Statistique Inférentielle

Mohamad Ghassany

2019-06-20

Contents

Introduction

Cours de Statistique Inférentielle

```r
icon::fa("rocket", size=5, animate="spin", color="#00b4a1")
```

```r
icon::fa("leanpub", size=5, flip="horizontal", color="#00b4a1")
```

```r
icon::fa("book-reader", size=5, color="#00b4a1")
```

Complément de formation Statistique

Variables Aléatoires Discrètes

0.1 Notion de variable aléatoire réelle (v.a.r.)

Après avoir réalisé une expérience aléatoire, il arrive bien souvent qu'on s'intéresse plus à une fonction du résultat qu'au résultat lui-même. Expliquons ceci au moyen des exemples suivants: lorsqu'on joue au dés, certains jeux accordent de l'importance à la somme obtenue sur deux dés, 7 par exemple, plutôt qu'à la question de savoir si c'est la paire (1,6) qui est apparue, ou (2,5), (3,4), (4,3), (5,2) ou plutôt (6,1). Dans le cas du jet d'une pièce, il peut être plus intéressant de connaître le nombre de fois où le côté pile est apparue plutôt que la séquence détaillée des jets pile et face. Ces grandeurs auxquelles on s'intéresse sont en fait des fonctions réelles définies sur l'ensemble fondamental et sont appelées ***variables aléatoires***.

Du fait que la valeur d'une variable aléatoire est déterminée par le résultat de l'expérience, il est possible d'attribuer une probabilité aux différentes valeurs que la variable aléatoire peut prendre.

Soient ε une expérience aléatoire et $(\Omega, \mathcal{A}, P)$ un espace probabilisé lié à cette expérience. Dans de nombreuses situations, on associe à chaque résultat $\omega \in \Omega$ un nombre réel noté $X(\omega)$; on construit ainsi une application $X : \Omega \to \mathbb{R}$. Historiquement, ε était un jeu et X représentait le gain du joueur.

Exemple: Un joueur lance un dé équilibré à 6 faces numérotées de 1 à 6, et on observe le numéro obtenu.

- Si le joueur obtient 1, 3 ou 5, il gagne 1 euro.
- S'il obtient 2 ou 4, il gagne 5 euros.
- S'il obtient 6, il perd 10 euros.

Selon l'expérience aléatoire (lancer d'un dé équilibré) l'ensemble fondamental est $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ et P l'équiprobabilité sur $(\Omega, \mathcal{A})$. Soit X l'application de Ω dans $\mathbb{R}$ qui à tout $\omega \in \Omega$ associe le gain correspondant. On a donc

- $X(1) = X(3) = X(5) = 1$
- $X(2) = X(4) = 5$
- $X(6) = -10$

On dit que X est une **variable aléatoire** sur Ω.

On peut s'intéresser à la probabilité de gagner 1 euro, c'est-à-dire d'avoir $X(\omega) = 1$, ce qui se réalise si et seulement si $\omega \in \{1, 3, 5\}$. La probabilité cherchée est donc égale à $P(\{1, 3, 5\}) = 1/2$. On écrira aussi $P(X = 1) = 1/2$.

On pourra donc considérer l'événement :

$$\{X = 1\} = \{\omega \in \Omega / X(\omega) = 1\} = \{\omega \in \Omega / X(\omega) \in \{1\}\} = X^{-1}(\{1\}) = \{1, 3, 5\}.$$

On aura du même $P(X = 5) = 1/3$ et $P(X = -10) = 1/6$. Ce que l'on peut présenter dans un tableau

x_i	-10	1	5
$p_i = P(X = x_i)$	1/6	1/2	1/3

Cela revient à considérer un nouvel ensemble d'événements élémentaires:

$$\Omega_X = X(\Omega) = \{-10, 1, 5\}$$

et à munir cet ensemble de la probabilité P_X définie par le tableau des $P(X = x_i)$ ci dessus. Cette nouvelle probabilité s'appelle **loi de la variable aléatoire** X.

Remarquer que

$$P(\bigcup_{x_i \in \Omega_X} \{X = x_i\}) = \sum_{x_i \in \Omega_X} P(X = x_i) = 1$$

Dans ce chapitre, nous traitons le cas où $X(\Omega)$ est dénombrable. La variable aléatoire est alors dite ***discrète***. Sa loi de probabilité, qui peut être toujours définie par sa fonction de répartition, le sera plutôt par les probabilités individuelles. Nous définirons les deux caractéristiques numériques principales d'une variable aléatoire discrète, l'espérance caractéristique de valeur centrale, et la variance, caractéristique de dispersion. Nous définirons aussi les couples de variables aléatoires.

0.2 Variables aléatoires discrètes

Définition, loi de probabilité

Définition 0.1. On dit qu'une variable aléatoire réelle (v.a.r.) X est ***discrète*** (v.a.r.d.) si l'ensemble des valeurs que prend X est fini ou infini dénombrable.

Si on suppose $X(\Omega)$ l'ensemble des valeurs de X qui admet un plus petit élément x_1. Alors la v.a.r.d. X est entièrement définie par:

- L'ensemble $X(\Omega)$ des valeurs prises par X, rangées par ordre croissant: $X(\Omega) = \{x_1, x_2, \ldots, x_i, \ldots\}$ avec $x_1 \leq x_2 \leq \ldots \leq x_i \leq \ldots$.
- La ***loi de probabilité*** définie sur $X(\Omega)$ par

$$p_i = P(X = x_i) \quad \forall\ i = 1, 2, \ldots$$

Remarques:

- Soit B un ensemble de $\mathbb{R}$,

$$P(X \in B) = \sum_{i / x_i \in B} p(x_i)$$

- En particulier

$$P(a < X \leq b) = \sum_{i / a < x_i \leq b} p(x_i)$$

- Bien sûr tous les $p(x_i)$ sont positives et $\sum_{i=1}^{\infty} p(x_i) = 1$.
- Si X ne prend qu'un petit nombre de valeurs, cette loi est généralement présentée dans un tableau.

Fonction de répartition d'une variable aléatoire discrète

Définition 0.2. On appelle *fonction de répartition* de la v.a. X, qu'on note $F(a)$ de la v.a.r.d. X, ou $F_X(a)$, la fonction définie pour tout réel a, $-\infty < a < \infty$, par

$$F(a) = P(X \leq a) = \sum_{i/x_i \leq a} P(X = x_i)$$

Cette valeur représente la probabilité de toutes les réalisations inférieures ou égales au réel a.

Propriétés: Voici quelques propriétés de cette fonction:

1. C'est une fonction en escalier (constante par morceaux).
2. $F(a) \leq 1$ car c'est une probabilité.
3. $F(a)$ est continue à droite.
4. $\lim_{a \to -\infty} F(a) = 0$ et $\lim_{a \to \infty} F(a) = 1$

La fonction de répartition caractérise la loi de X, autrement dit: $F_X = F_Y$ si et seulement si les variables aléatoires X et Y ont la même loi de probabilité.

Fonction de répartition et probabilités sur X

Tous les calculs de probabilité concernant X peuvent être traités en termes de fonction de répartition. Par exemple,

$$P(a < X \leq b) = F(b) - F(a) \qquad \text{pour tout } a < b$$

On peut mieux s'en rendre compte en écrivant $\{X \leq b\}$ comme union des deux événements incompatibles $\{X \leq a\}$ et $\{a < X \leq b\}$, soit

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$$

et ainsi

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

ce qui établit l'égalité ci dessus.

On peut déduire de F les probabilités individuelles par:

$$p_i = F(x_i) - F(x_{i-1}) \qquad \text{pour } 1 \leq i \leq n$$

Exemple: On joue trois fois à pile ou face. Soit X la variable aléatoire "nombre de pile obtenus". Ici $\Omega = \{P, F\}^3$, et donc

$$X(\Omega) = \{0, 1, 2, 3\}.$$

On a $card(\Omega) = 2^3 = 8$. Calculons par exemple $P(X = 1)$, c'est à dire la probabilité d'avoir exactement une pile.

$$X^{-1}(1) = \{(P, F, F), (F, P, F), (F, F, P)\}$$

D'où $P(X=1)=\frac{3}{8}$.

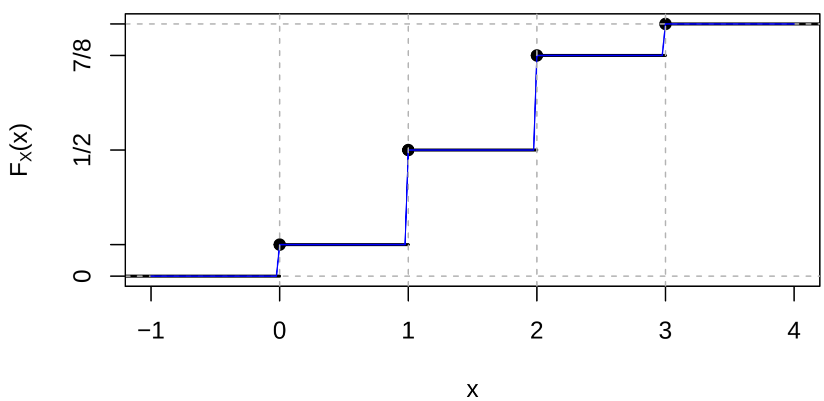
En procédant de la même façon, on obtient la loi de probabilité de X:

k	0	1	2	3
$P(X=k)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

La fonction de répartition de X est donc donnée par:

$$F(x)=\begin{cases} 0 & \text{si } x<0 \\ 1/8 & \text{si } 0\leq x<1 \\ 1/2 & \text{si } 1\leq x<2 \\ 7/8 & \text{si } 2\leq x<3 \\ 1 & \text{si } x\geq 3 \end{cases}$$

Le graphe de cette dernière est représentée dans la figure suivante:



Exemple: Soit A un événement quelconque. On appelle variable aléatoire *indicatrice* de cet événement A, la variable aléatoire définie par:

$$X(\omega)=\begin{cases} 1 & \text{si } \omega\in A \\ 0 & \text{si } \omega\in \bar{A} \end{cases}$$

et notée $X=1_A$. Ainsi:

$$P(X=1)=P(A)=p$$
$$P(X=0)=P(\bar{A})=1-p$$

La fonction de répartition de X est donc donnée par:

$$F(x)=\begin{cases} 0 & \text{si } x<0 \\ 1-p & \text{si } 0\leq x<1 \\ 1 & \text{si } x\geq 1 \end{cases}$$

On peut prendre par exemple le cas d'un tirage d'une boule dans une urne contenant 2 boules blanches et 3 boules noires. Soit A : "obtenir une boule blanche" et X la variable indicatrice de A. La loi de probabilité de X est alors

k	0	1
$P(X = k)$	$\frac{3}{5}$	$\frac{2}{5}$

et sa fonction de répartition est:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 3/5 & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

0.3 Moments d'une variable aléatoire discrète

Espérance mathématique

Définition 0.3. Pour une variable aléatoire discrète X de loi de probabilité $p(.)$, on définit l'*espérance* de X, notée $E(X)$, par l'expression

$$E(X) = \sum_{i \in \mathbb{N}} x_i p(x_i)$$

En termes concrets, l'espérance de X est la moyenne pondérée des valeurs que X peut prendre, les poids étant les probabilités que ces valeurs soient prises.

Reprenons l'exemple où on joue 3 fois à pile ou face. L'espérance de X ="nombre de pile obtenus" est égal à:

$$E(X) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = 1.5$$

Dans le cas de la loi uniforme sur $X(\Omega) = \{x_1, \ldots, x_k\}$, c'est à dire avec équiprobabilité de toutes les valeurs $p_i = 1/k$, on obtient:

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

et dans ce cas $E(X)$ se confond avec la moyenne arithmétique simple $\bar{x}$ des valeurs possibles de X.

Pour le jet d'un dé équilibré par exemple:

$$E(X) = \frac{1}{6} \sum_{i=1}^{6} i = \frac{7}{2} = 3.5$$

Espérance d'une fonction d'une variable aléatoire

Théorème 0.1 (Théorème du transfert) Si X est une variable aléatoire discrète pouvant prendre ses valeurs parmi les valeurs x_i, $i \geq 1$, avec des probabilités respectives $p(x_i)$, alors pour toute fonction réelle g on a

$$E(g(X)) = \sum_{i} g(x_i) p(x_i)$$

Exemple 0.1. Soit X une variable aléatoire qui prend une des trois valeurs $\{-1, 0, 1\}$ avec les probabilités respectives

$$P(X=-1)=0.2 \qquad P(X=0)=0.5 \qquad P(X=1)=0.3$$

Calculer $E(X^2)$.
Solution: . *Première approche*: Soit $Y=X^2$. La distribution de Y est donnée par

$$\begin{aligned} P(Y=1) &= P(X=-1)+P(X=1)=0.5 \\ P(Y=0) &= P(X=0)=0.5 \end{aligned}$$

Donc

$$E(X^2)=E(Y)=1(0.5)+0(0.5)=0.5$$

Deuxième approche: En utilisant le théorème

$$\begin{aligned} E(X^2) &= (-1)^2(0.2)+0^2(0.5)+1^2(0.3) \\ &= 1(0.2+0.3)+0(0.5)=0.5 \end{aligned}$$

Remarquer que

$$0.5=E(X^2)\neq (E(X))^2=0.01$$

Linéarité de l'espérance Propriétés de l'espérance

1. $E(X+a)=E(X)+a, \quad a\in\mathbb{R}$
   résultat qui se déduit de:
   $$\sum_i p_i(x_i+a)=\sum_i p_ix_i+\sum_i ap_i=\sum_i p_ix_i+a\sum_i p_i=\sum_i p_ix_i+a$$
2. $E(aX)=aE(X), \quad a\in\mathbb{R}$
   il suffit d'écrire:
   $$\sum_i p_iax_i=a\sum_i p_ix_i$$
3. $E(X+Y)=E(X)+E(Y)$, X et Y étant deux variables aléatoire.

On peut résumer ces trois propriétés en disant que l'espérance mathématique est linéaire:

$$E(\lambda X+\mu Y)=\lambda E(X)+\mu E(Y), \quad \forall\lambda\in\mathbb{R}, \forall\mu\in\mathbb{R}.$$

Variance

Définition 0.4. La variance est un indicateur mesurant la dispersion des valeurs x_i que peut prendre la v.a. X et son espérance $E(X)$. On appelle **variance** de X, que l'on note $V(X)$, la quantité

$$V(X)=E\big[(X-E(X))^2\big]$$

lorsque cette quantité existe.
C'est l'espérance mathématique du carré de la v.a. centrée $X-E(X)$.

On peut établir une autre formule pour le calcul de $V(X)$:

$$V(X)=E(X^2)-E^2(X)$$

Or:

$$\begin{aligned} V(X) &= E\left[X^2 - 2XE(X) + E^2(X)\right] \\ &= E(X^2) - E[2XE(X)] + E[E^2(X)] \\ &= E(X^2) - 2E^2(X) + E^2(X) \\ &= E(X^2) - E^2(X) \end{aligned}$$

On cherche $V(X)$ où X est le nombre obtenu lors du jet d'un dé équilibré. On a vu dans l'exemple que $E(X) = \frac{7}{2}$. De plus,

$$\begin{aligned} E(X^2) &= 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right) \\ &= \left(\frac{1}{6}\right)(91) = \frac{91}{6}. \end{aligned}$$

Et donc

$$V(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

Propriétés de la variance

1. $V(X) \geq 0$
2. $V(X + a) = V(X)$
   en effet:
   $$\begin{aligned} V(X+a) &= E\big[\,[X + a - E(X+a)]^2\,\big] \\ &= E\big[\,[X + a - E(X) - a]^2\,\big] \\ &= E\big[\,[X - E(X)]^2\,\big] \\ &= V(X). \end{aligned}$$
3. $V(aX) = a^2V(X)$
   en effet:
   $$\begin{aligned} V(aX) &= E\big[\,[aX - E(aX)]^2\,\big] \\ &= E\big[\,[aX - aE(X)]^2\,\big] \\ &= E\big[a^2\,[X - E(X)]^2\big] \\ &= a^2\big[E\,[X - E(X)]^2\big] \\ &= a^2V(X). \end{aligned}$$

Ecart-type

Définition 0.5. La racine carrée de $V(X)$ est appelée l'***écart-type*** de X, qui se note σ_X. On a

$$\sigma_X = \sqrt{V(X)}$$

σ_X s'exprime dans les mêmes unités de mesure que la variable aléatoire X.

A noter:

- L'écart type sert à mesurer la dispersion d'un ensemble de données.
- Plus il est faible, plus les valeurs sont regroupées autour de la moyenne.

- Exemple: La répartition des notes d'une classe. Plus l'écart type est faible, plus la classe est homogène.
- L'espérance et l'écart-type sont reliés par l'*inégalité de Bienaymé-Tchebychev*.

Inégalité de Bienaymé-Tchebychev

Théorème 0.2 Soit X une variable aléatoire d'espérance μ et de variance σ^2. Pour tout $\varepsilon > 0$, on a l'inégalité suivante:

$$P\left(|X - E(X)| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2}$$

On peut l'écrire autrement. Soit $k = \varepsilon/\sigma$.

$$P\left(|X - E(X)| \geq k\sigma\right) \leq \frac{1}{k^2}$$

Importance: Cette inégalité relie la probabilité pour X de s'écarter de sa moyenne $E(X)$, à sa variance qui est justement un indicateur de dispersion autour de la moyenne de la loi. Elle montre quantitativement que “plus l'écart type est faible, plus la probabilité de s'écarter de la moyenne est faible”.

Théorème 0.3 (Inégalité de Markov) Soit X une variable aléatoire à valeur non négatives. Pour tout réel $a > 0$

$$P(X > a) \leq \frac{E(X)}{a}$$

Moments non centrés et centrés

On appelle moment non centré d'ordre $r \in \mathbb{N}^*$ de X la quantité, lorsqu'elle existe:

$$m_r(X) = \sum_{i \in \mathbb{N}} x_i^r p(x_i) = E(X^r).$$

Le moment centré d'ordre $r \in \mathbb{N}^*$ est la quantité, lorsqu'elle existe:

$$\mu_r(X) = \sum_{i \in \mathbb{N}} p_i \left[x_i - E(X)\right]^r = E\left[X - E(X)\right]^r.$$

Les premiers moments sont:

$$m_1(X) = E(X), \quad \mu_1(X) = 0$$

$$\mu_2(X) = V(X) = m_2(X) - m_1^2(X)$$

0.4 Couple de variables aléatoires discrètes

Considérons deux variables aléatoires discrètes X et Y. Il nous faut pour modéliser le problème une fonction qui nous donne la probabilité que $(X = x_i)$ en même temps que $(Y = y_j)$. C'est la loi de probabilité conjointe.

Soit X et Y deux variables aléatoires réelles discrètes, définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ et que

$$X(\Omega) = \{x_1, x_2, \dots, x_l\}$$
$$Y(\Omega) = \{y_1, y_2, \dots, y_k\}$$
$$(l \text{ et } k \in \mathbb{N})$$

La ***loi du couple*** (X, Y), dite **loi de probabilité conjointe ou simultanée**, est entièrement définie par les probabilités:

$$p_{ij} = P(X = x_i; Y = y_j) = P(\{X = x_i\} \cap \{Y = y_j\})$$

On a

$$p_{ij} \geq 0 \quad \text{et} \quad \sum_{i=1}^{l} \sum_{j=1}^{k} p_{ij} = 1$$

Le couple (X, Y) s'appelle variable aléatoire à deux dimensions et peut prendre $l \times k$ valeurs.

Table de probabilité conjointe

Les probabilités p_{ij} peuvent être présentées dans un tableau à deux dimensions qu'on appelle table de probabilité conjointe:

Table 4: Table de probabilité conjointe

$X \backslash Y$	y_1	y_2	$\dots$	y_j	$\dots$	y_k
x_1	p_{11}	p_{12}		p_{1j}		p_{1k}
x_2	p_{21}	p_{22}		p_{2j}		p_{2k}
$\vdots$						
x_i	p_{i1}	p_{i2}		p_{ij}		p_{ik}
$\vdots$						
x_l	p_{l1}	p_{l2}		p_{lj}		p_{lk}

A la première ligne figure l'ensemble des valeurs de Y et à la première colonne figure l'ensemble des valeurs de X. La probabilité $p_{ij} = P(X = x_i; Y = y_j)$ est à l'intersection de la i^e et de la j^e colonne.

Lois marginales

Lorsqu'on connaît la loi conjointe des variables aléatoires X et Y, on peut aussi s'intéresser à la loi de probabilité de X seule et de Y seule. Ce sont les lois de probabilité marginales.

- Loi marginale de X:

$$p_{i.} = P(X = x_i) = P[\{X = x_i\} \cap \Omega] = \sum_{j=1}^{k} p_{ij} \qquad \forall\, i = 1, 2, \dots, l$$

- Loi marginale de Y:

$$p_{.j} = P(Y = y_j) = P[\Omega \cap \{Y = y_j\}] = \sum_{i=1}^{l} p_{ij} \qquad \forall\, j = 1, 2, \dots, k$$

On peut calculer les lois marginales directement depuis la table de la loi conjointe. La loi marginale de X est calculée en faisant les totaux par ligne, tandis que celle de Y l'est en faisant les totaux par colonne.

C'est le fait que les lois de X et Y individuellement puissent être lues dans les marges du tableau qui leur vaut leur nom de lois marginales.

Table 5: Table de probabilité conjointe avec les lois marginales

$X \backslash Y$	y_1	y_2	$\dots$	y_j	$\dots$	y_k	Marginale de X
x_1	p_{11}	p_{12}		p_{1j}		p_{1k}	$p_{1.}$
x_2	p_{21}	p_{22}		p_{2j}		p_{2k}	$p_{2.}$
$\vdots$							
x_i	p_{i1}	p_{i2}		p_{ij}		p_{ik}	$p_{i.}$
$\vdots$							
x_l	p_{l1}	p_{l2}		p_{lj}		p_{lk}	$p_{l.}$
Marginale de Y	$p_{.1}$	$p_{.2}$		$p_{.l}$		$p_{.k}$	1

On tire au hasard 3 boules d'une urne contenant 3 boules rouges, 4 blanches et 5 noires. X et Y désignent respectivement le nombre de boules rouges et celui de boules blanches tirées. Déterminer la loi de probabilité conjointe du couple (X, Y) ainsi que les lois marginales de X et de Y.

Lois conditionnelles

Pour chaque valeur y_j de Y telle que $p_{.j} = P(Y = y_j) \neq 0$ on peut définir la loi conditionnelle de X sachant $Y = y_j$ par

$$p_{i/j} = P(X = x_i / Y = y_j) = \frac{P(X = x_i; Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{.j}} \qquad \forall i = 1, 2, \dots, l$$

De même on définit la loi de Y sachant $X = x_i$ par

$$p_{j/i} = P(Y = y_j / X = x_i) = \frac{P(X = x_i; Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i.}} \qquad \forall j = 1, 2, \dots, k$$

Indépendance de variables aléatoires

Théorème 0.4 On dit que deux v.a.r.d sont indépendantes si et seulement si

$$P(X = x_i; Y = y_j) = P(X = x_i)P(Y = y_j) \qquad \forall\, i = 1, 2, \dots, l \text{ et } j = 1, 2, \dots, k$$

On montre que

$$P(\{X \in A\} \cap \{Y \in B\}) = P(\{X \in A\})P(\{Y \in B\}) \qquad \forall\ A \text{ et } B \in \mathcal{A}$$

Propriétés

Soit deux v.a.r.d. X et Y,

1. $E(X+Y) = E(X) + E(Y)$
2. Si X et Y sont indépendantes alors $E(XY) = E(X)E(Y)$. Mais la réciproque n'est pas toujours vraie.

Covariance

Soit X et Y deux v.a.r.d. On appelle ***covariance*** de X et de Y la valeur si elle existe de

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))] = \sum_i \sum_j (x_i - E(X))(y_j - E(Y))p_{ij}$$

qu'on peut calculer en utilisant la formule suivante

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Propriétés

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(aX_1 + bX_2, Y) = aCov(X_1,Y) + bCov(X_2,Y)$
- $V(X+Y) = V(X) + V(Y) + 2Cov(X,Y)$
- Si X et Y sont indépendantes alors
  - $Cov(X,Y) = 0$ (la réciproque n'est pas vraie)
  - $V(X+Y) = V(X) + V(Y)$ (la réciproque n'est pas vraie)

Coefficient de corrélation linéaire

On appelle coefficient de corrélation linéaire de X et de Y la valeur définie par

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

On peut montrer que

$$-1 \leq \rho(X,Y) \leq 1$$

Pour le montrer on peut partir du fait que la variance est toujours positive ou nulle. Donc $V(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}) \geq 0$ et $V(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}) \geq 0$.

Interprétation de ρ

- Le coefficient de corrélation est une mesure du degré de linéarité entre X et Y.
- Les valeurs de ρ proches de 1 ou -1 indiquent une linéarité quasiment rigoureuse entre X et Y.
- Les valeurs de ρ proche de 0 indiquent une absence de toute relation linéaire.
- Lorsque $\rho(X,Y)$ est positif, Y a tendance à augmenter si X en fait autant.
- Lorsque $\rho(X,Y) < 0$, Y a tendance à diminuer si X augmente.
- Si $\rho(X,Y) = 0$, on dit que ces deux statistiques sont non corrélées.

0.5 Lois usuelles discrètes

0.5.1 Loi uniforme discrète $\mathcal{U}(n)$

Définition 0.6. Une distribution de probabilité suit une loi uniforme lorsque toutes les valeurs prises par la variable aléatoire sont équiprobables. Si n est le nombre de valeurs différentes prises par la variable aléatoire alors on a:

$$P(X = x_i) = \frac{1}{n} \qquad \forall\, i \in \{1, \ldots, n\}$$

Exemple: La distribution des chiffres obtenus au lancer de dé (si ce dernier est non pipé) suit une loi uniforme dont la loi de probabilité est la suivante :

x_i	1	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Moments de loi uniforme discrète

Dans le **cas particulier** d'une loi uniforme discrète où chaque valeur de la variable aléatoire X correspond à son rang, i.e. $x_i = i \ \forall i \in \{1, \ldots, n\}$, on a:

$$E(X) = \frac{n+1}{2} \quad \text{et} \quad V(X) = \frac{n^2 - 1}{12}$$

La démonstration de ces résultats est établie en utilisant les égalités (cf. Annexe)

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \quad \text{et} \quad \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}.$$

En revenant à l'exemple du lancer du dé de cette section, on peut calculer directement les moments de X:

$$E(X) = \frac{6+1}{2} = 3.5$$

et

$$V(X) = \frac{6^2 - 1}{12} = \frac{35}{12} \simeq 2.92.$$

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

Définition 0.7. On réalise une expérience dont le résultat sera interprété soit comme un succès soit comme un échec. On définit alors la variable aléatoire X en lui donnant la valeur 1 lors d'un succès et 0 lors d'un échec (variable indicatrice). La loi de probabilité de X est alors

$$\begin{aligned} p(1) &= P(X = 1) = p \\ p(0) &= P(X = 0) = 1 - p = q \end{aligned} \tag{1}$$

où p est la probabilité d'un succès, $0 \leq p \leq 1$.

Une variable aléatoire X est dite de **Bernoulli** $X \sim \mathcal{B}(p)$ s'il existe un nombre $p \in]0, 1[$ tel que la loi de probabilité de X soit donnée par (1).

La fonction de répartition est définie par:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1-p & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1. \end{cases}$$

L'espérance la loi de Bernoulli est p, en effet

$$E(X) = 1 \times P(X=1) + 0 \times P(X=0) = P(X=1) = p$$

La variance la loi de Bernoulli est np, en effet

$$V(X) = E(X^2) - E^2(X) = p - p^2 = p(1-p) = pq$$

car

$$E(X^2) = 1^2 \times P(X=1) + 0^2 \times P(X=0) = P(X=1) = p$$

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

Décrite pour la première fois par *Isaac Newton* en 1676 et démontrée pour la première fois par le mathématicien suisse *Jacob Bernoulli* en 1713, la loi binomiale est l'une des distributions de probabilité les plus fréquemment rencontrées en statistique appliquée.

Supposons qu'on exécute maintenant n épreuves **indépendantes**, chacune ayant p pour probabilité de succès et $1-p$ pour probabilité d'échec. La variable aléatoire X qui compte **le nombre de succès** sur l'ensemble des n épreuves est dite variable aléatoire **binomiale** de paramètres n et p.

Une variable de Bernoulli n'est donc qu'une variable binomiale de paramètres $(1,p)$.

Définition 0.8. Si on effectue n épreuves successives indépendantes où on note à chaque fois la réalisation ou non d'un certain événement A, on obtient une suite de la forme $AA\bar{A}A\bar{A}\ldots\bar{A}AA$. Soit X le nombre de réalisations de A. On définit ainsi une v.a. X qui suit une loi binomiale de paramètres n et $p = P(A)$, caractérisée par $X(\Omega) = \{0, 1, \ldots, n\}$:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad 0 \leq k \leq n \tag{2}$$

On écrit $X \sim \mathcal{B}(n,p)$. Donc la loi binomiale modélise le nombre de réalisations de A (succès) obtenues lors de la répétition indépendante et identique de n épreuves de Bernoulli.

Pour établir (2) il faut remarquer que $\binom{n}{k}$ est le nombre d'échantillons de taille n comportant exactement k événements A, de probabilité p^k, indépendamment de l'ordre, et donc $n-k$ événements $\bar{A}$, de probabilité $(1-p)^{n-k}$.

Remarque: Il est possible d'obtenir aisément les valeurs des combinaisons de la loi binomiale en utilisant le triangle de Pascal.

En utilisant la formule du binôme de Newton, on vérifie bien que c'est une loi de probabilité:

$$\sum_{k=0}^{n} P(X=k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = [p + (1-p)]^n = 1$$

Exemple: On jette cinq pièces équilibrées. Les résultats sont supposés indépendants. Donner la loi de probabilité de la variable X qui compte le nombre de piles obtenus.

Moments de la loi Binomiale

Pour calculer facilement les moments de cette loi, nous allons associer à chaque épreuve i, $1 \leq i \leq n$, une v.a. de Bernoulli (variable indicatrice sur A):

$$1_A = X_i = \begin{cases} 1 & \text{si } A \text{ est réalisé} \\ 0 & \text{si } \bar{A} \text{ est réalisé} \end{cases}$$

On peut écrire alors: $X = \sum_{i=1}^{n} X_i = X_1 + X_2 + \ldots + X_n$, ce qui nous permet de déduire aisément:

$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = np$$

et

$$V(X) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) = np(1-p) \quad \text{car les v.a. } X_i \text{ sont indépendantes.}$$

Le calcul direct des moments de X peut s'effectuer à partir de la définition générale, mais de façon beaucoup plus laborieuse:

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} = np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\
&= np \sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} p^j (1-p)^{n-1-j} = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \\
&= np[p + (1-p)]^{n-1} = np
\end{aligned}$$

Pour obtenir $E(X^2)$ par un procédé de calcul identique, on passe par l'intermédiaire du moment factoriel $E[X(X-1)] = E(X^2) - E(X)$:

$$\begin{aligned}
E[X(X-1)] &= \sum_{k=0}^{n} k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} (1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j} \\
&= n(n-1)p^2 [p + (1-p)]^{n-2} = n(n-1)p^2
\end{aligned}$$

On en déduit alors:

$$E(X^2) = E[X(X-1)] + E(X) = n(n-1)p^2 + np,$$

puis:

$$\begin{aligned}
V(X) &= n(n-1)p^2 + np - (np)^2 \\
&= n^2p^2 + np(1-p) - n^2p^2 \\
&= np(1-p).
\end{aligned}$$

Le nombre de résultats pile apparus au cours de n jets d'une pièce de monnaie suit une loi binomiale $\mathcal{B}(n, 1/2)$:

$$P(X = k) = \binom{n}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{n-k} = \frac{\binom{n}{k}}{2^n}, \quad 0 \leq k \leq n$$

avec $E(X) = n/2$ et $V(X) = n/4$.

Le nombre N de boules rouges apparues au cours de n tirages avec remise dans une urne contenant deux rouges, trois vertes et une noire suit une loi binomiale $\mathcal{B}(n, 1/3)$:

$$P(N = k) = \binom{n}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{n-k} = \binom{n}{k}\frac{2^{n-k}}{3^n}, \quad 0 \leq k \leq n$$

avec $E(X) = n/3$ et $V(X) = 2n/9$.

Théorème 0.5 Si $X_1 \sim \mathcal{B}(n_1, p)$ et $X_2 \sim \mathcal{B}(n_2, p)$, les v.a. X_1 et X_2 étant indépendantes, alors $X_1 + X_2 \sim \mathcal{B}(n_1 + n_2, p)$. Ceci résulte de la définition d'une loi binomiale puisqu'on totalise ici le résultat de $n_1 + n_2$ épreuves indépendantes.

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

La loi de Poisson est découverte au début du XIX^e siècle par le magistrat français *Siméon-Denis Poisson*. Les variables aléatoires de Poisson ont un champ d'application fort vaste, en particulier du fait qu'on peut les utiliser pour approximer des variables aléatoires binomiales de paramètres (n, p) pour autant que n soit grand et p assez petit pour que np soit d'ordre de grandeur moyen.

Définition 0.9. Une v.a. X suit une loi de Poisson de paramètre $\lambda > 0$ si c'est une variable à valeurs entières, $X(\Omega) = \mathbb{N}$, donc avec une infinité de valeurs possibles, de probabilité:

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

Cette loi ne dépend qu'un seul paramètre réel positif λ, avec l'écriture symbolique $X \sim \mathcal{P}(\lambda)$.

Le développement en série entière de l'exponentielle $e^\lambda = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!}$ permet de vérifier qu'il s'agit bien d'une loi de probabilité:

$$\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1$$

Moments de loi de Poisson

Le calcul de l'espérance mathématique se déduit du développement en série entière de l'exponentielle:

$$\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} kP(X = k) = \sum_{k=1}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = \lambda e^{-\lambda}e^{\lambda} \\
&= \lambda.
\end{aligned}$$

Pour calculer la variance nous n'allons pas calculer $E(X^2)$ mais le moment factoriel $E[X(X-1)]$ qui s'obtient plus facilement, selon la méthode précédente:

$$\begin{aligned}
E[X(X-1)] &= \sum_{k=0}^{\infty} k(k-1)P(X=k) = \sum_{k=2}^{\infty} k(k-1)\, e^{-\lambda}\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)!} = \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} \\
&= \lambda^2 e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = \lambda^2 e^{-\lambda}e^{\lambda} = \lambda^2.
\end{aligned}$$

On en déduit:

$$\begin{aligned}
V(X) &= E(X^2) - E^2(X) = E[X(X-1)] + E(X) - E^2(X) \\
&= \lambda^2 + \lambda - \lambda^2 = \lambda.
\end{aligned}$$

Théorème 0.6 Si X et Y sont deux variables **indépendantes** suivant des lois de Poisson

$$X \sim \mathcal{P}(\lambda) \quad \text{et} \quad Y \sim \mathcal{P}(\mu)$$

alors leur somme suit aussi une loi de Poisson:

$$X + Y \sim \mathcal{P}(\lambda + \mu).$$

Exemple: Soit X la variable aléatoire associée au nombre de micro-ordinateurs vendus chaque jour dans le magasin. On suppose que X suit une loi de Poisson de paramètre $\lambda = 5$. On écrit alors $X \sim \mathcal{P}(5)$.
La probabilité associée à la vente de 5 micro-ordinateurs se détermine par :

$$P(X=5) = e^{-5}\frac{5^5}{5!} = e^{-5} \simeq 0.1755$$

La probabilité de vendre au moins 2 micro-ordinateurs est égal à:

$$P(X \geq 2) = 1 - \left(e^{-5}\frac{5^0}{0!} + e^{-5}\frac{5^1}{1!}\right) \simeq 0.9596$$

Le nombre moyen de micro-ordinateurs vendus chaque jour dans le magasin est égal à 5 puisque $E(X) = \lambda = 5$.

Approximation d'une loi binomiale

Le théorème de Poisson nous montre que si n est suffisamment grand et p assez petit, alors on peut approcher la distribution d'une loi binomiale de paramètres n et p par celle d'une loi de Poisson de paramètre $\lambda = np$, en effet

$$\text{si } n \to \infty \text{ et } p \to 0 \text{ alors } X : \mathcal{B}(n,p) \to \mathcal{P}(\lambda).$$

Une bonne approximation est obtenue si $n \geq 50$ et $np \leq 5$.

Dans ce contexte, la loi de Poisson est souvent utilisée pour modéliser le nombre de succès lorsqu'on répète un très grand nombre de fois une expérience ayant une chance très faible de réussir par une loi de Poisson (nombre de personnes dans la population française atteints d'une maladie rare, par exemple).

On cherche la probabilité de trouver au moins un centenaire parmi 200 personnes dans une population où une personne sur cent est un centenaire.

La probabilité $p = 1/100 = 0.01$ étant faible et $n = 200$ étant suffisamment grand, on peut modéliser le nombre X de centenaires pris parmi 200 personnes par la loi de Poisson de paramètre $\lambda = 200 \times 0.01 = 2$. Donc on a:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-2} \simeq 0.86$$

Soit une v.a. X telle que $X \sim \mathcal{B}(100, 0.01)$, les valeurs des probabilités pour k de 0 à 5 ainsi que leur approximation à 10^{-3} avec une loi de Poisson de paramètre $\lambda = np = 1$ sont données dans le tableau ci-dessous :

k	0	1	2	3	4	5
$P(X = k)$	0.366	0.370	0.185	0.061	0.015	0.000
Approximation	0.368	0.368	0.184	0.061	0.015	0.003

Dans le cas de cet exemple où $n = 100$ et $np = 1$, l'approximation de la loi binomiale par une loi de poisson donne des valeurs de probabilités identiques à 10^{-3} près.

Loi Géométrique ou de Pascal $\mathcal{G}(p)$

On effectue des épreuves successives indépendantes jusqu'à la réalisation d'un événement particulier A de probabilité $p = P(A)$ et on note X le nombre aléatoire d'épreuves effectuées. On définit ainsi une v.a. à valeurs entières de loi géométrique, ou de Pascal. A chaque épreuve est associé l'ensemble fondamental $\Omega = \{A, \bar{A}\}$ et l'événement $\{X = k\}$ pour $k \in \mathbb{N}^*$ est représenté par une suite de $k - 1$ événements $\bar{A}$, terminée par l'événement A:

$$\underbrace{\bar{A}\bar{A}\ldots\bar{A}}_{k-1}A$$

D'où:

$$P(X = k) = (1 - p)^{k-1}p \quad \forall\, k \in \mathbb{N}^* \tag{3}$$

Cette loi peut servir à modéliser des temps de vie, ou des temps d'attente, lorsque le temps est mesuré de manière discrète (nombre de jours par exemple).

En utilisant la série entière

$$\sum_{k=0}^{\infty} x^k = 1/(1-x) \quad \text{pour} \quad |x| < 1$$

on vérifie bien que c'est une loi de probabilité:

$$\begin{aligned}\sum_{k=1}^{\infty} P(X = k) = \sum_{k=1}^{\infty}(1-p)^{k-1}p = p\sum_{j=0}^{\infty}(1-p)^j \\ = p\frac{1}{1-(1-p)} = 1\end{aligned}$$

Moments de loi Géométrique

En dérivant la série entière (3) ci-dessus, on obtient $\sum_{k=1}^{\infty} kx^{k-1} = 1/(1-x)^2$. Ceci permet d'obtenir l'espérance:

$$E(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = \frac{p}{[1-(1-p)]^2} = \frac{1}{p}$$

- En d'autres termes, si des épreuves indépendantes ayant une probabilité p d'obtenir un succès sont réalisés jusqu'à ce que le premier succès se produise, le nombre espéré d'essais nécessaires est égal à $1/p$. Par exemple, le nombre espéré de jets d'un dé équilibré qu'il faut pour obtenir la valeur 1 est 6.

Le calcul de la variance se fait à partir du moment factoriel et en utilisant la dérivée seconde de la série entière (3): $\sum_{k=2}^{\infty} k(k-1)x^{k-2} = 2/(1-x)^3$, Donc

$$\begin{aligned} E[X(X-1)] &= \sum_{k=2}^{\infty} k(k-1)p(1-p)^{k-1} \\ &= p(1-p)\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} \\ &= \frac{2p(1-p)}{[1-(1-p)]^3} = \frac{2(1-p)}{p^2} \end{aligned}$$

d'où on déduit:

$$V(X) = E[X(X-1)] + E(X) - E^2(X) = \frac{1-p}{p^2}.$$

Si l'on considère la variable aléatoire X "nombre de naissances observées jusqu'à l'obtention d'une fille" avec p = 1/2 (même probabilité de naissance d'une fille ou d'un garçon), alors X suit une loi géométrique et on a pour tout $k \in \mathbb{N}^*$:

$$P(X = k) = (1 - 1/2)^{k-1}(1/2) = 1/2^k$$

avec $E(X) = 2$ et $V(X) = 2$.

Variables Aléatoires Continues

0.6 Densité d'une variable aléatoire continue

Dans les chapitres précédents nous avons traité des variables aléatoires discrètes, c'est-à-dire de variables dont l'univers est fini ou infini dénombrable. Il existe cependant des variables dont l'univers est infini non dénombrable. On peut citer par exemple, l'heure d'arrivée d'un train à une gare donnée ou encore la durée de vie d'un transistor. Désignons par X une telle variable.

Définition 0.10. X est une **variable aléatoire continue** s'il existe une fonction f non négative définie pour tout $x \in \mathbb{R}$ et vérifiant pour tout ensemble B de nombres réels la propriété

$$P(X \in B) = \int_B f(x)dx \tag{4}$$

La fonction f est appelée **densité de probabilité** de la variable aléatoire X.

Tous les problèmes de probabilité relatifs à X peuvent être traités grâce à f. Par exemple pour $B = [a, b]$, on obtient grâce à l'équation (4)

$$P(a \leq X \leq b) = \int_a^b f(x)dx \tag{5}$$

Graphiquement, $P(a \leq X \leq b)$ est l'aire de la surface entre l'axe de x, la courbe correspondante à $f(x)$ et les droites $x = a$ et $x = b$. Voire Figure 1 et Figure 2.

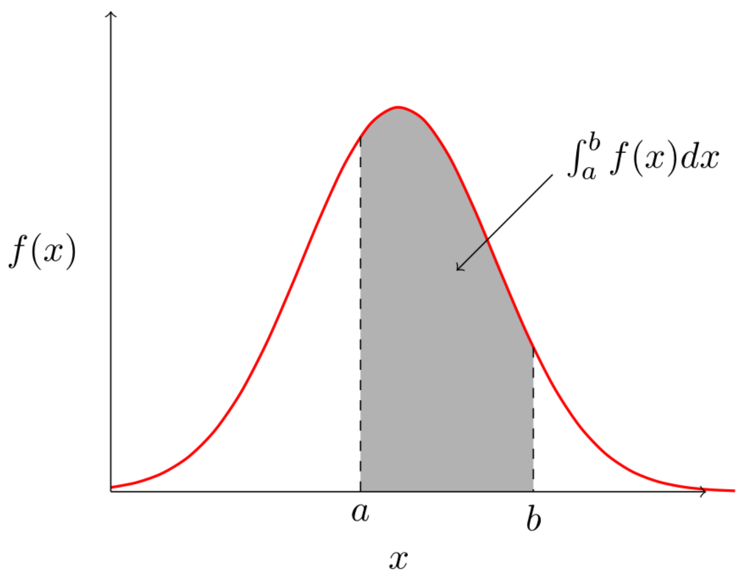


Figure 1: $P(a \leq X \leq B)$ = surface grisée

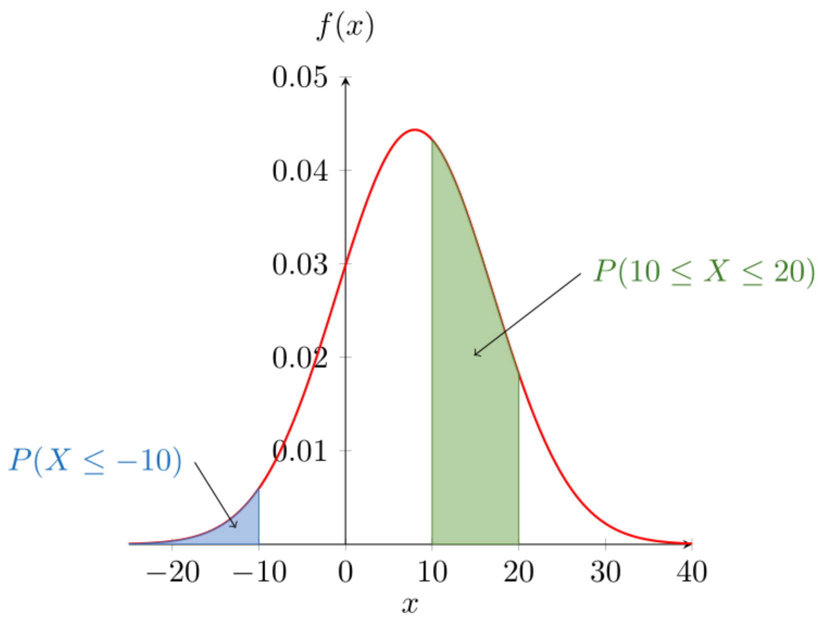


Figure 2: L'aire hachurée correspond à des probabilités. $f(x)$ étant une fonction densité de probabilité

Définition 0.11. Pour toute variable aléatoire continue X de densité f:

- $f(x) \geq 0 \quad \forall x \in \mathbb{R}$
- $\int_{-\infty}^{+\infty} f(x)dx = 1$
- Si l'on pose $a = b$ dans (5), il résulte

$$P(X = a) = \int_a^a f(x)dx = 0$$

- Ceci siginifie que la probabilité qu'une variable aléatoire continue prenne une valeur isolée fixe est toujours nulle. Aussi on peut écrire

$$P(X < a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

Soit X la variable aléatoire réelle de densité de probabilité

$$f(x) = \begin{cases} kx & \text{si} \quad 0 \leq x \leq 5 \\ 0 & \text{sinon} \end{cases}$$

1. Calculer k.
2. Calculer: $P(1 \leq X \leq 3), P(2 \leq X \leq 4)$ et $P(X < 3)$.

Soit X une variable aléatoire réelle continue ayant pour densité de probabilité

$$f(x) = \begin{cases} \frac{1}{6}x + k & \text{si} \quad 0 \leq x \leq 3 \\ 0 & \text{sinon} \end{cases}$$

1. Calculer k.
2. Calculer $P(1 \leq X \leq 2)$

0.7 Fonction de répartition d'une v.a.c

Définition 0.12. Si comme pour les variables aléatoires discrètes, on définit la fonction de répartition de X par:

$$\begin{aligned} F_X : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto F_X(a) = P(X \leq a) \end{aligned}$$

alors la relation entre la fonction de répartition F_X et la fonction densité de probabilité $f(x)$ est la suivante:

$$\forall \quad a \in \mathbb{R} \quad F_X(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

La fonction de répartition $F_X(a)$ est la **primitive** de la fonction densité de probabilité $f(x)$ (donc la densité d'une v.a.c est la dérivée de la fonction de répartition), et permet d'obtenir les probabilités associées à la variable aléatoire X, en effet:

Propriétés: Pour une variable aléatoire continue X:

- $F'_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x) = f(x)$.
- Pour tous réels $a \leq b$,

$$\begin{aligned} P(a < X < b) &= P(a < X \leq b) \\ &= P(a \leq X < b) \\ &= P(a \leq X \leq b) \\ &= F_X(b) - F_X(a) = \int_{a}^{b} f(x)dx \end{aligned}$$

La fonction de répartition correspond aux probabilités cumulées associées à la variable aléatoire continue sur l'intervalle d'étude (Figure 3).

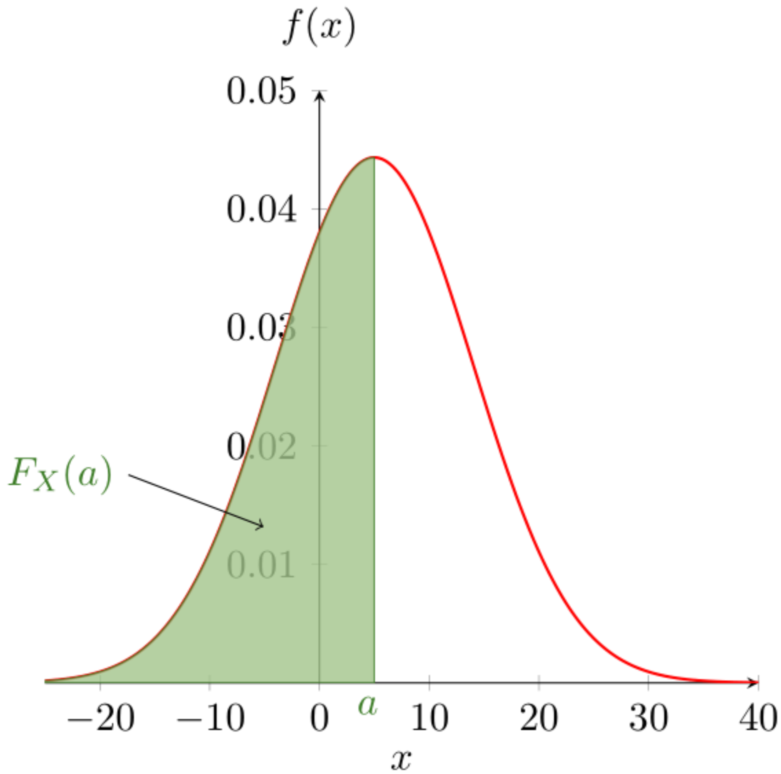


Figure 3: L'aire hachurée en vert sous la courbe de la fonction densité de probabilité correspond à la probabilité $P(X < a) = F_X(a)$ et vaut 0.5 car ceci correspond exactement à la moitié de l'aire totale sous la courbe

Propriétés: Les propriétés associées à la fonction de répartition sont les suivantes:

1. F_X est continue sur $\mathbb{R}$, dérivable en tout point où f est continue.
2. F_X est croissante sur $\mathbb{R}$.
3. F_X est à valeurs dans $[0, 1]$.
4. $\lim_{x \to -\infty} F_X(x) = 0$ et $\lim_{x \to +\infty} F_X(x) = 1$.

0.8 Fonction d'une variable aléatoire continue

Soit X une variable aléatoire continue de densité f_X et de fonction de répartition F_X. Soit h une fonction continue définie sur $X(\Omega)$, alors $Y = h(X)$ est une variable aléatoire.

Pour déterminer la densité de Y, notée f_Y, on commence par calculer la fonction de répartition de Y, notée F_Y, ensuite nous dérivons pour déterminer f_Y.

Calcul de densités pour $h(X) = aX + b$

$\forall \quad y \in \mathbb{R}$,

$$F_Y(y) = P(Y \leq y) = P(h(X) \leq y) = P(aX + b \leq y)$$

si $a > 0$,

$$F_Y(y) = P(aX + b \leq y) = P(X \leq \frac{y-b}{a}) = F_X(\frac{y-b}{a})$$

si $a < 0$,

$$F_Y(y) = P(aX + b \leq y) = P(X \geq \frac{y-b}{a}) = 1 - F_X(\frac{y-b}{a})$$

En dérivant on obtient la densité de Y

$$f_Y(y) = \frac{1}{|a|} f_X(\frac{y-b}{a}).$$

Calcul de densités pour $h(X) = X^2$

si $y < 0$, $F_Y(y) = P(Y \leq y) = 0$.
si $y > 0$,

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

En dérivant on obtient la densité de Y,

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} \left[f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right] & \text{si} \quad y \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Calcul de densités pour $h(X) = e^X$

si $y < 0$, $F_Y(y) = P(Y \leq y) = 0$.
si $y > 0$, $F_Y(y) = P(Y \leq y) = P(e^X \leq y) = P(X \leq \ln(y)) = F_X(\ln(y))$.

En dérivant on obtient la densité de Y

$$f_Y(y) = \begin{cases} \frac{1}{y} f\big(\ln(y)\big) & \text{si} \quad y \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Soit la v.a.c X ayant la fonction de densité

$$f_X(x) = \begin{cases} 2x & \text{si} \quad 0 \leq x \leq 1 \\ 0 & \text{sinon} \end{cases}$$

Déterminer la densité de: $Y = 3X + 1$, $Z = X^2$ et $T = e^X$.

0.9 Espérance et variance de variables aléatoires continues

Espérance d'une v.a.c

Définition 0.13. Si X est une variable aléatoire absolument continue de densité f, on appelle espérance de X, le réel $E(X)$, défini par:

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

si cette intégrale est convergente.

Les propriétés de l'espérance d'une variable aléatoire continue sont les mêmes que pour une variable aléatoire discrète.

Propriétés: Soit X une variable aléatoire continue,

- $E(aX + b) = aE(X) + b \qquad a \geq 0$ et $b \in \mathbb{R}$.
- Si $X \geq 0$ alors $E(X) \geq 0$.
- Si X et Y sont deux variables aléatoires définies sur un même univers Ω alors

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

Théorème 0.7 (Théorème de transfert) Si X est une variable aléatoire de densité $f(x)$, alors pour toute fonction réelle g on aura

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

Soit la v.a.c X ayant la fonction de densité

$$f_X(x) = \begin{cases} 2x & \text{si} \quad 0 \leq x \leq 1 \\ 0 & \text{sinon} \end{cases}$$

Calculer l'espérance des variables aléatoires $Y = 3X + 1$, $Z = X^2$ et $T = e^X$.

Variance d'une v.a.c

La variance d'une variable aléatoire $V(X)$ est l'espérance mathématique du carré de l'écart à l'espérance mathématique. C'est un paramètre de dispersion qui correspond au moment centré d'ordre 2 de la variable aléatoire X.

Définition 0.14. Si X est une variable aléatoire ayant une espérance $E(X)$, on appelle variance de X le réel

$$V(X) = E\big([X - E(X)]^2\big) = E(X^2) - [E(X)]^2$$

Si X est une variable aléatoire continue, on calcule $E(X^2)$ en utilisant le théorème 0.7,

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

Propriétés: Si X est une variable aléatoire admettant une variance alors:

- $V(X) \geq 0$, si elle existe.
- $\forall \quad a \in \mathbb{R}, V(aX) = a^2 V(X)$
- $\forall \quad (a,b) \in \mathbb{R}, V(aX+b) = a^2 V(X)$
- Si X et Y sont deux variables aléatoires indépendantes, $V(X+Y) = V(X) + V(Y)$

Définition 0.15 (Ecart-type)**.** Si X est une variable aléatoire ayant une variance $V(X)$, on appelle écart-type de X, le réel:

$$\sigma_X = \sqrt{V(X)}$$

0.10 Lois usuelles de v.a.c

Loi uniforme $U(a,b)$

La loi uniforme est la loi exacte de phénomènes continus uniformément répartis sur un intervalle.
Définition 0.16. La variable aléatoire X suit une loi uniforme sur le segment $[a,b]$ avec $a < b$ si sa densité de probabilité est donnée par

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si} \quad x \in [a,b] \\ 0 & \text{si} \quad x \notin [a,b] \end{cases} = \frac{1}{b-a} 1_{[a,b]}(x)$$

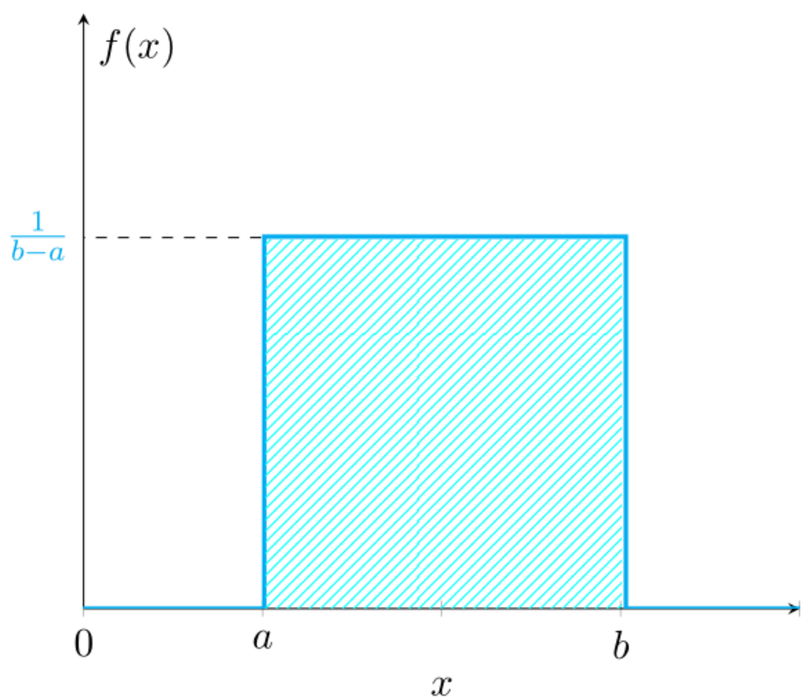


Figure 4: Fonction de densité de $U([a,b])$

Quelques commentaires:

1. La loi uniforme continue étant une loi de probabilité, l'aire hachurée en bleu sur la Figure 4 vaut 1.
2. La fonction de répartition associée à la loi uniforme continue est

$$F_X(x) = \begin{cases} 0 & \text{si} \quad x < a \\ \frac{x-a}{b-a} & \text{si} \quad a \leq x \leq b \\ 1 & \text{si} \quad x > b \end{cases}$$

Propriétés: Si X est une v.a.c qui suit la loi uniforme sur $[a, b]$:

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

Soit $X \sim U(0, 10)$. Calculer:

1. $P(X < 3)$
2. $P(X \geq 6)$
3. $P(3 < X < 8)$

Loi exponentielle $\mathcal{E}(\lambda)$

Définition 0.17. On dit qu'une variable aléatoire X est **exponentielle** (ou suit la loi exponentielle) de paramètre λ si sa densité est donnée par

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si} \quad x \geq 0 \\ 0 & \text{si} \quad x < 0 \end{cases} = \lambda e^{-\lambda x} 1_{\mathbb{R}^+}(x)$$

On dit $X \sim \mathcal{E}(\lambda)$

La fonction de répartition F d'une variable aléatoire exponentielle est donnée par

$$\text{Si } x \geq 0 \quad F(x) = P(X \leq x) = \int_0^x f(t)dt = \int_0^x \lambda e^{-\lambda t} dt = \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x}$$

Propriétés: Si $X \sim \mathcal{E}(\lambda)$

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

Cas d'utilisations de la loi exponentielle : Dans la pratique, on rencontre souvent la distribution exponentielle lorsqu'il s'agit de représenter le temps d'attente avant l'arrivée d'un événement spécifié. Une loi exponentielle modélise la durée de vie d'un phénomène sans mémoire, ou sans vieillissement, ou sans usure. En d'autres termes, le fait que le phénomène ait duré pendant un temps t ne change rien à son espérance de vie à partir du temps t. On dit qu'une variable aléatoire non négative X est **sans mémoire** lorsque

$$P(X > t + h | X > t) = P(X > h) \qquad \forall \quad t, h \geq 0$$

Par exemple, la durée de vie de la radioactivité ou d'un composant électronique, le temps qui nous sépare d'un prochain tremblement de terre ou du prochain appel téléphonique mal aiguillé sont toutes des variables aléatoires dont les distributions tendent en pratique à se rapprocher de distributions exponentielles.

Loi Normale ou de Laplace-Gauss $\mathcal{N}(\mu, \sigma^2)$

Définition 0.18. Une variable aléatoire X est dite **normale** avec paramètres μ et σ^2 si la densité de X est donnée par

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad \forall\, x \in \mathbb{R}$$

Avec $\mu \in \mathbb{R}$ et $\sigma \in \mathbb{R}^+$. On dit que $X \sim \mathcal{N}(\mu, \sigma^2)$.

Remarque: On admet que $\int_{-\infty}^{+\infty} f(x)dx = 1$ dans la mesure où l'intégration analytique est impossible.

Étude de la densité de la loi Normale

- La fonction f est paire autour d'un axe de symétrie $x = \mu$ car $f(x + \mu) = f(\mu - x)$.
- $f'(x) = 0$ pour $x = \mu$, $f'(x) < 0$ pour $x < \mu$ et $f'(x) > 0$ pour $x > \mu$

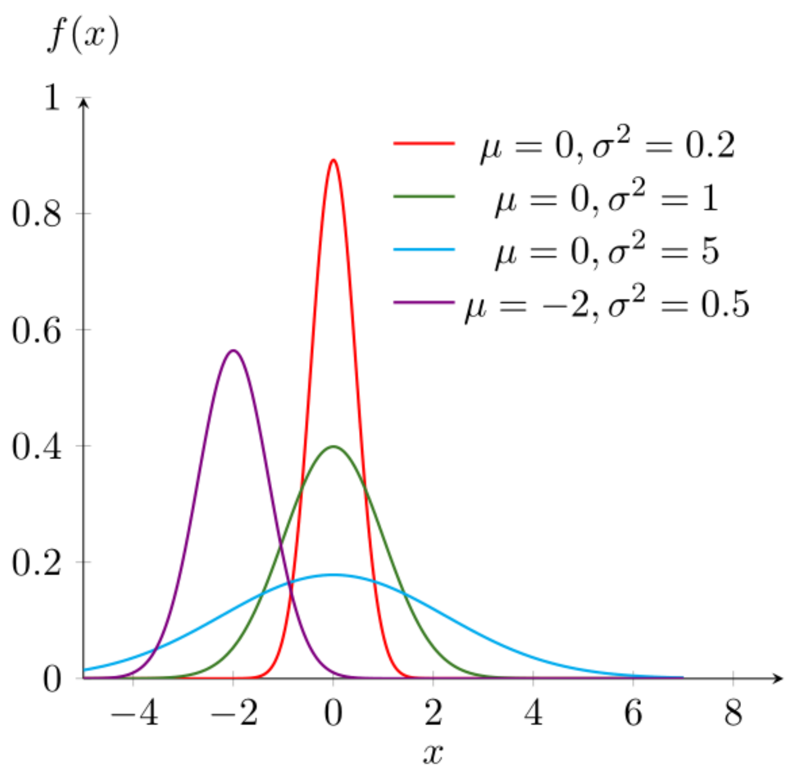


Figure 5: Représentation graphique de la densité d'une loi normale. Remarque: Le paramètre μ représente l'axe de symétrie et s le degré d'aplatissement de la courbe de la loi normale dont la forme est celle d'une courbe en cloche

Propriétés: Soit $X \sim \mathcal{N}(\mu, \sigma^2)$, on a:

- $E(X) = \mu$
- $V(X) = \sigma^2$

Théorème 0.8 (Stabilité de la loi normale) Soit X_1 et X_2 deux variables aléatoires normales et indépendantes de paramètres respectifs (μ_1, σ_1^2) et (μ_2, σ_2^2), alors leur somme $X_1 + X_2$ est une variable aléatoire normale de paramètres $(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Loi Normale centrée réduite $\mathcal{N}(0, 1)$

Définition 0.19. Une variable aléatoire continue X suit une **loi normale centrée réduite** si sa densité de probabilité est donnée par

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad \forall\, x \in \mathbb{R} \tag{6}$$

On dit $X \sim \mathcal{N}(0, 1)$.

Remarque: $E(X) = 0$ et $V(X) = 1$.

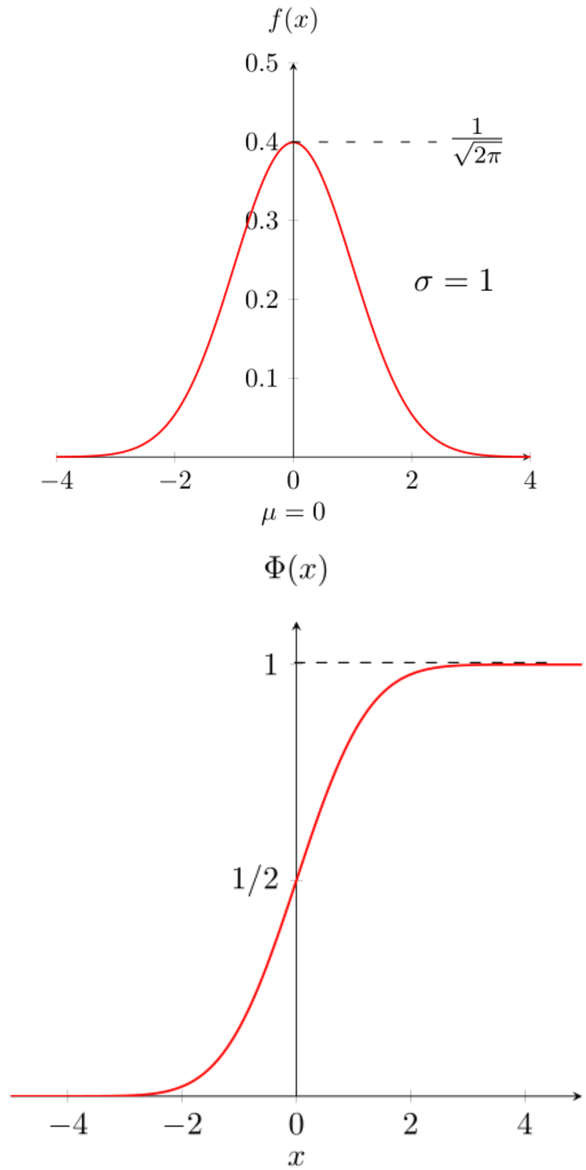


Figure 6: (gauche): Densité d'une loi normale centrée réduite $\mathcal{N}(0,1)$. (droite): Fonction de répartition de $\mathcal{N}(0,1)$.

Relation entre loi normale et loi normale centrée réduite

Théorème 0.9 (Relation avec la loi normale) Si X suit une loi normale $\mathcal{N}(\mu,\sigma^2)$, alors $Z = \frac{X-\mu}{\sigma}$ est une variable centrée réduite qui suit la loi normale centrée réduite $\mathcal{N}(0,1)$.

Calcul des probabilités d'une loi normale

La fonction de répartition de la loi normale réduite permet d'obtenir les probabilités associées à toutes variables aléatoires normales $\mathcal{N}(\mu,\sigma^2)$ après transformation en variable centrée réduite.
Définition 0.20. On appelle fonction Φ, la fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0,1)$,

telle que

$$\forall\ x \in \mathbb{R} \quad \Phi(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} f(t)dt$$

Propriétés: Les propriétés associées à la fonction de répartition Φ sont:

1. Φ est croissante, continue et dérivable sur $\mathbb{R}$ et vérifie:

   $\lim\limits_{x \to -\infty} \Phi(x) = 0$ et $\lim\limits_{x \to \infty} \Phi(x) = 1$

2. $\forall\ x \in \mathbb{R} \quad \Phi(x) + \Phi(-x) = 1$

3. $\forall\ x \in \mathbb{R} \quad \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$

Une application directe de la fonction Φ est la lecture des probabilités de la loi normale sur la table de la loi normale centrée réduite.

Soit X une variable aléatoire normale de paramètres $\mu = 3$ et $\sigma^2 = 4$. Calculer:

1. $P(X > 0)$
2. $P(2 < X < 5)$
3. $P(|X - 3| > 4)$

Approximation normale d'une répartition binomiale

Un résultat important de la théorie de probabilité est connu sous le nom de théorème limite de Moivre-Laplace. Il dit que pour n grand, une variable binomiale $\mathcal{B}(n, p)$ suivra approximativement la même loi qu'une variable aléatoire normale avec même moyenne et même variance. Ce théorème énonce que si "on standardise" une variable aléatoire binomiale $\mathcal{B}(n, p)$ en soustrayant d'abord sa moyenne np puis en divisant le résultat par son écart-type $\sqrt{np(1-p)}$, alors la variable aléatoire standardisée (de moyenne 0 et variance 1) suivra approximativement, lorsque n est grand, une distribution normale standard. Ce résultat fut ensuite progressivement généralisé par Laplace, Gauss et d'autres pour devenir le théorème actuellement connu comme théorème centrale limite qui est un des deux résultats les plus importants de la théorie de probabilités. Ce théorème sert de base théorique pour expliquer un fait empirique souvent relevé, à savoir qu'en pratique de très nombreux phénomènes aléatoires suivent approximativement une distribution normale.

On remarquera qu'à ce stade deux approximations de la répartition binomiale ont été proposées: l'approximation de Poisson, satisfaisante lorsque n est grand et lorsque np n'est pas extrême; l'approximation normale pour laquelle on peut montrer qu'elle est de bonne qualité lorsque $np(1-p)$ est grand (dès que $np(1-p)$ dépasse 10).

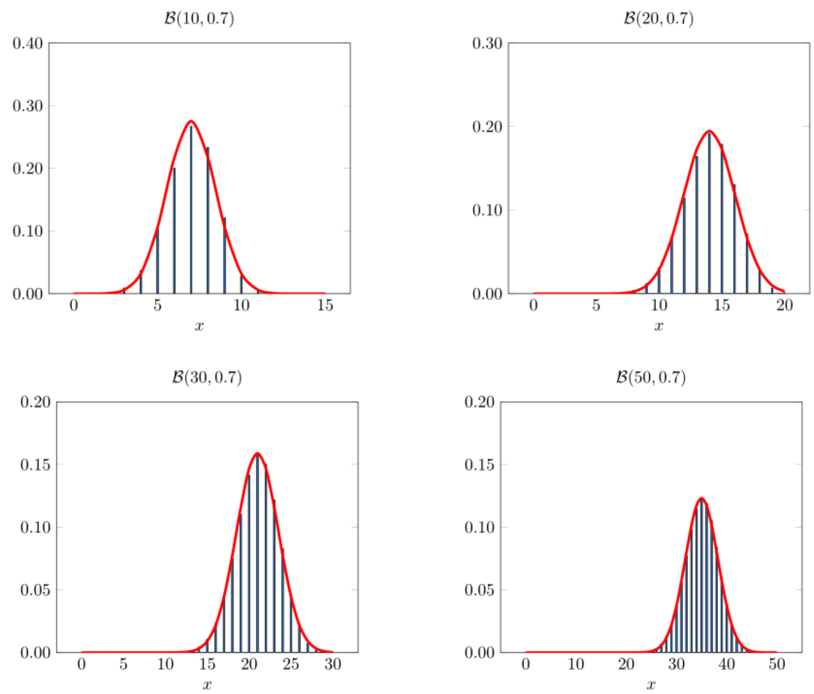


Figure 7: La loi de probabilité d'une variable aléatoire $B(n,p)$ devient de plus en plus **normale** à mesure que n augmente.

Loi de χ^2 de Pearson

Définition 0.21. Soit $X_1, X_2, \ldots, X_n$, n variables **normales centrées réduites**, et Y la variable aléatoire définie par

$$Y = X_1^2 + X_2^2 + \ldots + X_i^2 + \ldots + X_n^2 = \sum_{i=1}^{n} X_i^2$$

On dit que Y suit la loi de χ^2 (ou loi de Pearson) à n degrés de liberté, $Y \sim \chi^2(n)$

> La loi de χ^2 trouve de nombreuses applications dans le cadre de la comparaison de proportions, des tests de conformité d'une distribution observée à une distribution théorique et le test d'indépendance de deux caractères qualitatifs. Ce sont les *tests du khi-deux.*

Remarque: Si $n = 1$, la variable du χ^2 correspond au carré d'une variable normale centrée réduite $\mathcal{N}(0,1)$.

Propriétés: Si $Y \sim \chi^2(n)$, alors:

- $E(Y) = n$
- $V(Y) = 2n$

Loi de Student $St(n)$

Définition 0.22. Soit U une variable aléatoire suivant une loi **normale centrée réduite** $\mathcal{N}(0,1)$ et V une variable aléatoire suivant une loi de $\chi^2(n)$, U et V étant indépendantes, on dit alors que $T_n = \frac{U}{\sqrt{\frac{V}{n}}}$ suit une

loi de Student à n degrés de liberté. $T_n \sim St(n)$

La loi de Student est utilisée lors des tests de comparaison de paramètres comme la moyenne et dans l'estimation de paramètres de la population à partir de données sur un échantillon (Test de Student).

Loi de Fisher-Snedecor $\mathcal{F}(n, m)$

Définition 0.23. Soit U et V deux variables aléatoires indépendantes suivant une loi de χ^2 respectivement à n et m degrés de liberté.

On dit que $F = \frac{U/n}{V/m}$ suit une loi de Fisher-Snedecor à (n, m) degrés de liberté. $F \sim \mathcal{F}(n, m)$

La loi de Fisher-Snedecor est utilisée pour comparer deux variances observées et sert surtout dans les très nombreux tests d'analyse de variance et de covariance.

0.11 Couple de variables aléatoires continues

Densité conjointe

Définition 0.24. On dit que (X, Y) est un couple aléatoire continu s'il existe une fonction $f : \mathbb{R}^2 \to \mathbb{R}$ telle que pour tout $D \subseteq \mathbb{R}$ on a

$$P\{(X, Y) \in D\} = \iint_{(x,y)\in D} f(x, y)dxdy$$

Remarque: On a la condition de normalité $\iint_{\mathbb{R}^2} f(x, y)dxdy = 1$

La fonction f s'appelle **densité conjointe** de X et Y. Notons par A et B deux ensembles de nombres réels. En définissant $D = \{(x, y) : x \in A, y \in B\}$, on obtient

$$P(X \in A, Y \in B) = \int_A \int_B f(x, y)dxdy$$

La fonction de répartition du (X, Y) est définie par

$$F(a, b) = P(X \leq a, Y \leq b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y)dxdy$$

f est le dérivé de F: $f(a, b) = \frac{\partial^2}{\partial a \partial b} F(a, b)$

Soit (X, Y) un couple aléatoire continu de densité

$$f(x, y) = \begin{cases} axy^2 & \text{si} \quad 0 \leq x \leq y \leq 1 \\ 0 & \text{sinon} \end{cases}$$

Trouver la constante a.

Soit (X, Y) un couple aléatoire continu de densité

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & \text{si} \quad x > 0,\ y > 0 \\ 0 & \text{sinon} \end{cases}$$

Montrer que:

1. $P(X > 1, Y < 1) = e^{-1}(1 - e^{-2})$
2. $P(X < a) = 1 - e^{-a}$
3. $P(X < Y) = 1/3$

Densités marginales

Si on dispose de la densité du couple, on peut retrouver les densités de X et de Y, appelées les densités marginales:

Densité marginale de X:

$$f(x,.) = f_X(x) = \int_{\mathbb{R}} f(x,y)dy$$

Densité marginale de Y:

$$f(.,y) = f_Y(y) = \int_{\mathbb{R}} f(x,y)dx$$

Espérance d'une fonction du couple

Si (X, Y) est un couple continu de densité $f(x, y)$ et $g : \mathbb{R}^2 \to \mathbb{R}$ on a

$$E[g(X,Y)] = \iint_{\mathbb{R}^2} g(x,y)f(x,y)dxdy$$

Indépendance

Les v.a. X et Y sont indépendantes ssi $\forall\,(x, y) \in \mathbb{R}^2$ on a

$$f(x,y) = f_X(x)f_Y(y)$$

Distribution conditionnelle

Si (X, Y) est un couple continu de densité $f(x, y)$, on définit **densité conditionnelle de** X, sous la condition $Y = y$ et lorsque $f_Y(y) > 0$ par la relation

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

Supposons que X et Y aient pour densité conjointe

$$f(x,y) = \begin{cases} \frac{1}{y} e^{-x/y} e^{-y} & \text{si} \quad x > 0,\ y > 0 \\ 0 & \text{sinon} \end{cases}$$

1. Déterminer la densité conditionnelle de X lorsque $Y = y$.
2. Calculer $P(X > 1 | Y = y)$

Exercices

Combinatoire

Exercice 0.1. On tire simultanément 5 cartes d'un jeu de 52 cartes.

1. Combien de tirages différents peut-on obtenir?
2. Combien de tirages peut-on obtenir ? contenant:
    1. 5 carreaux ou 5 coeurs;
    2. 2 coeurs et 3 piques;
    3. au moins 1 roi;
    4. au plus 1 roi.

Exercice 0.2. On jette un dé équilibré 3 fois de suite, et on s'intéresse au total des points obtenus. De combien de façons peut-on obtenir:

1. un total égale à 16.
2. un total égale à 15.
3. un total au moins égale à 15.

Événements

Exercice 0.3. Soit A,B et C trois événements d'un espace probabilisable $(\Omega, \mathcal{A})$. Exprimer en fonction de A,B et C et des opérations ensemblistes (réunion, intersection et complémentaire) les événements ci-après:

1. A seul (parmi les 3 événements) se produit.
2. A et C se produisent, mais non B.
3. Les trois événements se produisent.
4. L'un au moins des 3 événements se produit.
5. Aucun des trois événements ne se produit.
6. Deux événements exactement se produisent.
7. Pas plus de deux événements ne se produisent.

Probabilité

Exercice 0.4. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Soient A,B et C trois événements quelconques. On pose : $E = A \cap \bar{B} \cap \bar{C}$ et $F = A \cap (B \cup C)$.

1. Montrer que E et F sont incompatibles.
2. Montrer que $E \cup F = A$.
3. Sachant que $P(A) = 0.6$; $P(A \cap B) = 0.2$; $P(A \cap C) = 0.1$ et $P(A \cap B \cap C) = 0.05$: Calculer $P(F)$ et $P(E)$.

Variables aléatoires discrètes

Exercice 0.5. On choisit deux boules au hasard dans une urne contenant 8 boules blanches, 4 boules noires et 2 boules oranges. Supposons que l'on reçoive 2 euros pour chaque boule noire tirée et que l'on perde 1 euro pour chaque boule blanche tirée. Désignons les gains nets par X.

1. Quelles sont les valeurs possibles pour X et les probabilités associées à ces valeurs ?
2. Quelle est l'espérance de X ?

Exercice 0.6. Une urne contient une boule qui porte le numéro 0, deux qui portent le numéro 1 et quatre qui portent le numéro 3. On extrait simultanément deux boules dans cette urne.

1. Déterminer la loi de probabilité de la variable aléatoire X qui représente la somme des nombres obtenus.
2. Déterminer la fonction de répartition de X.
3. Calculer $E(X)$, $V(X)$ et $\sigma(X)$.

Exercice 0.7. Soit X une v.a. qui suit la loi uniforme (e.g. équiprobabilité de valeurs de X) sur l'ensemble $X(\Omega) = \{-3, -2, 1, 4\}$.

1. Donner la loi de X.
2. Calculer $E(X)$ et $V(X)$.

   On définit la variable aléatoire $Y = (X + 1)^2$.

3. Donner $Y(\Omega)$ et la loi de Y.
4. Calculer $E(Y)$ de deux façons différents.

Exercice 0.8. Soient X et Y des variables aléatoires discrètes dont la loi jointe est donnée par le tableau suivant:

$X \backslash Y$	-1	0	2	5
0	0.10	0.05	0.15	0.05
1	0.15	0.20	0.25	0.05

1. Quelle est la loi marginale de X ?
2. Quelle est la loi marginale de Y ?
3. Calculer $P(Y \geq 0/X = 1)$.
4. Calculer $E(X)$, $E(Y)$, et $cov(X, Y)$.
5. Les variables X et Y sont elles indépendantes ?

Exercice 0.9. Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{N}^2$ tel que

$$\forall (p,q) \in \mathbb{N}^2, \quad P(X = p, Y = q) = \lambda \frac{p+q}{p!q!2^{p+q}}$$

1. Déterminer λ.
2. Calculer les lois marginales.
3. Les variables X et Y sont elles indépendantes ?

Exercice 0.10. Une urne contient 2 boules de numéro 20, 4 boules de numéro 10 et 4 boules de numéro 5.

1. Une épreuve consiste à tirer simultanément 3 boules de l'urne. Calculer la probabilité p que la somme des numéros tirés soit égale à 30.
2. On répéte cette épreuve 4 fois en remettant à chaque fois les trois boules tirés dans l'urne. Soit X la v.a. indiquant le nombre de tirages donnant une somme de numéros égale à 30.
    1. Quelle la loi de X. Donner son espérance et son écart-type.
    2. Déterminer la probabilité d'avoir au moins une fois la somme 30 dans les 4 tirages.

Exercice 0.11. Vous avez besoin d'une personne pour vous aider à déménager. Quand vous téléphonez à un ami, il y a une chance sur quatre qu'il accepte. Soit X la variable aléatoire qui représente le nombre d'amis que vous devrez contacter pour obtenir cette aide.

1. Déterminer la loi de probabilité de X.
2. Calculer $P(X \leq 3)$.
3. Calculer $E(X)$.

Exercice 0.12. Pour être sélectionné aux jeux olympiques, un athlète doit réussir deux fois à dépasser les minima fixés par sa fédération. Il a une chance sur trois de réussir à chaque épreuve à laquelle il participe. On note X la variable aléatoire qui représente le nombre d'épreuves auxquelles il devra participer pour être sélectionné.

1. Déterminer la loi de probabilité de X.
2. Si cet athlète ne peut participer qu'à quatre épreuves maximum, quelle est la probabilité qu'il soit sélectionné ?

Exercice 0.13. Un sac contient cinq jetons : deux sont numérotés 1 et les trois autres sont numérotés 2. On effectue une série illimitée de tirages avec remise d'un jeton dans le sac S. On désigne par Y la variable aléatoire égale au nombre de tirages effectués avant le tirage amenant un jeton numéroté 1 pour la première fois.

1.
    1. Justifier que la variable aléatoire $Z = Y + 1$ suit une loi usuelle que l'on précisera.
    2. En déduire la loi de probabilité de Y.
2.
    1. Préciser l'espérance mathématique et la variance de Z.
    2. En déduire l'espérance mathématique et la variance de Y.

Exercice 0.14. Le nombre de pannes d'électricité qui se produisent dans une certaine région au cours d'une période d'un an suit une loi de Poisson de paramètre $\lambda = 3$.

1. Calculer la probabilité qu'au cours d'une période d'un an il y a exactement une panne qui se produit.
2. En supposant l'indépendance des pannes d'une année à l'autre, calculer la probabilité qu'au cours des dix prochaines années il y ait au moins une année pendant laquelle il se produira exactement une panne.

Exercice 0.15. Un poste de radio a 2 types de pannes: transistor ou condensateur. Durant la première année d'utilisation, on désigne par:

X = nombre de pannes dues à une défaillance de transistor.
Y = nombre de pannes dues à une défaillance de condensateur.

On suppose que X et Y sont des v.a. indépendantes suivant des lois de Poisson de paramètres respectives $\lambda = 2$ et $\mu = 1$.

1. Calculer la probabilité qu'il y ait 2 pannes dues à une défaillance de transistor.
2. Calculer la probabilité qu'il y ait au moins une panne due à une défaillance de condensateur.
3. 1. Quelle est la loi du nombre $Z = X + Y$ de pannes durant la première année ?
   2. Déterminer la probabilité qu'il y ait 2 pannes de type quelconque.
   3. Calculer $P(Z = 3)$. Que peut-on remarquer ?
   4. Décrire les variations de $P(Z = k)$ en fonction de k.
   5. Donner le nombre moyen de pannes et la probabilité qu'il y ait au plus une panne durant cette période.

Variables alétoires continues

Exercice 0.16. Soit X une v.a.c. de densité f définie par: $f(x) = kx \times 1_{]0,2[}(x)$.

1. Déterminer la constante k.
2. Calculer $E(X)$ et $E(X^2)$.
3. On pose $Z = X^2$. Déterminer la densité de Z. Calculer $E(Z)$.

Exercice 0.17. Soit X une variable aléatoire continue dont la fonction de densité est donnée par:

$$f(x) = \begin{cases} c(1 - x^2) & \text{si} \quad -1 < x < 1 \\ 0 & \text{sinon} \end{cases}$$

1. Quelle est la valeur de c?
2. Quelle est la fonction de répartition de X?

Exercice 0.18. Soit X une variable aléatoire continue dont la fonction de densité est donnée par:

$$f(x) = \begin{cases} 0 & \text{si} \quad |x| > k > 0 \\ x + 1 & \text{si} \quad |x| \leq k \end{cases}$$

1. Déterminer k.
2. Calculer $E(X)$ et $E(X^2)$.
3. Déterminer la fonction de répartition de X.
4. Soit $Y = X^2$. Déterminer la fonction de répartition ainsi que la fonction de densité de Y.
5. Calculer $E(Y)$.

Exercice 0.19 (Variable aléatoire de densité paire)**.** Soit X une variable aléatoire réelle admettant une fonction **paire** f pour densité.

1. Calculer $P(X \leq 0)$ et $P(X \geq 0)$.
2. Montrer que la fonction de répartition F de X vérifie: $\forall\, x \in \mathbb{R}, F(x) = 1 - F(-x)$.
3. On admet que X admet une espérance, calculer $E(X)$.
4. Donner un exemple de densité paire.

Table de la loi Normale Centrée Réduite

Exercice 0.20. On note Φ la fonction de répartition de la loi normale centrée réduite.

1. Soit X une v.a. qui suit une loi normale centrée réduite, i.e. $X \sim \mathcal{N}(0,1)$. A l'aide de la table de la loi normale, calculer: $P(X>2), P(-1<X<1.5)$ et $P(X<0.5)$.

2. Soit Y une v.a. qui suite une loi normale: $Y \sim \mathcal{N}(\mu,\sigma^2) = \mathcal{N}(4,16)$. Calculer: $P(Y>2), P(-1< Y<1.5)$ et $P(Y<0.5)$.

3. Soit $U \sim \mathcal{N}(6,4)$. Calculer: $P(|U-4|<3)$ et $P(U>6|U>3)$.

Exercice 0.21. Une machine produit des pièces dont le diamètre X (en cm) est une variable aléatoire qui suit une loi normale d'espérance μ et de variance $\sigma^2 = (0.01)^2$. Quelle devrait être la valeur de μ de sorte que la probabilité qu'une pièce prise au hasard ait un diamètre supérieur à 3 cm, soit inférieur à 0.01?

Exercice 0.22. On envisage de construire à l'entrée d'une caserne une guérite dans laquelle pourra s'abriter la sentinelle en cas d'intempéries. Les sentinelles sont des appelés dont la taille est approximativement distribuée selon une loi normale d'espérance 175cm et d'écart-type 7cm. A quelle hauteur minimale doit se trouver le toit de la guérite, pour qu'un sentinelle pris au hasard ait une probabilité supérieure à 0.95 de s'y tenir debout?

Exercice 0.23 (Loi uniforme et loi exponentielle)**.** Soit U une v.a.c de loi unifrorme sur $[0,1]$. Montrer que la v.a. $X = -\ln U$ suit une loi exponentielle.

Exercice 0.24 (La loi exponentielle est sans mémoire)**.** On suppose que la durée de vie D, en jours, d'une ampoule, est une variable aléatoire de loi exponentielle de paramètre $\frac{1}{100}$.

1. Quelle est la durée de vie moyenne d'une ampoule?

2. Calculer la fonction de répartition de D. En déduire l'expression de $P(D>x)$.

3. On dit qu'une variable aléatoire est *sans mémoire* si $\forall l>0 \quad P(X \geq n+l|X \geq n) = P(X \geq l)$. Montrer que D est sans mémoire.

4. Quelle est la probabilité qu'une ampoule dure encore au moins 10 jours, sachant qu'à son n^{e} jour, elle marche encore?

Exercice 0.25 (Lois des v.a.r. 'min(X,Y)' et 'max(X,Y)')**.** Soit X et Y deux v.a.c de densités respectives f_X et f_Y et de fonctions de répartition respectives F_X et F_Y. On suppose que X et Y sont indépendantes. On pose:

$$Z = max(X,Y) \qquad \text{et} \qquad T = min(X,Y)$$

1. Exprimer les fonctions de répartition de Z et T à l'aide des fonctions de répartition F_X et F_Y.

2. Exprimer une densité de Z et une densité de T à l'aide de f_X, f_Y, F_X et F_Y.

Exercice 0.26 (Minimum et Maximum de deux lois exponentielles)**.** Soit X_1 et X_2 deux variables aléatoires indépendantes suivant une loi exponentielle de paramètres λ_1 et λ_2. On pose $X = min(X_1,X_2)$.

1. Montrer que X suit une loi exponentielle de paramètre $\lambda_1+\lambda_2$.

2. Deux guichets sont ouverts à une banque. Le temps de service au premier guichet (resp. au deuxième) suit une loi exponentielle de moyenne 20 min (resp. 30 min). Deux client rentrent simultanément, l'un choisit le guicher 1 et l'autre le guichet 2.

   1. En moyenne, après combien de temps sort le premier?

   2. En moyenne, après combien de temps sort le dernier?

Indication: On pourra utiliser la relation $X_1+X_2 = min(X_1,X_2)+max(X_1,X_2)$. La somme de deux nombres réels est égale à la somme de leur minimum et de leur maximum.

Exercice 0.27 (Fonction Gamma (Euler))**.** La fonction Gamma est définie sur $\mathbb{R}_+^*$ par:

$$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt$$

1. Montrer que $\Gamma(x+1) = x\Gamma(x)$, $\forall\, x > 0$.
2. Exprimer $\Gamma(x+n)$ en fonction de $\Gamma(x)$ pour $n \in \mathbb{N}$.
3. Calculer $\Gamma(1)$. En déduire $\Gamma(n+1)$ pour $n \in \mathbb{N}$.
4. En utilisant le changement de variable $t = u^2$, montrer qu'on a:
$$\Gamma(x) = 2\int_0^{+\infty} u^{2x-1}e^{-u^2}du$$
5. On suppose que:
$$\int_0^{+\infty} e^{-x^2}dx = \frac{\sqrt{\pi}}{2}$$
Calculer alor $\Gamma(\frac{1}{2})$.
6. Montrer que $\Gamma(n+\frac{1}{2}) = (n-\frac{1}{2})(n-\frac{3}{2})\ldots(\frac{1}{2})\Gamma(\frac{1}{2})$
7. En déduire la valeur de $\Gamma(n+\frac{1}{2})$.

Exercice 0.28 (Loi Gamma)**.** Pour $a > 0$ et $\lambda > 0$, deux constantes réelles, on définit la fonction $f_{a,\lambda}$ sur $\mathbb{R}$ par:

$$\forall\, x \in \mathbb{R}, \quad f_{a,\lambda}(x) = \frac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda x} \times 1_{\mathbb{R}_+}(x)$$

1. Montrer que $f_{a,\lambda}$ est bien une densité d'une v.a. X.
2. Calculer $E(X)$.

Couple de variables aléatoires continues

Exercice 0.29. Soit (X,Y) un couple de variables aléatoires dont la densité jointe est définie par:

$$f(x,y) = \begin{cases} \alpha x e^{-y} & \text{si} \quad x \in [0,1], y \in \mathbb{R}_+ \\ 0 & \text{sinon} \end{cases}$$

1. Déterminer α pour que $f(x,y)$ soit une fonction de densité.
2. Déterminer les densités marginales de X et Y.
3. X et Y sont elles indépendantes?

Exercice 0.30. Soit (X,Y) un vecteur aléatoire uniformément distribué dans l'ensemble

$$D = \{(x,y) \in \mathbb{R}^2, x \in [0,2]\,\text{et}\,y \in [0,4]\}$$

1. Déterminer la densité jointe du couple de variables aléatoires (X,Y).
2. Calculer $P(X \le 1, Y \le 2)$.
3. Déterminer les densités marginales de X et Y. Les v.a. X et Y sont elles indépendantes?

Statistique Inférentielle

Chapter 1

Introduction

La **statistique** est la science dont l'objet est de recueillir, de traiter et d'analyser des données issues de l'observation de phénomènes **aléatoires**, c'est-à-dire dans lesquels le hasard intervient.

L'analyse des données est utilisée pour **décrire** les phénomènes étudiés, **faire des prévisions** et **prendre des décisions** à leur sujet. En cela, la statistique est un outil essentiel pour la compréhension et la gestion des phénomènes complexes.

Les données étudiées peuvent être de toute nature, ce qui rend la statistique utile dans tous les champs disciplinaires et explique pourquoi elle est enseignée dans toutes les filières universitaires, de l'économie à la biologie en passant par la psychologie, et bien sûr les sciences de l'ingénieur.

Le point fondamental est que les données sont entâchées d'incertitudes et présentent des **variations** pour plusieurs raisons :

- le déroulement des phénomènes observés n'est pas prévisible à l'avance avec certitude (par exemple on ne sait pas prévoir avec certitude les cours de la bourse ou les pannes des voitures)
- toute mesure est entâchée d'erreur
- etc…

Il y a donc intervention du **hasard** et des **probabilités**. L'objectif essentiel de la statistique est de maîtriser au mieux cette incertitude pour extraire des informations utiles des données, par l'intermédiaire de l'analyse des variations dans les observations.

Les méthodes statistiques se répartissent en deux classes :

- La **statistique descriptive**, **statistique exploratoire** ou **analyse des données**, a pour but de résumer l'information contenue dans les données de façon synthétique et efficace. Elle utilise pour cela des représentations de données sous forme de graphiques, de tableaux et d'indicateurs numériques (par exemple des moyennes). Elle permet de dégager les caractéristiques essentielles du phénomène étudié et de suggérer des hypothèses pour une étude ultérieure plus sophistiquée. Les probabilités n'ont ici qu'un rôle mineur.
- La **statistique inférentielle** va au delà de la simple description des données. Elle a pour but de **faire des prévisions** et de **prendre des décisions** au vu des observations. En général, il faut pour cela proposer des **modèles probabilistes** du phénomène aléatoire étudié et savoir gérer les risques d'erreurs. Les probabilités jouent ici un rôle fondamental.

Pour le grand public, les statistiques désignent les résumés de données fournis par la statistique descriptive. Par exemple, on parle des "statistiques du chômage" ou des "statistiques de l'économie américaine". Mais on oublie en général les aspects les plus importants liés aux prévisions et à l'aide à la décision apportés par la statistique inférentielle.

L'informatique et la statistique sont deux éléments du **traitement de l'information** : l'informatique acquiert et traite l'information tandis que la statistique l'analyse. Les deux disciplines sont donc étroitement liées. En particulier, l'augmentation considérable de la puissance des ordinateurs et la facilité de transmission des données par internet ont rendu possible l'analyse de très grandes masses de données, ce qui nécessite l'utilisation de méthodes de plus en plus sophistiquées, connues sous le nom de **data mining**, **fouille de données** ou **Big Data**. Enfin, l'**informatique décisionnelle** ou **business intelligence** regroupe les outils d'aide à la décision devenus essentiels dans la gestion des entreprises. Ces outils nécessitent un recours important aux méthodes statistiques.

Plus généralement, tout ingénieur est amené à prendre des décisions au vu de certaines informations, dans des contextes où de nombreuses incertitudes demeurent. Il importe donc qu'un ingénieur soit formé aux techniques de gestion du risque et de traitement de données expérimentales.

La démarche statistique

La statistique et les probabilités sont les deux aspects complémentaires de l'étude des phénomènes aléatoires. Ils sont cependant de natures bien différentes.

Les **probabilités** peuvent être envisagées comme une branche des mathématiques pures, basée sur la théorie de la mesure, abstraite et complètement déconnectée de la réalité.

Les **probabilités appliquées** proposent des **modèles probabilistes** du déroulement de phénomènes aléatoires concrets. On peut alors, **préalablement à toute expérience**, faire des prévisions sur ce qui va se produire.

Par exemple, il est usuel de modéliser la durée de bon fonctionnement ou durée de vie d'un système, mettons une ampoule électrique, par une variable aléatoire X de loi exponentielle de paramètre λ. Ayant adopté ce modèle probabiliste, on peut effectuer tous les calculs que l'on veut. Par exemple :

- la probabilité que l'ampoule ne soit pas encore tombée en panne à la date t est $P(X > t) = e^{\lambda t}$.
- la durée de vie moyenne est $E(X) = 1/\lambda$.
- Si n ampoules identiques sont mises en fonctionnement en même temps, et qu'elles fonctionnent indépendamment les unes des autres, le nombre N_t d'ampoules qui tomberont en panne avant un instant t est une variable aléatoire de loi binomiale $\mathcal{B}(n, P(Xt)) = \mathcal{B}(n, 1e^{-\lambda t})$. Donc on s'attend à ce que, en moyenne, $E(N_t) = n(1e^{-\lambda t})$ ampoules tombent en panne entre 0 et t

Dans la pratique, l'utilisateur de ces ampoules est très intéressé par ces résultats. Il souhaite évidemment avoir une évaluation de leur durée de vie, de la probabilité qu'elles fonctionnent correctement pendant plus d'un mois, un an, etc... Mais si l'on veut utiliser les résultats théoriques énoncés plus haut, il faut d'une part pouvoir s'assurer qu'on a choisi un bon modèle, c'est-à-dire que la durée de vie de ces ampoules est bien une variable aléatoire de loi exponentielle, et, d'autre part, pouvoir calculer d'une manière ou d'une autre la valeur du paramètre λ. C'est la statistique qui va permettre de résoudre ces problèmes. Pour cela, il faut faire une expérimentation, recueillir des données et les analyser.

On met donc en place ce qu'on appelle un **essai** ou une **expérience**. On fait fonctionner en parallèle et indépendamment les unes des autres $n = 10$ ampoules identiques, dans les mêmes conditions expérimentales, et on relève leurs durées de vie. Admettons que l'on obtienne les durées de vie suivantes, exprimées en heures :

91.6	35.7	251	24.3	5.4	67.3	171	9.5	118	57.1

Notons $x_1, \ldots, x_n$ ces observations. Il est bien évident que la durée de vie des ampoules n'est pas prévisible avec certitude à l'avance. On va donc considérer que $x_1, \ldots, x_n$ sont les **réalisations** de variables aléatoires $X_1, \ldots, X_n$. Cela signifie qu'avant l'expérience, la durée de vie de la $i^{\text{ème}}$ ampoule est inconnue et que l'on traduit cette incertitude en modélisant cette durée par une variable aléatoire X_i. Mais après l'expérience, la durée de vie a été observée. Il n'y a donc plus d'incertitude, cette durée est égale au réel x_i. On dit que x_i est la réalisation de X_i sur l'essai effectué.

Puisque les ampoules sont identiques, il est naturel de supposer que les X_i sont de même loi. Cela signifie qu'on observe plusieurs fois le même phénomène aléatoire. Mais le hasard fait que les réalisations de ces variables aléatoires de même loi sont différentes, d'où la variabilité dans les données. Puisque les ampoules ont fonctionné indépendamment les unes des autres, on pourra également supposer que les X_i sont des variables aléatoires indépendantes. On peut alors se poser les questions suivantes :

1. Au vu de ces observations, est-il raisonnable de supposer que la durée de vie d'une ampoule est une variable aléatoire de loi exponentielle? Si non, quelle autre loi serait plus appropriée? C'est un problème de **choix de modèle** ou de **test d'adéquation**.
2. Si le modèle de loi exponentielle a été retenu, comment proposer une valeur (ou un ensemble de valeurs) vraisemblable pour le paramètre λ? C'est un problème d'**estimation paramétrique**.
3. Dans ce cas, peut-on garantir que λ est inférieur à une valeur fixée λ_0 ? Cela garantira alors que $E(X) = 1/\lambda \geq 1/\lambda_0$, autrement dit que les ampoules seront suffisamment fiables. C'est un problème de **test d'hypothèses paramétriques**.
4. Sur un parc de 100 ampoules, à combien de pannes peut-on s'attendre en moins de 50 h? C'est un problème de **prévision**.

Le premier problème central est celui de l'**estimation** : comment proposer, au vu des observations, une approximation des grandeurs inconnues du problème qui soit la plus proche possible de la réalité? La première question peut se traiter en estimant la fonction de répartition ou la densité de la loi de probabilité sous-jacente, la seconde revient à estimer un paramètre de cette loi, la troisième à estimer un nombre moyen de pannes sur une période donnée.

Le second problème central est celui des **tests d'hypothèses** : il s'agit de se prononcer sur la validité d'une hypothèse liée au problème : la loi est-elle exponentielle? λ est-il inférieur à λ_0? un objectif de fiabilité est-il atteint? En répondant oui ou non à ces questions, il est possible que l'on se trompe. Donc, à toute réponse statistique, il faudra associer le **degré de confiance** que l'on peut accorder à cette réponse. C'est une caractéristique importante de la statistique par rapport aux mathématiques classiques, pour lesquelles un résultat est soit juste, soit faux.

Pour résumer, la démarche probabiliste suppose que la nature du hasard est connue. Cela signifie que l'on adopte un modèle probabiliste particulier (ici la loi exponentielle), qui permettra d'effectuer des prévisions sur les observations futures. Dans la pratique, la nature du hasard est inconnue. La statistique va, au vu des observations, formuler des hypothèses sur la nature du phénomène aléatoire étudié. Maîtriser au mieux cette incertitude permettra de traiter les données disponibles. Probabilités et statistiques agissent donc en aller-retour dans le traitement mathématique des phénomènes aléatoires.

L'exemple des ampoules est une illustration du cas le plus fréquent où les données se présentent sous la forme d'une suite de nombres. C'est ce cas que nous traiterons dans ce cours, mais il faut savoir que les données peuvent être beaucoup plus complexes : des fonctions, des images, etc… Les principes et méthodes généraux que nous traiterons dans ce cours seront adaptables à tous les types de données.

Objectifs et plan du cours

Ce cours a pour but de présenter les principes de base d'une analyse statistique de données (description, estimation, tests), ainsi que les méthodes statistiques les plus usuelles. Ces méthodes seront toujours illustrées par des problèmes concrets. Le cours privilégie l'application à la théorie. Les méthodes présentées seront mises en œuvre à l'aide du logiciel R (https://www.r-project.org).

Le chapitre 2 présente les techniques de base en statistique descriptive, représentations graphiques et indicateurs statistiques. Le chapitre 3 est consacré aux problèmes d'estima- tion paramétrique ponctuelle, le chapitre 4 aux intervalles de confiance et le chapitre 5 aux tests d'hypothèses. Le dernier chapitre est consacré à une des méthodes statistiques les plus utilisées, la régression linéaire. Enfin, des annexes donnent quelques rappels de probabilités utiles en statistique, des tables des lois de probabilité usuelles et une courte introduction à R.

Chapter 2

Statistique descriptive

2.1 Introduction

La **statistique** est une méthode scientifique qui consiste à réunir des données chiffrées sur des ensembles nombreux, puis à analyser, à commenter et à critiquer ces données. Il ne faut pas confondre ***la*** statistique qui est la science qui vient d'être définie et ***une*** statistique qui est un ensemble de données chiffrées sur un sujet précis.

Les premières statistiques correctement élaborées ont été celles des recensements **démographiques**. Ainsi le vocabulaire statistique est essentiellement celui de la démographie.

Les ensembles étudiés sont appelés **population**. Les éléments de la population sont appelés **individus** ou unités statistiques. La population est étudiée selon un ou plusieurs **variables** (ou **caractères**).

2.2 Echantillonnage statistique

Pour recueillir des informations sur une population statistique, l'on dispose de deux méthodes :

- la **méthode exhaustive** ou **recensement** où chaque individu de la population est étudié selon le ou les caractères étudiés.
- la **méthode des sondages** ou **échantillonnage** qui conduit à n'examiner qu'une fraction de la population, un ***échantillon***.

Définition 2.1. L'**échantillonnage** représente l'ensemble des opérations qui ont pour objet de prélever un certain nombre d'individus dans une population donnée.

Pour que les résultats observés lors d'une étude soient généralisables à la population statistique, l'*échantillon doit être représentatif* de cette dernière, c'est à dire qu'il doit refléter fidèlement sa composition et sa complexité. Seul l'**échantillonnage aléatoire** assure la représentativité de l'échantillon.

> Un échantillon est qualifié d'**aléatoire** lorsque chaque individu de la population a une **probabilité connue et non nulle** d'appartenir à l'échantillon.
>
> Le cas particulier le plus connu est celui qui affecte à chaque individu la même probabilité d'appartenir à l'échantillon.

Echantillonnage aléatoire simple

Définition 2.2. L'échantillonnage aléatoire simple est une méthode qui consiste à prélever au hasard et de **façon indépendante**, n individus ou unités d'échantillonnage d'une population à N individus.

Chaque individu possède ainsi la **même probabilité** de faire partie d'un échantillon de n individus et chacun des échantillons possibles de taille n possède la même probabilité d'être constitué. L'échantillonnage aléatoire simple assure l'**indépendance des erreurs**, c'est-à-dire l'absence d'*autocorrélations* parmi les données relatives à un même caractère. Cette indépendance est indispensable à la validité de plusieurs **tests statistiques**.

Exemple : Les données météorologiques ne sont pas indépendantes puisque les informations recueillies sont d'autant plus identiques qu'elles sont rapprochées dans le temps et dans l'espace.

Il existe d'autres techniques d'échantillonnage que nous ne développerons pas dans ce cours comme l'échantillonnage systématique ou l'échantillonnage stratifié.

2.3 Statistique et Probabilités

Les concepts qui viennent d'être présentés sont les homologues de concepts du calcul des probabilités et il est possible de disposer en regard les concepts homologues (voir table ci-dessous).

Probabilités

Statistique

Espace fondamental

Population

Epreuve

Tirage (d'un individu), expérimentation

Evènement élémentaire

Individu, observation

Variable aléatoire

Caractère

Epreuves répétées

Echantillonnage

Nbre de répétitions d'une épreuve

Taille de l'échantillon, effectif total

Probabilité

Fréquence observée

Loi de probabilité

Distribution observée ou loi empirique

Espérance mathématique

Moyenne observée

Variance

Variance observée

Ainsi la notion de **caractère** se confond avec celle de **variable aléatoire**.

2.4 Description d'une série de valeurs

On considère ici une série (un ensemble) de valeurs, numériques, ou non, homogènes en ce sens qu'elles se réfèrent à une même variable et qu'elles ne sont pas structurées en sous-ensembles. Chaque valeur est associée à un individu statistique (unité statistique, observation). C'est le cas, par exemple, des notes obtenus par une promotion d'élèves à un examen. Dans cet exemple, lorsqu'il y a plusieurs examens, on peut vouloir considérer ensemble les notes d'un même élève. Les notes sont alors structurées en sous-ensembles et les méthodes à utiliser diffèrent de celles présentées ici.

Pour décrire une telle série, l'examen direct des valeurs n'est pas commode dès lors que ces valeurs sont un tant soir peu nombreuses. Pour cela, la statistique propose deux types d'outils : des graphiques et des indicateurs.

2.5 Graphiques

2.5.1 Variable qualitative nominale

Données

Diagramme en bâtons

Diagrammes circulaires

2.5.2 Variable qualitative ordinale

2.5.3 Variable quantitative (continue)

Diagramme en bâtons

Histogrammes

2.6 Indicateurs

2.6.1 Tendance centrale

Moyenne Médiane

2.6.2 Dispersion

Etendue Quantiles Ecart absolu moyen Ecart-type et variance Coefficient de variation

Chapter 3

Échantillonnage et Théorèmes limites

3.1 Échantillonnage

L'étude d'une caractéristique d'une pièce fabriquée en grand nombre (telle que la luminosité d'une ampoule, sa durée de vie ou encore le diamètre d'une pièce mécanique) relève, elle aussi, de la statistique descriptive. Il n'est toutefois pas possible de mesurer cette caractéristique sur toutes les pièces produites. Il est alors nécessaire de se limiter a l'étude des éléments d'un échantillon. Cet échantillon devra répondre a des critères particuliers pour pouvoir représenter la population toute entière dans l'étude statistique.

La démarche statistique présente plusieurs étapes :

- **Prélèvement d'un échantillon représentatif de la population ou échantillon aléatoire** par des techniques appropriées. Cela relève de la théorie de l'échantillonnage.
- **Étude des caractéristiques de cet échantillon**, issu d'une population dont on connaît la loi de probabilité. On s'intéresse principalement à ceux issus d'une population gaussienne.

3.2 Théorèmes limites

3.2.1 Introduction

Ce chapitre introduit trois résultats importants de la théorie asymptotique des probabilités: la loi faible des grands nombres, la loi forte des grands nombres et le théorème central limite, dans sa version pour variables aléatoires indépendantes et identiquement distribuées (X_i suivent la même loi pour $i = 1, \ldots, n$). Ce sont des résultats qui traitent les propriétés de la distribution de la **moyenne d'une suite de variables aléatoires** ($\overline{X}_n$).

Les deux lois des grands nombres énoncent les conditions sous lesquelles la moyenne d'une suite de variables aléatoires converge vers leur espérance commune et expriment l'idée que lorsque le nombre d'observations augmente, la différence entre la valeur attendue ($\mu = E(X_i)$) et la valeur observée ($\overline{X}_n$) tend vers zéro. De son côté, le théorème central limite établit que la distribution standardisée d'une moyenne tend asymptotiquement vers une loi normale, et cela même si la distribution des variables sous-jacentes est non normale. Ce résultat est central en probabilités et statistique et peut être facilement illustré (cf. figure 3.1). Indépendamment de la distribution sous-jacente des observations (ici une loi uniforme), lorsque n croît, la distribution de $\overline{X}_n$ tend vers une loi normale: on observe dans l'illustration la forme de plus en plus symétrique de la distribution ainsi que la concentration autour de l'espérance (ici $\mu = 0.5$) et la réduction de la variance.

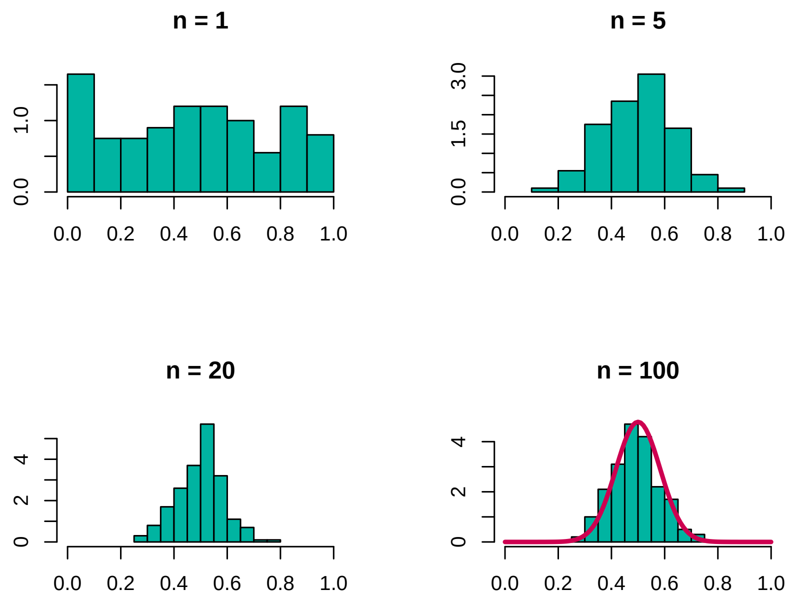


Figure 3.1: Illustration du théorème central limite: histogramme de la moyenne de 200 échantillons issus d'une loi uniforme sur l'intervalle (0,1) en fonction de la taille n de l'échantillon.

Les retombées pratiques de ces résultats sont importantes. En effet, la moyenne de variables aléatoires est une quantité qui intervient dans plusieurs procédures statistiques. Aussi, le résultat du théorème central limite permet l'approximation des probabilités liées à des sommes de variables aléatoires (e.g. méthode de Monte-Carlo). De plus, lorsque l'on considère des modèles statistiques, le terme d'erreur représente la somme de beaucoup d'erreurs (erreurs de mesure, variables non considérées, etc.). En prenant comme justification le théorème central limite, ce terme d'erreur est souvent supposé se comporter comme une loi normale.

3.2.2 Loi faible des grands nombres

Théorème 3.1 (Loi faible des grands nombres) Soit $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées. On suppose que $E(|X_i|) < \infty$ et que tous les X_i admettent la même espérance $E(X_i) = \mu$. Pour tout $\epsilon > 0$

$$\lim_{n \to \infty} P(|\overline{X}_n - \mu| > \epsilon) = 0,$$

c'est-à-dire $\overline{X}_n$ converge en probabilité vers μ.

3.2.3 Loi forte des grands nombres

Théorème 3.2 (Loi forte des grands nombres) Soit $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées. On suppose que $E(|X_i|) < \infty$ et que tous les X_i admettent la même espérance $E(X_i) = \mu$. Pour tout $\epsilon > 0$

$$P\big(\lim_{n \to \infty} \overline{X}_n = \mu\big) = 1$$

On dit que $\overline{X}_n$ converge presque sûrement vers μ.

3.2.4 Théorème central limite

Théorème 3.3 (Théorème central limite) Soit $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées, d'espérance μ et variance σ^2 finie. Alors

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{n\to\infty} \mathcal{N}(0,1) \tag{3.1}$$

en distribution.

On voit bien que, afin que la convergence se fasse, une standardisation est nécessaire: en effet, on peut voir le rapport dans (3.1) comme

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{var(\overline{X}_n)}}$$

Notes Historiques

La loi faible des grands nombres a été établie la première fois par J. Bernoulli pour le cas particulier d'une variable aléatoire binaire ne prenant que les valeurs 0 ou 1. Le résultat a été publié en 1713.

La loi forte des grands nombres est due au mathématicien E. Borel (1871- 1956), d'où parfois son autre appellation: théorème de Borel.

Le théorème central limite a été formulé pour la première fois par A. de Moivre en 1733 pour approximer le nombre de "piles" dans le jet d'une pièce de monnaie équilibrée. Ce travail a été un peu oublié jusqu'à ce que P.S. Laplace ne l'étende à l'approximation d'une loi binomiale par la loi normale dans son ouvrage *Théorie analytique des probabilités* en 1812. C'est dans les premières années du XX^e siècle que A. Lyapounov l'a redéfini en termes généraux et prouvé avec rigueur.

Chapter 4

Estimation ponctuelle

Chapter 5

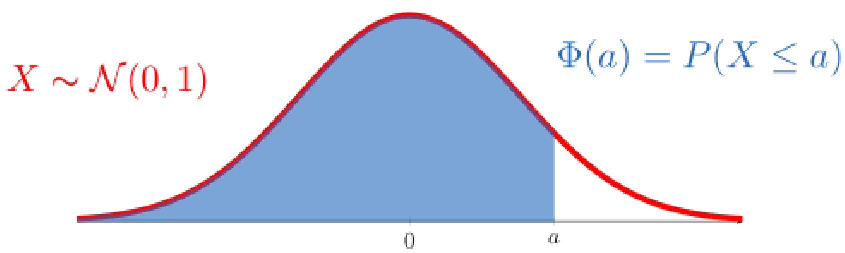
Intervalle de confiance

Chapter 6

Tests d'hpothèses

Appendix A

Table de la loi Normale centrée réduite



0

0.01

0.02

0.03

0.04

0.05

0.06

0.07

0.08

0.09

0

0.5000

0.5040

0.5080

0.5120

0.5160

0.5199

0.5239

0.5279

0.5319

0.5359

0.1

0.5398

0.5438

0.5478

0.5517

0.5557

0.5596

0.5636

0.5675

0.5714

0.5753

0.2

0.5793

0.5832

0.5871

0.5910

0.5948

0.5987

0.6026

0.6064

0.6103

0.6141

0.3

0.6179

0.6217

0.6255

0.6293

0.6331

0.6368

0.6406

0.6443

0.6480

0.6517

0.4

0.6554

0.6591

0.6628

0.6664

0.6700

0.6736

0.6772

0.6808

0.6844

0.6879

0.5

0.6915

0.6950

0.6985

0.7019

0.7054

0.7088

0.7123

0.7157

0.7190

0.7224

0.6

0.7257

0.7291

0.7324

0.7357

0.7389

0.7422

0.7454

0.7486

0.7517

0.7549

0.7

0.7580

0.7611

0.7642

0.7673

0.7704

0.7734

0.7764

0.7794

0.7823

0.7852

0.8

0.7881

0.7910

0.7939

0.7967

0.7995

0.8023

0.8051

0.8078

0.8106

0.8133

0.9

0.8159

0.8186

0.8212

0.8238

0.8264

0.8289

0.8315

0.8340

0.8365

0.8389

1

0.8413

0.8438

0.8461

0.8485

0.8508

0.8531

0.8554

0.8577

0.8599

0.8621

1.1

0.8643

0.8665

0.8686

0.8708

0.8729

0.8749

0.8770

0.8790

0.8810

0.8830

1.2

0.8849

0.8869

0.8888

0.8907

0.8925

0.8944

0.8962

0.8980

0.8997

0.9015

1.3

0.9032

0.9049

0.9066

0.9082

0.9099

0.9115

0.9131

0.9147

0.9162

0.9177

1.4

0.9192

0.9207

0.9222

0.9236

0.9251

0.9265

0.9279

0.9292

0.9306

0.9319

1.5

0.9332

0.9345

0.9357

0.9370

0.9382

0.9394

0.9406

0.9418

0.9429

0.9441

1.6

0.9452

0.9463

0.9474

0.9484

0.9495

0.9505

0.9515

0.9525

0.9535

0.9545

1.7

0.9554

0.9564

0.9573

0.9582

0.9591

0.9599

0.9608

0.9616

0.9625

0.9633

1.8

0.9641

0.9649

0.9656

0.9664

0.9671

0.9678

0.9686

0.9693

0.9699

0.9706

1.9

0.9713

0.9719

0.9726

0.9732

0.9738

0.9744

0.9750

0.9756

0.9761

0.9767

2

0.9772

0.9778

0.9783

0.9788

0.9793

0.9798

0.9803

0.9808

0.9812

0.9817

2.1

0.9821

0.9826

0.9830

0.9834

0.9838

0.9842

0.9846

0.9850

0.9854

0.9857

2.2

0.9861

0.9864

0.9868

0.9871

0.9875

0.9878

0.9881

0.9884

0.9887

0.9890

2.3

0.9893

0.9896

0.9898

0.9901

0.9904

0.9906

0.9909

0.9911

0.9913

0.9916

2.4

0.9918

0.9920

0.9922

0.9925

0.9927

0.9929

0.9931

0.9932

0.9934

0.9936

2.5

0.9938

0.9940

0.9941

0.9943

0.9945

0.9946

0.9948

0.9949

0.9951

0.9952

2.6

0.9953

0.9955

0.9956

0.9957

0.9959

0.9960

0.9961

0.9962

0.9963

0.9964

2.7

0.9965

0.9966

0.9967

0.9968

0.9969

0.9970

0.9971

0.9972

0.9973

0.9974

2.8

0.9974

0.9975

0.9976

0.9977

0.9977

0.9978

0.9979

0.9979

0.9980

0.9981

2.9

0.9981

0.9982

0.9982

0.9983

0.9984

0.9984

0.9985

0.9985

0.9986

0.9986

3

0.9987

0.9987

0.9987

0.9988

0.9988

0.9989

0.9989

0.9989

0.9990

0.9990

Par exemple, pour $x = 1.23$ (intersection de la ligne 1.2 et de la colonne 0.03), on obtient : $\Phi(1.23) \approx 0.8907$.