

Mathématiques

Statistiques

Stéphane Perret

Version 3.010

Table des matières

Introduction	1
1 Variables aléatoires	3
1.1 Variables aléatoires discrètes	3
1.1.1 Jets d'une pièce bien équilibrée	3
1.1.2 Jets d'un dé à six faces bien équilibré	5
1.1.3 Variables aléatoires discrètes, espérance et variance	8
1.1.4 Le théorème de Tchebychev	10
1.1.5 Jets de plusieurs dés à six faces bien équilibrés	11
1.1.6 La loi de Bernoulli	13
1.1.7 La loi binomiale	13
1.1.8 La loi des jets de k dés	14
1.2 Variables aléatoires continues	15
1.2.1 La loi uniforme	15
1.2.2 La loi normale	16
1.2.3 La fonction gamma	18
1.2.4 Les lois de Student	19
1.2.5 Les lois du chi-carré	20
1.2.6 Espérance et variance dans le cas continu	21
1.3 Les différentes formes de densité de distribution	22
1.4 Propriétés de l'espérance et de la variance (cas discret et continu)	24
1.5 Centrage et réduction	25
1.6 Fonction de répartition d'une variable aléatoire	26
1.6.1 Fonction de répartition d'une variable discrète	26
1.6.2 Fonction de répartition d'une variable continue	26
1.7 Deux propriétés essentielles de la loi normale	27
1.7.1 La loi normale permet d'approximer la loi binomiale	27
1.7.2 Le théorème de la limite centrale	29
1.8 Tables	30
1.8.1 Fonction de répartition de la loi normale	30
1.8.2 Les quantiles des lois de Student	31
1.8.3 Les quantiles des lois du chi-carré	32
2 Estimation de paramètres	33
2.1 Paramètres de position	33
2.2 Paramètres de dispersion	34
2.3 Les boîtes à moustaches	36
2.3.1 Les quartiles de Tukey et les quartiles de Freund et Perles	36
2.3.2 La boîte à moustaches de Tukey	37

2.4	Statistique, estimateur et estimation	38
2.5	La méthode du maximum de vraisemblance	39
2.6	Qualités d'un estimateur	40
3	Tests d'hypothèses	43
3.1	Tests d'hypothèses symétriques sur une moyenne	44
3.1.1	Tests d'hypothèses symétriques sur une moyenne, variance connue	44
3.1.2	Test d'hypothèses symétriques sur une moyenne, variance inconnue	48
3.1.3	Résumé et autres statistiques de tests sur les moyennes	50
3.2	Test du chi-carré : adéquation à une loi	52
3.3	Test du chi-carré : comparaison d'échantillons	55
3.4	Test du chi-carré : indépendance	57
3.5	La p -valeur associée à un test d'hypothèse	59
4	Intervalles de confiance	61
4.1	L'intervalle de confiance sur une moyenne, variance connue	62
4.2	L'intervalle de confiance sur une moyenne, variance inconnue	63
5	Régression linéaire	65
5.1	La droite des moindres carrés	66
5.2	Le coefficient de corrélation	68
5.3	Le coefficient de détermination	69
5.4	La droite des moindres carrés forcée à l'origine	70
5.4.1	Les coefficients de détermination et de corrélation	71
5.4.2	Preuves	71
5.5	Autres types de régression	72
5.5.1	Préambule : une autre vision de la régression linéaire	72
5.5.2	Régression quadratique	72
5.5.3	Régression hyperbolique	72
5.5.4	Régression exponentielle	73
5.5.5	Régression d'une puissance	73
5.5.6	Régression logarithmique	73
5.6	Preuves des théorèmes	74
5.6.1	Preuve des théorèmes des moindres carrés	74
5.6.2	Preuves de la relation «miraculeuse»	76
5.6.3	Les deux visions pour la droite de régression	76
5.6.4	Preuve des ingrédients pour le modèle linéaire	77
5.6.5	Preuve du théorème de retrouvailles	77
5.6.6	Preuve des ingrédients pour le modèle quadratique	78

Introduction

Le monde des statistiques se sépare en deux parties :

1. Les statistiques descriptives.

Elles consistent à travailler sur les données mesurées afin d'en extraire l'essentiel à l'aide de méthodes graphiques (histogrammes, boîtes à moustaches, droite des moindres carrés) ou de paramètres de positions et de dispersion (comme la moyenne, le mode, la médiane; l'étendue, la variance et l'écart type).

2. Les statistiques inférentielles.

Elles utilisent des résultats théoriques (comme la loi des grands nombres ou le théorème de la limite centrale) afin de pouvoir créer des modèles généraux à partir des informations données par les statistiques descriptives dans le but d'effectuer des analyses prédictives ou de déduire des comportements généraux d'une population.

Statistique descriptive	Statistique inférentielle
Mesures expérimentales	Modèle physique (formule)
Mesures météorologiques	Prévision du temps à court terme
Sondage sur un échantillon	Déduction sur la population globale
Récolte des données (SuperCard)	Agencement des produits dans les rayons
Tests placebos	Détermination de l'efficacité d'un médicament

Vastes champs d'applications

De nos jours, les probabilités et les statistiques sont utilisées de plus en plus dans beaucoup de domaines dont, entres autres :

Traitement du signal	Biologie	Physique	Jeux vidéos
Hydrologie et climatologie	Médecine	Sport	Marketing
Sciences économiques	Psychologie	Sciences sociales	Criminologie

Le danger des statistiques

Les statistiques font parties du monde des mathématiques : il s'agit d'une science exacte ! Néanmoins, la réalité étant ce qu'elle est (imparfaite telle l'homme qui l'étudie), la plupart des théorèmes en statistiques sont utilisés sans qu'on puisse être certains que les hypothèses soient vraiment vérifiées. Il est souvent moral de se dire que l'on peut appliquer tel ou tel théorème. C'est une faiblesse¹ des outils statistiques que l'on ne peut négliger ! Heureusement, cette faiblesse est compensée par la robustesse de la plupart de certains théorèmes.

1. Les modèles utilisés par les physiciens souffrent des mêmes faiblesses. La théorie de Newton a été surclassée par la relativité d'Einstein, qui a aussi été surclassée par la relativité générale (macrophysique). Cette dernière est actuellement incompatible avec la mécanique quantique (microphysique).

Chapitre 1

Variables aléatoires

Dans l'édition 2004 du petit Robert, on trouve sous «le hasard» la définition suivante.

«Cause fictive de ce qui arrive sans raison apparente ou explicable, souvent personnifiée au même titre que le sort, la fortune, etc. [...] Se dit à l'occasion d'un heureux concours de circonstances.»

Dans ce domaine que sont les statistiques, on étudie la structure qui se cache derrière quelques processus aléatoires. On y verra que lorsqu'on répète un tel processus un grand nombre de fois, «le hasard» s'efface pour laisser place à des lois statistiques.

1.1 Variables aléatoires discrètes

1.1.1 Jets d'une pièce bien équilibrée

Considérons une pièce de monnaie parfaitement bien équilibrée. Théoriquement, une telle pièce tombe une fois sur deux sur pile, et une fois sur deux sur face. En effet, par les propriétés de symétrie et d'homogénéité de la pièce, chacun de ses côtés à la même probabilité d'apparaître.

Il y a des pièces truquées dont les deux côtés sont les mêmes (deux côtés pile par exemple). On peut aussi truquer une pièce à l'aide d'une presse¹.

Effectuons 100 lancers de pièces en notant F pour face et P pour pile. On va étudier l'évolution de la proportion des piles sur le nombre de lancers.

Les 100 lancers sont

P	P	F	F	P	P	P	F	P	P	P	F	F	F	F	F	F	P	P	
P	F	P	F	F	F	P	P	F	F	P	F	F	F	P	F	P	P	P	F
F	P	F	F	P	P	P	F	F	F	P	P	P	P	P	P	P	F	P	P
P	P	F	F	P	F	P	P	P	P	P	F	P	F	F	F	P	P	F	F
F	P	P	P	F	F	F	P	P	P	F	F	P	P	P	F	F	P	F	P

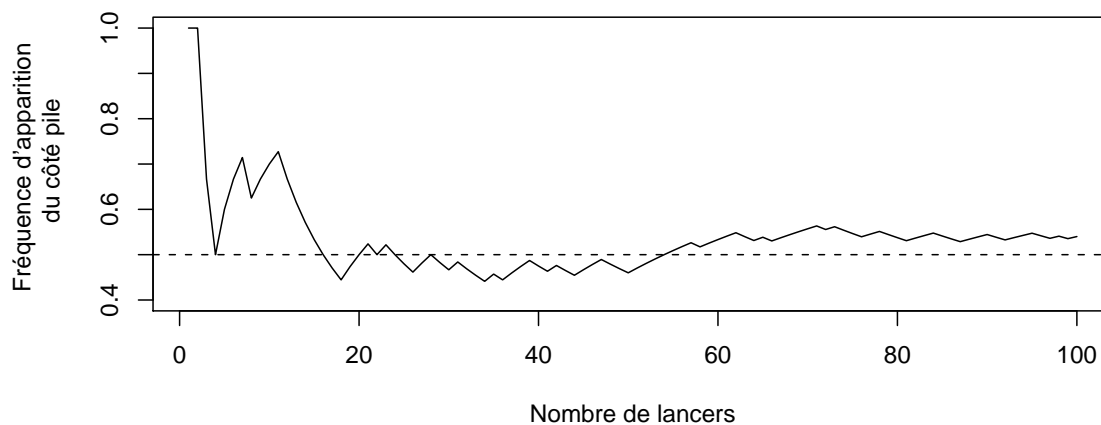
Comme le côté pile est sorti 54 fois, sa fréquence d'apparition est de 0.54.

On remarque aussi qu'il y a une série de 7 piles de suite, et une série de 7 faces de suite. Si on devait demander à un humain d'écrire 100 piles ou faces de suite, il hésiterait fortement à en mettre 7 de suite et pourtant cela arrive fréquemment².

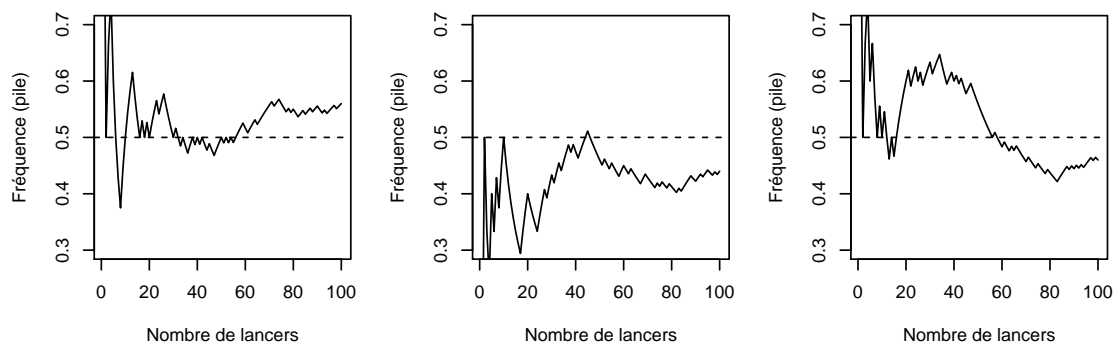
1. http://www.stat.columbia.edu/~cook/movabletype/archives/2008/01/a_sighting_of_t.html

2. Nicole Vogel montre, dans l'article «Peut-on imiter le hasard?» de la brochure APMEP 451, que sur une série de 100 lancers, il y a plus d'une chance sur deux d'avoir une série d'au moins 7 piles ou 7 faces. Disponible sur <https://www.apmep.fr/IMG/pdf/AAA04016.pdf>.

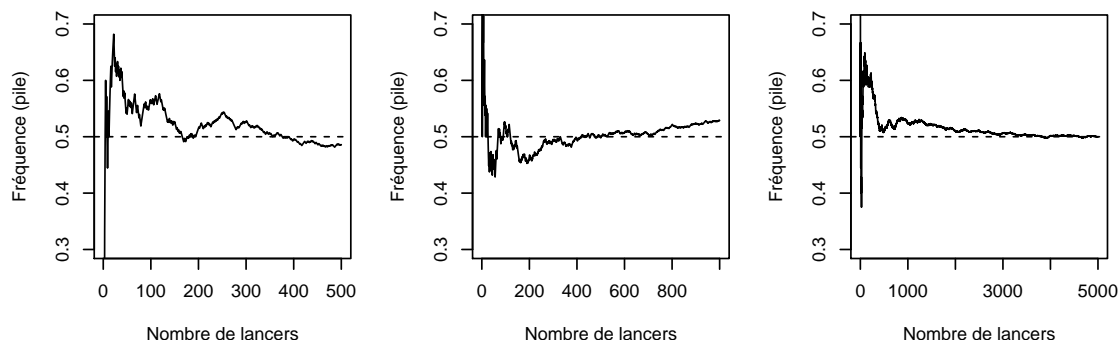
Voici la représentation graphique de l'évolution de la fréquence d'apparition du côté pile. On commence à 1, puisque le premier lancer a donné pile. Le deuxième lancer donne aussi pile, la fréquence reste à $\frac{2}{2} = 1$. Le troisième lancer donne F , ainsi la fréquence donne $\frac{2}{3} \cong 0.67$. Le quatrième lancer donne F , ainsi la fréquence donne $\frac{2}{4} \cong 0.5$. Là le graphe touche l'horizontale à hauteur $\frac{1}{2} = 0.5$. Rappelons qu'au 100-ième lancer, la fréquence valait 0.54.



Effectuons trois autres séries de 100 lancers pour sentir ce qui se passe.



Pour les trois graphes ci-dessous, on a d'abord lancé 500 pièces, puis 1000, puis 5000.



On voit que plus le nombre de lancers augmente, plus la fréquence s'approche de 0.5. Ce phénomène illustre le résultat fondamental suivant de la théorie des probabilités.

La loi des grands nombres

Lorsqu'on répète une expérience aléatoire de manière indépendante un grand nombre de fois, alors les proportions expérimentales se rapprochent des proportions théoriques, appelée *probabilités*.

1.1.2 Jets d'un dé à six faces bien équilibré

Considérons un dé à six faces bien équilibré. Théoriquement, la fréquence d'apparition de chaque face est de $\frac{1}{6}$. En effet, par les propriétés de symétrie et d'homogénéité du cube, chacune de ses faces a la même probabilité d'apparaître.

On trouve sur le marché des dés pipés qui ont un numéro qui sort dans le 80% des cas³.

En effectuant 60 lancers, on obtient les données suivantes.

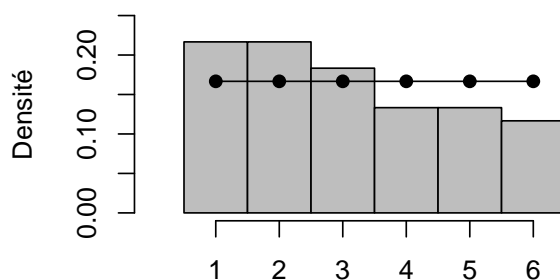
2 3 1 1 2 1 2 6 1 2 1 2 2 2 4 4 3 4 5 6
 3 2 5 3 6 5 1 4 4 2 3 6 2 5 6 4 3 5 1 1
 3 1 5 5 3 2 4 6 4 5 3 2 2 6 1 3 1 1 3 1

On construit, avec ces données, un *histogramme*. Pour chacune des issues possibles (ici les nombres de 1 à 6), on compte le nombre de fois qu'elle apparaît, c'est le nombre d'*occurrences*, puis on calcule la *densité* qui est le nombre d'occurrences divisé par le nombre total de jets.

face	1	2	3	4	5	6
nombre d'occurrences	13	13	11	8	8	7
densité	0.216	0.216	0.183	0.133	0.133	0.116

On fait ensuite un histogramme (voir le complément sur les histogrammes à la fin de cette sous-section en page 7).

Histogramme



L'histogramme permet de mieux visualiser les données. Ici, on voit que les faces 1, 2 et 3 sont sorties bien plus souvent. Il est possible que le dé soit pipé (avec une petite bille de plomb vers le sommet adjacent aux faces 4, 5 et 6).

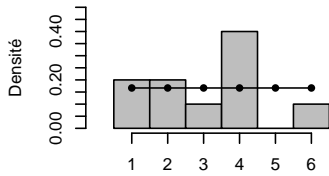
Néanmoins, ces données ne permettent pas de déduire que le dé est truqué. D'ailleurs un test du chi-carré indique que ce dé n'est probablement pas truqué (voir en page 52). Le test ne se trompe pas puisque ces données sont issues de simulations à partir d'un dé (informatique) non pipé.

La loi des grands nombres affirme que plus le nombre de lancers est grand plus les densités se rapprochent des probabilités (ou des densités théoriques) qui valent ici $\frac{1}{6} \cong 0.167$. Ces probabilités sont indiquées sur l'histogramme ci-dessus par les points reliés.

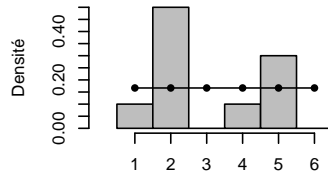
3. <http://www.toursdemagie.com/produit/magie/6-Des-Pipes---8-Des-truques.html>

En augmentant le nombre n de lancers effectués, on vérifie empiriquement la loi des grands nombres.

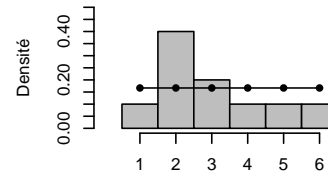
Histogramme pour $n = 10$



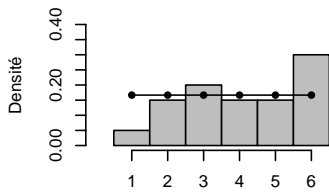
Histogramme pour $n = 10$



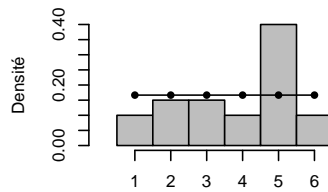
Histogramme pour $n = 10$



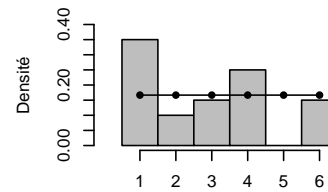
Histogramme pour $n = 20$



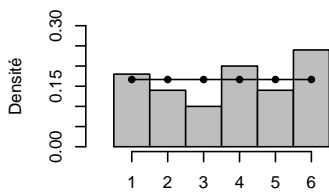
Histogramme pour $n = 20$



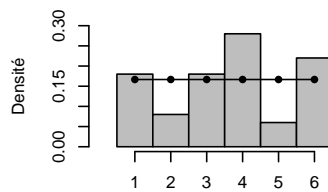
Histogramme pour $n = 20$



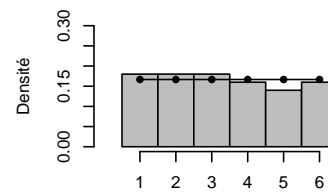
Histogramme pour $n = 50$



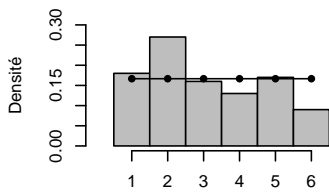
Histogramme pour $n = 50$



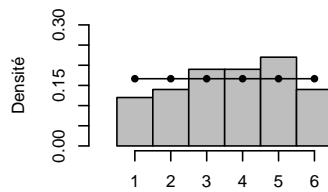
Histogramme pour $n = 50$



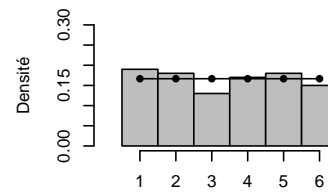
Histogramme pour $n = 100$



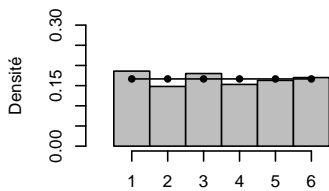
Histogramme pour $n = 100$



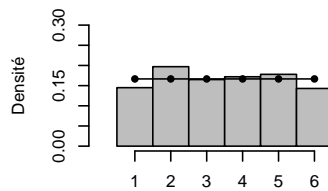
Histogramme pour $n = 100$



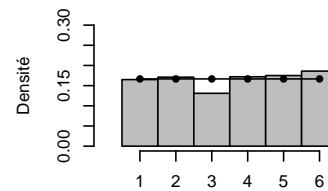
Histogramme pour $n = 1000$



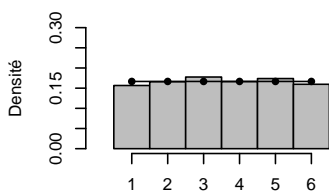
Histogramme pour $n = 1000$



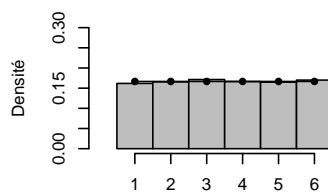
Histogramme pour $n = 1000$



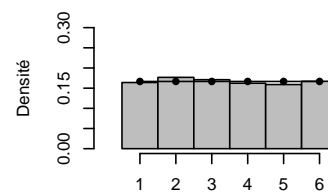
Histogramme pour $n = 5000$



Histogramme pour $n = 5000$



Histogramme pour $n = 5000$



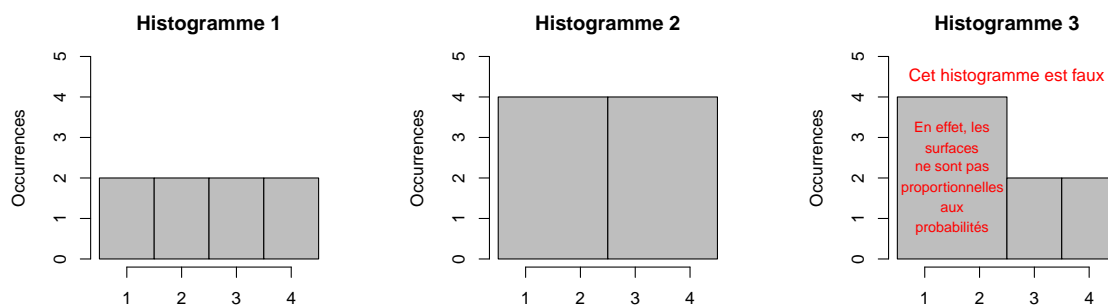
Compléments sur les histogrammes

Il y a deux sortes d'histogrammes : ceux qui ont en ordonnée les occurrences (c'est-à-dire le nombre de fois qu'une valeur mesurée apparaît ; dans l'exemple, il s'agissait du nombre montré par le dé) et ceux qui ont en ordonnée les densités.

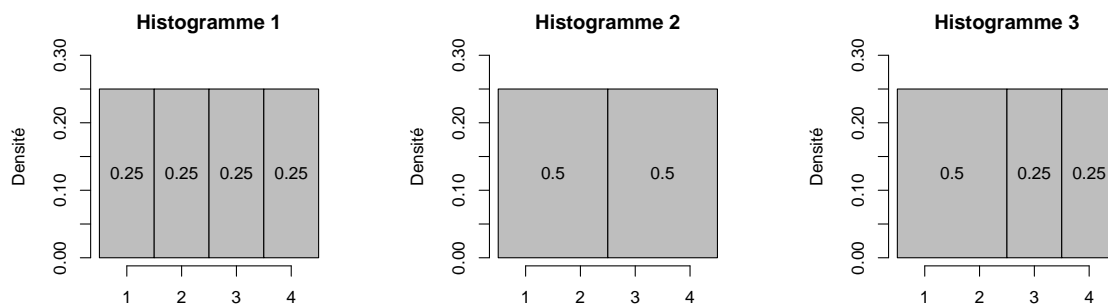
Imaginons que l'on lance un dé à quatre faces et que l'on obtienne les résultats suivants.

face	1	2	3	4
nombre d'occurrences	2	2	2	2
densité	0.25	0.25	0.25	0.25

Voici trois histogrammes des occurrences avec différentes façons de grouper les résultats.



Et ceux pour les densités (avec la même façon de grouper les résultats).



On voit le lien entre ces trois histogrammes et la densité de probabilité : les aires doivent correspondre aux probabilités (voir page 15).

Ces histogrammes s'obtiennent à l'aide du logiciel «R» (<http://www.r-project.org>) qui est la version gratuite du célèbre logiciel «S plus» (<http://stat.bell-labs.com/S>).

Voici ce qu'il faut taper dans «R» pour obtenir ces histogrammes (pour raccourcir le code, certains réglages ont été enlevés).

```
data <- c(1,1,2,2,3,3,4,4); layout(matrix(1:3, 1, 3, byrow=TRUE))
```

```
hist(data,breaks=c(0.5,1.5,2.5,3.5,4.5),prob=FALSE,col="gray",
      main=paste("Histogramme 1"),ylim=c(0,5),xlab="",ylab="Occurrences")
```

```
hist(data,breaks=c(0.5,2.5,4.5),prob=FALSE,col="gray",
      main=paste("Histogramme 2"),ylim=c(0,5),xlab="",ylab="Occurrences")
```

```
hist(data,breaks=c(0.5,2.5,3.5,4.5),prob=FALSE,col="gray",
      main=paste("Histogramme 3"),ylim=c(0,5),xlab="",ylab="Occurrences")
```

En mettant `prob=TRUE` au lieu de `prob=FALSE`, «R» passe directement des histogrammes des occurrences aux histogrammes des densités.

1.1.3 Variables aléatoires discrètes, espérance et variance

Lorsque les valeurs prises par la variable aléatoire X sont en nombre fini ou dénombrable, on dit que la variable aléatoire est *discrète*.

Définitions

L'ensemble des issues d'une expérience aléatoire est appelé *Univers* et noté Ω .

Une *variable aléatoire*⁴ est une fonction $X : \Omega \rightarrow \mathbb{R}$ qui associe à chaque issue de l'expérience en question un unique nombre réel.

On note $\mathbf{P}(X = x)$ pour la probabilité⁵ que la variable X prenne la valeur x (attention à bien distinguer majuscule (variable aléatoire) et minuscule (valeur prise par cette variable aléatoire)).

On définit l'*espérance* de X par $\mu = E(X)$ et la variance de X par $\sigma^2 = V(X)$ où

$$\mu = E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbf{P}(\omega) = \sum_{x \in \mathbb{R}} x \cdot \mathbf{P}(X = x)$$

On définit la *variance* de X par $\sigma^2 = V(X)$ où

$$\sigma^2 = V(X) = \sum_{x \in \mathbb{R}} (x - \mu)^2 \cdot \mathbf{P}(X = x)$$

On définit aussi l'*écart type* de X par $\sigma = \sqrt{\sigma^2}$.

Ces formules ne sont valables que lorsque $\mathbf{P}(X = x) \neq 0$ pour un nombre fini ou dénombrable⁶ de $x \in \mathbb{R}$ (ce qui est le cas pour une variable aléatoire discrète). C'est pour cette raison qu'on se permet de noter $\sum_{x \in \mathbb{R}}$.

Interprétations

1. L'espérance de X est la moyenne des issues pondérée par leurs probabilités.
2. La variance de X est la moyenne des carrés des écarts par rapport à l'espérance pondérée par les probabilités. Autrement dit, plus les valeurs prises par X s'éloignent de l'espérance, plus la variance augmente fortement. Quant aux effets des petits écarts sur la variance, ils sont presque négligeables.

Ces notions sont liées à la variable aléatoire, donc à la théorie des probabilités. Elles ne sont pas liées aux mesures obtenues lors d'expérimentations. On verra qu'on pourra estimer ces paramètres (espérance, variance et écart type) à l'aide de mesures dans le chapitre 2.

Subtilité! Contrairement à d'habitude, si X est une variable aléatoire, x est une valeur prise par cette variable aléatoire. Donc x est dans le domaine arrivée de X , qui est \mathbb{R} .

4. On comprendra mieux dans les chapitres qui suivent pourquoi on les appelle *variables aléatoires* alors que si on regarde la définition, il n'y a rien de variable, ni d'aléatoire.

5. Par rapport aux notions du cours DF, la probabilité que la variable aléatoire atteigne la valeur x est donnée par $\mathbf{P}(X = x) = \mathbf{P}(\{\omega \in \Omega : X(\omega) = x\})$.

6. dans ce cas, les séries sont infinies. Se référer au chapitre correspondant du cours OS.

Retour à l'exemple du jet d'un dé

En reprenant l'exemple précédent. On peut considérer la variable aléatoire X qui représente le nombre montré par le jet d'un dé.

On peut ainsi avoir une notation efficace pour les probabilités : la probabilité que le nombre montré par le dé soit entre 4 et 6 se note $\mathbb{P}(4 \leq X \leq 6)$. Puisque la probabilité d'avoir chaque nombre vaut $\frac{1}{6}$, on a

$$\mathbb{P}(4 \leq X \leq 6) = \mathbb{P}(X = 4) + \mathbb{P}(X = 5) + \mathbb{P}(X = 6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = 50\%$$

Le calcul de l'espérance est le suivant.

$$\begin{aligned} E(X) &= \mathbb{P}(X=1) + 2\mathbb{P}(X=2) + 3\mathbb{P}(X=3) + 4\mathbb{P}(X=4) + 5\mathbb{P}(X=5) + 6\mathbb{P}(X=6) \\ &= \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = \frac{7}{2} = 3.5 \end{aligned}$$

On vérifie que l'espérance n'est rien d'autre que la moyenne des résultats possibles.

$$E(X) = \frac{1+2+3+4+5+6}{6}$$

Le calcul de la variance est le suivant.

$$\begin{aligned} V(X) &= (1-\mu)^2 \cdot \mathbb{P}(X=1) + (2-\mu)^2 \cdot \mathbb{P}(X=2) + (3-\mu)^2 \cdot \mathbb{P}(X=3) \\ &\quad + (4-\mu)^2 \cdot \mathbb{P}(X=4) + (5-\mu)^2 \cdot \mathbb{P}(X=5) + (6-\mu)^2 \cdot \mathbb{P}(X=6) \\ &= \frac{(1-\frac{7}{2})^2}{6} + \frac{(2-\frac{7}{2})^2}{6} + \frac{(3-\frac{7}{2})^2}{6} + \frac{(4-\frac{7}{2})^2}{6} + \frac{(5-\frac{7}{2})^2}{6} + \frac{(6-\frac{7}{2})^2}{6} \\ &= \frac{\frac{25}{4}}{6} + \frac{\frac{9}{4}}{6} + \frac{\frac{1}{4}}{6} + \frac{\frac{1}{4}}{6} + \frac{\frac{9}{4}}{6} + \frac{\frac{25}{4}}{6} = \frac{70}{24} = \frac{35}{12} = 2.91\bar{6} \end{aligned}$$

L'écart type vaut $\sqrt{\frac{35}{12}} \cong 1.71$.

L'espérance de gain

Quand la variable aléatoire donne des valeurs en CHF, on parle d'*espérance de gain* au lieu d'espérance (mais cela ne change rien à sa définition).

Dans ce cas, l'unité de l'espérance est le CHF. Quant à l'unité de la variance, il s'agit de CHF² ! Ainsi, l'unité de l'écart type est le CHF. Il est donc plus naturel de parler de l'écart type lorsqu'il y a une unité.

Par exemple, imaginons que deux personnes Anka et Brad jouent à pile ou face. Si la pièce tombe sur pile, alors Anka donne 2 CHF à Brad ; si la pièce tombe sur face, alors c'est Brad qui donne 4 CHF à Anka. On peut traduire cette règle par une variable aléatoire $X : \{\text{pile, face}\} \rightarrow \mathbb{R}$ qui gère la fortune de Anka : $X(\text{pile}) = -2$ CHF et $X(\text{face}) = 4$ CHF.

Pour cette variable X , l'espérance vaut $E(X) = -2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{2} = 1$ CHF. La variance vaut $V(X) = (-2-1)^2 \cdot \frac{1}{2} + (4-1)^2 \cdot \frac{1}{2} = 9$ CHF². L'écart type vaut $\sigma = 3$ CHF.

L'espérance de gain mesure le gain moyen : en jouant 1000 fois à un tel jeu, Anka est en droit d'espérer gagner 1000 CHF (si elle n'est ni trop chanceuse, ni trop malchanceuse). L'écart type est en relation avec le risque.

Généralisation de l'espérance

Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction réelle et X une variable aléatoire, alors $g \circ X$ est une variable aléatoire notée $g(X)$ dont l'espérance est :

$$E(g(X)) = \sum_{\omega \in \Omega} g(X(\omega)) \cdot \mathbf{P}(\omega) = \sum_{x \in \mathbb{R}} g(x) \cdot \mathbf{P}(X = x)$$

Cela permet notamment d'écrire la variance comme une espérance.

$$V(X) = \sum_{x \in \mathbb{R}} (x - \mu)^2 \cdot \mathbf{P}(X = x) = E((X - \mu)^2)$$

1.1.4 Le théorème de Tchebychev

Soit X une variable aléatoire d'espérance μ et d'écart type σ . Alors, on a

$$\mathbf{P}(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2} \quad \text{pour tout } k > 0$$

Cas particuliers

1. Pour $k = 1$, on a $\mathbf{P}(\mu - \sigma \leq X \leq \mu + \sigma) \geq 0\%$.

Cette affirmation n'est pas particulièrement utile.

2. Pour $k = 2$, on a $\mathbf{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \geq 75\%$.

Ainsi il y a au moins 75% des valeurs de X qui sont dans l'intervalle $[\mu - 2\sigma, \mu + 2\sigma]$.

3. Pour $k = 3$, on a $\mathbf{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \geq 88.\bar{8}\%$.

Ainsi il y a au moins 88% des valeurs de X qui sont dans l'intervalle $[\mu - 3\sigma, \mu + 3\sigma]$.

Illustration dans l'exemple du jeu d'argent de la page précédente

Dans cet exemple, on a $\mu = 1$ et $\sigma = 3$. On examine les gains d'Anka pour une partie.

1. La probabilité que la somme gagnée soit dans l'intervalle $[\mu - \sigma, \mu + \sigma] = [-2, 4]$ doit être plus grande ou égale à 0%.

C'est vrai, car cette probabilité vaut 100%.

2. La probabilité que la somme gagnée soit dans l'intervalle $[\mu - 2\sigma, \mu + 2\sigma] = [-5, 7]$ doit être plus grande ou égale à 75%.

C'est vrai, car cette probabilité vaut 100%.

3. La probabilité que la somme gagnée soit dans l'intervalle $[\mu - 3\sigma, \mu + 3\sigma] = [-8, 10]$ doit être plus grande ou égale à 88.\bar{8}\%.

C'est vrai, car cette probabilité vaut 100%.

Conclusion Le théorème ne donne pas des valeurs optimales, mais il est vrai quelque soit la variable aléatoire. Ici, on a pris un exemple un peu trop simple, ce qui explique la faible performance du théorème.

7. Cette notation est abusive : $g(X)(\omega)$ n'a pas de sens ! La justification est celle-ci : si $X : \Omega \rightarrow \mathbb{R}$ et $g : \mathbb{R} \rightarrow \mathbb{R}$, alors $(g \circ X)(\omega) = g(X(\omega)) = g(x)$ avec $x = X(\omega) \in \text{Im}(X)$. D'où la notation $g \circ X = g(X)$.

1.1.5 Jets de plusieurs dés à six faces bien équilibrés

Considérons plusieurs dés à six faces bien équilibrés. On examine la somme obtenue des valeurs montrées par un jet de plusieurs dés.

Jets de deux dés

Voici un tableau représentant la somme, notée Σ , de chacune des $6^2 = 36$ issues.

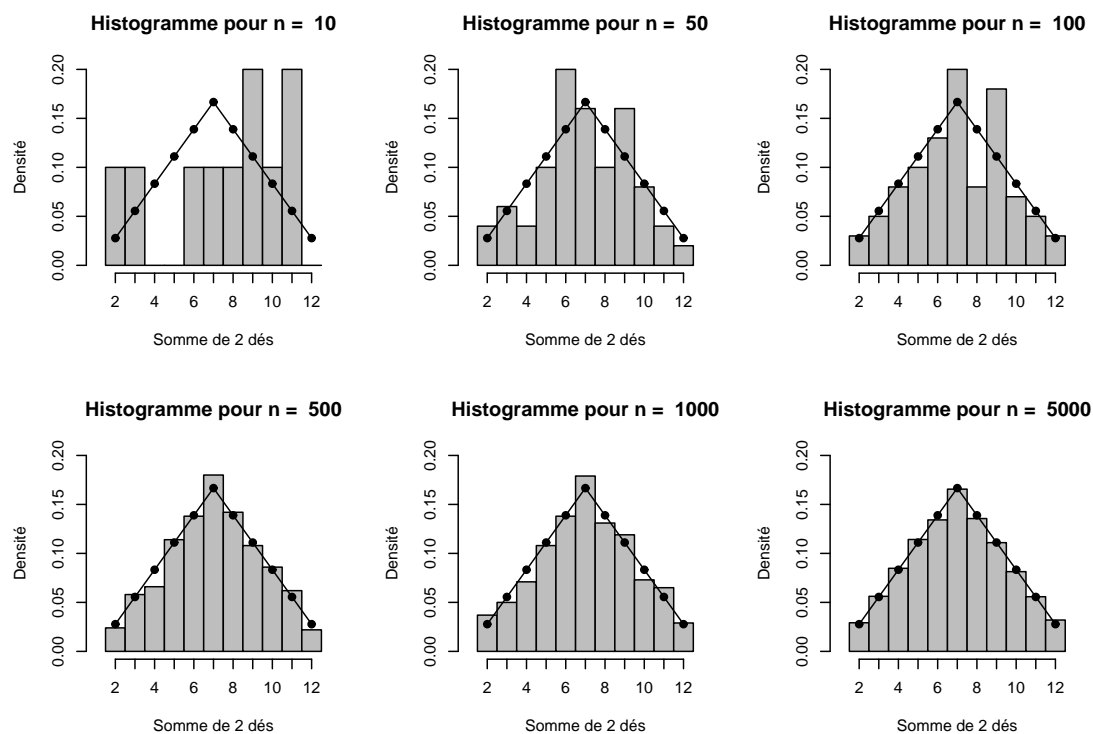
Issue	Σ	Issue	Σ	Issue	Σ	Issue	Σ	Issue	Σ	Issue	Σ
	2		3		4		5		6		7
	3		4		5		6		7		8
	4		5		6		7		8		9
	5		6		7		8		9		10
	6		7		8		9		10		11
	7		8		9		10		11		12

Si on suppose que les dés sont bien équilibrés, alors chaque issue a autant de chances de se produire (c'est pour cette raison qu'on a tenu compte de l'ordre). On trouve ainsi le tableau des probabilités suivant.

somme	2	3	4	5	6	7	8	9	10	11	12
probabilité	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

La loi des grands nombres affirme que plus le nombre n de lancers est grand plus les densités se rapprochent des probabilités ci-dessus. Ces probabilités sont indiquées sur les histogrammes ci-dessous par les points reliés.

Voici une vérification empirique de la loi des grands nombres.



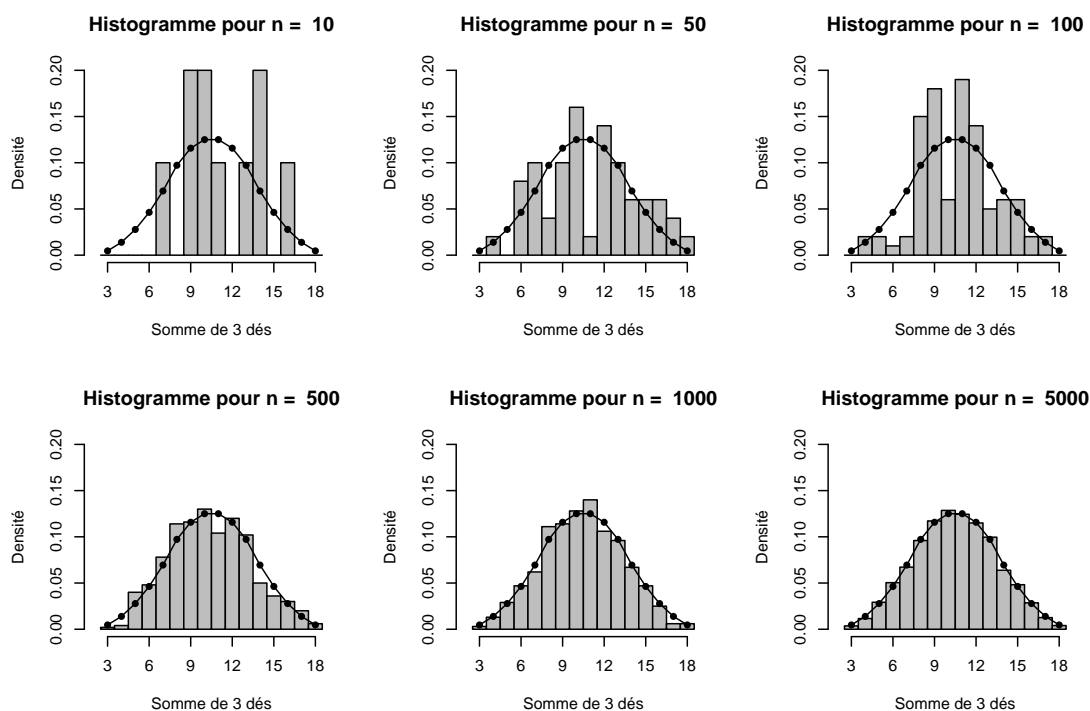
Jets de trois dés

Lorsqu'on jette 3 dés, il y a $6^3 = 216$ issues. On a le tableau de probabilités suivant.

somme	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
proba	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$	$\frac{27}{216}$	$\frac{25}{216}$	$\frac{21}{216}$	$\frac{15}{216}$	$\frac{10}{216}$	$\frac{6}{216}$	$\frac{3}{216}$	$\frac{1}{216}$

La loi des grands nombres affirme que plus le nombre n de lancers est grand plus les densités se rapprochent des probabilités ci-dessus. Ces probabilités sont indiquées sur les histogrammes ci-dessous par les points reliés.

Voici une vérification empirique de la loi des grands nombres.



Propriétés de l'espérance et de la variance

En reprenant l'exemple précédent. On peut considérer les variables aléatoires X_1 , X_2 et X_3 qui représentent le nombre montré par les dés numéro 1, 2 et 3 respectivement.

On note dans ce cas, la variable aléatoire $X = \sum_{i=1}^3 X_i = X_1 + X_2 + X_3$ qui représente la somme des nombres montrés par les trois dés.

Grâce aux probabilités qui se trouvent dans le tableau de l'exemple précédent, on a

$$\mathbf{P}(12 \leq X \leq 18) = \frac{25}{216} + \frac{21}{216} + \frac{15}{216} + \frac{10}{216} + \frac{6}{216} + \frac{3}{216} + \frac{1}{216} = \frac{81}{216} = 0.375$$

On utilise les propriétés suivantes (voir section 1.4).

$$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) \quad \text{et} \quad V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) \text{ si les } X_i \text{ sont indépendantes}$$

$$\text{Ainsi} \quad E(X) = E(X_1) + E(X_2) + E(X_3) = 3 \cdot \frac{7}{2} = 10.5 \implies \mu = 10.5$$

$$V(X) = V(X_1) + V(X_2) + V(X_3) = 3 \cdot \frac{35}{12} = 8.75 \implies \sigma \cong 2.96$$

1.1.6 La loi de Bernoulli

On considère une expérience aléatoire à deux issues : le succès et l'échec. À cette expérience, on associe la variable aléatoire X qui donne 1 en cas de succès et 0 en cas d'échec.

$$X : \{\text{succès, échec}\} \rightarrow \mathbb{R}; \quad \text{succès} \mapsto 1, \quad \text{échec} \mapsto 0$$

Si $p \in]0, 1[$ est la probabilité de succès, alors on a

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

On dit que la variable aléatoire X suit une loi de Bernoulli de paramètre p , notée $B(p)$.

Il est facile de montrer que $E(X) = p$ et $V(X) = p(1 - p)$.

Application

On tente de lancer une pièce de monnaie bien équilibrée sur pile. L'espérance vaut $\frac{1}{2}$ (on va réussir en moyenne une fois sur 2), la variance vaut $\frac{1}{4}$, l'écart type vaut $\frac{1}{2}$.

1.1.7 La loi binomiale

On considère une expérience aléatoire constituée de n épreuves successives identiques, indépendantes, à deux issues chacune : le succès et l'échec. Autrement dit, à la i -ième épreuve, on associe une variable aléatoire X_i qui suit une loi de Bernoulli $B(p)$ où $p \in]0, 1[$ est la probabilité d'un succès à chaque épreuve successive. À cette expérience, on associe la variable aléatoire X qui compte le nombre de succès, ainsi $X = \sum_{i=1}^n X_i$.

La probabilité d'avoir k succès (et donc $n - k$ échecs) est donnée grâce à la technique des anagrammes⁸ par la formule suivante (la i -ième case correspond au i -jet ; il y a n cases).

$$\begin{aligned} \mathbb{P}(X = k) &= \mathbb{P}\left(\underbrace{\begin{array}{|c|c|c|c|c|c|} \hline \text{succès} & \text{succès} & \dots & \text{succès} & \text{échec} & \dots & \text{échec} \\ \hline \end{array}}_{\substack{\text{permutations} \\ k \text{ succès} \quad n-k \text{ échecs}}}\right) \\ &= \binom{n}{k} p^k (1-p)^{n-k} \text{ où } \binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} & \text{si } k \in \{0, 1, 2, \dots, n\} \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

En fait⁹, $\binom{n}{k}$ est le coefficient de x^k dans le polynôme $(1+x)^n$.

On dit que la variable aléatoire X suit une binomiale de paramètres n et p , notée $B(n, p)$.

Remarque importante

Une variable qui suit une loi binomiale $B(n, p)$ est une somme de n variables qui suivent une loi de Bernoulli $B(p)$. Grâce aux propriétés de l'espérance et de la variance (voir section 1.4), on a $E(X) = np$ et $V(X) = np(1 - p)$.

Application

On compte le nombre de piles obtenus en lançant n fois une pièce de monnaie bien équilibrée. L'espérance vaut $\frac{n}{2}$ (le nombre de piles obtenus est en moyenne égal à la moitié des lancers), l'écart type vaut $\frac{n}{4}$.

8. Se référer à la technique des anagrammes du cours de probabilité (cours DF).

9. Par rapport au résultat de la page suivante, on a $(1+x)^n = (x^0 + x^1)^n$ car notre pièce de monnaie est comme un dé à deux faces dont les numéros sont 0 pour face et 1 pour pile.

1.1.8 La loi des jets de k dés

Cas particulier où $k = 5$

La variable aléatoire X compte la somme des résultats des jets de chacun des 5 dés.

On a évidemment $P(X = 5) = \left(\frac{1}{6}\right)^5 = \frac{1}{7776}$, puisque chacun des 5 dés doit tomber sur 1. De même, on a $P(X = 30) = \frac{1}{7776}$. Il est plus dur de calculer $P(X = k)$ avec $5 < m < 30$.

Regardons en détails comment calculer $P(X = 8)$ avec la méthode des anagrammes.

$$\begin{aligned}
 P(X = 8) &= \mathbf{P}\left(\overbrace{\boxed{1 \mid 1 \mid 2 \mid 2 \mid 2}}^{\text{perm.}}\right) + \mathbf{P}\left(\overbrace{\boxed{1 \mid 1 \mid 1 \mid 2 \mid 3}}^{\text{perm.}}\right) + \mathbf{P}\left(\overbrace{\boxed{1 \mid 1 \mid 1 \mid 1 \mid 4}}^{\text{perm.}}\right) \\
 &= \frac{5!}{2!3!} \cdot \left(\frac{1}{6}\right)^5 + \frac{5!}{3!1!1!} \cdot \left(\frac{1}{6}\right)^5 + \frac{5!}{4!1!} \cdot \left(\frac{1}{6}\right)^5 \\
 &= \left(\frac{1}{6}\right)^5 \left(\frac{5!}{2!3!} + \frac{5!}{3!1!1!} + \frac{5!}{4!1!}\right)
 \end{aligned}$$

On voit que ce qui n'est pas facile à trouver dans le cas général, c'est le contenu de la parenthèse qui contient les fractions avec des factorielles.

Mais en fait, ces fractions peuvent être expliquées¹⁰ ainsi.

$$\frac{5!}{2!3!} \quad \text{est le coefficient de } x_1^2 x_2^3 \text{ dans } (x_1 + x_2 + x_3 + x_4 + x_5 + x_6)^5$$

$$\frac{5!}{3!1!1!} \quad \text{est le coefficient de } x_1^3 x_2 x_3 \text{ dans } (x_1 + x_2 + x_3 + x_4 + x_5 + x_6)^5$$

$$\frac{5!}{4!1!} \quad \text{est le coefficient de } x_1^4 x_4 \text{ dans } (x_1 + x_2 + x_3 + x_4 + x_5 + x_6)^5$$

On utilise l'astuce suivante : en remplaçant x_1 par x , x_2 par x^2 , x_3 par x^3 , ..., x_6 par x^6 , on transforme les monômes à six variables ci-dessus en un polynôme à une seule variable. Ainsi, les trois monômes ci-dessus ($x_1^2 x_2^3$, $x_1^3 x_2 x_3$ et $x_1^4 x_4$) sont fusionnés sur un nouveau monôme à une seule variable dont la puissance correspond à la somme des dés, ici x^8 .

En effet, $x_1^2 x_2^3$ devient $x^2 x^6 = x^8$; $x_1^3 x_2 x_3$ devient $x^3 x^2 x^3 = x^8$; $x_1^4 x_4$ devient $x^4 x^4 = x^8$. Ainsi

$$\frac{5!}{2!3!} + \frac{5!}{3!1!1!} + \frac{5!}{4!1!} \quad \text{est le coefficient de } x^8 \text{ dans } (x + x^2 + x^3 + x^4 + x^5 + x^6)^5$$

Règle générale pour k dés

La probabilité de lancer k dés et d'obtenir un total de m est donnée par le coefficient x^m du polynôme $(x + x^2 + x^3 + x^4 + x^5 + x^6)^k$ multiplié par $\left(\frac{1}{6}\right)^k$. Ce coefficient est facilement déterminable à l'aide de Maxima : la commande qui établit la liste des coefficients est `makelist(coeff(ratsimp((x+x^2+x^3+x^4+x^5+x^6)^k), x, m), m, k, 6*k)`.

Par exemple, la probabilité de faire un total de 120 avec 30 dés vaut

$$2\,633\,742\,030\,711\,604\,837\,831 \cdot \frac{1}{221\,073\,919\,720\,733\,357\,899\,776} \cong 0.0119$$

10. On utilise les combinaisons vues dans le chapitre sur le dénombrement du cours DF

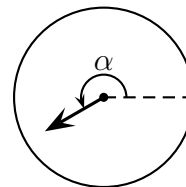
1.2 Variables aléatoires continues

Lorsque les valeurs prises par la variable aléatoire X varient dans un intervalle (ou une réunion d'intervalles), on dit que la variable aléatoire est *continue*.

1.2.1 La loi uniforme

Imaginons l'expérience aléatoire suivante : on fait tourner une aiguille fixée au milieu d'un disque. L'aiguille s'arrêtera de manière aléatoire.

Regardons la variable aléatoire X qui associe à chaque position de l'aiguille l'angle correspondant dans l'intervalle $[0, 2\pi[$.



La probabilité que l'angle soit exactement $\frac{\pi}{2}$ est nulle. On voit tout de même que la probabilité que l'angle soit entre $\frac{\pi}{4}$ et $\frac{3\pi}{4}$ devrait valoir $\frac{1}{4}$.

L'astuce consiste à se ramener à regarder des aires sous une courbe appelée *densité (continue) de probabilité* ou encore *distribution* et notée $f(x)$. On a

$$\underbrace{\mathbf{P}(a \leq X \leq b)}_{\text{Probabilité que } X \text{ prenne une valeur entre } a \text{ et } b} = \underbrace{\int_a^b f(x) dx}_{\text{Notation mathématique pour décrire l'aire entre } f \text{ et l'axe des } x, \text{ de } a \text{ jusqu'à } b}$$

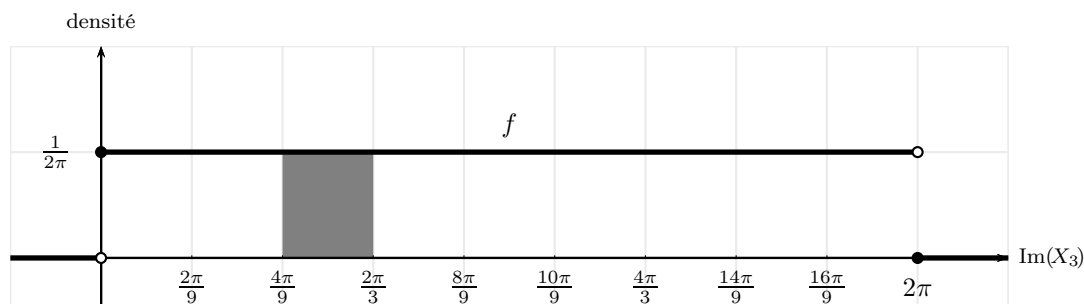
Comme une probabilité est un nombre entre 0 et 1 (ou entre 0% et 100%), on doit avoir

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \quad \text{et} \quad f(x) \geq 0 \text{ pour tout } x \in \mathbb{R}$$

La densité de la variable aléatoire X associée à cette expérience aléatoire est donnée par

$$f(x) = \begin{cases} \frac{1}{2\pi} & \text{si } x \in [0, 2\pi[\\ 0 & \text{si } x \notin [0, 2\pi[\end{cases}$$

Voici son graphe.



Cette densité est dite *uniforme*.

On a

$$\mathbf{P}\left(\frac{4\pi}{9} \leq X \leq \frac{2\pi}{3}\right) = \int_{\frac{4\pi}{9}}^{\frac{2\pi}{3}} f(x) dx = \frac{1}{9}$$

En effet, en observant le quadrillage, on constate que la zone grise définie sur l'intervalle $\left[\frac{4\pi}{9}, \frac{2\pi}{3}\right]$ a une proportion de $\frac{1}{9}$.

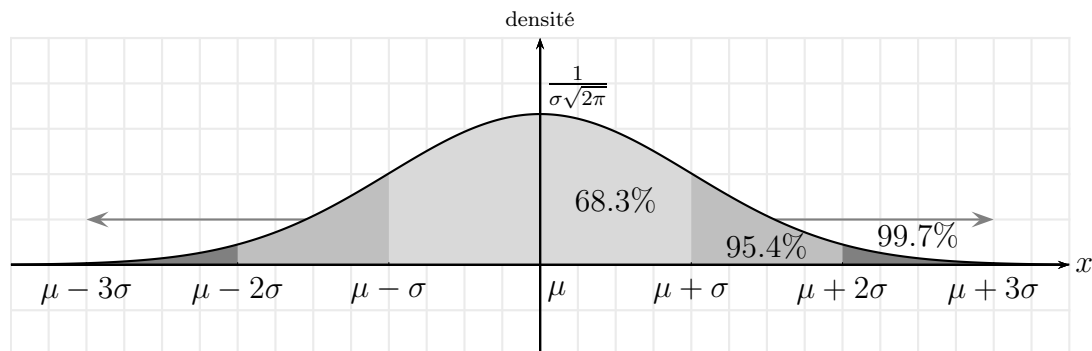
1.2.2 La loi normale

Il existe une loi, notée $\mathcal{N}(\mu, \sigma^2)$ et appelée *loi normale* (ou *loi de Laplace-Gauss*) qui, comme on le verra, permet d'approximer certaines probabilités.

Cette loi dépend de deux paramètres : l'espérance μ et la variance σ^2 (qui est le carré de l'écart type σ). Elle est donnée par la densité

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

où e est le nombre d'Euler qui vaut environ 2.71828. Voici le graphe de f .

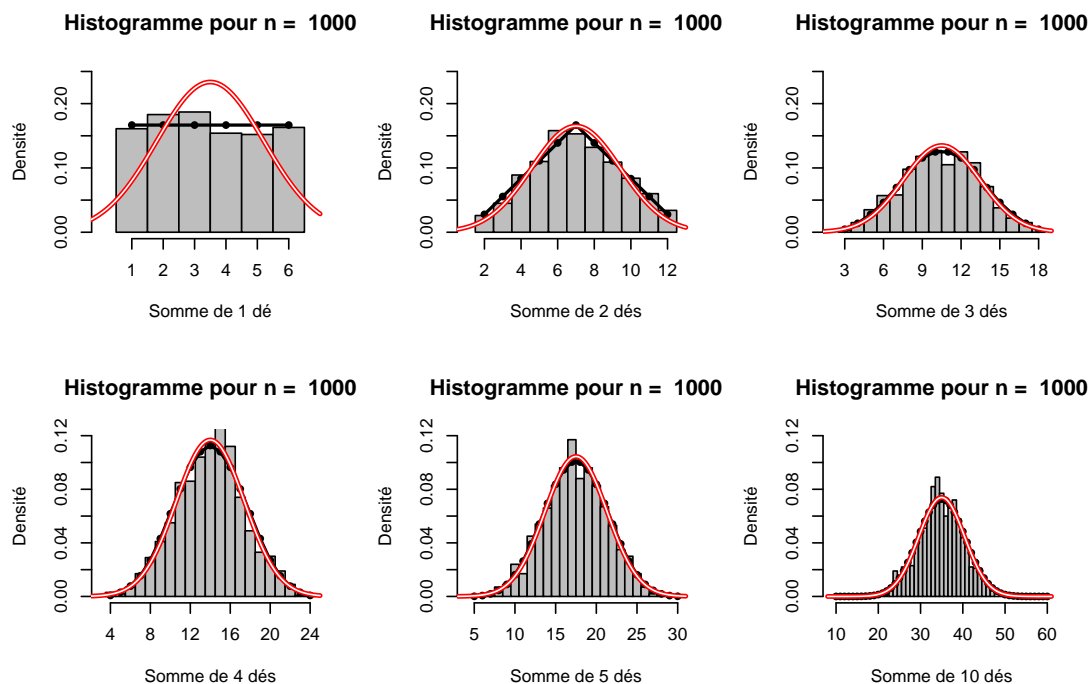


Le lecteur pourra comparer les pourcentages indiqués avec ceux donnés par le théorème de Tchebychev.

Même si cette loi est très importante, tous les phénomènes ne peuvent pas forcément être estimés par la loi normale.

Approximation de la somme d'un jet de dés

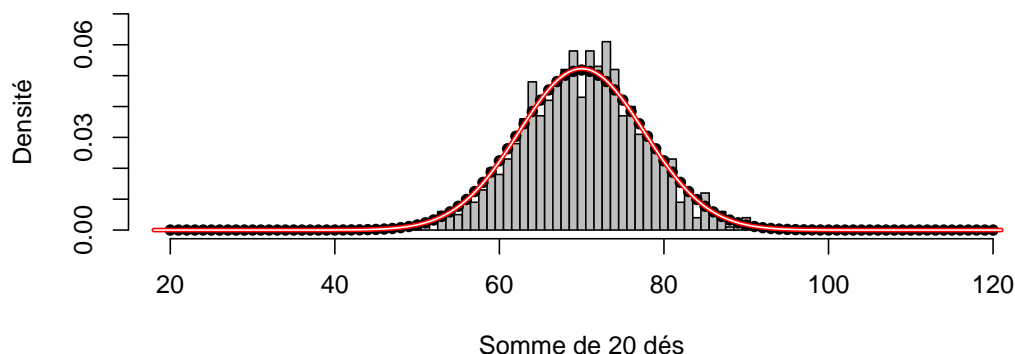
Superposons en rouge/blanc la loi normale $\mathcal{N}(k\mu, k\sigma^2)$ correspondant aux lancers de k dés pour $k \in \{1, 2, 3, 4, 5, 10\}$. Rappelons que $\mu = 3.5$ et $\sigma = \frac{35}{12}$.



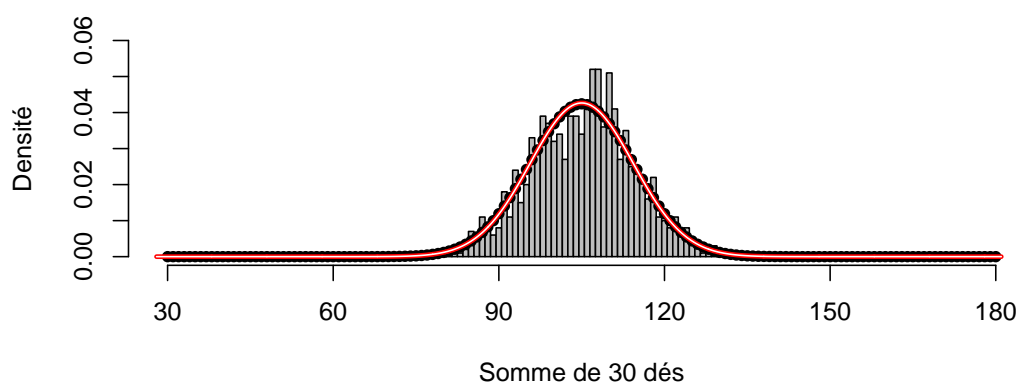
On voit que lorsque n grandit, l'approximation par la loi normale devient meilleure.

Encore quelques comparaisons.

Histogramme pour n = 1000



Histogramme pour n = 1000



Comparons la probabilité exacte d'avoir, en lançant 30 dés, une somme comprise entre 120 et 180 avec son approximation par la loi normale $\mathcal{N}(105, 87.5)$.

méthodes de calcul	réponse
vraies probabilités (voir section 1.1.8)	$\mathbb{P}(120 \leq X \leq 180) \cong 0.06067$
loi normale $\mathcal{N}(105, 87.5)$	$\int_{120}^{180} f(x) dx \cong 0.05440$
loi normale $\mathcal{N}(105, 87.5)$ avec correction de continuité (voir aussi la page 28)	$\int_{119.5}^{180.5} f(x) dx \cong 0.06056$

La correction de continuité consiste à enlever $\frac{1}{2}$ à la première borne d'intégration et à ajouter $\frac{1}{2}$ à la deuxième borne. En effet, si on veut calculer la probabilité $\mathbb{P}(X = 100)$ et qu'on calcule $\int_{100}^{100} f(x) dx$, on va trouver 0, parce qu'il n'y a pas d'aire. Si on veut approximer $\mathbb{P}(X = 100)$ à l'aide de la loi normale, il faut calculer l'intégrale $\int_{99.5}^{100.5} f(x) dx$, car les barres des histogrammes sont centrées en 100 mais sont de largeur 1.

1.2.3 La fonction gamma

La fonction *gamma* est définie par l'intégrale suivante.

$$\Gamma(n) = \int_0^{+\infty} x^{n-1} e^{-x} dx \quad n \in]0, +\infty[$$

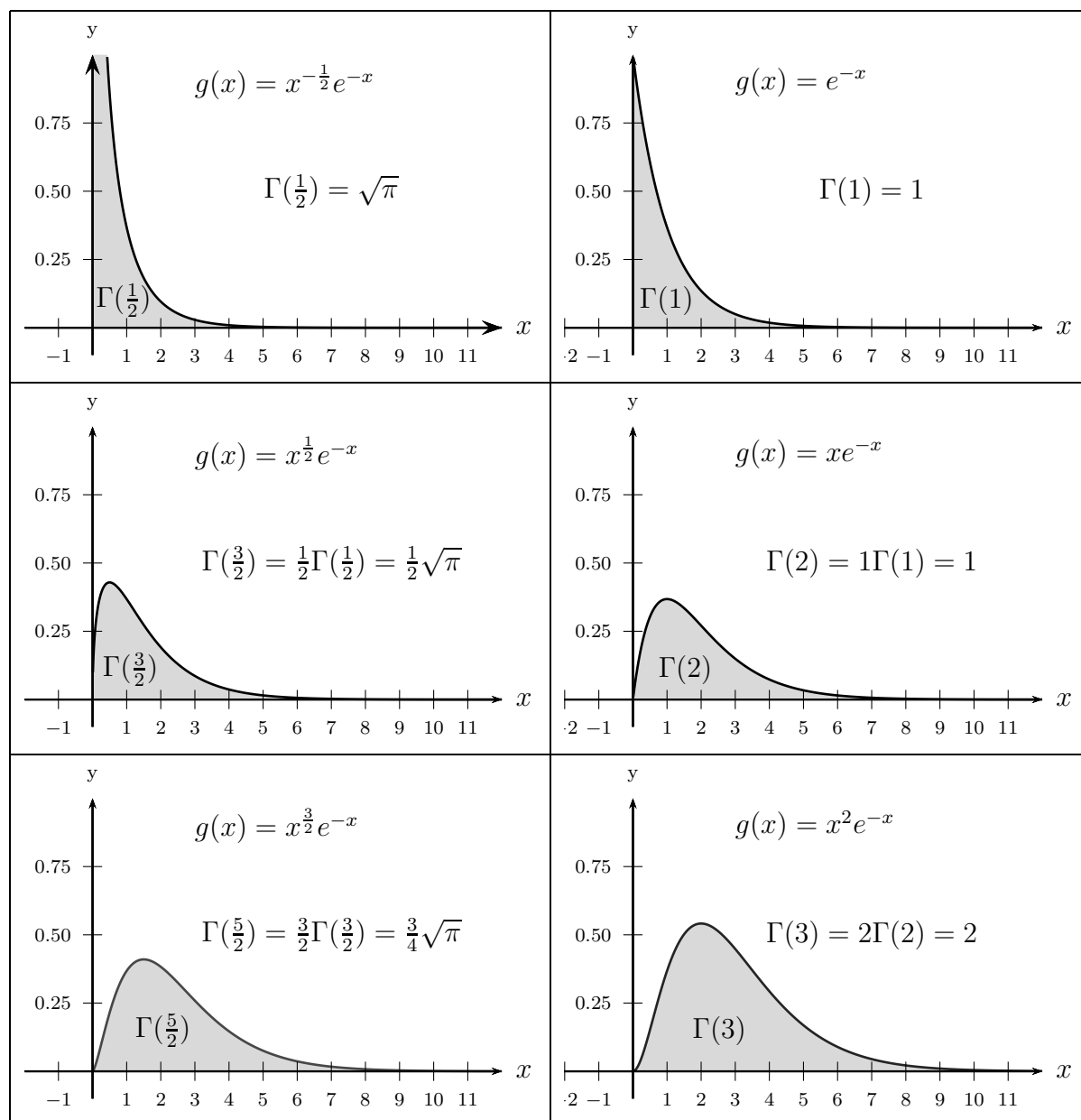
Propriétés de la fonction gamma.

$$\Gamma(n+1) = n \cdot \Gamma(n) \text{ pour tout } n > 0 \quad \text{avec} \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \quad \text{et} \quad \Gamma(1) = 1$$

Ces propriétés permettent de calculer $\Gamma\left(\frac{n}{2}\right)$ avec $n \geq 1$.

On voit aussi que lorsque n est un nombre naturel, $n \geq 1$, la fonction *gamma* est une généralisation de la factorielle, autrement dit qu'on a $\Gamma(n) = (n-1)!$.

En effet, la factorielle satisfait les propriétés $n! = n \cdot (n-1)!$ et $0! = 1$.



1.2.4 Les lois de Student

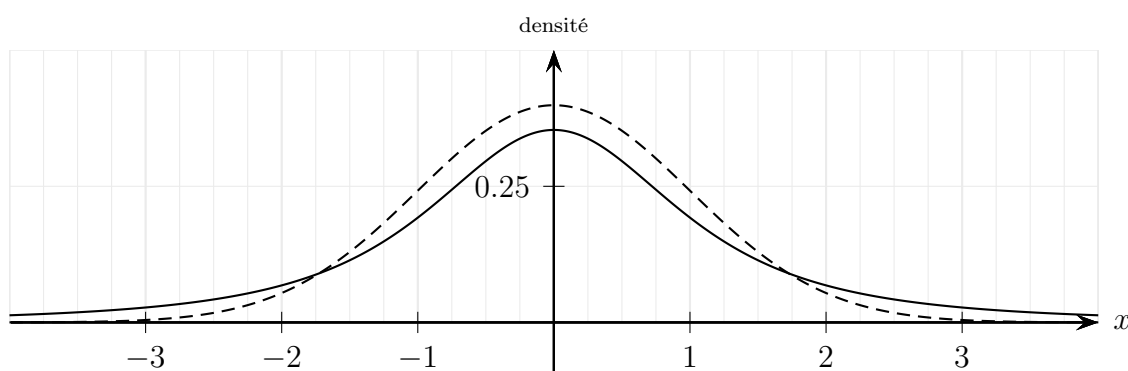
La densité de probabilité de la loi de Student T_ν à ν degrés de liberté est donnée par la fonction suivante.

$$f_{T_\nu}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{1}{\sqrt{\nu\pi}} \cdot \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}$$

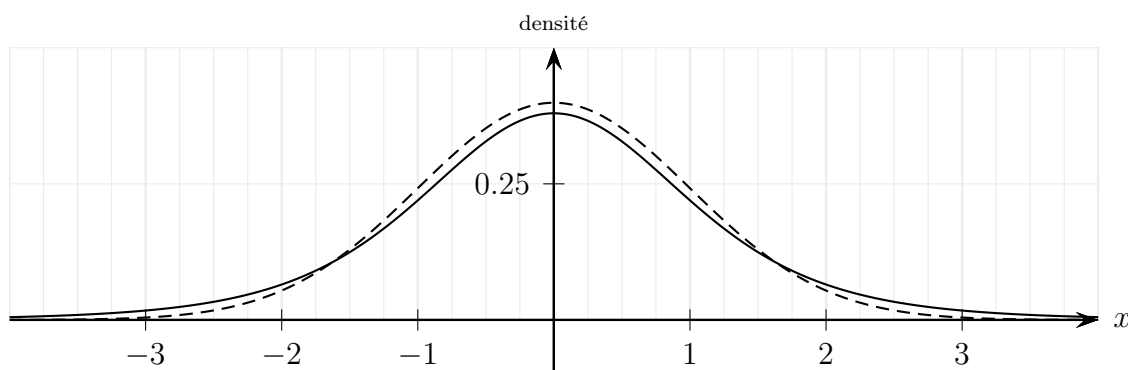
Il s'agit d'une distribution qui ressemble à celle de la loi normale centrée réduite, mais elle est plus haute sur les côtés.

Sur les graphiques suivants, on voit la loi normale centrée réduite $\mathcal{N}(0,1)$ en traitillés, les lois de Student pour $\nu = 2$, $\nu = 5$ et $\nu = 30$.

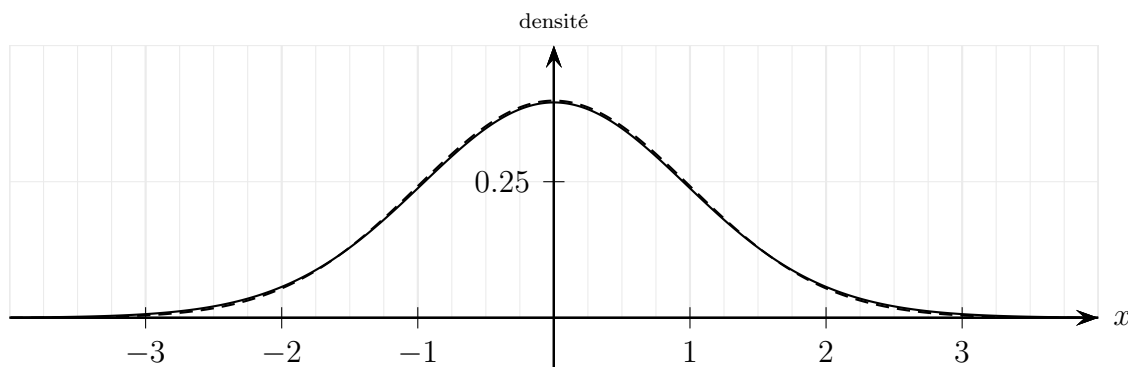
Comparaison en la loi normale $\mathcal{N}(0,1)$ et T_2 .



Comparaison en la loi normale $\mathcal{N}(0,1)$ et T_5 .

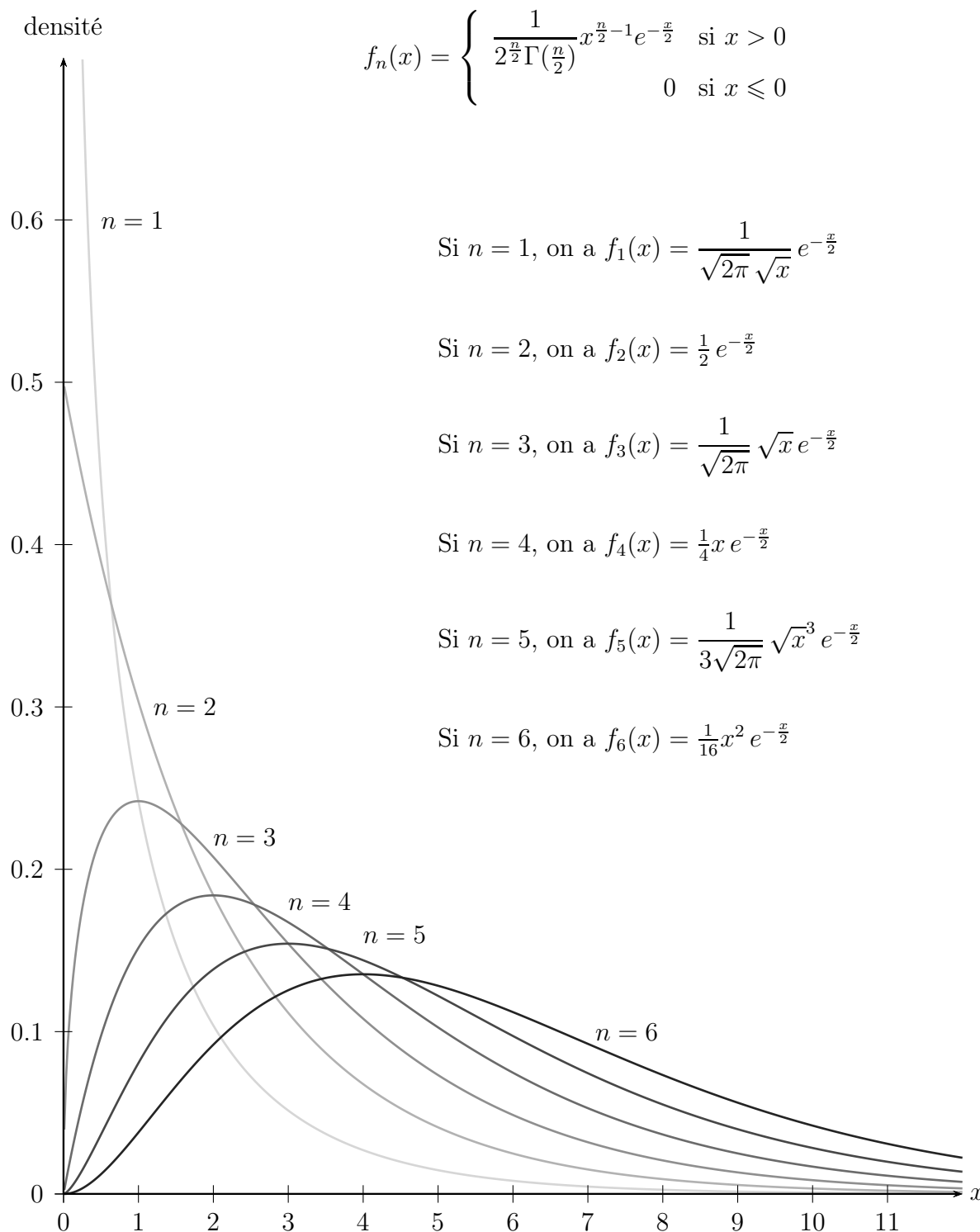


Comparaison en la loi normale $\mathcal{N}(0,1)$ et T_{30} .



1.2.5 Les lois du chi-carré

Il existe des lois, notées χ_n^2 , et appelées lois du chi-carré (prononcer «ki») avec n degrés de liberté. Voici les expressions fonctionnelles de leur densité et leur représentation graphique.



1.2.6 Espérance et variance dans le cas continu

Espérance d'une variable aléatoire continue

L'*espérance* d'une variable aléatoire X , notée $E(X)$ ou μ , est la moyenne pondérée des valeurs images de X par leur probabilité :

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx \quad (\text{en modifiant la formule du cas discret})$$

Variance et écart type d'une variable aléatoire continue

La *variance* d'une variable aléatoire X , notée $V(X)$, est la moyenne pondérée des carrés des écarts par rapport à l'espérance μ par leur probabilité.

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

L'*écart type* de X est défini par $\sigma(X) = \sqrt{V(X)}$.

Généralisation de l'espérance

Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction réelle et X une variable aléatoire, alors $g \circ X$ est une variable aléatoire notée¹⁰ $g(X)$ dont l'espérance est :

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

Cela permet notamment d'écrire la variance comme une espérance.

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

Exemple de calcul

Pour la variable aléatoire X définie en page 15, l'espérance est le résultat que l'on peut le plus souvent espérer (en moyenne) et vaut :

$$E(X) = \int_0^{2\pi} \frac{x}{2\pi} dx \stackrel{\text{TFCI}}{=} \frac{x^2}{4\pi} \Big|_0^{2\pi} = \frac{4\pi^2}{4\pi} = \pi$$

La variance est :

$$V(X) = \int_0^{2\pi} \frac{(x - \pi)^2}{2\pi} dx \stackrel{\text{TFCI}}{=} \frac{(x - \pi)^3}{6\pi} \Big|_0^{2\pi} = \frac{\pi^3}{6\pi} - \frac{(-\pi)^3}{6\pi} = \frac{\pi^2}{6} + \frac{\pi^2}{6} = \frac{\pi^2}{3}$$

Par conséquent, l'écart type vaut $\sigma(X) = \frac{\pi}{\sqrt{3}}$.

10. Cette notation est abusive : $g(X)(\omega)$ n'a pas de sens ! La justification est celle-ci : si $X : \Omega \rightarrow \mathbb{R}$ et $g : \mathbb{R} \rightarrow \mathbb{R}$, alors $(g \circ X)(\omega) = g(X(\omega)) = g(x)$ avec $x = X(\omega) \in \text{Im}(X)$. D'où la notation $g \circ X = g(X)$.

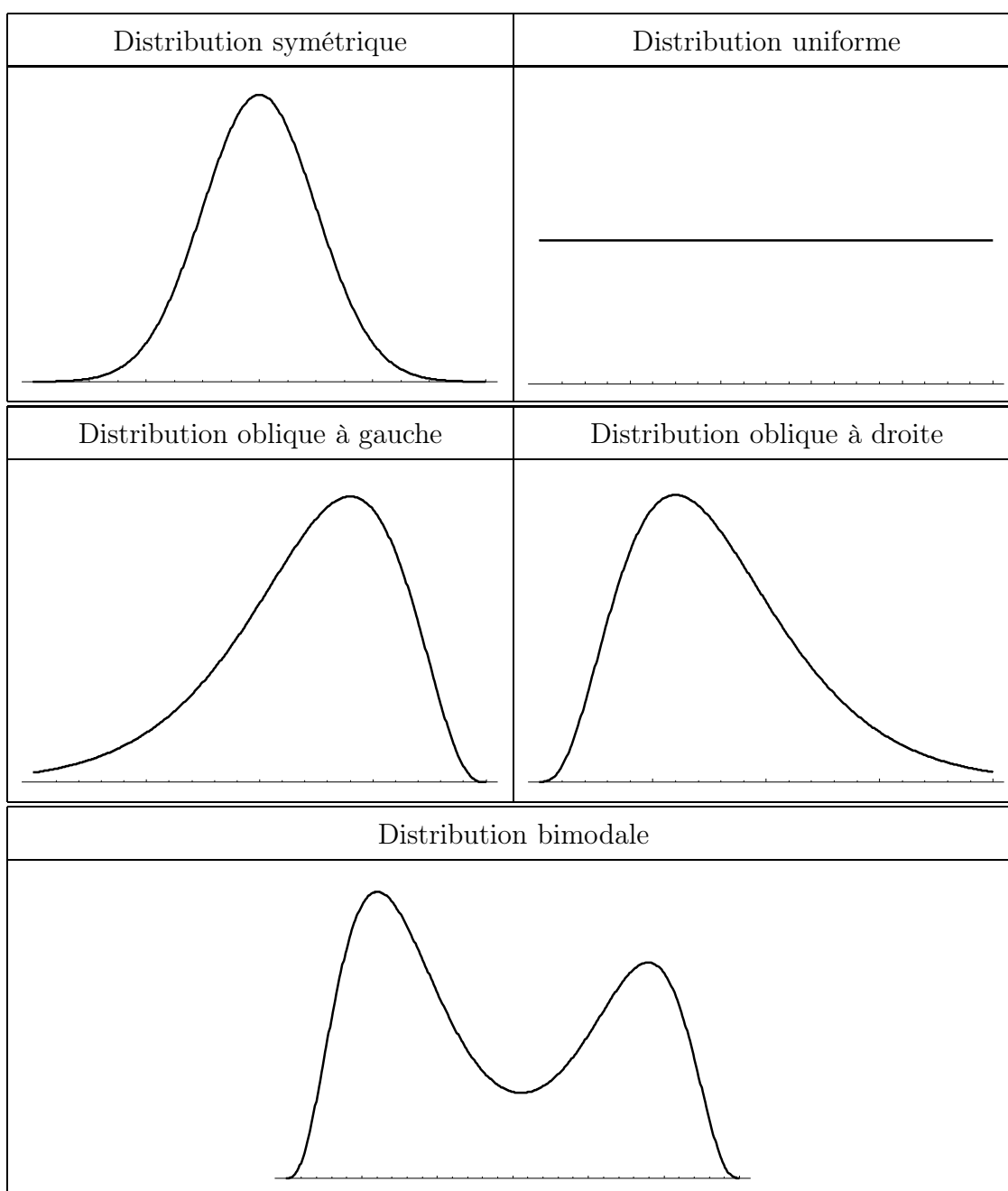
1.3 Les différentes formes de densité de distribution

Il s'agit d'une notion issue de la théorie des probabilités : une *densité* (ou *distribution*) est une description de la courbe que *devrait* suivre un histogramme lorsque les mesures utilisées pour faire l'histogramme sont en très grands nombre (grâce à la loi des grands nombres). Il s'agit d'une *courbe théorique* !

Tandis que l'histogramme est basé sur les observations, la distribution est théorique. Il faut donc faire preuve de prudence lorsqu'on dit que les données mesurées suivent une certaine distribution !

Dans les exemples précédents, les densités ont toujours été déterminées de manière exacte (loi de Bernoulli, loi binomiale ou loi multinomiale) ou approximative (loi normale, lois de Student, lois du chi-carré).

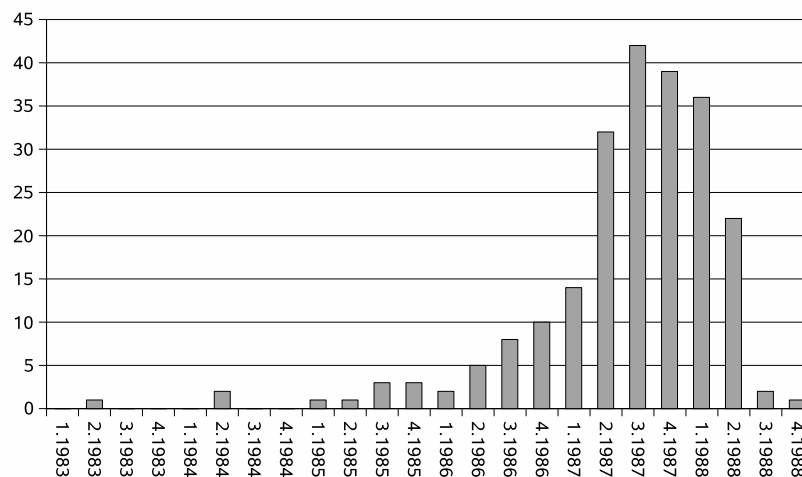
Voici les noms de certaines formes de distributions que l'on peut trouver dans la nature.



Date de naissance de élèves de première année (2003-2004)

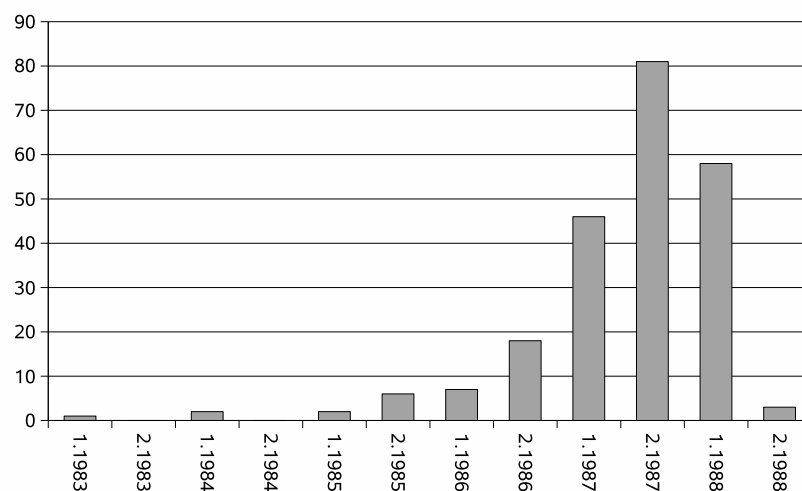
On peut regarder la date de naissance des élèves de première année. Il y a 224 élèves inscrits en août 2003. Voici l'histogramme où les classes sont des trimestres (trois mois).

Histogramme



On voit beaucoup mieux la nécessité de classer les données dans des classes, afin d'avoir un graphe lisible. Voici l'histogramme où les classes sont des semestres (six mois).

Histogramme



Il faut faire attention à l'interprétation que l'on donne à l'histogramme, puisque le choix des classes influence légèrement l'histogramme. Pour cette raison, il est important de ne pas faire de calculs à partir des nombres indiqués sur l'histogramme, mais plutôt à partir des données initiales.

Même si le choix des classes modifie l'histogramme, on s'aperçoit que la forme générale reste la même. Ici, on voit qu'en première année, il y a beaucoup d'élèves qui sont nés en 1987 et dans la première moitié de 1988.

La distribution est oblique à gauche, car il y a aura certes quelques élèves surdoués qui auront sauté une année (donc qui entreront au lycée très jeune), mais il y a aura bien plus d'élèves qui entreront au lycée un peu plus vieux (redoublement, réorientation professionnelle, ...).

1.4 Propriétés de l'espérance et de la variance

Propriétés de l'espérance

L'espérance est linéaire (cela provient des propriétés du symbole somme et de l'intégrale). Cela signifie que l'espérance satisfait :

$$\begin{aligned} \text{additivité :} & \quad E(X + Y) = E(X) + E(Y) \\ \text{homogénéité :} & \quad E(\lambda X) = \lambda E(X) \end{aligned}$$

De plus, si C est une variable aléatoire constante, on a $E(C) = C$.

Propriétés de la variance

La variance n'est pas linéaire. On a tout de même :

$$\begin{aligned} \text{additivité si } X \text{ et } Y \text{ sont indépendants :} & \quad V(X + Y) = V(X) + V(Y) \\ \text{homogénéité de degré 2 :} & \quad V(\lambda X) = \lambda^2 V(X) \\ \text{lorsque } C \text{ est une variable aléatoire constante :} & \quad V(C) = 0 \end{aligned}$$

Pour démontrer la formule de l'additivité, il faut définir la notion de variables aléatoires indépendantes (et aussi la notion de covariance) : cela dépasse le cadre de ce cours. Par contre, les deux autres propriétés sont évidentes.

Autre formule pour la variance

En utilisant les propriétés de l'espérance, on montre que $V(X) = E(X^2) - (E(X))^2$. En effet, en notant $\mu = E(X)$, on a :

$$V(X) = E\left((X - \mu)^2\right) = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - (E(X))^2$$

Théorème

On considère n variables aléatoires indépendantes X_1, X_2, \dots, X_n de même espérance μ et de même variance σ^2 . La moyenne de ces n variables aléatoires est aussi une variable aléatoire, notée \bar{X} et définie par $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$. Sous ces hypothèses, on a :

$$E(\bar{X}) = \mu \quad V(\bar{X}) = \frac{\sigma^2}{n}$$

Preuve

$$\text{On a :} \quad E(\bar{X}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) \stackrel{E \text{ lin.}}{=} \frac{\sum_{i=1}^n E(X_i)}{n} = \frac{\sum_{i=1}^n \mu}{n} = \frac{n\mu}{n} = \mu$$

Pour la variance :

$$V(\bar{X}) = V\left(\frac{\sum_{i=1}^n X_i}{n}\right) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right)$$

Point délicat : on a $V(\sum_{i=1}^n X_i) = \sum_{i=1}^n V(X_i)$ parce que les variables aléatoires X_i sont indépendantes. Par conséquent :

$$V(\bar{X}) = \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \quad \square$$

1.5 Centrage et réduction

Définitions

- Une variable aléatoire (discrète ou continue) est dite *centrée* si son espérance est nulle.
- Une variable aléatoire (discrète ou continue) est dite *réduite* si sa variance vaut 1.

Théorème

Si X est une variable aléatoire (discrète ou continue) d'espérance μ et de variance σ^2 , alors

$$\frac{X - \mu}{\sigma} \text{ est une variable centrée réduite}$$

Preuve

1. On a, par linéarité :

$$E\left(\frac{X - \mu}{\sigma}\right) = E\left(\frac{1}{\sigma}(X - \mu)\right) \stackrel{E \text{ lin.}}{=} \frac{1}{\sigma} \cdot E(X - \mu) \stackrel{E \text{ lin.}}{=} \frac{1}{\sigma} \cdot \overbrace{(E(X) - \mu)}^{=0} = 0$$

2. On commence par utiliser la formule $V(X) = E((X - E(X))^2)$ afin d'avoir :

$$V\left(\frac{X - \mu}{\sigma}\right) = E\left(\left(\frac{X - \mu}{\sigma} - \underbrace{E\left(\frac{X - \mu}{\sigma}\right)}_{=0 \text{ (voir ci-dessus)}}\right)^2\right)$$

On peut ainsi continuer le calcul :

$$\begin{aligned} V\left(\frac{X - \mu}{\sigma}\right) &= E\left(\left(\frac{X - \mu}{\sigma}\right)^2\right) = E\left(\frac{(X - \mu)^2}{\sigma^2}\right) = E\left(\frac{1}{\sigma^2} \cdot (X - \mu)^2\right) \\ &= \frac{1}{\sigma^2} \cdot E((X - \mu)^2) = \frac{1}{\sigma^2} \cdot E(X^2 - 2X\mu + \mu^2) \\ &\stackrel{E \text{ lin.}}{=} \frac{1}{\sigma^2} \cdot (E(X^2) - 2\mu E(X) + \mu^2) \stackrel{\mu = E(X)}{=} \frac{1}{\sigma^2} \cdot (E(X^2) - (E(X))^2) \end{aligned}$$

On termine en se souvenant de la formule $V(X) = E(X^2) - (E(X))^2$. En effet :

$$V\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2} \cdot (E(X^2) - (E(X))^2) = \frac{1}{\sigma^2} \cdot V(X) = \frac{1}{\sigma^2} \cdot \sigma^2 = 1$$

□

1.6 Fonction de répartition d'une variable aléatoire

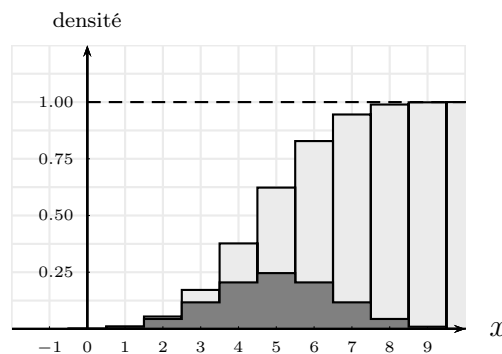
1.6.1 Fonction de répartition d'une variable discrète

On définit la fonction de répartition d'une variable discrète comme suit.

$$\phi(x) = \mathbf{P}(X \leq x) = \sum_{k=-\infty}^x \mathbf{P}(X = k) \quad \text{où } \mathbf{P}(X = k) \text{ est la densité discrète de probabilité}$$

Exemple

Ci-dessous, on trouve la loi binomiale $B(10, \frac{1}{2})$ (dont la densité est en gris foncé) et sa fonction de répartition (en gris clair).



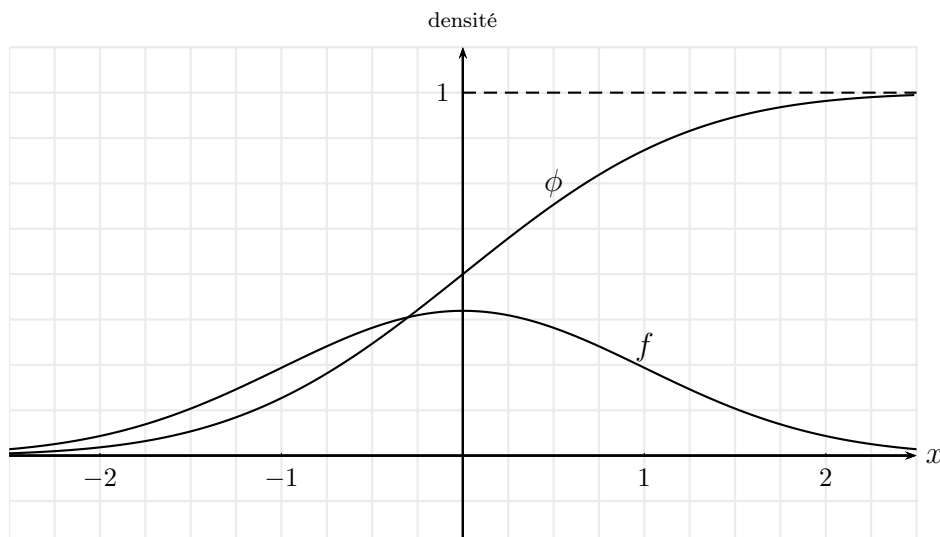
1.6.2 Fonction de répartition d'une variable continue

On définit la fonction de répartition d'une variable continue comme suit.

$$\phi(x) = \mathbf{P}(X \leq x) = \int_{-\infty}^x f(t) dt \quad \text{où } f(x) \text{ est la densité de probabilité}$$

Exemple

Ci-dessous, on trouve la loi normale $\mathcal{N}(0, 1)$ et sa fonction de répartition.



1.7 Deux propriétés essentielles de la loi normale

1.7.1 La loi normale permet d'approximer la loi binomiale

Le théorème central-limite de Laplace

Soit $a, b \in \mathbb{R}$ et $p \in]0, 1[$. Alors

La loi binomiale $B(n, p)$ suit approximativement une loi normale $\mathcal{N}(np, np(1-p))$

Autrement dit, pour n assez grand, on peut approximer la probabilité que la variable X , qui suit une loi binomiale $B(n, p)$ d'espérance np et de variance $np(1-p)$, soit entre a et b de la manière suivante.

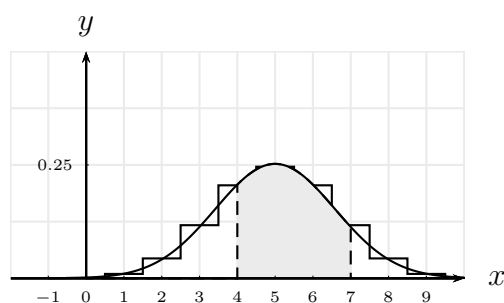
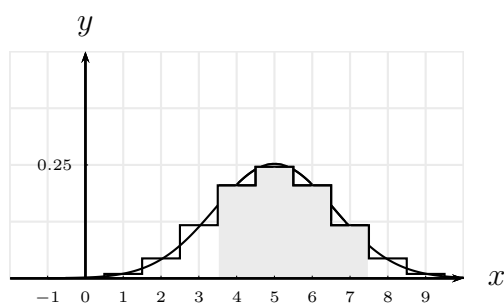
$$\begin{array}{l}
 \mathbf{P}(a \leq X \leq b) \stackrel{\substack{\text{approx par} \\ \cong \\ N \sim \mathcal{N}(np, np(1-p))}}{=} \mathbf{P}(a \leq N \leq b) \\
 \stackrel{\substack{\text{centrage} \\ = \\ \text{réduction}}}{=} \mathbf{P}\left(\frac{a - np}{\sqrt{np(1-p)}} \leq \frac{N - np}{\sqrt{np(1-p)}} \leq \frac{b - np}{\sqrt{np(1-p)}}\right) \\
 \stackrel{\substack{\text{fonction de} \\ \text{répartition}}}{=} \mathbf{P}\left(Z \leq \frac{b - np}{\sqrt{np(1-p)}}\right) - \mathbf{P}\left(Z \leq \frac{a - np}{\sqrt{np(1-p)}}\right) \\
 \stackrel{\substack{\text{notation} \\ = \\ \text{page 30}}}{=} \phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right)
 \end{array}$$

La plupart du temps, on considère que si $np \geq 5$ et si $n(1-p) \geq 5$ (autrement dit si n est suffisamment grand par rapport aux probabilités de succès p et d'échec $1-p$), alors l'approximation est de bonne qualité.

Exemple

Soit X une variable aléatoire qui suit la loi binomiale $B(10, \frac{1}{2})$ (on lance une pièce de monnaie bien équilibrée 10 fois de suite et on compte le nombre de piles obtenu). On a $np = n(1-p) = 5$. Ainsi, on estime $B(10, \frac{1}{2})$ avec la loi normale $\mathcal{N}(np, np(1-p)) = \mathcal{N}(5, 2.5)$.

En gris, on a $\mathbf{P}(4 \leq X \leq 7)$ que l'on voit sur le graphe de gauche et que l'on estime par la loi normale sur le graphe de droite.



On a :

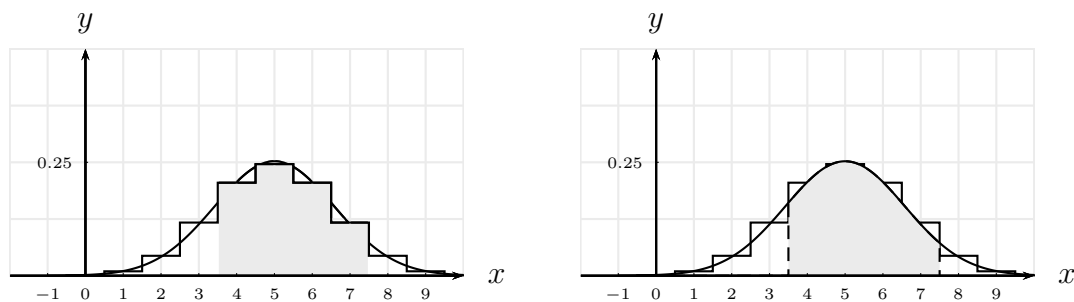
$$\begin{aligned}
 \mathbf{P}(4 \leq X \leq 7) &= \mathbf{P}(X = 4) + \mathbf{P}(X = 5) + \mathbf{P}(X = 6) + \mathbf{P}(X = 7) \\
 &= \binom{10}{4} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^6 + \binom{10}{5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^5 + \binom{10}{6} \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^4 + \binom{10}{7} \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^3 \\
 &= \frac{792}{1024} = \frac{99}{128} = 0.7734375
 \end{aligned}$$

L'approximation est donnée par

$$\mathbf{P}(4 \leq X \leq 7) \cong \phi\left(\frac{7-5}{\sqrt{2.5}}\right) - \phi\left(\frac{4-5}{\sqrt{2.5}}\right) \cong \phi(1.26) - \phi(-0.63) \cong 0.6319$$

Amélioration de l'estimation par correction de continuité

L'estimation de la page précédente n'est pas extraordinaire. Pour l'améliorer, on va légèrement modifier les bornes de l'intégrale. En effet, en allant de 4 à 7, l'intégrale précédente 'oublie' la moitié des deux rectangles aux extrémités. C'est pourquoi il est plus judicieux de partir à 3.5 et de finir à 7.5 comme le montre les schémas ci-dessous.



On a :

$$\mathbf{P}(4 \leq X \leq 7) = 0.7734375$$

L'approximation est maintenant bien meilleure

$$\mathbf{P}(4 \leq X \leq 7) \cong \phi\left(\frac{7.5-5}{\sqrt{2.5}}\right) - \phi\left(\frac{3.5-5}{\sqrt{2.5}}\right) \cong \phi(1.58) - \phi(-0.95) \cong 0.7718$$

Formule d'approximation améliorée

La formule d'approximation améliorée est donc dans le cas général donnée par

$$\mathbf{P}\left(a \leq X \leq b\right) \cong \phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Remarque

Si n est très grand, les deux approximations (l'améliorée et celle du théorème) seront les deux très proches.

1.7.2 Le théorème de la limite centrale

Le *théorème de la limite centrale*, aussi appelé *théorème central limite* est l'un des résultats les plus importants de la théorie des probabilités. De façon informelle, ce théorème¹¹ donne une estimation de l'erreur que l'on commet en approchant la moyenne théorique μ par la moyenne arithmétique \bar{x} .

Théorème de la limite centrale

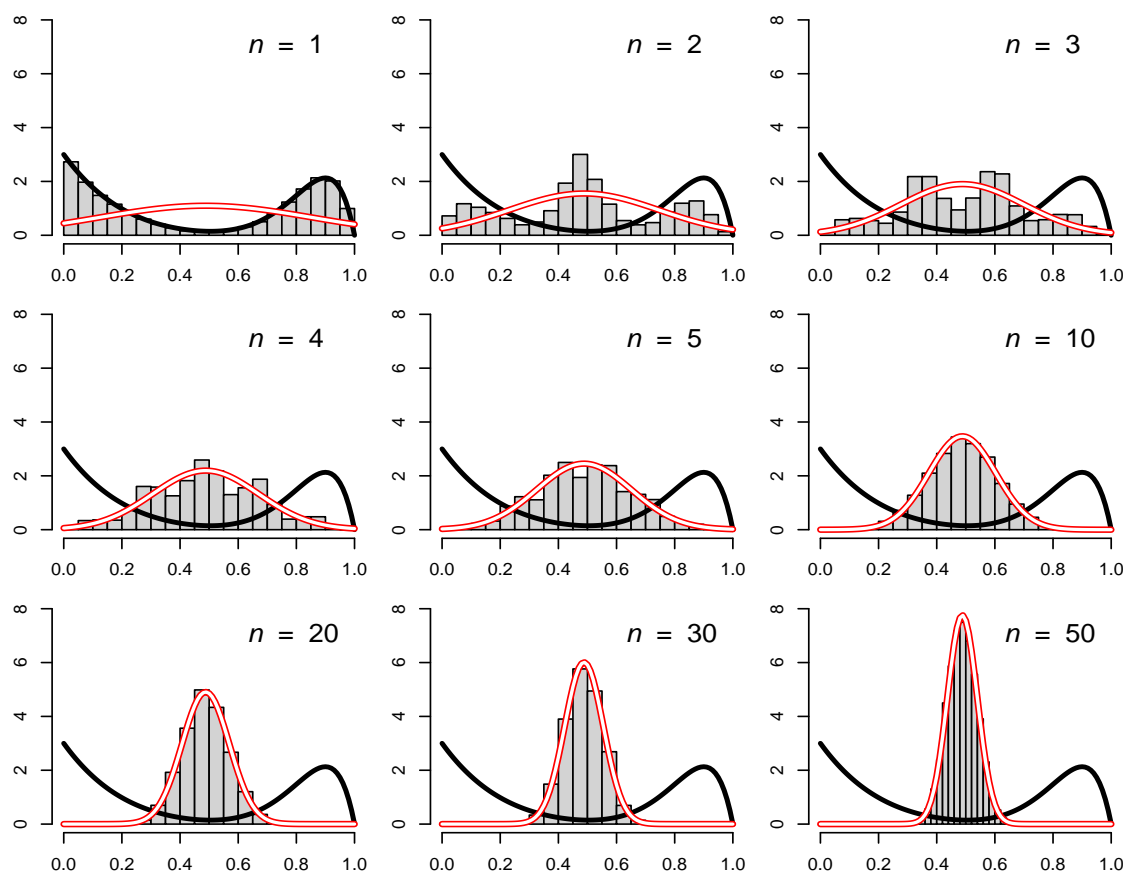
Supposons qu'on a n variables aléatoires X_1, \dots, X_n indépendantes et qui suivent une même loi de probabilité d'espérance μ et de variance σ^2 .

Le théorème central limite dit que, si n est suffisamment grand (la plupart du temps $n \geq 30$ suffit), alors la variable aléatoire $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ qui correspond à la moyenne des variables X_1, \dots, X_n suit approximativement une loi normale $\mathcal{N}(\mu, \frac{\sigma^2}{n})$.

Simulation numérique

En noir, la loi d'origine. En rouge/blanc, la loi normale du théorème. En gris, l'histogramme des moyennes de 10 000 échantillons de taille n .

Pour $n = 1$, l'histogramme suit la loi d'origine, puisque la moyenne d'une mesure est égale à la mesure elle-même. Puis plus n grandit, plus l'histogramme cesse de suivre la loi d'origine pour se rapprocher de la loi normale (qui devient de plus en plus étroite, car $\frac{\sigma^2}{n}$ devient de plus en plus petit, et haute, car son aire vaut toujours 1).



11. Ce phénomène a d'abord été observé par Gauss qui l'appelait loi des erreurs ; mais ce dernier n'en a pas donné de démonstration rigoureuse. La preuve du théorème a été apportée par deMoivre et Laplace ; le théorème porte donc parfois leurs noms.

1.8 Tables

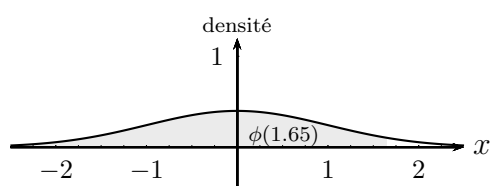
1.8.1 Fonction de répartition de la loi normale

La fonction de répartition $\phi(x)$ de la loi normale centrée réduite $\mathcal{N}(0, 1)$ est une intégrale qui n'admet pas de primitive explicite. On recourt ainsi à une table pour donner des valeurs approximatives.

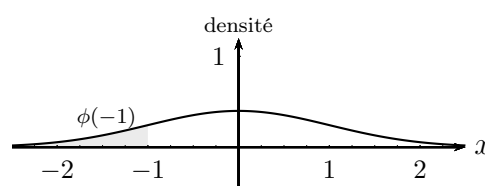
$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	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	0.7703	0.7734	0.7764	0.7793	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	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	0.8906	0.8925	0.8943	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	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	0.9986	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

On lit les décimales dans les lignes, et les centièmes en colonnes. Par exemple, on trouve l'approximation arrondie à 4 décimales de $\phi(1.65)$, qui vaut 0.9505, à l'intersection de la ligne 1.6 et de la colonne 0.05. Si x est négatif, on utilise la relation $\phi(-x) = 1 - \phi(x)$ qui provient du fait que la densité est paire et que l'aire sous la courbe vaut 1.



$$\phi(1.65) \cong 0.9505$$



$$\phi(-1) = 1 - \phi(1) \cong 1 - 0.8413 = 0.1587$$

1.8.2 Les quantiles des lois de Student

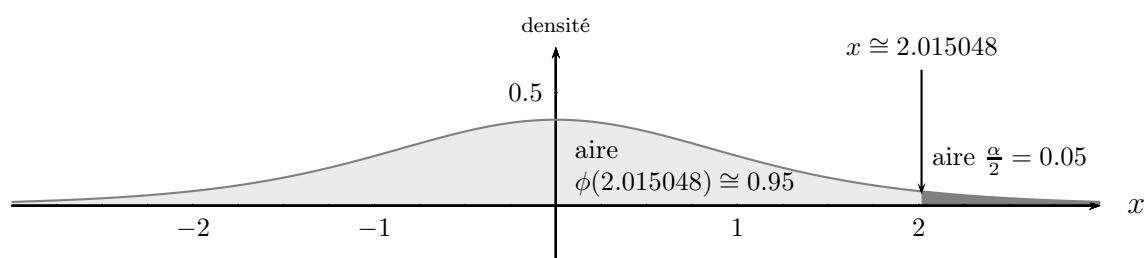
La fonction de répartition ϕ_ν de la loi de Student T_ν à ν degrés de liberté est une intégrale qui n'admet pas de primitive explicite (sauf si $\nu = 1$).

$$\phi_\nu(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \cdot \frac{1}{\sqrt{\nu\pi}} \cdot \int_{-\infty}^x \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt, \quad x \in \mathbb{R}$$

Contrairement à la table de la loi normale, il est coutume de préférer une table des quantiles (c'est-à-dire une table qui donne la valeur de x lorsqu'on connaît la valeur de l'aire $\phi_\nu(x)$).

$\nu \backslash \phi_\nu(x)$	0.60	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.302653	6.964557	9.924843	31.59906
3	0.276671	0.764892	1.637744	2.353363	3.182446	4.540703	5.840909	12.92398
4	0.270722	0.740697	1.533206	2.131847	2.776445	3.746947	4.604095	8.610302
5	0.267181	0.726687	1.475884	2.015048	2.570582	3.364930	4.032143	6.868827
6	0.264835	0.717558	1.439756	1.943180	2.446912	3.142668	3.707428	5.958816
7	0.263167	0.711142	1.414924	1.894579	2.364624	2.997952	3.499483	5.407883
8	0.261921	0.706387	1.396815	1.859548	2.306004	2.896459	3.355387	5.041305
9	0.260955	0.702722	1.383029	1.833113	2.262157	2.821438	3.249836	4.780913
10	0.260185	0.699812	1.372184	1.812461	2.228139	2.763769	3.169273	4.586894
11	0.259556	0.697445	1.363430	1.795885	2.200985	2.718079	3.105807	4.436979
12	0.259033	0.695483	1.356217	1.782288	2.178813	2.680998	3.054540	4.317791
13	0.258591	0.693829	1.350171	1.770933	2.160369	2.650309	3.012276	4.220832
14	0.258213	0.692417	1.345030	1.761310	2.144787	2.624494	2.976843	4.140454
15	0.257885	0.691197	1.340606	1.753050	2.131450	2.602480	2.946713	4.072765
16	0.257599	0.690132	1.336757	1.745884	2.119905	2.583487	2.920782	4.014996
17	0.257347	0.689195	1.333379	1.739607	2.109816	2.566934	2.898231	3.965126
18	0.257123	0.688364	1.330391	1.734064	2.100922	2.552380	2.878440	3.921646
19	0.256923	0.687621	1.327728	1.729133	2.093024	2.539483	2.860935	3.883406
20	0.256743	0.686954	1.325341	1.724718	2.085963	2.527977	2.845340	3.849516
21	0.256580	0.686352	1.323188	1.720743	2.079614	2.517648	2.831360	3.819277
22	0.256432	0.685805	1.321237	1.717144	2.073873	2.508325	2.818756	3.792131
23	0.256297	0.685306	1.319460	1.713872	2.068658	2.499867	2.807336	3.767627
24	0.256173	0.684850	1.317836	1.710882	2.063899	2.492159	2.796940	3.745399
25	0.256060	0.684430	1.316345	1.708141	2.059539	2.485107	2.787436	3.725144
26	0.255955	0.684043	1.314972	1.705618	2.055529	2.478630	2.778715	3.706612
27	0.255858	0.683685	1.313703	1.703288	2.051831	2.472660	2.770683	3.689592
28	0.255768	0.683353	1.312527	1.701131	2.048407	2.467140	2.763262	3.673906
29	0.255684	0.683044	1.311434	1.699127	2.045230	2.462021	2.756386	3.659405
30	0.255605	0.682756	1.310415	1.697261	2.042272	2.457262	2.749996	3.645959
31	0.255532	0.682486	1.309464	1.695519	2.039513	2.452824	2.744042	3.633456
32	0.255464	0.682234	1.308573	1.693889	2.036933	2.448678	2.738481	3.621802
33	0.255399	0.681997	1.307737	1.692360	2.034515	2.444794	2.733277	3.610913
34	0.255339	0.681774	1.306952	1.690924	2.032245	2.441150	2.728394	3.600716
35	0.255281	0.681564	1.306212	1.689572	2.030108	2.437723	2.723806	3.591147
inf	0.253347	0.674490	1.281552	1.644854	1.959964	2.326348	2.575829	3.290527

Si, par exemple, pour un test d'hypothèses symétriques, le seuil de signification α est égal à 10%, c'est-à-dire que $\alpha = 0.1$, alors on a $\phi_\nu(x) = 1 - \frac{0.1}{2} = 0.95$. On trouve $x = \phi_\nu^{-1}(0.95)$ en regardant la valeur se trouvant dans la colonne correspondant à 0.95 et dans la ligne correspondant aux degrés de liberté ν . Si, de plus, on choisit $\nu = 5$, on trouve $x \cong 2.015048$.



1.8.3 Les quantiles des lois du chi-carré

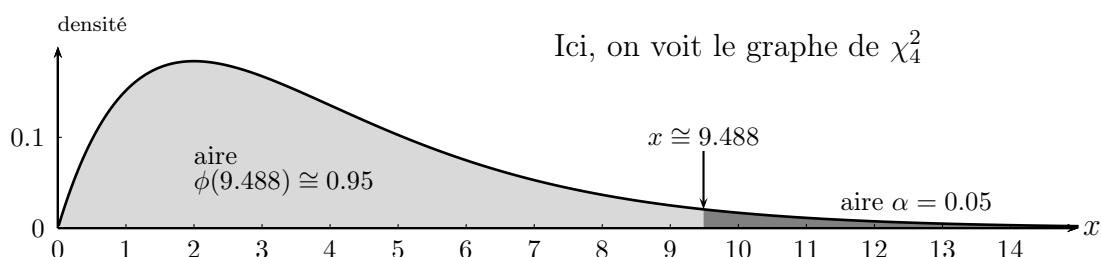
La fonction de répartition ϕ_ν de la loi du chi-carré χ_ν^2 à ν degrés de liberté est une intégrale qui n'admet pas de primitive explicite (sauf si ν est pair).

$$\phi_\nu(x) = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \int_0^x t^{\frac{\nu}{2}-1} e^{-\frac{t}{2}} dt, \quad x \in [0, +\infty[$$

Comme pour la table des lois de Student, il est coutume de préférer une table des quantiles (c'est-à-dire une table qui donne la valeur de x lorsqu'on connaît la valeur de l'aire $\phi_\nu(x)$).

$\nu \backslash \phi_\nu(x)$	0.500	0.750	0.800	0.900	0.950	0.990	0.995	0.999
1	0.455	1.323	1.642	2.706	3.841	6.635	7.879	10.828
2	1.386	2.773	3.219	4.605	5.991	9.210	10.597	13.816
3	2.366	4.108	4.642	6.251	7.815	11.345	12.838	16.266
4	3.357	5.385	5.989	7.779	9.488	13.277	14.86	18.467
5	4.351	6.626	7.289	9.236	11.07	15.086	16.75	20.515
6	5.348	7.841	8.558	10.645	12.592	16.812	18.548	22.458
7	6.346	9.037	9.803	12.017	14.067	18.475	20.278	24.322
8	7.344	10.219	11.030	13.362	15.507	20.090	21.955	26.124
9	8.343	11.389	12.242	14.684	16.919	21.666	23.589	27.877
10	9.342	12.549	13.442	15.987	18.307	23.209	25.188	29.588
11	10.341	13.701	14.631	17.275	19.675	24.725	26.757	31.264
12	11.340	14.845	15.812	18.549	21.026	26.217	28.300	32.909
13	12.340	15.984	16.985	19.812	22.362	27.688	29.819	34.528
14	13.339	17.117	18.151	21.064	23.685	29.141	31.319	36.123
15	14.339	18.245	19.311	22.307	24.996	30.578	32.801	37.697
16	15.338	19.369	20.465	23.542	26.296	32.000	34.267	39.252
17	16.338	20.489	21.615	24.769	27.587	33.409	35.718	40.790
18	17.338	21.605	22.760	25.989	28.869	34.805	37.156	42.312
19	18.338	22.718	23.900	27.204	30.144	36.191	38.582	43.820
20	19.337	23.828	25.038	28.412	31.410	37.566	39.997	45.315
21	20.337	24.935	26.171	29.615	32.671	38.932	41.401	46.797
22	21.337	26.039	27.301	30.813	33.924	40.289	42.796	48.268
23	22.337	27.141	28.429	32.007	35.172	41.638	44.181	49.728
24	23.337	28.241	29.553	33.196	36.415	42.980	45.559	51.179
25	24.337	29.339	30.675	34.382	37.652	44.314	46.928	52.620
30	29.336	34.800	36.250	40.256	43.773	50.892	53.672	59.703
35	34.336	40.223	41.778	46.059	49.802	57.342	60.275	66.619
40	39.335	45.616	47.269	51.805	55.758	63.691	66.766	73.402
45	44.335	50.985	52.729	57.505	61.656	69.957	73.166	80.077
50	49.335	56.334	58.164	63.167	67.505	76.154	79.490	86.661
55	54.335	61.665	63.577	68.796	73.311	82.292	85.749	93.168
60	59.335	66.981	68.972	74.397	79.082	88.379	91.952	99.607
65	64.335	72.285	74.351	79.973	84.821	94.422	98.105	105.988
70	69.334	77.577	79.715	85.527	90.531	100.425	104.215	112.317
75	74.334	82.858	85.066	91.061	96.217	106.393	110.286	118.599
80	79.334	88.130	90.405	96.578	101.879	112.329	116.321	124.839
85	84.334	93.394	95.734	102.079	107.522	118.236	122.325	131.041
90	89.334	98.650	101.054	107.565	113.145	124.116	128.299	137.208
95	94.334	103.899	106.364	113.038	118.752	129.973	134.247	143.344

Si, par exemple, le seuil de signification α vaut 5%, c'est-à-dire que $\alpha = 0.05$, alors on a $\phi_\nu(x) = 1 - 0.05 = 0.95$. On trouve $x = \phi_\nu^{-1}(0.95)$ en regardant la valeur se trouvant dans la colonne correspondant à 0.95 et dans la ligne correspondant aux degrés de liberté ν . Si, de plus, on choisit $\nu = 4$, on trouve $x \cong 9.488$.



Chapitre 2

Estimation de paramètres

Considérons des variables aléatoires X_1, \dots, X_n dont les mesures lors d'une expérience sont données par x_1, \dots, x_n . En général, on désire que X_1, \dots, X_n suivent une même loi qu'une variable aléatoire X d'espérance μ et de variance σ^2 .

Si on reprend l'exemple de nos 60 lancers d'un dé (à la section 1.1.2), X_i est la variable aléatoire qui donne le résultat du dé au i -ième jet, et x_i le résultat obtenu¹ lors du i -ième jet en réalisant l'expérience. On avait par exemple, $x_1 = 2$ et $x_{60} = 1$. Quant à X , il s'agit de la variable aléatoire qui représente le nombre montré par le jet de ce dé; son espérance vaut $\mu = 3.5$ et sa variance σ^2 vaut environ 2.92 (voir section 1.1.3).

2.1 Paramètres de position

Ces paramètres permettent de situer nos mesures x_1, \dots, x_n .

1. La moyenne arithmétique

La *moyenne arithmétique*, notée \bar{x} , est définie par

$$\boxed{\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i} \quad \left(= \frac{x_1 + x_2 + x_3 + x_4 + \dots + x_{n-1} + x_n}{n} \right)$$

C'est le paramètre de position qui est utilisé pour estimer la note semestrielle des élèves du point de vue de l'évaluation dans le bulletin scolaire.

La variable aléatoire correspondante à la moyenne arithmétique est notée ainsi

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

\bar{X} est un *estimateur* de l'espérance théorique μ , et \bar{x} est son *estimation*.

Remarque

C'est avec cet estimateur qu'on estime votre niveau dans chaque discipline lors de chaque semestre.

1. On voit mieux pourquoi on parle de variable aléatoire : la variable aléatoire X_i n'a rien d'aléatoire, mais sa réalisation x_i est aléatoire.

2. La médiane

Pour calculer la *médiane*, il faut classer les mesures de la plus petite à la plus grande. La médiane est la valeur du milieu lorsque le nombre des mesures est impair. C'est la moyenne arithmétique des deux valeurs autour du milieu lorsque le nombre des mesures est pair.

Par exemple

Mesures	Médiane
1 2 3 4 5	3
1 2 3 4 5 6	3.5
1 1 1 3 3 3	2

3. Le mode

Le *mode* est la valeur la plus fréquemment atteinte. Il est possible qu'une série de mesures ait plusieurs modes. Pour cette raison, le mode est un sous-ensemble des valeurs possibles.

Par exemple

Mesures	Mode
1 1 2 2 2 4 5	{2}
1 1 1 3 3	{1}
1 1 1 3 3 3	{1, 3}

2.2 Paramètres de dispersion

Ces paramètres permettent de mesurer l'écartement des mesures x_1, \dots, x_n .

1. L'étendue

L'*étendue* est la longueur du *domaine de variation* qui est le plus petit intervalle contenant toutes les valeurs.

Par exemple

Mesures	Domaine de variation	étendue
1 2 2 4 5	[1, 5]	4
1 1 3 3 3 3	[1, 3]	2
1 1 3 3 7	[1, 7]	6
2 2 2 3 3 7	[2, 7]	5

2. La variance

La *variance*, notée s_n^2 , est la moyenne des carrés des écarts par rapport à \bar{x} .

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \left(= \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} \right)$$

Par exemple, si les mesures sont $x_1 = 1$, $x_2 = 2$, $x_3 = 2$, $x_4 = 4$ et $x_5 = 6$. Alors la moyenne arithmétique vaut $\bar{x} = 3$ et on peut calculer la variance ainsi

$$\begin{aligned} s_n^2 &= \frac{(1-3)^2 + (2-3)^2 + (2-3)^2 + (4-3)^2 + (6-3)^2}{5} \\ &= \frac{4 + 1 + 1 + 1 + 9}{5} = \frac{16}{5} = 3.2 \end{aligned}$$

La variable aléatoire correspondante à la variance est notée ainsi

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

S_n^2 est un *estimateur* de la variance théorique σ^2 , et s_n^2 est son *estimation*.

3. L'écart type

L'*écart type* est égal à la racine carrée de la variance. Il est logiquement noté s_n .

$$s_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad \left(= \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}} \right)$$

Dans l'exemple précédent, on avait la variance $s_n^2 = 3.2$. Ainsi l'écart type correspondant vaut $s_n = \sqrt{3.2} \cong 1.79$.

De même, la variable aléatoire correspondante à l'écart type est notée ainsi

$$S_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

S_n est un *estimateur* de l'écart type théorique σ , et s_n est son *estimation*.

L'estimateur sans biais de la variance et de l'écart type

L'estimateur de l'espérance μ donné par

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

est un bon estimateur : on dit qu'il est sans biais (voir section 1).

Par contre, l'estimateur de la variance σ^2 donné par

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

a un biais, il a en effet tendance à sous-estimer σ^2 .

On corrige ce biais en utilisant l'estimateur, noté S_{n-1} ou plus simplement S , défini par

$$S^2 = S_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \left(= \frac{n}{n-1} S_n^2 \right)$$

Pour les mêmes raisons, l'estimateur sans biais de l'écart type σ est

$$S = S_{n-1} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

2.3 Les boîtes à moustaches

Inventées par John Tukey et John Wilder en 1977, les boîtes à moustaches, aussi appelées diagrammes en boîte ou boîtes de Tukey, sont très utiles pour donner une représentation simple à comprendre des données mesurées (un peu comme un histogramme).

2.3.1 Les quartiles de Tukey et les quartiles de Freund et Perles

Les *quartiles* permettent de séparer les données en quatre parties égales ; c'est une généralisation de la médiane qui sépare les données en deux parties égales.

Le quartile inférieur sépare le premier quart des données ; le deuxième quartile est la médiane : il sépare la moitié des données ; le quartile supérieur sépare le dernier quart des données.

Voici deux manières de calculer les quartiles à partir d'une série de données.

1. Les quartiles utilisés pour faire les boîtes à moustaches.

Cette méthode a été établie par John Tukey en 1983.

On met les n données dans l'ordre et on coupe les données en deux ensembles sur lesquels on calcule la médiane.

- (a) Si n est impair, il y a une valeur centrale (la médiane), et on coupe les données en deux ensembles en mettant la médiane dans chacun des deux ensembles. Le quartile inférieur est la médiane du premier ensemble ; le quartile supérieur est la médiane du deuxième ensemble.
- (b) Si n est pair, il y a deux valeurs centrales (la médiane est la moyenne arithmétique de ces deux valeurs), et on coupe en deux ensembles en mettant dans chaque ensemble la valeur centrale correspondante.

Voici des exemples pour cette première méthode.

Mesures	Les ensembles	Quartile inférieur	Quartile supérieur
1 2 4 5	{ 1, 2 } { 4, 5 }	1.5	4.5
1 3 5 5 7	{ 1, 3, 5 } { 5 , 5, 7 }	3	5
1 3 4 6 7 9	{ 1, 3, 4 } { 6 , 7, 9 }	3	7
1 3 5 6 7 9 15	{ 1, 3, 5, 6 } { 6 , 7, 9, 15 }	4	8

Sous «R», on trouve ces quartiles en utilisant la commande `fivenum` (qui donne le minimum, le quartile inférieur, la médiane, le quartile supérieur, le maximum).

2. Les quartiles utilisés par les logiciels «Excel», «OpenOffice» et «R».

Cette méthode a été établie par John Freund et Benjamin Perles en 1987.

On met les n données dans l'ordre. Le quartile inférieur est la donnée numéro $\frac{1n+3}{4}$, le quartile supérieur est la donnée numéro $\frac{1+3n}{4}$. Si ces numéros ne sont pas entiers, on utilise une interpolation linéaire pour trouver les quartiles inférieur et supérieur (voir page suivante).

Lorsque n est impair, on retrouve les quartiles de Tukey.

Concernant l'interpolation linéaire

Si on cherche la donnée numéro 3.25, on calcule la moyenne pondérée $\frac{3x_3+x_4}{4}$ (on voit l'analogie avec les notes : 3.25 est la moyenne des notes 3, 3, 3 et 4).

Si on cherche la donnée numéro 3.5, on calcule la moyenne pondérée $\frac{2x_3+2x_4}{4} = \frac{x_3+x_4}{2}$.

Si on cherche la donnée numéro 3.75, on calcule la moyenne pondérée $\frac{x_3+3x_4}{4}$.

Reprenons les mêmes exemples que pour la première méthode.

Mesures	$\frac{1n+3}{4}$	$\frac{3n+1}{4}$	Quartile inférieur	Quartile supérieur
1 2 4 5	1.75	3.25	1.75	4.25
1 3 5 5 7	2	4	3	5
1 3 4 6 7 9	2.25	4.75	3.25	6.75
1 3 5 6 7 9 15	2.5	5.5	4	8

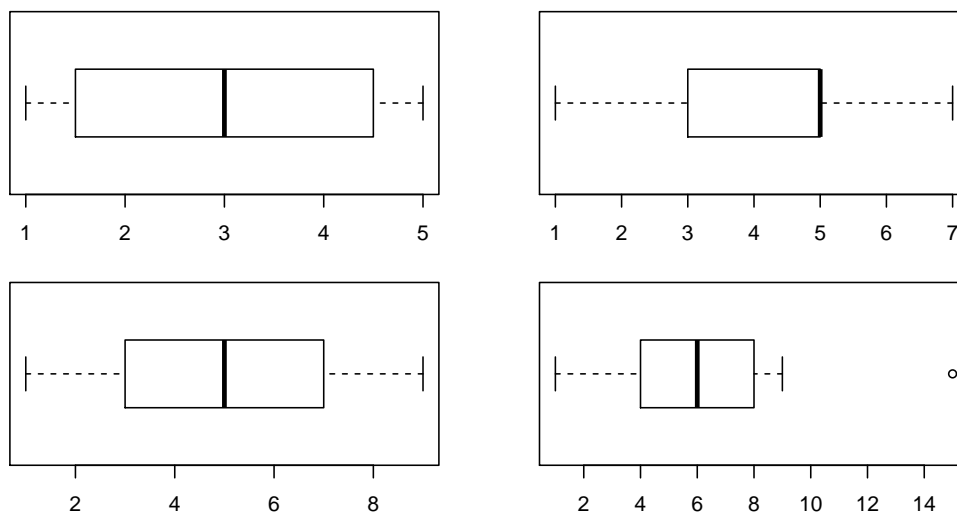
Sous «R», on trouve ces quartiles en utilisant la commande `quantile` (qui donne le minimum, le quartile inférieur, la médiane, le quartile supérieur, le maximum).

Sous «OpenOffice», on les trouve à l'aide de la commande `quantile(zone;nombre)` (qui donne le minimum pour `nombre = 0`, le quartile inférieur pour `nombre = 1`, la médiane pour `nombre = 2`, le quartile supérieur pour `nombre = 3`, le maximum pour `nombre = 4`).

2.3.2 La boîte à moustaches de Tukey

C'est une boîte dont les côtés correspondent aux quartiles inférieur et supérieur ; on indique la médiane par un trait à l'intérieur de cette boîte. Des moustaches sortent de la boîte pour indiquer les valeurs minimale et maximale. Ces moustaches² font au plus 150% de la longueur la boîte, sinon on prend les valeurs minimale ou maximale suivantes et on indique les valeurs extrêmes par des points : ce sont les valeurs atypiques.

Voici les boîtes à moustaches qui correspondent aux exemples ci-dessus (avec la première méthode). Sur le dernier exemple, la longueur de la boîte vaut 4 ($= 8 - 4$) et la dernière valeur est atypique (car elle est à distance 7 (> 6) du quartile supérieur).



Sous «R», on peut dessiner ces boîtes à moustaches en utilisant la commande `boxplot`.

2. Selon les contextes, les moustaches peuvent être construites de manière plus théoriques.

2.4 Statistique, estimateur et estimation

Une pièce de monnaie est jetée six fois de suite. Imaginons que l'on obtienne la suite $(P; P; F; P; F; P)$ où P signifie que la pièce est tombée sur le côté pile et F sur le côté face.

Du point de vue des probabilités, à chaque fois qu'on lance une pièce de monnaie, on a une probabilité $p \in [0, 1]$ d'obtenir le côté pile et $q = 1 - p$ d'obtenir le côté face. Si $p = 1$, cela signifie que la pièce est truquée et qu'elle a deux côtés pile. Si $p = \frac{2}{3}$, cela signifie que la pièce n'est pas bien équilibrée et qu'elle tombe deux fois plus souvent sur pile que face. Si $p = \frac{1}{2}$ cela signifie que la pièce est bien équilibrée. Il n'existe probablement aucune pièce de monnaie sur terre qui soit parfaitement bien équilibrée, néanmoins on se permet la plupart du temps de supposer que $p = \frac{1}{2}$!

À une telle expérience, on fait correspondre une famille de variables aléatoires X_i qui indiquent 1 si le i -ième jet tombe sur pile et 0 s'il tombe sur face. Ainsi, la suite obtenue se traduit par $(x_1, x_2, x_3, x_4, x_5, x_6) = (1, 1, 0, 1, 0, 1)$ où les x_i sont les données observées. Comme la pièce de monnaie reste à la même à chaque nouveau lancé, on peut raisonnablement supposer que chacune de ces variables aléatoires suit la même loi (c'est une loi de Bernoulli comme vue en section 1.1.6) donnée par

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

Dans ce cas, il était facile de trouver la distribution (comme la plupart du temps). Il est déjà plus difficile d'estimer son paramètre, ici p !

Définition

Toute fonction T des variables aléatoires X_1, X_2, \dots, X_n est appelée *statistique*. Lorsqu'une telle fonction est utilisée pour estimer des paramètres d'une même distribution des variables aléatoires X_1, \dots, X_n , on dit que $T(X_1, \dots, X_n)$ est un *estimateur*.

Notations

1. Si $T = T(X_1, \dots, X_n)$ est l'estimateur du paramètre inconnu, $t = T(x_1, \dots, x_n)$ est la *réalisation de cet estimateur*, aussi appelée *estimation*, à l'aide des données observées.
2. Lorsqu'on estime un paramètre θ , on a l'habitude d'écrire son estimateur avec un chapeau : $\hat{\theta}$.

Remarque

Une statistique T est une variable aléatoire !

Exemple

Dans notre exemple, il est naturel d'estimer p à l'aide de l'estimateur suivant :

$$\hat{P} = \hat{P}(X_1, \dots, X_6) = \frac{\text{nombre de pile}}{\text{nombre total de jets}} = \frac{\sum_{i=1}^6 X_i}{6} = \bar{X}$$

La réalisation de cet estimateur est dans notre cas l'estimation $\hat{p} = \frac{\sum_{i=1}^6 x_i}{6} = \frac{4}{6} = \frac{2}{3}$.

2.5 La méthode du maximum de vraisemblance

Une des techniques parmi les plus populaires pour trouver des estimateurs est celle dite du *maximum de vraisemblance*. Si X_1, \dots, X_n est une famille de variables aléatoires indépendantes issues de la même densité $f(x; \theta)$ (où θ est le paramètre de la densité f). La fonction de vraisemblance³ est définie par :

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta)$$

Justification de cette formule dans le cas discret

Il faut penser à $L(\theta; x_1, \dots, x_n)$ comme étant la probabilité que l'échantillon x_1, \dots, x_n se produise. Dans le cas discret on a

$$L(\theta; x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1 \text{ et } X_2 = x_2 \text{ et } \dots \text{ et } X_n = x_n)$$

Comme les X_i sont indépendants, on a $L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i) = \prod_{i=1}^n f(x_i; \theta)$.

L'estimateur du maximum de vraisemblance

L'estimateur du maximum de vraisemblance⁴ de θ , noté $\hat{\theta}_{\text{MLE}}$, est l'estimateur pour lequel la fonction de vraisemblance est la plus grande. Autrement dit, il s'agit de la valeur du paramètre pour laquelle l'échantillon observé est le plus vraisemblable.

Pour chercher cet estimateur, on doit trouver le maximum de la fonction $L(\theta; x_1, \dots, x_n)$. C'est une optimisation classique : un symptôme d'un maximum est un zéro de la dérivée. En pratique, il peut être plus facile de trouver le maximum du logarithme de la fonction de vraisemblance (en effet, la fonction \ln est croissante et transforme les produits en somme ; ainsi $\ln(L(\theta; x_1, \dots, x_n)) = \sum_{i=1}^n \ln(f(x_i; \theta))$).

Exemple : les MLEs de la loi normale

On considère la loi normale de paramètres μ et σ dont la densité est donnée en page 16. En exercice, on trouvera⁵ les estimateurs du maximum de vraisemblance suivants.

$$\hat{\mu}_{\text{MLE}} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{et} \quad \hat{\sigma}_{\text{MLE}} = \sqrt{\frac{\sum_{i=1}^n (X_i - \hat{\mu}_{\text{MLE}})^2}{n}}$$

Théorème Si f est une fonction d'un estimateur θ , alors $\widehat{f(\theta)}_{\text{MLE}} = f(\hat{\theta}_{\text{MLE}})$.

Idée de la preuve

Le MLE $\hat{\theta}_{\text{MLE}}$ est la valeur pour laquelle la fonction de vraisemblance est maximale. On nomme $\nu = f(\theta)$ un nouveau paramètre. On doit exprimer la fonction de vraisemblance en fonction de ce nouveau paramètre ; on a $\theta = f^{-1}(\nu)$ (on crée une fonction réciproque de f en choisissant un domaine de définition). Ainsi $L(\theta) = L(f^{-1}(\nu))$. Cette fonction admet un maximum lorsque $f^{-1}(\hat{\nu}_{\text{MLE}}) = \hat{\theta}_{\text{MLE}}$, c'est-à-dire $\hat{\nu}_{\text{MLE}} = f(\hat{\theta}_{\text{MLE}})$.

3. En anglais, vraisemblance se dit *likelihood*. Ainsi la fonction de vraisemblance est notée L .

4. En anglais, estimateur du maximum de vraisemblance se dit **maximum likelihood estimator** (MLE).

5. Lorsqu'on cherche à optimiser une fonction à plusieurs variables, on cherche les valeurs des variables qui annulent le *gradient* (c'est le vecteur dont la i -ème composante est la dérivée de la fonction par rapport à la i -ième variable uniquement).

2.6 Qualités d'un estimateur

Le biais d'un estimateur

Soit $\hat{\theta}$ un estimateur d'un paramètre θ . Le *biais* de $\hat{\theta}$ est défini par :

$$b_{\theta}(\hat{\theta}) = E(\hat{\theta} - \theta) \stackrel{\substack{\text{linéarité de} \\ \text{l'espérance}}}{=} \downarrow E(\hat{\theta}) - \theta$$

Un estimateur $\hat{\theta}$ est dit *sans biais* lorsque $b_{\theta}(\hat{\theta}) = 0$. Autrement dit, lorsqu'en moyenne l'estimateur est égal au paramètre que l'on cherche à estimer.

Exemple : les estimateurs les plus célèbres

On suppose qu'on a n variables aléatoires X_i indépendantes et suivant toutes une même loi d'espérance μ et de variance σ^2 . Alors l'espérance μ peut être estimée sans biais par l'estimateur \bar{X} et la variance σ^2 peut être estimée sans biais par l'estimateur S^2 . Ces estimateurs sont définis comme suit :

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{et} \quad S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Dans le cas d'une loi normale $\mathcal{N}(\mu, \sigma^2)$, on remarque que \bar{X} n'est rien d'autre que l'estimateur de vraisemblance de μ , tandis que $S^2 = \frac{n}{n-1} \hat{\sigma}_{\text{MLE}}^2$.

1. L'estimateur de la moyenne est sans biais

En effet, on a :

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) \stackrel{\substack{\text{linéarité de} \\ \text{l'espérance}}}{=} \downarrow \frac{\sum_{i=1}^n E(X_i)}{n} = \frac{\sum_{i=1}^n \mu}{n} = \frac{n\mu}{n} = \mu$$

Donc l'estimateur \bar{X} de l'espérance μ est sans biais, car $b_{\mu}(\bar{X}) = E(\bar{X}) - \mu = 0$.

2. L'estimateur de la variance est sans biais

En effet, on a :

$$\begin{aligned} E(S^2) &= E\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}\right) \stackrel{\text{lin.}}{=} \frac{1}{n-1} \sum_{i=1}^n E((X_i - \bar{X})^2) \\ &= \frac{1}{n-1} \sum_{i=1}^n E(X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &\stackrel{\text{lin.}}{=} \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - 2E\left(\sum_{i=1}^n X_i \cdot \bar{X}\right) + \sum_{i=1}^n E(\bar{X}^2) \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - 2nE\left(\frac{\sum_{i=1}^n X_i}{n} \cdot \bar{X}\right) + nE(\bar{X}^2) \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - 2nE(\bar{X} \cdot \bar{X}) + nE(\bar{X}^2) \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2) \right) \end{aligned}$$

Or, d'après la seconde formule de la variance, on a $V(X) = E(X^2) - E(X)^2$. Ainsi, on a la formule suivante pour n'importe quelle variable aléatoire X :

$$E(X^2) = V(X) + E(X)^2 \quad \star$$

Cela permet de reprendre le calcul précédent :

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2) \right) \\ &\stackrel{\star}{=} \frac{1}{n-1} \left(\sum_{i=1}^n (V(X_i) + E(X_i)^2) - n(V(\bar{X}) + E(\bar{X})^2) \right) \end{aligned}$$

Puisque l'espérance de chaque X_i est égale à μ , que la variance de chaque X_i est égale à σ^2 et que l'estimateur \bar{X} est sans biais, on a :

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} \left(\sum_{i=1}^n (\sigma^2 + \mu^2) - n(V(\bar{X}) + \mu^2) \right) \\ &= \frac{1}{n-1} (n(\sigma^2 + \mu^2) - n(V(\bar{X}) + \mu^2)) = \frac{1}{n-1} (n\sigma^2 - nV(\bar{X})) \end{aligned}$$

Or, grâce au théorème de la page 24 qui dit que $V(\bar{X}) = \frac{\sigma^2}{n}$ (parce que les variables aléatoires sont indépendantes), on peut finalement montrer que S^2 est sans biais, car :

$$E(S^2) = \frac{1}{n-1} (n\sigma^2 - nV(\bar{X})) = \frac{1}{n-1} \left(n\sigma^2 - n\frac{\sigma^2}{n} \right) = \frac{1}{n-1} (n-1)\sigma^2 = \sigma^2$$

L'erreur quadratique moyenne de l'erreur d'un estimateur

Soit $\hat{\theta}$ un estimateur d'un paramètre θ . L'erreur quadratique moyenne de $\hat{\theta}$ est définie⁶ par :

$$\text{MSE}_\theta(\hat{\theta}) = E((\hat{\theta} - \theta)^2) \stackrel{\star}{=} V(\hat{\theta}) + (b_\theta(\hat{\theta}))^2$$

Où $V(\hat{\theta})$ est la *variance* de $\hat{\theta}$ (rappelons qu'un estimateur est une variable aléatoire).

Théorème (sans preuve)

On suppose qu'on a n variables aléatoires X_i indépendantes et qui suivent toutes une loi normale $\mathcal{N}(\mu, \sigma^2)$ d'espérance μ et de variance σ^2 . On a :

$$\text{MSE}_\mu(\bar{X}) = V(\bar{X}) = \frac{\sigma^2}{n} \quad \text{et} \quad \text{MSE}_{\sigma^2}(S^2) = V(S^2) = \frac{2\sigma^4}{n-1}$$

6. En anglais, on parle de **mean squared error**, abrégée MSE.

Chapitre 3

Tests d'hypothèses

«Un statisticien est une personne dont l'ambition principale est d'avoir tort dans 5% des cas.»
Anonyme

Les *tests d'hypothèses* sont utiles pour vérifier si une affirmation sur un modèle théorique correspondant à une expérience aléatoire est cohérente avec les mesures effectuées.

Dans une étude statistique, on peut se demander si les mesures observées peuvent correspondre à une certaine réalité. Par exemple, est-ce que la série de jets d'une pièce de monnaie ($P; P; F; P; F; P$) peut correspondre à une pièce bien équilibrée ?

Le but d'un test d'hypothèses est de confronter deux hypothèses entre elles : l'*hypothèse nulle* H_0 et l'*hypothèse alternative* H_1 . Les *hypothèses* sont des énoncés qui concernent un paramètre d'une population. La confrontation s'effectue à l'aide d'un estimateur du paramètre en question, appelé *statistique de test*.

Dans l'exemple des jets de pièce de monnaie (voir page 38), le paramètre en question est la probabilité p que la pièce tombe sur pile. Son estimateur est donné par $\hat{P} = \bar{X}$. L'hypothèse nulle sera "la pièce est bien équilibrée ($p = \frac{1}{2}$)" et l'hypothèse alternative sera "la pièce n'est pas bien équilibrée ($p \neq \frac{1}{2}$)".

On va traiter plusieurs cas :

1. Tests d'hypothèses sur une moyenne ou une proportion.

(a) Variance connue ou variance inconnue.

Selon le fait que la variance est connue ou inconnue, on utilisera une loi différente pour effectuer les calculs.

(b) Tests symétriques ou asymétriques.

Dans le cours, on présente les tests symétriques, les tests asymétriques seront vus en exercices.

Les tests d'hypothèses sur une proportion sont des tests où la variance est connue, parce que des lois de Bernoulli et binomiale apparaissent.

2. Test d'hypothèses de comparaison de moyenne : données appariées ou non.

3. Les tests du chi-carré

- a) pour l'adéquation à une loi ; b) pour la comparaison d'échantillon ;
c) pour l'indépendance.

Il existe d'autres tests d'hypothèses. Par exemple, les tests permettant d'inférer la variance σ^2 d'une population. Ces tests utilisent d'autres lois de probabilité, comme, par exemple, la distribution de Fisher à deux paramètres.

3.1 Tests d'hypothèses symétriques sur une moyenne

3.1.1 Tests d'hypothèses symétriques sur une moyenne, variance connue

Considérons des variables aléatoires X_1, \dots, X_n indépendantes qui suivent une même loi d'espérance μ (c'est la moyenne théorique qu'on cherche à tester) et de variance σ^2 (supposée connue).

Dans ce modèle de tests d'hypothèses, on teste l'hypothèse «la moyenne théorique vaut μ_0 » à partir des n mesures effectuées.

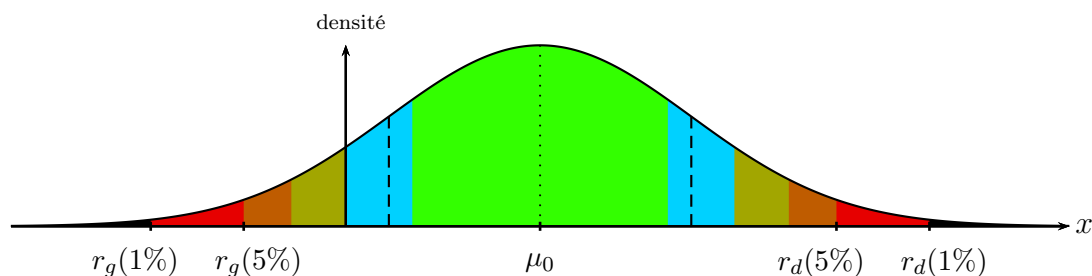
On se trouve face à deux alternatives, appelées *hypothèses*, qui sont

$$\begin{array}{ll} \text{hypothèse nulle} & \text{hypothèse alternative} \\ H_0 : \mu = \mu_0 & \text{et} \quad H_1 : \mu \neq \mu_0 \end{array}$$

Comme dans un raisonnement par l'absurde, on suppose qu'on se trouve sous l'hypothèse H_0 et on regarde si les données mesurées permettent d'en tirer une contradiction.

Sous l'hypothèse H_0 , les variables aléatoires X_i sont d'espérance μ_0 .

Par le théorème de la limite centrale, $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ suit approximativement une loi normale $\mathcal{N}(\mu_0, \frac{\sigma^2}{n})$. En traitillés, on voit les bornes de l'intervalle $[\mu_0 - \frac{\sigma}{\sqrt{n}}, \mu_0 + \frac{\sigma}{\sqrt{n}}]$; on remarque ainsi que lorsque n grandit la courbe se resserre.



On estime \bar{X} par la moyenne des mesures effectuées $\bar{x} = \frac{x_1 + \dots + x_n}{n}$ (attention à bien faire la différence entre les majuscules et les minuscules). L'estimation \bar{x} donnée par les mesures va donc se trouver quelque part sous la loi.

Il y a une probabilité de 60% que \bar{x} tombe dans la zone ■, il y a une probabilité de 20% que \bar{x} tombe dans la zone ■, il y a une probabilité de 10% que \bar{x} tombe dans la zone ■, il y a une probabilité de 5% que \bar{x} tombe dans la zone ■, il y a une probabilité de 4% que \bar{x} tombe dans la zone ■, il y a une probabilité de 1% que \bar{x} tombe dans la zone ■.

Plus on se trouve dans une zone éloignée de μ_0 donc dans l'ordre : ■, ■, ■, ■, ■, ■; plus les mesures sont en contradiction avec l'hypothèse H_0 . On décide ainsi du critère suivant.

On rejette l'hypothèse H_0 au seuil de signification 5% si \bar{x} se trouve dans la zone ■ ou ■.

On rejette l'hypothèse H_0 au seuil de signification 1% si \bar{x} se trouve dans la zone ■.

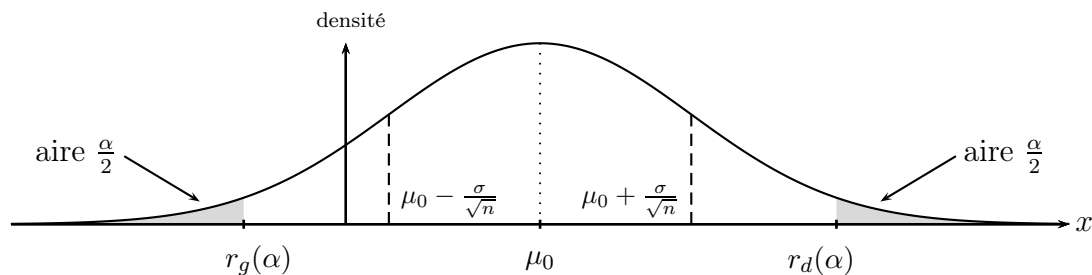
Le seuil de signification du test est la probabilité α que l'on a de rejeter l'hypothèse H_0 sachant que H_0 est vraie.

$$\mathbb{P}(\text{rejeter } H_0 \mid H_0 \text{ est vraie}) = \alpha$$

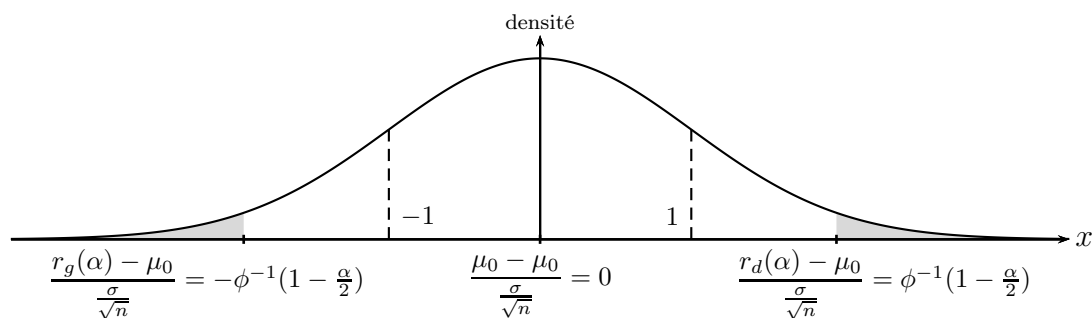
La distribution ci-dessous est la distribution sous l'hypothèse H_0 , donc α se voit sur le dessin : c'est l'aire sous la distribution dans la zone de rejet de H_0 (en grisé).

On montre dans le cours OS que si on a une variable N qui suit une loi normale $\mathcal{N}(\mu_0, \frac{\sigma^2}{n})$, alors la variable $\frac{N - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ suit la loi normale $\mathcal{N}(0, 1)$. On dit que N a été *centrée-réduite*.

Si la densité de \bar{X} est



Alors la densité de $\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ est (l'échelle a changé)



où $\phi(x) = \mathbf{P}(Z \leq x)$ est la fonction de répartition de la variable aléatoire Z qui suit la loi normale $\mathcal{N}(0, 1)$. Les valeurs de ϕ et ϕ^{-1} se trouvent dans la table de la page 30.

On rejette H_0 au seuil de signification α si

$$\begin{array}{ll} \text{pour la variable aléatoire } \bar{X} & \text{pour la variable aléatoire } \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \\ \bar{x} < r_g(\alpha) & \left| \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| > \phi^{-1}\left(1 - \frac{\alpha}{2}\right) \\ \text{ou } \bar{x} > r_d(\alpha) & \end{array}$$

En particulier, si $\alpha = 5\%$, le critère est $\left| \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| > \phi^{-1}(0.975) \stackrel{\text{table}}{\cong} 1.96$.

De même, si $\alpha = 1\%$, le critère est $\left| \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| > \phi^{-1}(0.995) \stackrel{\text{table}}{\cong} 2.58$.

Exemple

Le test “Gauche-Droite” de Jean Piaget a pour but de vérifier l’acquisition par l’enfant des notions gauche-droite à différents points de vue et d’évaluer ainsi son niveau de socialisation et subjectivisation. Les enfants de 7 ans obtiennent en moyenne 12 comme résultat avec un écart type de 3.4. On applique ces tests à 25 enfants gauchers de 7 ans choisis au hasard et on obtient un résultat moyen de 13.4.

À partir de ces tests, peut-on affirmer qu’il y a une différence significative entre les gauchers et les droitiers ?

Réponse

Ici, on met en doute le fait que la moyenne théorique des gauchers est égale à celle des droitiers. On suppose par ailleurs que l’écart type théorique est le même pour les deux populations¹. L’approximation par le théorème de la limite centrale est de bonne qualité car $n = 25$. On peut donc effectuer le test avec les alternatives suivantes.

$$\begin{array}{ll} \textit{hypothèse nulle} & \textit{hypothèse alternative} \\ H_0 : \mu = 12 & \text{et} \quad H_1 : \mu \neq 12 \end{array}$$

Prétendre que $\mu = 12$ revient à prétendre que les gauchers suivent le même modèle que celui de l’ensemble de la population (donc a fortiori celui des droitiers).

Les données livrent une moyenne $\bar{x} = 13.4$. On a ainsi

$$\left| \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| = \left| \frac{13.4 - 12}{\frac{3.4}{\sqrt{25}}} \right| \cong 2.059$$

On voit qu’on rejette H_0 au seuil 5%, mais qu’on ne peut pas rejeter H_0 au seuil 1% (en fait \bar{x} est tombé dans la zone ■, mais pas dans la zone ■).

Remarques

1. En supposant que H_0 soit vraie, et qu’un grand nombre d’examineurs aient fait passer ces tests à d’autres groupes de 25 gauchers, alors 1% des examineurs auraient eu une moyenne inférieure à 10.25 ou supérieure à 13.75 et 5% des examineurs auraient eu une moyenne inférieure à 10.7 ou supérieure à 13.3.
2. On ne fixe pas le seuil de signification après avoir fait le test, mais avant. Ceci afin d’éviter de régler le seuil pour que le test donne le résultat escompté après avoir effectué les mesures.
3. Il est habituel de prendre 1% pour confirmer H_0 et 5% pour infirmer H_0 , mais ces seuils sont arbitraires. En cas de doute sur la conclusion du test, il est important d’aller regarder de plus près comment les mesures ont été effectuées. On peut aussi refaire l’expérience sur un autre échantillon. Si lors des mesures, une erreur a été effectuée, le test ne sert plus à rien.
4. Pour qu’un test d’hypothèses soit utile, il faut que les mesures aient été effectuées sur un échantillon représentatif de la population.

1. Si le lecteur n’est pas d’accord avec cette hypothèse, il peut utiliser le test de Student mis au point par William Gosset (voir section suivante).

Les différents types d'erreurs

Il y a deux probabilités α et β qui décrivent deux types de risque qu'on rencontre en effectuant un test d'hypothèses. On a déjà parlé de α qui est aussi appelé le *seuil de signification d'un test d'hypothèse*, mais il y a aussi β qui est défini ci-dessous.

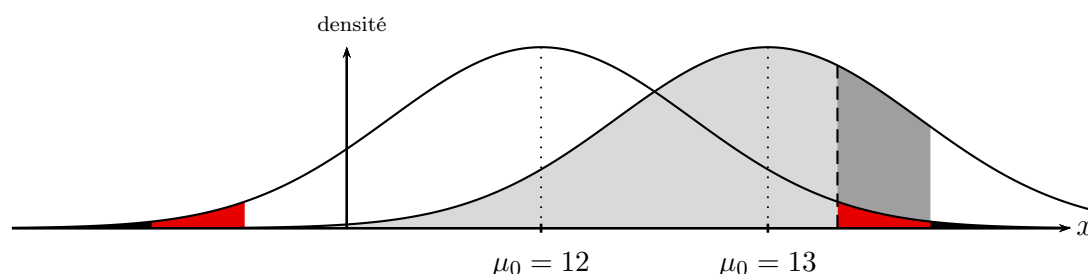
$$\begin{aligned}\alpha &= \mathbf{P}(\text{rejeter } H_0 \mid H_0 \text{ est vraie}) \\ \beta &= \mathbf{P}(\text{ne pas rejeter } H_0 \mid H_0 \text{ est fautive})\end{aligned}$$

Comme il est difficile de calculer β , dans la pratique on calcule plutôt des β_{μ_i} comme définis ci-dessous.

	H_0 est vraie	H_0 est fautive		
	$\mu = \mu_0$	$\mu = \mu_1$...	$\mu = \mu_k$
On rejette H_0	Mauvaise décision Probabilité α	Bonne décision Probabilité $1 - \beta_{\mu_1}$...	Bonne décision Probabilité $1 - \beta_{\mu_k}$
On ne rejette pas H_0	Bonne décision Probabilité $1 - \alpha$	Mauvaise décision Probabilité β_{μ_1}	...	Mauvaise décision Probabilité β_{μ_k}

Le seuil α est une probabilité appelée *risque de première espèce* et les nombres β_{μ_i} , tout comme β , sont appelées *risques de deuxième espèce*.

Calcul d'un risque de deuxième espèce (suite de l'exemple précédent)



Pour un seuil $\alpha = 5\%$, le risque de deuxième espèce avec $H_1 : \mu = 13$ donne $\beta_{13} \cong 69\%$ (c'est la proportion de la zone en gris clair déterminée par les zones ■ et ■). Pour un seuil $\alpha = 1\%$, le risque de deuxième espèce avec $H_1 : \mu = 13$ donne $\beta_{13} \cong 87\%$ (c'est la proportion de la zone en gris clair et foncé déterminée par la zone ■).

Autrement dit, au seuil $\alpha = 5\%$. Si H_0 est vrai, la bonne décision est de ne pas rejeter l'hypothèse; sa probabilité est de 95%. La mauvaise décision a une probabilité de 5%.

Par contre (au même seuil α), si au lieu de H_0 , c'est $H_1 : \mu = 13$ qui est vrai, alors la bonne décision est de rejeter H_0 , la probabilité est maintenant de 31%. La mauvaise décision a une probabilité de 69%.

On le voit sur cet exemple, si le risque de première espèce diminue, alors le risque de deuxième espèce augmente.

Pour diminuer le risque de deuxième espèce, il faudrait augmenter le seuil α et par conséquent le risque de première espèce.

C'est un compromis qu'il faut savoir accepter.

3.1.2 Test d'hypothèses symétriques sur une moyenne, variance inconnue

On suppose que les variables aléatoires X_1, \dots, X_n qui correspondent aux n valeurs observées sont indépendantes et suivent une loi normale $\mathcal{N}(\mu, \sigma^2)$ où les paramètres μ et σ sont inconnus.

Sous ces hypothèses, l'estimateur \bar{X} suit une loi normale de paramètres $\mathcal{N}(\mu, \frac{\sigma^2}{n})$.

On teste l'hypothèse nulle $H_0 : \mu = \mu_0$ contre l'hypothèse alternative $H_1 : \mu \neq \mu_0$.

Comme avant, on se place sous l'hypothèse nulle et on contemple. Malheureusement, on ne peut pas utiliser la loi $\mathcal{N}(\mu_0, \frac{\sigma^2}{n})$ pour calculer les probabilités puisque σ est inconnu. Heureusement, William Gosset, brasseur et mathématicien britannique trouva une loi qui permit de travailler sous l'hypothèse H_0 . Il publia son résultat sous le pseudonyme de Student.

Théorème (sans preuve)

Si X_1, \dots, X_n sont des variables aléatoires indépendantes qui suivent une loi normale $\mathcal{N}(\mu, \sigma^2)$, alors la variable aléatoire $\frac{\bar{X} - \mu}{S/\sqrt{n}}$ suit une distribution de Student avec $n - 1$ degrés de liberté.

Théorème

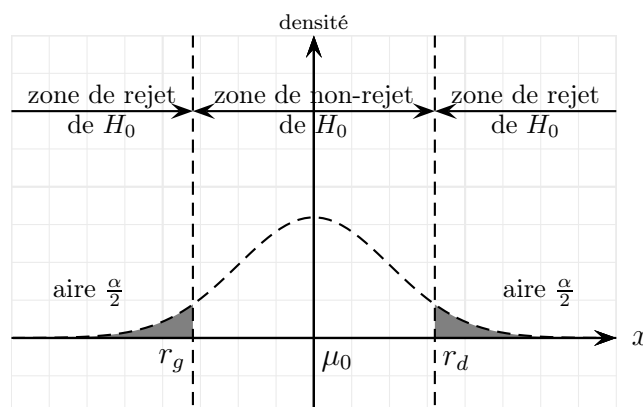
On rejette l'hypothèse H_0 au seuil de signification α si³

$$\bar{x} \notin \left[\mu_0 - \phi_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}}, \mu_0 + \phi_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}} \right] \quad \text{ou} \quad \left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| > \phi_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right)$$

où $\phi_{n-1}(x)$ est la fonction de répartition de la loi de Student à $n - 1$ degrés de liberté.

Preuve

On se fixe le seuil de signification α . Sous l'hypothèse H_0 , on sait que \bar{X} suit une distribution normale $\mathcal{N}(\mu_0, \frac{\sigma^2}{n})$ de la forme suivante :



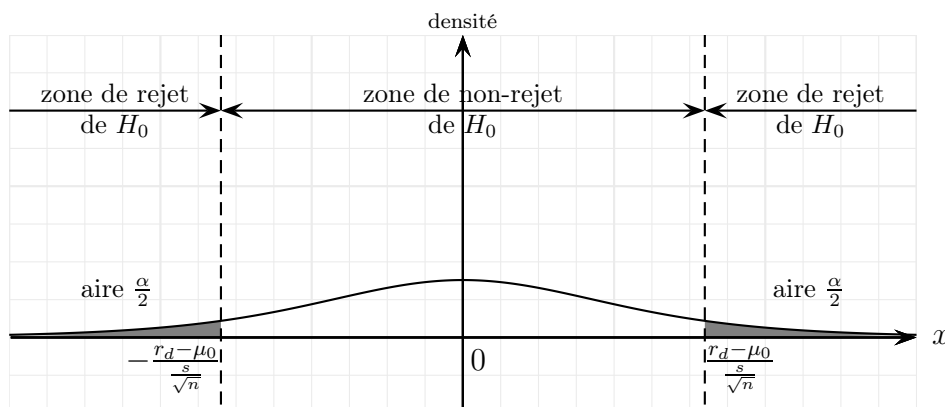
2. Ce n'est pas la variable centrée réduite de \bar{X} , car on a remplacé σ par son estimateur sans biais S .

3. Dans l'énoncé de ce théorème, s est l'estimation de l'estimateur S , tout comme \bar{x} est l'estimation de l'estimateur \bar{X} .

Malheureusement, comme σ est inconnu, on ne peut pas utiliser cette loi pour calculer la valeur exacte de r_d sous l'hypothèse H_0 . Mais grâce à Gosset, on sait que la variable aléatoire $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ suit approximativement une distribution de Student avec $n - 1$ degrés de liberté.

$$1 - \frac{\alpha}{2} = \mathbb{P}\left(\frac{\bar{X}-\mu}{S/\sqrt{n}} \leq \frac{r_d-\mu}{s/\sqrt{n}}\right) \stackrel{\substack{\text{on se place sous} \\ \text{l'hypothèse } H_0}}{\cong} \mathbb{P}\left(\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \leq \frac{r_d-\mu_0}{s/\sqrt{n}}\right) \cong \phi_{n-1}\left(\frac{r_d-\mu_0}{s/\sqrt{n}}\right)$$

Voici la représentation de cette loi de Student :



Donc, on a :

$$\boxed{\frac{r_d-\mu_0}{s/\sqrt{n}} = \phi_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right)} \iff r_d - \mu_0 = \phi_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}} \iff \boxed{r_d = \mu_0 + \phi_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}}}$$

Donc, on rejette l'hypothèse H_0 au seuil de signification α si l'estimation \bar{x} est plus grande que r_d ou plus petite que $r_g = \mu_0 - \phi_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}}$. \square

3.1.3 Résumé et autres statistiques de tests sur les moyennes

Dans les tableaux ci-dessous, on note $z_\alpha = \phi^{-1}(\alpha)$ où ϕ est la fonction de répartition de la loi normale centrée réduite et $t_{\nu,\alpha} = \phi_\nu^{-1}(\alpha)$ où ϕ_ν est la fonction de répartition de la loi de Student à ν degrés de liberté.

Tests d'inférence d'une espérance (rappel)

On considère n variables aléatoires indépendantes X_1, X_2, \dots, X_n qui suivent toute la même loi normale $\mathcal{N}(\mu, \sigma^2)$.

H_0	Statistique de test	H_1	Région de rejet
$\mu = \mu_0$	$Z = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$ σ connu	$\mu \neq \mu_0$ $\mu < \mu_0$ $\mu > \mu_0$	$Z \leq z_{\frac{\alpha}{2}}$ ou $Z \geq z_{1-\frac{\alpha}{2}}$ $Z \leq z_\alpha$ $Z \geq z_{1-\alpha}$
$\mu = \mu_0$	$T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}}$ σ inconnu	$\mu \neq \mu_0$ $\mu < \mu_0$ $\mu > \mu_0$	$T \leq t_{n-1, \frac{\alpha}{2}}$ ou $T \geq t_{n-1, 1-\frac{\alpha}{2}}$ $T \leq t_{n-1, \alpha}$ $T \geq t_{n-1, 1-\alpha}$

Tests de comparaison de deux espérances

On considère m variables aléatoires indépendantes X_1, X_2, \dots, X_m qui suivent toute la même loi normale $\mathcal{N}(\mu_X, \sigma_X^2)$ et n variables aléatoires indépendantes Y_1, Y_2, \dots, Y_n qui suivent toute la même loi normale $\mathcal{N}(\mu_Y, \sigma_Y^2)$. On suppose en outre que ces familles de variables sont indépendantes.

H_0	Statistique de test	H_1	Région de rejet
$\mu_X - \mu_Y = d_0$	$Z = \frac{(\bar{X} - \bar{Y}) - d_0}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}}$ σ_X et σ_Y connus	$\mu_X - \mu_Y \neq d_0$ $\mu_X - \mu_Y < d_0$ $\mu_X - \mu_Y > d_0$	$Z \leq z_{\frac{\alpha}{2}}$ ou $Z \geq z_{1-\frac{\alpha}{2}}$ $Z \leq z_\alpha$ $Z \geq z_{1-\alpha}$
$\mu_X - \mu_Y = d_0$	$T = \frac{(\bar{X} - \bar{Y}) - d_0}{\sqrt{\frac{S_p^2}{m} + \frac{S_p^2}{n}}}$ $\sigma_X = \sigma_Y$ inconnus $S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$	$\mu_X - \mu_Y \neq d_0$ $\mu_X - \mu_Y < d_0$ $\mu_X - \mu_Y > d_0$	$T \leq t_{m+n-2, \frac{\alpha}{2}}$ ou $T \geq t_{m+n-2, 1-\frac{\alpha}{2}}$ $T \leq t_{m+n-2, \alpha}$ $T \geq t_{m+n-2, 1-\alpha}$
$\mu_X - \mu_Y = d_0$	$T = \frac{(\bar{X} - \bar{Y}) - d_0}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}}$ $\sigma_X \neq \sigma_Y$ inconnus $\hat{\nu} = \frac{\left(\frac{S_X^2}{m} + \frac{S_Y^2}{n}\right)^2}{\frac{\left(\frac{S_X^2}{m}\right)^2}{m-1} + \frac{\left(\frac{S_Y^2}{n}\right)^2}{n-1}}$	$\mu_X - \mu_Y \neq d_0$ $\mu_X - \mu_Y < d_0$ $\mu_X - \mu_Y > d_0$	$T \leq t_{\hat{\nu}, \frac{\alpha}{2}}$ ou $T \geq t_{\hat{\nu}, 1-\frac{\alpha}{2}}$ $T \leq t_{\hat{\nu}, \alpha}$ $T \geq t_{\hat{\nu}, 1-\alpha}$

Tests de comparaison de deux espérances d'échantillons appariés

On considère n variables aléatoires indépendantes X_1, X_2, \dots, X_n qui suivent toute la même loi normale $\mathcal{N}(\mu_X, \sigma_X^2)$ et n variables aléatoires indépendantes Y_1, Y_2, \dots, Y_n qui suivent toute la même loi normale $\mathcal{N}(\mu_Y, \sigma_Y^2)$. On suppose de plus que pour chaque i , il y a une dépendance entre X_i et Y_i : on dit que X_i et Y_i sont appariés.

On sait (c'est un théorème) que les variables aléatoires $D_i = Y_i - X_i$ suivent aussi une loi normale $\mathcal{N}(\mu_D, \sigma_D^2)$ où $\mu_D = \mu_Y - \mu_X$. On est ainsi ramené au test d'inférence de l'espérance de D_i qui se trouve sur la page précédente (celui où la variance n'est pas supposée connue).

H_0	Statistique de test	H_1	Région de rejet
$\underbrace{\mu_Y - \mu_X}_{=\mu_D} = d_0$	$Z = \frac{\bar{D} - d_0}{\sqrt{\frac{\sigma_D^2}{n}}}$ σ_D connu	$\mu_Y - \mu_X \neq d_0$ $\mu_Y - \mu_X < d_0$ $\mu_Y - \mu_X > d_0$	$Z \leq z_{\frac{\alpha}{2}}$ ou $Z \geq z_{1-\frac{\alpha}{2}}$ $Z \leq z_\alpha$ $Z \geq z_{1-\alpha}$
$\underbrace{\mu_Y - \mu_X}_{=\mu_D} = d_0$	$T = \frac{\bar{D} - d_0}{\sqrt{\frac{S_D^2}{n}}}$ σ_D inconnu	$\mu_Y - \mu_X \neq d_0$ $\mu_Y - \mu_X < d_0$ $\mu_Y - \mu_X > d_0$	$T \leq t_{n-1, \frac{\alpha}{2}}$ ou $T \geq t_{n-1, 1-\frac{\alpha}{2}}$ $T \leq t_{n-1, \alpha}$ $T \geq t_{n-1, 1-\alpha}$

3.2 Test du chi-carré : adéquation à une loi

Considérons des variables aléatoires X_1, \dots, X_n indépendantes qui suivent une même loi qu'une variable aléatoire X . On cherche à tester la densité de probabilité (ou distribution) de X .

On découpe les valeurs possibles pour X en k classes A_1, \dots, A_k . On note N_i les variables aléatoires qui comptent le nombre de mesures qui tombent dans la classe A_i .

tableau des probabilités théoriques

A_1	\dots	A_k	total
$\mathbb{P}(X \in A_1)$	\dots	$\mathbb{P}(X \in A_k)$	1

Dans ce modèle de test d'hypothèses, on teste l'hypothèse «la variable X a une certaine densité de probabilité» à partir des n mesures effectuées. À l'aide du découpage en classes, cette hypothèse est reformulée par «les probabilités théoriques $\mathbb{P}(X \in A_i)$ valent p_i ».

On se trouve face à deux alternatives qui sont

$$H_0 : \mathbb{P}(X \in A_i) = p_i \text{ pour tout } i \quad \text{et} \quad H_1 : \text{il existe } i \text{ tel que } \mathbb{P}(X \in A_i) \neq p_i$$

Comme dans un raisonnement par l'absurde, on suppose qu'on se trouve sous l'hypothèse H_0 et on regarde si les données mesurées permettent d'en tirer une contradiction.

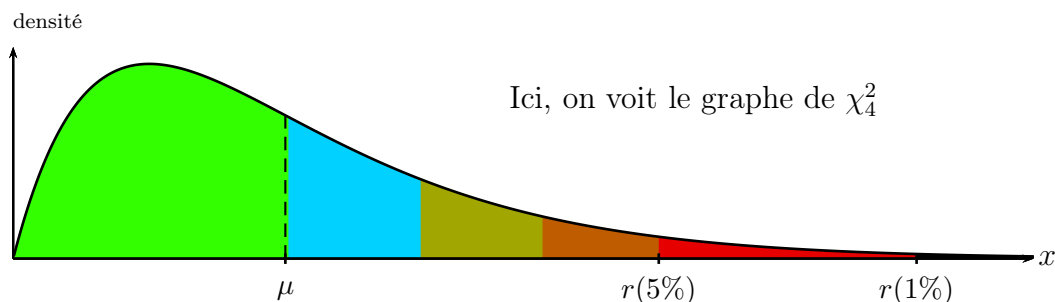
Sous l'hypothèse H_0 , on peut établir le tableau des effectifs théoriques.

effectifs mesurés				effectifs théoriques			
A_1	\dots	A_k	total	A_1	\dots	A_k	total
n_1	\dots	n_k	n	np_1	\dots	np_k	n







Les mathématiciens ont montré que si les effectifs théoriques np_i sont de taille au moins 5, alors la variable aléatoire D^2 suivante suit approximativement une loi du chi-carré à $k - 1$ degrés de liberté.







$$D^2 = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i}$$

On a $k - 1$ degrés de liberté, car les variables aléatoires N_1 à N_k sont liées entre-elles par la condition : $\sum_i N_i = n$.



On estime D^2 à l'aide des mesures effectuées $d^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$ (attention à bien faire la différence entre les majuscules et les minuscules). L'estimation d^2 donnée par les mesures va donc se trouver quelque part sous la loi.

Il y a une probabilité de 60% que d^2 tombe dans la zone , il y a une probabilité de 20% que d^2 tombe dans la zone , il y a une probabilité de 10% que d^2 tombe dans la zone , il y a une probabilité de 5% que d^2 tombe dans la zone , il y a une probabilité de 4% que d^2 tombe dans la zone , il y a une probabilité de 1% que d^2 tombe dans la zone .

Plus on se trouve dans une zone éloignée de 0 donc dans l'ordre : , , , , , ; plus les mesures sont en contradiction avec l'hypothèse H_0 . On décide ainsi du critère suivant.

On rejette l'hypothèse H_0 au seuil de signification 5% si d^2 se trouve dans la zone  ou .

On rejette l'hypothèse H_0 au seuil de signification 1% si d^2 se trouve dans la zone .

On rejette H_0 au seuil de signification α si

$$d^2 > \phi_{k-1}^{-1}(1 - \alpha)$$

où $\phi_{k-1}(x) = \mathbb{P}(D^2 \leq x)$ est la fonction de répartition de la variable aléatoire D^2 qui suit la loi du chi-carré χ_{k-1}^2 avec $k - 1$ degrés de liberté. Les valeurs de cette fonction ϕ_{k-1} sont données par la table de la page 32.

En particulier, si $\alpha = 5\%$, le critère est $d^2 > \phi_{k-1}^{-1}(0.95) \stackrel{\text{table}}{=} 11.07$ si $k = 6$.

De même, si $\alpha = 1\%$, le critère est $d^2 > \phi_{k-1}^{-1}(0.99) \stackrel{\text{table}}{=} 15.09$ si $k = 6$.

Remarque sur les degrés de liberté

Si lors du calcul des probabilités p_i , on doit utiliser k estimateurs, il faut encore enlever k degrés de liberté.

Ce peut être le cas lorsque, par exemple, on calcule les p_i avec une loi normale où il faut d'abord estimer l'espérance et la variance.

Remarque sur le risque de deuxième espèce

Dans le cas d'un test du χ^2 , le risque de deuxième espèce ne se calcule pas aussi facilement que dans le cas d'un test sur une moyenne. Cela dépasse le cadre de ce cours.

Exemples

Reprenons l'exemple du début du cours où on a lancé un dé 60 fois. Les classes choisies sont $A_i = \{i\}$ de sorte que A_i corresponde à l'apparition de la face i .

1. Testons l'alternative suivante.

$$\begin{aligned} H_0 &: \text{le dé est bien équilibré, c'est-à-dire } p_i = \frac{1}{6}. \\ H_1 &: \text{le dé n'est pas bien équilibré.} \end{aligned}$$

On peut donc construire les tableaux suivants.

effectifs mesurés							effectifs théoriques								
face	1	2	3	4	5	6	total	face	1	2	3	4	5	6	total
n_i	13	13	11	8	8	7	60	np_i	10	10	10	10	10	10	60

Pour que l'approximation soit bonne, il faut que $np_i \geq 5$ pour chaque i . C'est le cas ici. On a

$$d^2 = \sum_{i=1}^6 \frac{(n_i - np_i)^2}{np_i} = \underbrace{\frac{(13 - 10)^2}{10}}_{\text{pour la face 1}} + \underbrace{\frac{(13 - 10)^2}{10}}_{\text{pour la face 2}} + \dots + \underbrace{\frac{(7 - 10)^2}{10}}_{\text{pour la face 6}} = 3.6$$

Comme $d^2 < \phi_5^{-1}(0.95) \cong 11.07$, on ne peut pas rejeter l'hypothèse H_0 au seuil 5%. Il n'y a pas assez de preuves pour dire que le dé n'est pas bien équilibré.

2. Imaginons qu'on pense qu'il y a une bille de plomb vers le sommet adjacent aux faces 4, 5 et 6 qui fait que le dé montre pour le 75% des lancers les faces 1, 2 ou 3.

On en déduit que $p_1 = p_2 = p_3 = \frac{3}{12}$ et que $p_4 = p_5 = p_6 = \frac{1}{12}$ (le total des p_i vaut bien 1, et $p_1 + p_2 + p_3 = 3(p_4 + p_5 + p_6)$ tout comme 75% vaut $3 \cdot 25\%$).

Testons l'alternative suivante.

$$\begin{aligned} H_0 &: \text{le dé est truqué comme ci-dessus.} \\ H_1 &: \text{le dé n'est pas truqué comme ci-dessus.} \end{aligned}$$

On peut donc construire les tableaux suivants.

effectifs mesurés							effectifs théoriques								
face	1	2	3	4	5	6	total	face	1	2	3	4	5	6	total
n_i	13	13	11	8	8	7	60	np_i	15	15	15	5	5	5	60

Pour que l'approximation soit bonne, il faut que $np_i \geq 5$ pour chaque i . C'est le cas ici. On a

$$d^2 = \sum_{i=1}^6 \frac{(n_i - np_i)^2}{np_i} = \underbrace{\frac{(13 - 15)^2}{15}}_{\text{pour la face 1}} + \underbrace{\frac{(13 - 15)^2}{15}}_{\text{pour la face 2}} + \dots + \underbrace{\frac{(7 - 5)^2}{5}}_{\text{pour la face 6}} = 6$$

Comme $d^2 < \phi_5^{-1}(0.95) \cong 11.07$, on ne peut pas rejeter l'hypothèse H_0 au seuil 5%. Il n'y a pas assez de preuves pour dire que le dé n'est pas truqué de cette façon.

Les résultats des 60 jets ne sont pas assez différents de ce qu'on aurait pu obtenir avec un dé bien équilibré ou un dé truqué de la façon ci-dessus. La valeur mesurée d^2 indiquerait que la situation la plus probable est que le dé est bien équilibré (la p -valeur est meilleure sur le premier exemple ($\cong 60.8\%$) que sur le deuxième ($\cong 30.6\%$)).

Néanmoins, le mieux serait de lancer ce dé encore 60 fois et de refaire ces tests sur les 120 lancers obtenus.

3.3 Test du chi-carré : comparaison d'échantillons

Pour chaque échantillon i , $1 \leq i \leq k$, considérons des variables aléatoires $X_1^{(i)}, \dots, X_{n^{(i)}}^{(i)}$ indépendantes qui suivent une même loi qu'une variable aléatoire $X^{(i)}$. On cherche à tester si les lois $X^{(i)}$ suivent toutes une même loi qu'une variable aléatoire X .

On découpe les valeurs possibles pour les échantillons en l classes A_1, \dots, A_l (les mêmes pour chaque échantillon). On note $N_j^{(i)}$ les variables aléatoires qui comptent le nombre de mesures de l'échantillon i qui tombent dans la classe A_j . On note aussi $N_j = \sum_i N_j^{(i)}$.

tableau des probabilités théoriques

	A_1	\dots	A_l	totaux
échantillon 1	$\mathbf{P}(X^{(1)} \in A_1)$	\dots	$\mathbf{P}(X^{(1)} \in A_l)$	1
\vdots	\vdots		\vdots	\vdots
échantillon k	$\mathbf{P}(X^{(k)} \in A_1)$	\dots	$\mathbf{P}(X^{(k)} \in A_l)$	1
moyenne	$\frac{1}{k} \sum_i \mathbf{P}(X^{(i)} \in A_1)$	\dots	$\frac{1}{k} \sum_i \mathbf{P}(X^{(i)} \in A_l)$	1

Dans ce modèle de tests d'hypothèses, on teste l'hypothèse «les échantillons suivent tous la même loi» à partir des mesures effectuées sur chaque échantillon. Cette hypothèse est reformulée par «les $X^{(i)}$ suivent une même loi X pour tout i ».

On se trouve face à deux alternatives qui sont

$$H_0 : \mathbf{P}(X^{(i)} \in A_j) = \mathbf{P}(X \in A_j) \text{ pour tout } i$$

$$H_1 : \text{il existe } i \text{ tel que } \mathbf{P}(X^{(i)} \in A_j) \neq \mathbf{P}(X \in A_j)$$

Comme dans un raisonnement par l'absurde, on suppose qu'on se trouve sous l'hypothèse H_0 et on regarde si les données mesurées permettent d'en tirer une contradiction.

Sous l'hypothèse H_0 , les moyennes ci-dessus deviennent

$$\frac{1}{k} \sum_i \mathbf{P}(X^{(i)} \in A_j) = \frac{1}{k} \sum_{i=1}^k \mathbf{P}(X \in A_j) = \mathbf{P}(X \in A_j)$$

et on peut établir le tableau des effectifs théoriques en notant pour simplifier n pour le nombre total de mesures $n = \sum_i n^{(i)}$ et $p_j = \mathbf{P}(X \in A_j)$.

effectifs mesurés

	A_1	\dots	A_l	totaux
éch. 1	$n_1^{(1)}$	\dots	$n_l^{(1)}$	$n^{(1)}$
\vdots	\vdots	$n_j^{(i)}$	\vdots	\vdots
éch. k	$n_1^{(k)}$	\dots	$n_l^{(k)}$	$n^{(k)}$
totaux	n_1	\dots	n_l	n

effectifs théoriques

	A_1	\dots	A_l	totaux
éch. 1	$n^{(1)}p_1$	\dots	$n^{(1)}p_l$	$n^{(1)}$
\vdots	\vdots	$n^{(i)}p_j$	\vdots	\vdots
éch. k	$n^{(k)}p_1$	\dots	$n^{(k)}p_l$	$n^{(k)}$
totaux	np_1	\dots	np_l	n

Les mathématiciens ont montré que si les effectifs théoriques $n^{(i)}p_j$ sont de taille au moins 5, alors la variable aléatoire D^2 suivante suit approximativement une loi du χ^2 à $kl - k$ degrés de liberté (on enlève à kl un degré par échantillon, car les variables aléatoires $N_1^{(i)}$ à $N_l^{(i)}$ sont liées entre-elles par la condition : $\sum_j N_j^{(i)} = n^{(i)}$).

$$D^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(N_j^{(i)} - n^{(i)}p_j)^2}{n^{(i)}p_j}$$

Lorsque que les paramètres p_j sont inconnus : on les estime⁴ par $p_j = \frac{n_j}{n}$. Ces estimateurs sont naturels, ils font en sorte que les totaux des deux tableaux soient identiques.

Dans ce cas, il faut recalculer les degrés de liberté, il faut encore enlever $(l - 1)$ degrés de liberté (les estimateurs p_1 à p_l sont liés entre-eux par la condition $\sum_j p_j = 1$). Ainsi le nombre de degré de liberté vaut $kl - k - (l - 1) = kl - k - l + 1 = k(l - 1) - (l - 1) = (k - 1)(l - 1)$.

On estime D^2 à l'aide des mesures effectuées $d^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_j^{(i)} - n^{(i)} p_j)^2}{n^{(i)} p_j}$ (attention à bien faire la différence entre les majuscules et les minuscules). L'estimation d^2 donnée par les mesures va donc se trouver quelque part sous la loi.

L'allure de la loi de probabilité et le critère de rejet sont les mêmes que pour le test d'adéquation (attention aux degrés de liberté qui changent).

Exemple

Cet exemple est basé sur les relevés concernant les déchets urbains du Jura en 2010, 2000 et 1994. Les déchets sont subdivisés en cinq catégories : les déchets urbains combustibles ; les déchets compostables ; le papier et le carton ; le verre ; l'aluminium, le fer blanc, la ferraille.

Ici les échantillons correspondent aux trois années susmentionnées, et on va tester si la façon dont les types de déchets sont répartis évolue avec les années (si c'est le cas, on peut éventuellement affirmer que les habitants ont pris conscience de l'importance de trier les déchets).

H_0 : la répartition des déchets dans les cinq catégories est la même pour les trois ans.

H_1 : il y a deux ans au moins pour lesquels les déchets sont répartis différemment.

Voici les tableaux des effectifs mesurés et celui des effectifs théoriques (calculés à partir de ces mesures sous l'hypothèse H_0). Les unités sont en kilogrammes par habitant et par année. Les erreurs d'arrondis ont été lissées (pour que les totaux jouent).

effectifs mesurés							effectifs théoriques						
	comb.	comp.	papier	verre	fer	totaux		comb.	comp.	papier	verre	fer	totaux
1994	284	19	28	33	6	370	1994	222	58	43	35	12	370
2000	265	75	52	47	23	462	2000	277	73	54	44	14	462
2010	252	117	75	48	13	505	2010	302	80	58	49	16	505
totaux	801	211	155	128	42	1337	totaux	801	211	155	128	42	1337

Pour que l'approximation par la loi du chi-carré soit bonne, il faut que les effectifs théoriques soient plus grands ou égaux à 5. C'est le cas ici, et on a

$$d^2 = \sum_{i=1}^3 \sum_{j=1}^5 \frac{(n_{i,j} - np_i q_j)^2}{np_i q_j} = \frac{(284 - 222)^2}{222} + \frac{(19 - 58)^2}{58} + \dots + \frac{(13 - 16)^2}{16} \cong 89$$

(sans arrondir les effectifs théoriques, on trouve environ 88.952)

Les paramètres p_j ont été estimés, on se retrouve avec $(3 - 1)(5 - 1) = 8$ degrés de liberté.

Comme $d^2 > \phi_2^{-1}(0.99) \cong 20.090$, on rejette l'hypothèse H_0 au seuil de signification de 1% (c'est-à-dire avec une probabilité de 1% de chance de rejeter à tort).

Les données laissent à penser que les citoyens trient de mieux en mieux leurs déchets.

4. Il s'agit des estimateurs du maximum de vraisemblance du cours OS, notés $\hat{p}_{j\text{MLE}}$.

3.4 Test du chi-carré : indépendance

Considérons des couples de variables aléatoires $(X_1; Y_1), \dots, (X_n; Y_n)$ indépendantes qui suivent les mêmes lois que le couple de variables aléatoires $(X; Y)$. On cherche à tester l'indépendance entre X et Y .

On découpe les valeurs possibles pour X en k classes A_1, \dots, A_k , et celles pour Y en l classes B_1, \dots, B_l . On note $N_{i;j}$ les variables aléatoires qui comptent le nombre de couples de mesures qui tombent dans $(A_i; B_j)$. On note aussi $N_{i;\heartsuit} = \sum_j N_{i;j}$ et $N_{\heartsuit;j} = \sum_i N_{i;j}$.

tableau des probabilités théoriques

	B_1	\dots	B_l	totaux
A_1	$\mathbb{P}((X \in A_1) \cap (Y \in B_1))$	\dots	$\mathbb{P}((X \in A_1) \cap (Y \in B_l))$	$\mathbb{P}(X \in A_1)$
\vdots	\vdots		\vdots	\vdots
A_k	$\mathbb{P}((X \in A_k) \cap (Y \in B_1))$	\dots	$\mathbb{P}((X \in A_k) \cap (Y \in B_l))$	$\mathbb{P}(X \in A_k)$
totaux	$\mathbb{P}(Y \in B_1)$	\dots	$\mathbb{P}(Y \in B_l)$	1

Dans ce modèle de tests d'hypothèses, on teste l'hypothèse «les variables X et Y sont indépendantes» à partir des n mesures effectuées. En utilisant le découpage en classes et la théorie des probabilités, cette hypothèse est reformulée par «les probabilités théoriques satisfont $\mathbb{P}((X \in A_i) \cap (Y \in B_j)) = \mathbb{P}(X \in A_i) \cdot \mathbb{P}(Y \in B_j)$ ».

On se trouve face à deux alternatives qui sont

$$H_0 : \mathbb{P}((X \in A_i) \cap (Y \in B_j)) = \mathbb{P}(X \in A_i) \cdot \mathbb{P}(Y \in B_j) \text{ pour tout } i \text{ et } j$$

$$H_1 : \text{il existe } i \text{ et } j \text{ tels que } \mathbb{P}((X \in A_i) \cap (Y \in B_j)) \neq \mathbb{P}(X \in A_i) \cdot \mathbb{P}(Y \in B_j)$$

Comme dans un raisonnement par l'absurde, on suppose qu'on se trouve sous l'hypothèse H_0 et on regarde si les données mesurées permettent d'en tirer une contradiction.

Sous l'hypothèse H_0 , on peut établir le tableau des effectifs théoriques en notant pour simplifier $p_i = \mathbb{P}(X \in A_i)$ et $q_j = \mathbb{P}(Y \in B_j)$.

effectifs mesurés

	B_1	\dots	B_l	totaux
A_1	$n_{1;1}$	\dots	$n_{1;l}$	$n_{1;\heartsuit}$
\vdots	\vdots	$n_{i;j}$	\vdots	\vdots
A_k	$n_{k;1}$	\dots	$n_{k;l}$	$n_{k;\heartsuit}$
totaux	$n_{\heartsuit;1}$	\dots	$n_{\heartsuit;l}$	n

effectifs théoriques

	B_1	\dots	B_l	totaux
A_1	np_1q_1	\dots	np_1q_l	np_1
\vdots	\vdots	np_iq_j	\vdots	\vdots
A_k	np_kq_1	\dots	np_kq_l	np_k
totaux	nq_1	\dots	nq_l	n

Les mathématiciens ont montré que si les effectifs théoriques np_iq_j sont de taille au moins 5, alors la variable aléatoire D^2 suivante suit approximativement une loi du χ^2 à $kl - 1$ degrés de liberté (on enlève à kl un degré de liberté, car les variables aléatoires $N_{i;j}$ sont liées entre-elles par la condition : $\sum_{i,j} N_{i;j} = n$).

$$D^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(N_{i;j} - np_iq_j)^2}{np_iq_j}$$

Lorsque que les paramètres p_i et q_j sont inconnus : on les estime⁵ par $p_i = \frac{n_{i\heartsuit}}{n}$ et $q_j = \frac{n_{\heartsuit j}}{n}$. Ces estimateurs sont naturels, ils font en sorte que les totaux des deux tableaux soient identiques.

Dans ce cas, il faut recalculer les degrés de liberté, il faut encore enlever $(k-1) + (l-1)$ degrés de liberté (les estimateurs p_i et q_j sont liés entre-eux par les conditions $\sum_i p_i = 1$ et $\sum_j q_j = 1$). Ainsi le nombre de degré de liberté vaut $kl - 1 - (l-1) - (k-1) = kl - k - l + 1 = k(l-1) - (l-1) = (k-1)(l-1)$.

On estime D^2 à l'aide des mesures effectuées $d^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{i,j} - np_i q_j)^2}{np_i q_j}$ (attention à bien faire la différence entre les majuscules et les minuscules). L'estimation d^2 donnée par les mesures va donc se trouver quelque part sous la loi.

L'allure de la loi de probabilité et le critère de rejet sont les mêmes que pour le test d'adéquation (attention aux degrés de liberté qui changent).

Exemple

Cet exemple est tiré du livre «Statistics» (2e édition) écrit par Freeman, Pisani, Purves et Adhikari, aux éditions Norton, international student edition.

Une étude basée sur 2237 américains âgés de 25 à 34 a permis de montrer que les femmes sont plus souvent droitières que les hommes.

L'étude consistait à faire un test d'hypothèse sur l'indépendance entre le sexe d'une personne et le fait qu'elle soit droitière ou gauchère (ou ambidextre).

On teste l'alternative suivante.

H_0 : le sexe d'une personne est indépendant du fait qu'elle soit gauchère ou droitière.

H_1 : il n'y a pas indépendance.

Voici les données observées (effectifs mesurés) et les effectifs théoriques (calculés à partir de ces mesures sous l'hypothèse H_0).

effectifs mesurés				effectifs théoriques			
	femmes	hommes	totaux		femmes	hommes	totaux
droitiers	1070	934	2004	droitiers	1048	956	2004
gauchers	92	113	205	gauchers	107	98	205
ambidextres	8	20	28	ambidextres	15	13	28
totaux	1170	1067	2237	totaux	1170	1067	2237

Pour que l'approximation par la loi du chi-carré soit bonne, il faut que les effectifs théoriques soient plus grands ou égaux à 5. C'est le cas ici, et on a

$$d^2 = \sum_{i=1}^3 \sum_{j=1}^2 \frac{(n_{i,j} - np_i q_j)^2}{np_i q_j} = \frac{(1070 - 1048)^2}{1048} + \frac{(934 - 956)^2}{956} + \dots + \frac{(20 - 13)^2}{13} \cong 12$$

(sans arrondir les effectifs théoriques, on trouve environ 11.806)

Les paramètres p_i et q_j ont été estimés, on se retrouve avec $(3-1)(2-1) = 2$ degrés de liberté.

Comme $d^2 > \phi_2^{-1}(0.99) \cong 9.210$, on rejette l'hypothèse H_0 au seuil de signification de 1% (c'est-à-dire avec une probabilité de 1% de chance de rejeter à tort).

Les données laissent penser que les femmes sont plus souvent droitières que les hommes.

5. Il s'agit des estimateurs du maximum de vraisemblance du cours OS, notés $\hat{p}_{i\text{MLE}}$ et $\hat{q}_{j\text{MLE}}$.

3.5 La p -valeur associée à un test d'hypothèse

La p -valeur est la probabilité sous l'hypothèse nulle H_0 que la statistique de test⁶ soit au moins aussi "extrême" que la valeur observée à partir des données.

La signification du mot "extrême" dépend de la façon dont la probabilité indiquée par le seuil de signification α est définie dans le test d'hypothèse en question.

Exemple d'un test bilatéral avec une distribution symétrique

Rappelons que les distributions des lois normales, de Student et binomiales sont des distributions symétriques.

Dans le cas des tests bilatéraux où la statistique de test T possède une distribution symétrique, le seuil de signification α satisfait la condition suivante :

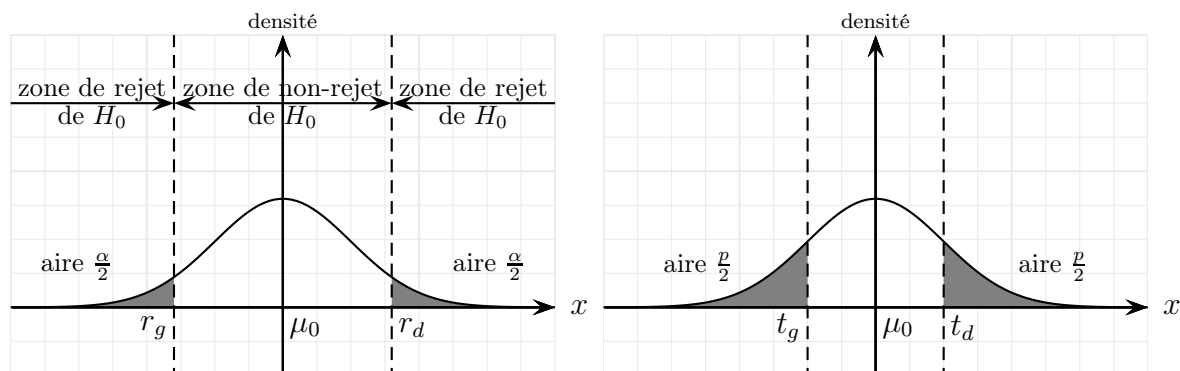
$$\alpha = \underbrace{\mathbb{P}(T < r_g)}_{\frac{\alpha}{2}} + \underbrace{\mathbb{P}(T > r_d)}_{\frac{\alpha}{2}} \quad \text{avec} \quad r_g = F^{-1}\left(\frac{\alpha}{2}\right) \quad \text{et} \quad r_d = F^{-1}\left(1 - \frac{\alpha}{2}\right)$$

où $F(x)$ est la fonction de répartition associée à la statistique de test T . Les nombres r_g et r_d sont les limites de la zone de rejet de H_0 .

De façon similaire, on définit la p -valeur par
$$p = \underbrace{\mathbb{P}(T < t_g)}_{\frac{p}{2}} + \underbrace{\mathbb{P}(T > t_d)}_{\frac{p}{2}}$$

où t_g est la valeur à gauche qui est obtenue par symétrie de t (observation de la statistique de test T) par rapport à l'axe de symétrie de la distribution de T et où t_d est la valeur symétrique de t à droite (l'une des deux est égale à t !).

Description graphique d'un tel test d'hypothèse (pour cette situation, H_0 ne peut pas être rejetée au seuil de signification α).



Utilité de la p -valeur

Par construction de la p -valeur, on rejette H_0 au seuil de signification α lorsque $p < \alpha$. En pratique, on peut aussi utiliser le tableau suivant :

p -valeur	évidence contre H_0	seuil de signification associé
$p\text{-valeur} \geq 0.10$	négligeable	$10\% < \alpha$
$0.10 > p\text{-valeur} \geq 0.05$	faible	$5\% < \alpha \leq 10\%$
$0.05 > p\text{-valeur} \geq 0.01$	modérée	$1\% < \alpha \leq 5\%$
$0.01 > p\text{-valeur} \geq 0.001$	forte	$0.1\% < \alpha \leq 1\%$
$0.001 > p\text{-valeur}$	très forte	$\alpha \leq 0.1\%$

6. Les *statistiques de test* sont les variables aléatoires utilisées pour les tests d'hypothèses.

Chapitre 4

Intervalles de confiance

Dans une étude statistique, on peut se demander si les mesures observées peuvent correspondre à une certaine réalité. Par exemple, est-ce que notre série de jets d'une pièce de monnaie ($P; P; F; P; F; P$) peut correspondre à une pièce bien équilibrée?

Pour répondre à cette question, on cherche à estimer la probabilité p que la pièce montre pile. Si on arrive à établir que $p = \frac{1}{2}$, alors la pièce est parfaitement équilibrée. Malheureusement, il est extrêmement difficile, voire impossible, de montrer que $p = \frac{1}{2}$.

On peut tenter d'estimer p à l'aide d'un estimateur \hat{P} , mais la probabilité $\mathbf{P}(\hat{P} = p)$ que \hat{P} soit exactement égal au paramètre p est, en général, extrêmement petite, voire nulle¹. Par conséquent, il est nécessaire de s'intéresser à la probabilité

$$\mathbf{P}(\hat{P} \geq p - r \text{ et } \hat{P} \leq p + r) = \mathbf{P}(\hat{P} - r \leq p \leq \hat{P} + r) = \mathbf{P}([\hat{P} - r, \hat{P} + r] \ni p)$$

Cette probabilité s'approche de 1 au fur et à mesure que le nombre positif r devient grand. Néanmoins, plus r grandit, plus l'approximation de p par un intervalle perd en précision. Il faut donc trouver des compromis.

Définition

Un *intervalle de confiance* est un intervalle $[G, D]$ qui contient le paramètre θ à estimer avec une certaine probabilité β (G et D sont des variables aléatoires).

$$\mathbf{P}([G, D] \ni \theta) = \beta$$

On dit que β est le *seuil de confiance* ou encore la *probabilité de couverture*, tandis que $\alpha = 1 - \beta$ est appelé le *risque d'erreur*.

Construction d'un intervalle de confiance

Pour construire un intervalle de confiance pour le paramètre θ , on commence par choisir le seuil de confiance β désiré (ce qui est équivalent à choisir un risque d'erreur $\alpha = 1 - \beta$). Ensuite, il faut réussir à déterminer le nombre positif r correspondant en utilisant la distribution théorique de l'estimateur $\hat{\theta}$.

Comme d'habitude, on choisit un seuil de confiance de 95% dans une démarche infirmative, et un seuil de confiance de 99% dans une démarche confirmative.

1. De plus, la probabilité pour que deux estimations \hat{p}_1 et \hat{p}_2 de l'estimateur \hat{P} soient égales est aussi, en général, extrêmement faible.

4.1 L'intervalle de confiance sur une moyenne, variance connue

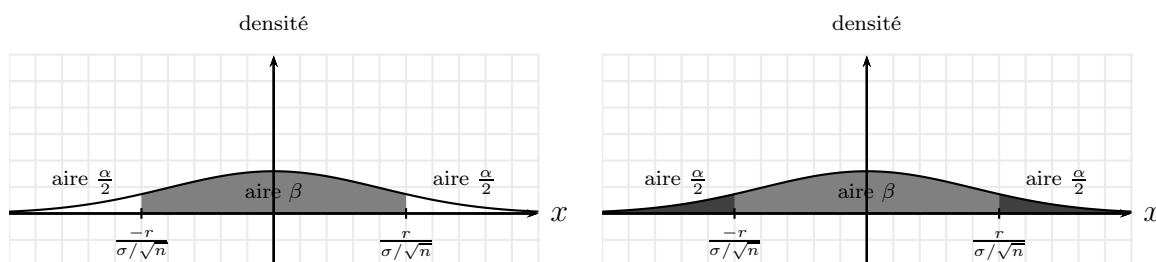
On considère n variables aléatoires indépendantes, notées X_1, X_2, \dots, X_n , qui suivent toutes une loi normale $\mathcal{N}(\mu, \sigma^2)$ où seul le paramètre σ est connu. Pour estimer μ , on prend le MLE \bar{X} qui est aussi sans biais. Sous ces hypothèses, \bar{X} suit une loi normale de paramètres $\mathcal{N}(\mu, \frac{\sigma^2}{n})$. Afin de déterminer l'intervalle de confiance, on va devoir centrer-réduire cet estimateur de sorte à pouvoir utiliser la table concernant la fonction de répartition ϕ de la loi normale $\mathcal{N}(0, 1)$. On trouvera l'intervalle de confiance suivant :

$$\left[\bar{X} - \phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + \phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}} \right]$$

En effet, on cherche $r > 0$ tel que :

$$\begin{aligned} 1 - \alpha &= \beta = \mathbf{P}([\bar{X} - r, \bar{X} + r] \ni \mu) = \mathbf{P}(\bar{X} - r \leq \mu \leq \bar{X} + r) \\ &= \mathbf{P}(\bar{X} \geq \mu - r \text{ et } \bar{X} \leq \mu + r) = \mathbf{P}(\mu - r \leq \bar{X} \leq \mu + r) \\ &= \mathbf{P}(-r \leq \bar{X} - \mu \leq r) = \mathbf{P}\left(\frac{-r}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{r}{\frac{\sigma}{\sqrt{n}}}\right) \end{aligned}$$

On s'est maintenant ramené à la loi normale centrée réduite Z .



On peut ensuite utiliser la fonction de répartition ϕ de la loi normale centrée réduite.

$$\begin{aligned} 1 - \alpha &= \mathbf{P}\left(\frac{-r}{\frac{\sigma}{\sqrt{n}}} \leq Z \leq \frac{r}{\frac{\sigma}{\sqrt{n}}}\right) = \mathbf{P}\left(Z \leq \frac{r}{\frac{\sigma}{\sqrt{n}}}\right) - \mathbf{P}\left(Z < \frac{-r}{\frac{\sigma}{\sqrt{n}}}\right) \\ &= \mathbf{P}\left(Z \leq \frac{r}{\frac{\sigma}{\sqrt{n}}}\right) - \underbrace{\left(1 - \mathbf{P}\left(Z \leq \frac{r}{\frac{\sigma}{\sqrt{n}}}\right)\right)}_{= 1 - \mathbf{P}\left(Z \leq \frac{r}{\frac{\sigma}{\sqrt{n}}}\right)} = 2\mathbf{P}\left(Z \leq \frac{r}{\frac{\sigma}{\sqrt{n}}}\right) - 1 \\ &= 2\phi\left(\frac{r}{\frac{\sigma}{\sqrt{n}}}\right) - 1 \quad \text{car } \mathbf{P}\left(Z < \frac{-r}{\frac{\sigma}{\sqrt{n}}}\right) = \mathbf{P}\left(Z > \frac{r}{\frac{\sigma}{\sqrt{n}}}\right) \text{ par symétrie} \end{aligned}$$

Finalement, on isole r :

$$\begin{aligned} 1 - \alpha &= 2\phi\left(\frac{r}{\frac{\sigma}{\sqrt{n}}}\right) - 1 \iff 2\phi\left(\frac{r}{\frac{\sigma}{\sqrt{n}}}\right) = 2 - \alpha \iff \phi\left(\frac{r}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \frac{\alpha}{2} \\ &\iff \frac{r}{\frac{\sigma}{\sqrt{n}}} = \phi^{-1}\left(1 - \frac{\alpha}{2}\right) \iff r = \phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}} \end{aligned}$$

4.2 L'intervalle de confiance sur une moyenne, variance inconnue

On considère n variables aléatoires indépendantes, notées X_1, X_2, \dots, X_n , qui suivent toutes une loi normale $\mathcal{N}(\mu, \sigma^2)$ où le paramètre σ est aussi inconnu. Pour estimer μ , on prend le MLE \bar{X} qui est aussi sans biais. Sous ces hypothèses, \bar{X} suit une loi normale de paramètres $\mathcal{N}(\mu, \frac{\sigma^2}{n})$.

Malheureusement, comme on ne connaît pas σ , ce n'est pas très utile. Mais, en utilisant l'estimateur sans biais de σ , noté S , on peut utiliser le résultat de William Gosset : la variable aléatoire² $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ suit une distribution de Student avec $n - 1$ degrés de liberté.

Afin de déterminer l'intervalle de confiance, on procède comme ci-dessus de sorte à pouvoir utiliser la table concernant la fonction de répartition ϕ_{n-1} de la loi de Student à $n - 1$ degrés de liberté.

On trouvera l'intervalle de confiance suivant :

$$\left[\bar{X} - \phi_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{S}{\sqrt{n}}, \bar{X} + \phi_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{S}{\sqrt{n}} \right]$$

2. Ce n'est pas la variable centrée réduite de \bar{X} , car on a remplacé σ par son estimateur sans biais S .

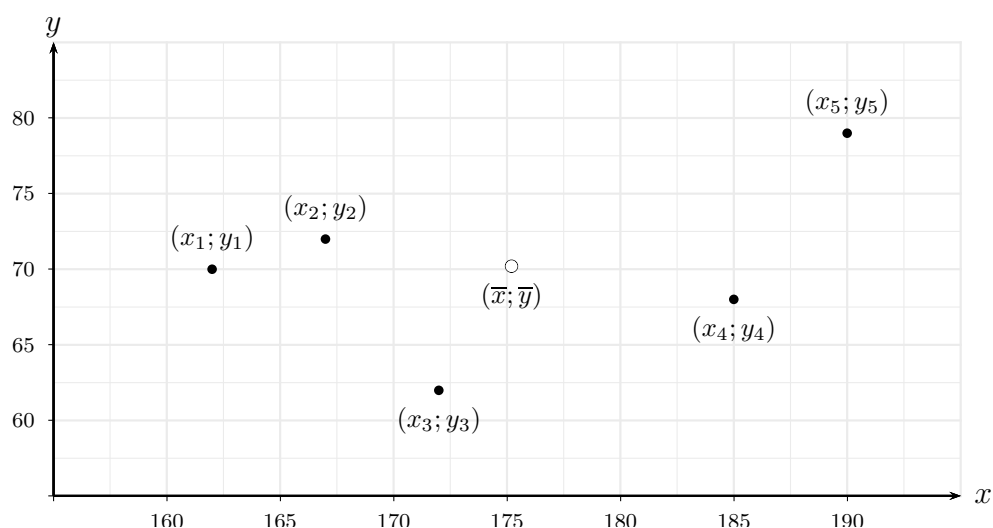
Chapitre 5

Régression linéaire

Supposons que l'on se donne deux caractéristiques X et Y sur une même population. Par exemple, on pourrait mesurer la taille en centimètres pour X et le poids en kilogrammes pour Y des lycéens de première année. Pour cela, on mesure les caractéristiques X et Y sur un échantillon aléatoire de taille n . On obtient ainsi des observations à deux coordonnées (x_i, y_i) pour $i \in \{1, 2, \dots, n\}$. Pour cet exemple, prenons $n = 5$.

élève	i	1	2	3	4	5	moyennes
taille en cm	x_i	162	167	172	185	190	$\bar{x} = 175.2$
poids en kg	y_i	70	72	62	68	79	$\bar{y} = 70.2$

La méthode la plus simple pour observer la relation entre X et Y est de représenter ces points dans le plan où l'axe horizontal représente la caractéristique X et l'axe vertical la caractéristique Y . Une telle représentation est appelée un *diagramme de dispersion*.



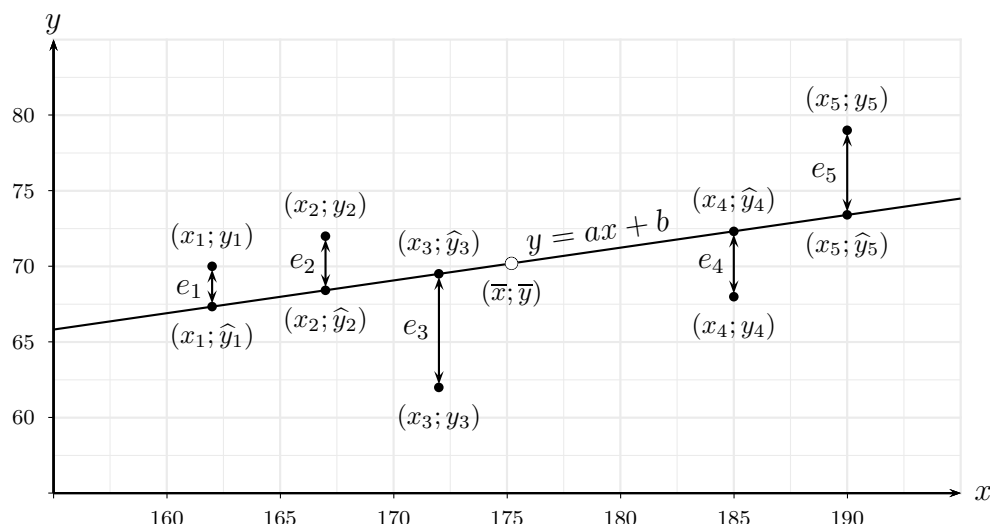
Si la relation entre X et Y est exacte, alors on devrait pouvoir trouver, pour une mesure x_i donnée, l'UNIQUE valeur pour y_i . Ainsi, Y serait une FONCTION de X ($y = f(x)$).

Malheureusement (ou heureusement), il se trouve que dans la plupart des cas, la relation n'est pas exacte (par exemple deux individus de même taille n'ont pas exactement le même poids). Néanmoins, même s'il n'y a pas de relation exacte, il se pourrait qu'il y ait une relation théorique et que, dans chaque mesure, il y ait une part aléatoire.

Dans un tel contexte, on dit que Y est la *variable dépendante*, et X est la *variable indépendante*.

Pour commencer, on va regarder s'il y a une chance pour que la relation (exacte ou non) entre X et Y soit affine (le graphe d'une fonction affine est une droite).

On va donc essayer de faire passer une droite "au mieux" parmi les points (x_i, y_i) .



Les notations de la régression linéaire

y_i	y_i est la mesure effective associée à x_i .
$y = ax + b$	Il s'agit du modèle affine théorique entre les caractéristiques X et Y . Il faut déterminer les valeurs des bons paramètres a et b .
$\hat{y}_i = ax_i + b$	\hat{y}_i est l'approximation théorique par le modèle affine associée à x_i .
$e_i = y_i - \hat{y}_i$	e_i est l'erreur entre la mesure effective y_i et son approximation théorique \hat{y}_i associée à la i -ième mesure. Les e_i sont appelés les <i>résidus</i> associés au modèle.

On a ainsi.

$$e_i = y_i - \hat{y}_i \iff y_i = \hat{y}_i + e_i \iff y_i = \overbrace{ax_i + b}^{\text{modèle linéaire}} + e_i$$

5.1 La droite des moindres carrés

Dans ce cas le modèle théorique $y = ax + b$ est construit de manière à ce que la somme des carrés des résidus soit la plus petite possible.

Autrement dit, on veut a et b tels que la somme $\sum_{i=1}^n e_i^2$ soit minimale.

Pourquoi les carrés

1. L'élevation au carré néglige les signes. Ainsi une erreur négative ne sera pas compensée par une erreur positive.
2. L'élevation au carré réduit les petits écarts (car $(\frac{1}{2})^2 = \frac{1}{4}$; $(\frac{1}{4})^2 = \frac{1}{16}$) et amplifie les grands écarts (car $2^2 = 4$; $4^2 = 16$).

Bien sûr, il existe d'autres méthodes, comme la méthode des moindres valeurs absolues où l'on cherche les paramètres a et b qui minimisent $\sum_{i=1}^n |e_i|$. Mais les calculs associés à cette méthode sont plus compliqués.

Un résultat pratique

Les deux sommes suivantes sont égales.

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}$$

Preuve

On commence par développer la somme de gauche (les sommes ayant toujours les mêmes indices, on les notera juste \sum). À la fin du calcul, on utilise $n\bar{x} = \sum x_i$ et $n\bar{y} = \sum y_i$ (*).

$$\begin{aligned} \sum (x_i - \bar{x})(y_i - \bar{y}) &= \sum (x_i y_i - \bar{x} y_i - x_i \bar{y} + \bar{x} \bar{y}) \\ &\stackrel{(P_2)}{=} \sum x_i y_i - \sum \bar{x} y_i - \sum x_i \bar{y} + \sum \bar{x} \bar{y} \\ &\stackrel{(P_3)}{=} \sum x_i y_i - \bar{x} \sum y_i - \bar{y} \sum x_i + \bar{x} \bar{y} \sum 1 \\ &\stackrel{(*)}{=} \sum x_i y_i - \bar{x} n \bar{y} - \bar{y} n \bar{x} + n \bar{x} \bar{y} \\ &= \sum x_i y_i - n \bar{x} \bar{y} \end{aligned}$$

Notations

On va ainsi noter¹ chacune de ces deux sommes σ_{XY} .

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sigma_{XY} = \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}$$

En remplaçant Y par X ou X par Y , on a les formules suivantes.

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sigma_{XX} = \sum_{i=1}^n x_i^2 - n \bar{x}^2 \quad \text{et} \quad \sum_{i=1}^n (y_i - \bar{y})^2 = \sigma_{YY} = \sum_{i=1}^n y_i^2 - n \bar{y}^2$$

Théorème des moindres carrés

Les valeurs des paramètres a et b pour la droite des moindres carrés $y = ax + b$ sont :

$$\boxed{a = \frac{\sigma_{XY}}{\sigma_{XX}}} \quad \text{et} \quad \boxed{b = \bar{y} - a \bar{x}}$$

Ces formules ne sont valables que s'il existe au moins deux x_i qui ont des valeurs différentes (sinon, il y a une division par zéro dans la formule pour a).

Conséquences graphiques

Dans la section 5.6, on montre deux propriétés graphiquement intéressantes.

1. la droite des moindres carrés $y = ax + b$ passe par $(\bar{x}; \bar{y})$, qui est le *centre de gravité* des points $(x_i; y_i)$. En effet, on a la relation $\bar{y} = a\bar{x} + b$.
2. la droite des moindres carrés $y = ax + b$ est telle que la somme des résidus est nulle. En effet, la relation $\sum \hat{y}_i = \sum y_i$ est équivalente à $\sum e_i = 0$.

Ainsi, la droite de régression est agréable à regarder.

1. Par rapport aux notations en probabilités, on a $\sigma_{XX} = n\sigma^2(X)$ et $\sigma_{YY} = n\sigma^2(Y)$ où $\sigma(X)$ et $\sigma(Y)$ représentent les écarts types respectifs de X et de Y . De même, $\sigma_{XY} = n\text{Cov}(X, Y)$ où $\text{Cov}(X, Y)$ est la covariance de X et de Y . De plus, si on travaille avec des échantillons, les estimateurs de la variance et de la covariance sont sans biais lorsqu'on remplace n par $(n - 1)$.

5.2 Le coefficient de corrélation

On est maintenant capable de faire passer “au mieux” une droite parmi un nuage de points selon la méthode des moindres carrés. Cela ne nous dit toujours pas s’il y a une relation (ne serait-ce que linéaire) entre les caractères X et Y . Pour cela, les mathématiciens ont inventé un outil : il s’agit du *coefficient de corrélation* défini par

$$\rho = \frac{\sigma_{XY}}{\sqrt{\sigma_{XX}}\sqrt{\sigma_{YY}}}$$

Propriétés de ce coefficient

Le coefficient de corrélation est toujours compris entre -1 et 1 .

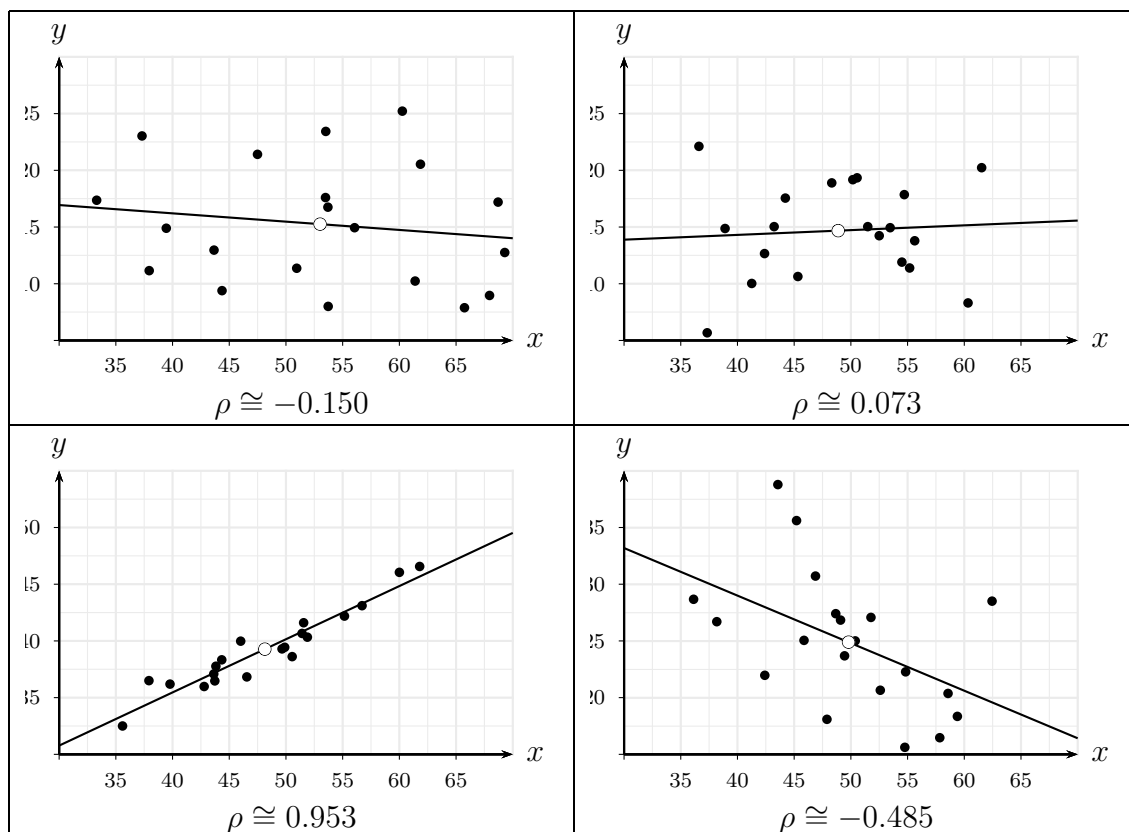
$$-1 \leq \rho \leq 1$$

C’est un outil qui permet de mesurer si la relation est linéaire, presque linéaire ou très peu linéaire.

- Lorsque ρ est proche de 0, alors la relation n’est pas linéaire. Soit les variables sont indépendantes, soit il y a un autre type de relation (voir page 72).
- Plus ρ est proche de 1, plus les points sont proches de la droite des moindres carrés (qui sera de pente positive).
- Plus ρ est proche de -1 , plus les points sont proches de la droite des moindres carrés (qui sera de pente négative).

Moralité Plus $|\rho|$ est proche de 1, meilleure est l’approximation par la droite des moindres carrés. On dit alors que X et Y sont *corrélés* (ou *linéairement dépendants*).

Exemples



5.3 Le coefficient de détermination

Relation évidente

Les valeurs y_i et \bar{y} proviennent directement des données. Les valeurs \hat{y}_i sont obtenues à partir du modèle. Ces trois valeurs sont liées par la relation suivante.

$$\underbrace{(y_i - \hat{y}_i)}_{\text{résidu}} + \underbrace{(\hat{y}_i - \bar{y})}_{\substack{\text{dépend aussi} \\ \text{du modèle}}} = \underbrace{(y_i - \bar{y})}_{\substack{\text{ne dépend que} \\ \text{des données}}}$$

Relation «miraculeuse»

Cette relation est vraie lorsqu'on utilise le modèle de la droite des moindres carrés.

$$\underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{\text{variation due aux résidus}} + \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{\text{variation due au modèle}} = \underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{\substack{\text{variation totale} \\ \text{(car somme des} \\ \text{deux autres)}}}$$

Définition du coefficient de détermination

Le coefficient de détermination, noté R^2 , est défini par

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Propriétés de ce coefficient

Lorsque la relation «miraculeuse» est vraie, le coefficient de détermination détermine le rapport entre la variation due au modèle et la variation totale. C'est donc le pourcentage de la variation due au modèle dans la variation totale.

Pour cette raison, le coefficient de détermination est toujours compris entre 0 et 1.

$$0 \leq R^2 \leq 1$$

- Lorsque R^2 est proche de 0, près du 100% de la variation totale est expliquée par la variation due aux résidus. Cela signifie que le modèle n'est pas adapté aux données.
- Lorsque R^2 est proche de 1, près du 100% de la variation totale est expliquée par la variation due au modèle. Cela signifie que le modèle est bien adapté aux données.

Théorème de retrouvailles

Lorsque le modèle est celui de la droite des moindres carrés, le coefficient de détermination est égal au carré du coefficient de corrélation. Autrement dit

$$R^2 = \rho^2$$

5.4 La droite des moindres carrés forcée à l'origine

Dans ce cas le modèle théorique $y = ax + b$ est construit de manière à ce que

1. la droite passe par l'origine du plan ;
2. la somme des carrés des résidus soit la plus petite possible.

Autrement dit, on veut a et b tels que

1. $b = 0$, ainsi le modèle est $y = ax$.
2. la somme $\sum_{i=1}^n e_i^2$ soit minimale.

Théorème des moindres carrés pour la version forcée à l'origine

La valeur du paramètre a pour la droite des moindres carrés $y = ax$ est :

$$a = \frac{\sigma_{XY}^{(0)}}{\sigma_{XX}^{(0)}} \quad \text{où} \quad \sigma_{XY}^{(0)} = \sum_{i=1}^n x_i y_i \quad \text{et} \quad \sigma_{XX}^{(0)} = \sum_{i=1}^n x_i^2$$

Cette formule n'est valable que s'il existe au moins deux x_i qui ont des valeurs différentes (sinon, il y a une division par zéro dans la formule pour a).

Ennuis

1. la droite des moindres carrés $y = ax$ ne passe pas forcément par $(\bar{x}; \bar{y})$, qui est le *centre de gravité* des points $(x_i; y_i)$.
2. la droite des moindres carrés $y = ax$ ne vérifie pas forcément la relation $\sum \hat{y}_i = \sum y_i$, et ainsi la somme des résidus $\sum e_i$ n'est pas forcément nulle.

Ainsi, à l'œil, cette droite peut paraître un peu bizarre.

Exemple

Lors d'une expérience d'osmose en biologie, un arbre chimique grandit dans une solution à 30% de saccharose. On cherche à déterminer la vitesse moyenne de croissance de l'arbre chimique durant les 15 premières minutes qui sera la pente de la droite des moindres carrés. Lors des mesures, des élèves ont obtenus les nombres suivants.

temps (min)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
hauteur (cm)	0.0	1.5	2.9	4.3	5.5	6.6	7.8	9.0	10.3	11.4	12.5	13.7	14.7	15.8	16.8	17.7

On trouve

$$a \cong 1.23 \text{ cm} \cdot \text{min}^{-1}$$

Les relations des pages 68 et 69 ne sont pas toujours vraies dans le cas du modèle $y = ax$. L'exemple précédent infirme chacune de ces formules.

Néanmoins, on peut les retrouver si, dans ces relations, on remplace \bar{x} et \bar{y} par 0, comme on le voit à la page suivante. Ce qui explique la notation avec les exposants ⁽⁰⁾.

Bien évidemment, si le centre de gravité $(\bar{x}; \bar{y})$ est l'origine du plan, les droites des moindres carrés $y = ax + b$ et le modèle qui force la droite à l'origine sont les mêmes.

5.4.1 Les coefficients de détermination et de corrélation

Relation évidente

$$\underbrace{(y_i - \hat{y}_i)}_{\text{résidu}} + \underbrace{\hat{y}_i}_{\text{dépend aussi du modèle}} = \underbrace{y_i}_{\text{ne dépend que des données}}$$

Relation «miraculeuse»

$$\underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{\text{variation due aux résidus}} + \underbrace{\sum_{i=1}^n \hat{y}_i^2}_{\text{variation due au modèle par rapport à 0}} = \underbrace{\sum_{i=1}^n y_i^2}_{\text{variation totale (car somme des deux autres)}}$$

Les coefficients de détermination et de corrélations

$$R^2 = \frac{\sum_{i=1}^n \hat{y}_i^2}{\sum_{i=1}^n y_i^2} \quad \text{et} \quad \rho = \frac{\sigma_{XY}^{(0)}}{\sqrt{\sigma_{XX}^{(0)}} \sqrt{\sigma_{YY}^{(0)}}} \quad \text{où} \quad \begin{aligned} \sigma_{XY}^{(0)} &= \sum_{i=1}^n x_i y_i \\ \sigma_{XX}^{(0)} &= \sum_{i=1}^n x_i^2 \\ \sigma_{YY}^{(0)} &= \sum_{i=1}^n y_i^2 \end{aligned}$$

Théorème de retrouvailles

Même dans ce modèle où la droite des moindres carrés est forcée à l'origine, on a

$$R^2 = \rho^2$$

5.4.2 Preuves

Même si on a perdu l'ingrédient $\hat{y}_i = \sum y_i$. On conserve l'ingrédient $\sum \hat{y}_i^2 = \sum \hat{y}_i y_i$.
En effet

$$\begin{aligned} \sum_{i=1}^n \hat{y}_i^2 &= \sum_i (ax_i)^2 = a^2 \sum_i x_i^2 = \left(\frac{\sum_i x_i y_i}{\sum_i x_i^2} \right)^2 \sum_i x_i^2 = \frac{(\sum_k x_k y_k)(\sum_i x_i y_i)}{\sum_k x_k^2} \\ &= \sum_i \left(\frac{\sum_k x_k y_k}{\sum_k x_k^2} x_i y_i \right) = \sum_i \left(\underbrace{ax_i}_{\hat{y}_i} y_i \right) = \sum_{i=1}^n \hat{y}_i y_i \end{aligned}$$

Preuve de la relation «miraculeuse»

On peut maintenant prouver la relation «miraculeuse».

$$\begin{aligned} \sum (y_i - \hat{y}_i)^2 + \sum \hat{y}_i^2 &= \sum y_i^2 - 2 \sum y_i \hat{y}_i + \sum \hat{y}_i^2 + \sum \hat{y}_i^2 \\ &\stackrel{\text{ingrédient}}{=} \sum y_i^2 - 2 \sum \hat{y}_i^2 + \sum \hat{y}_i^2 + \sum \hat{y}_i^2 = \sum y_i^2 \end{aligned}$$

Preuve du théorème de retrouvailles

$$R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\sum (ax_i)^2}{\sum y_i^2} = a^2 \cdot \frac{\sum x_i^2}{\sum y_i^2} = \left(\frac{\sigma_{XY}^{(0)}}{\sigma_{XX}^{(0)}} \right)^2 \cdot \frac{\sigma_{XX}^{(0)}}{\sigma_{YY}^{(0)}} = \frac{\sigma_{XY}^{(0)2}}{\sigma_{XX}^{(0)} \sigma_{YY}^{(0)}} = \rho^2$$

5.5 Autres types de régression

5.5.1 Préambule : une autre vision de la régression linéaire

On suppose que les données sont liées par une relation linéaire, non exacte, de la façon suivante.

$$y_i = ax_i + b + e_i$$

En suivant la méthode des moindres carrés, on trouve le minimum de $\sum_{i=1}^n e_i^2$ en annulant le gradient² : cela revient à résoudre le système suivant d'inconnues a et b .

$$\begin{cases} \sum_{i=1}^n y_i x_i = a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i = a \sum_{i=1}^n x_i + b n \end{cases}$$

5.5.2 Régression quadratique

On suppose que les données sont liées par une relation quadratique, non exacte, de la façon suivante.

$$y_i = ax_i^2 + bx_i + c + e_i$$

En suivant la méthode des moindres carrés, on trouve le minimum de $\sum_{i=1}^n e_i^2$ en annulant le gradient² : cela revient à résoudre le système suivant d'inconnues a , b et c .

$$\begin{cases} \sum_{i=1}^n y_i x_i^2 = a \sum_{i=1}^n x_i^4 + b \sum_{i=1}^n x_i^3 + c \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n y_i x_i = a \sum_{i=1}^n x_i^3 + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i = a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i + c n \end{cases}$$

5.5.3 Régression hyperbolique

On suppose que les données sont liées par une relation hyperbolique, non exacte, de la façon suivante.

$$y_i = \frac{1}{ax_i + b + e_i} \iff \frac{1}{y_i} = ax_i + b + e_i$$

La méthode des moindres carrés peut s'appliquer³ en utilisant le changement de variable suivant :

$$z = \frac{1}{y} \iff y = \frac{1}{z} \quad \text{et} \quad z_i = \frac{1}{y_i} \iff y_i = \frac{1}{z_i}$$

On trouve que :

$$\left(\iff z = \frac{1}{ax + b} \right) \quad \text{où} \quad \boxed{a = \frac{\sigma_{XZ}}{\sigma_{XX}}} \quad \text{et} \quad \boxed{b = \bar{z} - a\bar{x}}$$

2. Le gradient est une notion vue à l'université qui ne peut être expliquée qu'en deuxième année de lycée (il faut savoir dériver).

3. La *vraie* méthode des moindres carrés consisterait à minimiser la somme des carrés des ε_i donnés par la relation $y_i = \frac{1}{ax_i + b} + \varepsilon_i$. Néanmoins les calculs sont ici très complexes.

5.5.4 Régression exponentielle

On suppose que les données sont liées par une relation exponentielle, non exacte, de la façon suivante.

$$y_i = b \cdot a^{x_i} \cdot 10^{e_i} \iff \log(y_i) = \log(a)x_i + \log(b) + e_i$$

La méthode des moindres carrés peut s'appliquer⁴ en utilisant le changement de variable suivant :

$$w = \log(y) \iff y = 10^w \quad \text{et} \quad w_i = \log(y_i) \iff y_i = 10^{w_i}$$

On trouve que :

$$\left(\begin{array}{l} y = b \cdot a^x \\ \iff w = \log(a)x + \log(b) \end{array} \right) \text{ où } \begin{array}{l} \log(a) = \frac{\sigma_{XW}}{\sigma_{XX}} \\ \iff a = \exp_{10} \left(\frac{\sigma_{XW}}{\sigma_{XX}} \right) \end{array} \quad \text{et} \quad \begin{array}{l} \log(b) = \bar{w} - \log(a) \bar{x} \\ \iff b = 10^{\bar{w} - \log(a) \bar{x}} \end{array}$$

5.5.5 Régression d'une puissance

On suppose que les données sont liées par une puissance, non exacte, de la façon suivante.

$$y_i = b \cdot x_i^a \cdot 10^{e_i} \iff \log(y_i) = a \log(x_i) + \log(b) + e_i$$

La méthode des moindres carrés peut s'appliquer⁵ en utilisant le changement de variable suivant :

$$\begin{array}{l} w = \log(y) \iff y = 10^w \quad \text{et} \quad w_i = \log(y_i) \iff y_i = 10^{w_i} \\ v = \log(x) \iff x = 10^v \quad \text{et} \quad v_i = \log(x_i) \iff x_i = 10^{v_i} \end{array}$$

On trouve que :

$$\left(\begin{array}{l} y = b \cdot x^a \\ \iff w = av + \log(b) \end{array} \right) \text{ où } \begin{array}{l} a = \frac{\sigma_{VW}}{\sigma_{VV}} \\ \iff b = 10^{\bar{w} - a \bar{v}} \end{array} \quad \text{et} \quad \begin{array}{l} \log(b) = \bar{w} - a \bar{v} \\ \iff b = 10^{\bar{w} - a \bar{v}} \end{array}$$

5.5.6 Régression logarithmique

On suppose que les données sont liées par une relation logarithmique, non exacte, de la façon suivante.

$$y_i = a \log(x_i) + b + e_i$$

Cette fois la *vraie* méthode des moindres carrés peut s'appliquer en utilisant le changement de variable suivant :

$$v = \log(x) \iff x = 10^v \quad \text{et} \quad v_i = \log(x_i) \iff x_i = 10^{v_i}$$

$$\text{On trouve que : } \begin{array}{l} y = a \log(x_i) + b \\ \iff y = av + b \end{array} \text{ où } \begin{array}{l} a = \frac{\sigma_{VY}}{\sigma_{VV}} \\ \iff b = \bar{y} - a \bar{v} \end{array}$$

4. La *vraie* méthode des moindres carrés consisterait à minimiser la somme des carrés des ε_i donnés par la relation $y_i = b \cdot a^{x_i} + \varepsilon_i$. Néanmoins les calculs sont ici très complexes.

5. La *vraie* méthode des moindres carrés consisterait à minimiser la somme des carrés des ε_i donnés par la relation $y_i = b \cdot x_i^a + \varepsilon_i$. Néanmoins les calculs sont ici très complexes.

5.6 Preuves des théorèmes

5.6.1 Preuve des théorèmes des moindres carrés

Rappel sur les paraboles

Une parabole d'expression fonctionnelle $p(x) = \alpha x^2 + \beta x + \gamma$ avec $\alpha > 0$ a un minimum pour $x = -\frac{\beta}{2\alpha}$.

Preuve du théorème sur les moindres carrés

Trouvons une valeur de b tel que la somme des résidus au carré soit minimale. En d'autres termes, on veut trouver le minimum de l'expression suivante.

$$\sum_{i=1}^n e_i^2$$

En utilisant le fait que $\hat{y}_i = ax_i + b$, on peut rendre cette somme dépendante du paramètre b . C'est pourquoi, cette somme est momentanément appelée $S(b)$.

$$\begin{aligned} S(b) &= \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (ax_i + b))^2 \\ &= \sum_{i=1}^n ((y_i - ax_i) - b)^2 = \sum_{i=1}^n ((y_i - ax_i)^2 - 2(y_i - ax_i)b + b^2) \\ &= \sum_{i=1}^n (y_i - ax_i)^2 - 2b \sum_{i=1}^n (y_i - ax_i) + \sum_{i=1}^n b^2 \\ &= \sum_{i=1}^n (y_i - ax_i)^2 - 2b \left(\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i \right) + \sum_{i=1}^n b^2 \end{aligned}$$

Or $\sum_{i=1}^n y_i = n\bar{y}$ et $\sum_{i=1}^n x_i = n\bar{x}$, ainsi on a

$$\begin{aligned} S(b) &= \sum_{i=1}^n (y_i - ax_i)^2 - 2b(n\bar{y} - an\bar{x}) + nb^2 \\ &= \underbrace{n}_{\alpha > 0} b^2 - \underbrace{2n(\bar{y} - a\bar{x})}_{\beta} b + \underbrace{\sum_{i=1}^n (y_i - ax_i)^2}_{\gamma} \end{aligned}$$

Par le rappel ci-dessus, la valeur de b qui minimise $S(b)$ est donnée par

$$b = \frac{2n(\bar{y} - a\bar{x})}{2n} = \bar{y} - a\bar{x}$$

Maintenant qu'on a établi la relation $b = \bar{y} - a\bar{x}$, l'expression de la droite des moindres carrés est ainsi devenue

$$y = ax + b = ax + \bar{y} - a\bar{x} = a(x - \bar{x}) + \bar{y}$$

Il faut maintenant trouver a tel que la somme des résidus au carré soit minimale. En d'autres termes, on veut trouver le minimum de l'expression suivante.

$$\sum_{i=1}^n e_i^2$$

En utilisant le fait que $\hat{y}_i = ax_i + b = a(x_i - \bar{x}) + \bar{y}$, on peut rendre cette somme uniquement dépendante du paramètre a . C'est pourquoi, on a décidé d'appeler la somme $S(a)$.

$$S(a) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \left(y_i - (a(x_i - \bar{x}) + \bar{y}) \right)^2 = \sum_{i=1}^n \left(y_i - a(x_i - \bar{x}) - \bar{y} \right)^2$$

On cherche à trouver a tel que $S(a)$ est le plus petit possible. On sait que $S(a) \geq 0$ (car une somme de nombres positifs (ou nuls) ne peut être que positive (ou nulle)).

On a donc

$$\begin{aligned} S(a) &= \sum_{i=1}^n \left(y_i - a(x_i - \bar{x}) - \bar{y} \right)^2 = \sum_{i=1}^n \left((y_i - \bar{y}) - a(x_i - \bar{x}) \right)^2 \\ &= \sum_{i=1}^n \left((y_i - \bar{y})^2 - 2a(x_i - \bar{x})(y_i - \bar{y}) + a^2(x_i - \bar{x})^2 \right) \\ &= \underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{\sigma_{YY}} - 2a \underbrace{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}_{\sigma_{XY}} + a^2 \underbrace{\sum_{i=1}^n (x_i - \bar{x})^2}_{\sigma_{XX}} \\ &= \underbrace{\sigma_{XX}}_{>0} a^2 - 2\sigma_{XY}a + \sigma_{YY} \end{aligned}$$

Par le rappel ci-dessus, la valeur de a qui minimise $S(a)$ est donnée par

$$a = \frac{2\sigma_{XY}}{2\sigma_{XX}} = \frac{\sigma_{XY}}{\sigma_{XX}}$$

Mais, il faut que $\sigma_{XX} > 0$ afin d'avoir un minimum. C'est le cas s'il y a au moins deux x_i qui sont différents. \square

Preuve pour la version forcée à l'origine

Trouvons une valeur de a tel que la somme des résidus au carré soit minimale. En d'autres termes, on veut trouver le minimum de l'expression suivante.

$$\begin{aligned} S(a) &= \sum_{i=1}^n (y_i - ax_i)^2 = \sum_{i=1}^n (y_i^2 - 2ax_iy_i + a^2x_i^2) = \underbrace{\sum_{i=1}^n y_i^2}_{\sigma_{YY}^{(0)}} - 2a \underbrace{\sum_{i=1}^n x_iy_i}_{\sigma_{XY}^{(0)}} + a^2 \underbrace{\sum_{i=1}^n x_i^2}_{\sigma_{XX}^{(0)}} \\ &= \underbrace{\sigma_{XX}^{(0)}}_{>0} a^2 - 2\sigma_{XY}^{(0)}a + \sigma_{YY}^{(0)} \end{aligned}$$

Par le rappel ci-dessus, la valeur de a qui minimise $S(a)$ est donnée par

$$a = \frac{2\sigma_{XY}^{(0)}}{2\sigma_{XX}^{(0)}} = \frac{\sigma_{XY}^{(0)}}{\sigma_{XX}^{(0)}}$$

Mais, il faut que $\sigma_{XX}^{(0)} > 0$ afin d'avoir un minimum. C'est le cas s'il y a au moins deux x_i qui sont différents. \square

5.6.2 Preuves de la relation «miraculeuse»

On a besoin de deux ingrédients.

$$\boxed{\sum \hat{y}_i = \sum y_i \quad \text{et} \quad \sum \hat{y}_i^2 = \sum \hat{y}_i y_i} \quad \star$$

Si le modèle vérifie ces deux ingrédients, alors la relation «miraculeuse» est vraie.

En effet, on a

$$\begin{aligned} & \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2 \\ = & \sum (y_i^2 - 2\hat{y}_i y_i + \hat{y}_i^2) + \sum (\hat{y}_i^2 - 2\hat{y}_i \bar{y} + \bar{y}^2) \\ = & \sum y_i^2 - 2 \underbrace{\sum \hat{y}_i y_i}_{\star} + \sum \hat{y}_i^2 + \sum \hat{y}_i^2 - 2\bar{y} \underbrace{\sum \hat{y}_i}_{\star} + n\bar{y}^2 \\ \stackrel{\star}{=} & \sum y_i^2 - 2 \sum \hat{y}_i^2 + \sum \hat{y}_i^2 + \sum \hat{y}_i^2 - 2\bar{y} \sum y_i + n\bar{y}^2 \\ \stackrel{\sum y_i = n\bar{y}}{=} & \sum y_i^2 - 2 \sum \hat{y}_i^2 + \sum \hat{y}_i^2 + \sum \hat{y}_i^2 - 2n\bar{y}^2 + n\bar{y}^2 \\ = & \sum y_i^2 - \sum \hat{y}_i^2 + \sum \hat{y}_i^2 - n\bar{y}^2 \\ = & \sum y_i^2 - n\bar{y}^2 \\ \stackrel{\text{notation } \sigma_{YY}}{=} & \sum (y_i - \bar{y})^2 \\ & \text{page 67} \end{aligned}$$

5.6.3 Les deux visions pour la droite de régression

À la page 72, on affirme que a et b satisfont le système suivant.

$$\begin{cases} \sum y_i x_i = a \sum x_i^2 + b \sum x_i \\ \sum y_i = a \sum x_i + b n \end{cases} \iff \begin{cases} \sum y_i x_i = a \sum x_i^2 + b n \bar{x} \\ n \bar{y} = a n \bar{x} + b n \end{cases}$$

À la page 67, on a donné les valeurs suivantes de a et b .

$$a = \frac{\sigma_{XY}}{\sigma_{XX}} \quad \text{et} \quad b = \bar{y} - a \bar{x}$$

On retrouve ces coefficients en résolvant le système d'équation. En effet, la deuxième ligne est équivalente à

$$\bar{y} = a \bar{x} + b \iff b = \bar{y} - a \bar{x}$$

De plus si, à la première ligne, on soustrait \bar{x} fois la deuxième, cette première ligne devient

$$\begin{aligned} \sum y_i x_i - n \bar{x} \bar{y} &= a \sum x_i^2 - a n \bar{x}^2 \iff \sum y_i x_i - n \bar{x} \bar{y} = a (\sum x_i^2 - n \bar{x}^2) \\ \iff \sigma_{XY} &= a \sigma_{XX} \iff a = \frac{\sigma_{XY}}{\sigma_{XX}} \end{aligned}$$

5.6.4 Preuve des ingrédients pour le modèle linéaire

On se rappelle que a et b sont solutions du système

$$(\star) : \begin{cases} \sum y_i x_i = a \sum x_i^2 + b \sum x_i \\ \sum y_i = a \sum x_i + b n \end{cases}$$

Preuve du premier ingrédient

On utilise le fait que $\hat{y}_i = ax_i + b$, on développe et on observe le système.

$$\sum \hat{y}_i = \sum (ax_i + b) = a \sum x_i + b n \stackrel{(\star)}{=} \sum y_i$$

Preuve du deuxième ingrédient

On utilise le fait que $\hat{y}_i = ax_i + b$, on développe et on observe le système.

$$\begin{aligned} \sum \hat{y}_i^2 &= \sum (ax_i + b)^2 = \sum (a^2 x_i^2 + 2abx_i + b^2) = a^2 \sum x_i^2 + 2ab \sum x_i + b^2 n \\ &= a(a \sum x_i^2 + b \sum x_i) + ab \sum x_i + b^2 n \\ &= a(a \sum x_i^2 + b \sum x_i) + b(a \sum x_i + b n) \\ &\stackrel{(\star)}{=} a \sum y_i x_i + b \sum y_i \\ &= \sum y_i (ax_i + b) \\ &= \sum y_i \hat{y}_i = \sum \hat{y}_i y_i \end{aligned}$$

5.6.5 Preuve du théorème de retrouvailles

On se rappelle que $a = \frac{\sigma_{XY}}{\sigma_{XX}}$ et $b = \bar{y} - a\bar{x}$.

$$\begin{aligned} R^2 &= \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (ax_i + b - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (ax_i + \bar{y} - a\bar{x} - \bar{y})^2}{\sum (y_i - \bar{y})^2} \\ &= \frac{\sum (ax_i - a\bar{x})^2}{\sum (y_i - \bar{y})^2} = a^2 \cdot \frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2} = \left(\frac{\sigma_{XY}}{\sigma_{XX}} \right)^2 \cdot \frac{\sigma_{XX}}{\sigma_{YY}} = \frac{\sigma_{XY}^2}{\sigma_{XX}^2} \cdot \frac{\sigma_{XX}}{\sigma_{YY}} \\ &= \frac{\sigma_{XY}^2}{\sigma_{XX}\sigma_{YY}} = \left(\frac{\sigma_{XY}}{\sqrt{\sigma_{XX}\sigma_{YY}}} \right)^2 = \rho^2 \end{aligned}$$

5.6.6 Preuve des ingrédients pour le modèle quadratique

On se rappelle que a , b et c sont solutions du système

$$(\star) : \begin{cases} \sum y_i x_i^2 = a \sum x_i^4 + b \sum x_i^3 + c \sum x_i^2 \\ \sum y_i x_i = a \sum x_i^3 + b \sum x_i^2 + c \sum x_i \\ \sum y_i = a \sum x_i^2 + b \sum x_i + c n \end{cases}$$

Preuve du premier ingrédient

On utilise le fait que $\hat{y}_i = ax_i^2 + bx_i + c$, on développe et on observe le système.

$$\sum \hat{y}_i = \sum (ax_i^2 + bx_i + c) = a \sum x_i^2 + b \sum x_i + c n \stackrel{(\star)}{=} \sum y_i$$

Preuve du deuxième ingrédient

On utilise le fait que $\hat{y}_i = ax_i^2 + bx_i + c$, on développe et on observe le système.

$$\begin{aligned} \sum \hat{y}_i^2 &= \sum (ax_i^2 + bx_i + c)^2 \\ &= \sum (a^2 x_i^4 + 2abx_i^3 + b^2 x_i^2 + 2acx_i^2 + 2bcx_i + c^2) \\ &= a^2 \sum x_i^4 + 2ab \sum x_i^3 + b^2 \sum x_i^2 + 2ac \sum x_i^2 + 2bc \sum x_i + c^2 n \\ &= a(a \sum x_i^4 + b \sum x_i^3 + c \sum x_i^2) \\ &\quad + ab \sum x_i^3 + b^2 \sum x_i^2 + ac \sum x_i^2 + 2bc \sum x_i + c^2 n \\ &= a(a \sum x_i^4 + b \sum x_i^3 + c \sum x_i^2) \\ &\quad + b(a \sum x_i^3 + b \sum x_i^2 + c \sum x_i) \\ &\quad + ac \sum x_i^2 + bc \sum x_i + c^2 n \\ &= a(a \sum x_i^4 + b \sum x_i^3 + c \sum x_i^2) \\ &\quad + b(a \sum x_i^3 + b \sum x_i^2 + c \sum x_i) \\ &\quad + c(c \sum x_i^2 + b \sum x_i + c n) \\ &\stackrel{(\star)}{=} a \sum y_i x_i^2 + b \sum y_i x_i + c \sum y_i \\ &= \sum y_i (ax_i^2 + bx_i + c) \\ &= \sum y_i \hat{y}_i = \sum \hat{y}_i y_i \end{aligned}$$