Probabilités - Cours

Département Informatique IUT Montpellier

Année 2022-2023

Table des matières

1 Le langage des Probabilités **1**

1 Introduction . . . 1

2 Événements . . . 2

3 Lien entre probabilités et statistique . . . 3

4 Probabilités . . . 4

4.1 Définition . . . 4

4.2 L'équiprobabilité . . . 4

5 Probabilités conditionnelles . . . 4

5.1 Définition . . . 4

5.2 Quelques formules . . . 5

5.3 Evénements indépendants . . . 6

6 Exemples . . . 6

2 Variables aléatoires discrètes **9**

1 Définition et vocabulaire . . . 9

1.1 Exemple . . . 9

1.2 Définition d'une variable aléatoire . . . 10

2 Fonction de répartition d'une variable réelle . . . 11

3 Moyenne et variance d'une variable réelle . . . 12

4 Couple aléatoire . . . 13

5 Variables aléatoires indépendantes . . . 14

6 Propriétés de l'espérance et de la variance . . . 15

6.1 Transformation affine . . . 15

6.2 Variance d'une somme . . . 15

7 Lois discrètes usuelles . . . 16

7.1 Loi de Bernoulli . . . 16

7.2 Loi uniforme sur un ensemble fini de réels . . . 17

7.3 Loi binomiale . . . 17

7.4 Loi hypergéométrique . . . 19

7.5 Loi géométrique 20
7.6 Loi de Poisson 21

3 Variables aléatoires continues **23**
1 Généralités 23
2 Fonction de répartition 25
3 Espérance et variance 27
4 Une loi à densité usuelle : la loi normale 29
4.1 Densité de la loi normale centrée réduite $\mathbb{N}(0, 1)$ 29
4.2 Loi normale $\mathbb{N}(\mu, \sigma)$ 33
4.3 Somme de variables aléatoires normales 35

4 Théorème Central Limite **37**
1 Somme de variables aléatoires i.i.d. 37
2 TCL pour la somme de variables aléatoires i.i.d. 39
3 Applications : approximation de la loi binomiale 40
4 TCL pour la moyenne de variables aléatoires i.i.d. 42

Chapitre 1

Le langage des Probabilités

1 Introduction

La théorie des probabilités fournit des modèles mathématiques permettant l'étude d'expériences dont le résultat ne peut être prévu avec une totale certitude. En voici quelques exemples :

Expérience	**Résultat observable**
Lancer d'un dé	Un entier $k \in \{1, ..., 6\}$
Prélèvement de n objets en sortie d'une chaîne de production	Nombre d'objets défectueux dans l'échantillon
Questionnaire à 100 questions binaires	Suite ω de 100 réponses, $\omega \in \{\text{oui, non}\}^{100}$
Lancer d'une pièce jusqu'à la première obtention de pile	Un entier $k \in \mathbb{N}$: le temps d'attente du premier succès
Lancer d'une fléchette sur une cible	Point d'impact

Bien que le résultat de chacune de ces expériences soit imprévisible, l'observation et l'intuition nous amènent à penser que ces phénomènes obéissent à certaines lois. Par exemple si on jette 12000 fois le dé, on s'attend à ce que le nombre d'apparitions du numéro «2» soit voisin de 2000.

La *théorie des probabilités* permet de donner un sens précis à ces considérations. La *statistique* permet de confronter les modèles probabilistes avec la réalité observée afin de les valider ou de les invalider. Par exemple, si quelqu'un a 40 bonnes réponses sur 100 questions binaires, est-il légitime de considérer qu'il a « mieux fait » que le hasard ?

2 Événements

La théorie des probabilités utilise le langage des ensembles pour modéliser une expérience aléatoire. Nous noterons Ω un ensemble dont les éléments représentent tous les résultats possibles ou **événements élémentaires** d'une expérience aléatoire donnée. Cet ensemble est appelé l'**univers** des possibles. Les **événements** sont représentés par des parties de Ω.

Exemple 1.1
Si l'on jette un dé, l'événement

$$A : \text{« obtention d'un chiffre pair »}$$

n'est pas élémentaire. Il est composé des trois événements élémentaires $\{2\}$, $\{4\}$, $\{6\}$:

$$A = \{2,\, 4,\, 6\}, \quad \text{avec de plus } \Omega = \{1,\, 2,\, 3,\, 4,\, 5,\, 6\}\ .$$

Exemple 1.2
Si on lance trois fois une pièce de monnaie, les événements élémentaires sont des triplets comme $(f,\, f,\, p)$ indiquant le résultat précis de chacun des trois lancers.

$$\Omega = \{p,\, f\}^3\ .$$

L'événement B : « obtention de face au premier des trois lancers » est composé :

$$B = \{(f,\, f,\, p),\, (f,\, f,\, f),\, (f,\, p,\, f),\, (f,\, p,\, p)\}\ .$$

Avec ce mode de représentation, les opérations logiques sur les événements : «et», «ou», « négation » se traduisent par des opérations ensemblistes : intersection, réunion, complémentaire. Voici un tableau de correspondance entre les deux langages.

Notations	Vocabulaire ensembliste	Vocabulaire probabiliste
$\emptyset$	ensemble vide	événement impossible
Ω	ensemble plein	événement certain
ω	élément de Ω	événement élémentaire
A	sous-ensemble de Ω	événement
$A \subset B$	A inclus dans B	A implique B
$A \cup B$	réunion de A et B	A ou B
$A \cap B$	intersection de A et B	A et B
$\bar{A}$	complémentaire de A dans Ω	événement contraire de A
$A \cap B = \emptyset$	A et B sont disjoints	A et B sont incompatibles

3 Lien entre probabilités et statistique

De nombreux programmes d'ordinateurs utilisent des nombres "au hasard". Nous nous intéressons ici aux programmes de simulation où l'on essaie de recréer pour les étudier des situations faisant intervenir le hasard.
Pour mieux appréhender le hasard, on peut commencer par réaliser un grand nombre de simulations d'une même expérience aléatoire. Simuler n fois la même expérience aléatoire permet d'obtenir un échantillon de taille n.

Définition 1.1 (Fréquence d'un événement)
Soit A un événement élémentaire associé à une expérience aléatoire. Si dans un échantillon de taille n, l'événement A est apparu m fois, alors la fréquence de cet événement dans l'échantillon est : $f(A) = \frac{m}{n}$.
Si A est un événement, la fréquence de A est égale à la somme des fréqquences élémentaires qui le composent.

$$f(A) = \frac{\textit{nombre d'apparition de } A}{n}$$

Proposition 1.1 (Propriétés des fréquences)
i) $f(\overline{A}) = 1 - f(A)$;
ii) Si A et B sont incompatibles, $f(A \cup B) = f(A) + f(B)$;
iii) Si A et B sont quelconques, $f(A \cup B) = f(A) + f(B) - f(A \cap B)$.

Proposition 1.2 (Interprétation simplifiée de la loi des grands nombres)
Soit A un événement associé à une expérience aléatoire. Si on simule cette expérience, alors plus la taille de l'échantillon est importante, plus la fréquence d'apparition de A va se stabiliser. Cette valeur autour de laquelle la fréquence d'apparition se stabilise est dite fréquence théorique.

Exemple 1.3
Si on lance un dé, la fréquence théorique d'apparition du 1 est de $\frac{1}{6}$ soit environ 16.67 % de chances d'obtenir un 1. Voici un exemple de simulations du lancer d'un dé. Pour chaque échantillon, on relève la fréquence d'apparition du 1. La simulation a été effectuée avec 20, 100, 200 , 1000 puis 5000 lancers.

Nombre de lancers	20	100	200	1000	5000
Fréquence	0.25	0.181	0.155	0.1666	0.1669

La fréquence d'apparition du 1, a tendance à se stabiliser autour de la valeur 0.1667.

La notion de probabilité généralise et prolonge le concept de fréquence théorique d'apparition.

4 Probabilités

4.1 Définition

Définition 1.2
Soit Ω un ensemble et F l'ensemble des parties de Ω. On appelle ***probabilité*** *sur (Ω, F), toute application $\mathbf{P}$ de F dans $[0,1]$ vérifiant :*
i) $\mathbf{P}(\Omega) = 1$
ii) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$ *si A et B sont deux événements incompatibles.*
Le triplet $(\Omega,\ F,\ \mathbf{P})$ s'appelle ***espace probabilisé****.*

La probabilité hérite des propriétés des fréquences.

Proposition 1.3 (Propriétés générales d'une probabilité)
Toute probabilité $\mathbf{P}$ sur $(\Omega,\ F)$ vérifie les propriétés suivantes :

1. $\mathbf{P}(\emptyset) = 0$;
2. $\forall\, A \in F,\ \mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$;
3. $\forall\, A \in F,\ \forall\, B \in F,\ A \subset B \Rightarrow \mathbf{P}(A) \leq \mathbf{P}(B)$;
4. $\forall\, A \in F,\ \forall\, B \in F,\ \mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

4.2 L'équiprobabilité

Définition 1.3
Dans un univers contenant N événements élémentaires, on dit qu'il y a équiprobabilité lorsque tous les événements élémentaires ont la même probabilité de se produire. Cette probabilité vaut alors $\frac{1}{N}$.

Proposition 1.4
Dans le cas d'équiprobabilité, pour tout événement A , on a

$$\mathbf{P}(A) = \frac{\textit{nombre de cas favorables}}{\textit{nombre de cas possibles}}$$

5 Probabilités conditionnelles

5.1 Définition

On considère un espace probabilisé $\Omega, P(\Omega), P)$ et un ensemble B tel que $P(B) > 0$. Pour $A \in P(\Omega)$, on appelle probabilité conditionnelle de A sachant B ou probabilité de A conditionnellement à B, le nombre

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

L'application

$$\mathbf{P}(.|B)\ :\ P(\Omega) \to\ [0,1]$$

$$A \longmapsto \mathbf{P}(A|B)$$

est une nouvelle probabilité que l'on appelle conditionnelle par rapport à B. Toutes les propriétés vues au paragraphe 4 sont donc valables pour $\mathbf{P}(A|B)$

5.2 Quelques formules

i) On déduit directement de la définition de probabilité conditionnelle, cette égalité :
$\mathbf{P}(A \cap B) = \mathbf{P}(A|B) \times \mathbf{P}(B)$.

On considère un espace probabilisé $\Omega, P(\Omega), P)$ et un événement A. Si (B_i) est une partition (finie ou dénombrable) de l'événement B.
ii) *Formule des probabilités totales* :

$$\mathbf{P}(A) = \sum_i \mathbf{P}(B_i)\mathbf{P}(A|B_i).$$

iii) On déduit des propriétés i) et ii), la *formule de Bayes* : Pour tout j,

$$\mathbf{P}(B_j|A) = \frac{\mathbf{P}(A|B_j) \times \mathbf{P}(B_j)}{\sum_i \mathbf{P}(B_i)\mathbf{P}(A|B_i)}.$$

Exemple 1.4
Parmi les étudiants de deuxième année de BUT, 25 % apprécient l'enseignement des mathématiques. La probabilité qu'un étudiant aimant cet enseignement ait la moyenne au partiel, est de 0,8. La probabilité qu'un étudiant n'aimant pas cet enseignement ait la moyenne au partiel, est de 0,1.
Quelle est la probabilité qu'un étudiant pris au hasard ait la moyenne au partiel ?
Soient les événements A : "l'étudiant a la moyenne au partiel " et B : "l'étudiant apprécie les maths".

$$\mathbf{P}(A) = \mathbf{P}(A \cap B)\ +\ \mathbf{P}(A \cap \bar{B})$$

$$\mathbf{P}(A) = \mathbf{P}(A|B) \times \mathbf{P}(B)\ +\ \mathbf{P}(A|\bar{B}) \times \mathbf{P}(\bar{B}).$$

D'où $\mathbf{P}(A) = \frac{8}{10} \times \frac{1}{4}\ +\ \frac{1}{10} \times \frac{3}{4}\ =\ \frac{11}{40}$.

Sachant que l'étudiant tiré au hasard a eu la moyenne au partiel, quelle est la probabilité qu'il apprécie les maths ?

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(A|B) \times \mathbf{P}(B)}{\mathbf{P}(A)} = \frac{8}{11}.$$

5.3 Evénements indépendants

Attention : La formule $\mathbf{P}(A \cap B) = \mathbf{P}(A) \times \mathbf{P}(B)$ est en général fausse sauf si les deux événements A et B sont indépendants.

Définition 1.4
Deux événements sont indépendants si

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

Les trois propriétés suivantes sont équivalentes :
i) $\mathbf{P}(A|B) = \mathbf{P}(A)$
ii) $\mathbf{P}(B|A) = \mathbf{P}(B)$
iii) $\mathbf{P}(A \cap B) = \mathbf{P}(A) \times \mathbf{P}(B)$.

Exemple 1.5
On jette deux fois le même dé. Les événements

$$A = \{ \text{ obtenir un chiffre pair au premier lancer } \}$$
$$B = \{ \text{ obtenir le chiffre 6 au deuxième lancer } \}$$

sont indépendants.

En effet, en prenant $\Omega = \{1, 2, 3, 4, 5, 6\}^2$, $F = \mathcal{P}(\Omega)$ et $\mathbf{P}$ l'équiprobabilité, on a :

$$\mathbf{P}(A) = \frac{3 \times 6}{36} = \frac{1}{2}, \qquad \mathbf{P}(B) = \frac{6 \times 1}{36} = \frac{1}{6},$$
$$\mathbf{P}(A \cap B) = \frac{3 \times 1}{36} = \frac{1}{12}, \qquad \mathbf{P}(A)\,\mathbf{P}(B) = \frac{1}{2} \times \frac{1}{6}\ .$$

6 Exemples

Exemple 1.6
On effectue une partie de pile ou face en trois coups. Quelle est la probabilité d'obtenir pile aux premier et troisième lancers ?

On peut modéliser cette expérience en prenant $\Omega = \{f,\, p\}^3$ et pour famille d'événements observables $F = \mathcal{P}(\Omega)$ l'ensemble de toutes les parties de Ω. La pièce étant supposée équilibrée, aucun des 8 triplets de résultats possibles n'est favorisé ou défavorisé par rapport aux autres. Nous choisissons donc $\mathbf{P}$ de sorte que tous les événements élémentaires aient même probabilité (hypothèse d'équiprobabilité), soit :

$$\forall\, \omega \in \Omega, \quad \mathbf{P}(\{\omega\}) = \frac{1}{\mathrm{Card}(\Omega)} = \frac{1}{2^3}\ .$$

L'événement B dont on veut calculer la probabilité s'écrit :

$$B = \{(p,\, f,\, p); (p,\, p,\, p)\}\ .$$

D'où :

$$\mathbf{P}(B) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}\ .$$

Exemple 1.7

On fait remplir un questionnaire à 10 questions binaires. Quelle est la probabilité qu'un candidat répondant au hasard obtienne au moins 8 bonnes réponses ?

On choisit ici :

$$\Omega = \{\text{oui, non}\}^{10}, \qquad F = \mathcal{P}(\Omega)\ .$$

Si le candidat répond complètement au hasard, on peut considérer que chacune des 2^{10} grilles de réponses possibles a la même probabilité d'apparaître (hypothèse d'équiprobabilité sur Ω). Pour tout $B \subset \Omega$, on a alors :

$$\begin{aligned}\mathbf{P}(B) &= \frac{\mathrm{Card}(B)}{\mathrm{Card}(\Omega)} \\ &= \frac{\binom{10}{8} + \binom{10}{9} + \binom{10}{10}}{2^{10}} = \frac{56}{1024} = 0,0547\ .\end{aligned}$$

Exemple 1.8 (Contrôle de production)

On prélève au hasard un échantillon de 3 pièces dans une production totale de 10 pièces comprenant en tout 4 pièces défectueuses. Quelle est la probabilité de :

$$A_j = \{\text{il y a exactement } j \text{ pièces défectueuses dans l'échantillon}\}\ ?$$

On prend pour Ω l'ensemble de toutes les parties à 3 éléments d'un ensemble à 10 éléments (ensemble de tous les échantillons possibles de taille 3), $F = \mathcal{P}(\Omega)$ et $\mathbf{P}$ l'équiprobabilité sur Ω. Il suffit alors de dénombrer tous les échantillons ayant exactement j pièces défectueuses. Un tel échantillon se construit en prenant j pièces dans le sous–ensemble des défectueuses ($\binom{4}{j}$ choix possibles) et en complétant par $(3-j)$ pièces prises dans le sous–ensemble des non–défectueuses ($\binom{6}{3-j}$ choix possibles). On en déduit :

$$\mathbf{P}(A_j) = \begin{cases} \dfrac{\binom{4}{j}\binom{6}{3-j}}{\binom{10}{3}} & \text{pour } j \in \{0,\ 1,\ 2,\ 3\}, \\ 0 \quad \text{sinon} . \end{cases}$$

Chapitre 2

Variables aléatoires discrètes

1 Définition et vocabulaire

1.1 Exemple

Dans de nombreux jeux, on fait intervenir le hasard en observant la somme des points marqués par deux dés. Considérons le jet d'un dé blanc et d'un dé rouge et notons S la somme des points obtenus. On modélise cette expérience en prenant l'équiprobabilité sur l'univers

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

Un événement élémentaire ω est un couple $(i,\, j)$ où i est le résultat du dé blanc et j, le résultat du dé rouge.
Le tableau suivant résume tous les cas possibles. Pour tout $\omega = (i,\, j)$, on a $S(\omega) = i + j$.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

On a défini une application S de Ω dans l'ensemble des points possibles $\{2,\, 3,\, 4,\, 5,\, 6,\, 7,\, 8,\, 9,\, 10,\, 11,\, 12\}$. On dit que S est une **variable aléatoire** sur Ω. Ce qui nous intéresse ici n'est pas ω mais $S(\omega)$. On cherche donc à déterminer la probabilité que la somme des points soit égale à k pour tout k entier de 2 à 12.
En prenant l'équiprobabilité pour $\mathcal{P}(\Omega)$, on obtient :

k	2	3	4	5	6	7	8	9	10	11	12
$\mathbf{P}(S=k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

On considère donc un nouvel ensemble d'événements élémentaires noté

$$\Omega' = \{2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11,\ 12\}\ ,$$

muni de la probabilité $\mathbf{P}_S$ définie dans le tableau ci-dessus. Cette nouvelle probabilité s'appelle **loi de la variable** S.

1.2 Définition d'une variable aléatoire

Soient Ω et Ω' deux ensembles finis ou dénombrables et X une application de Ω vers Ω'. Lorsqu'on choisit une structure d'espace probabilisé discret $(\Omega,\ \mathcal{P}(\Omega),\ \mathbf{P})$, on dit que X est une variable aléatoire sur $(\Omega,\ \mathcal{P}(\Omega),\ \mathbf{P})$ et à valeurs dans Ω'. C'est alors la probabilité $\mathbf{P}$ qui est utilisée pour donner un sens aux probabilités d'événements de la forme «$X \in A$», où A est inclus dans Ω', ce qui permet de définir une nouvelle probabilité sur Ω' que l'on note $\mathbf{P}_X$.

$$\forall\, A \in \mathcal{P}(\Omega'), \qquad \mathbf{P}_X(A) = \mathbf{P}(X \in A) = \mathbf{P}\{\omega : \omega \in \Omega \quad \text{et} \quad X(\omega) \in A\}\ .$$

Une variable aléatoire peut donc être vue comme un couple formé d'une application X et d'une probabilité sur son ensemble de départ.
$(\Omega',\ \mathcal{P}(\Omega'),\ \mathbf{P}_X)$ est un nouvel espace probabilisé, construit à partir de $(\Omega,\ \mathcal{P}(\Omega),\ \mathbf{P})$ et de X. On dit que la loi $\mathbf{P}_X$ est la loi de probabilité de X.

Définition 2.1 (Ensemble discret)
*Un ensemble E est dit **discret** s'il est possible de décrire tous ses éléments en les indexant (c'est-à-dire en les numérotant). En général, l'ensemble sera soit fini, soit l'ensemble des entiers naturels $\mathbb{N}$.*

Définition 2.2 (Variables aléatoire discrète)
*Soit un espace probabilisé d'espace fondamental Ω et de mesure de probabilité $\mathbf{P}$ un ensemble discret. On appelle **variable aléatoire** sur cet espace, toute application de Ω dans $\mathbb{R}$ telle que :*

$$\Omega \to \mathbb{R}$$

$$\omega \to X(\omega)$$

A chaque événement élémentaire ω de Ω correspond un nombre réel x associé à la variable aléatoire X. (On note v.a. pour variable aléatoire.)

*Si $E = \{x_1,\ x_2,\ x_3, ...\}$, la **loi de probabilité** de X est donnée par les réels*
$p_n = \mathbf{P}(X = x_n) = \mathbf{P}(\{\omega \in \Omega \text{ tel que } X(\omega) = x_n\})$.

Sauf mention contraire, nous ne considérons ici que le cas où le nombre de valeurs de X ayant une probabilité non nulle est fini ;le cas où le nombre de valeurs de X ayant une probabilité non nulle est infini dénombrable nécessite la notion en analyse de “séries numériques”, thème que nous n'abordons pas dans ce module.

2 Fonction de répartition d'une variable réelle

Soit X une variable aléatoire réelle. La fonction de répartition de X, notée F, donne la probabilité $F(x)$ pour que X tombe à gauche de x.

$$\begin{aligned} F : \mathbb{R} &\longrightarrow [0,1] \\ x &\longmapsto F(x) = \mathbf{P}(X \leq x) = \mathbf{P}_X(]-\infty, x]) \ . \end{aligned}$$

Par complémentaire, on a bien sûr $\mathbf{P}(X > x) = 1 - F(x)$.

F est une fonction en escalier, croissante, continue à droite et vérifiant

$$\lim_{x \to -\infty} F(x) = 0 \qquad \text{et} \qquad \lim_{x \to +\infty} F(x) = 1 \ .$$

Exemple 2.1
Soit la loi de probabilité suivante :

Valeurs de X	2	5	6	8	10	12
Probabilités	0,05	0,1	0,2	0,4	0,15	0,1

On obtient la fonction de répartition suivante :
- si $x < 2$, alors $F(x) = 0$
- si $2 \leq x < 5$ alors $F(x) = 0,05$
- si $5 \leq x < 6$ alors $F(x) = 0,15$
- si $6 \leq x < 8$ alors $F(x) = 0,35$
- si $8 \leq x < 10$ alors $F(x) = 0,75$
- si $10 \leq x < 12$ alors $F(x) = 0,9$
- si $x \geq 12$ alors $F(x) = 1$

3 Moyenne et variance d'une variable réelle

Elles sont définies de manière similaire à celles des variables statistiques mais les variables aléatoires pouvant prendre une infinité de valeurs, elles n'existent pas toujours. À la place du mot «moyenne», on utilise souvent le mot «espérance» en particulier quand on parle de gain ou de durée de vie.

Espérance. Soit X une variable aléatoire réelle discrète dont la distribution de probabilité est donnée par le tableau

Valeurs de X	x_1	x_2	...	x_n	...
Probabilités	p_1	p_2	...	p_n	...

1er cas : le nombre de valeurs possibles pour X est fini et égal à n. Alors l'espérance de X existe et vaut

$$\mathbf{E}(X) = \sum_{i=1}^{n} p_i\, x_i \ .$$

2ème cas : le nombre de valeurs de X ayant une probabilité non nulle est infini dénombrable.
Si la série basée sur la suite $(p_i\, x_i)_{i\geq 1}$ est convergente alors sa somme est l'espérance de X

$$\mathbf{E}(X) = \sum_{i=1}^{+\infty} p_i\, x_i \ .$$

Sinon X n'a pas d'espérance.

Nous n'étudions pas dans ce cours, d'exemple qui illustre ce cas car il nécessite la notion d'analyse sur les séries. Cependant, nous retrouverons à la fin de ce chapitre, deux lois usuelles de variables pouvant prendre un nombre infini de valeurs.

Variance.
Si la variable aléatoire réelle X admet une espérance $E(X)$ et si la variable $(X - E(X))^2$ admet aussi une espérance, on l'appelle variance de X et on la note $\mathbf{V}(X)$. On a alors

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

Lorsque $\mathbf{V}(X)$ existe, on note $\sigma(X)$ sa racine carrée et on l'appelle écart–type de X.

Exemples

Exemple 1.1 : Espérance de la variable S somme des deux dés

$$E(S) = \sum_{i=1}^{11} p_i\, x_i = \frac{252}{36} = 7\ .$$

Variance de la variable S somme des deux dés

$$\mathbf{E}(S^2) = \sum_{i=1}^{11} p_i\, x_i^2 = \frac{1974}{36}.$$

D'où

$$\mathbf{V}(S) = \frac{1974}{36} - 7^2\ = \frac{210}{36}$$

On en déduit $\sigma(S) = 2,41$.

Exemple 1.2

$$E(X) = \sum_{i=1}^{6} p_i\, x_i = \frac{77}{10} = 7,7\ .$$

$$\mathbf{E}(S^2) = \sum_{i=1}^{n} p_i\, x_i^2 = 64,9$$

D'où

$$\mathbf{V}(X) = 64,9\ -\ 7,7^2 = 5,61$$

On en déduit $\sigma(X) = 2,37$

L'espérance s'interprète comme ce que l'on peut espérer obtenir en moyenne si on répète un grand nombre de fois l'expérience aléatoire. Par exemple, pour un jeu d'argent, une espérance positive caractérisera un jeu favorable aux joueurs, une espérance négative un jeu défavorable, une espérance nulle un jeu équitable.
L'écart-type est une quantité réelle positive, utilisée pour caractériser la répartition d'une v.a. autour de sa moyenne.

4 Couple aléatoire

Définition 2.3

Soient X et Y deux variables aléatoires discrètes définies sur le même espace probabilisé $(\Omega,\, \mathcal{P}(\Omega),\, \mathbf{P})$, l'application :

$$\begin{aligned} \Omega &\longrightarrow \mathbb{R}^2 \\ \omega &\longmapsto (X(\omega),\, Y(\omega)) \end{aligned}$$

est appelée couple aléatoire discret des marginales X et Y et notée $(X,\, Y)$.

Définition 2.4
La loi $\mathbf{P}_{X,Y}$ *du couple* (X, Y) *est la probabilité définie sur l'ensemble des parties de* $\mathbb{R}^2$ *par :*

$$\forall B \subset \mathbb{R}^2, \quad \mathbf{P}_{X,Y}(B) = \mathbf{P}\left(\{\omega \in \Omega, (X(\omega), Y(\omega)) \in B\}\right).$$

Les lois $\mathbf{P}_X$ *et* $\mathbf{P}_Y$ *des v.a.* X *et* Y *sont appelées lois marginales du couple.*

Dans la suite, les ensembles de valeurs possibles pour les v.a. marginales X et Y seront notés :

$$X(\Omega) = \{x_0, x_1, ..., x_i, ...\} \text{ et } Y(\Omega) = \{y_0, y_1, ..., y_j, ...\}.$$

La loi du couple (X, Y) est caractérisée par les probabilités

$$\mathbf{P}_{X,Y}(\{(x_i, x_j)\}) = \mathbf{P}((X = x_i) \cap (Y = y_j)), \qquad \text{pour } x_i \in X(\Omega), y_j \in Y(\Omega)\ .$$

Proposition 2.1
Si (X, Y) *est un couple aléatoire, ses lois marginales peuvent se calculer par :*

$$\forall x_i \in X(\Omega), \qquad \mathbf{P}_X(x_i) = \sum_{y_j \in Y(\Omega)} \mathbf{P}((X = x_i) \cap (Y = y_j)),$$
$$\forall y_j \in Y(\Omega), \qquad \mathbf{P}_Y(y_j) = \sum_{x_i \in X(\Omega)} \mathbf{P}((X = x_i) \cap (Y = y_j))\ .$$

5 Variables aléatoires indépendantes

Définition 2.5 (Indépendance de deux variables aléatoires)
Soient X *et* Y *deux variables aléatoires discrètes définies sur le même espace probabilisé* $(\Omega, \mathcal{P}(\Omega), \mathbf{P})$. *On dit que les variables aléatoires* X *et* Y *sont indépendantes si et seulement si*

$$(\forall x_i \in X(\Omega)), (\forall y_j \in Y(\Omega)), \quad \mathbf{P}((X = x_i) \cap (Y = y_j)) = \mathbf{P}(X = x_i)\mathbf{P}(Y = y_j)\ .$$

Proposition 2.2
Soient X *et* Y *deux variables aléatoires discrètes indépendantes,* f *et* g *deux fonctions dont les domaines de définition contiennent respectivement* $X(\Omega)$ *et* $Y(\Omega)$. *Alors les variables* $f(X)$ *et* $f(Y)$ *sont indépendantes.*

6 Propriétés de l'espérance et de la variance

6.1 Transformation affine

Soit X une variable aléatoire réelle qui admet une espérance $\mathbf{E}(X)$ et une variance $\mathbf{V}(X)$. Si $(a,\, b) \in \mathbb{R}^2$ alors

$$\begin{aligned} \mathbf{E}(a\,X+b) &= a\,\mathbf{E}(X)+b, \\ \mathbf{V}(a\,X+b) &= a^2\mathbf{V}(X). \end{aligned}$$

6.2 Variance d'une somme

Si X et Y sont deux variables aléatoires définies sur le même espace probabilisé et qui ont toutes deux une espérance et une variance alors $(X+Y)$ a une espérance et une variance et on a

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

où $cov(X,Y)$, la covariance de $(X,\, Y)$ est définie par

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

Attention ! Comme l'espérance et la variance, la covariance n'existe pas toujours.

Proposition 2.3
Soient X et Y deux variables aléatoires discrètes ***indépendantes****.*
Alors $cov(X,\, Y) = 0$.

Attention ! La réciproque est fausse.
Mais la contraposée de la propriété nous permet de déduire :
Si $cov(X,\, Y) \neq 0$, alors les variables X et Y ne sont pas indépendantes.

Exemple 2.2
On donne la loi du couple de variable aléatoire (X,Y).

X \ Y	0	1	2	$\mathbf{P}(X=x_i)$
0	0,2	0,1	0,1	0,4
1	0,2	0,2	0,2	0,6
$\mathbf{P}(Y=x_j)$	0,4	0,3	0,3	1

Ayant déterminé les lois marginales de X et de Y, nous calculons l'espérance, la variance de chacune de ces variables et leur covariance.
$\mathbf{E}(X) \ = \ 0 \times \mathbf{P}(X=0) \ + \ 1 \times \mathbf{P}(X=1)$. D'où $\mathbf{E}(X) = 0,6$.
$\mathbf{E}(Y) \ = \ 0 \times \mathbf{P}(Y=0) \ + \ 1 \times \mathbf{P}(Y=1) \ + \ 2 \times \mathbf{P}(Y=2)$. D'où $\mathbf{E}(Y) = 0,9$.
$\mathbf{E}(XY) = \sum_{i=1}^{2} \sum_{j=1}^{3} \mathbf{P}((X = x_i) \cap (Y = y_j))$
$\mathbf{E}(XY) = 0,6$
Donc $\mathbf{Cov}(X,Y) = 0,6 - 0,6 \times 0,9 = 0,06$.
Comme $\mathbf{Cov}(X,Y) \neq 0$, on peut en déduire que les variables X et Y ne sont pas indépendantes.
Mais pour cela, il suffisait de remarquer que, par exemple,

$$\mathbf{P}((X=0) \cap (Y=0)) \neq \mathbf{P}(X=0)\mathbf{P}(Y=0).$$

7 Lois discrètes usuelles

Soit X une variable aléatoire définie sur un espace probabilisé $(\Omega,\ \mathcal{P}(\Omega),\ \mathbf{P})$.

7.1 Loi de Bernoulli

Définition 2.6
La v.a. X suit une loi de Bernoulli de paramètre p ($p \in [0;\,1]$) si elle ne prend que des valeurs 0 et 1 avec :

$$\mathbf{P}(X=1) = p$$
$$\mathbf{P}(X=0) = 1 - p = q\ .$$

Notation : $X \sim B(p)$.

Elle permet la modélisation de toute expérience où l'on a 2 possibilités (pile ou face, réussite ou échec ...). On peut remarquer que $X^2 = X$, donc X^2 suit également la loi $\sim B(p)$ (de même pour X^n, $n \geq 1$). On a donc

$$E(X) = p \quad \text{et} \quad E(X^2) = p$$

$$V(X) = p - p^2 = p(1-p).$$

Exemple 2.3
Dans une urne, il y a 3 boules rouges et 7 boules noires. On choisit au hasard 1 boule dans cette urne. On s'intéresse à la probabilité d'obtenir 1 boules rouge.
Soit X la v.a. prenant la valeur 1 si la boule tirée est rouge (succès) et 0 sinon.
$X \sim B(\frac{3}{10})$

D'où la probabilité recherchée :

$$\mathbf{P}(X = 1) = \frac{3}{10}.$$

Exemple 2.4
On lance un dé équilibré.
Soit X la v.a. prenant la valeur 1 si le chiffre qui apparaît est pair (succès) et 0 sinon.
$X \sim B(\frac{1}{2})$
D'où la probabilité recherchée :

$$\mathbf{P}(X = 1) = \frac{1}{2}.$$

7.2 Loi uniforme sur un ensemble fini de réels

Définition 2.7
La v.a. X suit la loi uniforme sur l'ensemble des réels $\{x_1,\, x_2, ...,\, x_n\}$ si :

$$\forall\, k \in \mathbb{N}^*, \quad \mathbf{P}(X = x_k) = \frac{1}{n}\ .$$

Exemple 2.5
On lance un dé équilibré. Le nombre de points indiqué par le dé suit la loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$.
Soit X la v.a. désignant le nombre de points.

$$\mathbf{P}(X = i) = \frac{1}{6}, \qquad i \in \{1, 2, 3, 4, 5, 6\}.$$

L'espérance d'une loi uniforme sur $\{x_1,\, x_2, ...,\, x_n\}$ est égale à

$$E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

et sa variance

$$V(X) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \ - \ (E(X))^2.$$

7.3 Loi binomiale

Définition 2.8
La v.a. X suit une loi binomiale de paramètres n et p ($n \in \mathbb{N}$ et $p \in [0;\, 1]$) si l'ensemble des valeurs est $X(\Omega) = \{0,\, 1, ...,\, n\}$ et :

$$\forall\, k \ = \ 0, \ ..., n, \quad \mathbf{P}(X = k) = \binom{n}{k} p^k\, (1-p)^{n-k}\ .$$

Notation : $X \sim B(n,\, p)$.

La formule ci–dessus définit bien une loi de probabilité puisque les $\binom{n}{k} p^k\,(1-p)^{n-k}$ sont positifs et

$$\sum_{k=0}^{n} \binom{n}{k} p^k\,(1-p)^{n-k} = (p + (1-p))^n = 1^n = 1$$

en appliquant la formule du binôme de Newton.

Espérance et variance de la loi binomiale $B(n,\, p)$.
Les espérances de X et X^2 sont données par

$$E(X) = \sum_{k=0}^{n} k\mathbf{P}(X = k) = \sum_{k=0}^{n} k \binom{n}{k} p^k\,(1-p)^{n-k}$$

$$E(X^2) = \sum_{k=0}^{n} k^2\mathbf{P}(X = k) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k\,(1-p)^{n-k}$$

qui ne se prêtent pas très bien au calcul. Pour évaluer $E(X)$ et $V(X)$ il vaut mieux utiliser un raisonnement probabiliste fondé sur des propriétés de variables aléatoires. Pour cela on utilise le fait que la loi binomiale est celle du nombre X de succès dans la répétition de n expériences de Bernoulli ayant la même probabilité de succès p. On note Y_i le résultat de l'expérience numéro i. La loi de chaque variable Y_i est une loi de Bernoulli.

Valeurs de Y_i	0	1
Probabilités	p	$(1-p)$

On a donc $E(Y_i) = p$ et $V(Y_i) = p(1-p)$. Comme X est le nombre de total de succès parmi les n expériences, on peut écrire que

$$X = \sum_{i=1}^{n} Y_i$$

L'espérance d'une somme est égale à la somme des espérances, on a donc

$$E(X) = \sum_{i=1}^{n} E(Y_i) = np.$$

La variance d'une somme de variables aléatoires *indépendantes* est égale à la somme des variances donc

$$V(X) = \sum_{i=1}^{n} V(Y_i) = np(1-p).$$

Exemple 2.6
Dans une urne, il y a 3 boules rouges et 7 boules noires. On choisit au hasard et successivement 8 boules dans cette urne, en remettant chaque fois la boule tirée dans l'urne. Si l'on suppose l'indépendance des tirages, quelle est la probabilité d'obtenir 5 boules rouges ?
Notons X la v.a. " nombre de boules rouges obtenues ". X représente le nombre de succès (" la boule tirée est rouge "). lorsqu'on répète de façon indépendante $n = 8$ expériences de Bernoulli de même paramètre $p = \frac{3}{10}$.
$X \sim B(8, \frac{3}{10})$
D'où la probabilité recherchée :

$$\mathbf{P}(X = 5) = \binom{8}{5} \left(\frac{3}{10}\right)^5 \left(1 - \frac{3}{10}\right)^3 .$$

7.4 Loi hypergéométrique

Alors que la loi binomiale intervient dans les tirages avec remise, la loi hypergéométrique correspond aux tirages sans remise.

Exemple 2.7
Dans une production totale de N objets dont M sont défectueux, on prélève au hasard un échantillon de n objets (tirage sans remise). Soit X le nombre aléatoire d'objets défectueux dans l'échantillon. Quelle est sa loi ?

On peut prendre comme espace Ω l'ensemble de tous les échantillons possibles (toutes les parties à n éléments d'un ensemble de cardinal N) muni de l'équiprobabilité. Chaque échantillon a ainsi une probabilité $\frac{1}{\binom{N}{n}}$ d'être choisi. Les échantillons réalisant l'événement $\{X = k\}$ sont ceux qui contiennent k objets défectueux et $n - k$ objets non défecteux. Il y a $\binom{M}{k}\binom{N-M}{n-k}$ échantillons possibles. Ceci n'est réalisé que pour $0 \leq k \leq M$ et $0 \leq n - k \leq N - M$. On a donc :

$$\begin{aligned} \mathbf{P}(X = k) &= \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \qquad \text{si } 0 \leq k \leq \min(M, n) \text{ et} \\ &= 0 \text{ sinon.} \end{aligned}$$

Définition 2.9
La loi définie ci-dessus s'appelle la loi hypergéométrique de paramètres N, n et $\frac{M}{N}$. N est l'effectif de la population totale, M celui de la sous-population à laquelle on s'intéresse et n la taille de l'échantillon.
Notation : $X \sim H(N, n, p)$ avec $p = \frac{M}{N}$.

On peut montrer que

$$E(X) = np \quad et$$
$$V(X) = np(1-p)\frac{N-n}{N-1}.$$

Pour une taille d'échantillon n fixée, plus N et M sont grands, moins les tirages sans remise diffèrent des tirages avec remise.

Exemple 2.8
Dans une urne, il y a 3 boules rouges et 7 boules noires. On choisit au hasard et simultanément 8 boules dans cette urne. Quelle est la probabilité d'obtenir 2 boules rouges ?
Notons X la v.a. "nombre de boules rouges obtenues". X représente le nombre de succès ("la boule tirée est rouge") lors de $n = 8$ tirages de boules parmi $N = 10$.
$X \sim H(10,\, 8,\, \frac{3}{10})$.
D'où la probabilité recherchée :

$$\mathbf{P}(X=2) = \frac{\binom{3}{2}\binom{7}{6}}{\binom{10}{8}}$$

7.5 Loi géométrique

Définition 2.10 (v.a. de loi géométrique)
Soit $p \in]0,1[$. On considère une épreuve de Bernoulli dont la probabilité de succès est p et celle d'échec $q = 1-p$.
On renouvelle cette épreuve de manière indépendante jusqu'au premier succès. Soit X le numéro de la première épreuve où l'on obtient un succès. Les valeurs de X sont les entiers naturels non nuls 1, 2, ... On dit que X suit une loi géométrique de paramètre p.
La probabilité que le premier succès soit obtenu au rang k est :

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

Notation : $X \sim G(p)$.

Remarque. La suite $((1-p)^{k-1}\, p)_{k\geq 1}$ est géométrique de premier terme p et de raison $1-p$.

A l'aide de la théorie des séries numériques, on montre que

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

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

Exemple 2.9
Une machine de casino est réglée pour que le joueur ait une probabilité de gain de 0,4 à chaque partie. Si un joueur s'installe à cette machine, la probabilité de gagner pour la première fois à la troisième partie est $0,4 \times 0,6^2 = 0,144$.
L'espérance vaut $\frac{1}{0,4} = 2,5$ à $0,1$ près, ce qui signifie que sur un grand nombre de joueurs, il faut en moyenne 2,5 parties pour gagner.

7.6 Loi de Poisson

Définition 2.11
Soit $\lambda > 0$ un réel. On dit qu'une variable aléatoire X suit la loi de Poisson de paramètre $\lambda > 0$ si $X(\Omega) = \mathbb{N}$ et

$$\forall\, k \in \mathbb{N}, \quad \mathbf{P}(X = k) = \frac{\exp(-\lambda)\, \lambda^k}{k!} \ .$$

Notation : $X \sim P(\lambda)$.

Exemple 2.10
On a effectué une étude sur une voiture de marque A. L'étude révèle que le nombre de pannes pendant une durée d'un an suit une loi de Poisson de paramètre 3. On note X la variable aléatoire donnant le nombre de pannes en un an.
La probabilité que la voiture n'ait aucune panne en un an est

$$\mathbf{P}(X = 0) = \frac{\exp(-3)\, 3^0}{0!} \approx 0.05$$

Une des raisons de l'importance de cette loi est le théorème de convergence de la loi binomiale vers la loi de Poisson. Lorsqu'on répète un très grand nombre de fois des expériences de Bernoulli indépendantes de même paramètre p très petit, on peut utiliser la loi de probabilité de Poisson. En général, $n > 30$ et $np \in [0.1, 18]$ donnent une bonne approximation.
Elle s'applique par exemple au nombre d'erreurs commises lorsque la probabilité d'erreur est très petite et le nombre de répétitions indépendantes très grand (fautes de frappe dans l'ensemble d'un livre, pièces défectueuses lorsque les tests sont nombreux et la production de bonne qualité, ...), au nombre d'appels à un serveur ou au nombre de clients qui arrivent dans un magasin pendant un temps donné, etc.

Exemple 2.11
Une secrétaire a la probabilité 0.001 de faire une erreur en frappant chaque caractère d'un texte. On suppose que chaque frappe de caractère est indépendante des autres et on appelle N le nombre d'erreurs observées sur une page de 3200 caractères. N suit une loi binomiale.
$N \sim B(3200,\, 0.001)$
Comme $np = 3.2 \in [0.1, 18]$, la loi de Poisson est une bonne approximation de la loi binomiale.
La probabilité que la secrétaire ne fasse aucune faute est

$$\mathbf{P}(X = 0) \approx \frac{\exp(-3.2)\, 3.2^0}{0!} \approx 0.04$$

On peut montrer que si X suit la loi $P(\lambda)$ alors

$$E(X) = \lambda \quad \text{et} \quad V(X) = \lambda.$$

Chapitre 3

Variables aléatoires continues

1 Généralités

On a vu dans le chapitre "Variables discrètes", comment décrire le comportement d'une variable aléatoire à valeur dans un ensemble $\{x_1, x_2, \ldots\}$ fini. Déterminer une loi de probabilité $\mathbf{P}$ sur $\{x_1, x_2, \ldots\}$ revenait à déterminer les valeurs $\mathbf{P}(X = x_i)$. Il se peut que l'ensemble des valeurs prises par une v.a. X soit un intervalle I de $\mathbb{R}$. Par exemple, on étudie la durée d'une communication téléphonique ou la durée de vie d'un appareil. X peut alors prendre une infinité de valeurs, la probabilité de chacune d'elles serait nulle. Les événements intéressants ne sont plus « obtenir tel réel » mais plutôt « obtenir un réel compris entre a et b » noté $\mathbf{P}(X \in [a, b])$ ou $\mathbf{P}(a \leq X \leq b)$.

Définition 3.1 (Variable aléatoire continue)
Soit un espace probabilisé d'espace fondamental Ω *et de mesure de probabilité* $\mathbf{P}$. *On dit qu'une variable aléatoire* $X : \Omega \to \mathbb{R}$ *est* ***continue*** *lorsque l'ensemble de ses valeurs est* $\mathbb{R}$ *ou un intervalle de* $\mathbb{R}$.

Exemple 3.1
Considérons la durée d'une communication téléphonique. Cette durée est aléatoire et on cherche un modèle mathématique qui nous permettrait de répondre par exemple aux questions suivantes :
- Quelle est la probabilité pour que la durée ne dépasse pas x minutes ?
- Quelle est la durée moyenne d'une communication ? Son écart–type ?
- Quelle est la probabilité pour que la communication dure encore x minutes sachant qu'elle a déjà duré y minutes ?

On voudrait proposer comme modèle pour la durée une *fonction* positive qui possède les caractéristiques suivantes
- elle serait d'autant plus élevée sur un intervalle, que la probabilité de tomber dans cet intervalle est forte ;

— la probabilité de tomber dans cet intervalle serait égale à la surface comprise entre la courbe et l'axe des abscisses, sur cet intervalle ;
— par conséquence, la surface totale sous la courbe devrait valoir 1.

Un telle fonction devrait nécessairement être positive ou nulle. On décide ici de se restreindre à des fonctions pour lesquelles la surface est calculable, c'est–à–dire les fonctions que l'on sait intégrer. On justifie ainsi la définition de l'objet introduit ci–après.

Définition 3.2 (Fonction de densité)
On appelle fonction de densité *une fonction f définie sur $\mathbb{R}$ et à valeurs réelles telle que*
— $(\forall\, x \in \mathbb{R}), \quad f(x) \geq 0$
— *f est continue sauf éventuellement en un nombre fini de points*
— $\displaystyle\int_{-\infty}^{\infty} f(x)\, dx = 1$

Définition 3.3
Soit f une fonction de densité. On appelle loi de probabilité de densité f, la loi définie pour tout intervalle A de $\mathbb{R}$ par

$$\mathbf{P}(X \in A) = \int_A f(x)\, dx \ .$$

Remarque 3.1
Dans cette définition, chacune des bornes de l'intervalle A peut être infinie. De plus, la probabilité de l'événement $\{X = a\}$ est nulle pour tout réel a. En effet, on peut par abus de langage utiliser la définition avec « l'intervalle » réduit au point a. On a alors comme probabilité une intégrale sur un intervalle réduit à un seul point, qui est nulle.

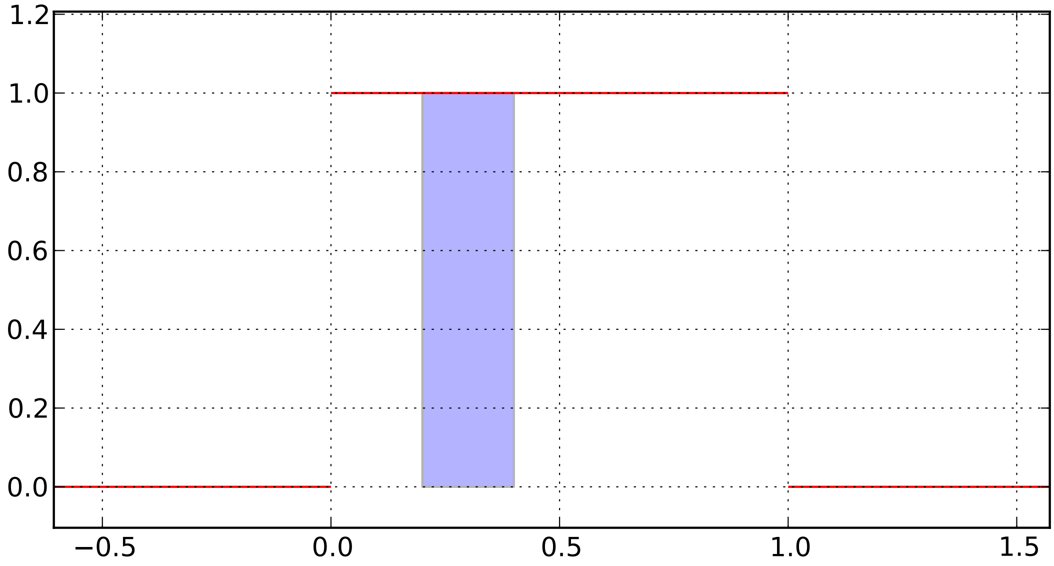
Lecture d'une densité : pour poursuivre notre point de vue de la modélisation, on a maintenant les interprétations suivantes :
— les régions de $\mathbb{R}$ de forte (resp. faible) densité sont les régions de forte (resp. faible) probabilité d'occurrence des réalisations de X. C'est–à–dire que X «tombera» plus souvent dans des zones de forte densité.
— les régions où f s'annule sont des valeurs interdites pour X.

Exemple 3.2 (Loi uniforme)
Les variables aléatoires continues les plus simples sont celles dont la densité est

donnée par la fonction :

$$\begin{aligned} f\,:\,\mathbb{R} &\longrightarrow \mathbb{R}^+ \\ x &\longmapsto f(x) = \begin{cases} 1 \text{ si } 0 \leq x \leq 1 \\ 0 \text{ sinon.} \end{cases} \end{aligned}$$

La représentation graphique de cette fonction est donnée par la figure ci–dessous :



Des réalisations d'une variable aléatoire d'une telle loi sont données par la touche `random` d'une calculatrice. Il s'agit d'un nombre réel, compris entre 0 et 1. On peut lire sur le graphique qu'aucune zone de l'intervalle $[0;\,1]$ n'est privilégiée, d'où le nom de loi *uniforme*. Cette fonction est une densité car elle est

— positive ou nulle (elle vaut 0 ou 1),
— continue sauf en 0 et 1,
— d'intégrale 1 (surface d'un carré de côté 1).

En utilisant les surfaces de rectangles, on notamment que si X est une variable aléatoire de loi uniforme sur $[0;\,1]$

$$\mathbf{P}(X \in [0;\,\frac{1}{2}]) = \mathbf{P}(X \in [\frac{1}{2};\,1]) = \frac{1}{2}, \quad \mathbf{P}(X < 0) = \mathbf{P}(X > 1) = 0\ .$$

2 Fonction de répartition

On a vu en probabilités discrètes que la fonction de répartition était un objet qui caractérisait complètement une variable aléatoire. Un tel objet existe également pour les variables aléatoires à densité.

Définition 3.4
Soit X une variable aléatoire continue de densité f donnée. On appelle fonction de répartition de X la fonction définie sur $\mathbb{R}$ et à valeur dans $[0;\,1]$

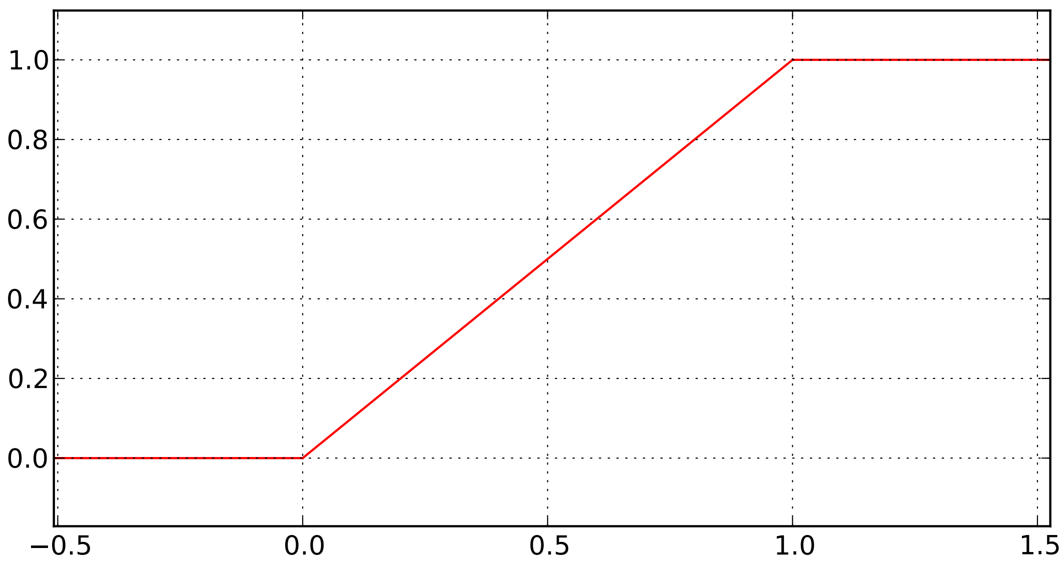
$$\begin{aligned} F : \mathbb{R} &\longrightarrow [0;\,1] \\ t &\longmapsto F(t) = \mathbf{P}(X \leq t) = \int_{-\infty}^{t} f(x)\,dx \;. \end{aligned}$$

Proposition 3.1
La fonction de répartition
- *est croissante sur $\mathbb{R}$,*
- *est continue partout mais dérivable seulement en tout point où f est continue,*
- *a pour limite 0 en $-\infty$ et 1 en $+\infty$.*

Loi uniforme.
Pour l'exemple de la loi uniforme, on peut encore utiliser les surfaces de rectangles pour calculer les intégrales. On obtient la fonction de répartition

$$(\forall\, t \in \mathbb{R}), \quad F(t) = \begin{cases} 0 & \text{si } t < 0 \\ t & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } t > 1 \end{cases} \quad .$$

Sa représentation graphique est :



Lecture de la fonction de répartition. Le comportement d'une variable aléatoire se lit aussi sur le graphique de sa fonction de répartition : les zones de forte probabilité sont les zones de *forte croissance* de F. Les zones où F est constante sont celles où la «masse» de probabilité ne s'accroît pas. La v.a. X ne tombe pas dans ces zones.

Définition 3.5 (Quantiles)
Soit $p \in]0;1[$. *On appelle quantile d'ordre* p *le plus petit réel* t_p *tel que*

$$F(t_p) = p \ .$$

On distingue en particulier la médiane Me *qui est le quantile d'ordre* $\frac{1}{2}$.
Les trois quartiles Q_1, Q_2 *et* Q_3 *sont les quantiles d'ordre respectif* $\frac{1}{4}$, $\frac{1}{2}$ *et* $\frac{3}{4}$.

Les trois quartiles partagent $\mathbb{R}$ en quatre intervalles de probabilités égales. On peut aussi dire qu'ils partagent le graphique de la densité en quatre parties de surfaces égales.
Le quantile d'ordre 0,95 est la valeur t telle que $\mathbf{P}(X \leq t) = 0.95$.

3 Espérance et variance

Rappelons que pour une variable aléatoire discrète, l'espérance était définie par

$$\mathbf{E}(X) = \sum_i x_i \, \mathbf{P}(X = x_i).$$

Pour avoir une bonne intuition de l'objet à introduire, on considère que les valeurs de x_i sont très nombreuses et on imagine un petit intervalle autour d'un x_i d'amplitude $\mathrm{d}x$. Pour cet x_i et son intervalle, on peut faire l'approximation

$$\mathbf{P}(x_i < X < x_i + dx) \sim f(x_i) \, \mathrm{d}x$$

où f est une densité appropriée. Cela revient à disperser la masse de probabilité placée en x_i sur l'intervalle infinitésimal autour de x_i. Il reste à sommer les contributions de chacun de ces intervalles, ce qui correspond à la notion d'intégration. On justifie donc ainsi la définition suivante :

Définition 3.6
On appelle espérance *de la variable aléatoire* X *de densité* f, *le réel* **s'il existe**

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x \, f(x) \, dx \ .$$

Plus généralement, si ϕ est une application de $\mathbb{R}$ dans $\mathbb{R}$ on définit de même

$$\mathbf{E}(\phi(X)) = \int_{-\infty}^{\infty} \phi(x)\, f(x)\, dx \ .$$

Il faut remarquer que l'espérance n'existe pas toujours car l'intégrale n'est pas forcément finie.

Les propriétés de linéarité de l'intégrale donnent les résultats intuitifs suivants sur l'espérance :

Proposition 3.2
Dans la mesure où les quantités ci–dessous existent, on a pour $a \in \mathbb{R}$

$$\mathbf{E}(a\, X) = a\, \mathbf{E}(X) \qquad et \qquad \mathbf{E}(X + a) = \mathbf{E}(X) + a.$$

Lorsqu'une variable aléatoire X à densité admet une espérance $\mathbf{E}(X)$, on peut se questionner sur la dispersion des réalisations de X autour de cette valeur espérée. Cette dispersion est mesurée, entre autre, par la variance définie ainsi :

Définition 3.7
Soit X une variable aléatoire de densité f, admettant une espérance $\mathbf{E}(X)$. On appelle variance *de X, la valeur positive réelle si elle existe, donnée par*

$$\mathbf{V}(X) = \mathbf{E}((X - \mathbf{E}(X))^2) = \int_{-\infty}^{\infty} (x - \mathbf{E}(X))^2\, f(x)\, dx.$$

*On définit également l'*écart-type *comme la racine carrée de la variance.*

Formule de calcul. La définition dans la forme donnée ci–dessus est utile pour saisir la signification de la variance en terme de dispersion mais n'est pas pratique pour le calcul. On montre, en développant le carré dans l'intégrale, que si la variance existe, alors on a

$$\begin{aligned}\mathbf{V}(X) &= \mathbf{E}(X^2) - (\mathbf{E}(X))^2 \\ &= \int_{-\infty}^{\infty} x^2\, f(x)\, \mathrm{d}x - \left(\int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x\right)^2\end{aligned}$$

qui s'énonce «l'espérance du carré moins le carré de l'espérance». On vérifiera à titre d'exercice que l'espérance d'une variable aléatoire de loi uniforme sur [0, 1] est égale à $\frac{1}{12}$.

Les résultats suivants sont familiers pour les variances déjà rencontrées en statistique descriptive et en probabilités discrètes. Nous les énonçons donc sans démonstration.

Proposition 3.3
Dans la mesure où les quantités ci–dessous existent, on a pour $a \in \mathbb{R}$

$$\mathbf{V}(a\,X) = a^2\,\mathbf{V}(X) \qquad \textit{et} \qquad \mathbf{V}(X+a) = \mathbf{V}(X)\ .$$

Parmi les lois à densité de variables aléatoires continues, la loi normale que nous étudions dans ce chapitre occupe une place importante. L'une des raisons est qu'elle apparaît de manière naturelle dans un des piliers de la théorie des probabilités : le théorème central limite (que l'on verra dans le chapitre suivant). Elle est proposée pour la première fois comme modèle des erreurs de mesures par le mathématicien allemand Gauss. À cette époque, l'expression de sa densité est déjà connue par les travaux antérieurs de De Moivre, poursuivis par Laplace. Pour cette raison, on rencontrera parfois la loi normale sous le nom de loi de Laplace–Gauss, loi gaussienne, loi de De Moivre–Laplace–Gauss etc ... La forme caractéristique «en cloche» de sa densité est observée dans les études statistiques de nombreux phénomènes réels.

4 Une loi à densité usuelle : la loi normale

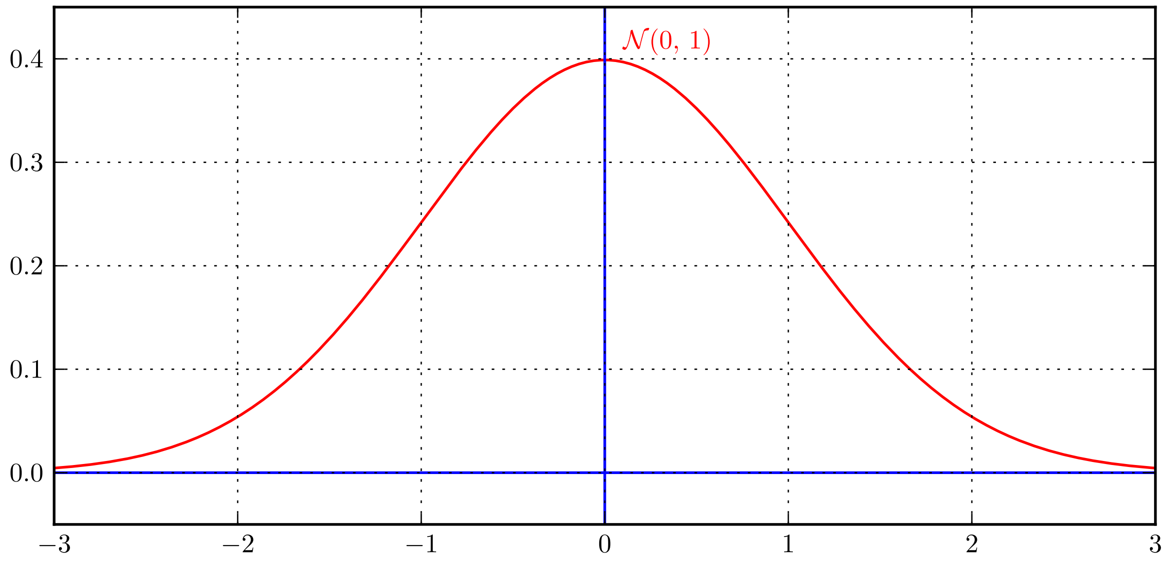
4.1 Densité de la loi normale centrée réduite $\mathbb{N}(0,\,1)$

Définition 3.8
On suppose donné un espace probabilisé $(\Omega,\,\mathcal{P}(\Omega),\,\mathbf{P})$ qui ne sera pas précisé. Une variable aléatoire continue Z est dite de loi normale centrée réduite si sa fonction de densité est donnée par

$$\begin{aligned} f\,:\,\mathbb{R} &\longrightarrow \mathbb{R}^+ \\ x &\longmapsto f(x) = \frac{1}{\sqrt{2\,\pi}}\,e^{-\frac{x^2}{2}}\,. \end{aligned}$$

On notera $Z \sim \mathbb{N}(0,\,1)$.

La représentation graphique de cette densité est donnée ci–après.



Cette définition prétend que la fonction définie ci–dessus est bien une densité. Or, s'il est clair qu'elle est continue sur $\mathbb{R}$ (comme composition et produit de fonctions continues) et positive, il reste à prouver que

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x = 1.$$

En tant que fonction continue sur $\mathbb{R}$, la fonction f admet des primitives mais **on n'en connaît pas** d'expression analytique. La démonstration de ce résultat ne s'appuie donc pas sur un simple calcul de primitive mais sur une technique d'intégration qui ne relève pas de ce cours. On admet donc ce résultat important, ainsi que les suivants qui sont du même type.

Proposition 3.4 (Moments d'une normale centrée réduite)
Si $Z \sim \mathbb{N}(0,\ 1)$, *alors*

$$\mathbf{E}(Z) = 0 \qquad \textit{et} \qquad \mathbf{V}(Z) = 1\ .$$

Il faut noter que la densité f est une fonction paire. Sa représentation graphique admet donc l'axe des ordonnées comme axe de symétrie ce qui suffit pour assurer que l'espérance soit nulle.
L'espérance est aussi la médiane.

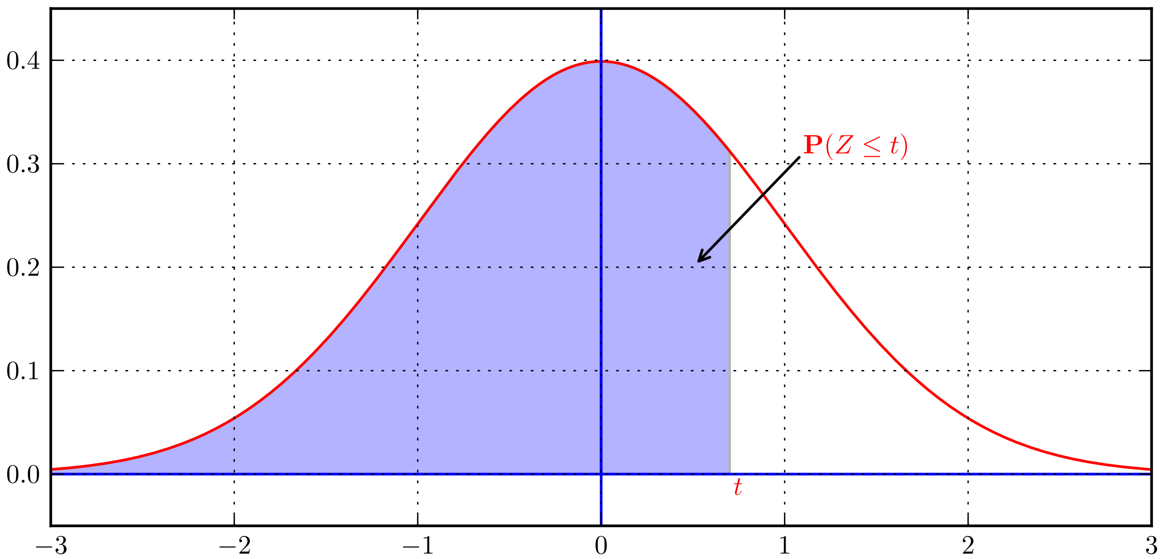
Fonction de répartition, table

Comme toute variable aléatoire à densité, la fonction de répartition de la loi de Z est définie par

$$\begin{aligned} F : \mathbb{R} &\longrightarrow [0;\, 1] \\ t &\longmapsto F(t) = \mathbf{P}(X \leq t) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{t} x\, e^{-\frac{x^2}{2}}\, \mathrm{d}x\ . \end{aligned}$$

En pratique, on a recours à des algorithmes d'intégration numériques pour obtenir les valeurs de ces probabilités. Les sorties de ces algorithmes sont rassemblées dans des tables qui donnent la valeur de la fonction de répartition pour suffisamment de t. On distingue deux cas d'utilisation de ces tables.

Lecture directe On se donne $t \in \mathbb{R}$ et on cherche $F(t) \in [0;\, 1]$. Deux cas se présentent.
$\boxed{\text{Si } t \geq 0 :}$ la lecture est immédiate.
Par l'argument de symétrie, la valeur lue sera forcément plus grande que $\frac{1}{2}$.



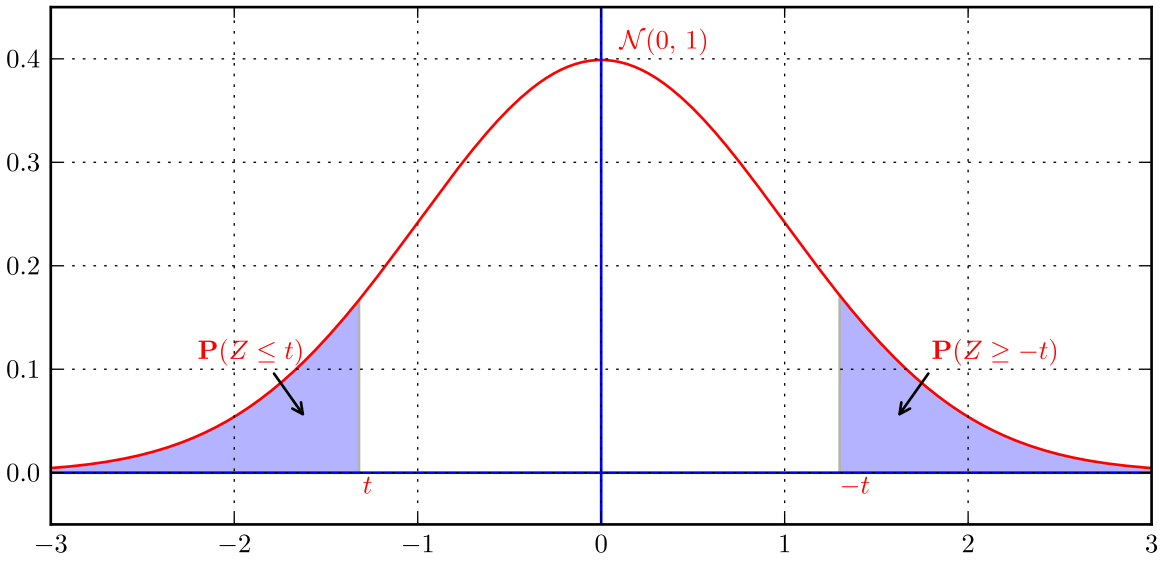
Par exemple, on lit dans la table que :

$$\mathbf{P}(Z \leq 1,56) = 0,9406\ .$$

$\boxed{\text{Si } t < 0 :}$ l'argument de symétrie implique que

$$\mathbf{P}(Z \leq t) = \mathbf{P}(Z \geq -t) = 1 - \mathbf{P}(Z \leq -t)\ .$$

La valeur obtenue sera forcément plus petite que $\frac{1}{2}$.

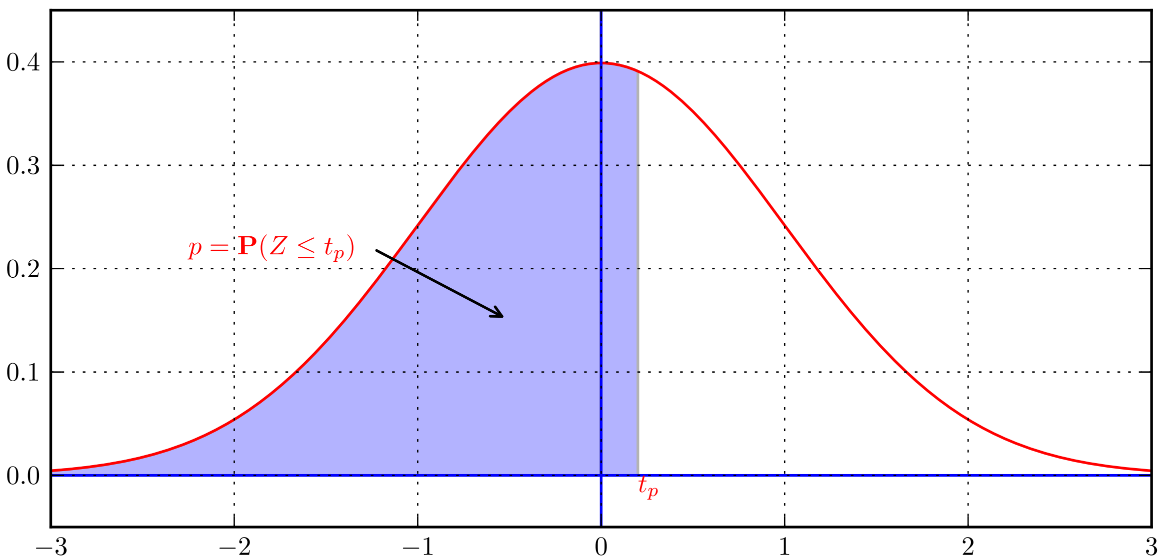


Par exemple, on lit dans la table que :

$$\mathbf{P}(Z \leq -1,78) = 1 - \mathbf{P}(Z \leq 1,78) = 1 - 0,9625 = 0,0375 \ .$$

Lecture inverse (quantile) On se donne maintenant $p \in [0;\, 1]$ et on cherche le quantile d'ordre p, c'est–à–dire le réel t_p tel que $\mathbf{P}(Z \leq t_p) = p$. Ici encore, on distingue deux cas.

$\boxed{\text{Si } p \geq \frac{1}{2} :}$ le quantile recherché est nécessairement positif. La lecture se fait en recherchant dans le corps de la table la valeur la plus proche de p. Le quantile se lit alors dans les marges.

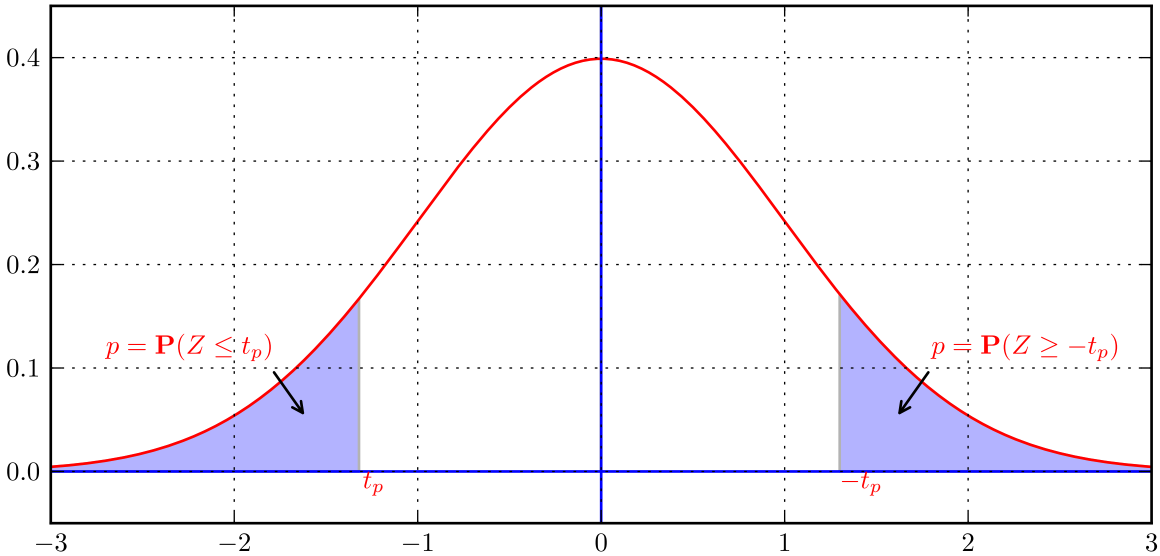


On obtient par exemple que $t_{0,65} \simeq 0,39$.

$\boxed{\text{Si } p < \frac{1}{2} :}$ le quantile recherché est nécessairement négatif. Par l'argument de symétrie, le quantile t_p et son opposé sont liés par la relation

$$p = \mathbf{P}(Z \leq t_p) = \mathbf{P}(Z \geq -t_p)$$

c'est–à–dire $t_p = -t_{1-p}$.



On est alors ramené au cas précédent en recherchant le quantile d'ordre $1 - p \geq \frac{1}{2}$, dont on prendra l'opposé.
On lit par exemple que $t_{0,27} = -t_{0,73} = -0,61$.

4.2 Loi normale $\mathbb{N}(\mu,\, \sigma)$

La loi normale que nous venons d'étudier est centrée–réduite c'est-à-dire que son espérance est nulle et sa variance 1. On opère maintenant une transformation affine pour obtenir une espérance et une variance quelconque.
Soient $\mu \in \mathbb{R}$ et σ un réel **strictement positif**. On s'intéresse à la variable $X = \mu + \sigma\, Z$ où $Z \sim \mathbb{N}(0,\, 1)$. On se rappelle que la densité d'une variable aléatoire peut s'obtenir comme la dérivée de sa fonction de répartition. On calcule donc la fonction de répartition F_X de X en fonction de F_Z, celle de Z.

$$\begin{aligned} F_X(x) = \mathbf{P}(X \leq x) &= \mathbf{P}(\mu + \sigma\, Z \leq x) \\ &= \mathbf{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) \\ &= F_Z\left(\frac{x-\mu}{\sigma}\right) \ . \end{aligned}$$

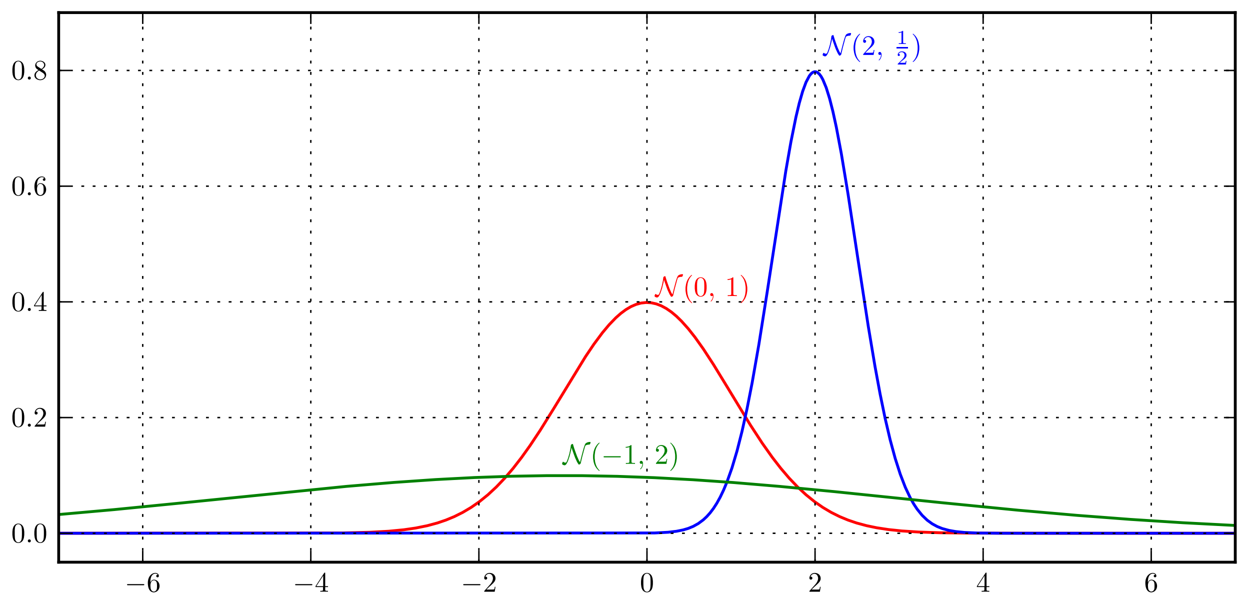
Il suffit alors de dériver cette dernière composition de fonction et d'utiliser le fait que la dérivée de F_Z est précisément la densité de la loi normale centrée réduite.

$$\begin{aligned} f_X(x) &= F'_Z\left(\frac{x-\mu}{\sigma}\right)\left(\frac{x-\mu}{\sigma}\right)' \\ &= f_Z\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma} \\ &= \frac{1}{\sqrt{2\,\pi}\,\sigma}\,e^{-\dfrac{(x-\mu)^2}{2\,\sigma^2}}\ . \end{aligned}$$

Cette dernière expression est appelée *densité de la loi normale d'espérance* μ *et d'écart–type* σ, notée $\mathbb{N}(\mu,\,\sigma)$. Le calcul de l'espérance et de la variance est en effet immédiat :

$$\begin{aligned} \mathbf{E}(X) &= \mathbf{E}(\mu+\sigma\,Z) = \mu+\sigma\,\mathbf{E}(Z) = \mu \\ \mathbf{V}(X) &= \mathbf{V}(\mu+\sigma\,Z) = \sigma^2\,\mathbf{V}(Z) = \sigma^2\ . \end{aligned}$$

En étudiant la densité de la loi normale, on peut remarquer que l'allure générale de la courbe reste la même quels que soient l'espérance et l'écart–type. Un déplacement de l'espérance induit un décalage (translation) de la courbe et un changement d'écart–type induit une dilatation. Plus l'écart–type est grand, plus la courbe «s'étale». À l'inverse la courbe devient très «pointue» si l'écart–type est petit.



Fonction de répartition

Il est hors de question de produire de nombreuses tables pour les fonctions de répartition de plusieurs lois normales d'espérance et de variance quelconque. C'est de plus inutile : en effet, puisque la loi $\mathbb{N}(\mu,\, \sigma)$ a été obtenue par une transformation affine de la loi $\mathbb{N}(0,\, 1)$, la transformation affine inverse permet d'obtenir la loi $\mathbb{N}(0,\, 1)$ à partir de la loi $\mathbb{N}(\mu,\, \sigma)$.

Proposition 3.5

Si $X \sim \mathbb{N}(\mu,\, \sigma)$, alors

$$Z = \frac{X-\mu}{\sigma} \sim \mathbb{N}(0,\, 1).$$

Il suffit de reprendre à l'envers le calcul qui a permis d'obtenir $\mathbb{N}(\mu,\, \sigma)$ au moyen de $\mathbb{N}(0,\, 1)$.

Proposition 3.6

La fonction de répartition de la loi $\mathbb{N}(\mu,\, \sigma)$ s'exprime au moyen de la fonction de répartition de la loi $\mathbb{N}(0,\, 1)$ grâce à la relation :

$$\begin{aligned} F_X(t) = \mathbf{P}(X \leq t) &= \mathbf{P}\left(\frac{X-\mu}{\sigma} \leq \frac{t-\mu}{\sigma}\right) \\ &= \mathbf{P}\left(Z \leq \frac{t-\mu}{\sigma}\right) \\ &= F_Z\left(\frac{t-\mu}{\sigma}\right) \quad . \end{aligned}$$

On remarque que l'espérance est encore la médiane.

4.3 Somme de variables aléatoires normales

Proposition 3.7

*Si $X_1 \sim \mathbb{N}(\mu_1,\, \sigma_1)$ et $X_2 \sim \mathbb{N}(\mu_2,\, \sigma_2)$ sont deux variables aléatoires définies sur le même espace et **indépendantes**, alors $X_1 + X_2 \sim \mathbb{N}(\mu_1 + \mu_2,\, \sqrt{\sigma_1^2 + \sigma_2^2})$.*

Il faut remarquer que l'espérance et la variance d'une somme de variables aléatoires sont déjà connues. Ce qui est remarquable dans ce résultat, est que la loi de la somme reste normale. Ce n'est pas le cas pour toutes les lois.
Par ailleurs, une conséquence de cette proposition est que **toute combinaison linéaire non–nulle de normales indépendantes est de loi normale.**

Chapitre 4

Théorème Central Limite

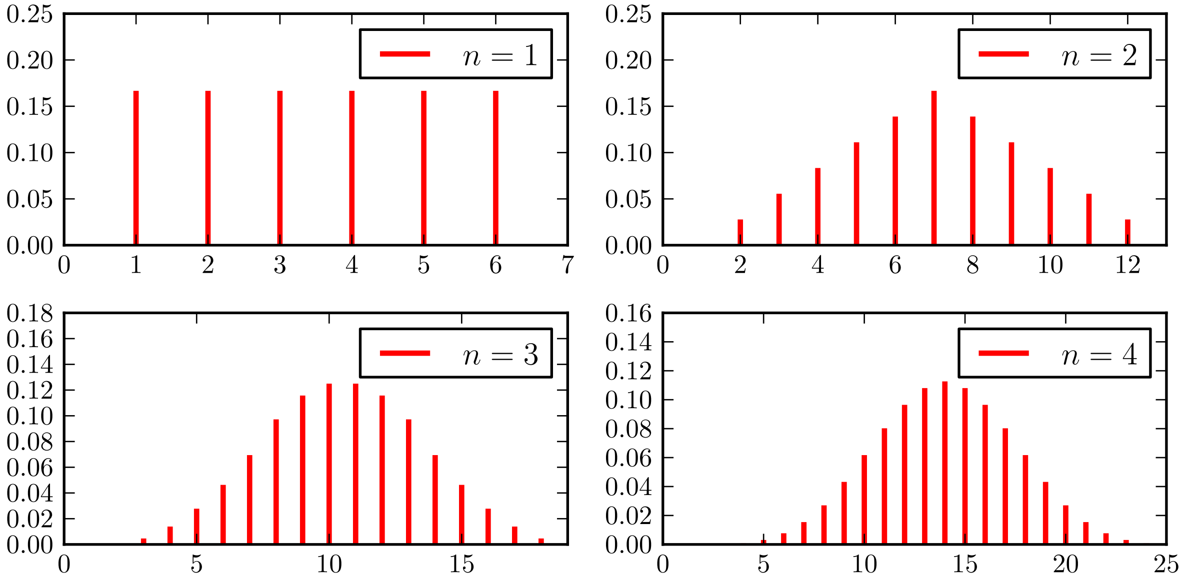
Le théorème central limite est l'un des théorèmes les plus importants de la théorie des probabilités. De façon simplifiée, il affirme que si de nombreux phénomènes aléatoires indépendants et de même loi se cumulent, la loi du résultat global n'est pas très différente d'une loi normale.

1 Somme de variables aléatoires i.i.d.

Dans tout le chapitre, on suppose fixé un espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$ qui ne sera pas précisé et une suite de variables aléatoires X_1, X_2, ..., définies sur Ω et à valeurs dans $E \subset \mathbb{R}$, discret ou non. On suppose que toutes ces variables aléatoires sont indépendantes et de même loi (i.i.d.). On définit la variable aléatoire S_n à valeurs dans $\mathbb{R}$ par

$$S_n = X_1 + \cdots + X_n = \sum_{i=1}^{n} X_i \ .$$

Exemple 4.1
On dispose d'un dé ordinaire et on considère la suite d'expériences qui consiste à jeter le dé indéfiniment. On peut considérer le lancer numéro i comme le résultat de la variable aléatoire X_i de loi uniforme sur $\{1, \ldots, 6\}$. Il s'agit de variables discrètes indépendantes et de même loi. La variable somme S_n est alors simplement le cumul des points jusqu'au lancer numéro n. C'est une variable aléatoire à valeurs dans $\{n, n+1, \ldots, 6n\}$. Intuitivement, il est impensable que la loi de S_n soit toujours uniforme. En effet, la valeur $6\,n$ par exemple sera très rarement obtenue, car il faut pour cela obtenir 6 à tous les lancers.
Le graphique ci–dessous représente la loi de probabilité de S_n pour les quatre premières valeurs de n.



On peut tout d'abord remarquer que dès les faibles valeurs de n, la forme de la loi de S_n rappelle la courbe en cloche de la loi normale. Pour mettre en évidence cette similitude de forme, il est nécessaire de recaler les courbes les unes par rapport aux autres en abscisses. On va donc chercher une mise à l'échelle adéquate.

On note μ et σ^2 l'espérance et la variance commune des X_i. Les deux résultats suivants sont déjà connus sur S_n grâce aux chapitres précédents :

$$\mathbf{E}(S_n) = n\,\mu \qquad \text{et} \qquad \mathbf{V}(S_n) = n\,\sigma^2, \qquad \sigma(S_n) = \sqrt{n}\,\sigma\ .$$

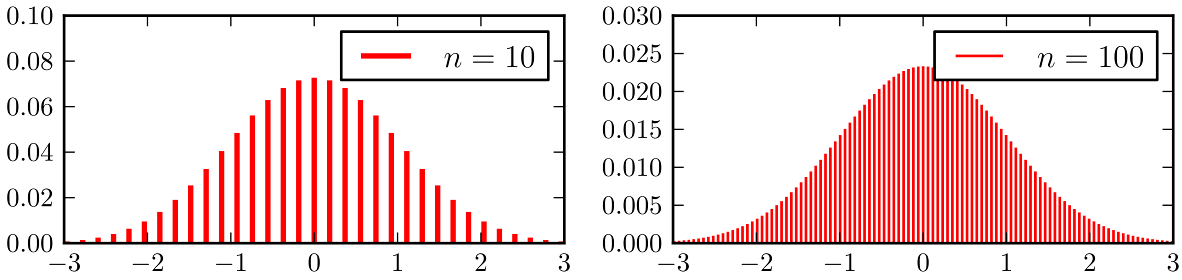
Rappelons que le résultat concernant l'espérance reste vrai même si les variables ne sont pas indépendantes mais pas celui concernant la variance.
Pour comparer graphiquement les lois des S_n pour différentes valeurs de n, il est naturel de considérer des variables de même espérance et de même écart–type. Pour cela on définit

$$Z_n = \frac{S_n - n\,\mu}{\sqrt{n}\,\sigma}.$$

Grâce aux propriétés de l'espérance et de la variance, on a

$$\mathbf{E}(Z_n) = 0 \qquad \text{et} \qquad \mathbf{V}(Z_n) = 1\ .$$

Cette transformation revient à décaler les S_n pour les centrer puis à les renormaliser pour qu'ils aient tous la même variance. Ainsi, la forme du graphique est conservée : seules les graduations des axes ont changé. Sur l'exemple des dés, on obtient après mise à l'échelle :



Sur ces deux graphiques, on distingue une «convergence» vers une forme stable qui ressemble fortement à une densité de loi normale. Le théorème central limite affirme que cette convergence a bien lieu (en un sens à préciser) et ceci pour toute suite de variables i.i.d. dès lors qu'elles admettent une espérance et une variance.

2 TCL pour la somme de variables aléatoires i.i.d.

Théorème 4.1 (TCL pour la somme)
Soient $X_1, X_2, \ldots$ des variables aléatoires indépendantes et identiquement distribuées définies sur le même espace probabilisé, d'espérance et de variance communes μ et σ^2 respectivement. Alors si n est suffisamment grand,

$$Z_n = \frac{S_n - n\,\mu}{\sqrt{n}\,\sigma} \simeq \mathbb{N}(0,\,1)\ .$$

Remarquons qu'on a également l'écriture :

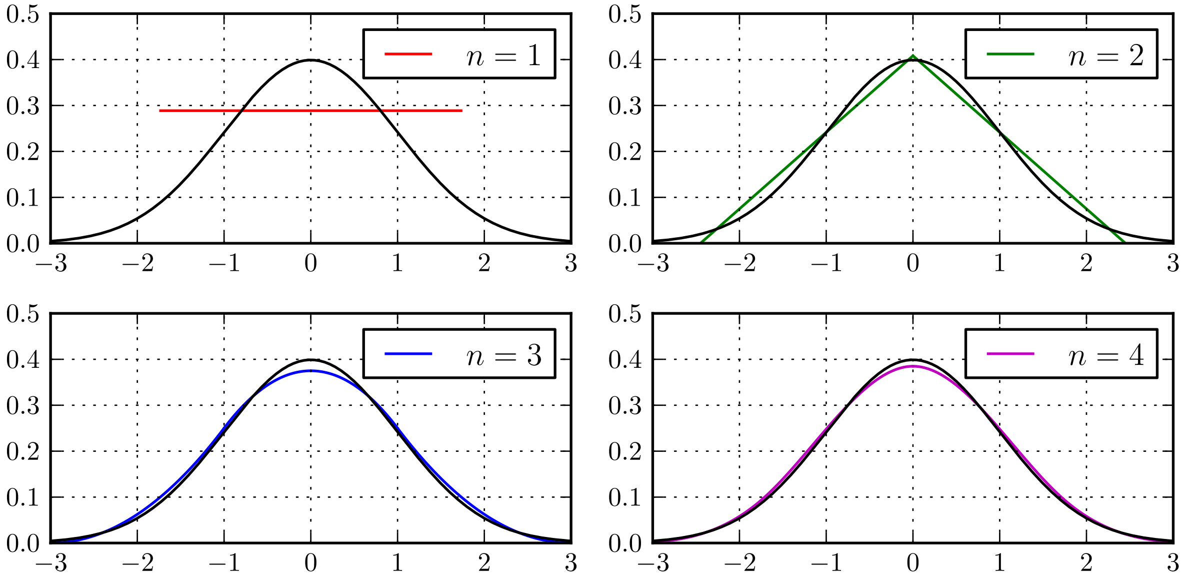
$$Z_n = \frac{S_n - \mathbf{E}(S_n)}{\sigma(S_n)}\ .$$

La condition «n suffisamment grand» est sujette à adaptation suivant le contexte d'application. Une possibilité est de choisir $n \geq 100$. D'une manière générale, on peut retenir que pour une loi continue, symétrique et unimodale, $n = 30$ est suffisant. Dès que l'une de ces conditions n'est pas satisfaite, il est déconseillé d'appliquer le TCL avec une confiance absolue. Le calcul de probabilités d'événements rares doit être pris avec mesure. Il est possible de donner une signification mathématiquement rigoureuse du signe $\simeq$. Nous retiendrons simplement que si A est un intervalle (ou une réunion d'intervalles) de $\mathbb{R}$, alors la probabilité de l'événement $Z_n \in A$ ne diffère que très peu de celle de l'événement $Z \in A$ si n est grand, où Z est de loi $\mathbb{N}(0,\,1)$.
Nous pouvons illustrer ce théorème pour l'exemple des dés par le tableau suivant :

t	0	0,5	1	1,5
$\mathbf{P}(Z_1 \leq t)$	0,5	0,6667	0,8333	1
$\mathbf{P}(Z_4 \leq t)$	0,5563	0,6644	0,841	0,946
$\mathbf{P}(Z_{10} \leq t)$	0,5363	0,676	0,8435	0,942
$\mathbf{P}(Z_{100} \leq t)$	0,5117	0,6904	0,847	0,9322
$\mathbf{P}(Z_{1000} \leq t)$	0,5037	0,6947	0,8435	0,9344
$\mathbf{P}(Z_{10000} \leq t)$	0,5012	0,6917	0,8409	0,9334
$\mathbf{P}(Z \leq t)$	0,5	0,6915	0,8413	0,9332

Remarquons que la dernière ligne s'appuie sur un calcul d'intégrale : il s'agit de lecture dans la table de la loi normale. Les premières lignes sont des sommes ordinaires. Par ailleurs, on observe que si la convergence a bien lieu, elle est assez lente. En fait, la vitesse de convergence dépend fortement de la loi initiale.

Exemple 4.2 (Loi uniforme)
On donne ci–dessous la représentation graphique de la densité des sommes centrées et normalisée Z_n pour les 4 premières valeurs de n. On a superposé la densité de la loi normale centrée réduite sur chaque graphique. Il est clair que la convergence a lieu assez tôt.



3 Applications : approximation de la loi binomiale

On a déjà vu que si X_1, $X_2, \ldots$ était une suite de variable aléatoires i.i.d. dont la loi commune est la loi de Bernoulli $\mathbf{B}(p)$ où p est la probabilité de succès, alors le nombre de succès S_n lors des n premières expériences est de loi binomiale $\mathcal{B}(n, p)$.

On a remarqué l'écriture :

$$S_n = \sum_{i=1}^{n} X_i$$

de sorte que le TCL s'applique. L'espérance et la variance de S_n sont connues car il s'agit d'une loi binomiale :

$$\begin{aligned}
\mathbf{E}(S_n) &= n\,p = n\,\mathbf{E}(X_1)\\
\mathbf{V}(S_n) &= n\,p\,(1-p) = n\,\mathbf{V}(X_1)\\
\sigma(S_n) &= \sqrt{n\,p\,(1-p)} = \sqrt{n}\,\sigma(X_1)\ .
\end{aligned}$$

On en déduit que si n est grand,

$$Z_n = \frac{S_n - n\,p}{\sqrt{n\,p\,(1-p)}} \simeq \mathbb{N}(0,\ 1)\ .$$

Exemple 4.3
Dans une urne, il y a 3 boules rouges et 7 boules noires. On choisit au hasard et successivement 50 boules dans cette urne, en remettant chaque fois la boule tirée dans l'urne. Si l'on suppose l'indépendance des tirages, quelle est la probabilité d'obtenir entre 12 et 20 boules rouges ?
Notons S_{50} la v.a. " nombre de boules rouges obtenues ". S_{50} représente le nombre de succès (" la boule tirée est rouge ") lorsqu'on répète de façon indépendante $n = 50$ expériences de Bernoulli de même paramètre $p = \frac{3}{10}$.
$S_{50} \sim B(50,\ \frac{3}{10})$
$E(S_{50}) = 15$ et $V(S_{50}) = 10,5$

$$\begin{aligned}
\mathbf{P}(12 \le S_{50} \le 20) &= \mathbf{P}\left(\frac{12-20}{\sqrt{10,5}} \le \frac{S_{50}-15}{\sqrt{10,5}} \le \frac{20-15}{\sqrt{10,5}}\right)\\
&= \mathbf{P}\left(\frac{-3}{\sqrt{10,5}} \le Z_n \le \frac{5}{\sqrt{10,5}}\right)\\
&\simeq \mathbf{P}\left(\frac{-3}{\sqrt{10,5}} \le Z \le \frac{5}{\sqrt{10,5}}\right) \qquad \text{où } Z \sim \mathbb{N}(0,\ 1)\\
&= F_{0,1}(1,54) - F_{0,1}(-0,92)
\end{aligned}$$

en notant $F_{0,1}$ la fonction de répartition de la loi normale centrée–réduite.
D'où la probabilité recherchée :

$$\mathbf{P}(12 \le S_{50} \le 20) = 0,9382 - (1 - 0,8212) = 0,7594.$$

4 TCL pour la moyenne de variables aléatoires i.i.d.

Toujours dans le même contexte, considérons la variable aléatoire appelée la moyenne empirique des n premières variables aléatoires X_i et notée $\bar{X}_n$:

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

On remarque que $\bar{X}_n$ et S_n ne diffèrent que d'un facteur $\frac{1}{n}$. On en déduit

$$\begin{aligned}
\mathbf{E}(\bar{X}_n) &= \mathbf{E}\left(\frac{1}{n} S_n\right) = \frac{1}{n} n\, \mu = \mu \\
\mathbf{V}(\bar{X}_n) &= \mathbf{V}\left(\frac{1}{n} S_n\right) = \frac{1}{n^2} n\, \sigma^2 = \frac{\sigma^2}{n} \\
\sigma(\bar{X}_n) &= \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} \ .
\end{aligned}$$

On a donc

$$S_n = n\, \bar{X}_n, \quad \mathbf{E}(S_n) = n\, \mathbf{E}(\bar{X}_n), \quad \sigma(\bar{S}_n) = n\, \sigma(\bar{X}_n) \ .$$

Ceci donne une autre écriture de Z_n en fonction de $\bar{X}_n$:

$$Z_n = \frac{S_n - \mathbf{E}(S_n)}{\sigma(S_n)} = \frac{n\, \bar{X}_n - n\, \mathbf{E}(\bar{X}_n)}{n\, \sigma(\bar{X}_n)} = \frac{\bar{X}_n - \mathbf{E}(\bar{X}_n)}{\sigma(\bar{X}_n)} \ .$$

On obtient ainsi une autre écriture du TCL :

Théorème 4.2 (TCL pour la variable aléatoire 'moyenne empirique')
Soient $X_1, X_2, \ldots$ des variables aléatoires indépendantes et identiquement distribuées définies sur le même espace probabilisé, d'espérance et de variance communes μ et σ^2 respectivement. Alors

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \simeq \mathbb{N}(0,\, 1) \ .$$

Il ne s'agit pas d'un théorème différent mais du même théorème écrit sous une forme différente. L'utilité de cette écriture est que pour de nombreuses applications, la moyenne empirique apparaît plus naturellement que la somme. C'est le cas par exemple, si l'on considère la variable aléatoire $\hat{p}_n$ qui est la proportion de succès sur n expériences de Bernoulli i.i.d.

Exemple 4.4
Dans une urne, il y a 3 boules rouges et 7 boules noires. On choisit au hasard et successivement 50 boules dans cette urne, en remettant chaque fois la boule tirée dans l'urne. Si l'on suppose l'indépendance des tirages, quelle est la probabilité d'obtenir entre 24 % et 40 % de boules rouges ?
Notons $\bar{X}_{50}$ la v.a. "proportion de boules rouges obtenues" et S_{50} la v.a. "nombre de boules rouges obtenues".

$$\bar{X}_{50} = \frac{1}{50} S_{50}$$

Comme dans l'exemple précédent,
$S_{50} \sim B(50,\ \frac{3}{10})$
$E(S_{50}) = 15$ et $V(S_{50}) = 10,5$.
D'où
$E(\bar{X}_{50}) = \frac{1}{50} E(S_{50}) = 0,3$ et $V(\bar{X}_{50}) = \frac{1}{50^2} E(S_{50}) = 0,0042$.

$$\begin{aligned}
\mathbf{P}(0,24 \leq \bar{X}_{50} \leq 0,4) &= \mathbf{P}\left(\frac{0,24 - 0,3}{\sqrt{0,0042}} \leq \frac{\bar{X}_{50} - 0,3}{\sqrt{0,0042}} \leq \frac{0,4 - 0,3}{\sqrt{0,0042}}\right) \\
&\simeq \mathbf{P}\left(-0,92 \leq Z \leq 1,54\right) \qquad \text{où } Z \sim \mathbb{N}(0,\ 1) \\
&= F_{0,1}(1,54) - F_{0,1}(-0,92)
\end{aligned}$$

en notant $F_{0,1}$ la fonction de répartition de la loi normale centrée–réduite.
D'où la probabilité recherchée :

$$\mathbf{P}(0,24 \leq \bar{X}_{50} \leq 0,40) = 0,9382 - (1 - 0,8212) = 0,7594.$$