{n} \binom{n}{p} f^{(p)} g^{(n-p)}.$$

1.4 Exemples d'application : La ruine du joueur

Deux joueurs A et B jouent une partie de pile ou face ; la partie consiste en une succession de lancers indépendants les uns des autres. On suppose que $P(\text{“A gagne”}) = p$ et $P(\text{“B gagne”}) = q = 1-p$. À chaque coup, le perdant donne 1 franc au gagnant. Au départ, A et B disposent des sommes respectives a et b, et ils ne peuvent pas abandonner la partie avant que l'un des deux joueurs soit ruiné. On cherche la probabilité $R(a, b)$ pour que A soit ruiné. On notera $\mathcal{A}(a, b)$ l'événement “A est ruiné” et $S(a, b)$ la probabilité que B soit ruiné.

Étude du premier coup

On note $C_1 =$"A gagne le premier coup". Alors $\{C_1, \overline{C_1}\}$ est un système complet d'événements, et donc

$$R(a,b) = P(\mathcal{A}(a,b)|C_1)P(C_1) + P(\mathcal{A}(a,b)|\overline{C_1})P(\overline{C_1}). \tag{1.1}$$

Or $P(\mathcal{A}(a,b)|C_1) = R(a+1,b-1)$ et $P(\mathcal{A}(a,b)|\overline{C_1}) = R(a-1,b+1)$. De plus le capital total $s = a+b$ est constant, ce qui signifie que $b = s-a$ et donc, en posant $u(a) = R(a,b) = R(a,s-a)$ (1.1) devient

$$R(a,b) = p \times u(a+1) + q \times u(a-1). \tag{1.2}$$

Le polynôme caractéristique de (1.2) est $pX^2 - X + q$ qui admet pour racines $\frac{q}{p}$ et 1.

Premier cas $p \neq q$

Alors $u(a)$ s'écrit $\lambda + \mu \left(\frac{q}{p}\right)$. Comme on doit avoir $u(0) = 1$ (A est ruiné) et $u(a+b) = 0$ (B est ruiné),on trouve

$$R(a,b) = \frac{q^a(p^b - q^b)}{p^{a+b} - q^{a+b}} \text{ et } S(a,b) = \frac{p^b(p^a - q^a)}{p^{a+b} - q^{a+b}}.$$

On remarque que $R(a,b) + S(a,b) = 1$ et donc, avec une probabilité 1, la partie s'achève.

Deuxième cas $p = q$

Alors 1 est racine double ce qui signifie que $u(a)$ est du type $\lambda + a\mu$. on a toujours comme conditions aux limites $u(0) = 1$ et $u(a+b) = 0$, ce qui donne

$$R(a,b) = \frac{b}{a+b} \text{ et } S(a,b) = \frac{a}{a+b}.$$

Ainsi si B possède 90F et A possède 10F, A a 90% de chances d'être ruiné. On a toujours $R+S = 1$, donc la partie s'achève aussi dans ce cas avec une probabilité 1.

Conclusions

On suppose que B est infiniment riche. Si $p \leq q$ alors A sera ruiné avec une probabilité 1. Si $p > q$ A a une probabilité $\left(\frac{q}{p}\right)^a$ de gagner....
Que se passe-t-il si on suppose que les joueurs sont des entreprises et que le capital est en fait un nombre de parts de marché ?

Chapitre 2

variables aléatoires et loi

2.1 Variables aléatoires réelles

2.1.1 Variables aléatoires et fonctions de répartitions

Définition 2.1.1. *Soit* $(\Omega, \mathcal{B}, P)$ *un espace probabilisé. On appelle application mesurable toute application* $X : \Omega \to \mathbb{R}$ *telle que l'image réciproque de tout intervalle ouvert* $]a, b[$ *par* X *soit un événement .*

Définition 2.1.2. *Soit* $(\Omega, \mathcal{B}, P)$ *un espace probabilisé. On appelle variable aléatoire (va) toute application mesurable* $X : \Omega \to \mathbb{R}$.
Si X *est une va on notera* $X(\Omega)$ *l'image de* Ω *par* X*, et on l'appelle univers image.*

Exemples

1. On jette deux dés discernables ; $\Omega = \{(i, j), 1 \leq i, j \leq 6\}$ et on pose $X : \Omega \to \mathbb{R}$ définie par $X((i, j)) = i + j$ (somme des deux dés). L'univers image est $X(\Omega) = \{2, 3, \ldots, 12\}$.
2. On jette une pièce une infinité de fois. On a $\Omega = \{\underline{x} = (x_n)_{n\in\mathbb{N}^*}, x_n = \text{ Pile ou Face}\}$, qui est un ensemble infini non dénombrable. On pose $X(\underline{w}) =$ rang d'apparition du premier pile (ou 0 s'il n'apparaît jamais). L'univers image est $\mathbb{N}$.

2.1.2 Loi d'une va

Si X est une va sur Ω et $A \subset \mathbb{R}$, on notera $\{X \in A\} \stackrel{def}{=} \{w \in \Omega, X(w) \in A\}$. En particulier on aura $\{X \geq a\} = \{w \in \Omega, X(w) \geq a\}$ et $\{X = a\} = \{w \in \Omega, X(w) = a\}$.

Définition 2.1.3. *Soient Ω un univers et P une probabilité sur $\mathcal{B}$. Si X est une va sur Ω, la probabilité image de P par X définie par*

$$P_X(]a,b[) \stackrel{def}{=} P(X \in]a,b[)$$

s'appelle la loi (de probabilité) de X.

On peut vérifier que cela définit bien une probabilité sur $\mathbb{R}$, muni de la tribu borélienne (tribu engendrée par les intervalles ouverts). Il s'agit bien d'une façon de compter par tranche.

Définition 2.1.4. *Soit X une variable aléatoire réelle. La fonction $t \mapsto P(X \leq t)$ s'appelle la fonction de répartition de X et on la note F_X.*

Proposition 2.1.5. *Soient X une va discrète sur Ω et F_X sa fonction de répartition. Alors*

1. *Pour tout $t \in \mathbb{R}$, $0 \leq F_X(t) \leq 1$.*
2. *F_X est croissante sur $\mathbb{R}$.*
3. *$\lim_{t\to-\infty} F_X(t) = 0$ et $\lim_{t\to+\infty} F_X(t) = 1$.*
4. *Si $a \leq b$ alors $P(\{a \leq X < b\}) = F_X(b) - F_X(a)$.*
5. *F_X est continue à gauche.*

Définition 2.1.6. *Deux variables aléatoires X et Y sont dites indépendantes si pour tout couple (t,u), les événements $\{X \leq t\}$ et $\{Y \leq u\}$ sont indépendants.*

2.2 Les variables aléatoires discrètes

Définition 2.2.1. *Soient Ω un univers fini ou dénombrable et P une probabilité sur $\mathcal{P}(\Omega)$ de distribution (p_n). Si X est une va sur Ω, l'application*

$$\begin{array}{rcl} X(\Omega) & \to & [0,1] \\ a & \mapsto & P(\{X=a\}) \end{array}$$

s'appelle la loi (de probabilité) de X.

On remarque que

$$\sum_{a\in X(\Omega)} P(\{X=a\}) = \sum_{a\in X(\Omega)} \sum_{w,X(w)=a} P(\{w\}) = \sum_{w\in\Omega} P(\{w\}) = 1,$$

par un raisonnement (sommes partielles) déjà vu.

Exemple

Dans l'exemple "somme de deux dés", on a ...

Dans l'exemple "rang d'apparition du premier pile" on pose $P(X = n > 0) = \frac{1}{2^n}$ ce qui définit bien une probabilité sur $X(\Omega)$.

Exemples

Variable de Bernoulli de paramètre p. C'est une va X à valeur dans $\{0, 1\}$ telle que

$$P(X = 1) = p \text{ et } P(X = 0) = q \overset{def}{=} 1 - p.$$

Applications : jeu de pile ou face.

Loi binomiale. Une va X suit la loi binomiale $B(n, p)$ de paramètres $n \in \mathbb{N}^*$ et $p \in [0, 1]$ si elle est telle que pour tout $k \in [0, n]$

$$\mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}.$$

On marquera $X \rightsquigarrow B(n, p)$ pour dire que la loi de X est $B(n, p)$.
Applications : itérations indépendantes de jeu de pile ou face.

Loi hypergéométrique Une va X suit la loi hypergéométrique de paramètres p, $q = 1 - p$ et A si

$$\mathbb{P}(X = k) = \frac{\binom{pA}{k}\binom{qA}{n-k}}{\binom{A}{n}}.$$

On note alors $X \rightsquigarrow \mathcal{H}(p, A)$.
Applications : tirage de n boules parmi A avec une proportion p de boules rouges et $q = 1 - p$ de boules noires.

Loi de Poisson Une va X suit la loi de Poisson de paramètre $\lambda \in \mathbb{R}_+^*$ si elle est telle que pour tout entier $n \geq 0$,

$$\mathbb{P}(X = n) = \frac{\lambda^n e^{-\lambda}}{n!}.$$

On note alors $X \rightsquigarrow P(\lambda)$. Applications : comptages

2.2.1 Fonction génératrice

Nous allons introduire un outils très pratique pour avoir des information sur un va,et donc pour pouvoir comparer (du point de vue des lois) deux va.
Soit X une va discrète à valeur dans $\mathbb{N}$. Si t est dans $[0, 1]$ on pose

$$G_X(t) \overset{def}{=} \sum_{n=0}^{+\infty} P(X = n) t^n = \mathbb{E}[t^X],$$

où t^X désigne la fonction

$$\begin{array}{ccc} \Omega & \to & \mathbb{R} \\ \omega & \mapsto & t^{X(\omega)}. \end{array}$$

Définition 2.2.2. *L'application $G : [0,1] \to \mathbb{R}$ définie par la valeur de $G_X(t)$ donné ci-dessus s'appelle la fonction génératrice de la va discrète X.*

Il faut vérifier que cette fonction est bien définie, c'est à dire que pour tout $t \in [0,1]$ la série $G_X(t)$ converge bien.

Or on se souvient que $\sum_{n=0}^{+\infty} P(X=n)$ converge et vaut 1. De plus pour tout n,

$$0 \leq P(X=n)t^n \leq P(X=n),$$

ce qui prouve bien que la suite de terme général $\sum_{n=0}^{N} P(X=n)t^n$ est croissante et majorée par 1, donc elle converge.

On remarquera que si $X(\Omega)$ est fini, alors $G_X(t)$ est un polynôme en t. De plus, on peut démontrer que $G_X(t)$ peut aussi être défini pour $-1 \leq t \leq 0$.

Exemples

1. Dans l'exemple 1, on trouve $G_X(t) = \dfrac{t^2 + 2t^3 + 3t^4 + 4t^5 + 5t^6 + 6t^7 + 5t^8 + \ldots + t^{12}}{36}$.
2. Dans l'exemple 2 on trouve $G_X(t) = \sum_{n=1}^{+\infty} \left(\dfrac{t}{2}\right)^n = \dfrac{t}{2-t}$.

On admettra la proposition suivante.

Proposition 2.2.3. *Pour tout $k \in \mathbb{N}$, G_X est k-fois dérivable en 0 et*

$$G_X^{(k)}(0) = k!P(X=k).$$

La preuve fait intervenir l'unicité de la décomposition en série entière, quand elle existe. On remarquera que la preuve est évidente si G_X est un polynôme.

Corollaire 2.2.4. *Si $G_X = G_Y$ alors X et Y ont la même loi.*

Attention, deux va peuvent avoir la même loi mais être tr̀es différentes. Considérons par exemple une puce qui se déplace en sautant dans la segment $[-5,5]$. On suppose que chaque saut est de taille 1, et que la probabilité de présence en chaque entier est une fonction paire, c'est à dire que $P(puce\ en\ n) = P(puce\ en\ -n)$. Considérons maintenant les deux va de $[-5,5]$ dans $\mathbb{Z}$ respectivement définies par $X(n) = n$ (position de la puce) et $Y(n) = -n$ (opposé de la position de la puce). Alors X et Y sont différentes mais ont la même loi.

2.3 Introduction aux probabilités et variables aléatoires absolument continues

2.3.1 Exemple

On choisit un point M au hasard dans le segment $[a, b]$, et on note X la va égale à l'abscisse de M. L'univers des possibles est $[a, b]$, et l'événement (si ça en est un) "$X = t$" a –intuitivement– une probabilité nulle.on comprend alors qu'on ne peut pas espérer décrire l'expérience (en terme de probabilité) en donnant la probabilité de choisir tel ou tel point. Pour contourner la difficulter, il faut essayer de modéliser l'expérience en donnant la probabilité de choisir non pas un point précis, mais un ensemble suffisamment gros de points. Il faut aussi pouvoir faire cela pour une classe intéressante d'ensembles. En termes plus mathématiques il faut choisir un tribu qui ne soit pas la tribu complète, mais qui soit néanmoins suffisament fine. L'idée est donc d'introduire la tribu générée par les intervalles.
Si $a \leq u \leq v \leq b$, on voudrait avoir

$$P(u \leq X \leq v) = \frac{long([u, v])}{long([a, b])} = \frac{v - u}{a - b}.$$

Compte tenu de cette définition, on remarque que $P(u \leq X \leq v) = P(u \leq X < v) = P(u < X < v) = P(u < X \leq v)$.

$$\text{Ainsi } P(X \leq t) = \begin{cases} 0 \text{ si } t \leq a \\ \frac{t-a}{b-a} \text{ si } a \leq t \leq b \\ 1 \text{ si } t \geq b \end{cases}$$

Définition 2.3.1. *On dira que $t \mapsto P(X \leq t)$ est la fonction de répartition de X et on la notera F_X.*

On remarque que dans cet exemple, $F_X(t) = \displaystyle\int_{-\infty}^{t} f(x)dx$, où

$$f(x) = \begin{cases} 0 \text{ si } t < a \\ \frac{1}{b-a} \text{ si } a \leq t \leq b \\ 0 \text{ si } t > b \end{cases},$$

ce qui entraine $P(u \leq X \leq v) = \int_u^v f(t)dt$.

Définition 2.3.2. *La fonction f s'appelle densité de probabilité de X.*

Comme f est constante, on dit que X suit la loi uniforme sur $[a, b]$. On remarque qu'on a aussi $\displaystyle\int_{-\infty}^{+\infty} f(t)dt = 1$, que f est positive et que F_X est la primitive de f qui s'annule en $-\infty$.

2.3.2 Densité et fonctions de répartition.

Définition 2.3.3. *On appelle densité de probabilité toute fonction $f : \mathbb{R} \to \mathbb{R}$ continue par morceaux, positive, telle que*

$$\int_{-\infty}^{+\infty} f(t)dt = 1.$$

Définition 2.3.4. *Soit X une application de Ω dans $\mathbb{R}$. Dire que X est une va absolument continue sur Ω signifie qu'il existe une densité de probabilité f telle que pour tout t dans $\mathbb{R}$ on ait*

$$P(\{X \leq t\}) = \int_{-\infty}^{t} f(u)du.$$

On remarque que dans cette définition on fait une sommation par tranche. Si X est une va absolument continue de densité f, la quantité $F_X(t) \stackrel{def}{=} P(X \leq t) = \int_{-\infty}^{t} f(u)du$ sera appelée fonction de répartition de la va.

Proposition 2.3.5. *Soit X une va absolument continue de fonction de répartition F_X et densité f. Alors :*
(1) Si $a < b$, $P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f(t)dt$.
(2) Pour tout $a \in \mathbb{R}$, $P(X = a) = 0$.
(3) $P(a \leq X < b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X \leq b)$.

Démonstration. Preuve du (1). On a $\{X \leq b\} = \{X \leq a\} \sqcup \{a < X < b\}$. Il suffit alors de dire ce que valent $P(X < a)$ et $P(X < b)$.
Preuve de (2). Comme f est continue par morceaux, elle est bornée sur toute partie $[m, M]$. On en déduit qu'il existe K tel pour tout $h \in [0, 1]$,

$$P(a \leq X < a + h) = \int_a^{a+h} f(t)dt \leq Kh.$$

Cela entraine en particulier que $P(X = a) \leq Kh$, pour tout h, et donc cela prouve (2).
Preuve de (3). Il suffit de voir que les différences entre les 4 probabilités sont comprises entre $-P(X = a) - P(X = b)$ et $P(X = a) + P(X = b)$, qui sont nulles par le (2). $\square$

Proposition 2.3.6. *Soit F_X la fonction de répartition d'une va X absolument continue. Alors :*
(1) F_X est continue et dérivable en tout point de continuité de f ; on a alors $F'_X(a) = f(a)$ si a est un point de continuité de f.
(2) F_X est croissante sur $\mathbb{R}$.
(3) F_X est à valeur dans $[0, 1]$.
(4) $\lim_{t \to +\infty} F_X(t) = 1$ et $\lim_{t \to -\infty} F_X(t) = 0$.

Démonstration. Le (1) résulte du cours d'analyse. Les propositions (2) et (3) sont des simples conséquences de la définition de F_X. La proposition (4) résulte de la définition de F_X et du cours d'analyse. □

2.3.3 Exemples de lois

Loi uniforme

Celle dans un intervalle de $\mathbb{R}$ vient d'être vue. On peut aussi définir une loi uniforme dans une partie de $\mathbb{R}^n$. Typiquement la loi uniforme dans le disque $D(0,1)$, ou la carré $[0,1] \times [0,1]$....On y reviendra plus tard.

Loi de Cauchy

Définition 2.3.7. *On dit que la va X suit la loi de Cauchy si c'est une va absolument continue de densité* $t \mapsto f(t) = \frac{1}{\pi}\frac{1}{1+t^2}$.

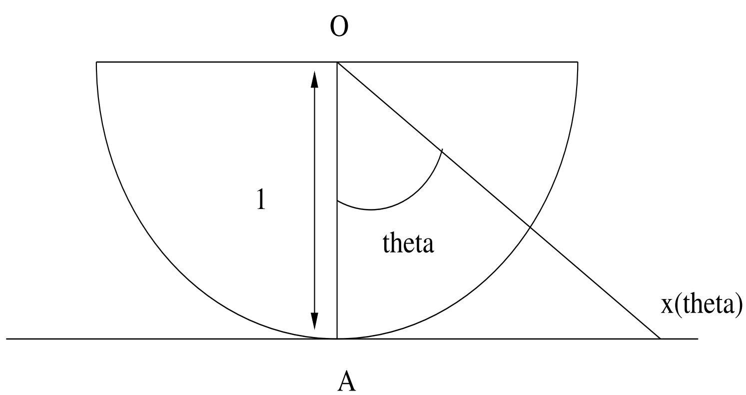
On vérifie que cela définit bien une probabilité, c'est à dire que f est continue par morceaux positive et $\int_{\mathbb{R}} f(t)dt = 1$.
f est évidemment positive, continue, et on a

$$\int_{-a}^{+a} \frac{dt}{1+t^2} = [\arctan t]_{-a}^{+a} = 2\arctan(a).$$

Si $a \to +\infty$, on trouve $\displaystyle\int_{-\infty}^{+\infty} \frac{dt}{1+t^2} = \pi$.
Calcul de la fonction de répartition :
Par définition on a $F_X(t) = \displaystyle\frac{1}{\pi}\int_{-\infty}^{t} \frac{du}{1+u^2}$. En remplaçant la borne $-\infty$ par $-b$, puis en calculant cette nouvelle intégrale, et en passant à la limite on trouve

$$F_X(t) = \frac{1}{\pi}\arctan(t) + \frac{1}{2}.$$

Interprétation géométrique.



On appelle $X(\theta)$ le point sur le droite D tel que $\overrightarrow{OX(\theta)}$ fasse un angle θ avec $\overrightarrow{0A}$. Si θ suit la loi uniforme sur $]-\frac{\pi}{2}\,;\,\frac{\pi}{2}[$, alors X suit la loi de Cauchy.

Variantes : loi de Cauchy avec un paramètre $a > 0$; $f_a(x) = \frac{a}{\pi}\frac{1}{a^2+x^2}$.

Loi de Laplace

Définition 2.3.8. *On dit que la va X suit la loi de Laplace si c'est une va absolument continue de densité $t \mapsto f(t) = \frac{1}{2}e^{-|t|}$.*

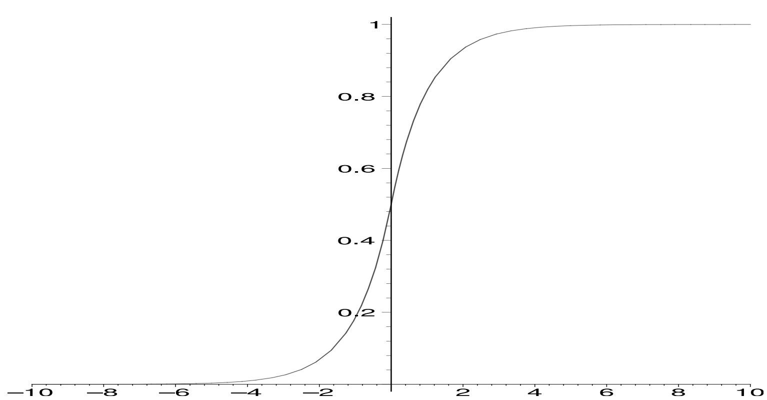
On vérifie que cela définit bien une probabilité.
f est évidemment positive, continue, et on a

$$\int_{-a}^{+a} e^{-|t|}dt = 2\int_0^{+a} e^{-t}dt = 2(1-e^{-a}) \longrightarrow_{a\to+\infty} 2.$$

Cela montre bien que f est une densité de probabilité.

Calcul de la fonction de répartition :
Par définition on a $F_X(t) = \frac{1}{2}\int_{-\infty}^{t} e^{-|u|}du$. Si $t \leq 0$, cette intégrale se calcule facilement, et si $t > 0$, on la coupe en deux parties. On trouve donc :

$$F_X(t) = \begin{cases} \frac{1}{2}e^t \text{ si } t \leq 0 \\ 1 - \frac{1}{2}e^{-t} \text{ si } t \geq 0 \end{cases}$$



Loi exponentielle

Définition 2.3.9. *On dit que la va X suit la loi exponentielle de paramètre $\lambda > 0$ si c'est une va absolument continue de densité*

$$f(t) = \begin{cases} 0 \; si \; t < 0 \\ \lambda e^{-\lambda t} \; si \; t \geq 0 \end{cases}$$

On vérifie que cela définit bien une probabilité.
f est évidemment positive et est continue par morceaux. De plus on a

$$\int_{-a}^{+a} f(t)dt = \int_{0}^{+a} \lambda e^{-\lambda t} dt = (1 - e^{-\lambda a}) \longrightarrow_{a \to +\infty} 1.$$

Cela montre bien que f est une densité de probabilité.

Calcul de la fonction de répartition :

Clairement, $F_X(t) = 0$ si $t \leq 0$. De plus pour $t \geq 0$, $F_X(t) = \lambda \int_0^t e^{-\lambda t} du = 1 - e^{-\lambda t}$.

Proposition 2.3.10. *Si X suit la loi exponentielle de paramètre λ, alors pour tout $t > 0$ et pour tout $h \geq 0$,*

$$P(X \leq t + h | X \geq t) = P(X \leq h).$$

Démonstration. Par définition

$$\begin{aligned} P(X \leq t + h | X > t) &= \frac{P(t < X \leq t + h)}{P(t < X)} \\ &= \frac{F_X(t+h) - F_X(t)}{1 - F_X(t)} \\ &= \frac{e^{-\lambda t} - e^{-\lambda(t+h)}}{e^{-\lambda t}} \\ &= 1 - e^{-\lambda h} = P(X \leq h). \end{aligned}$$

□

En fait cette propriété caractérise la loi exponentielle. Elle s'exprime par la relation

$$F(t + h) - F(t) = (1 - F(t))F(h) \tag{2.1}$$

si on divise par h est qu'on fait tendre h vers 0, en acceptantl'hypothèse F dérivable en 0, on trouve une équation différentielle

$$F'(t) + \kappa F(t) = \kappa, \tag{2.2}$$

qui donne comme solution (compte-tenu des conditions aux limites) $F(t) = 1 - e^{\kappa t}$ et κ paramètre négatif. On retrouve bien la loi exponentielle de paramètre $\lambda = -\kappa$.

Application. Le nombre de véhicules qui passent à un péage entre les instants 0 et T suit la loi de Poisson de paramètre T. On souhaite donner la loi du temps d'attente entre 2 véhicules.
Si T est cette va, on aura

$$P(T > t) = P(\text{entre les instants 0 et } T \text{ il ne passe aucun véhicule}),$$

ce qui permet d'avoir $P(T \leq t) = 1 - e^{-T}$.

2.4 Espérance, variance.

2.4.1 Espérance

Définition 2.4.1. *Soit X une va discrète sur un univers Ω. On appelle espérance (parfois moyenne) de X la valeur*

$$\mathbb{E}[X] \stackrel{def}{=} \sum_{a \in X(\Omega)} aP(X = a),$$

à condition que cette quantité existe.

Si X est une va, on note parfois son espérance (si elle existe) $\overline{X}$. cette notation est malheureuse car elle fait penser à l'événement *non* X. Dans ce cours d'introduction, nous noterons donc l'espérance $\underline{X}$.

Exemples

1. On trouve $\mathbb{E}[X] = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + \ldots + 12 \times \frac{1}{36} = 7$. Ceci signifie que en moyenne,la somme des deux chiffres vaut 7.
2. $\mathbb{E}[X] = \sum_{n=1}^{+\infty} n \frac{1}{2^n} = 2$.

On appelle $\mathcal{L}^1$ l'ensemble des va absolument continues, X, de densité de probabilité f telles que $\displaystyle\int_{-\infty}^{+\infty} |x| f(x) dx$ existe (dans $\mathbb{R}$). On verra plus loin que $\mathcal{L}^1$ est un $\mathbb{R}$-espace vectoriel.

Définition 2.4.2. *Soit X une va absolument continue, de densité f. Si X est dans $\mathcal{L}^1$ alors on note $\mathbb{E}[X]$ la quantité $\int_{-\infty}^{+\infty} x f(x) dx$ (qui existe) et on l'appelle espérance de X*

L'existence de l'intégrale résulte du cours d'analyse.

Proposition 2.4.3. *Soient X et Y deux va réelles définies sur un même espace probabilisé $(\Omega, \mathcal{B}, P)$ et admettant chacune une espérance.*
Alors pour tout a, b dans $\mathbb{R}$, $aX + bY$ admet une espérance et $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$

La preuve est admise (pour le moment). Il est faux de croire que la densité de $X + Y$ sera $f + g$ (si f est la densité de X et g celle de Y), ne serait-ce que parce que $\int_{\mathbb{R}} (f + g) = 2$ et non pas 1. Néanmoins cette proposition montre que $\mathcal{L}^1$ est un $\mathbb{R}$-espace vectoriel.

Comme le résultat précédent on admettra le suivant, qui sera toutefois démontré en TD dans certains cas ($\varphi(x) = ax + b$, $= x^2$, $= e^x$ si X suit la loi de Laplace).

Théorème 2.4.1. *Soient X une va absolument continue de densité f, et $\varphi : \mathbb{R} \to \mathbb{R}$ dérivable. Alors la va $Y = \varphi(X)$ est une va absolument continue ; elle est dans $\mathcal{L}^1$ si et seulement si*

$$\int_{-\infty}^{+\infty} |\varphi(x)| f(x) dx < +\infty,$$

et on aura alors $\mathbb{E}[Y] = \displaystyle\int_{-\infty}^{+\infty} \varphi(x) f(x) dx$.

2.4.2 Variance

Définition 2.4.4. *Soit X une va réelle admettant une espérance $\bar{X}$. On appelle variance de X la quantité (si elle existe)*

$$V(X) = \mathbb{E}[(X - \bar{X})^2],$$

et écart type de X le réel $\sigma(X) = \sqrt{V(X)}$ (aussi écart quadratique moyen).

Proposition 2.4.5. *Soit X une va réelle admettant une variance. Alors*
(1) $V(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
(2) pour tout a, b $V(aX + b) = a^2 V(X)$.

Démonstration. Il suffit d'utiliser le théorème 2.4.1 et de faire comme pour les va discrètes. □

On admettra que si X est une va dont la variance est nulle, alors elle est constante.

On peut aussi définir l'ensemble $\mathcal{L}^2$ des va X telles que X^2 soit dans $\mathcal{L}^1$. On a même :

Proposition 2.4.6. *$\mathcal{L}^2$ est un sous-espace vectoriel de $\mathcal{L}^1$. Si X et Y sont dans $\mathcal{L}^2$, alors la va XY est dans $\mathcal{L}^1$ et on a* **l'inégalité de Cauchy-Schwarz** *:*

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}.$$

Démonstration. Comme $|XY| \leq \frac{1}{2}(X^2 + Y^2)$, si X et Y sont dans $\mathcal{L}^2$, la va XY est bien dans $\mathcal{L}^1$. De plus, par linéarité de l'espérance on a

$$0 \leq \mathbb{E}[(\lambda X + Y)^2] = \lambda^2 \mathbb{E}[X^2] + 2\lambda \mathbb{E}[XY] + \mathbb{E}[Y^2],$$

ce qui prouve que le discriminant de ce polynôme est négatif ou nul. On trouve ainsi l'inégalité demandé. On en déduit donc facilement que $\mathcal{L}^2$ est un sous-espace de $\mathcal{L}^1$. □

Complément pour les va discrètes

Proposition 2.4.7. *Si X est une va discrète qui admet un espérance et une variance, alors $\mathbb{E}[X] = G'_X(1)$ et $V(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2$.*

2.5 Calculs pour les variables classiques

2.5.1 Variable de Bernoulli

Définition 2.5.1. *Soient Ω un univers (fini ou dénombrable) et P une probabilité sur $\mathcal{P}(\Omega)$. On appelle variable de Bernoulli de paramètre $p \in [0,1]$ toute va discrète X sur Ω à valeur dans $\{0,1\}$ telle que*

$$P(X=1) = p \text{ et } P(X=0) = q \overset{def}{=} 1-p.$$

Exemples

(1) On joue à pile ou face avec une probabilité p de tomber sur pile et $q = 1-p$ de tomber sur face.
(2) On lance un dé, avec une probabilité p d'obtenir un chiffre pair et $q = 1-p$ d'obtenir un chiffre impair.
(3) On lance un dé, avec une probabilité p d'obtenir un chiffre multiple de 3 et $q = 1-p$ d'obtenir un chiffre non multiple de 3.

Proposition 2.5.2. *Si X est une va de Bernoulli de paramètre p, alors $\mathbb{E}[X] = p$, $V(X) = pq$ et $G_X(t) = pt + q$.*

Démonstration. Clairement, $\mathbb{E}[X] = 0.q+1.p = p$. De plus $V(x) = \mathbb{E}[X^2]-\mathbb{E}[X]^2 = p - p^2 = p(1-p) = pq$. Enfin, $G_X(t) = \mathbb{E}[t^X] = qt^0 + pt = pt + q$. □

2.5.2 Loi binômiale

Définition 2.5.3. *Soient n un entier positif, p dans $[0,1]$ et $q = 1-p$. On appelle loi binômiale de paramètres n et p la loi de toute va discrète $X : \Omega \to \{0,1,\ldots,n\}$ telle que pour tout $k \in [0,n]$*

$$P(X=k) = \binom{n}{k} p^k q^{n-k}.$$

On la notera $B(n,p)$ et on maquera $X \rightsquigarrow B(n,p)$ pour dire que la loi de X est $B(n,p)$.

Il faut quand même vérifier que $B(n,p)$ est bien une loi, c'est à dire que cela définit une probabilité sur l'univers image $X(\Omega)$.
On a $\sum_{k=0}^{n} P(X=k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1.$

Exemple

On jour n fois à pile ou face ; Chaque tirage est indépendant, et il y a une probabilité p de tirer pile et q de tirer face. Si $X =$"nombre de fois où on tire pile", alors $X \rightsquigarrow B(n,p)$.
En effet calculons $P(X = k)$. Il faut choisir les k fois parmi n où l'on tire pile. Il y a $\binom{n}{k}$ choix possibles. Pour chacune de ces fois il y a une probabilité p de tirer pile. Pour chacune des $n-k$ fois restantes, il y a une probabilité q de tirer face. On trouve donc bien

$$P(X = k) = \binom{n}{k} p^k q^{n-k}.$$

Nous allons généraliser ce résultat :

Proposition 2.5.4. *Soient $X_1, \ldots, X_n$ n va de Bernoulli de paramètre p indépendantes entre elles. Si $S = X_1 + \ldots + X_n$ alors S suit la loi binômiale $B(n,p)$.*

Démonstration. On clacule $P(S = k)$. Or $S = k$ signifie que k des va X_i valent 1 et $n-k$ valent 0. On refait le raisonnement du jeu de pile ou face. □

Proposition 2.5.5. *Si S suit la loi binômiale $B(n,p)$, alors $\mathbb{E}[S] = np$, $V(S) = npq$ et $G_S(t) = (pt+q)^n$.*

Démonstration. On calcule d'abord l'espérance :

$$\begin{aligned}
\mathbb{E}[S] &= \sum_{k=0}^{n} kP(X = k) \\
&= \sum_{k=0}^{n} \binom{n}{k} kp^k q^{n-k} \\
&= p \sum_{k=0}^{n} \binom{n}{k} kp^{k-1} q^{n-k} \\
&= p \frac{\partial}{\partial x}(x+q)^n_{|p} \\
&= pn(p+q)^{n-1} = np.
\end{aligned}$$

Calculons maintenant le variance.

$$\begin{aligned}
V(S) &= \mathbb{E}[S^2] - n^2p^2 \\
&= \sum_{k=0}^{n} \binom{n}{k} k^2 p^k q^{n-k} - n^2p^2 \\
&= p^2 \sum_{k=0}^{n} \binom{n}{k} k(k-1)p^{k-2} q^{n-k} + p \sum_{k=0}^{n} \binom{n}{k} kp^{k-1} q^{n-k} - n^2p^2 \\
&= p^2 n(n-1) + np - n^2p^2 \\
&= npq.
\end{aligned}$$

On calcule enfin $G_S(t)$.

$$\begin{aligned} G_S(t) &= \mathbb{E}[t^S] \\ &= \sum_{k=0}^{n} t^k P(X=k) \\ &= \sum_{k=0}^{n} C_n^k t^k p^k q^{n-k} \\ &= (pt+q)^n. \end{aligned}$$

On remarque qu'on aurait pu calculer directement $G_S(t)$, et en tirer $\mathbb{E}[S] = G'(1)$ et $V(S) = G''_S(1) + G'_S(1) - G'_S(1)^2$. □

Théorème 2.5.1. *Soient S et S' deux va indépendantes de loi respective $B(n,p)$ et $B(m,p)$. Alors $S+S'$ suit la loi $B(n+m,p)$.*

Démonstration. Comme S et S' sont indépendantes, on a

$$G_{S+S'}(t) = G_S(t)G_{S'}(t) = (pt+q)^n(pt+q)^m.$$

Ainsi $S+S' \rightsquigarrow B(n+m,p)$. □

2.5.3 Loi de Poisson

Définition 2.5.6. *On dit qu'une va discrète X définie sur un univers infini Ω à valeur dans $\mathbb{N}$ suit la loi de Poisson de paramètre λ si pour tout n,*

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

On note alors $X \rightsquigarrow P(\lambda)$.

Comme pour le loi binômiale, il faut vérifier qu'on définit bien une probabilité sur l'univers image $X(\Omega)$.
On a $\displaystyle\sum_{k=0}^{+\infty} P(X=n) = \sum_{k=0}^{+\infty} \frac{\lambda^n e^{-\lambda}}{n!} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^n}{n!} = 1$.
Cette loi sert à décrire toute une catégorie de phénomènes, tels le comptage de véhicules franchissant un péage, le nombre d'appels téléphoniques reçus par un standard, le nombre de défauts d'un objet fabriqué en série (voir en fin de cours). Mais la justification de la loi de Poisson se verra comme loi qui approxime la loi binômiale.

Proposition 2.5.7. *Si X suit une loi de Poisson de paramètre λ, alors $\mathbb{E}[X] = \lambda$, $V(X) = \lambda$ et $G_X(t) = e^{\lambda(t-1)}$.*

Démonstration. On calcule d'abord la fonction génératrice.

$$\begin{aligned} G_X(t) &= \sum_{k=0}^{+\infty} t^n P(X=n) \\ &= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(t\lambda)^n}{n!} \\ &= e^{-\lambda} e^{t\lambda}. \end{aligned}$$

On obtient alors les valeurs en dérivant. □

Théorème 2.5.2. *Si X et Y sont deux va discrètes indépendantes qui suivent les loi $P(\lambda)$ et $P(\mu)$, alors $X+Y \rightsquigarrow P(\lambda+\mu)$.*

Démonstration. $G_{X+Y}(t) = G_X(t)G_Y(t) = e^{\lambda(t-1)}e^{\mu(t-1)} = e^{(\lambda+\mu)(t-1)}$. □

Approximation de la loi binômiale

On suppose qu'on joue à pile ou face n fois, que la loi du résultat est $B(n,p)$. On suppose que n devient très grand, avec un p petit, c'est à dire que np est supposé rester constant valant θ. Alors

$$\begin{aligned} P(X=k) &= C_n^k p^k (1-p)^{n-k} \\ &= \frac{n(n-1)\dots(n-k+1)}{k!} p^k (1-p)^{n-k} \\ &= \frac{(np)^k (1-\frac{np}{n})^n}{k!} (1-\frac{1}{n})\dots(1-\frac{k-1}{n}) \frac{1}{(1-\frac{np}{n})^k} \\ &= \frac{\theta^k (1-\frac{\theta}{n})^n}{k!} (1-\frac{1}{n})\dots(1-\frac{k-1}{n}) \frac{1}{(1-\frac{\theta}{n})^k} \end{aligned}$$

En passant à la limite en n, on trouve donc

$$P(X=k) \sim \frac{\theta^k e^{-\theta}}{k!}.$$

Ainsi la loi de Poisson de paramètre λ approxime la loi binômiale de paramètres n et λ/n si n est grand.

Proposition 2.5.8. *Si $n \geq 50$ et si $\lambda = np \leq 5$ on peut remplacer la loi binômiale par la loi de Poisson en faisant une erreur de moins de 10^{-2} sur chaque p_k.*

2.5.4 Loi uniforme sur $[a, b]$

Si X suit la loi uniforme sur $[a, b]$, calculer son espérance et sa variance revient à calculer l'intégrale de polynômes sur un intervalle borné. Ainsi X est dans $\mathcal{L}^1$ et dans $\mathcal{L}^2$.

On a

$$\mathbb{E}[X] := \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{b-a}\frac{b^2-a^2}{2} = \frac{a+b}{2}.$$

$$V(X) = \frac{(b-a)^2}{12}.$$

2.5.5 Loi de Cauchy

Si X suit la loi de Cauchy sur $\mathbb{R}$ son espérance, si elle existe doit être égale à

$$\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x}{1+x^2}dx.$$

Cette intégrale ne converge pas en $\pm\infty$, donc X n'admet ni espérance ni variance.

2.5.6 Loi de Laplace

Soit X une va qui suit la loi de Laplace. On note que les intégrales $\int^{\pm\infty} xe^{-|x|}\,dx$ convergent. Par parité, on obtient

$$\mathbb{E}[X] = \frac{1}{2}\int_{-\infty}^{+\infty} xe^{-|x|}\,dx = 0.$$

Pour la variance, $\int^{\pm\infty} x^2e^{-|x|}\,dx$ convergent et on trouve donc

$$V(X) = \int_0^{+\infty} x^2e^{-x}\,dx = 2.$$

2.5.7 Loi exponentielle

Si X suit la loi exponentielle de paramètre λ, $\int^{+\infty} x^n e^{-\lambda x}\,dx$ converge pour tout entier n. Donc X admet une espérance et une variance. Les calculs donnent

$$\mathbb{E}[X] = \frac{1}{\lambda} \quad V(X) = \frac{1}{\lambda^2}.$$

2.6 La loi normale (ou loi de Gauss)

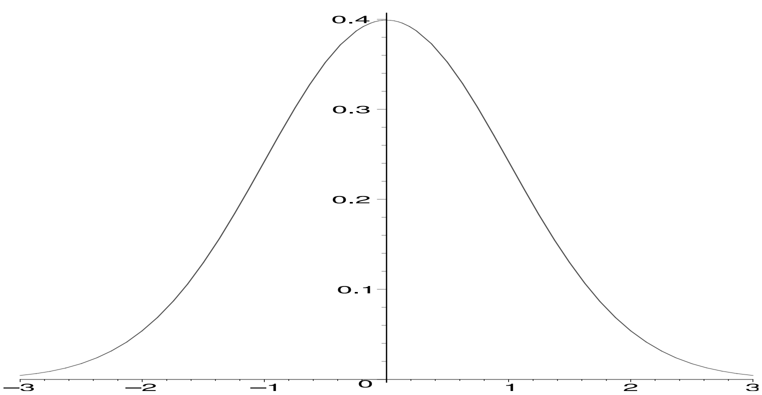
2.6.1 La loi normale centrée réduite

Définition 2.6.1. *Une va X suit la loi normale réduite si c'est une va absolument continue admettant pour densité la fonction*

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

On remarque que f est continue paire et positive. Ainsi $\int_{-\infty}^{+\infty} f(x)dx = 2\frac{1}{\sqrt{2\pi}} \int_0^{+\infty} e^{-\frac{1}{2}x^2} dx$. En faisant $u = \frac{x}{\sqrt{2}}$ on trouve 1.

Étude de la densité. On remarque que f est décroissante sur $\mathbb{R}_+$, et présente un uniquepoint d'inflexion pour $x = 1$. Sa courbe est une courbe de Gauss.



Espérance et variance. L'espérance existe car $f(x) =_\infty o(\frac{1}{x^3})$. Comme f est paire, on a alors

$$\mathbb{E}[X] = 0,$$

ce qui signifie que X est centrée.
De même $f(x) =_\infty o(\frac{1}{x^4})$, ce qui montre que la variance existe. On peut la calculer par intégrations par partie successives. On peut aussi remarquer que $f''(x) = (x^2 - 1)f(x)$, ce qui permet de calculer $\int_0^{+\infty} x^2 f(x)dx$. On trouve alors

$$V(X) = 1,$$

ce qui signifie que X est réduite.

Fonction de répartition. On notera π la fonction de répartition d'une va suivant le loi normale centrée réduite :

$$\pi(t) = P(X < t) = p(X \leq t) = \int_{-\infty}^{t} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} dx.$$

Proposition 2.6.2. *Avec les notations précédentes :*
(1) $\pi(0) = 0,5$.
(2) π *est dérivable sur* $\mathbb{R}$, *de dérivé* $\pi'(x) = \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}$, *et est croissante.*
(3) $\lim_{t\to+\infty} \pi(t) = 1$ *et* $\lim_{t\to-\infty} \pi(t) = 0$.
(4) Pour tout t, $\pi(t) + \pi(-t) = 1$ *et* $\pi(t) - \pi(-t) = P(|X| < t) = 2\pi(t) - 1$.

Démonstration. Seul le dernier point n'est pas évident. Or $\pi(-t) = \int_{-\infty}^{-t} f(t)dt$ et comme f est paire, on trouve $\pi(-t) = \int_{+t}^{+\infty} f(t)dt$, ce qui montre que $\pi(t)+\pi(-t) = 1$.
De plus $\pi(t) - \pi(-t) = \int_{-t}^{t} f(t)dt = P(|X| < t)$... □

Emploi des tables

1. $P(|X| \leq 1) = 2\pi(1) - 1 = 2 * 0,8413 - 1 = 0,6826$;
   $P(|X| \leq 2) = 2\pi(2) - 1 = 2 * 0,9772 - 1 = 0,9544$.
2. Réciproquement, α étant fixé, on cherche t tel que $P(|X| \leq t) \leq \alpha$. Cela revient à chercher t tel que $\pi(t) = \frac{1+\alpha}{2}$.
   Application numérique.

$\alpha = 0,90$	$\pi(t) = 0,95$	$t = 1,65$	
$\alpha = 0,95$	$\pi(t) = 0,975$	$t = 1,96$	$P(|X| \leq 1,96) = 0,95$
$\alpha = 0,99$	$\pi(t) = 0,995$	$t = 2,58$	$P(|X| \leq 2,58) = 0,99$
$\alpha = 0,999$	$\pi(t) = 0,9995$	$t = 3,3$	$P(|X| \leq 3,3) = 0,999$.

   α s'appelle le coefficient de sécurité en statistique.

2.6.2 La loi normale $\mathcal{N}(m, \sigma^2)$

Définition 2.6.3. *Une va absolument continue* X *suit la loi normale (ou loi de Gauss) de paramètres* (m, σ^2) *si elle admet pour densité la fonction définie par*

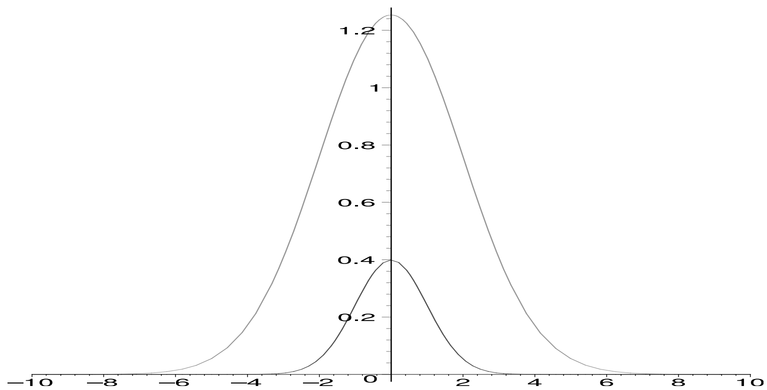
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}.$$

On notera alors $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$.

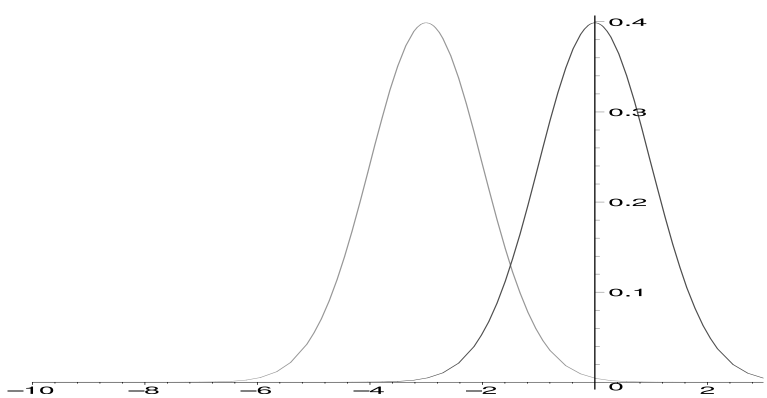
Par exemple si X suit la loi normale réduite, on notera $X \rightsquigarrow \mathcal{N}(0, 1)$. Il convient de vérifier que f ainsi définie est bien une densité de probabilité.
Clairement f est positive et continue. De plus, en effectuant le changement de variable $u = \frac{x-m}{\sigma}$, on montre que

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

<u>*Étude de la densité*</u>. Le maximum de f est atteint pour $x = m$ et vaut $\frac{1}{\sigma\sqrt{2\pi}}$. Ainsi si m reste fixe, mais si σ augmente, le maximum diminue :



Si σ reste fixe, mais m varie, alors la “bosse” se déplace sans changer la valeur du maximum :



Proposition 2.6.4. *Soit X qui suit la loi normale $\mathcal{N}(m, \sigma^2)$. Alors la va $Y = \dfrac{X-m}{\sigma}$ suit la loi normale réduite.*

Démonstration. Il suffit de calculer $P(Y < t)$.

$$\begin{aligned}
P(Y < t) &= P(\frac{X-m}{\sigma} < t) \\
&= P(X < \sigma t + m) \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\sigma t + m} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}u^2} du,
\end{aligned}$$

en effectuant le changement de variable $u = \frac{x-m}{\sigma}$. □

La va Y s'appelle la va réduite associée à X.

Corollaire 2.6.5. *Si $X \leadsto \mathcal{N}(m, \sigma^2)$ alors $\mathbb{E}[X] = m$ et $V(X) = \sigma^2$.*

Démonstration. On note Y la va réduite associée à X. Alors

$$0 = \mathbb{E}[Y] = \mathbb{E}[\frac{X-m}{\sigma}] = \frac{\mathbb{E}[X]-m}{\sigma}.$$

De plus $1 = V(Y) = V(\dfrac{X-m}{\sigma}) = \dfrac{1}{\sigma^2}V(X)$. $\square$

On peut aussi calculer la dispersion d'une va qui suit la loi $\mathcal{N}(m, \sigma^2)$.
On remarque que si $Y = \dfrac{X-m}{\sigma}$, alors

$$|Y| \leq t \iff m - t\sigma \leq X \leq m + t\sigma.$$

Ainsi, $P(m-\sigma \leq X \leq m+\sigma) = 0,6826$ et $P(m-2\sigma \leq X \leq m+2\sigma) = 0,9544$.

Exemple

On suppose que la taille X des individus d'une population suit la loi $\mathcal{N}(m, \sigma^2)$. Calculer m et σ sachant que 4% des individus mesurent moins de 1m60 et 11% des individus mesurent plus de 1m80.
On pose $X = \sigma Y + m$, où $Y \rightsquigarrow \mathcal{N}(0,1)$. Alors
$P(X \leq 160) = P(Y \leq \frac{160-m}{\sigma}) = \pi(\frac{160-m}{\sigma}) = 0,04$. Donc $\pi(\frac{m-160}{\sigma}) = 0,96$, ce qui donne $\frac{m-160}{\sigma} = 1,76$.
De même $P(X \geq 180) = P(Y \geq \frac{180-m}{\sigma}) = 1 - \pi(\frac{180-m}{\sigma}) = 0,11$.
On trouve donc $\pi(\frac{180-m}{\sigma}) = 0,89$, soit $\frac{180-m}{\sigma} = 1,23$.
En combinant les deux égalités on trouve $m = 172$ et $\sigma = 6,7$.

Chapitre 3

Vecteurs aléatoires, indépendance, calculs de lois.

3.1 Introduction et définitions de vecteurs aléatoires

3.1.1 Introduction

Comme nous l'avons vu précédemment, si X et Y sont deux variables aléatoires de densité respective, f et g, il n'apparaît pas clairement que d'une part la va $X+Y$ soit absolument continue, et d'autre part que sa densité s'exprime de façon simple (algébriquement parlant) en fonction de f et de g.
Il n'est donc pas facile de prouver le résultat suivant déjà vu :

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

En outre, dans le cas de va discrètes, on a $V(X+Y) = V(X)+V(Y)+2Cov(X,Y)$, et on aimerait bien avoir une formule similaire dans le cas des va absoluments continues. En fait, une bonne façon pour étudier l'interdépendance de 2 va X et Y, est d'étudier les propriétés probabilistiques du vecteur $\begin{pmatrix} X \\ Y \end{pmatrix}$.

3.1.2 Intégrales multiples

Il ne s'agit pas là de faire un vrai cours, mais simplement d'introduire les objets nécessaire à l'étude des vecteurs aléatoires.

— Si f est une fonction de 2 variables réelles, l'intégrale $\int_a^b \int_c^d f(x,y)dxdy$ vaut $\int_a^b G(y)dy$ où $G(y)$ vaut $\int_c^d f(x,y)dx$, la valeur de y étant un paramètre dans cette dernière intégrale. Il est donc important de constater l'ordre $dx\,dy$ qui donne l'ordre d'intégration.

Exemple

$$\begin{aligned}
\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}} \cos(x+y)dxdy &= \int_0^{\frac{\pi}{2}} [\sin(y+x)]_0^{\frac{\pi}{2}} dy \\
&= \int_0^{\frac{\pi}{2}} \sin(y+\frac{\pi}{2}) - \sin(y)dy \\
&= [-\cos(y+\frac{\pi}{2})]_0^{\frac{\pi}{2}} + [\cos(y)]_0^{\frac{\pi}{2}} \\
&= 0
\end{aligned}$$

Parfois certaines intégrales multiples sont de fausses intégrales multiples ; c'est par exemple le cas si f est à variables séparées, c'est à dire du type $f(x,y) = f_1(x)f_2(y)$. On aura alors

$$\int_a^b \int_c^d f(x,y)dxdy = \int_a^b f_1(x)dx \int_c^d f_2(y)dy.$$

— Comme pour les intégrales dans $\mathbb{R}$ nous nous restreignons aux intégrales absoluments convergentes, c'est à dire que les parties positives et négatives de f ont une intégrale convergente. Ceci nous permet de n'étudier que des intégragles doubles de fonctions positives ; on les supposes suffisament régulières (au moins $\mathcal{C}^0$ par morceaux). On a alors le théorème de Fubini.

  Théorème 3.1.1 (Fubini)**.** $\displaystyle\int_a^b \int_c^d f(x,y)dxdy = \int_c^d \int_a^b f(x,y)dydx.$

  L'explication est qu'il revient au même de sommer en ligne ou en colonne. Cela justifie aussi l'écriture

$$\int_a^b \int_c^d f(x,y)dxdy = \iint_{[a,b]\times[c,d]} f(x,y)dxdy,$$

  qui exprime qu'on somme sur un rectangle. Il faut néanmoins faire attention aux bornes des intégrales. Par exemple on ne peut pas appliquer le théorème de Fubini dans une intégrale du type $\displaystyle\int_0^1 \int_0^x f(x,y)dydx$ puisque l'intégrale $\displaystyle\int_0^x \int_0^1 f(x,y)dxdy$ n'a aucun sens.

— Jacobien d'une transformation. On considère un $\mathcal{C}^1$-difféomorphisme ϕ entre une région A de $\mathbb{R}^2$ et son image $B = \phi(A)$. Ceci signifie que ϕ est une bijection $\mathcal{C}^1$ entre A et B et que ϕ^{-1} est aussi $\mathcal{C}^1$. On peut écrire matriciellement

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \phi(x,y).$$

On appelle alors jacobien de ϕ la déterminant $J = \begin{vmatrix} \frac{\partial X}{\partial x} & \frac{\partial X}{\partial y} \\ \frac{\partial Y}{\partial x} & \frac{\partial Y}{\partial y} \end{vmatrix}$. La formule de changement de variable s'écrit alors

$$\iint_{B=\phi(A)} f(X,Y)dXdY = \iint_A f \circ \phi(x,y)|J|dxdy.$$

Exemple
le changement de variable en, polaire est classique. Si ρ et θ sont les coordonnées polaires, on a $x = \rho\cos\theta$ et $y = \rho\sin\theta$. On trouve comme jacobien

$$\begin{vmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{vmatrix} = \rho.$$

— La même formule est encore valable pour des changements de variables en dimension supérieure. Le jacobien étant alors celui associé à la transformation.
— Il existe des formules d'intégrations par parties. Elles sont néanmoins compliquées et nous ne les utiliserons pas ; il n'est donc pas nécessaire d'en inventer une en examen.

En guise d'exemple nous allons calculer une intégale déjà vue : $\int_0^{+\infty} e^{-x^2}dx$. On appelle I cette intégrale. On a alors

$$\begin{aligned} I^2 &= \int_0^{+\infty} e^{-x^2}dx \int_0^{+\infty} e^{-x^2}dx \\ &= \int_0^{+\infty} e^{-x^2}dx \int_0^{+\infty} e^{-y^2}dy \\ &= \iint_{[0,+\infty[\times[0,+\infty[} e^{-(x^2+y^2)}dxdy \\ &= \int_0^{\frac{\pi}{2}} \int_0^{+\infty} e^{-\rho^2}\rho d\rho d\theta \\ &= \frac{\pi}{4}. \end{aligned}$$

Ceci nous permet de retrouver $\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}dx = \sqrt{2\pi}$. On remarque que la fonction est paire donc l'intégrale est le double de l'intégrale sur $[0,+\infty[$. Cette dernière vaut $\sqrt{2}I$ (changement de variable $x \mapsto \frac{x}{\sqrt{2}}$). D'où

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

3.1.3 Vecteurs aléatoires

Définition 3.1.1. *On appelle densité de probabilité sur $\mathbb{R}^n$ toute fonction $f : \mathbb{R}^n \to \mathbb{R}$ "continue par morceaux", positive, telle que*

$$\int_{\mathbb{R}^n} f(t)dt = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(t_1, \dots, t_n)dt_1 \dots dt_n = 1.$$

Définition 3.1.2. *Soient $(\Omega, \mathcal{B}, P)$ un espace probabilisé et $X = (X_1, \dots, X_n)$ une famille de n variables aléatoires réelles. On dira que X est un vecteur aléatoire absolument continue sur Ω s'il existe une densité de probabilité, f, telle que pour tout $(x_1, \dots, x_n) \in \mathbb{R}^n$,*

$$P\Big(\bigcap_{1\le i\le n} \{X_i \le x_i\}\Big) \overset{def}{=} \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(t_1, \dots, t_n)dt_1 \dots dt_n.$$

Exemple
le vecteur (X, Y) qui suit la loi uniforme sur le carré $A = [0, 1]^2$ est le vecteur de densité la fonction indicatrice du carré, $\mathbb{1}_A$. Plus généralement si B est un ouvert de $\mathbb{R}^2$ le vecteur (X, Y) qui suit la loi uniforme sur B est le vecteur de densité de probabilité la fonction indicatrice de B (valant 1 sur B et 0 ailleurs).

Proposition 3.1.3. *Si (X, Y) est un vecteur absolument continu de densité f, alors il existe un ensemble de surface strictement positive, A, sur lequel la densité f ne s'annule pas.*

Nous ne donnerons pas la preuve de ce résultat ; nous retenons néanmoins qu'un vecteur aléatoire qui est presque sûrement localisé sur un ensemble de surface nulle ne peut pas être absolument continu.

Proposition 3.1.4. *Soit X une va dans $\mathbb{R}^2$ de densité f. Posons $X = (Y, Z)$. Alors Y et Z sont deux va dans $\mathbb{R}$ absolument continues et de densités respectives les fonctions f_Y et f_Z données par*

$$f_Y(y) = \int_{-\infty}^{+\infty} f(y, \xi)d\xi, \quad f_Z(z) = \int_{-\infty}^{+\infty} f(\xi, z)d\xi.$$

Les deux fonctions f_Y et f_Z s'appellent les densités marginales de f.

ATTENTION la réciproque à cette proposition est fausse ; Y et Z peuvent être deux va absolument continues avec densité sans que la va vectorielle $X = (Y, Z)$ soit avec densité. En effet, prenons $Y = Z$, Y suivant la loi normale (par exemple) et $\Delta = \{(x, x), x \in \mathbb{R}\}$ la diagonale de $\mathbb{R}^2$. Clairement $P(X \in \Delta) = 1$, et si X avait une densité f, on devrait avoir $P(X \in \Delta) = \int_{\mathbb{R}^2} \mathbb{1}_\Delta(x)f(x)dx = 0$.

Démonstration. $P(Y \le y) = P(X \in] - \infty, y] \times \mathbb{R}) = \int_{-\infty}^{y} \int_{-\infty}^{+\infty} f(\zeta, \xi)d\xi d\zeta = \int_{-\infty}^{y} f_Y(\zeta)d\zeta.$ □

Exemple
On parlera de vecteur aléatoire (X, Y) de densité uniforme sur le disque $D(0,1)$, sur le carré $[0,1] \times [0,1]$... comme étyant le vecteur dont la densité est l'indicatrice du disque, du carré, etc.

3.2 Variables aléatoires indépendantes et covariance

3.2.1 Indépendance

Définition 3.2.1. *Soient X et Y deux va absolument continues définies sur un même espace probabilisé $(\Omega, \mathcal{B}, P)$. Les deux va sont dites indépendantes si et seulement si pour tout couple d'intervalles de $\mathbb{R}$ (en fait de boréliens) I et J, les événements $\{X \in I\}$ et $\{Y \in J\}$ sont indépendants, ie*

$$P(X \in I \text{ et } Y \in J) = P(X \in I)P(Y \in J).$$

On remarquera que cette définition s'étend à l'indépendance de plusieurs variables entre elles. On donne d'abord une caractérisation importante des va indépendantes :

Proposition 3.2.2. *Deux va réelles X et Y de densité respectives f et g sont indépendantes si et seulement si la va vectorielle $Z = (X, Y)$ admet comme densité la fonction $(x, y) \mapsto h(x, y) = f(x)g(y)$.*

Démonstration. Si X et Y sont indépendantes on a

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) = \int_{-\infty}^{x} f(\xi)d\xi \int_{-\infty}^{y} f(\zeta)d\zeta = \int_{-\infty}^{x}\int_{-\infty}^{y} h(\xi, \zeta)d\xi d\zeta,$$

ce qui signifie que Z admet h comme densité.
Réciproquement, si Z admet h comme densité, le calcul de dessus prouve que pour tout x et pour tout y on a

$$P(X \in]-\infty, x] \cap Y \in]-\infty, y]) = P(X \leq x)P(Y \leq y).$$

Tout intervalle $[a, b]$ de $\mathbb{R}$ s'écrit $]-\infty, b] \setminus]-\infty, a[$, on obtient la propriété voulue qui définie l'indépendance (au moins pour tous les intervalles). $\square$

Proposition 3.2.3. *Soient X et Y deux va indépendantes, absolument continues, définies sur un même espace probabilisé $(\Omega, \mathcal{B}, P)$, et g et h deux fonctions mesurables de $\mathbb{R}$ dans $\mathbb{R}$.*
(1) Les va $g(X)$ et $h(Y)$ sont indépendantes.
(2) si de plus $g(X)$ et $h(Y)$ sont dans $\mathcal{L}^1$, alors $g(X)h(Y)$ est dans $\mathcal{L}^1$ et $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$.

Démonstration. La preuve du (1) est immédiate si on prend dans la définition les boréliens.
La preuve du (2) est facile si g et h sont des fonctions indicatrices (d'un intervalle cela suffit), puis si ce sont des fonctions étagées, et enfin si ce sont des fonctions positives. Ceci prouve que l'égalité est vrai pour $|g|$ et $|h|$, ce qui prouve que $g(X)h(Y)$ est dans $\mathcal{L}^1$. On obtient alors l'égalité voulue par différence ($g = g^+ - g^-$). $\square$

Une application directe est la théorème suivant

Théorème 3.2.1. *Soient X et Y deux va absolument continues, définies sur un même espace probabilisé $(\Omega, \mathcal{B}, P)$ et admettant chacune une espérance et une variance. Si de plus X et Y sont indépendantes, alors*
(1) $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
(2) $V(X+Y) = V(X) + V(Y)$.

la géométrie à la rescousse

Parfois la géométrie permet de constater directement que deux va X et Y ne sont pas indépendantes. Par exemple prenons le vecteur (X,Y) qui suit la loi uniforme sur le disque unité $\mathbb{D}$ (de centre 0 et de rayon 1). Alors X et Y ne sont pas indépendantes. Prenons en effet deux petits intervalles du type $I = J =]1-\varepsilon, 1[$. Si ε est suffisament petit le carré $I \times J$ est en-dehors du disque $\mathbb{D}$ donc $P((X,Y) \in I \times J) = 0$. Mais la probabilité d'avoir $X \in I$ est (à une constante multiplicative près) la surface du disque dont la première coordonnée est dans I, surface non nulle ; de même $P(Y \in J)$ n'est pas nulle. Donc

$$0 = P(X \in I \cap Y \in J) \neq P(X \in I)P(Y \in J).$$

Applications

Calcul de la loi d'une somme de va indépendantes On se donne deux va réelles absolument continues, X et Y, de densite respective f et g et indépendantes. Le but est de calculer la densité de $X+Y$.
On commence par poser $Z = (X,Y)$ et $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ l'application définie par $\phi(x,y) = (x, x+y)$. L'application ϕ est un difféomorphisme de $\mathbb{R}^2$ de Jacobien en tout point $J(x,y) = 1$. En particulier elle est inversible et le Jacobien de son inverse est 1 également.
On cherche une fonction $h : \mathbb{R}^2 \to \mathbb{R}$ telle que

$$\mathbb{P}(\phi(Z) \in B) = \iint_B h(a,b)\,dadb. \tag{3.1}$$

En effectuant le changement de variable $(a,b)=\phi(u,v)$ on trouve

$$\begin{aligned}\mathbb{P}(\phi(Z)\in B) &= \mathbb{P}(Z\in\phi^{-1}(B))\\ &= \iint_{\phi^{-1}(B)} f(u)g(v)dudv\\ &= \iint_B f(a)g(b-a)dadb. \qquad (3.2)\end{aligned}$$

La loi de $X+Y$ est donc la seconde marginale de h et en rapprochant (3.2) et (3.1) on trouve que la loi de $X+Y$ a pour densité

$$f_{X+Y}(z)=\int_{-\infty}^{+\infty} f(a)g(z-a)da.$$

On retiendra

> Si X et Y sont deux va réelles absolument continues de densité respective f et g et indépendantes, la loi de $X+Y$ a pour densité
>
> $$f_{X+Y}(z)=\int_{-\infty}^{+\infty} f(a)g(z-a)da.$$

Somme de va normales La somme de deux va indépendantes qui suivent chacune une loi normale est encore un va qui suit une loi normale. Notons $X \rightsquigarrow \mathcal{N}(m,\sigma^2)$ et $Y \rightsquigarrow \mathcal{N}(m',s^2)$. Alors la densité de

$X+Y$ est

$$f(z)=\frac{1}{2\pi\sigma s}\int_{-\infty}^{+\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} e^{-\frac{(z-x-m')^2}{2s^2}} dx.$$

On regroupe les termes en x qui se mettent sous un début de carré.

On effectue alors le changement de variable $u=x\frac{\sqrt{\sigma^2+s^2}}{s\sigma}+C$ (avec la bonne constante C).

L'intégrale se calcule et tue un $\sqrt{2\pi}$, le changement de variable ayant apporté un $\sqrt{\sigma^2+s^2}$. On retiendra

> Si $X \rightsquigarrow \mathcal{N}(m,\sigma^2)$ et $Y \rightsquigarrow \mathcal{N}(m',s^2)$, et si X et Y sont indépendantes alors $X+Y$ suit la loi normale $\mathcal{N}(m+m',\sigma^2+s^2)$

3.2.2 Covariance

Définition 3.2.4. *Si X et Y sont deux va dans $\mathcal{L}^2$ on appelle covariance des deux va, le réel $Cov(X,Y)=\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]$, et coefficient de corrélation de ces va le nombre*

$$\rho=\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}.$$

Tout comme pour les va discrètes, ρ est dans l'intervalle $[-1, 1]$ (c'est une conséquence de Cauchy-Schwarz) et quantifie le lien entre les 2 va.
On remarquera que si X et Y sont deux va réelles, un simple calcul donne

$$\boxed{V(X+Y) = V(X) + V(Y) + 2Cov(X,Y).}$$

Définition 3.2.5. *Si $X = (X_1, \ldots, X_n)$ est un vecteur aléatoire, et si chaque X_i est dans $\mathcal{L}^2$, on appelle matrice des covariances de X la matrice $n \times n$ d'éléments $c_{ij} = Cov(X_i, X_j)$.*

Exemple
Pour un vecteur aléatoire (X, Y) de $\mathbb{R}^2$ la matrice des convariances sera la matrice

$$\begin{pmatrix} V(x) & Cov(X,Y) \\ Cov(X,Y) & V(Y) \end{pmatrix}.$$

3.2.3 Application : une preuve du théorème de Stone Weierstrass

Chapitre 4

Quelques théorèmes de limites.

4.1 Fonction caractéristique

4.1.1 Rappel sur les complexes

Dans toute cette sous-section a est un réel non nul. On rappelle que le nombre complexe e^{iat} désigne le complexe de module 1 $\cos(at) + i\sin(at)$. L'application $t \mapsto e^{iat}$ est dérivable de dérivée $t \mapsto iae^{iat}$. Elle admet aussi une primitive sur $\mathbb{R}$, $t \mapsto \frac{1}{ia}e^{iat}$. Ces résultats se prouvent par retour à l'écriture avec cos et sin. On les utilisera directement sans les redémontrer.

Ainsi pour une fonction f, le terme $\int_a^b e^{iax}f(x)dx$ vaut simplement $\int_a^b \cos(ax)f(x)dx+$ $i\int_a^b \sin(ax)f(x)dx$.

4.1.2 Définition

Définition 4.1.1. *Soit X une variable aléatoire définie sur un espace $(\Omega, \mathcal{B}, P)$. Si X est discrète, on pose $X(\Omega) = \{x_n\}$ et on appelle fonction caratéristique de X la fonction*

$$t \mapsto \phi_X(t) = \sum_n e^{itx_n}P(x_n) = \mathbb{E}[e^{itX}].$$

Si X est absolument continue de densité f, on appelle fonction caratéristique de X la fonction

$$t \mapsto \phi_X(t) = \int_{-\infty}^{+\infty} e^{itx}f(x)dx = \mathbb{E}[e^{itX}].$$

Il convient dans les 2 cas de vérifier que la fonction est bien définie. On remarque que les 2 sommes sont absolument convergentes donc convergentes.

Théorème 4.1.1. *Soient X et Y deux va définies sur un même espace $(\Omega, \mathcal{B}, P)$. Si les fonctions caractéristiques de X et Y coïncident, alors X et Y ont la même loi.*

Démonstration. Nous n'allons pas donner la preuve de ce résultat. L'idée peut se comprendre avec la formule avec l'intégrale. Ce que le théorème explique c'est que si f et g sont deux fonctions telles que pour tout t,

$$\int \cos(tx)f(x)dx = \int \cos(tx)g(x)dx,$$

alors $f = g$. $\square$

Exemples pour les loi usuelles Si X est une variable de Bernoulli de paramètre p alors $\phi_X(t) = q + pe^{it}$.
Si X suit la loi binomiale $B(n,p)$ alors $\phi_X(t) = \displaystyle\sum_{k=0}^{n} C_n^k p^k e^{itk} q^{n-k} = (q + pe^{it})^n$. Si X suit la loi de Poisson de paramètre λ, alors $\phi_X(t) = \displaystyle\sum_n e^{-\lambda}\frac{\lambda^n}{n!}e^{itn}$. Si X suit la loi uniforme sur $[a,b]$, alors $\phi_X(t) = \dfrac{1}{it}\dfrac{e^{itb} - e^{ita}}{b-a}$. Si X suit la loi de Cauchy sur $\mathbb{R}$, alors $\phi_X(t) = e^{-|t|}$. Si X suit la loi $\mathcal{N}(0,1)$ sur $\mathbb{R}$, alors
$dispphi_X(t) = e^{-\frac{1}{2}t^2}$. Si X suit la loi exponentielle de paramètre λ, alors

$$\phi_X(t) = \frac{\lambda}{\lambda - it}.$$

Proposition 4.1.2. *Si a et b sont deux réels, on a*

$$\phi_{aX+b}(t) = e^{itb}.\phi_X(at).$$

On en conclut que si X suit la loi $\mathcal{N}(m,\sigma^2)$ alors la fonction caractéristique de X est

$$\phi_X(t) = e^{it\sigma m}e^{-\frac{1}{2}\sigma^2 t^2}.$$

4.1.3 Propriétés

Proposition 4.1.3. *Soit X une va de fonction caractéristique ϕ_X. Alors*

1. $\phi_X(0) = 1$.
2. ϕ_X *est continue.*
3. *Si X est dans $\mathcal{L}^1$, alors ϕ_X est dérivable.*
4. *Si X est dans $\mathcal{L}^2$, alors ϕ_X est 2 fois dérivable.*

Démonstration. $\phi_X(0) = \mathbb{E}[e^{i0.X}] = \mathbb{E}[1] = 1$.
Considérons que X ait pour densité f. Soient t et h 2 réels, avec h proche de 0. Un calcul donne

$$|\phi_X(t) - \phi_X(t+h)| \leq \int_{\mathbb{R}} 2|\sin(htx/2)|f(x)dx.$$

en découpant cette intégrale en 2 parties, une sur $[-a, a]$, l'autre sur le reste de $\mathbb{R}$ on montre qu'elle tend vers 0 si h tend vers 0. L'intégrale sur $[-a, a]$ est en effet petite car x est borné et $\sin(0) = 0$. l'autre partie est petite car f est très petite (et F_X converge en $\pm\infty$).
Si X est dans $\mathcal{L}^1$ alors $\displaystyle\int xf(x)dx$ converge. De plus $\dfrac{\partial e^{itx}}{\partial x} f(x) = ixf(x)$, donc on peut utiliser le théorème de dérivation sous le signe $\displaystyle\int$ pour conclure que

$$\phi'_X(t) = \int_{\mathbb{R}} \frac{\partial e^{itx}}{\partial x} f(x)dx.$$

La même démonstration marche pour la dérivée seconde. □

Tout comme pour la fonction génératrice d'une variable discrète à valeurs entières, la fonction caractéristique permet de retrouver la moyenne et la variance d'une variable, lorsqu'elles existent. Mais son intérêt principal est qu'elle transforme le produit de convolution en un produit :

Proposition 4.1.4. *Si X et Y sont deux va indépendantes, alors ϕ_{X+Y} vaut $\phi_X \times \phi_Y$.*

Démonstration. Il suffit de remarquer que pour tout t,

$$\phi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX}e^{itY}] = \mathbb{E}[e^{itX}]\mathbb{E}[e^{itY}],$$

puisque X et Y sont indépendantes. □

Attention la réciproque est fausse. En effet, si X suit la loi de Cauchy sa fonction caractéristique est donnée par $\phi_X(t) = e^{-|t|}$. De plus X n'est pas indépendant d'elle-même ; posons $Y = \frac{X+X}{2}$. Nous avons $\phi_Y(t) = \phi_X(t) = \phi_{X/2}(t) \times \phi_{X/2}(t)$.

4.1.4 Un exemple d'application

Nous terminons par un exemple d'application des propriétés de la fonction caractéristique. Supposons qu'une va X soit telle que sa fonction caractéristique soit périodique de période T. Alors on va "montrer" que X est discrète :
Nous avons donc $\phi(T) = 1$, ce qui s'écrit

$$\int_{-\infty}^{+\infty} \cos(Tx)f(x)dx + i\int_{-\infty}^{+\infty} \sin(Tx)f(x)dx = \int_{-\infty}^{+\infty} f(x)dx. \tag{4.1}$$

Ceci ne peut être valable que si la partie imaginaire est nulle et si nous avons

$$\int_{-\infty}^{+\infty} (\cos(Tx) - 1)f(x)dx = 0. \tag{4.2}$$

la fonction dans l'intégrale de 4.2 est négative et d'intégrale nulle ; elle est donc nulle (presque sûrement). Ceci montre que la "fonction" f est nulle partout sauf sur $\frac{2\pi}{T}\mathbb{Z}$. Donc la variable est discrète.

4.2 Loi faible des grands nombres

4.2.1 Inégalité de Markov

Proposition 4.2.1. *Soit X une va discrète ou absolument continue. Alors pour tout $k > 0$,*

$$P(|X| > k) \leq \frac{1}{k^2}\mathbb{E}[X^2].$$

Démonstration. Si X est discrète. On note $\{x_i, i \in \mathbb{N}\}$ les valeurs prises par X et p_i la distribution de probabilité $p_i = P(X = x_i)$ sur $X(\Omega)$. Fixons $k > 0$ et notons $J = \{i, |x_i| > k\}$. Alors

$$\begin{aligned} P(|X| > k) &= \sum_{i \in J} p_i \\ &\leq \sum_{i \in J} p_i \frac{x_i^2}{k^2} \\ &\leq \sum_{i \in \mathbb{N}} p_i \frac{x_i^2}{k^2} = \frac{1}{k^2}\mathbb{E}[X^2]. \end{aligned}$$

Si X est absolument continue. Soit f sa densité. Alors

$$\begin{aligned} P(|X| > k) &= P(X > k) + P(X < -k) \\ &= \int_k^{+\infty} f(t)dt + \int_{-\infty}^{-k} f(t)dt \\ &\leq \int_k^{+\infty} \frac{x^2}{k^2} f(t)dt + \int_{-\infty}^{-k} \frac{x^2}{k^2} f(t)dt \\ &\leq \frac{1}{k^2}\mathbb{E}[X^2]. \end{aligned}$$

□

4.2.2 Inégalité de Bienaymé-Tchébycheff

Théorème 4.2.1 (Bienaymé-Tchébycheff)**.** *Soit X une va discrète ou absolument continue admettant une espérance $\underline{X}$. Alors pour tout $k > 0$*

$$P(|X - \underline{X}| > k) \leq \frac{V(X)}{k^2}.$$

Démonstration. On applique l'inégalité de Markov à $X - \underline{X}$. Bien sur cette inégalité n'a de l'intérêt que si $V(X)$ est finie. □

L'inégalité de Bienaymé-Tchébycheff donne

$$P(\underline{X} - t\sigma \leq X \leq \underline{X} + t\sigma) \geq 1 - \frac{1}{t^2},$$

et cela sans avoir besoin de connaitre la loi de X. En particulier on trouve

$$P(\underline{X} - 2\sigma \leq X \leq \underline{X} + 2\sigma) \geq 75\%,$$

$$P(\underline{X} - 4\sigma \leq X \leq \underline{X} + 4\sigma) \geq 93,75\%.$$

Application à l'analyse

On veut montrer que $\int_0^t e^{-x^2/2}dx \geq \sqrt{\frac{\pi}{2}}(1 - \frac{1}{t^2})$.

On note X une va suivant la loi $\mathcal{N}(0,1)$. Alors $P(|X| < t) \geq 1 - \dfrac{1}{t^2}$ devient

$$2\frac{1}{\sqrt{2\pi}}\int_0^t e^{-x^2/2}dx \geq 1 - \frac{1}{t^2},$$

ce qui s'écrit aussi $\int_0^t e^{-x^2/2}dx \geq \sqrt{\frac{\pi}{2}}(1 - \frac{1}{t^2})$.

4.2.3 Loi faible des grands nombres

Théorème 4.2.2. *Soient $X_1, X_2, \ldots, X_n, \ldots$ une suite de variables aléatoires discrètes ou absolument continues, indépendantes et de même loi (admettant donc une même espérance m et une même variance σ^2). Alors*

$$\forall h > 0 \quad \lim_{n\to+\infty} P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - m\right| \leq h\right) = 1.$$

Démonstration. On note $S_n = X_1 + X_2 + \ldots + X_n$. Par linéarité $\mathbb{E}[S_n] = nm$ et par indépendance $V(S_n) = n\sigma^2$. On pose $Y_n = \frac{1}{n}S_n$, et on lui applique l'inégalité de Bienaymé-Tchébycheff.

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - m\right| > h\right) &\leq \frac{1}{h^2}V(\frac{S_n}{n}) \\ &\leq \frac{\sigma^2}{nh^2} \\ \Longrightarrow P\left(\left|\frac{S_n}{n} - m\right| \leq h\right) &\geq 1 - \frac{\sigma^2}{nh^2}. \end{aligned}$$

□

Exemple

On joue à pile ou face avec une probabilité p de faire face. On pose

$$X_i = \begin{cases} 0 \text{ si le } i\text{ème jet est pile;} \\ 1 \text{ si le } i\text{ème jet est face.} \end{cases}$$

Les X_i sont des variables de Bernoulli toutes indépendantes, de même espérance p et de variance $pq \stackrel{def}{=} \sigma^2$. La va $S_n = X_1 + \ldots + X_n$ suit donc la loi binômiale $B(n,p)$, et elle représenta la fréquence ϕ_n d'apparition de "face" en n jets.
On trouve alors $\lim_{n\to+\infty} P(|\phi_n - p| \leq h) = 1$, et même on a

$$P(|\phi_n - p| \leq h) \geq 1 - \frac{\sigma^2}{nh^2} \leq 1 - \frac{1}{4nh^2},$$

car $pq \leq 0,25$.
Si $p = q = 0,5$ et $h = 10^{-2}$, il faut 50000 coups pour être sûr à 95% que la fréquence d'apparition de face diffère de moins de 10^{-2} de $\frac{1}{2}$.

Application à l'analyse

Si S_n suit la loi $B(n,p)$ et si h est strictement positif, alors

$$P(|\frac{S_n}{n} - p| > h) = \sum_{k, |\frac{k}{n} - p| > h} C_n^k p^k (1-p)^{n-k}.$$

Théorème 4.2.3 (Weierstrass)**.** *Soit $f : [0,1] \to \mathbb{R}$ continue. Pour $n \geq 0$ on pose*

$$B_n(f)(x) = \sum_{k=0}^{n} C_n^k f(\frac{k}{n}) x^k (1-x)^{n-k}.$$

Alors la suite de polynômes $(B_n(f))$ converge uniformément vers f sur $[0,1]$.

Démonstration. Si x est dans $[0,1]$ on a $f(x) - B_n(f)(x) = \sum_{k=0}^{n} C_n^k (f(x) - f(\frac{k}{n})) x^k (1-x)^{n-k}$. comme f est continue sur $[0,1]$ elle est bornée par M et uniformément continue. Pour $\varepsilon > 0$ fixé, il existe un $\delta > 0$ tel que

$$|x - y| \leq \delta \Longrightarrow |f(x) - f(y)| \leq \varepsilon.$$

On trouve alors

$$\begin{aligned} |f(x) - B_n(f)(x)| &\leq \sum_{k, |\frac{k}{n} - p| \leq \delta} C_n^k |f(x) - f(\frac{k}{n})| x^k (1-x)^{n-k} \\ &\quad + \sum_{k, |\frac{k}{n} - p| > \delta} C_n^k |f(x) - f(\frac{k}{n})| x^k (1-x)^{n-k} \\ &\leq \varepsilon \sum_{k=0}^{n} C_n^k x^k (1-x)^{n-k} + 2M \frac{1}{4n\delta^2}. \end{aligned}$$

Si n est suffisament grand, le dernier majorant est plus petit que 2ε, et ce indépendament de x.

□

4.3 Théorème de la limite centrée

4.3.1 Le TCL

Théorème 4.3.1. *Soit (X_n) une suite de variables aléatoires discrètes ou absolument continues, indépendantes et suivant toutes la même loi. On note m leur espérance commune, σ^2 leur variance et $S_n = X_1 + \ldots + X_n$. Alors*

$$\lim_{n\to+\infty} P\left(a \leq \frac{S_n - nm}{\sigma\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Démonstration. Nous commençons par rappeler que $(1 + \frac{x}{n})^n$ converge vers e^x lorsque n tend vers $+\infty$.

Posons $Y_n = \dfrac{S_n - nm}{\sigma\sqrt{n}}$. Alors la règle de calcul pour les fonctions caractéristiques et l'indépendance des X_i donne

$$\phi_{Y_n}(t) = (\phi_X(\frac{t}{\sigma\sqrt{n}})e^{-i\frac{tm}{\sigma\sqrt{n}}})^n.$$

Pour t fixé, $\frac{t}{\sigma\sqrt{n}}$ tend vers 0 lorsque n tend vers $+\infty$, et nous allons faire un développement limité de $\phi_{Y_n}(t)$.

Nous avons $\phi_X(\dfrac{t}{\sigma\sqrt{n}}) = 1 + \dfrac{t}{\sigma\sqrt{n}}\phi'(0) + \dfrac{t^2}{2n\sigma^2}\phi''(\theta/n)$ par la formule de Taylor.

De même nous avons $e^{-i\frac{tm}{\sigma\sqrt{n}}} = 1 - i\dfrac{tm}{\sigma\sqrt{n}} - \dfrac{t^2m^2}{2n\sigma^2} + o(1/n)$.

de plus $\phi'(0) = im$ et $\phi''(0) = -\mathbb{E}[X^2]$. On trouve donc

$$\phi_{Y_n}(t) = (1 + \frac{t^2}{2n\sigma^2}(\phi''(\theta/\sqrt{n}) + m^2) + o(1/n))^n.$$

Si n tend vers $+\infty$, cette quantité converge vers $e^{-\frac{t^2}{2}}$, qui est bien la fonction caractéristique de la loi normale centrée réduite. □

Remarques :

1. La limite ne dépend pas de la loi des X_i.
2. $\mathbb{E}[S_n] = nm$ et $V(S_n) = n\sigma^2$, donc la va $T_n \stackrel{def}{=} \dfrac{S_n - nm}{\sigma\sqrt{n}}$ est centrée et réduite. Le théorème dit que quand n devient grand, la loi de T_n "tend" vers $\mathcal{N}(0,1)$.
3. Si n est grand, il est difficile d'avoir la loi exacte de $S_n = X_1 + \ldots + X_n$. On peut approximer T_n par $\mathcal{N}(0,1)$.

Exemple

Une machine à calculer fait une erreur d'au plus $\pm\varepsilon$ à chaque opération. Quelle est la probabilité pour que l'erreur soit de moins de 1 au bout de 200 opérations ?

On pose X_i l'erreur commise au $i$ème calcul. Les X_i sont supposés indépendants et suivent la loi uniforme sur $[-\varepsilon, \varepsilon]$. On a $\mathbb{E}[X_i] = 0$ et $V(X_i) = \varepsilon^2/3$. On trouve donc

$$P(-1 \leq S \leq 1) = 2\pi(\frac{\sqrt{3}}{\sqrt{200}\varepsilon}) - 1.$$

Si $\varepsilon = 0,1$, alors la probabilité est d'environ $2\pi(0,3873) - 1$ soit environ 30% de certitude.

4.3.2 Approximation de la loi binomiale

Théorème 4.3.2. *Soit* (S_n) *une suite de va chacune suivant la loi* $B(n, p)$*. Alors*

$$\lim_{n\to+\infty} P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Démonstration. S_n peut s'écrire comme étant $X_1 + \ldots + X_n$, où les X_i sont indépendantes de Bernoulli de paramètre p. On applique le TCL. □

Applications

Au jeu de pile ou face, on note ϕ_n la fréquence d'apparition de face. Le théorème précédent dit que

$$\lim_{n\to+\infty} P\left(a \leq \frac{\phi_n - p}{\sqrt{pq/n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Si $p = q = 0,5$, combien faut-il de coups pour être sûr à 95% que la fréquence diffère de 0,5 de moins de 10^{-2} ?
On veut $P(-10^{-2} \leq \phi_n - 1/2 \leq 10^{-2}) \geq 0,95$. Or

$$P\left(-a \leq \frac{\phi_n - 1/2}{\sqrt{1/4n}} \leq a\right) \sim \frac{1}{\sqrt{2\pi}} \int_{-a}^a e^{-x^2/2} dx = 2\pi(a) - 1.$$

On cherche d'abord a pour avoir $2\pi(a) - 1 \geq 0,95$, ce qui donne $a = 1,96$. On cherche ensuite n pour que $\frac{0,98}{\sqrt{n}} \leq 10^{-2}$, soit $n = 9604$ coups.

Une association comporte 100 membres. Chacun utilise le téléphone en moyenne 10 minutes par heure. Quel nombre minimum de lignes l'association doit-elle posséder pour qu'à un instant donné, la probabilité de saturation des lignes soit inférieure à 5% ?
On fixe l'instant t_0, et on pose $X_k = \begin{cases} 1 \text{ si le membre } k \text{ téléphone à l'instant } t_0 \\ 0 \text{ sinon} \end{cases}$. Il y a donc $S = \sum_{k=1}^{100} X_k$ personnes qui téléphonent à l'instant t_0. On cherche le nombre N de lignes tel que $P(S > N) \leq 5\%$.
De plus $p \overset{def}{=} P(X_k = 1)$ vaut 1/6=10/60 et $q = 1 - p$ vaut 5/6. De plus

$$S > N \iff \frac{S - 100p}{\sqrt{100pq}} > \frac{N - 100p}{\sqrt{100pq}},$$

ce qui revient à chercher N tel que

$$1 - \pi(\frac{N - 100p}{\sqrt{100pq}}) \leq 0,05.$$

Cela donne au final $N = 23$.

4.3.3 Approximation de la loi de Poisson

Théorème 4.3.3. *Soit (X_λ) une famille de va suivant chacune $P(\lambda)$. Alors*

$$\lim_{\lambda \to +\infty} P\left(a \leq \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Si λ est grand, calculer $P(\lambda + a\sqrt{\lambda} \leq X \leq \lambda + b\sqrt{\lambda})$ demande beaucoup de calculs (de l'ordre de $O(\sqrt{\lambda})$), alors que l'approximation se calcule plus facilement avec les tables.

Exemple
Le nombre d'urgences traitées par jour dans un hôpital suit la loi de Poisson de paramètre 2.

1. Probabilité de traiter 70 urgences par mois ?
   Si X est le nombre mensuel d'urgence on a $X \rightsquigarrow P(2 \times 30) = P(60)$. De plus $X > 70$ donne $\frac{X-60}{\sqrt{60}} > \frac{10}{\sqrt{60}}$, ce qui nous amène à calculer $1 - \pi(\frac{10}{\sqrt{60}})$. On trouve au final une probabilité de 9,85%.
2. Probabilité de traiter moins de 55 urgences par mois ?
   on se ramène à $1 - \pi(\frac{5}{\sqrt{60}})$, soit 25,78%.

4.3.4 Application : Intervalle de confiance et test de la moyenne

On jette une pièce 10000 fois, et on trouve 4810 piles et 5190 faces. On veut savoir si la pièce est truquée ou non, en ayant une certitude de plus de 95%.
On note p la probabilité d'obtenir pile et S_n le nombre de piles en n lancers ; S_n suit la loi $B(n, p)$. On note enfin ϕ_n la fréquence observée d'apparition de piles. Par approximation de la loi binômiale, si n est grand,on peut approximer

$$P(\left|\frac{S_n - np}{\sqrt{npq}}\right| < \alpha) \sim \frac{1}{\sqrt{2\pi}} \int_{-\alpha}^{\alpha} e^{-x^2/2} dx.$$

On veut une certitude à 5% près, ce qui donne $2\pi(\alpha) - 1 = 0,95$, soit $\alpha = 1,96$. De plus $pq < 1/4$ et donc

$$\{\left|\frac{S_n - np}{\sqrt{npq}}\right| < \alpha\} \subset \{\phi_n - \frac{\alpha}{2\sqrt{n}} \leq p \leq \phi_n + \frac{\alpha}{2\sqrt{n}}\}.$$

Cela donne donc avec $\alpha = 1,96$,

$$P(\phi_n - \frac{1,96}{2\sqrt{n}} \leq p \leq \phi_n + \frac{1,96}{2\sqrt{n}}) \geq 0,95.$$

Avec les valeurs numériques on trouve une probabilité de 95% pour que $0,4712 \leq p \leq 0,4908$, ce qui nous permet de dire qu'avec 95% de chances, la pièce est truquée. En effet $\{\text{la piece est truquée}\} = \{p \neq 0,5\}$ qui contient l'ensemble $\{0,4712 \leq p \leq 0,4908\}$. Donc

$$P(\text{la pièce est truquée}) \geq P(0,4712 \leq p \leq 0,4908) \geq 0,95.$$

Remarque 1. Supposons que l'on trouve 4910 piles et 5090 faces. Le même calcul donne $P(0,4812 \leq p \leq 0,5008) \geq 0,95$. Ceci ne permet pas pour autant de conclure quant au trucage éventuel de la pièce.
$\frac{1,96}{2} = 0,92 < 1$, ce qui donne $\{\phi_n - \frac{1,96}{2\sqrt{n}} \leq p \leq \phi_n + \frac{1,96}{2\sqrt{n}}\} \subset \{\phi_n - \frac{1}{\sqrt{n}} \leq p \leq \phi_n + \frac{1}{\sqrt{n}}\}$.
On retiendra :

$$P(\phi_n - \frac{1}{\sqrt{n}} \leq p \leq \phi_n + \frac{1}{\sqrt{n}}) \geq 0,95.$$

4.3.5 Récapitulation des approximations usuelles