Probabilités & Statistiques

David Hébert
hebert.iut@gmail.com

2022

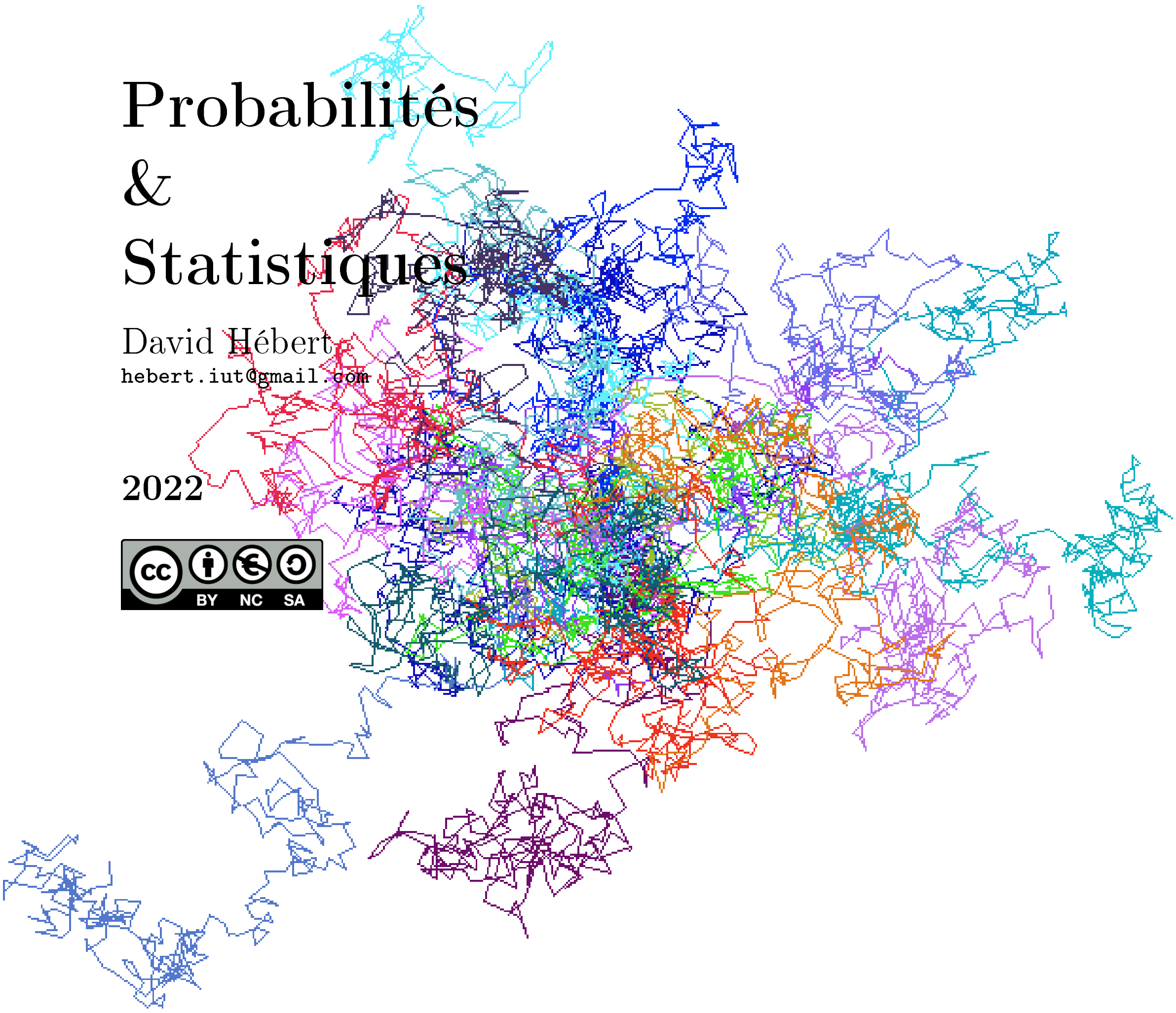


Table des matières

1. Probabilités discrètes

Aléatoire, univers, probabilités

Définir l'aléatoire est très difficile. Tentons une approche (que vous pourrez oublier au prochain paragraphe). En réalisant une expérience il peut y avoir deux natures de production :

- soit on sait comment va finir cette expérience de manière sûre et certaine. Dans ce cas on parle d'expérience **déterministe**. Par exemple si, sur la planète Terre, dans des conditions normales de pression, température etc on lance une pièce de monnaie alors elle tombe. On est sûre et certain qu'elle va retomber (c'est la loi de la gravitation qui le garantie).
- soit on ne sait pas comment va finir cette expérience mais on a des hypothèses de résultat. Par exemple, toujours avec la pièce, une fois lancée elle tombe soit sur pile soit sur face, mais, dans des conditions *normale* il est impossible de déterminer de manière sûre et certaine le résultat de cette expérience. On parle alors d'expérience **aléatoire**.

Étudier les résultats d'une expérience déterministe : c'est la statistique !

Les probabilités sont quant à elles, la science des expériences aléatoires. Le calcul des probabilités commence toujours avec une expérience aléatoire comme par exemple

- Lancer un dès
- Tirer une boule dans une urne
- Prendre trois cartes au hasard dans un jeu de carte
- ...

Définition

On appel **univers des possibilités** l'ensemble des issues possibles d'une expérience aléatoire. On le note en générale Ω.

Bon... nous avons tenter une définition mais elle souffre un petit peu de cadrage. Prenons par exemple l'expérience aléatoire "*On lance une pièce équilibrée*". L'univers des possibilités est alors $\Omega = \{\mathsf{Pile}, \mathsf{Face}\}$. Mais on pourrait aussi dire $\Omega = \{\mathsf{Pile}, \mathsf{Face}, \textit{Un condor passe au moment du lancé de pièce et l'emporte}\}$ sauf que ce dernier évènement bien que fort sympathique est improbable (dans des conditions normal de lancer de pièce dans une salle de classe par exemple et non dans une volière de condors ^_^). Dans l'univers des possibilités on ne considère que les *possibilités* c'est à dire ce qui a une *probabilité* de se produire. Qu'est-ce qu'une probabilité direz-vous. Oui c'est pas très propre de définir une notion avec une notion qu'on va définir après... d'un autre côté définir le hasard est difficile !

Bref, retenons que Ω désigne l'univers de ce qui peut se passer de raisonnable lors d'une expérience aléatoire. Par exemple :

- Si on lance trois pièces de monnaies alors

  $$\Omega = \{\mathsf{PPP}, \mathsf{PPF}, \mathsf{PFP}, \mathsf{FPP}, \mathsf{PFF}, \mathsf{FPF}, \mathsf{FFP}, \mathsf{FFF}\}$$

  où on note P pour pile de F pour face.
- Si on tire une carte dans un jeu de 52 cartes alors

  $$\Omega = \{\text{As de pique}, \text{Deux de pique}, \ldots\}$$

- Si on lance un dé à six faces alors

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

- ...

Définition

Soit Ω l'univers des possibilités d'une expérience aléatoire. Un **évènement** est un sous-ensemble de Ω.

Par exemple en lançant trois pièces de monnaies le sous-ensemble $\{\mathsf{FFP}, \mathsf{FPF}, \mathsf{PFF}\}$ correspond à l'évènement "obtenir exactement un pile".

Univers, sous-ensemble etc sont le cadre très rigoureux des probabilités : c'est très pratique, très puissant et suffisamment propre mathématiquement pour faire l'objet de théorème impressionnant. MAIS, il y a toujours un MAIS, le bagage théorique pour arriver à ce niveau de confort est très lourd, bien trop lourd pour ce cours : tribut, mesure, boréliens... bien au delà de ce que nous attendons ici.

Fort heureusement le calcul des probabilités est assez instinctif. Tout le monde comprend la phrase : "En lançant un dé, il y a une chance sur 6 d'obtenir un deux". Du coup, on peut présenter la probabilité de manière un peu moins rigoureuse mais beaucoup plus instinctive (en réalité ça s'appelle la *loi des grands nombres*).

Définition

Soit A un évènement relatif à une expérience aléatoire. On défini la probabilité de l'évènement A comme

$$P(A) = \lim_{\infty} \frac{\text{nombre de succès}}{\text{nombre de tentative}}$$

Nous avons lancer une pièce de monnaie équilibrée (c'est pour indiquer qu'elle n'est pas truquée). Nous considérons l'évènement E="Obtenir un pile". Nous avons réalisé plusieurs fois cette expérience pour estimer $P(E)$. Voici les résultats

Sur 25 lancés : 16 piles, donc $P(E) \simeq \dfrac{16}{25} = 0.64$

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

Sur 100 lancés : 49 piles, donc $P(E) \simeq \dfrac{49}{100} = 0.49$

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

Sur 200 lancés : 102 piles, donc $P(E) \simeq \dfrac{102}{200} = 0.51$

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

En continuant encore et encore cette expérience nous arriverons à montrer que $P(E) = 0.5$ que nous pouvons poétiquement traduire en "*la probabilité d'obtenir pile en lançant une pièce de monnaie équilibrée est d'une chance sur deux*". Dans la pratique on ne raisonne jamais avec cette définition par limite. On utilise de manière cachée les concepts ensemblistes. Par exemple en réalisant cette expérience, il y a deux issues possibles (pile ou face) et le seul résultat qui nous intéresse ne concerne qu'un de ces deux évènements (pile) on a donc un cas favorable sur deux cas possible : $\dfrac{1}{2}$.

Autre exemple (autre expérience aléatoire) : on lance un dé équilibré et on s'intéresse à l'événement E="on obtient un 3 ou un 6". Dans cette expérience il y a 6 issues possibles (1, 2, 3, 4, 5 ou 6) mais seulement deux qui sont favorables à E (à savoir 3 et 6). Deux cas sont alors favorables à l'évènement sur 6 cas possibles ; en conclusion : $P(E) = \frac{2}{6}$.

On ramène le calcul des probabilités à un comptage de possibilités. Avant d'avancer, il nous faut donc nous armer d'outils de comptage... d'ailleurs on dit plutôt *dénombrement* que comptage.

Outils du dénombrement

Définition

Pour tout entier n strictement positif on définit **factorielle de** n notée $n!$ comme le produit des n premiers entiers consécutifs.

Par exemple $3! = 1 \times 2 \times 3 = 6$, $5! = 1 \times 2 \times 3 \times 4 \times 5 = 120$.
On conviens, pour le bien être de l'univers mathématique, que $0! = 1$

Proposition

Soit $n \in \mathbb{N}$ alors $(n+1)! = (n+1) \times n!$

Démonstration. C'est une conséquence triviale de la définition. □

Définition

Soient n et p des entiers tel que $0 \leqslant p \leqslant n$. On défini A_n^p appelé l'**arrangement de** p **parmi** n le nombre

$$A_n^p = \frac{n!}{(n-p)!}$$

L'arrangement A_n^p correspond au produit des p nombres partant n en décrémentant. Par exemple $A_7^3 = 7 \times 6 \times 5 = 210$.

Définition

Soient n et p des entiers tel que $0 \leqslant p \leqslant n$. On défini C_n^p appelé la **combinaison de** p **parmi** n le nombre

$$C_n^p = \frac{n!}{p!(n-p)!} = \frac{A_n^p}{p!}$$

Par exemple $C_5^2 = \frac{5!}{2!3!} = 10$.

Proposition Triangle de Pascal

Soient n et p des entiers tel que $0 \leqslant p < n$.

$$C_n^p + C_n^{p+1} = C_{n+1}^{p+1}$$

Démonstration.

$$
\begin{aligned}
C_n^p + C_n^{p+1} &= \frac{n!}{p!(n-p)!} + \frac{n!}{(p+1)!(n-(p+1))!} \\
&= \frac{n!}{p!(n-p)!} + \frac{n!}{(p+1)!(n-p-1)!} \\
&= \frac{n! \times (p+1)}{p!(n-p)! \times (p+1)} + \frac{n! \times (n-p)}{(p+1)!(n-p-1)! \times (n-p)} \\
&= \frac{n! \times (p+1)}{(p+1)!(n-p)!} + \frac{n! \times (n-p)}{(p+1)!(n-p)!} \\
&= \frac{n!(p+1) + n!(n-p)}{(p+1)!(n-p)!} \\
&= \frac{n![(p+1)+(n-p)]}{(p+1)!(n-p)!} \\
&= \frac{n![n+1]}{(p+1)!(n-p)!} \\
&= \frac{(n+1)!}{(p+1)!(n-p)!} \\
&= \frac{(n+1)!}{(p+1)!(n+1-p-1)!} \\
&= \frac{(n+1)!}{(p+1)!((n+1)-(p+1))!} \\
&= C_{n+1}^{p+1}
\end{aligned}
$$

□

On parle du triangle de Pascal (du au mathématicien, physicien, philosophe etc Blaise Pascal) car ces nombres s'obtiennent de manière récurrentes en dessinant un tableau. Les lignes représentent le n et les colonnes le p.

n/p	0	1	2	3	4
0					
1					
2					
3					
4					

On initialise ce tableau en mettant des 1 sur la diagonale et dans la première colonne ainsi que des 0 au dessus de la diagonale

n/p	0	1	2	3	4
0	1	0	0	0	0
1	1	1	0	0	0
2	1		1	0	0
3	1			1	0
4	1				1

La formule s'illustre alors en sommant deux termes consécutifs sur une ligne pour obtenir celui de la ligne suivante

n/p	0	1	2	3	4
0	1	0	0	0	0
1	$\boxed{1}$	$\boxed{1}$	0	0	0
2	1	2	1	0	0
3	1			1	0
4	1				1

n/p	0	1	2	3	4
0	1	0	0	0	0
1	1	1	0	0	0
2	1	$\boxed{2}$	$\boxed{1}$	0	0
3	1		3	1	0
4	1				1

etc

Calculs des probabilités

On considère une urne avec n boules numérotées de 1 à n. On va tirer p boules dans cette urne de plusieurs manières différentes.

Premier cas : tirage avec remise.

Il y a n^p tirages différents.

Prenons par exemple $n = 4$ et $p = 2$. On se demande combien de tirage différent sont possibles de 2 boules dans une urne qui en contient 4. On commence par tirer une première boule, il y a 4 possibilités différentes. Puisque cette boule tirée est remise dans l'urne (car c'est le cas que nous traitons ici) alors lors du second tirage nous avons encore 4 possibilités. La liste de tous les cas possibles est

$$\begin{array}{cccc} 11 & 12 & 13 & 14 \\ 21 & 22 & 23 & 24 \\ 31 & 32 & 33 & 34 \\ 41 & 42 & 43 & 44 \end{array}$$

Second cas : tirage sans remise. Avec cette formulation on suppose que l'ordre compte.

Il y a A_n^p tirages différents.

Prenons par exemple $n = 4$ et $p = 2$. On se demande combien de tirage différent sont possibles de 2 boules dans une urne qui en contient 4. On commence par tirer une première boule, il y a 4 possibilités différentes. Puisque cette boule tirée n'est pas remise dans l'urne (car c'est le cas que nous traitons ici) alors lors du second tirage nous n'avons plus que 3 possibilités. La liste de tous les cas possibles est

$$\begin{array}{cccc} & 12 & 13 & 14 \\ 21 & & 23 & 24 \\ 31 & 32 & & 34 \\ 41 & 42 & 43 & \end{array}$$

Troisième cas : tirage simultané. Le mot simultané est équivalent à sans remise et sans ordre.

Il y a C_n^p tirages différents.

Prenons par exemple $n = 4$ et $p = 2$. On se demande combien de tirage différent sont possibles de 2 boules dans une urne qui en contient 4. On sait que si l'ordre compte on en a A_4^2. Mais dans ce dénombrement une différence a été faite entre 12 et 21 qui sont, dans le cas simultané le même tirage. Il faut donc diviser le nombre A_4^2 par $2 = 2!$ pour obtenir le résultat ce qui correspond exactement, par définition, à C_4^2. La liste de tous les cas possibles est

$$\begin{array}{ccc} 12 & 13 & 14 \\ & 23 & 24 \\ & & 34 \end{array}$$

Toutes les expériences aléatoires, comme presque tous les jeux combinatoires, peuvent s'identifier à un tirage de boule dans une urne. Par exemple combien de code a 4 chiffres est-il possible de composer avec un clavier numérique classique ? Cela reviens à se demander, dans une urne avec 10 boules, combien de tirage de 4 boules avec remise peut-on faire ? La réponse est $10^4 = 10\ 000$.

Autre exemple. On tire deux cartes au hasards dans un jeu qui en contient 52 (tirage simultané). Combien de jeu différent est-il possible d'obtenir ? Cela s'identifie à un tirage simultané de 2 boules dans une urne qui en contient 52 donc $C_{52}^2 = 1\ 326$.

Pour le calcul des probabilités, on raisonne comme nous l'avons suggéré dans l'introduction. La probabilité d'un évènement correspond au ratio du nombre de cas favorable divisé par le nombre de cas possible.

Toujours avec l'expérience aléatoire de tirer deux cartes dans un jeu de 52 (tirage simultané), considérons l'évènement E="On obtient deux piques". Alors $P(E) = \dfrac{C_{13}^2}{C_{52}^2}$. En effet il y a en tout C_{52}^2 tirage possibles (on prend 2 cartes parmi 52), d'où le dénominateur. Mais il y a C_{13}^2 manières d'avoir deux piques (sur les 52 cartes il y a 13 piques et on en souhaite 2) d'où le numérateur.

Considérons l'évènement F="Obtenir au moins un cœur". Pour calculer $P(F)$ utilisons quelques éléments théoriques.

Proposition

Soit Ω l'univers des possibilités d'une expériences aléatoire. Notons A un évènement et $\overline{A}$ son contraire.

1. $P(\Omega) = 1$
2. $P(\varnothing) = 0$
3. $P(A) + P(\overline{A}) = 1$
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Pour déterminer la probabilité de l'évènement F nous allons passer par son contraire $\overline{F}$="Obtenir aucun cœur". On a $P(\overline{F}) = \dfrac{C_{39}^2}{C_{52}^2}$ car 39 cartes ne sont pas des cœurs. En conclusion $P(F) = 1 - P(\overline{F}) = 1 - \dfrac{C_{39}^2}{C_{52}^2}$.

Un exemple

Considérons une urne avec 5 boules : deux boules rouges numérotées 1 et 2, deux boules bleues numérotées 1 et 2 et une boule verte avec le numéro 1. L'expérience aléatoire consiste à tirer avec remise trois boules dans l'urne.

> Quel est la probabilité de l'évènement "*on tire exactement une boule rouge ou exactement une boule numérotée d'un 1.*" ?

On considère les évènements suivants :

— A="On obtient exactement une boule rouge".

— B="On obtient exactement une boule numérotée d'un 1".

Déterminons la probabilité de l'évènement A. Il y a 5^3 tirages possible, il suffit de déterminer le nombre de tirage favorable à l'évènement A. Il faut donc compter combien donne exactement une boule rouge. Il y a trois cas :

1. On tire une boule rouge au premier tirage donc nécessairement deux boules non rouge pour les autres tirage; ce qui donne $2^1 \times 3^2 = 18$ puisque dans l'urne deux boules sont rouges et 3 ne le sont pas (et qu'il s'agit d'un tirage avec remise).
2. On tire une boule rouge au second tirage donc le premier et le troisième sont non rouge; ce qui donne : $3^1 \times 2^1 \times 3^1 = 18$.
3. On tire une boule rouge au troisième tirage; ce qui donne $3^2 \times 2^1 = 18$.

Finalement $P(A) = \dfrac{18+18+18}{5^3} = \dfrac{54}{125} = 0.432$.

De la même manière pour l'évènement B : $P(B) = \dfrac{3^1 \times 2^2 + 2^1 \times 3^1 \times 2^1 + 2^2 \times 3^1}{5^3} = \dfrac{36}{125} = 0.288$.

Déterminons à présent $P(A \cap B)$. L'évènement $A \cap B$ est "*On obtient exactement une boule rouge et une boule marquée de 1*". On peut diviser cette évènement en deux partie : tirer trois boules dont le 1 rouge (et donc deux qui ne sont ni 1 ni rouge), on notera E_1 cet évènement ou tirer trois boules avec un 1 et un rouge (et une ni 1 ni rouge) mais sans le 1 rouge que l'on notera E_2. Avec notre construction nous avons $P(A \cap B) = P(E_1) + P(E_2)$. Comme pour l'évènement A puisqu'il y a un seul 1 rouge et une seule boule qui n'est ni 1 ni rouge, on a $P(E_1) = \dfrac{1^1 \times 1^2 + 1^1 \times 1^1 \times 1^1 + 1^2 \times 1^1}{5^3} = \dfrac{3}{125} = 0.024$.

De la même manière : il y a une seule boule rouge qui n'est pas le 1 rouge, il y a deux boules numérotées 1 qui ne sont pas rouge et toujours une seule boule qui n'est ni rouge ni 1; de plus il y a 6 tirages différents possibles d'une boule rouge (qui n'est pas 1), d'un 1 (qui n'est pas rouge) et d'une boule qui n'est ni un rouge ni un 1 d'où $P(E_1) = 6 \times \dfrac{1^1 \times 2^1 \times 1^1}{5^3} = \dfrac{12}{125} = 0.096$.

Au final $P(A \cap B) = 0.024 + 0.096 = 0.12$.

Pour conclure la question de cet exercice, on utilise une formule introduite dans le précédent paragraphe :

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.432 + 0.288 - 0.12 \\ &= 0.72 - 0.12 \\ &= 0.6 \end{aligned}$$

Indépendance et conditionnement

Définition

On dira que deux évènements A et B sont indépendants si

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

Dans l'exemple précédent les évènements A et B ne sont pas indépendants car d'une part $P(A) \times P(B) = 0.432 \times 0.288 = 0.124416$ et d'autre part $P(A \cap B) = 0.12$.

Imaginons que nous lancions deux pièces de monnaies équilibrées. On note A l'évènement "*la première pièce donne un pile*" et B l'évènement "*la seconde pièce donne face*". Alors l'expérience s'identifie à tirer deux boules (la première pièce et la seconde pièce) dans une urne qui contient deux boules (pile et face). Le tirage est avec remise. Il y a donc 2^2 cas possible que l'on peut facilement énumérer $\Omega = \{PP, PF, FP, FF\}$. On observe alors que deux cas sont favorables à A ; précisément PP et FP d'où $P(A) = \frac{2}{4}$ et deux cas sont favorables à B (FF, PF) d'où $P(B) = \frac{2}{4}$. L'évènement $A \cap B$ correspond à "la première pièce donne pile et la seconde donne face" qui ne peut être que PF d'où $P(A \cap B) = \frac{1}{4}$. Alors

$$P(A \cap B) = \frac{1}{4} = \frac{2}{4} \times \frac{2}{4} = P(A) \times P(B)$$

et les évènements A et B sont indépendants.

Définition

Soient A et B deux évènements tel que $P(B) \neq 0$. On définit la **probabilité conditionnelle de** A **sachant** B notée $P_B(A)$ ou $P(A|B)$ par la formule

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

Reprenons le dernier exemple avec le lancer des deux pièces. On a $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2}$.

Proposition

Deux évènements A et B de probabilités non nuls sont indépendants si et seulement si

$$P(A|B) = P(A)$$

Démonstration. Cela vient de la définition de conditionnement $P(A|B) = \frac{P(A \cap B)}{P(B)}$ qui se réécrit $P(A|B) \times P(B) = P(A \cap B)$.

- Si A et B sont indépendants alors $P(A|B)P(B) = P(A \cap B) = P(A)P(B)$ et en simplifiant par $P(B)$ on trouve $P(A|B) = P(A)$.
- Si $P(A|B) = P(A)$ alors $P(A)P(B) = P(A|B)P(B) = P(A \cap B)$.

□

Voici un outil de la théorie des ensembles nécessaire à la formulation de la *probabilité totale*. C'est conceptuellement assez technique. Pour l'illustrer ce principe considérons l'expérience aléatoire suivante :

> On dispose de deux urnes. Un première U_1 contenant une boule blanche et quatre boules noires et une seconde U_2 avec trois boules blanche et deux boules noire.
>
> On lance un dé équilibré. Si il tombe sur un 6 on tire au hasard une boule dans l'urne U_1 sinon dans l'urne U_2.

Dans la suite du chapitre nous allons tenter de répondre à la question suivante : après avoir réalisé cette expérience nous avons tirer une boule noire. Quel est la probabilité d'avoir obtenu un 6 lors du lancé de dé ?

Définition

Soit E un ensemble. Une **partition** de E est la donnée de sous-ensemble E_i tel que leur union donne E et ont deux à deux des intersections vide.

Les partitions permettent de diviser l'univers des possibilités en plusieurs parties parfois plus simple à comprendre et dont les calculs de probabilités sont plus triviaux.

Dans notre exemple nous pouvons par exemple scinder l'univers des possibilités en deux (partitionner Ω en deux sous-ensembles) : soit on a obtenu un 6 lors du lancé de dé soit pas.

Théorème Formule de la probabilité totale

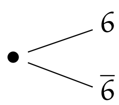
Soit A un évènement de $(\Omega_i)_i$ une partition de l'univers des possibilités Ω.

$$P(A) = \sum_i P(A|\Omega_i)P(\Omega_i)$$

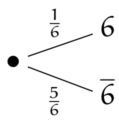
Mais oui que ça fait peur, que c'est incompréhensible, illisible ! Pas de panique nous allons tout de suite comprendre et voir comment, dans la pratique, une telle horreur s'utilise.

Un outil très élégant de ce type de calcul est **un arbre pondéré**. C'est un petit dessin qui représente l'expérience. Toute les expériences ne peuvent pas toujours être vu sur un arbre mais quand on cherche a appliquer la formule de la probabilité totale c'est qu'un arbre ne doit pas être loin.

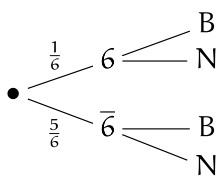
Nous commençons l'expérience par lancer un dé dont l'issu est soit un 6 soit pas un 6.



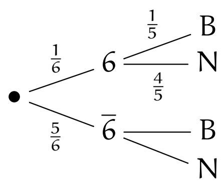
On augmente cet arbre en mettant sur les branches les probabilités d'obtenir l'évènement visé. Il y a trivialement une chance sur 6 d'obtenir un 6 et donc 5 chance sur 6 de ne pas l'obtenir.



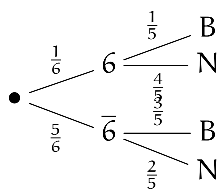
Une fois fait on tire une boule dont l'issu est la couleur blanche (B) ou noire (N) :



On doit faire apparaître les probabilités sur les arêtes. Considérons l'arête entre le 6 et le B. Par construction, la probabilité correspond à obtenir une boule blanche *sachant* que l'on a obtenu un 6 au lancé de dé. Il s'agit de la probabilité conditionnelle de B sachant 6. Inutile ici de revenir à la définition puisque si l'on a obtenu un 6 alors on a tiré au hasard une boule dans l'urne U_1, qui contient une seule boule blanche parmi les 5 d'où



De même pour les arêtes issues de $\overline{6}$.



La formule de probabilité totale permet alors de calculer $P(N)$ en suivant tous les chemins menant à l'évènement N. On multiplie les branches consécutives et on additionne les chemins. Ici

$$P(N) = \frac{1}{6} \times \frac{4}{5} + \frac{5}{6} \times \frac{2}{5} = \frac{14}{30}$$

Répondons à présent à la question : nous avons tirer une boule noire. Quel est la probabilité d'avoir obtenu un 6 lors du lancé de dé ? Cette question revient à calculer $P(6|N)$.

$$P(6|N) = \frac{P(6 \cap N)}{P(N)} = \frac{P(6)P(N|6)}{P(N)} = \frac{\frac{1}{6} \times \frac{4}{5}}{\frac{14}{30}} = \frac{\frac{4}{30}}{\frac{14}{30}} = \frac{4}{14} = \frac{2}{7} \simeq 0.286$$

Il y a environ 30% de chance d'avoir obtenue un 6 si à la fin de l'expérience on a tiré une boule noire.

Variables aléatoires

Nous commençons ce nouveau chapitre avec la définition ce ce qu'est une variable aléatoire. En préambule de cette définition, rappelons-nous les quelques mots marquant l'entrée des enfer de Dante : "*vous qui entrez ici, abandonner tout espoir*".

Définition

Soit Ω l'univers des possibilités d'une expérience aléatoire. Une **variable aléatoire réelle sur** Ω est la donnée d'une fonction $X : \Omega \to \mathbb{R}$.

Voila ! Comment ça va vous ? Alors oui ça pique. Comme nous l'avons légèrement survolé dans l'introduction du chapitre sur les probabilités, les concepts théoriques de cette branche des mathématiques que sont les probabilités sont très lourd mais très puissants. La définition de variable aléatoire présentée ici est aussi une de ces lourdeurs. Dans la pratique on n'utilise jamais (oh grand jamais) un tel concept fonctionnelle. On se rappel une formulation beaucoup plus simple de variable aléatoire : c'est une autre manière de regarder l'univers. C'est ce qu'est une fonction dans le fond ! Une fonction *transforme* l'espace de départ en l'espace d'arrivé. C'est ce que fait une variable aléatoire : comprendre, interpréter une expérience aléatoire d'une autre manière.

Prenons un exemple fameux : on lance deux dès non truqués. Alors l'univers des possibilités de cet expérience est l'ensemble des couples avec les nombres entiers de 1 à 6 que l'on peut énumérer :

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

On va considérer la variable aléatoire $X(a, b) = a + b$... oups pardon, j'ai dit qu'on n'utilise presque jamais la considération fonctionnelle des variables aléatoires. On considère donc la variable aléatoire où l'on fait la somme des valeurs obtenue par les dés.

Voila c'est tout. C'est pas plus compliqué : on ne regarde pas le résultat de l'expérience aléatoire en tant que tel mais modifié par une considération plus intéressante. On mélange ensuite ce concept aux calculs des probabilités classiques en se demandant par exemple quel est la probabilité d'obtenir 3 à cette somme dont la formulation devient maintenant : que vaut $P(X = 3)$?

Allez, faisons nous peur encore une fois.

Définition

Soit X une variable aléatoire réelle sur un univers de possibilités Ω d'une expérience aléatoire. Pour $x \in \mathbb{R}$, on pose

$$P(X = x) = P(\{\omega \in \Omega | X(\omega) = x\})$$

Bon pour être honnête cette définition n'est donnée que pour vos faire peur ! Oubliez là immédiatement (à moins que...). Retenons simplement que déterminer $P(X = x)$ reviens à déterminer quelles sont les évènements de Ω qui amène à x.

Dans notre exemple, quelles sont les les évènements de Ω dont la sommes fait 3. Un bref raisonnement nous permet de voir que deux cas sont favorables à un tel résultat 12 et 21. Deux cas favorables sur 36 cas possible, on en déduit que $P(X = 3) = \frac{2}{36}$.

Toujours avec le même exemple que vaut $P(X = 1)$? Deux dès sont lancés, et donne au moins 1 chacun donc la somme est au moins 2. La c'est facile : $P(X = 1) = 0$ (cela n'arrive jamais). De la même manière

$P(X=-1)=0$, $P(X=\ln(2))=0$ etc... Toutes les valeurs réelles ne sont pas intéressantes à étudier pour X. Bim définition !

Définition

Soit X une variable aléatoire réelle sur un univers de possibilités Ω d'une expérience aléatoire. On défini le **support** de X comme l'ensemble des valeurs de probabilité non nulle. On le note $\operatorname{Supp}(X)$:

$$\operatorname{Supp}(X)=\left\{x\in\mathbb{R}\;\middle|\;P(X=x)\neq 0\right\}$$

Dans notre exemple on trouve que $\operatorname{Supp}(X)=\{2,3,4,5,6,7,8,9,10,11,12\}$.

Définition

La **loi** d'une variable aléatoire réelle X est la donnée de $P(X=x)$ pour tout $x\in\operatorname{Supp}(X)$.

Dans la pratique, quand le support n'est pas trop grand, on synthétise la loi dans un tableau :

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

Comme nous pouvons l'observer dans l'exemple précédent la somme des probabilités fait toujours 1.

Proposition

$$\sum_{x\in\operatorname{Supp}(X)} P(X=x)=1$$

Espérance

La notion d'espérance correspond, dans le cadre déterministe des statistiques, à la notion de moyenne.

Définition

Soit X une variable aléatoire réelle. On note $E(X)$ l'**espérance** de la variable X, défini par

$$E(X)=\sum_{x\in\operatorname{Supp}(X)} xP(X=x)$$

L'espérance de X représente la valeur moyenne de X. Dans notre exemple de lancé de deux dé, l'espérance représente la moyenne des résultat possible.

Voici les résultats de 200 lancés (de la somme) que nous analysons comme une donnée statistique discrète :

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

— Il y a 9 tirages qui donnent un 2
— Il y a 14 tirages qui donnent un 3
— Il y a 11 tirages qui donnent un 4
— Il y a 19 tirages qui donnent un 5
— Il y a 28 tirages qui donnent un 6
— Il y a 39 tirages qui donnent un 7
— Il y a 27 tirages qui donnent un 8
— Il y a 21 tirages qui donnent un 9
— Il y a 12 tirages qui donnent un 10
— Il y a 15 tirages qui donnent un 11
— Il y a 5 tirages qui donnent un 12

La valeur moyenne de cette série statistique est $\frac{1390}{200} \simeq 6.95$. Déterminons la moyenne du coté probabiliste, c'est à dire l'espérance :

$$E(X) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} + 7 \times \frac{6}{36} + 8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} = 7$$

Variance et écart-type

La variance et l'écart-type en statistique ont elles aussi leurs équivalents dans l'univers non déterministe des probabilités.

Définition

La **variance** d'une variable aléatoire X, notée $V(X)$ est défini par

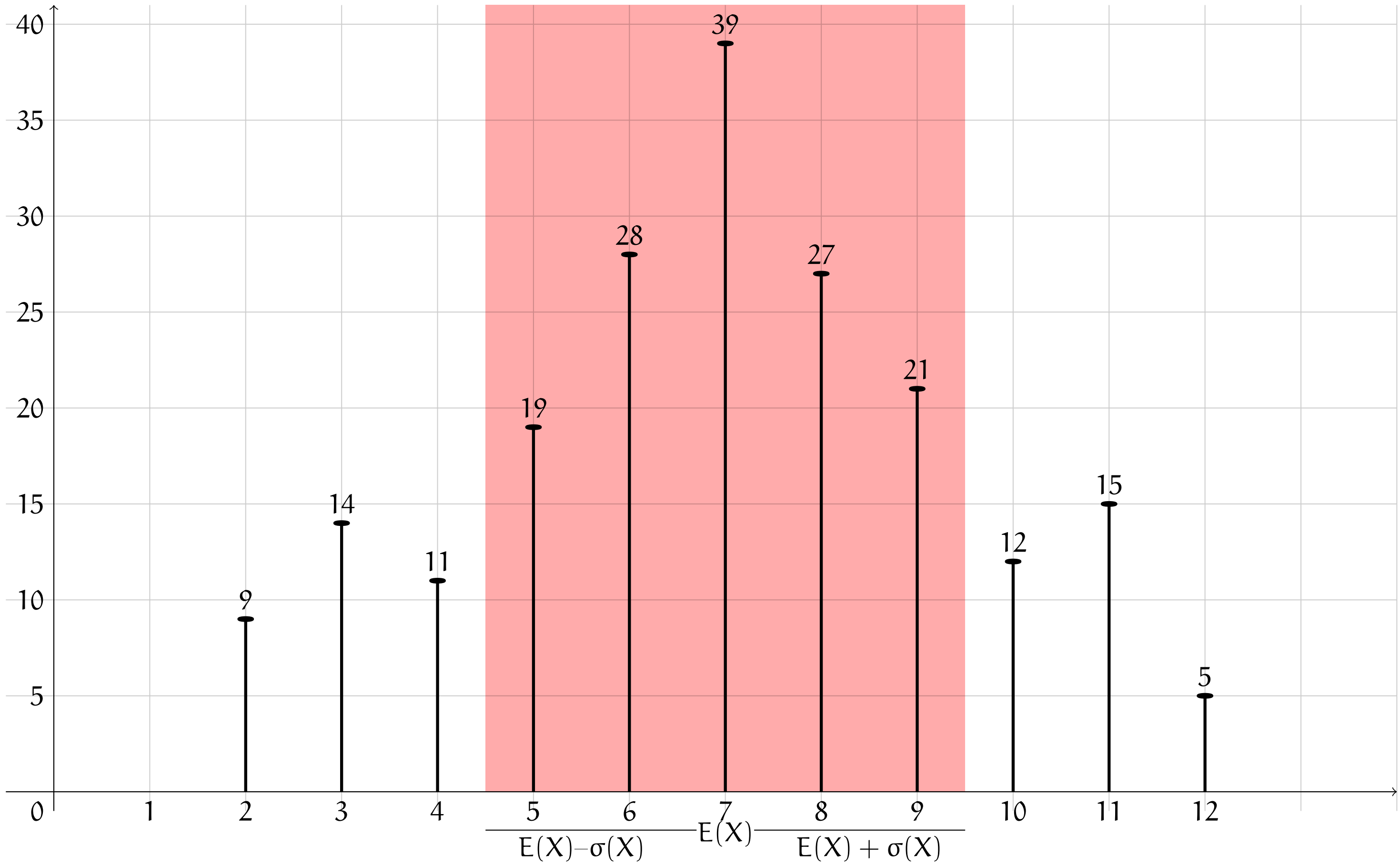
$$V(X) = E\left((X - E(X))^2\right)$$

L'**écart-type** $\sigma(X)$ est la racine carré de la variance.

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

L'écart-type estime l'écart probable entre une possibilité et la moyenne.

Reprenons les valeurs simulées au paragraphe précédent et déterminons la variance statistique de ce jeu de donnée. Un calcul passionnant donne que la variance est $\frac{1235.5}{200} \simeq 6.1775$ et donc l'écart-type vaut approximativement 2.5. Nous avions trouvé une moyenne de 6.95 que nous pouvons arrondir à 7. Alors il faut comprendre que la plupart des résultats de cette expérience aléatoire sont autour de 7 à plus ou moins 2.5 prés. En effet si on représente les données dans un diagramme en bâtons nous observons bien que plus de 60% des simulations sont entre 4.5 et 9.5. Les autres évènements ne sont pas improbables mais plus rare.



Dans la pratique on utilise la formule suivante pour calculer la variance.

Proposition Formule de Köning-Huygens

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

Démonstration.

$$\begin{aligned}
V(X) &= E\left((X - E(X))^2\right) \\
&= \sum_{x \in Supp(X)} (x - E(X))^2 P(X = x) \\
&= \sum_{x \in Supp(X)} (x^2 - 2xE(X) + E(X)^2) P(X = x) \\
&= \sum_{x \in Supp(X)} x^2 P(X = X) - \sum_{x \in Supp(X)} 2xE(X)P(X = x) + \sum_{x \in Supp(X)} E(X)^2 P(X = x) \\
&= \sum_{x \in Supp(X)} x^2 P(X = X) - 2E(X) \sum_{x \in Supp(X)} xP(X = x) + E(X)^2 \sum_{x \in Supp(X)} P(X = x) \\
&= E(X^2) - 2E(X)E(X) + E(X)^2 \times 1 \\
&= E(X^2) - E(X)^2
\end{aligned}$$

□

Déterminons à présent la valeur exacte et la variance et de l'écart-type de l'exemple que nous traitons.

$$E(X^2) = 2^2 \times \frac{1}{36} + 3^2 \times \frac{2}{36} + 4^2 \times \frac{3}{36} + 5^2 \times \frac{4}{36} + 6^2 \times \frac{5}{36} + 7^2 \times \frac{6}{36} + 8^2 \times \frac{5}{36} + 9^2 \times \frac{4}{36} + 10^2 \times \frac{3}{36} + 11^2 \times \frac{2}{36} + 12^2 \times \frac{1}{36} = \frac{329}{6}$$

$$V(X) = E(X^2) - E(X)^2 = \frac{329}{6} - 49 = \frac{35}{6}$$

$$\sigma(X) = \sqrt{\frac{35}{6}} \simeq 2.4152294577$$

Loi de Bernoulli

Définition

Soit $p \in [0;1]$. On dira qu'une variable aléatoire réelle X suit une **loi de Bernoulli**, notée

$$X \sim \mathcal{B}(p)$$

si le support de X est composée de deux valeurs, que l'on peut ramener à 0 et 1 tel que $P(X=1) = p$.

La loi de Bernoulli est appelé également la loi de *succès-échec*. Le paramètre p représentant la probabilité d'un succès. Puisque la somme des probabilités vaut toujours 1, on en déduit que $P(X=0) = 1-p$. On obtient aussi facilement les caractéristiques suivantes :

Théorème

Soit $X \sim \mathcal{B}(p)$.

1. $\text{Supp}\,(X) = \{0;1\}$.
2. La loi est $P(X=0) = 1-p$ et $P(X=1) = p$.
3. $E(X) = p$.
4. $V(X) = p(1-p)$.
5. $\sigma(X) = \sqrt{p(1-p)}$.

Voici un exemple. On joue au jeu suivant :

> On lance un dé. Si on obtient un nombre impaire on a perdu. Si on obtient un nombre paire on relance le dé. Si, à ce second lancé, on obtient un 6 on gagne sinon on perd.

Jouer à ce jeu coûte 10€. Si on gagne, on gagne 100€. Quel est le gain moyen de ce jeu ?

Nous somme dans le cas d'un succès-échec (gagner ou perdre). Déterminons la probabilité de gagner. Gagner c'est obtenir un nombre impaire au premier lancé, soit un chance sur 2, puis le 6 au second lancé soit une chance sur 6. Finalement la probabilité de gagner est $\frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$ et la probabilité de perdre et donc de $\frac{11}{12}$. Ainsi le nombre moyen de parti gagné est $p = \frac{1}{12}$. C'est à dire qu'on gagne 100€ une fois sur 12. Pour être précis on gagne 90€ car la partie à coûté 10€. Inversement 11 fois sur 12 on perd 10€. Si on fait la moyenne on arrive au gain moyen $-10 \times \frac{11}{12} + 90 \times \frac{1}{12} = -\frac{20}{12} = -\frac{5}{3}$. En moyenne, on perd environ $2,33$€. Cela signifie que si l'on joue 1000 fois, parfois on va gagner, souvent on va perdre et en moyenne on va perdre $2.33 \times 1000 = 2\ 330$€.

Si on reprend le même problème avec un gain de 150€ en cas de victoire alors le gain moyen sera $-10 \times \frac{11}{12} + 140 \times \frac{1}{12} = \frac{30}{12} = \frac{5}{2}$. Avec cette modalité, on va toujours perdre souvent mais les rare fois où l'on va gagner on va gagner suffisamment pour équilibrer notre porte feuille. Précisément si l'on joue 1000 fois, on va en moyenne, gagner $2\ 500$€.

Dans les jeux de casino, ou les jeux à gratter de la française des jeux, cette même étude est mené pour déterminer le prix du jeu. Il est calculé de telle sorte que la banque a toujours un gain positif et donc le joueur est, en moyenne, toujours perdant !

Loi Binomiale

Définition

Soient n un entier strictement positif et $p \in [0;1]$. On dira que X suit une loi binomiale, notée

$$X \sim \mathcal{B}(n, p)$$

si son support se ramène aux entiers entre à et n et que pour tout $k \in \{0, \dots, n\}$,

$$P(X = k) = C_n^k p^k (1-p)^{n-k}$$

La loi binomiale est appelée également la loi de n succès-échec.

Théorème

Soit $X \sim \mathcal{B}(n, p)$.

1. $\text{Supp}\,(X) = \{0, 1, \dots, n\}$.
2. La loi est $P(X = k) = C_n^k p^k (1-p)^{n-k}$.
3. $E(X) = np$.
4. $V(X) = np(1-p)$.
5. $\sigma(X) = \sqrt{np(1-p)}$.

Démonstration. Calculons l'espérance.

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{n} kP(X=k) \\
&= \sum_{k=0}^{n} kC_n^k p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \qquad \text{pour } k=0 \text{ le terme vaut } 0 \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n(n-1)!}{(k-1)!((n-1)-(k-1))!} p^k (1-p)^{(n-1)-(k-1)} \\
&= n\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} p^k (1-p)^{(n-1)-(k-1)} \\
&= n\sum_{k'=0}^{n-1} \frac{(n-1)!}{k'!((n-1)-k')!} p^{k'+1} (1-p)^{(n-1)-k'} \qquad \text{changement de variable : } k'=k-1 \\
&= n\sum_{k'=0}^{n-1} C_{n-1}^{k'} p^{k'+1} (1-p)^{(n-1)-k'} \\
&= n\sum_{k'=0}^{n-1} C_{n-1}^{k'} p p^{k'} (1-p)^{(n-1)-k'} \\
&= np\sum_{k'=0}^{n-1} C_{n-1}^{k'} p^{k'} (1-p)^{(n-1)-k'} \\
&= np(p+1-p)^{n-1} \qquad \text{binôme de Newton} \\
&= np
\end{aligned}
$$

Calculons l'espérance de X^2.

$$
\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2 P(X=k) \\
&= \sum_{k=0}^{n} k^2 C_n^k p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \frac{n(n-1)!}{(k-1)!((n-1)-(k-1))!} p^k (1-p)^{(n-1)-(k-1)} \\
&= n \sum_{k=1}^{n} k \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} p^k (1-p)^{(n-1)-(k-1)} \\
&= n \sum_{k'=0}^{n-1} (k'+1) \frac{(n-1)!}{k'!((n-1)-k')!} p^{k'+1} (1-p)^{(n-1)-k'} \\
&= n \left[\sum_{k'=0}^{n-1} k' \frac{(n-1)!}{k'!((n-1)-k')!} p^{k'+1} (1-p)^{(n-1)-k'} + \sum_{k'=0}^{n-1} \frac{(n-1)!}{k'!((n-1)-k')!} p^{k'+1} (1-p)^{(n-1)-k'} \right] \\
&= n \left[\sum_{k'=1}^{n-1} k' \frac{(n-1)!}{k'!((n-1)-k')!} p^{k'+1} (1-p)^{(n-1)-k'} + \sum_{k'=0}^{n-1} \frac{(n-1)!}{k'!((n-1)-k')!} p p^{k'} (1-p)^{(n-1)-k'} \right] \\
&= n \left[\sum_{k'=1}^{n-1} \frac{(n-1)(n-2)!}{(k'-1)!((n-2)-(k'-1))!} p^{k'+1} (1-p)^{(n-2)-(k'-1)} + p \sum_{k'=0}^{n-1} C_{n-1}^{k'} p^{k'} (1-p)^{(n-1)-k'} \right] \\
&= n \left[(n-1) \sum_{k''=0}^{n-2} \frac{(n-2)!}{k''!((n-2)-k'')!} p^{k''+2} (1-p)^{(n-2)-k''} + p(p+1-p)^{n-1} \right] \\
&= n \left[(n-1) \sum_{k''=0}^{n-2} C_{n-2}^{k''} p^{k''+2} (1-p)^{(n-2)-k''} + p \right] \\
&= n \left[(n-1) \sum_{k''=0}^{n-2} C_{n-2}^{k''} p^2 p^{k''} (1-p)^{(n-2)-k''} + p \right] \\
&= n \left[(n-1) p^2 \sum_{k''=0}^{n-2} C_{n-2}^{k''} p^{k''} (1-p)^{(n-2)-k''} + p \right] \\
&= n \left[(n-1)p^2 (p+1-p)^{n-2} + p \right] \\
&= n \left[(n-1)p^2 + p \right]
\end{aligned}
$$

En appliquant la formule de Köning-Huygens, on arrive à

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

□

Voici un exemple. Un questionnaire à choix multiple est composé de 8 questions. Pour chacune d'elles 4 réponses sont proposées et une seule est juste. Un étudiant répond au hasard à chacune des 8 questions de ce QCM. Quelle est la probabilité qu'il ai exactement 4 réponses correcte.

On note X la variable aléatoire comptant le nombre de bonne réponse. Alors $X \sim \mathcal{B}\left(8; \frac{1}{4}\right)$. En effet, il s'agit de 8 succès ou échec (bonne réponse ou mauvaise réponse). Un succès étant une bonne réponse. Puisque l'étudiant répond au hasard parmi les 4 solutions proposées il a une chance sur 4 d'avoir la réponse correcte.

La question reviens alors à se demander que vaut $P(X = 4)$. D'après la formule on a

$$P(X = 4) = C_8^4 \left(\frac{1}{4}\right)^4 \left(1 - \frac{1}{4}\right)^{8-4} = 70 \frac{1}{4^4} \frac{3^4}{4^4} = \frac{70 \times 3^4}{4^8} = \frac{5670}{65536} \simeq 0.087$$

Il y a donc environ 9% de chance d'avoir la moyenne... pas très fameux.

Combien de réponse seront juste en moyenne ? Il s'agit de calculer $E(X) = 8 \times \frac{1}{4} = 2$. En répondant au hasard, on aura en moyenne deux bonnes réponses sur les 8. Il a donc de forte chance d'avoir, en répondant au hasard, d'avoir au moins une réponse juste. Déterminons cette probabilité. Avoir au moins une réponse juste reviens à calculer $P(X \geqslant 1)$. Passons par l'évènement contraire : $P(X \geqslant 1) = 1 - P(X < 1)$ mais puisque X ne prend que des valeurs entières strictement positives si $X < 1$ alors $X = 0$.

$$P(X \geqslant 1) = 1 - P(X = 0) = 1 - C_8^0 \left(\frac{1}{4}\right)^0 \left(1 - \frac{1}{4}\right)^{8-0} = 1 - \left(\frac{3}{4}\right)^8 = 1 - \frac{3^8}{4^8} = 1 - \frac{6561}{65536} \simeq 0.899$$

Il y environ 90% de chance d'avoir au moins une réponse correcte mais moins de 10% d'en avoir 4.

Loi Géométrique

Définition

Soit $p \in [0; 1]$. On dira que X suit une loi géométrique, notée

$$X \sim \mathcal{G}(p)$$

si son support est l'ensemble des entiers strictement positif et que pour tout $k \in \mathbb{N}_{>0}$,

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

La loi géométrique est également appelé la loi du premier succès.

Théorème

Soit $X \sim \mathcal{G}(p)$.

1. $\mathrm{Supp}(X) = \mathbb{N}_{>0}$.
2. La loi est $P(X = k) = p(1-p)^{k-1}$.

3. $E(X) = \frac{1}{p}$.
4. $V(X) = \frac{1-p}{p^2}$.
5. $\sigma(X) = \frac{\sqrt{1-p}}{p}$.

Démonstration. Cette preuve fait appel à la notion de série convergente.

Il est bien connue que $\frac{1}{1-x} = \sum_{k=0}^{+\infty} x^k$ pour tout $x \in]-1;1[$. En dérivant cette expression on arrive à la formule

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

En dérivant une seconde fois

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

Déterminons à présent l'espérance :

$$\begin{aligned}
E(X) &= \sum_{k=1}^{+\infty} kP(X=k) \\
&= \sum_{k=1}^{+\infty} kp(1-p)^{k-1} \\
&= p\sum_{k=1}^{+\infty} k(1-p)^{k-1} \\
&= p\frac{1}{(1-(1-p))^2} \\
&= p\frac{1}{p^2} \\
&= \frac{1}{p}
\end{aligned}$$

Calculons également $E(X^2)$

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

Finalement

$$
\begin{aligned}
V(X) &= E(X^2) - E(X)^2 \\
&= (1-p)\frac{2}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2 \\
&= \frac{2}{p^2} - \frac{2}{p} + \frac{1}{p} - \frac{1}{p^2} \\
&= \frac{1}{p^2} - \frac{1}{p} \\
&= \frac{1-p}{p^2}
\end{aligned}
$$

□

Prenons par exemple le problème de l'homme ivre : un homme ivre rentre chez lui. Son trousseau de clef comporte 10 clefs mais une seule ouvre la porte de sa maison. L'homme essaye les clefs une par une pour ouvrir la porte de sa maison, mais l'ivresse est tel qu'il oublie s'il a déjà essayé une clef ou non.

Combien de clef va-t-il essayer en moyenne avant de pouvoir rentrer dans sa maison ?

Notons X la variable aléatoire qui compte le nombre de tentative avant de trouver la bonne clef. Il s'agit donc d'un expérience aléatoire "premier succès", c'est à dire que X est une loi géométrique. Pour déterminer son paramètre il faut connaitre la probabilité d'un succès, c'est à dire la probabilité de tirer la bonne clef dans son trousseau de 10 clef. Comme le tirage de cette clef est aléatoire, il a une chance sur dix de tirer la bonne. Ainsi $X \sim \mathcal{G}\left(\frac{1}{10}\right)$.

La question posée concerne la moyenne de cette variable qui est d'après le théorème précédent $\frac{1}{p}$, c'est à dire dans ce cas 10.

En moyenne, il va essayer 10 clefs avant de pouvoir rentrer chez lui.

3. Probabilités Continues

Introduction

Considérons l'expérience aléatoire : "On tire au hasard un nombre réelle entre 0 et 1" et considérons l'évènement A : "Le nombre est 0.35498416". Quel est la probabilité de A ?

Si on applique le même procédé que celui des chapitres précédents, on dirait : il y a une infinité de valeur possible pour le nombre tiré (il y a une infinité de nombre réel entre 0 et 1) ou autrement dis la cardinalité $\sharp[0;1] = \infty$. D'autre part, il n'y a qu'une valeur qui vaut 0.35498416, donc

$$P(A) = \frac{1}{\infty}$$

et il on convient que cette fraction vaut 0 parce que $\lim\limits_{n\to+\infty} \frac{1}{n} = 0$ par exemple.

Autrement dis : il est impossible d'obtenir 0.35498416 lorsqu'on prend un nombre au hasard dans l'intervalle $[0,1]$. Bon... soit !

Considérons à présent l'évènement B : "le nombre tiré est entre 0.2 et 0.7". De la même manière, si l'on cherche à déterminer la probabilité de cet évènement, il faut compter combien d'éléments sont dans B. Mais il apparait claire, qu'il y a une infinité de nombre réel entre 0.2 et 0.7 (la cardinatlité $\sharp[0.2;0.7] = \infty$). On écrirait donc

$$P(B) = \frac{\sharp[0.2;0.7]}{\sharp[0;1]} = \frac{\infty}{\infty}$$

Et là, c'est un problème ! C'est une forme indéterminée ! Il faut essayer autre chose.

Voici une idée : l'intervalle $[0;1]$ est composé de 10 morceaux de taille 0.1. Précisément :

$$[0;0.1] \quad [0.1;0.2] \quad [0.2;0.3] \quad [0.3;0.4] \quad [0.4;0.5] \quad [0.5;0.6] \quad [0.6;0.7] \quad [0.7;0.9] \quad [0.8;0.9] \quad [0.9;1]$$

Chacun de ces intervalles contient le même nombre de points, une infinité certes, mais la même *taille d'infini*. En d'autre terme $\sharp[0.4;0.5] = \sharp[0.1;0.2]$. Ou encore par une savante formule :

$$\sharp[0;1] = 10\sharp[0;0.1]$$

De la même manière $\sharp[0.2;0.7] = 5\sharp[0;0.1]$. Du coup on aurait envie d'écrire et de simplifier la fraction

$$P(B) = \frac{\sharp[0.2;0.7]}{\sharp[0;1]} = \frac{5\sharp[0;0.1]}{10\sharp[0;0.1]} = \frac{5}{10} = \frac{1}{2}$$

Tout cela est bien beau mais dans l'égalité précédente nous avons simplifier la fraction par $\sharp[0;0.1] = \infty$ et ce n'est pas très propre, mais assez légitime.

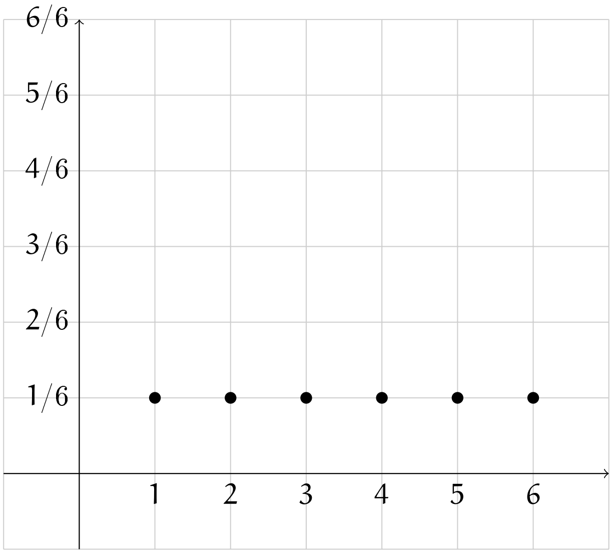
L'idée est de se débarrasser de ce dangereux infini [1].

Revenons aux bases. Qu'est-ce qu'une probabilité ? C'est une fonction qui va *peser* les ensembles ! Et la seule manière que nous avons de peser des ensembles c'est de passer par le calcul de leur cardinalité. Le poids d'un ensemble A (du point de vue la logique on parle donc d'*évènement*) est $\frac{\sharp A}{\sharp\Omega}$.

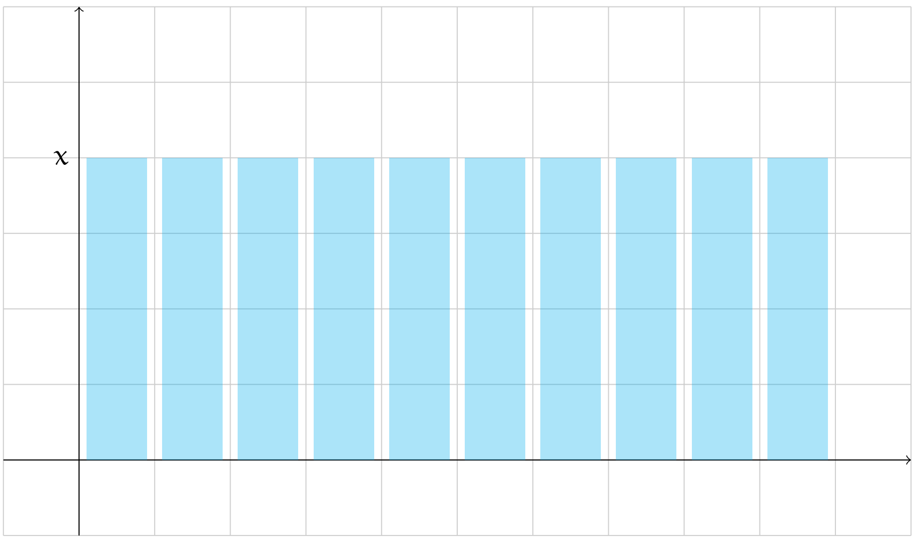
C'est ici que nous allons changer de paradigme. Nous allons passer par une fonction qui va représenter le poids d'un ensemble. Dans le cadre de ce cours, nous allons nous intéresser uniquement à *peser* les domaines de $\mathbb{R}$.

Reprenons l'exemple d'un dé équilibré. Représentons la fonction de poids

1. Les amoureux des concepts mathématiques pourront aller voir ce que sont les ordinaux et ne plus craindre l'infini !



Par analogie avec ce graphique reprenons l'exemple précédent de nombre réel tiré au hasard. Nous avions divisé l'intervalle $[0;1]$ en 10 morceaux de *même poids*. Ce poids est $\sharp[0;0.1]$ qui vaut l'infini. L'infini ferais un dessin pas très élégant. Pour faire simple, notons $x = \sharp[0;0.1]$. Alors la fonction de poids se représente comme suit



La somme des probabilité devant être égale à 1 nous serions tenter d'écrire $10x = 1$ donc $x = 0.1$ mais cela dépend bien évidement de la manière dont nous avons coupé $[0;1]$. Nous aurions pu également dire que $\sharp[0;1] = 100\sharp[0;0.01]$ ou $\sharp[0;1] = 1000\sharp[0;0.001]$ qui donnerait des valeurs de x différentes.

Cela dépend donc de la taille du découpage, ou plutôt la taille de l'intervalle $[0;0.1]$. La tentation n'est donc pas $10x = 1$ mais Nombre d'intervalle $\times$ taille de l'intervalle $\times\, x = 1$. Ainsi, peut importe la partition de $[0;1]$, $x = 1$.

Et en fait Nombre d'intervalle $\times$ taille de l'intervalle est la formule de l'aire d'un rectangle.

Ca y est on l'a! Notre nouveau paradigme pour *peser* des intervalles, c'est l'aire; et l'aire c'est une intégrale!

Fonction de densité

Définition

Soit $p : \mathbb{R} \longrightarrow \mathbb{R}$ une application. On dira que p est une fonction de densité si les trois conditions suivantes sont respectées :

- $\forall x \in \mathbb{R},\ p(x) \geqslant 0$
- La fonction p est continue sauf sur un nombre dénombrable de points.

— $\int_{\mathbb{R}} p(x)\ dx = 1$

Détaillons cette définition.

- Pour commencer, cette fonction est en fait une application, c'est à dire qu'elle doit être définie sur $\mathbb{R}$, en particulier elle ne peut pas avoir de valeur interdite.
- La première condition "$\forall x \in \mathbb{R},\ p(x) \geqslant 0$" ne traduit pas autre chose que l'application p est un poids ; elle ne peut pas être négative.
- La seconde condition est une hypothèse un peu technique pour garantir que nous pouvons calculer l'intégrale.
- La dernière condition est une reformulation de "*la somme des probabilités vaut 1*". En effet, lorsque nous étions en probabilité discrète, "*la somme des probabilités vaut 1*" s'écrivait $\sum_{k=0}^{n} P(X=k) = 1$.

  Dans le cadre continue, la notion de sommation ne se traduit plus par $\sum$ mais par $\int$ (d'ailleurs historiquement, lorsque B. Riemann a inventé[2] la notion de calcul d'aire il écrivait la lettre S pour *somme* ; la manipulation rapide et intensive de ce symbole l'a transformé au cour du temps en ce S déformé qu'est $\int$).

En particulier cette dernière remarque laisse à penser que tous les outils, concepts et formules que nous avons établie dans le cas discret avec le symbole $\sum$ ont leur pendant continue avec le symbole $\int$.

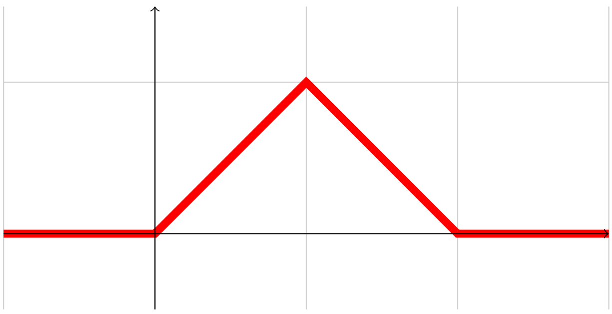
Un petit outil.

Définition

Soit D un domaine de $\mathbb{R}$. On note $\mathbb{1}_D$ la fonction

$$\begin{array}{rcl} \mathbb{1}_D : \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \begin{cases} 1 & \text{si } x \in D \\ 0 & \text{sinon} \end{cases} \end{array}$$

Considérons la fonction $p(x) = x\mathbb{1}_{[0;1]}(x) + (2-x)\mathbb{1}_{[1;2]}$ et vérifions qu'il s'agit bien d'une fonction de densité.



Le graphique nous convainc sans peine que la fonction est continue et positive sur $\mathbb{R}$. Vérifions que "*la somme des probabilités vaut 1*" :

$$\begin{aligned} \int_{\mathbb{R}} p(x)\ dx &= \int_{-\infty}^{0} p(x)\ dx + \int_{0}^{1} p(x)\ dx + \int_{1}^{2} p(x)\ dx + \int_{2}^{+\infty} p(x)\ dx \\ &= \int_{-\infty}^{0} 0\ dx + \int_{0}^{1} x\ dx + \int_{1}^{2} 2-x\ dx + \int_{2}^{+\infty} 0\ dx \\ &= 0 + \left[\frac{x^2}{2}\right]_0^1 + \left[-\frac{(2-x)^2}{2}\right]_1^2 + 0 \\ &= 0 + \left(\frac{1^2}{2} - \frac{0^2}{2}\right) + \left(-\frac{(2-2)^2}{2} + \frac{(2-1)^2}{2}\right) + 0 \\ &= 1 \qquad \text{CQFD} \end{aligned}$$

2. ... ou découvert

Variables aléatoires continues

Définition

Soit Ω l'univers des possibilités d'une expérience aléatoire dont les issues sont des nombres réelles, c'est à dire $\Omega \subseteq \mathbb{R}$.

Une variable aléatoire $X : \Omega \longrightarrow \mathbb{R}$ est la donnée d'une fonction de densité p. On note

$$X \sim p$$

Une variable aléatoire réelle est donc la donnée d'une fonction de répartition. Mathématiquement pour tout $\omega \in \Omega,\ X(\omega) = p(\omega)$ et $X(\omega) = 0$ pour les $\omega \notin \Omega$.

En imitant le cas discret nous avons les définitions suivantes.

Définition

Soit $X \sim p$.

— Le **support** de X, noté $\mathrm{Supp}(X)$ est l'ensemble des valeurs non nulle de la fonction de densité.

$$\mathrm{Supp}(X) = \left\{ x \in \mathbb{R} \middle| p(x) \neq 0 \right\}$$

— Soit A un évènement, c'est à dire $A \subseteq \Omega$.

$$P(X \in A) = \int_A p(x)\ dx$$

— L'**espérance** de la variable X est

$$E(X) = \int_{\mathbb{R}} xp(x)\ dx$$

— La **variance** de la variable X est

$$V(X) = \int_{\mathbb{R}} (x - E(X))^2 p(x)\ dx$$

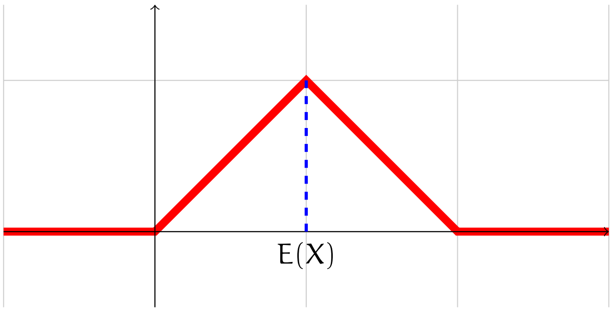
— L'**écart-type** de la variable X est

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

Reprenons l'exemple de la fonction de densité $p(x) = x\mathbb{1}_{[0;1]}(x) + (2-x)\mathbb{1}_{[1;2]}$ et déterminons sa moyenne, variance et écart-type.

$$\begin{aligned}
E(X) &= \int_{\mathbb{R}} xp(x)\ dx \\
&= \int_{-\infty}^{0} xp(x)\ dx + \int_0^1 xp(x)\ dx + \int_1^2 xp(x)\ dx + \int_2^{+\infty} xp(x)\ dx \\
&= \int_{-\infty}^{0} 0\ dx + \int_0^1 x^2\ dx + \int_1^2 x(2-x)\ dx + \int_2^{+\infty} 0\ dx \\
&= 0 + \left[\frac{x^3}{3}\right]_0^1 + \left[x^2 - \frac{x^3}{3}\right]_1^2 + 0 \\
&= \left[\frac{1^3}{3} - \frac{0^3}{3}\right]_0^1 + \left[\left(2^2 - \frac{2^3}{3}\right) - \left(1^2 - \frac{1^3}{3}\right)\right]_1^2 \\
&= \left(\frac{1}{3}\right) + \left(\left(4 - \frac{8}{3}\right) - \left(1 - \frac{1}{3}\right)\right) \\
&= 1
\end{aligned}$$

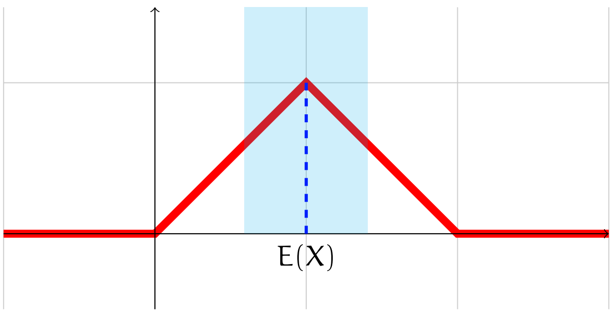
Graphiquement, l'espérance représente le *point d'équilibre de la fonction de densité*, c'est à dire, la valeur de x sur laquelle il faut poser son doigt pour tenir la fonction en équilibre.



$$
\begin{aligned}
V(X) &= \int_{\mathbb{R}} (x-1)^2 p(x)\ dx \\
&= \int_{-\infty}^{0} (x-1)^2 p(x)\ dx + \int_{0}^{1} (x-1)^2 p(x)\ dx + \int_{1}^{2} (x-1)^2 p(x)\ dx + \int_{2}^{+\infty} (x-1)^2 p(x)\ dx \\
&= \int_{-\infty}^{0} 0\ dx + \int_{0}^{1} (x-1)^2 x\ dx + \int_{1}^{2} (x-1)^2 (2-x)\ dx + \int_{2}^{+\infty} 0\ dx \\
&= 0 + \int_{0}^{1} x^3 - 2x^2 + x\ dx + \int_{1}^{2} -x^3 + 4x^2 - 5x + 2\ dx + 0 \\
&= \left[\frac{x^4}{4} - 2\frac{x^3}{3} + \frac{x^2}{2}\right]_0^1 + \left[-\frac{x^4}{4} + 4\frac{x^3}{3} - 5\frac{x^2}{2} + 2x\right]_1^2 \\
&= \left[\left(\frac{1^4}{4} - 2\frac{1^3}{3} + \frac{1^2}{2}\right) - \left(\frac{0^4}{4} - 2\frac{0^3}{3} + \frac{0^2}{2}\right)\right] + \left[\left(-\frac{2^4}{4} + 4\frac{2^3}{3} - 5\frac{2^2}{2} + 2\times 2\right) - \left(-\frac{1^4}{4} + 4\frac{1^3}{3} - 5\frac{1^2}{2} + 2\times 1\right)\right] \\
&= \left(\frac{1}{4} - \frac{2}{3} + \frac{1}{2}\right) + \left(\left(-\frac{16}{4} + \frac{32}{3} - \frac{20}{2} + 4\right) - \left(-\frac{1}{4} + \frac{4}{3} - \frac{5}{2} + 2\right)\right) \\
&= \frac{1}{4} - \frac{2}{3} + \frac{1}{2} - 4 + \frac{32}{3} - 10 + 4 + \frac{1}{4} - \frac{4}{3} + \frac{5}{2} - 2 \\
&= \frac{1}{6}
\end{aligned}
$$

Et donc l'écart-type $\sigma(X) = \frac{1}{\sqrt{6}} = \frac{\sqrt{6}}{6} \simeq 0.408$.

L'écart-type permet de déterminer ou se trouve la partie avec le plus de poids. Dans notre exemple elle se trouve autour de la moyenne à plus ou moins $\frac{\sqrt{6}}{6}$, précisément sur l'intervalle $I = \left[1 - \frac{\sqrt{6}}{6}; 1 + \frac{\sqrt{6}}{6}\right]$.



Pour s'en convaincre calculons la probabilité que X soit dans l'intervalle I.

$$\begin{aligned}
p(X\in I) &= \int_I p(x)\ dx \\
&= 2\int_{1-\frac{\sqrt{6}}{6}}^{1} p(x)\ dx \qquad \text{par symétrie} \\
&= 2\int_{1-\frac{\sqrt{6}}{6}}^{1} x\ dx \\
&= 2\left[\frac{x^2}{2}\right]_{1-\frac{\sqrt{6}}{6}}^{1} \\
&= \left[x^2\right]_{1-\frac{\sqrt{6}}{6}}^{1} \\
&= 1^2-\left(1-\frac{\sqrt{6}}{6}\right)^2 \\
&= 1-\left(1-2\frac{\sqrt{6}}{6}+\frac{6}{36}\right) \\
&= 1-1+\frac{2\sqrt{6}}{6}-\frac{1}{6} \\
&= \frac{2\sqrt{6}-1}{6} \\
&\simeq 65\%
\end{aligned}$$

Comme dans le cas discret nous aurons pu simplifier le calcul de la variance en appliquant la formule Köning-Huygens

Proposition Formule de Köning-Huygens

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

Démonstration. Il suffit de reprendre la démonstration dans le cas discret et remplacer le symbole $\sum$ par $\int$. □

Nous laissons au lecteur le passionnant calcul de $E(X^2)=\int_{\mathbb{R}} x^2p(x)\ dx$ et de s'assurer que $E(X^2)-E(X)^2$ donne bien $\frac{1}{6}$.

L'inconvénient de ces calculs réside dans la détermination d'une primitive. Dériver une fonction est une opération moins compliqué que l'intégration. Du point de vu des probabilité on parle de fonction de répartition.

Fonctions de répartitions

Définition

Soit $F:\mathbb{R}\longrightarrow\mathbb{R}$ une application. On dira que F est une **fonction de répartition** si les conditions suivantes sont vérifiées.

- L'application F est continue sur tout $\mathbb{R}$ et dérivable sauf sur un nombre dénombrable de point.
- L'application F est croissante.
- $\lim_{t\to-\infty} F(t)=0$.
- $\lim_{t\to+\infty} F(t)=1$.

Vérifions que la fonction F suivante est une fonction de répartition.

$$\begin{array}{rcl} \mathsf{F} : \mathbb{R} & \longrightarrow & \mathbb{R} \\ t & \longmapsto & \begin{cases} 0 & \text{si } t<0 \\ \dfrac{t^2}{2} & \text{si } t \in [0;1[\\ 1-\dfrac{(2-t)^2}{2} & \text{si } t \in [1;2[\\ 1 & \text{si } t \geqslant 2 \end{cases} \end{array}$$

Nous aurions pu également écrire

$$F(t) = \frac{t^2}{2}\mathbb{1}_{[0;1[}(t) + 1 - \frac{(2-t)^2}{2}\mathbb{1}_{[1;2[}(t) + \mathbb{1}_{[2;+\infty[}(t)$$

Clairement les conditions sur les limites sont vérifiées.

La fonctions est constante partout sauf entre 0 et 2. De plus sur $[0;1]$ c'est évidemment une parabole croissante. Une étude rapide permet de vérifier que sur $[1;2]$ la parabole $1-\dfrac{(2-t)^2}{2}$ est également croissante. En définitive, la fonction F est croissante sur $\mathbb{R}$.

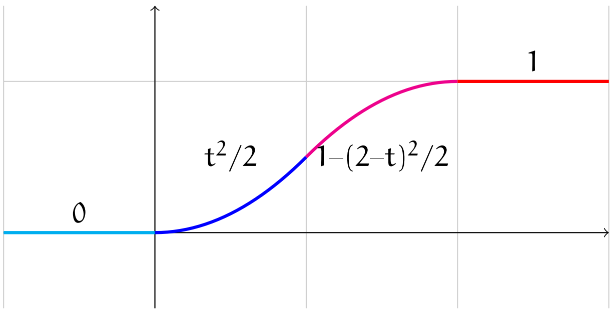
Il est également trivial que la fonction F est dérivable partout sauf éventuellement en 0, 1 et 2.

On observe que $\lim\limits_{t\to 0^-} F(t) = \lim\limits_{t\to 0^-} 0 = 0$ et d'autre par $\lim\limits_{t\to 0^+} F(t) = \lim\limits_{t\to 0^+} \dfrac{t^2}{2} = 0$ donc la fonction est continue en 0.

De la même manière, on a $\lim\limits_{t\to 1^-} F(t) = \lim\limits_{t\to 1^-} \dfrac{t^2}{2} = \dfrac{1}{2}$ et $\lim\limits_{t\to 1^+} F(t) = \lim\limits_{t\to 1^+} 1 - \dfrac{(2-t)^2}{2} = \dfrac{1}{2}$ ce qui prouve que F est continue en 1.

Finalement $\lim\limits_{t\to 2^-} F(t) = \lim\limits_{t\to 2^-} 1 - \dfrac{(2-t)^2}{2} = 1$ qui est bien la limite à droite de 2 et donc la fonction F est bien continue en 2.

Sans trop de peine on pourrait montrer que F est également dérivable sur tout $\mathbb{R}$ mais cela n'a pas d'intérêt ici.



Voici le liens qui existe entre fonction de répartition et fonction de densité.

Théorème

Soit $X \sim p$. La fonction

$$F_X(t) = P(X<t) = \int_{-\infty}^{t} p(x)\ dx$$

est une fonction de répartition. En particulier $F'_X(x) = p(x)$ pour les valeurs de x pour lesquelles F_X est dérivable.

Démonstration. La fonction de densité étant presque continue et l'opération d'intégration régularisante, la fonction F_X qui est une application puisque p l'est, est une fonction continue et dérivable sauf sur les points de discontinuité de p.

De plus comme la densité p est toujours positive, la fonction F_X est croissante (on ajoute de l'aire positive quand le t augmente).

La limite en $-\infty$ est trivialement nulle (il n'y a plus d'aire) et la limite en $+\infty$, correspond à "*la somme des probabilités vaut* 1" et vaut donc 1. $\square$

Corollaire

Soit X une variable aléatoire continue réelle de fonction de répartition F_X alors

$$P(X \in [a;b]) = F_X(b) - F_X(a)$$

Démonstration. Il s'agit de la reformulation du lien entre intégrale et primitive. □

Par exemple, la fonction de répartition de la densité $p(x) = x\mathbb{1}_{[0;1]}(x) + (2-x)\mathbb{1}_{[1;2]}$ est la fonction

$$F(t) = \frac{t^2}{2}\mathbb{1}_{[0;1[}(t) + \left(1 - \frac{(2-t)^2}{2}\right)\mathbb{1}_{[1;2[}(t) + \mathbb{1}_{[2;+\infty[}(t)$$

Lois classiques

Nous présentons ici un formulaire des lois classique en probabilités continue. Tous les résultats proposées se démontrent par des calculs plus ou moins savant qui sont laissés au lecteur [3]. *Sur les graphiques, les densités sont en rouges et les répartitions en bleues.*

Uniforme. Soit $a<b$ des nombres réels (finis)

Notation : $X \sim \mathcal{U}_{[a,b]}$

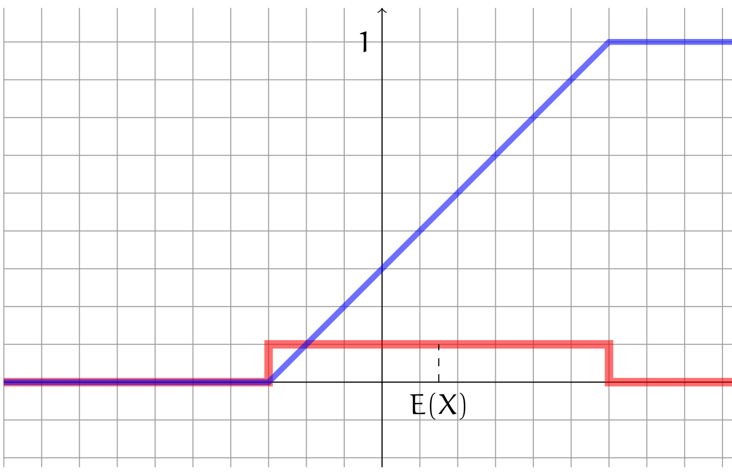
Support : $\mathrm{Supp}(X) = [a, b]$

Densité : $p(x) = \dfrac{1}{b-a}\mathbb{1}_{[a,b]}(x)$

Répartition : $F_X(t) = \dfrac{t-a}{b-a}\mathbb{1}_{[a,b]}(x) + \mathbb{1}_{]b,+\infty[}(x)$

Espérance : $E(X) = \dfrac{a+b}{2}$

Variance : $V(X) = \dfrac{(b-a)^2}{12}$



Exponentielle. Soit $\lambda>0$.

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

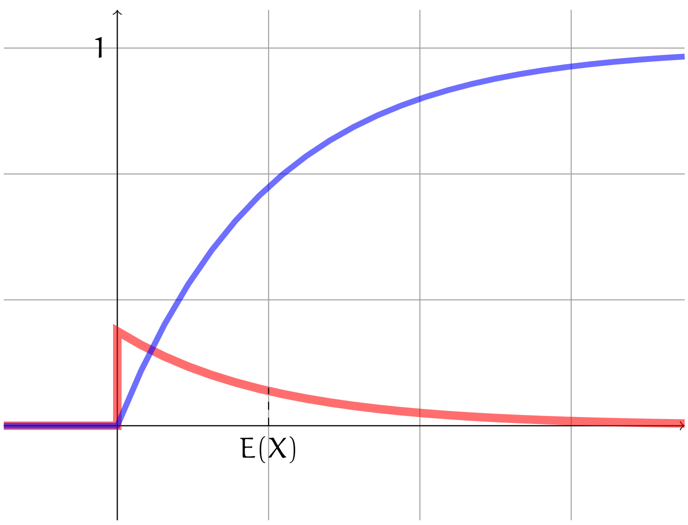
Support : $\mathrm{Supp}(X) = [0; +\infty[$

Densité : $p(x) = \lambda e^{-\lambda x}\mathbb{1}_{[0;+\infty]}(x)$

Répartition : $F_X(t) = (1 - e^{-\lambda t})\mathbb{1}_{[0;+\infty]}(t)$

Espérance : $E(X) = \dfrac{1}{\lambda}$

Variance : $V(X) = \dfrac{1}{\lambda}$



3. C'est un piège, ne vous lancez pas dans ces calculs !

Cauchy. Soient $a>0$ et $x_0 \in \mathbb{R}$.

Notation : $X \sim \mathcal{C}(x_0, a)$

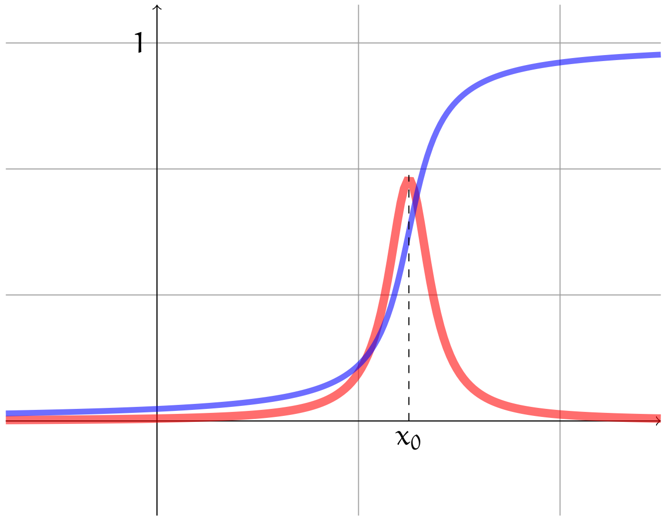
Support : $\mathrm{Supp}(X) = \mathbb{R}$

Densité : $p(x) = \dfrac{1}{\pi}\dfrac{a}{(x-x_0)^2 + a^2}$

Répartition : $F_X(t) = \dfrac{1}{\pi}\mathrm{Arctan}\left(\dfrac{x-x_0}{a}\right) + \dfrac{1}{2}$

Espérance : $E(X)$ n'existe pas (mais x_0 est raccourci efficace).

Variance : $V(X)$ n'existe pas (mais a est un raccourci efficace).



Normale. Soient $\sigma>0$ et $\mu \in \mathbb{R}$.

Notation : $X \sim \mathcal{N}(\mu, \sigma)$

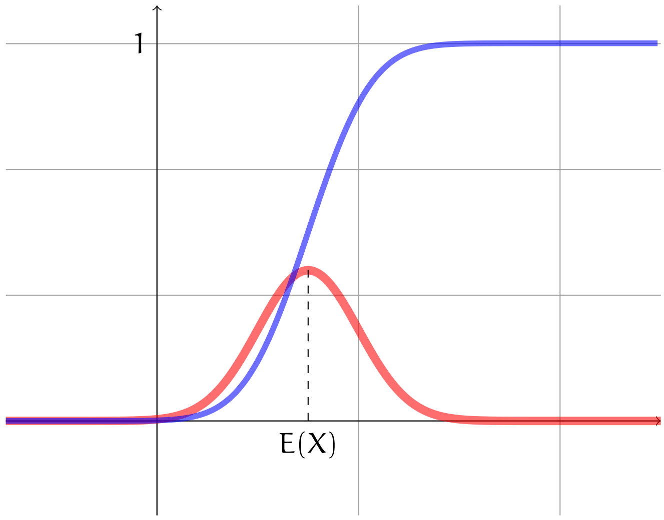
Support : $\mathrm{Supp}(X) = \mathbb{R}$

Densité : $p(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Répartition : Ne s'exprime avec des expressions algébrique simple.

Espérance : $E(X) = \mu$.

Variance : $V(X) = \sigma^2$.



Chi deux. Soit n un entier strictement positif appelé le degré de liberté.

Notation : $X \sim \chi^2_n$

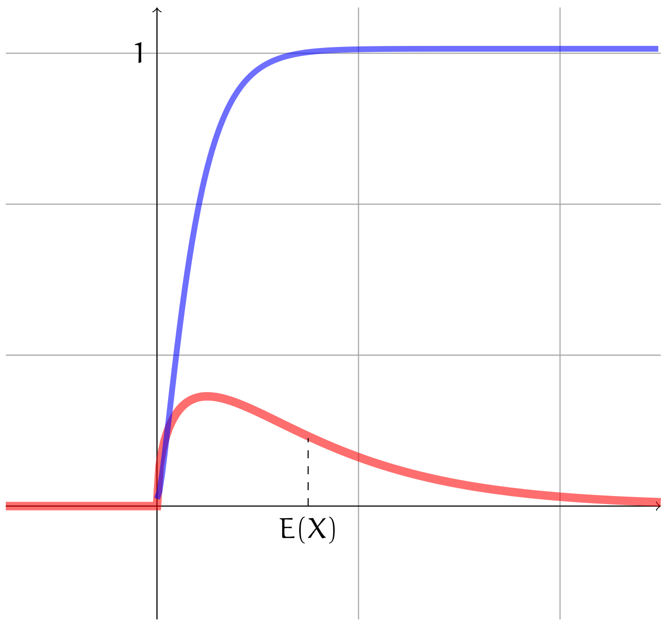
Support : $\mathrm{Supp}(X) = \mathbb{R}$

Densité : $p(x) = \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{\mathbb{R}_+}(x)$

Répartition : Existe, mais de forme complexe.

Espérance : $E(X) = n$

Variance : $V(X) = 2n$



Student. Soit n un entier strictement positif appelé le degré de liberté.

Notation : $X \sim \mathcal{T}_n$

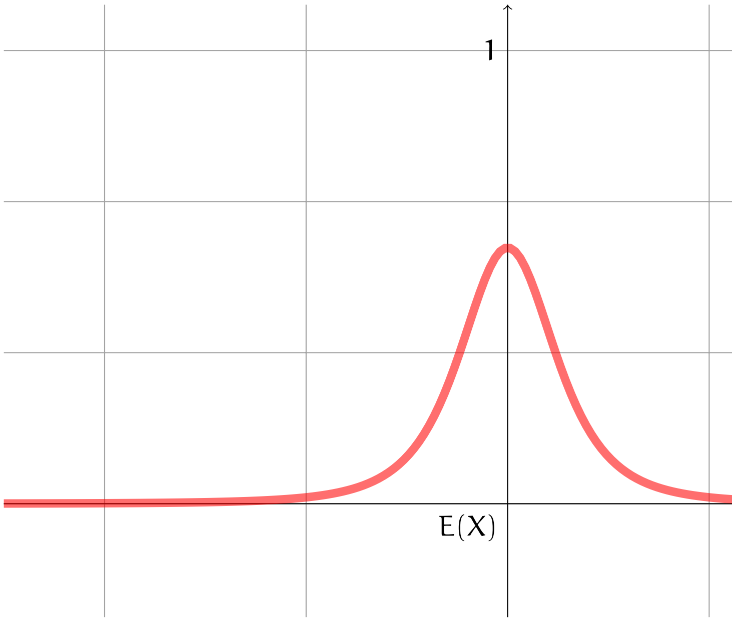
Support. $\mathrm{Supp}(X) = \mathbb{R}$

Densité : $p(x) = \dfrac{1}{\sqrt{n\pi}} \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \dfrac{x^2}{n}\right)^{-\frac{n+1}{2}}$

Répartition : Existe, mais de forme complexe.

Espérance : $E(X) = \begin{cases} \text{indéterminée} & \text{si } n = 1 \\ 0 & \text{si } n > 1 \end{cases}$

Variance : $V(X) = \begin{cases} \text{indéterminée} & \text{si } n = 1 \\ +\infty & \text{si } n = 2 \\ \dfrac{n}{n-2} & \text{si } n > 2 \end{cases}$



4. Statistiques unidimentionnelles

Introduction

Voici une définition de statistique :

> La statistique est d'un point de vue théorique une science, une méthode et une technique. La statistique comprend : la collecte des données, le traitement des données collectées, l'interprétation des données et la présentation afin de rendre les données compréhensibles par tous.

Dans ce cours nous nous intéresserons à un type de statistique en particulier : celle qui n'étudie qu'un caractère ou qu'une variable à la fois ; on parle alors de statistique unidimensionnelle.

Définition

Une étude statistique unidimensionnelle porte sur une caractéristique bien définie que l'on désigne par **caractère** ou **variable** et qui est présente chez chacun des éléments ou individus d'un ensemble donné appelé **population**.

Par exemple la population peut être les étudiants d'une classe et le caractère peut être les notes à l'examen de fin d'année.

On distingue deux types de caractères.

Définition

Une variable, ou caractère, statistique est dite **qualitative** si ses valeurs s'expriment de façon littérale ou par un codage sur lequel les opérations arithmétiques n'ont pas de sens.

Par exemple le sexe des personnes interrogées, le numéro de leur département de naissances (bien que cela soit des nombres et que les opérations arithmétiques usuelles soient valides, il n'y a aucun sens à considérer la somme de numéro de département ou la moyenne de ces numéros ; il s'agit ici d'un codage), leur situation familiale, la mention recalé, passable, assez bien, bien et très bien que peut avoir un étudiant à un examen. Dans ce dernier exemple on dit que le caractère est *ordinal* car on peut tout de même ordonner les valeurs du caractères. Dans les autres exemples, on parle de caractère, ou variable, *nominale* (ne sont décrit que par leur nom).

Définition

Une variable, ou caractère, statistique est dite **quantitative** si ses valeurs sont des nombres sur lesquels les opérations arithmétiques ont un sens. Elle peut être de deux formes :

- **Discrète :** si elle ne prend qu'un nombre fini de valeur. Ces valeurs sont appelées des **modalités**.
- **Continue :** si elle prend ses valeurs dans un intervalle. Ces intervalles sont appelées des **classes**.

Définition

Une série statistique est l'ensemble des modalités ou classes correspondant à tous les individus de la population considérée.

Série statistique à caractère discret

Dans la suite de ce chapitre, on fixe une série statistique à caractère discret s. Cela signifie que s est un ensemble fini de nombres réels. Il existe donc des nombres $\alpha_i \in \mathbb{R}$ tel que $s = \{\alpha_1, \alpha_2, ..., \alpha_n\}$

On note k le nombre de modalité différente et $x_1, x_2, ..., x_k$ ces différentes modalités ordonnées dans l'ordre permettant au mieux d'observer la série (dans la plupart des cas c'est dans l'ordre croissant).

Pour illustrer les définitions et notions nous utiliserons l'exemple suivant jusqu'à la fin du chapitre :

- La population étudiée est un groupe de TD de 30 étudiants.

- Le caractère étudié est les résultats obtenus à l'examen de mathématiques. Les notes, sur 20, sont les suivantes :

12	11	7	10	9	3
12	15	8	8	14	11
7	2	0	18	11	14
16	11	9	12	11	11
15	10	15	7	14	10

Le nombre de modalité différente est de 13 (*i.e.* $k = 13$ tandis que $n = 30$) et les différentes modalités sont $x_1 = 0$, $x_2 = 2$, $x_3 = 3$, $x_4 = 7$, $x_5 = 8$, $x_6 = 9$, $x_7 = 10$, $x_8 = 11$, $x_9 = 12$, $x_{10} = 14$, $x_{11} = 15$, $x_{12} = 16$, $x_{13} = 18$. A noter que toute modalité est une valeur mais toute valeur n'est pas une modalité. Par exemple 12 est une valeur et aussi une modalité mais 17 est une valeur sans être une modalité ; 17 est une valeur pour le caractère (une note) mais n'est pas une modalité de la série statistique car aucun des x_i ne vaut 17.

Effectif et fréquence

Définition

Le nombre d'élément de la série s est appelé **l'effectif total** de la série statistique s.

Définition

Soit x_i une modalité de la série statistique s. Le nombre n_i de répétition de x_i dans la série s est appelé **l'effectif** de x_i.

Dans la pratique, on représente ces résultats dans un tableau :

Caractères	x_1	x_2	$\cdots$	x_k
Effectifs	n_1	n_2	$\cdots$	n_k

Dans notre exemple, l'effectif total vaut 30 et les effectifs sont :

Notes	0	2	3	7	8	9	10	11	12	14	15	16	18
Effectifs	1	1	1	3	2	2	3	6	3	3	3	1	1

Par construction on a la proposition suivante.

Proposition

Notons n_i l'effectif de la modalité x_i.

$$n = \sum_{i=1}^{k} n_i$$

On vérifie en effet que dans notre exemple $1 + 1 + 1 + 3 + 2 + 2 + 3 + 6 + 3 + 3 + 3 + 1 + 1 = 30$.

Définition

Soit x_i une modalité de la série statistique s. On appelle **fréquence** relative à la modalité x_i le rapport de l'effectif de la modalité x_i avec l'effectif total.

$$f_i = \frac{n_i}{n}$$

Naturellement puisque la somme des effectifs vaut l'effectif total, la somme des fréquence, vaut 1 :

$$\sum_{i=1}^{k} f_i = 1$$

Ce dernier résultat montre en fait que la somme des $p_i = 100f_i$ fait 100 et donc que les p_i décrivent le pourcentage de l'effectif total ayant x_i pour caractère. On complète alors le tableau :

Caractères	x_1	x_2	$\cdots$	x_k
Effectifs	n_1	n_2	$\cdots$	n_k
Fréquences	f_1	f_2	$\cdots$	f_k
Pourcentages	p_1	p_2	$\cdots$	p_k

Ce qui donne dans notre exemple :

Notes	0	2	3	7	8	9	10	11	12	14	15	16	18
Effectifs	1	1	1	3	2	2	3	6	3	3	3	1	1
Fréquences	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{1}{10}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{30}$	$\frac{1}{30}$
Pourcentages	$\frac{10}{3}\%$	$\frac{10}{3}\%$	$\frac{10}{3}\%$	10%	$\frac{20}{3}\%$	$\frac{20}{3}\%$	10%	20%	10%	10%	10%	$\frac{10}{3}\%$	$\frac{10}{3}\%$

On interprète cela en observant, par exemple, que 20% des étudiants on obtenu un $11/20$ à leur examen.

Définition

L'**effectif cumulé** croissant (resp. décroissant) pour la modalité x_i est la somme des effectifs qui lui sont inférieurs (resp. supérieurs).

$$N_i = \textstyle\sum_{j=1}^{i} n_j$$
$$(\text{resp.} \quad N_i = \textstyle\sum_{j=i}^{k} n_j)$$

On observe en particulier que $N_k(\text{resp. } N_1) = \displaystyle\sum_{j=1}^{k} n_j = n$ (l'effectif total). On complète le tableau :

Caractères	x_1	x_2	$\cdots$	x_k
Effectifs	n_1	n_2	$\cdots$	n_k
Effectifs cumulés croissants	n_1	$n_1 + n_2$	$\cdots$	$n_1 + \cdots + n_k$
Effectifs cumulés décroissants	$n_1 + \cdots + n_k$	$n_2 + \cdots + n_k$	$\cdots$	n_k

Avec notre exemple :

Notes	0	2	3	7	8	9	10	11	12	14	15	16	18
Effectifs	1	1	1	3	2	2	3	6	3	3	3	1	1
Effectifs cumulés croissants	1	2	3	6	8	10	13	19	22	25	28	29	30
Effectifs cumulés décroissant	30	29	28	25	23	21	18	12	9	6	3	2	1

On peut interpréter ces résultats en observant, par exemple, que 10 étudiants ont obtenu une note strictement inférieure à 10.

Définition

La **fréquence cumulée** croissante (resp. décroissante) pour la modalité x_i est la somme des fréquences qui lui sont inférieures (resp. supérieures).

$$F_i = \sum_{j=1}^{i} f_j$$
$$\left(\text{resp.} \quad N_i = \sum_{j=i}^{k} f_j\right)$$

En générale on considèrera davantage les pourcentages que les fréquences en posant $P_i = \sum_{j=1}^{i} p_j$ (resp. $P_i = \sum_{j=i}^{K} p_j$). On les représente de même dans le tableau ce qui donne dans notre exemple :

Notes	0	2	3	7	8	9	10	11	12	14	15	16	18
Effectifs	1	1	1	3	2	2	3	6	3	3	3	1	1
Fréquences	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{1}{10}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{30}$	$\frac{1}{30}$
Pourcentages	$\frac{10}{3}\%$	$\frac{10}{3}\%$	$\frac{10}{3}\%$	10%	$\frac{20}{3}\%$	$\frac{20}{3}\%$	10%	20%	10%	10%	10%	$\frac{10}{3}\%$	$\frac{10}{3}\%$
Effectifs cumulés croissants	1	2	3	6	8	10	13	19	22	25	28	29	30
Fréquences cumulés croissantes	$\frac{1}{30}$	$\frac{1}{15}$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{4}{15}$	$\frac{1}{3}$	$\frac{13}{30}$	$\frac{19}{30}$	$\frac{11}{15}$	$\frac{5}{6}$	$\frac{14}{15}$	$\frac{29}{30}$	1
Pourcentages cumulés croissants	$\frac{10}{3}\%$	$\frac{20}{3}\%$	10%	20%	$\frac{80}{3}\%$	$\frac{100}{3}\%$	$\frac{130}{3}\%$	$\frac{190}{3}\%$	$\frac{220}{3}\%$	$\frac{250}{3}\%$	$\frac{280}{3}\%$	$\frac{290}{3}\%$	100%
Effectifs cumulés décroissant	30	29	28	25	23	21	18	12	9	6	3	2	1
Fréquences cumulés décroissantes	1	$\frac{29}{30}$	$\frac{14}{15}$	$\frac{5}{6}$	$\frac{23}{30}$	$\frac{7}{10}$	$\frac{3}{5}$	$\frac{2}{5}$	$\frac{3}{10}$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{1}{15}$	$\frac{1}{30}$
Pourcentages cumulés décroissants	100%	$\frac{290}{3}\%$	$\frac{280}{3}\%$	90%	80%	$\frac{220}{3}\%$	$\frac{200}{3}\%$	$\frac{170}{3}\%$	$\frac{110}{3}\%$	$\frac{80}{3}\%$	$\frac{50}{3}\%$	$\frac{20}{3}\%$	$\frac{10}{3}\%$

On interprète cela en observant, par exemple, que 80% des étudiants on obtenu une note supérieur ou égale à 8.

Représentation des données

Il existe plusieurs manière de représenter une série statistique à caractère discret.

Diagramme en bâtons. On trace les segments $\left\{\left[(x_i, n_i); (x_i, 0)\right]\right\}_{i \in [\![1;k]\!]}$ où les x_i désignent les modalités et n_i les effectifs associés. Avec notre exemple cela donne :

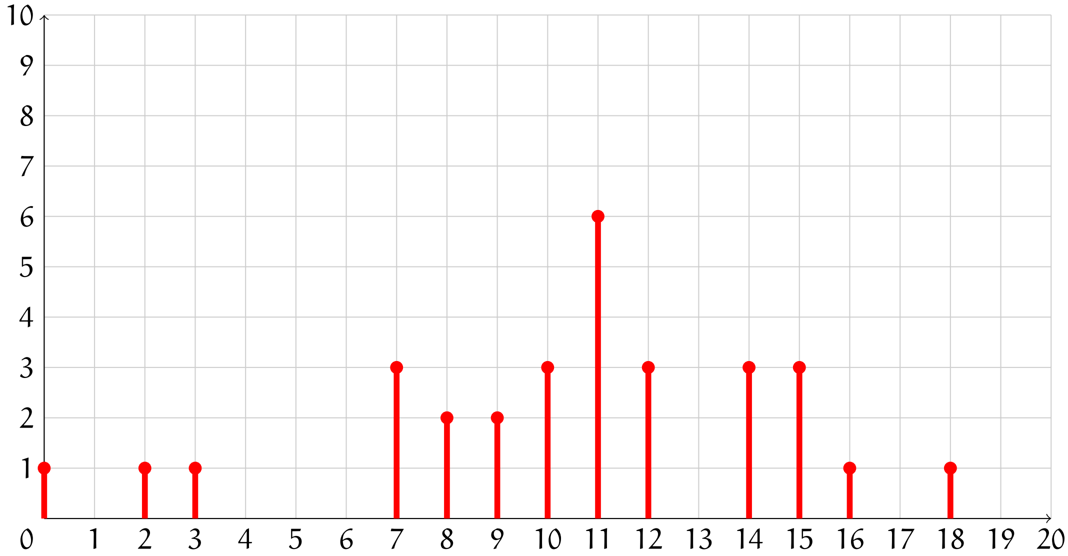


Diagramme en tuyau d'orgue. On procède comme le diagramme en bâtons à ceci près que l'on dessine des rectangles pour chaque modalité ; pour ne pas confondre avec les histogrammes (dont nous parlerons plus loin) on marque un espace entre chaque rectangle. Pour mieux illustrer la statistique, on peut indiquer les effectifs au dessus des rectangles. Dans notre exemple cela donne :

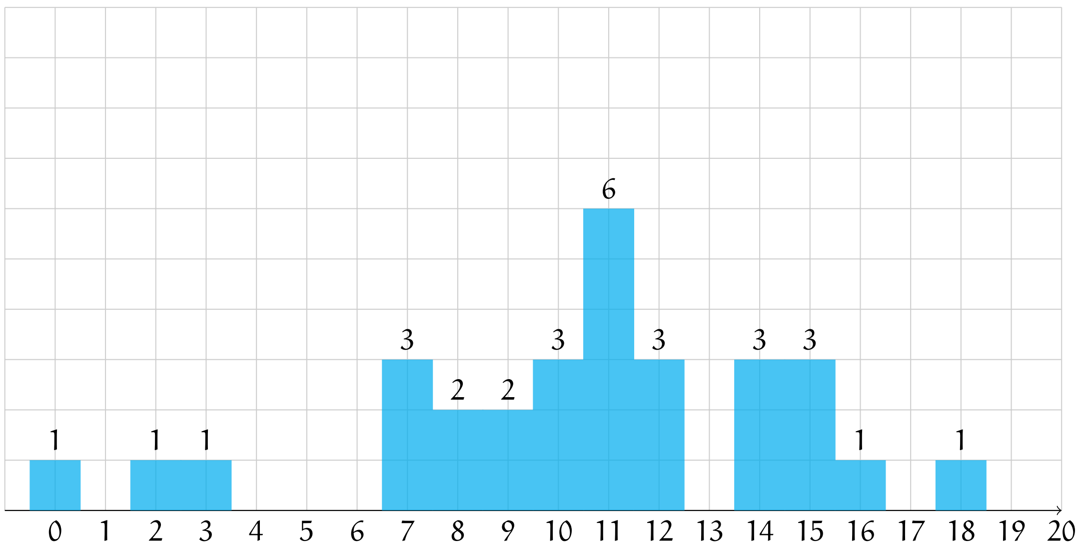


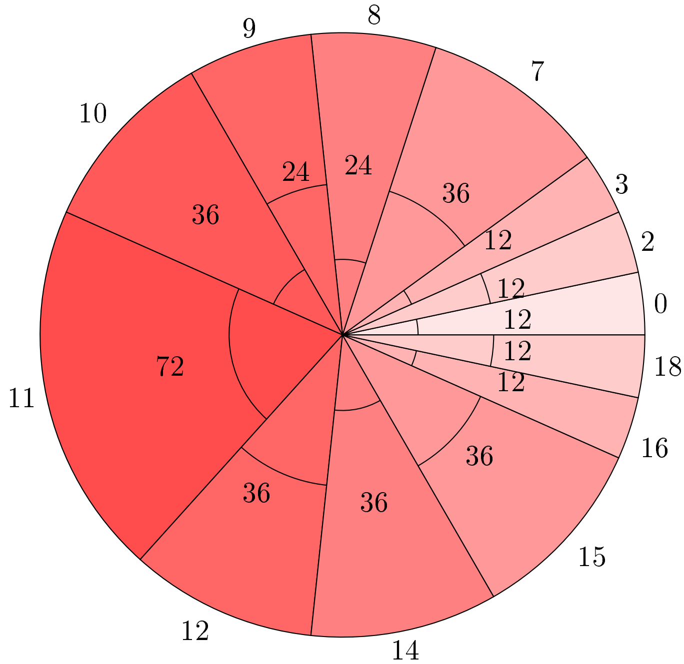
Diagramme circulaire. Pour chaque modalité x_i, on détermine l'angle en degré correspondant par la formule $\vartheta_i = n_i \frac{360}{n}$ où n désigne l'effectif total et n_i l'effectif de la modalité x_i.

Puisque la somme des ϑ_i vaut 360 chaque angles correspond à une partie d'un disque. On représente alors ces angles dans un disque en indiquant à quelle modalité correspond l'angle.

Dans notre exemple, on commence tout d'abord à déterminer les angles, en arrondissant à l'unité (et en s'arrangeant pour la somme des angles fasses bien 360 degrés).

Notes	0	2	3	7	8	9	10	11	12	14	15	16	18
Effectifs	1	1	1	3	2	2	3	6	3	3	3	1	1
Angles	12	12	12	36	24	24	36	72	36	36	36	12	12

Le diagramme circulaire correspondant est alors :



Caractéristiques de position

Le mode.

Définition

Le **mode** de S est la modalité avec le plus grand effectif.

Dans notre exemple le mode vaut 11.

La moyenne.

Définition

La **moyenne** de s, notée $\overline{s}$, est définie par la formule

$$\overline{s} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$$

où les n_i désignent l'effectif de la modalité x_i et n l'effectif total.

Dire qu'une statistique a $\overline{s}$ pour moyenne s'interprète en observant que c'est comme si tous les individus de la population étudiée avaient pour modalité $\overline{s}$.

Dans notre exemple, la moyenne vaut $\dfrac{313}{30} = 10.4\underline{3}$

La médiane. La médiane est la modalité qui sépare l'effectif en deux.

Définition

Soit $i_2 \in [\![1;k]\!]$ l'indice tel que $N_{i_2-1} < \dfrac{n}{2} \leqslant N_{i_2}$ où n désigne l'effectif total et N_i l'effectif cumulé croissant de la modalité x_i. La modalité x_{i_2} est appelé la **médiane** de la série s.

Il se peut que la médiane soit exactement entre deux modalités ; dans ce cas, on définit la médiane comme étant la valeur moyenne de ces deux modalités.

Dans notre exemple la médiane vaut 11. Cela s'interprète en observant que environ (c'est en effet une approximation car plusieurs individu peuvent avoir la modalité de la médiane) la moitié des étudiants ont obtenus une note inférieur à 11 et l'autre moitié supérieur à 11.

Les quantiles. La médiane sépare l'effectif en deux. On peut généraliser cette décomposition en remplaçant 2 par un autre nombre.

Définition

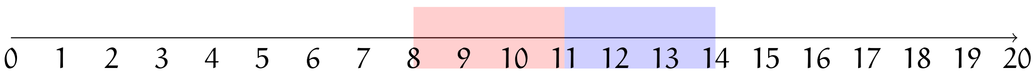
Soient $Q \in \mathbb{N}_{>1}$ et $q \in [\![1; Q-1]\!]$. Le $q^{\text{ième}}$ **quantile d'ordre Q** est la modalité x_{i_q} dont l'indice est tel que $N_{i_q-1} < \dfrac{n}{Q} \leqslant N_{i_q}$ où n désigne l'effectif total et N_i l'effectif cumulé croissant de la modalité x_i.

Dans la pratique trois quantiles sont étudiés :

La médiane. C'est le premier quantile d'ordre 2.

Les quartiles. On choisit de séparer l'effectif en quatre ($Q = 4$). Dans ce cas, le second quantile d'ordre 4 est la médiane. On s'attarde alors à calculer le premier quantile et le troisième quantile d'ordre 4 respectivement nommé **premier quartile** et **troisième quartile**. On représente généralement les quartiles dans un **diagramme en boîte** (également appelé boîte à moustache) : sur un axe représentant les modalités, on trace un rectangle dont deux des cotés opposés marquent respectivement le premier et le dernier quartile. On marque aussi la médiane.

Dans notre exemple, le premier quartile vaut 8 et le troisième 14.



Les déciles. On prend $Q = 10$.

Caractéristiques de dispersion

L'étendue.

Définition

L'**étendue** et_S d'une série statistique s est la différence entre le plus grande modalité et la plus petite.

$$et_s = \text{Max}(x_i | x_i \in s) - \text{Min}(x_i | x_i \in s).$$

L'étendue permet de mesurer si la série statistique est concentrée autour de sa moyenne ou plutôt dispersée : plus l'étendue est petite plus la série est concentré autour de sa moyenne et inversement. Dans notre exemple l'étendue est de 18. Cette série est donc dispersée autour de sa moyenne.

L'intervalle inter-quartile. Dans notre exemple, l'étendue de 18 nous indique que la série statistique est dispersée autour sa moyenne. L'intervalle inter-quartile permet de savoir s'il y a plus de modalité au dessus de la moyenne ou en dessous.

Définition

L'**intervalle inter-quartile** d'une série statistique est la différence entre le troisième et le premier quartile.

La variance et l'écart-type. Pour mieux observer la dispersion des modalités, on calcul l'écart-type. On va étudier les écarts entre chaque modalité avec la moyenne.

Définition

La **variance** d'une série statistique s est le nombre

$$v_s = \frac{1}{n} \sum_{i=1}^{k} n_i (x_i - \overline{s})^2$$

Dans la pratique on calcul la variance à l'aide de la formule suivante.

Proposition

Soit S une série statistique. Considérons S^2 la série ou toutes les modalités sont mis au carré. Alors

$$v_s = \overline{s^2} - \overline{s}^2 = \left(\frac{1}{n}\sum_{i=1}^{k} n_i x_i^2\right) - \overline{s}^2$$

Démonstration.

$$\begin{aligned}
v_S &= \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{S})^2 \\
&= \frac{1}{n}\sum_{i=1}^{k} n_i(x_i^2 - 2x_i\overline{S} + \overline{S}^2) \\
&= \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \frac{1}{n}\sum_{i=1}^{k} n_i 2x_i\overline{S} + \frac{1}{n}\sum_{i=1}^{k} n_i\overline{S}^2 \\
&= \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - 2\overline{S}\frac{1}{n}\sum_{i=1}^{k} n_i x_i + \overline{S}^2\frac{1}{n}\sum_{i=1}^{k} n_i \\
&= \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - 2\overline{S}\overline{S} + \overline{S}^2 \\
&= \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \overline{S}^2.
\end{aligned}$$

□

Pour "renormaliser" cette donnée (le passage au carré), on considère plus souvent l'écart-type.

Définition

L'**écart-type** d'une série statistique est définie comme la racine carré de la variance :

$$\sigma_s = \sqrt{v_s}$$

Série statistique à caractère continue

En général, les deux raisons principales qui peuvent amener à considérer comme continue une variable sont le grand nombre d'observation distinctes (trop grand pour une étude discrète) ou le caractère sensible d'une variable (salaire, age d'une femme, etc).

Dans ce chapitre on fixe une série statistique à caractère continue s. On note k le nombre de classe et chaque classe sera noté $[b_i; b_{i+1}[$ (les intervalles pouvant être fermés ou ouverts ; la seule règle à respecter est qu'une valeur ne peut être considérée que dans une seule classe). Pour illustrer les notions de ce chapitre nous considèrerons l'age des 121 employés d'une entreprise

26	22	41	43	18	31	34	28	26	21	44
52	60	62	34	38	23	31	40	58	60	33
33	26	28	30	29	29	29	29	33	35	33
26	42	24	22	44	41	47	30	49	32	37
26	51	28	55	52	61	47	22	19	27	25
35	33	25	34	43	42	41	30	29	27	51
52	31	32	29	25	21	31	41	21	31	51
32	22	42	52	23	44	50	51	29	29	29
28	27	29	35	43	49	57	57	57	31	33
33	48	49	22	18	19	20	21	22	23	23
23	19	44	55	33	48	28	42	54	25	29

Le nombre de modalité étant grand, on choisit une étude continue. On représente alors les données dans un tableau.

Classe	$[18;23[$	$[23;28[$	$[28;33[$	$[33;38[$	$[38;43[$	$[43;48[$	$[48;53[$	$[53;58[$	$[58;63]$
Effectif	16	18	28	15	10	9	14	6	5

Liens avec le cas discret

Définition

Soit s une série statistique à caractère continue.

- La **borne inférieur** de la classe $[b_i;b_{i+1}[$ est b_i.
- La **borne supérieur** de la classe $[b_i;b_{i+1}[$ est b_{i+1}.
- Le **centre de classe** de $[b_i;b_{i+1}[$ est $\dfrac{b_i+b_{i+1}}{2}$.
- L'**amplitude** de la classe $[b_i;b_{i+1}[$ est $b_{i+1}-b_i$.

Dans la pratique, on complète le tableau en rajoutant le centre des classes.

Classe	$[18;23[$	$[23;28[$	$[28;33[$	$[33;38[$	$[38;43[$	$[43;48[$	$[48;53[$	$[53;58[$	$[58;63]$
Centre des classes	20.5	25.5	30.5	35.5	40.5	45.5	50.5	55.5	60.5
Effectif	16	18	28	15	10	9	14	6	5

Définition

La **série statistique discrète associé** à s est la série dont les modalités sont les centres de classe et les effectifs correspondant aux classes respectives.

On peut donc appliquer dans ce cadre les définitions d'effectifs, effectif total, effectifs cumulés, fréquences, fréquences cumulées.

Représentations des données

En considérant les statistiques continues de manière discrète, on peut utiliser les représentations introduites dans le précédent chapitre : diagramme en batons, en tuyau d'orgue, circulaire etc.

Il existe également une représentation propre au caractère continue : l'histogramme.

Chaque classe est représenté par un rectangle dont la base est délimitée par les bornes correspondante et dont la hauteur est la densité d'effectif.

Définition

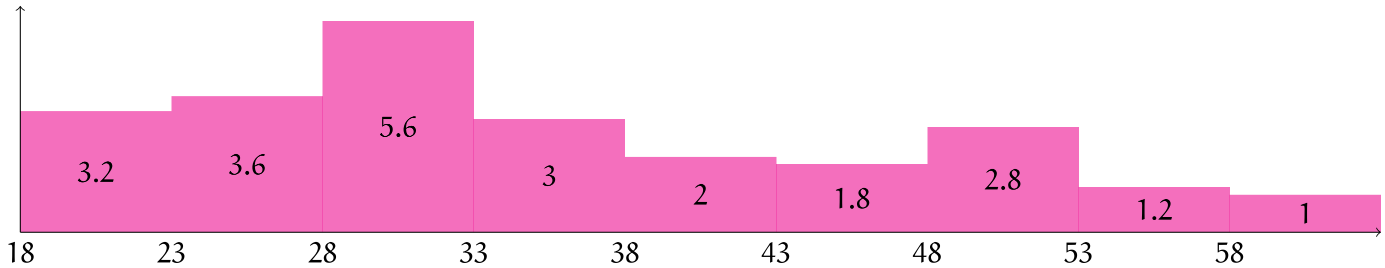
Soit S une série statistique à caractère continue. La **densité d'effectif** de la classe $[b_i; b_{i+1}[$ est le rapport entre l'effectif du centre de classe correspondant par l'amplitude de la classe.

$$\frac{n_i}{b_{i+1} - b_i}$$

Puisque la hauteur d'un rectangle est la densité d'effectif, l'aire d'un rectangle de l'histogramme, qui est le produit de la hauteur $\dfrac{n_i}{b_{i+1} - b_i}$ par la longueur $b_{i+1} - b_i$, est égale à l'effectif ; ceci permet donc une meilleur illustration de la série étudiée.

Avec notre exemple, cela donne :

Classe	$[18;23[$	$[23;28[$	$[28;33[$	$[33;38[$	$[38;43[$	$[43;48[$	$[48;53[$	$[53;58[$	$[58;63]$
Centre des classes	20.5	25.5	30.5	35.5	40.5	45.5	50.5	55.5	60.5
Effectif	16	18	28	15	10	9	14	6	5
Densité d'effectif	3.2	3.6	5.6	3	2	1.8	2.8	1.2	1



Caractéristiques de position et de dispersion liés au cas discret

Classe modale. La version continue du mode est la classe modale.

Définition

La **classe modale** de s est la classe du plus grand effectif.

Dans notre exemple, la classe modale est $[28;33[$.

De manière équivalente, la classe modale d'une série continue est la classe correspondant au mode de la série discrète associée.

Moyenne, variance et écart-type. La moyenne (resp. la variance, resp. l'écart-type) d'une série statistique continue, est la moyenne (resp. la variance, resp. l'écart-type) de la série statistique discrète associée.

Étendue.

Définition

L'**étendue** de s est la différence entre la plus grande borne supérieur et la plus petite borne inférieur.

$$et_s = \mathrm{Max}\Big(\mathrm{Sup}([b_i; b_{i+1}[\,\big|\, i \in [\![1;k]\!])\Big) - \mathrm{Min}\Big(\mathrm{Inf}([b_i; b_{i+1}[\,\big|\, i \in [\![1;k]\!])\Big).$$

Médiane et quantiles

Définition

Soient $Q \in \mathbb{N}_{>1}$ et $q \in [\![1; Q-1]\!]$. La $q^{\text{ième}}$ **classe-quantile d'ordre Q** est la classe $[b_{i_q}; b_{i_q+1}[$ dont l'indice i_q est tel que $N_{i_q-1} < \frac{n}{Q} \leqslant N_{i_q}$ où n désigne l'effectif total et N_i l'effectif cumulé croissant de la classe $[b_i; b_{i+1}[$.

Avec notre exemple :

Classe	$[18;23[$	$[23;28[$	$[28;33[$	$[33;38[$	$[38;43[$	$[43;48[$	$[48;53[$	$[53;58[$	$[58;63]$
Centre des classes	20.5	25.5	30.5	35.5	40.5	45.5	50.5	55.5	60.5
Effectif	16	18	28	15	10	9	14	6	5
Effectif cumulé croissant	16	34	62	76	87	96	110	116	121

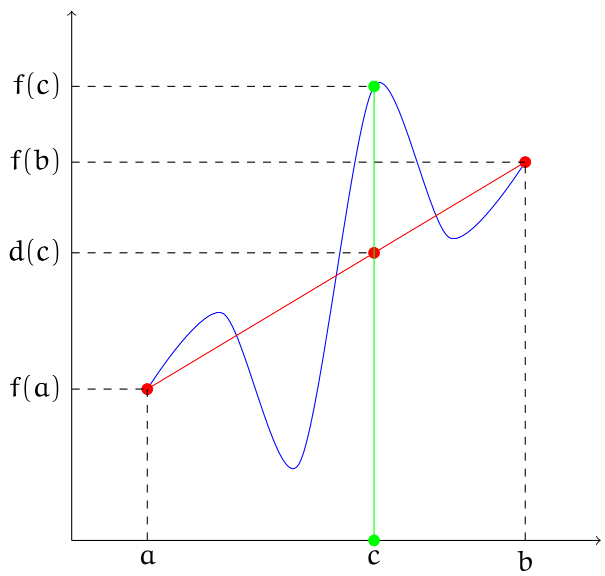
Dans ce cas la classe médiane est $[28;33[$. Le centre de classe étant 30.5, on pourrait dire que la médiane de cette série est 30.5. On peut cependant raffiner ce résultat par interpolation linéaire.

Interpolation Linéaire. Soient $a < b$ deux nombres réels et $f : [a, b] \to \mathbb{R}$ une fonction continue dont on connait les valeurs de $f(a)$ et $f(b)$. La question est d'approximer la valeur $f(c)$ pour $c \in [a, b]$.

Pour cela on va considérer la droite d passant par les points $(a, f(a))$ et $(b, f(b))$ et répondre $f(c) \simeq d(c)$. L'équation de la droite d est :

$$\frac{f(b) - f(a)}{b - a}(x - a) + f(a).$$

Ainsi $f(c) \simeq \frac{f(b) - f(a)}{b - a}(c - a) + f(a)$.



En utilisant le principe de l'interpolation linéaire, on peut approximer la valeur de la médiane. Dans notre exemple, la classe médiane est $[28;33[$. Il y a 34 individus qui ont moins de 28 ans et 62 moins de 33. Cela définit les points pour l'interpolation linéaire. Le point à interpoler étant $60.5 = \frac{121}{2}$. En appliquant la formule, on obtient que la médiane de cette série est

$$\frac{33 - 28}{62 - 34}(60.5 - 34) + 28 \simeq 32.7$$

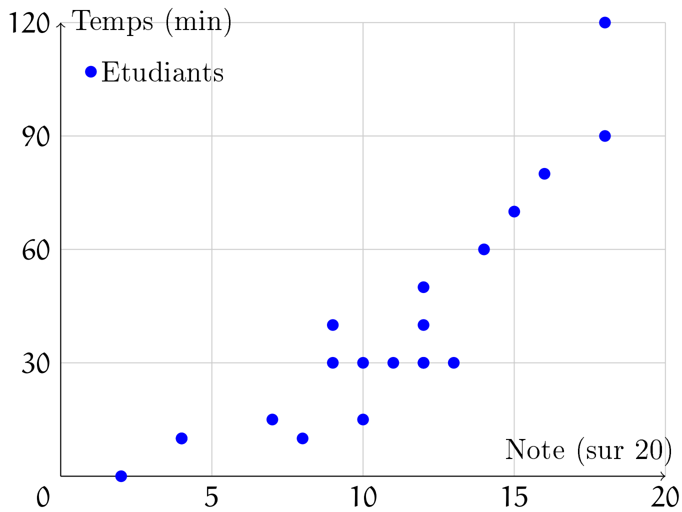
On raisonne de même pour les autres quantiles.

5. Modélisation linéaire simple

Exemple introductif

On considère dans une classe de 20 étudiants la note obtenues à l'examen de statistique (sur 20) et au temps de travail personnel par semaine (en minutes) dans cette matière. Les données sont synthétisées dans le tableau suivant :

Elève	Note	Temps
1	12	30
2	10	15
3	7	15
4	8	10
5	2	0
6	4	10
7	12	30
8	14	60
9	13	30
10	4	10
11	16	80
12	18	90
13	18	120
14	12	50
15	11	30
16	10	30
17	9	30
18	9	40
19	12	40
20	15	70



On note x la donnée statistique correspondant à la note et y celle du temps de travail.

Pour observer ces données on les a représenté dans un nuage de point ; les valeurs de x en abscisses et y en ordonnés.

On observe que plus la note est bonne, plus le temps de travail est important et inversement (surtout inversement).

Pour cela on considère le temps non pas comme une donnée statistique mais comme une variable aléatoire dont on cherche la loi. Nous nous plaçons dans le cadre d'un modèle linéaire simple.

Définition

Soient $x = (x_1, \ldots, x_n)$ et $y = (y_1, \ldots, y_n)$ des données statistiques.

La variable x est appelée **variable exogène** ou *expliquée*.

La variable y est appelée **variables endogène** ou *à expliquer*.

Une **modélisation linéaire simple** consiste à considérer les variables aléatoires

$$Y_i = ax_i + b + \varepsilon_i$$

où es ε_i sont des variables aléatoires i.i.d. appelés **termes d'erreurs** et suivent une loi normale $\mathcal{N}(0, \sigma)$.

Dans la pratique, on note par des lettres majuscules les variables aléatoires et en minuscules les données statistiques.

Remarque :

1. L'écart-type σ ne dépend pas de i (c'est à dire des réalisations). On parle d'homoscédasticité.

2. Le terme d'erreur est indépendant de la variable exogène.
3. $\forall i,\ Y_i \sim \mathcal{N}(ax_i + b, \sigma)$

Nous avons trois valeurs à estimer (a, b et σ) de tel sorte que les y_i s'écartent le moins possible de $ax_i + b$

La covariance

On rappel que si $x = (x_1, \ldots, x_n)$ est une donné statistique, on note $\overline{\alpha x + \beta} = \frac{1}{n}\sum_{i=1}^{n} \alpha x_i + \beta$ la moyenne. On montre très facilement que $\overline{\alpha x + \beta} = \alpha\overline{x} + \beta$. De la même manière, on rappel que l'espérance d'une variable aléatoire X est linéaire : $\mathbb{E}(\alpha X + \beta) = \alpha\mathbb{E}(X) + \beta$.

Ceci étant, définissons la covariance.

Définition

Soient X et Y des variables aléatoires. On définit la **covariance** de X et Y comme :

$$\mathrm{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

Soient $x = (x_1, \ldots, x_n)$ et $y = (y_1, \ldots, y_n)$ des données statistiques. On définit la **covariance** de x et y comme :

$$\sigma_{x,y} = \overline{(x - \overline{x})(y - \overline{y})}$$

Lemme :

Soient X, Y et Z des variables aléatoires réelles et α et β des nombres réels.

1. $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
2. $X \perp\!\!\!\perp Y \Rightarrow \mathrm{Cov}(X, Y) = 0$
3. $\mathrm{Cov}(X, X) = \mathbb{V}(X)$
4. $\mathrm{Cov}(X + \alpha, Y) = \mathrm{Cov}(X, Y)$
5. $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$
6. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
7. $\mathbb{V}(\alpha X + \beta Y) = \alpha^2\mathbb{V}(X) + 2\alpha\beta\mathrm{Cov}(X, Y) + \beta^2\mathbb{V}(Y)$

Soient $x = (x_1, \ldots, x_n)$ et $y = (y_1, \ldots, y_n)$ des données statistiques et α et β des nombres réels. Notons $xy = (x_1y_1, \ldots x_ny_n)$.

1. $\sigma_{x,y} = \overline{xy} - \overline{x}.\overline{y}$
2. $\sigma_{x,x} = \sigma_x^2$
3. $\sigma_{x+\alpha,y} = \sigma_{x,y}$
4. $\sigma_{x+y,z} = \sigma_{x,z} + \sigma_{y,z}$
5. $\sigma_{x,y} = \sigma_{y,x}$
6. $\sigma_{\alpha x+\beta y} = \alpha^2\sigma_x^2 + 2\alpha\beta\sigma_{x,y} + \beta^2\sigma_y^2$

Démonstration. Nous ne réalisons les démonstration que dans le cas probabiliste. Le cas déterministe (de la statistique) étant, aux notations près, le même.

La première formule se déduit de la linéarité de l'espérance :

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\
&= \mathbb{E}(XY - X\mathbb{E}(Y) - Y\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y)) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)
\end{aligned}$$

Pour la seconde viens du fait que si X et Y sont indépendants alors $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

La troisième et la quatrième découle trivialement des définitions, quand à la dernière elle vient de la linéarité de l'espérance :

$$\begin{aligned}
\mathbb{V}(\alpha X+\beta Y) &= \mathbb{E}\left((\alpha X+\beta Y-\mathbb{E}(\alpha X+\beta Y))^2\right)\\
&= \mathbb{E}\left((\alpha X+\beta Y-\alpha\mathbb{E}(X)-\beta\mathbb{E}(Y))^2\right)\\
&= \mathbb{E}\left((\alpha(X-\mathbb{E}(X))+\beta(Y-\mathbb{E}(Y)))^2\right)\\
&= \mathbb{E}\left(\alpha^2(X-\mathbb{E}(X))^2+2\alpha\beta(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))+\beta^2(Y-\mathbb{E}(Y))^2\right)\\
&= \alpha^2\mathbb{E}\left((X-\mathbb{E}(X))^2\right)+2\alpha\beta\mathbb{E}\left((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right)+\beta^2\mathbb{E}\left((Y-\mathbb{E}(Y))^2\right)\\
&= \alpha^2\mathbb{V}(X)+2\alpha\beta\mathrm{Cov}(X,Y)+\beta^2\mathbb{V}(Y)
\end{aligned}$$

□

Théorème Inégalité de Cauchy-Schwartz

Soient X et Y des variables aléatoires réelles.

$$\mathrm{Cov}(X,Y)^2 \leqslant \mathbb{V}(X)\mathbb{V}(Y)$$

Soient $x=(x_1,\dots,x_n)$ et $y=(y_1,\dots,y_n)$ des données statistiques.

$$\sigma_{x,y}^2 \leqslant \sigma_x^2\sigma_y^2$$

Démonstration. C'est la même preuve dans le cas déterministe et dans le cas probabiliste. Nous n'en faisons qu'une.

Pour tout nombre réel t, $\sigma_{tx+y}^2 \geqslant 0$. Or d'après le précédent lemme, $\sigma_{tx+y}^2 = t^2\sigma_x^2 + 2t\sigma_{x,y} + \sigma_y^2$. Ce polynôme de degré 2 est donc positif ou nul, donc son discriminant est négatif ou nul, c'est à dire $4\sigma_{x,y}^2 - 4\sigma_x^2\sigma_y^2 \leqslant 0$ ce qui prouve l'inégalité. □

EMCO : l'estimateur des moindres carrés ordinaires

On rappel que l'on cherche à estimer a, b et σ tel que pur tout i, $Y_i = ax_i + b + \varepsilon_i$ où $\varepsilon_i \simeq \mathcal{N}(0,\sigma)$. Bien sur pour que ce modèle soit le plus proche des valeurs observée, il faut que les réalisations y_i s'écartent le moins possible des $ax_i + b$. Un moyen est de passer par la mesure de l'erreur quadratique. On considère donc

$$EQ((x,y),(a,b)) = \sum_{i=1}^{n}\varepsilon_i^2 = \sum_{i=1}^{n}(y_i - ax_i - b)^2$$

Théorème

Soient $x=(x_1,\dots,x_n)$ et $y=(y_1,\dots,y_n)$ des données statistiques.

Le minimum de $\displaystyle\sum_{i=1}^{n}(y_i - ax_i - b)^2$ est atteint lorsque :

$$a = \hat{a} = \frac{\sigma_{x,y}}{\sigma_x^2} \qquad b = \hat{b} = \overline{y} - \hat{a}\overline{x}$$

La droite d'équation $d(x) = \hat{a}x + \hat{b}$ est appelé la **droite de régression linéaire**.

Démonstration. On cherche à minimiser $f(a,b) = \displaystyle\sum_{i=1}^{n}(y_i - ax_i - b)^2$. Il faut pour cela déterminer les valeurs pour lesquelles les dérivées partielles en a et en b s'annulent simultanément.

$$\begin{cases} \dfrac{\partial f}{\partial a} = 0 \\ \dfrac{\partial f}{\partial b} = 0 \end{cases} \Rightarrow \begin{cases} \displaystyle\sum_{i=1}^{n} -2x_i(y_i - ax_i - b) = 0 \\ \displaystyle\sum_{i=1}^{n} -2(y_i - ax_i - b) = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \displaystyle\sum_{i=1}^{n} x_iy_i - ax_i^2 - bx_i = 0 \\ \displaystyle\sum_{i=1}^{n} y_i - ax_i - b = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \overline{xy} - a\overline{x^2} - b\overline{x} = 0 \\ \overline{y} - a\overline{x} - b = 0 \end{cases}$$

La seconde équation montre qu'il suffit de trouver $\hat{a}$ pour trouver $\hat{b}$.

Si $\overline{x} \neq 0$. Dans ce cas, on peut multiplier la seconde ligne par $\overline{x}$ pour arriver à $\overline{x}.\overline{y} - a\overline{x}^2 - b\overline{x} = 0$. En la soustrayant à la première ligne, on arrive à

$$\underbrace{(\overline{xy} - \overline{x}.\overline{y})}_{\sigma_{x,y}} - a\underbrace{(\overline{x^2} - \overline{x})}_{\sigma_x^2} = 0$$

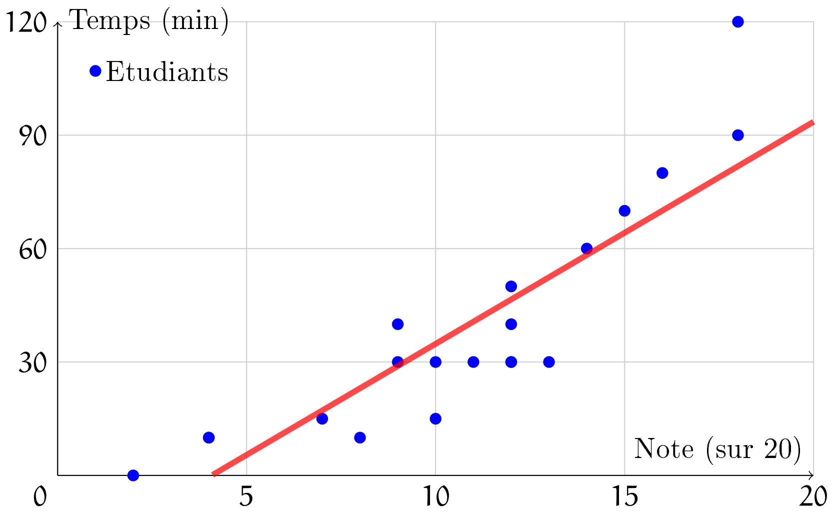
ce qui prouve le résultat.

Si $\overline{x} = 0$. Dans ce cas la première équation devient $\sigma_{x,y} - a\sigma_x^2 = 0$ ce qui prouve aussi le résultat.

□

Dans notre exemple (des notes et du temps de travail), la droite de régression linéaire est

$$d(x) = 5,8744582882x - 23,9441495125$$



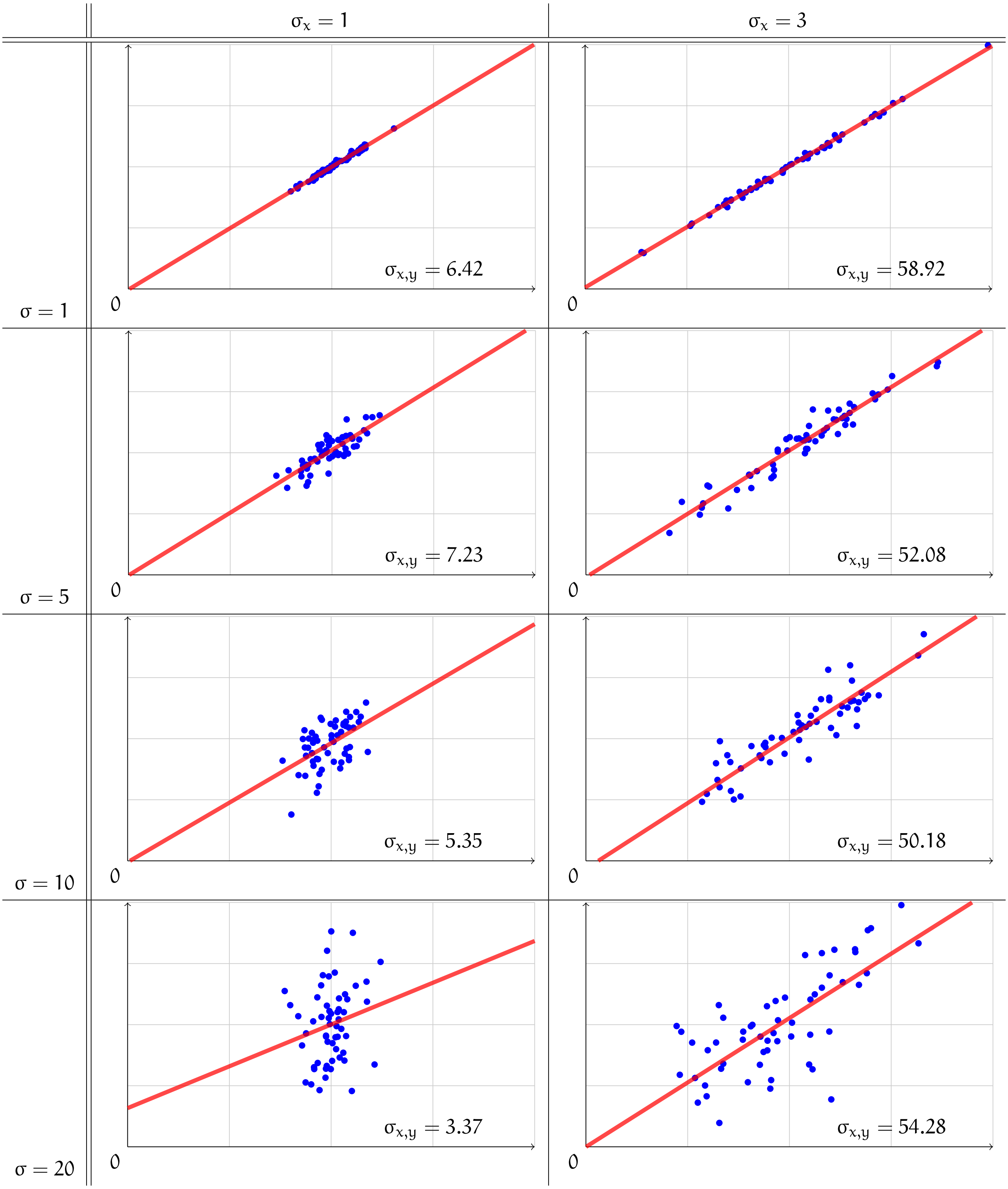
Il y a à peu près, autant de point au dessus qu'en dessous de la droite de régression linéaire.

Si on observe, on constate que $\hat{a}$ est inversement proportionnelle [4] à la variance de x. De sorte que plus cette variance est petite plus le coefficient directeur de la droite de régression linéaire sera grande

De même, par construction, plus σ sera petit, plus les point seront proche de la droite et inversement.

On peut faire ses observations sur les simulations suivantes :

4. Pas tout à fait car le numérateur fait aussi intervenir les données x.



Résidus et corrélation

Dans une modélisation linéaire simple $Y_i = ax_i + b + \varepsilon_i$ les variables aléatoires ε_i sont appelés les termes d'erreurs (du modèle) à ne pas confondre avec les résidus.

Définition

Soient $x = (x_1, \ldots, x_n)$ et $y = (y_1, \ldots, y_n)$ des données statistiques, $Y_i = ax_i + b + \varepsilon_i$ une modélisation linéaire simple et $\hat{a} = \dfrac{\sigma_{x,y}}{\sigma_x^2}$ et $\hat{b} = \overline{y} - \hat{a}\overline{x}$. Notons $\hat{y}_i = \hat{a}x_i + \hat{b}$.

On appel **résidus du modèle** les valeurs $\hat{\varepsilon}_i = y_i - \hat{y}_i$

Proposition

Les résidus d'une modélisations linéaire simple ont une moyenne nulle.

Démonstration.

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i &= \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{a}x_i - \hat{b} \\
&= \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{a}x_i - \overline{y} + \hat{a}\overline{x} \\
&= \overline{y} - \hat{a}\overline{x} - \overline{y} + \hat{a}\overline{x} \\
&= 0
\end{aligned}$$

□

Pour estimer les valeurs de a et de b (par $\hat{a}$ et $\hat{b}$) nous avons minimiser l'erreur quadratique du modèle. On peut chercher à mesurer la qualité de cette estimation en mesurant les résidus. Comme ils sont de moyenne nulle, on va regarder leur variance.

Définition

Avec les notations précédentes on définit le **coefficient de détermination du modèle**, noté R^2, par :

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

Proposition

Avec les notations précédentes

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

En particulier $R^2_{x,y} \in [0, 1]$.

Démonstration. Il suffit de montrer que $\sum_{i=1}^{n}(y_i-\overline{y})^2 = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2$

$$\begin{aligned}
\sum_{i=1}^{n}(y_i-\overline{y})^2 &= \sum_{i=1}^{n}(y_i-\hat{y}_i+\hat{y}_i-\overline{y})^2 \\
&= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}(y_i-\hat{y}_i)(\hat{y}_j-\overline{y}) + \sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2 \\
&= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + 2\sum_{i=1}^{n}(y_i-\hat{y}_i)\sum_{j=i+1}^{n}(\hat{y}_j-\overline{y}) + \sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2 \\
&= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + 2\underbrace{\sum_{i=1}^{n}\hat{\varepsilon}_i}_{=0}\sum_{j=i+1}^{n}(\hat{y}_j-\overline{y}) + \sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2 \\
&= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2
\end{aligned}$$

□

Remarque : Interprétation :

Si $R^2_{x,y}$ est proche de 1 alors $\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ est proche de 0 ce qui signifie que le modèle est très proche des valeurs : c'est un bon modèle.

Si $R^2_{x,y}$ est proche de 0 alors $\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2$ est proche de 0 et les $\hat{y}_i$ approche la moyenne : ce modèle n'est pas bon.

Définition

Soient $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$ des données statistiques.

Le coefficient de corrélation linéaire simple (ou de Pearson) de x et y est noté $r_{x,y}$ et est définit par

$$r_{x,y} = \frac{\sigma_{x,y}}{\sigma_x\sigma_y}$$

Le coefficient de corrélation linéaire multiple de x et y est noté $R_{x,y}$ et est définit par

$$R_{x,y} = \sqrt{R^2_{x,y}}$$

Proposition

Avec les notations précédentes

$$r_{x,y} = sg(\hat{a})R_{x,y}$$

où $sg(\hat{a})$ désigne le signe de $\hat{a}$.

Démonstration. On observe que $r_{x,y} = \frac{\sigma_{x,y}}{\sigma_x\sigma_y} = \frac{\sigma_{x,y}}{\sigma_x^2}\frac{\sigma_x}{\sigma_y} = \hat{a}\frac{\sigma_x}{\sigma_y}$. D'où

$$
\begin{aligned}
r_{x,y}^2 &= \hat{a}^2\frac{\sigma_x^2}{\sigma_y^2} \\
&= \hat{a}^2\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sigma_y^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{a}x_i-\hat{a}\overline{x})^2}{\sigma_y^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{a}x_i+\hat{b}-\hat{a}\overline{x}-\hat{b})^2}{\sigma_y^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i-\overline{y})}{\sigma_y^2} \\
&= R_{x,y}^2
\end{aligned}
$$

□

En définitive pour déterminer si un modèle linéaire simple est un bon modèle, on peut calculer au choix le coefficient de corrélation simple ou multiple et le comparer à 1. Plus ce coefficient est proche de 1, plus le modèle est bon.

Dans notre exemple, il vaut $78,23\%$. Le modèle choisi est plutôt bon.

Théorème de Gauss-Markov

Plaçons-nous dans un cadre un peut plus théorique. L'objectif de ce chapitre est de construire les EMCO et de montrer qu'il s'agit d'estimateur convergent. On fixe dans la suite des données statistique x et y dont on suppose avoir une infinité de réalisation (c'est juste pour la théorie, dans la pratique, c'est inutile et irréalisable) ainsi qu'un modèle linéaire simple $Y_i = ax_i + b + \varepsilon_i$ où les $\varepsilon_i \sim \mathcal{N}(0,\sigma)$ sont i.i.d. .

Proposition

Les estimateurs suivants

$$A_n = \frac{\sigma_{x,Y}}{\sigma_x^2} = a + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\frac{x_i-\overline{x}}{\sigma_x^2}$$

et

$$B_n = \overline{Y}_n - A_n\overline{x} = b + \overline{\varepsilon}_n + (A_n - a)\overline{x}$$

sont des estimateurs convergents et dans biais de a et b. De plus

$$\mathbb{V}(A_n) = \frac{\sigma^2}{n}\left(\frac{1}{\sigma_x^2}\right) \qquad \text{et} \qquad \mathbb{V}(B_n) = \frac{\sigma^2}{n}\left(1+\frac{\overline{x}^2}{\sigma_x^2}\right)$$

Démonstration.

$$
\begin{aligned}
\mathbb{V}(A_n) &= \mathbb{V}\left(a + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\frac{x_i-\overline{x}}{\sigma_x^2}\right) \\
&= \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\frac{x_i-\overline{x}}{\sigma_x^2}\right) \\
&= \frac{1}{n^2}\mathbb{V}\left(\sum_{i=1}^{n}\varepsilon_i\frac{x_i-\overline{x}}{\sigma_x^2}\right)
\end{aligned}
$$

$$\begin{aligned}\mathbb{E}(A_n) &= a+\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\frac{x_i-\overline{x}}{\sigma_x^2}\right)\\&= a+\frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n}\varepsilon_i\frac{x_i-\overline{x}}{\sigma_x^2}\right)\\&= a+\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\varepsilon_i)\frac{x_i-\overline{x}}{\sigma_x^2}\\&= a\end{aligned}$$

Ainsi A_n est un estimateur sans biais et de variance qui tend vers 0. D'après le cours, A_n est un estimateur convergent de a.

$$\begin{aligned}\mathbb{E}(B_n) &= \mathbb{E}\left(\overline{Y}_n-A_n\overline{x}\right)\\&= \mathbb{E}\left(\overline{Y}_n\right)-\mathbb{E}(A_n)\overline{x}\\&= a\overline{x}+b-a\overline{x}\\&= b\end{aligned}$$

$$\begin{aligned}\mathbb{V}(B_n) &= \mathbb{V}\left(\overline{Y_n}-A_n\overline{x}\right)\\&= \mathbb{V}\left(\overline{Y_n}\right)-2\overline{x}\mathrm{Cov}\left(\overline{Y_n},A_n\right)+\overline{x}^2\mathbb{V}(A_n)\end{aligned}$$

Or

- $$\begin{aligned}\mathbb{V}\left(\overline{Y_n}\right) &= \mathbb{V}\left(a\overline{x}+b+\overline{\varepsilon_n}\right)\\&= \mathbb{V}\left(\overline{\varepsilon_n}\right)\\&= \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\right)\\&= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(\varepsilon_i)\\&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2\\&= \frac{\sigma^2}{n}\end{aligned}$$

- $$\begin{aligned}\mathrm{Cov}\left(\overline{Y_n},A_n\right) &= \mathrm{Cov}\left(\overline{\varepsilon}_n,\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\frac{x_i-\overline{x}}{\sigma_x^2}\right)\\&= \frac{1}{n}\sum_{i=1}^{n}\frac{x_i-\overline{x}}{\sigma_x^2}\mathrm{Cov}(\overline{\varepsilon}_n,\varepsilon_i)\\&= \frac{1}{n}\sum_{i=1}^{n}\frac{x_i-\overline{x}}{\sigma_x^2}\frac{\sigma^2}{n}\\&= \frac{\sigma^2}{n\sigma_x^2}\frac{1}{n}\sum_{i=1}^{n}x_i-\overline{x}\\&= \frac{\sigma^2}{n\sigma_x^2}\times 0\\&= 0\end{aligned}$$

  Où on a remarqué que $\mathrm{Cov}(\overline{\varepsilon}_n,\varepsilon_i)=\dfrac{1}{n}\sum_{j=1}^{n}\mathrm{Cov}(\varepsilon_j,\varepsilon_i)$. Si $i\neq j$ alors par indépendance, $\mathrm{Cov}(\varepsilon_j,\varepsilon_i)=0$ et si $i=j$, $\mathrm{Cov}(\varepsilon_i,\varepsilon_i)=\mathbb{V}(\varepsilon_i)=\sigma^2$. D'où $\mathrm{Cov}(\overline{\varepsilon}_n,\varepsilon_i)=\dfrac{\sigma^2}{n}$.

Donc finalement

$$\begin{aligned}\mathbb{V}(B_n) &= \mathbb{V}\left(\overline{Y_n}-A_n\overline{x}\right)\\&= \mathbb{V}\left(\overline{Y_n}\right)+\overline{x}^2\mathbb{V}(A_n)\\&= \frac{\sigma^2}{n}+\overline{x}^2\frac{\sigma^2}{n}\frac{1}{\sigma_x^2}\\&= \frac{\sigma^2}{n}\left(1+\frac{\overline{x}^2}{\sigma_x^2}\right)\end{aligned}$$

Ainsi B_n est un estimateur sans biais et de variance qui tend vers 0. D'après le cours, B_n est un estimateur convergent de b.

□

On dis que ces estimateurs sont *BLUE* : *Best Linear Unbiased Estimator*. Ce sont les plus *efficace* : il est impossible d'obtenir des estimateurs avec une variance plus faible[5].

Estimation de σ

Avec nos hypothèses, les estimateurs A_n et B_n sont des combinaisons linéaires de loi normale donc ils suivent des lois normales.

L'estimateur A_n suit une loi normale de paramètre a et $\sigma(A_n)=\sigma\dfrac{1}{\sqrt{n}\sigma_x}$.

L'estimateur B_n suit une loi normale de paramètre b et $\sigma(B_n)=\sigma\dfrac{\sqrt{1+\dfrac{\overline{x}^2}{\sigma_x^2}}}{\sqrt{n}}$.

5. Ce qui est assez difficile à mathématiquement démontrer.

Dans un cas comme dans l'autre si l'on cherche à estimer les variances (ou les écart-types), il est nécessaire d'estimer σ.

Théorème

Soit $\hat{\varepsilon}_i = Y_i - \hat{y}_i = ax_i + b + \varepsilon_i - \hat{a}x_i - \hat{b}$.

$$\mathbb{E}\left(\sum_{i=1}^{n} \hat{\varepsilon}_i^2\right) = (n-2)\sigma^2$$

Démonstration. Admise □

Le 2 du $n-2$ viens du fait que deux paramètres ont été estimés : $\hat{a}$ et $\hat{b}$ ce qui diminue donc le degrés de liberté de cette égalité.

Corollaire

La variable aléatoire

$$S_n = \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{n-2}$$

est un estimateur convergent et sans biais de σ^2. En particulier :

1. $S_n^a = S_n \dfrac{1}{n\sigma_x^2}$ est un estimateur convergent et sans biais de $\mathbb{V}(A_n)$.
2. $S_n^b = S_n \dfrac{\sigma_x^2 + \overline{x}}{n\sigma_x^2}$ est un estimateur convergent sans biais de $\mathbb{V}(B_n)$.

De plus $(n-2)\dfrac{S_n}{\sigma^2} \sim \chi^2(n-2)$.

Démonstration. L'estimateur S_n est sans biais donc asymptotiquement sans biais et sa variance tend vers 0. Il est donc convergent. C'est le même raisonnement pour S_n^a et S_n^b.

Enfin $\varepsilon_i \sim \mathcal{N}(0, \sigma)$ de sorte que $\dfrac{\hat{\varepsilon}_i}{\sigma} \sim \mathcal{N}(0,1)$. Alors

$$(n-2)\frac{S_n}{\sigma^2} = \sum_{i=1}^{n} \frac{\hat{\varepsilon}_i^2}{\sigma^2}$$

est donc la somme du carré de loi normale centrée réduite. C'est une distribution du χ^2 (par définition) à $n-2$ degrés de libertés. □

Corollaire

$$\frac{A_n - a}{\sqrt{S_n^a}} \sim \mathcal{T}(n-2) \qquad \text{et} \qquad \frac{B_n - b}{\sqrt{S_n^b}} \sim \mathcal{T}(n-2)$$

Démonstration. On raisonne pour A_n, c'est un raisonnement identique pour B_n.

$$\frac{A_n - a}{\sqrt{S_n^a}} = \frac{\dfrac{A_n - a}{\sigma(A_n)}}{\dfrac{\sqrt{S_n^a}}{\sigma(A_n)}} = \frac{\mathcal{N}(0,1)}{\sqrt{\dfrac{S_n^a}{\mathbb{V}(A_n)}}} = \frac{\mathcal{N}(0,1)}{\sqrt{\dfrac{S_n \dfrac{1}{n\sigma_x^2}}{\dfrac{\sigma^2}{n}\dfrac{1}{\sigma_x^2}}}}$$

$$= \frac{\mathcal{N}(0,1)}{\sqrt{\dfrac{S_n}{\sigma^2}}} = \frac{\mathcal{N}(0,1)}{\sqrt{\dfrac{\sigma^2 \dfrac{\chi^2(n-2)}{n-2}}{\sigma^2}}} = \frac{\mathcal{N}(0,1)}{\sqrt{\dfrac{\chi^2(n-2)}{n-2}}}$$

□

Corollaire

Soient $0 < \beta < \alpha < 1$, $t_1 = Q_{\mathcal{T}(n-2)}(\beta)$ et $t_2 = Q_{\mathcal{T}(n-2)}(1 - \alpha + \beta)$ alors

$$\left[A_n - t_2\sqrt{S_n^a}; A_n - t_1\sqrt{S_n^a}\right]$$

est un intervalle de confiance $1 - \alpha$ de a et

$$\left[B_n - t_2\sqrt{S_n^b}; B_n - t_1\sqrt{S_n^b}\right]$$

est un intervalle de confiance de niveau $1 - \alpha$ de b.

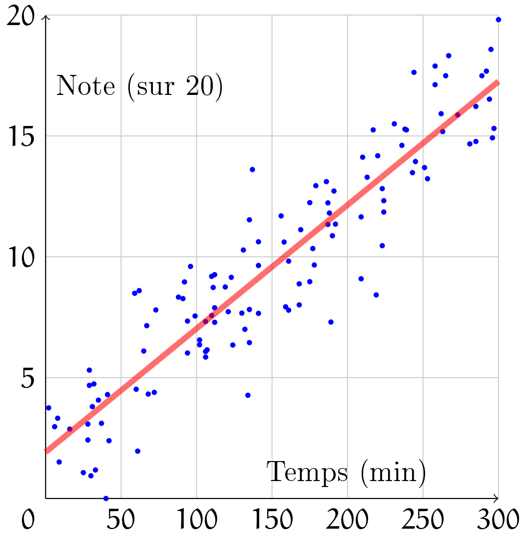
Exemple introductif

On dispose de données sur une promotion de 120 étudiants : le temps qu'ils passent à travailler, leur moyenne en mathématiques et leur moyenne en statistique :

#	Temps	Math	Stat
1	88	8.33	5.84
2	123	9.15	9.13
3	28	3.08	4.93
4	209	11.65	11.34
5	296	14.92	14.64
6	168	8.88	8.96
7	135	7.82	9.16
8	231	15.5	12.58
9	251	13.69	11.08
10	107	6.15	6.84
11	40	0	0
12	263	15.18	14.32
13	62	8.6	5.1
14	220	14.18	10.1
15	130	7.67	9.92
16	106	7.32	6.23
17	135	11.53	6.57
18	121	7.73	10.5
19	37	3.11	5.46
20	110	7.57	6.17
21	91	8.27	7.78
22	96	9.6	7.39
23	223	12.82	10.8
24	132	7.0	5.83
25	175	8.97	11.39
26	161	9.82	8.75
27	239	15.25	13.9
28	72	4.39	1.03
29	158	10.61	10.41
30	16	2.87	1.32
31	224	11.85	9.07
32	273	15.87	12.45
33	178	9.66	10.93
34	186	13.11	14.72
35	111	8.73	7.7
36	223	10.46	14.0
37	285	16.22	16.99
38	219	8.42	7.77
39	134	4.27	11.09
40	106	5.85	4.49
41	188	11.81	10.17
42	32	4.74	0.22
43	169	11.12	11.56
44	187	11.34	11.69
45	8	3.32	2.41
46	300	19.81	16.41
47	94	6.02	8.0
48	131	10.28	12.91
49	28	2.42	3.75
50	213	13.29	7.44
51	2	3.75	3.69
52	25	1.07	6.27
53	217	15.25	12.96
54	137	13.61	5.15
55	190	10.87	14.09
56	175	12.24	8.19
57	6	2.97	8.33
58	191	12.72	11.46
59	210	14.12	9.01
60	224	12.32	12.76
61	294	16.52	16.77
62	285	14.77	13.91
63	102	6.56	3.59
64	33	1.18	0
65	192	11.35	11.74
66	92	8.96	6.41
67	236	14.61	10.52
68	68	4.32	4.26
69	245	13.94	11.57
70	61	1.96	5.59
71	243	13.48	10.24
72	29	5.31	4.87
73	187	12.23	6.94
74	177	10.34	9.06
75	65	6.1	0.62
76	141	9.64	8.5
77	9	1.51	5.24
78	112	9.26	7.14
79	189	7.3	6.96
80	102	6.36	6.78
81	295	18.58	12.56
82	292	17.68	15.56
83	267	18.32	11.49
84	73	7.8	7.03
85	41	4.3	3.7
86	59	8.49	7.66
87	106	6.08	6.27
88	30	0.94	4.02
89	244	17.63	8.87
90	99	7.55	2.67
91	168	8.01	8.68
92	265	17.49	15.03
93	179	12.94	12.89
94	29	4.68	3.69
95	258	17.89	12.66
96	112	7.89	6.91
97	35	4.07	3.31
98	141	7.66	7.12
99	238	15.28	13.77
100	31	3.8	4.35
101	161	7.79	7.06
102	156	11.69	10.64
103	135	6.45	7.15
104	141	10.62	9.43
105	60	4.52	5.27
106	124	6.35	9.04
107	110	9.19	6.62
108	209	9.09	9.38
109	119	8.75	7.68
110	253	13.23	12.32
111	112	7.29	6.63
112	281	14.67	13.4
113	262	15.92	13.55
114	289	17.49	14.23
115	42	2.39	2.04
116	297	15.31	14.58
117	67	7.15	4.92
118	94	7.34	5.77
119	258	17.12	11.08
120	159	7.93	6.89

On peut commencer à réaliser une études en étudiants les caractères, deux par deux : le temps et la moyenne de math, le temps et la moyenne de stat et la moyenne de math et la moyenne de stat. A cette fin, on utilise les outils que nous avons développé lors dans le cas de la "dimension 2". Voici ce que l'analyse donne.

Notes de math en fonction du temps.



On a que $R^2 = 0.85163$ ce qui prouve que le modèle est bon. Les estimations ponctuelles donnent :

$$\hat{a} = 0.05112 \quad \hat{b} = 1.91537$$

De même on estime la variance du terme d'erreur : $\hat{\sigma} = 3.2522$.On en déduit les intervalles de confiance (symétrique) à l'aide de loi de Student à 118 degrés de liberté :

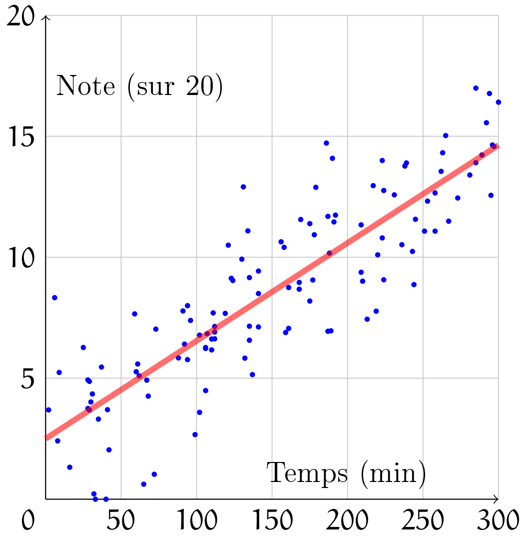
Niveau 95% :

$$a \in [0.04723; 0.05502] \qquad b \in [1.58911; 2.24163]$$

Niveau 99% :

$$a \in [0.04598; 0.05627] \qquad b \in [1.48403; 2.34671]$$

Notes de stat en fonction du temps.



On a que $R^2 = 0.73994$ ce qui prouve que le modèle est bon. Les estimations ponctuelles donnent :

$$\hat{a} = 0.04045 \quad \hat{b} = 2.49993$$

De même on estime la variance du terme d'erreur : $\hat{\sigma} = 4.10682$.On en déduit les intervalles de confiance (symétrique) à l'aide de loi de Student à 118 degrés de liberté :

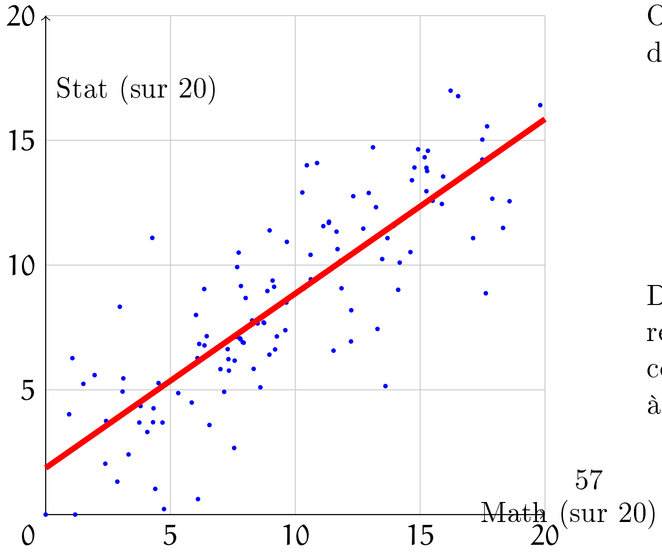
Niveau 95% :

$$a \in [0.03608; 0.04482] \qquad b \in [2.1333; 2.86656]$$

Niveau 99% :

$$a \in [0.03467; 0.04623] \qquad b \in [2.01522; 2.98464]$$

Notes de stat en fonction des notes de math.



On a que $R^2 = 0.67835$ ce qui prouve que le modèle est bon. Les estimations ponctuelles donnent :

$$\hat{a} = 0.69909 \quad \hat{b} = 1.86883$$

De même on estime la variance du terme d'erreur : $\hat{\sigma} = 5.0794$.On en déduit les intervalles de confiance (symétrique) à l'aide de loi de Student à 118 degrés de liberté :

Niveau 95% : $a \in [0.61133; 0.78685]$ $b \in [1.36966; 2.368]$

Niveau 99% : $a \in [0.58307; 0.81511]$ $b \in [1.20889; 2.52877]$

Évidemment les étudiants qui passent du temps à étudier ont de bonnes notes en mathématiques et en statistiques. Mais les statistiques utilisant des outils mathématiques plus des concept propre à cette matière, il est raisonnable de penser que les notes de stat sont "en lien" avec non seulement le temps de travail des étudiants mais aussi de leur niveau en math. Nous sommes donc naturellement amené à penser qu'il existe une règle de la forme

$$s_i = a + bt_i + cm_i$$

En imitant ce que nous avons fait au précédent chapitre, nous pouvons modéliser le problème en rajoutant un terme d'erreur et donc arriver au modèle linéaire :

$$S_i = a + bt_i + cm_i + \varepsilon_i$$

où $\varepsilon_i \sim \mathcal{N}(0, \sigma)$.

L'idée n'est pas du tout de reproduire le schéma précédent mais d'utiliser les technologies mathématiques pour exprimer ce modèle dans l'univers qui lui conviens le mieux : celui des matrices.

Avant de généraliser en dimension n quelconque, détaillons sur cet exemple.

Formulation matricielle : l'exemple

On choisi de réécrire Le modèle $S_i = a + bt_i + cm_i + \varepsilon_i$ sous la forme $S = Xm + e$ où, dans le cas de notre exemple, on a précisément :

$$\begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \\ S_7 \\ S_8 \\ S_9 \\ S_{10} \\ S_{11} \\ S_{12} \\ S_{13} \\ S_{14} \\ S_{15} \\ S_{16} \\ S_{17} \\ S_{18} \\ S_{19} \\ S_{20} \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 & 88 & 8.33 \\ 1 & 123 & 9.15 \\ 1 & 28 & 3.08 \\ 1 & 209 & 11.65 \\ 1 & 296 & 14.92 \\ 1 & 168 & 8.88 \\ 1 & 135 & 7.82 \\ 1 & 231 & 15.5 \\ 1 & 251 & 13.69 \\ 1 & 107 & 6.15 \\ 1 & 40 & 0 \\ 1 & 263 & 15.18 \\ 1 & 62 & 8.6 \\ 1 & 220 & 14.18 \\ 1 & 130 & 7.67 \\ 1 & 106 & 7.32 \\ 1 & 135 & 11.53 \\ 1 & 121 & 7.73 \\ 1 & 37 & 3.11 \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{16} \\ \varepsilon_{17} \\ \varepsilon_{18} \\ \varepsilon_{19} \\ \varepsilon_{20} \\ \vdots \end{pmatrix}$$

Ainsi au lieu de chercher les trois paramètre a, b et c "en dimension 1", on c'est ramener à déterminer un seul paramètre vectoriel m ici de dimension 3.

Exactement comme dans le paragraphe précédent, on cherche à minimiser les termes d'erreurs ε_i précisément :

$$\sum \varepsilon_i^2$$

Pour cela, encore une fois, plaçons nous dans le cadre vectoriel de l'équation $S = Xm + e$ et on observe que

$$\sum \varepsilon_i^2 = {}^t ee$$

où ${}^tM$ désigne la transposée de la matrice M, c'est à dire l'opération qui inverse les lignes et les colonnes.

Matriciellement (tout comme si on était en dimension 1 avec des nombres réels), $S = Xm + e$ implique que $e = S - Xm$. Laissons nous guider par les calculs :

$$\begin{aligned}
{}^t ee &= {}^t(S - Xm)\,(S - Xm) \\
&= \left({}^tS - {}^t(Xm)\right)(S - Xm) \\
&= \left({}^tS - {}^tm\,{}^tX\right)(S - Xm) \\
&= {}^tSS - {}^tSXm - {}^tm\,{}^tXS + {}^tm\,{}^tXXm
\end{aligned}$$

Observons les termes de cette dernière expression.

${}^tSS = \sum S_i^2$

$$\begin{aligned}
{}^tSXm &= \begin{pmatrix} S_1 & S_2 & S_3 & S4 & \cdots \end{pmatrix} \begin{pmatrix} 1 & 88 & 8.33 \\ 1 & 123 & 9.15 \\ 1 & 28 & 3.08 \\ 1 & 209 & 11.65 \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \\
&= \begin{pmatrix} S_1 & S_2 & S_3 & S4 & \vdots \end{pmatrix} \begin{pmatrix} a + 88b + 8.33c \\ a + 123b + 9.15c \\ a + 28b + 3.08c \\ a + 209b + 11.65c \\ \cdots \end{pmatrix} \\
&= S_1(a + 88b + 8.33c) + S_2(a + 123b + 9.15c) + S_3(a + 28b + 3.08c) + S_4(a + 209b + 11.65c) + \cdots
\end{aligned}$$

$$\begin{aligned}
{}^tm\,{}^tXS &= \begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & \cdots \\ 88 & 123 & 28 & 209 & \cdots \\ 8.33 & 9.15 & 3.08 & 11.65 & \cdots \end{pmatrix} \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ \vdots \end{pmatrix} \\
&= \begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} S_1 & S_2 & S_3 & S_4 & \cdots \\ 88S_1 & 123S_2 & 28S_3 & 209S_4 & \cdots \\ 8.33S_1 & 9.15S_2 & 3.08S_3 & 11.65S_4 & \cdots \end{pmatrix} \\
&= (aS_1 + 88bS_1 + 8.33cS_1) + (aS_2 + 123bS_2 + 9.15cS_2) + (aS_3 + 28bS_3 + 3.08cS_3) + (aS_4 + 209bS_4 + 11.65
\end{aligned}$$

En particulier, on observe, quitte à réordonner les termes, que ${}^tm\,{}^tXS = {}^tSXm$.

$$
\begin{aligned}
{}^t m {}^t X X m &= \begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & \cdots \\ 88 & 123 & 28 & 209 & \cdots \\ 8.33 & 9.15 & 3.08 & 11.65 & \cdots \end{pmatrix} \begin{pmatrix} 1 & 88 & 8.33 \\ 1 & 123 & 9.15 \\ 1 & 28 & 3.08 \\ 1 & 209 & 11.65 \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \\
&= \begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} 120 & 18042 & 1152.24 \\ 18042 & 3555336 & 216323.34 \\ 1152.24 & 216323.34 & 13650.2368 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \\
&= 120a^2 + 36084ab + 3555336b^2 + 2304.48ac + 432647.bc + 13650.2c^2
\end{aligned}
$$

On commence à sentir les notations s'appesantir... Finalement grâce à l'observation ${}^t m {}^t X S = {}^t S X m$, on a, plus simplement

$$
{}^t ee = {}^t SS - 2 {}^t m {}^t XS + {}^t m {}^t XXm
$$

Demander de minimiser cette égalité reviens à se demander pour quelle valeurs de m on a $\frac{\partial {}^t ee}{\partial m} = 0$. En dérivant "comme si m était une variable" [6] on arrive à

$$
\frac{\partial {}^t ee}{\partial m} = 0 \iff -2 {}^t XS + 2 {}^t XXm = 0 \iff {}^t XXm = {}^t XS \iff m = ({}^t XX)^{-1} {}^t XS
$$

A condition que ${}^t XX$ soit inversible, ce qui est le cas dans notre exemple. On trouve :

$$
m = ({}^t XX)^{-1} {}^t XS = \begin{pmatrix} 2.1734001594080326 \\ 0.03173376655529636 \\ 0.17047782118341104 \end{pmatrix}
$$

ce qui se lit en ligne par des estimations (en regardant les réalisations de S) $\hat{a} = 2.1734001594080326$, $\hat{b} = 0.03173376655529636$ et $\hat{c} = 0.17047782118341104$.

Bien on trouve des valeurs... mais est-ce que le modèle est bon ? En d'autre terme dans ce cadre, quel est l'équivalent du R^2.

Dans le cadre classique, $R^2 = \frac{\sum (\hat{y}_i - \overline{y})^2}{\sum (y_i - \overline{y})^2}$ ce qui se transpose sans problème ici (mais trouve un lien moins évident avec la corrélation de Pearson). Dans notre exemple, on peut donc donner de la valeur à notre modèle par le calcul

$$
R^2 = \frac{\sum (\hat{s}_i - \overline{s})^2}{\sum (s_i - \overline{s})^2} = \frac{\sum (\hat{a} + \hat{b} t_i + \hat{c} m_i - \overline{s})^2}{\sum (s_i - \overline{s})^2} \simeq 0.74592
$$

On peut raisonnablement penser que le modèle est correcte.

Ca suffit ! On a compris que tout les outils que nous avions en dimension 2 se transpose assez facilement avec le langage matricielle. Estimateurs, R^2 et tout leur petit copain se transposent assez bien.

Formulation matricielle

On généralise donc les observations et résultats établie précédemment. Dans ce contexte, on cherche à établir une relation linéaire simple entre une variable endogène à l'aide de p variables exogènes.

Définition

Soient p données statistiques $x_k = (x_{1,k}, \ldots, x_{n,k})$ et $y = (y_1, \ldots, y_n)$ des données statistiques.

Les variables x_k sont appelées **variables exogènes** ou *expliquées*.

La variable y est appelée **variable endogène** ou *à expliquer*.

6. Et c'est incroyable : c'est mathématiquement valide !

Une **modélisation linéaire multiple** consiste à considérer les variables aléatoires

$$Y_i = a_0 + a_1 x_{i,1} + a_2 x_{i,2} + \cdots + a_p x_{i,p} + \varepsilon_i$$

où es ε_i sont des variables aléatoires i.i.d. appelés **termes d'erreurs** et suivent une loi normale $\mathcal{N}(0, \sigma)$.

Comme dans le cas de la dimension 2 on suppose qu'il y a indépendance entre les termes d'erreurs et les variables exogène.

Proposition

Avec les notations de la définition précédente, un modèle linéaire multiple

$$Y_i = a_0 + a_1 x_{i,1} + a_2 x_{i,2} + \cdots + a_p x_{i,p} + \varepsilon_i$$

est équivalent à

$$Y = xm + e$$

Où x est une matrice à n lignes et $p+1$ colonnes, m le vecteur a estimer de dimension $p+1$ et e est un vecteur gaussien de dimension n.

Démonstration.

$$x = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \cdots x_{1,p} & \\ 1 & x_{2,1} & x_{2,2} & \cdots x_{2,p} & \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots x_{n,p} & \end{pmatrix} \qquad m = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} \qquad e = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

□

Remarque : On conserve le principe de notation : les minuscules pour les donnée déterminées et les majuscules pour les variables aléatoires.

Remarque : Comme nous l'avons observé avec l'exemple précédent, pour pouvoir estimer les paramètres, il est nécessaire que ${}^t xx$ soit inversible. C'est une condition assez souvent respectée mais qui peut être source d'erreur. Demander que ${}^t xx$ soit inversible équivaut à demander que x soit de rang maximale ; c'est à dire $p+1$ où encore que $\det({}^t xx) \neq 0$.

Remarque : Il est bien sure nécessaire que le nombre d'observation n soit strictement supérieur aux nombres de paramètre à estimer.

Ceci étant, on copie/colle les même énoncés, les preuves étant strictement identiques ou se déduisent des observations précédentes de l'exemple de ce chapitre.

Théorème

Avec les notations précédentes, $e = Y - xm$ et le minimum de ${}^t ee$ est atteint lorsque

$$m = \hat{m} \stackrel{\text{def}}{=} ({}^t xx)^{-1} {}^t xy$$

Définition

Avec les notations précédentes, on appel **résidu du modèle** le vecteur $\hat{e} = y - \hat{y}$ où $\hat{y} = x\hat{m}$

Définition

Avec les notations précédentes on définit le **coefficient de détermination du modèle**, noté R^2, par :

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

Proposition

Avec les notations précédentes

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

En particulier $R^2_{x,y} \in [0, 1]$.

Proposition

Avec les notations précédentes,

$$M_n = ({}^t xx)^{-1} {}^t xY$$

est un estimateur sans biais de m.

Démonstration.

$$\begin{aligned}
\mathbb{E}(M_n) &= \mathbb{E}\left(({}^t xx)^{-1} {}^t xY\right) \\
&= \mathbb{E}\left(({}^t xx)^{-1} {}^t x(xm + e)\right) \\
&= \mathbb{E}\left(({}^t xx)^{-1} {}^t xxm + ({}^t xx)^{-1} {}^t xe\right) \\
&= ({}^t xx)^{-1} {}^t xxm + \mathbb{E}\left(({}^t xx)^{-1} {}^t xe\right) \\
&= m + \mathbb{E}\left(({}^t xx)^{-1} {}^t xe\right) \\
&= m + ({}^t xx)^{-1} {}^t x \underbrace{\mathbb{E}(e)}_{=0} \\
&= m
\end{aligned}$$

□

Pour réaliser de l'inférence statistiques il nous faut une forme de *covariance* multivarié.

Matrice variance-covariance

Définition

Soit X un vecteur de probabilité. On définie la matrice de variance-covariance, notée $\mathbb{V}(X)$ par la formule

$$\mathbb{V}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^t (X - \mathbb{E}(X))\right)$$

Soit $A \simeq \mathcal{N}(1;1)$ et $B = 2A$ alors

$$\mathbb{V}\left(\begin{pmatrix} A \\ B \end{pmatrix}\right) = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

Proposition

1. La matrice de variance-covariance d'un vecteur aléatoire X vaut à l'intersection de la ligne i et de la colonne j $Cov(X_i, X_j)$.
2. Les éléments diagonaux sont les variances des coordonnées du vecteurs aléatoires.
3. La matrice de variance-covariance est symétrique à valeurs propres positives ou nulles.

Inférences des estimateurs

Dans le cas de la modélisation linéaire simple il existait un lien *simple* entre variances des estimateurs $\hat{a}$ et $\hat{b}$ avec la variance des termes d'erreurs. C'est ce lien qui a permis de déterminer les intervalles de confiance et autres joyeusetés.

Essayons, dans le cadre multivarié, de déterminer ce lien.

Nous partons du modèle $Y = xm + e$ où nous avons observé que $M_n = ({}^t xx)^{-1} {}^t xY$ est un estimateur sans biais mais pas (encore) convergent.

Proposition

Avec les notations précédentes, $\mathbb{V}(M_n) = \sigma^2 ({}^t xx)^{-1}$

Démonstration. Commençons par rappeler que ε_i sont i. i. d. et de moyenne nulle. En particulier $\mathbb{V}(e) = \sigma^2 Id$ puisque $Cov(\varepsilon_i, \varepsilon_j) = 0$ pour $i \neq j$ et $Cov(\varepsilon_i, \varepsilon_j) = \sigma^2$ sinon.

Rappelons également que x est une donnée déterministe (non aléatoire).

De plus $M_n - \mathbb{E}(M_n) = M_n - m = ({}^t xx)^{-1} {}^t xY - m = ({}^t xx)^{-1} {}^t x(xm + e) - m = ({}^t xx)^{-1} ({}^t xx)m + ({}^t xx)^{-1} {}^t xe - m = ({}^t xx)^{-1} {}^t xe$

$$\begin{aligned}
\mathbb{V}(M_n) &= \mathbb{E}\left((M_n - \mathbb{E}(M_n))^t (M_n - \mathbb{E}(M_n))\right) \\
&= \mathbb{E}\left(\left(({}^t xx)^{-1} {}^t xe\right)^t \left(({}^t xx)^{-1} {}^t xe\right)\right) \\
&= \mathbb{E}\left(({}^t xx)^{-1} {}^t xe {}^t ex^t (({}^t xx)^{-1})\right) \\
&= ({}^t xx)^{-1} {}^t x\mathbb{E}\left(e^t e\right) x^t (({}^t xx)^{-1}) \\
&= ({}^t xx)^{-1} {}^t x\sigma^2 Id x^t (({}^t xx)^{-1}) \\
&= \sigma^2 ({}^t xx)^{-1} \underbrace{({}^t xx)^t (({}^t xx)^{-1})}_{=Id} \\
&= \sigma^2 ({}^t xx)^{-1}
\end{aligned}$$

□

Corollaire

Si la matrice $({}^t xx)^{-1}$ tend vers la matrice nulle alors l'estimateur M_n est convergent.

Démonstration. Nous avons déjà observer que M_n est sans biais et par hypothèse sa variance tend vers 0. Donc M_n est un estimateur convergent. □

De la même manière que la modélisation linéaire simple nous pouvons estimer le paramètre σ :

Théorème

Soit $\boldsymbol{\varepsilon} = Y - \hat{y}$, alors

$$\mathbb{E}\left(\boldsymbol{\varepsilon}\right) = (n - (p+1))\sigma^2$$

Démonstration. Admise □

Corollaire

La variable aléatoire

$$S_n = \frac{{}^t\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}{n-(p+1)}$$

est un estimateur convergent et sans biais de σ^2. De plus :

1. $S_n^M = S_n({}^t xx)^{-1}$ est un estimateur convergent et sans biais de $\mathbb{V}\left(M_n\right)$.
2. En particulier, les coefficient diagonaux de S_n^M sont des estimateurs convergent et sans biais de la variance des coordonnées de M_n.
3. $(n-(p+1))\dfrac{S_n}{\sigma^2} \sim \chi^2(n-(p+1))$

Démonstration. Il s'agit des même arguments que pour la modélisation linaire simple. □

Corollaire

Soit $(M_n)_i$ la i-ieme coordonnée de M_n et S_i^M le coefficient diagonale de la i-ième ligne de S_n^m La variable aléatoire $\dfrac{M_{n,i} - m_i}{\sqrt{S_i^M}}$ suit une loi de Student à $n-(p+1)$ degrés de libertés.

Démonstration. Il s'agit des même arguments que pour la modélisation linaire simple. □

Corollaire

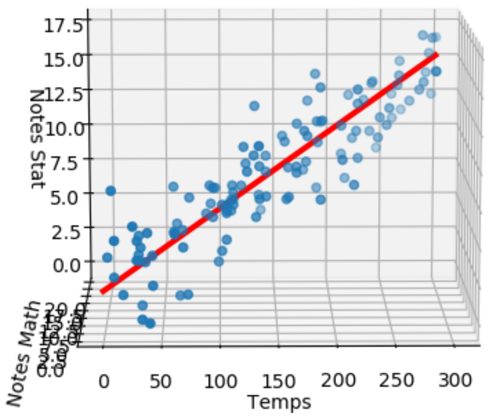
Soient $0 < \beta < \alpha < 1$, $t_1 = Q_{\mathcal{T}(n-(p+1))}(\beta)$ et $t_2 = Q_{\mathcal{T}(n-(p+1))}(1-\alpha+\beta)$ alors

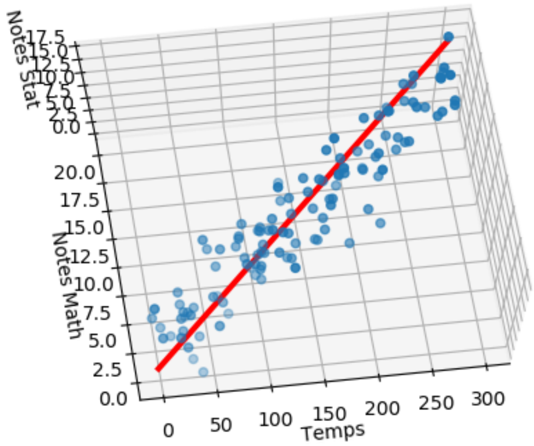
$$\left[(M_n)_i - t_2\sqrt{S_i^M}; (M_n)_i - t_1\sqrt{S_i^M}\right]$$

est un intervalle de confiance $1-\alpha$ de m_i.

Analyse multivarié de l'exemple introductif

Dans le cadre de notre exemple, on peut tenter un visualisation en trois dimension avec le temps de travail en abscisse, les notes en mathématiques en ordonnée et les notes de statistiques pour la côte.





De plus on a :

$$^{t}xx = \begin{pmatrix} 120 & 18042 & 1152.24 \\ 18042 & 3555336 & 216323.34 \\ 1152.24 & 216323.34 & 13650.2368 \end{pmatrix}$$

$$(^{t}xx)^{-1} = \begin{pmatrix} 0.04472 & 7.676e-05 & -0.005 \\ 7.676e-05 & 8e-06 & 0 \\ -0.005 & 0 & 0.003 \end{pmatrix}$$

On arrive alors a

$$\hat{m} = \begin{pmatrix} 2.17 \\ 0.032 \\ 0.17 \end{pmatrix} = \begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{pmatrix}$$

On estime donc le modèle $\hat{s}_i = \hat{a} + \hat{b}t_i + \hat{c}m_i$.

On a $R^2 = \dfrac{\sum(\hat{s}_i - \bar{s})^2}{\sum(s_i - \bar{s})^2} \simeq 0.746$ ce qui permet de dire que le modèle est viable.

Ce qui permet de déduire la matrice de variance covariance de l'estimateur M_n :

$$\mathbb{V}(\hat{M}_n) = \begin{pmatrix} 0.1809515800112538 & 0.0003106069843879191 & -0.020196806317842753 \\ 0.0003106069843879191 & 3.2362850206272714e-05 & -0.0005390920134243928 \\ -0.020196806317842753 & -0.0005390920134243928 & 0.010544604304957319 \end{pmatrix}$$

Ce qui permet d'obtenir les estimations de la variance de chacun des paramètres, par lecture des éléments diagonaux :

$$\widehat{\sigma_a^2} = 0.18095 \Rightarrow \quad \hat{\sigma_a} = 0.42538$$
$$\widehat{\sigma_b^2} = 3.2e-05 \Rightarrow \quad \hat{\sigma_b} = 0.00569$$
$$\widehat{\sigma_c^2} = 0.01054 \Rightarrow \quad \hat{\sigma_c} = 0.10269$$

En utilisant la loi de Student à $117 = 120 - 3$ degrés de liberté, on obtient les intervalles de confiances symétrique de niveau 95% des caractères à estimer :

$$a \in [1.331, 3.016]$$

$$b \in [0.02, 0.043]$$

$$c \in [-0.033, 0.374]$$

7. Annexes

7.1 Estimateurs

Nous disposons d'une pièce mais ne savons pas si elle est truquée. Nous la lançons 10 fois et faisons les observations suivantes :

$$\mathsf{P} \quad \mathsf{P} \quad \mathsf{F} \quad \mathsf{P} \quad \mathsf{F} \quad \mathsf{P} \quad \mathsf{P} \quad \mathsf{F} \quad \mathsf{P} \quad \mathsf{F}$$

Notons p la probabilité inconnue que la pièce tombe sur P. Naturellement, la fréquence empirique des observations permet d'approcher la valeur de p par $\dfrac{6}{10} = 0.6$. Mais 10 nouveau lancer vont très certainement aboutir à d'autre issue et donc une autre valeur possible de p. De plus cette estimation de p, à quelle point est-elle proche de sa vraie valeur ?

Cadrons le concept d'estimation et construisons des solutions.

Notation et définition

Définition

Soient $\mathbb{P}$ une loi de probabilité réelle et $n \in \mathbb{N}_{>0}$. On appel **échantillon de la loi** $\mathbb{P}$ un n-uplet de variable aléatoire indépendante de même loi $\mathbb{P}$.

Il faut distinguer le mot échantillon correspondant classiquement en statistique aux données observées du mot que nous définissons ici : il s'agit d'un modèle probabiliste. Il est d'accoutumé de noter les échantillons (du modèle) en majuscule et les donnés en minuscule. Ainsi $(X_1, \dots, X_n)$ est un échantillon d'une loi $\mathbb{P}$ tandis que $(x_1, \dots, x_n)$ est une observation de donnée.

Dans la pratique, la loi $\mathbb{P}$ dépend d'un paramètre inconnue ϑ et c'est ce paramètre (réel) que l'on cherche a estimer.

Définition

Soit $(X_1, \dots, X_n)$ un échantillon d'une loi de probabilité $\mathbb{P}_\vartheta$ dépendant d'un paramètre ϑ. Un **estimateur** T_n est une fonction de l'échantillon :

$$T_n = f(X_1, \dots, X_n)$$

Une **estimation de** ϑ est la valeur prise par l'estimateur sur une réalisation $(x_1, \dots, x_n)$ de l'échantillon.

$$\hat{\vartheta} = f(x_1, \dots, x_n)$$

Un estimateur est une variable aléatoire fonction de l'échantillon et une réalisation de cette estimateur est appelée une estimation !

Dans l'absolue n'importe quelle fonction f serait acceptable. Mais il faut donner un sens au mot *estimer*. C'est à dire se rapprocher de la vraie valeur de ϑ.

Dans la cas de pièce dont on ne connait pas la valeur du paramètre p, nous avions *estimer*, que cela pouvait être 0.6. Nous avons en effet considérer la fréquence empirique comme estimateur :

$$f(X_1, \dots, X_n) = \frac{X_1 + \dots + X_n}{n}$$

En effet codifions[7] l'observation

$$\mathsf{P} \quad \mathsf{P} \quad \mathsf{F} \quad \mathsf{P} \quad \mathsf{F} \quad \mathsf{P} \quad \mathsf{P} \quad \mathsf{F} \quad \mathsf{P} \quad \mathsf{F}$$

par

$$1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0$$

Alors $\hat{p} = f(1, 1, 0, 1, 0, 1, 1, 0, 1, 0) = \dfrac{6}{10}$.

La loi des grand nombre stipule que plus on va réaliser cette expérience (faire grandir n), plus la valeur empirique de p (estimée ici par $\hat{p}$) va se rapprocher de p. Cela motive la définition suivante.

7. Principalement parce que nous travaillons avec des variables quantitatives

Définition

Soient une loi de probabilité $\mathbb{P}_\vartheta$ dépendant d'un paramètre ϑ et T_n un estimateur de ϑ. On dira que T_n est un **estimateur convergent** si

$$\forall \epsilon > 0, \lim_{n \to +\infty} \mathbb{P}\left(\left|T_n - \vartheta\right| > \varepsilon\right) = 0$$

Autrement dis : il est presque impossible, pour de grande valeur de n, que T_n s'éloigne de ϑ. On peut de plus composer les estimateurs par des fonctions continues.

Théorème

Soient une loi de probabilité $\mathbb{P}_\vartheta$ dépendant d'un paramètre ϑ, T_n un estimateur de ϑ et f une fonction continue sur $\mathbb{R}$. Alors $f(T_n)$ est un estimateur convergent de $f(\vartheta)$.

Démonstration. Soit Ω l'univers des possibilités.
L'hypothèse de continuité se traduit par :

$$\forall \varepsilon > 0, \ \exists \eta_\varepsilon > 0, \quad \left(|x - \vartheta| \leqslant \eta_\varepsilon \Rightarrow |f(x) - f(\vartheta)| \leqslant \varepsilon\right)$$

En particulier la contraposé de l'implication est vraie, c'est à dire $\left(|f(x) - f(\vartheta)| > \varepsilon \Rightarrow |x - \vartheta| > \eta_\varepsilon\right)$. Notons

$$A_n = \left\{\omega \in \Omega \middle| \ |f\left(T_n(\omega)\right) - f(\vartheta)| > \varepsilon\right\}$$

$$B_n = \left\{\omega \in \Omega \middle| \ |T_n(\omega) - \vartheta| > \eta_\varepsilon\right\}$$

Alors l'hypothèse de continuité implique que $A_n \subseteq B_n$ et l'hypothèse de convergence de l'estimateur T_n implique que $\lim_{n \to +\infty} \mathbb{P}_\vartheta(B_n) = 0$. Dans ce cas

$$\mathbb{P}_\vartheta\left(|f\left(T_n(\omega)\right) - f(\vartheta)| > \varepsilon\right) = \mathbb{P}_\vartheta(A_n) \leqslant \mathbb{P}_\vartheta(B_n)$$

En passant à la limite dans cette inégalité on obtient bien que $\lim_{n \to +\infty} \mathbb{P}_\vartheta\left(|f\left(T_n(\omega)\right) - f(\vartheta)| > \varepsilon\right) = 0$ ce qui prouve que $f(T_n)$ est un estimateur convergent de $f(\vartheta)$. □

Voici un estimateur classique de l'espérance.

Proposition Loi faible des grands nombre

Soit $(X_1, \ldots, X_n)$ un échantillon d'une loi $\mathbb{P}$ d'espérance μ. Alors

$$f(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

est un estimateur convergent de μ.

Démonstration. C'est exactement la loi faible des grand nombre dont on pourra retrouver une preuve en annexe. □

Dans la pratique on note $\overline{X_n}$ cet estimateur.

Qualité d'un estimateur

Imaginons par exemple que le caractère *Age* des données se trouvant en annexe suit une loi normale. Il y a deux paramètres à estimer : la moyenne et la variance. Mais la proposition précédente permet d'obtenir la moyenne en réalisant la moyenne empirique des ages : 17.57. Mais ce critère de convergence à la limite n'apporte pas d'information sur la véracité de cette estimation. On ignore à quel point l'estimation est loin/proche de la vrai valeur.

Définition

Soient une loi de probabilité $\mathbb{P}_\vartheta$ dépendant d'un paramètre ϑ et T_n un estimateur de ϑ. On appel erreur quadratique, noté $\mathsf{EQ}(T_n,\vartheta)$ l'espérance :

$$\mathsf{EQ}(T_n,\vartheta) = \mathbb{E}\left((T_n-\vartheta)^2\right)$$

Dans la pratique on cherche un estimateur avec l'erreur quadratique la plus faible.

Théorème

Soient une loi de probabilité $\mathbb{P}_\vartheta$ dépendant d'un paramètre ϑ et T_n un estimateur de ϑ. Si $\lim\limits_{n\to+\infty} \mathsf{EQ}(T_n,\vartheta)=0$ alors T_n est un estimateur convergent de ϑ.

Démonstration. L'inégalité de Markov stipule que pour tout réel $a>0$ et pour toute variable aléatoire réelle Z presque surement positive on a $\mathbb{P}(Z\geqslant a)\leqslant \dfrac{\mathbb{E}(Z)}{a}$. Appliquons cette inégalité pour $a=\varepsilon^2$ où $\varepsilon>0$ est fixé et $Z=|T_n-\vartheta|^2$. On a alors

$$\mathbb{P}(|T_n-\vartheta|\geqslant\varepsilon)=\mathbb{P}(|T_n-\vartheta|^2\geqslant\varepsilon^2)\leqslant\frac{\mathbb{E}\left(|T_n-\vartheta|^2\right)}{\varepsilon^2}=\frac{\mathsf{EQ}(T_n,\vartheta)}{\varepsilon^2}$$

Puisque l'erreur quadratique tend vers 0, il en va de même pour $\mathbb{P}(|T_n-\vartheta|\geqslant\varepsilon)$ ce qui prouve que T_n est un estimateur convergent. □

On mesure aussi la qualité d'un estimateur par la *vitesse* dont il approche le paramètre.

Définition

Soient une loi de probabilité $\mathbb{P}_\vartheta$ dépendant d'un paramètre ϑ et T_n un estimateur de ϑ.

1. Le **biais** de l'estimateur est
$$\mathsf{B}(T_n,\vartheta)=\mathbb{E}(T_n-\vartheta)$$
2. L'estimateur est dis **sans biais** si $\mathsf{B}(T_n,\vartheta)=0$.
3. L'estimateur est dis **asymptotiquement sans biais** si $\lim\limits_{n\to+\infty}\mathsf{B}(T_n,\vartheta)=0$.

La linéarité de l'espérance permet de souligner que $\mathsf{B}(T_n,\vartheta)=\mathbb{E}(T_n)-\vartheta$.

Proposition

Soient une loi de probabilité $\mathbb{P}_\vartheta$ dépendant d'un paramètre ϑ et T_n un estimateur de ϑ.

$$\mathsf{EQ}(T_n,\vartheta)=\mathbb{V}(T_n)+\mathsf{B}(T_n,\vartheta)^2$$

Démonstration.

$$\begin{aligned}
\mathsf{EQ}(T_n,\vartheta) &= \mathbb{E}\left((T_n-\vartheta)^2\right)\\
&= \mathbb{E}\left((T_n-\mathbb{E}(T_n)+\mathbb{E}(T_n)-\vartheta)^2\right)\\
&= \mathbb{E}\left((T_n-\mathbb{E}(T_n))^2+2(T_n-\mathbb{E}(T_n))(\mathbb{E}(T_n)-\vartheta)+(\mathbb{E}(T_n)-\vartheta)^2\right)\\
&= \mathbb{E}\left((T_n-\mathbb{E}(T_n))^2\right)+2\mathbb{E}\left((T_n-\mathbb{E}(T_n))(\mathbb{E}(T_n)-\vartheta)\right)+\mathbb{E}\left((\mathbb{E}(T_n)-\vartheta)^2\right)\\
&= \mathbb{V}(T_n)+2\mathbb{E}\left((T_n-\mathbb{E}(T_n))\right)(\mathbb{E}(T_n)-\vartheta)+(\mathbb{E}(T_n)-\vartheta)^2\\
&= \mathbb{V}(T_n)+2\mathbb{E}(T_n)-\mathbb{E}(T_n))(\mathbb{E}(T_n)-\vartheta)+(\mathsf{B}(T_n,\vartheta))^2\\
&= \mathbb{V}(T_n)+\mathsf{B}(T_n,\vartheta)^2
\end{aligned}$$

□

Corollaire

Si un estimateur est asymptotiquement sans biais et que sa variance tend vers 0 alors il est convergent.

Démonstration. Les hypothèses impliquent que l'erreur quadratique tend vers 0 ce qui implique que l'estimateur est convergent. □

Intervalles de dispersion

L'erreur quadratique mesure la répartition de l'estimateur autour du paramètre. Les intervalles de dispersions permettent de raffiner l'information sur la qualité d'un estimateur.

Définition

Soit X une variable aléatoire réelle. La **fonction quantile** de X est définie par

$$\begin{array}{rcl} Q_X : [0,1] & \longrightarrow & \mathbb{R} \\ t & \longmapsto & \min\left(u \in \mathbb{R} \middle| \mathbb{P}(X \leqslant u) \geqslant t\right) \end{array}$$

Dans le cas où X est une v.a.r discrète, $Q_X(0.25)$ représente le premier quartile, $Q_X(0.5)$ la médiane et $Q_X(0.75)$ le troisième quartile.

Prenons par exemple une loi exponentielle de paramètre λ : $X \sim \mathcal{E}(\lambda)$. Alors pour tout $u \geqslant 0$

$$\begin{aligned} \mathbb{P}(X \leqslant u) &= \int_0^u \lambda e^{-\lambda x}\, dx \\ &= \left[-e^{-\lambda x}\right]_0^u \\ &= 1 - e^{-\lambda u} \end{aligned}$$

Pour $t \in [0,1]$ fixé, on cherche le plus petit u tel que $\mathbb{P}(X \leqslant u) \geqslant t$ c'est à dire le plus petit u tels que $1 - e^{-\lambda u} \geqslant t$. Cette inégalité équivaut à $u \geqslant -\frac{1}{\lambda}\ln(1-t)$. Le plus petit u est donc trivialement $-\frac{1}{\lambda}\ln(1-t)$. En conclusion

$$Q_{\mathcal{E}(\lambda)}(t) = -\frac{1}{\lambda}\ln(1-t)$$

Lemme : La fonction quantile de n'importe quelle variable aléatoire réelle est croissante.

Démonstration. Soient $0 \leqslant t_1 \leqslant t_2 \leqslant 1$. Notons

$$A_i = \left\{u \in \mathbb{R} \middle| \mathbb{P}(X \leqslant u) \geqslant t_i\right\}$$

pour $i \in \{1,2\}$.

Soit $u \in A_2$ alors $\mathbb{P}(X \leqslant u) \geqslant t_1 \geqslant t_2$ et donc $u \in A_2$. Ceci implique que $A_2 \subseteq A_1$ et donc $\min(A_1) \leqslant \min(A_2)$ soit encore $Q_X(t_1) \leqslant Q_X(t_2)$. □

Ce lemme permet de justifier la définition suivante[8].

Définition

Soient X une variable aléatoire réelle et $\alpha \in [0,1]$. Tout intervalle de la forme

$$\left[Q_X(\beta); Q_X(1-\alpha+\beta)\right]$$

où $0 \leqslant \beta \leqslant \alpha$ est appelé **intervalle de dispersion de niveau** $1-\alpha$.

8. Puisque $\alpha \leqslant 1$, $1-\alpha \geqslant 0$ et donc $1-\alpha+\beta \geqslant \beta$

L'**amplitude de dispersion** est la taille de l'intervalle $Q_X(1-\alpha+\beta) - Q_X(\beta)$.

Par exemple, avec $X \sim \mathcal{E}(\lambda)$, les intervalles de dispersions de niveau $1-\alpha$ ont pour forme $\left[-\frac{1}{\lambda}\ln(1-\beta); -\frac{1}{\lambda}\ln(\alpha-\beta)\right]$. Leurs amplitudes sont $\frac{1}{\lambda}\ln\left(\frac{1-\beta}{\alpha-\beta}\right)$.

Remarque : Lorsque la variable X est discrète (càd définie sur $\mathbb{N}$), on corrige les erreurs d'arrondi en remplaçant la borne droite $Q_X(1-\alpha+\beta)$ par $Q_X(1-\alpha+\beta)+1$.

Proposition

Soient X une variable aléatoire réelle, $\alpha \in [0,1]$ et $0 \leqslant \beta \leqslant \alpha$.

$$\mathbb{P}\left(X \in \left[Q_X(\beta); Q_X(1-\alpha+\beta)\right]\right) = 1-\alpha$$

Démonstration. Pour simplifier, supposons que le minimum est atteint, c'est à dire $\mathbb{P}(X \leqslant Q_X(t)) = t$.

$$\begin{aligned}
\mathbb{P}\left(X \in \left[Q_X(\beta); Q_X(1-\alpha+\beta)\right]\right) &= \mathbb{P}\left(Q_X(\beta) \leqslant X \leqslant Q_X(1-\alpha+\beta)\right) \\
&= \mathbb{P}\left(X \leqslant Q_X(1-\alpha+\beta)\right) - \mathbb{P}\left(X \leqslant Q_X(\beta)\right) \\
&= 1-\alpha+\beta-\beta \\
&= 1-\alpha
\end{aligned}$$

□

Définition

Un intervalle de dispersion de niveau $1-\alpha$ sera dit :
- → **unilatérale inférieur** si $\beta = 0$,
- → **unilatérale supérieur** si $\beta = \alpha$,
- → **symétrique** si $\beta = \frac{\alpha}{2}$,
- → **optimal** si son amplitude est minimale.

Par exemple, toujours avec une loi exponentielle de paramètre λ les intervalles de niveau $1-\alpha$
- → unilatérale inférieur ont pour forme $\left[0; -\frac{1}{\lambda}\ln(\alpha)\right]$
- → unilatérale supérieur ont pour forme $\left[-\frac{1}{\lambda}\ln(1-\alpha); +\infty\right[$
- → symétrique ont pour forme $\left[-\frac{1}{\lambda}\ln\left(1-\frac{\alpha}{2}\right); -\frac{1}{\lambda}\ln\left(\frac{\alpha}{2}\right)\right]$
- → optimaux sont les intervalles unilatérale supérieur. En effet on peut sans peine démontrer que $\beta \mapsto \frac{1}{\lambda}\ln\left(\frac{1-\beta}{\alpha-\beta}\right)$ atteint son minimum lorsque $\beta = \alpha$.

Classiquement α prend la valeur 0.05.

Lemme : Soient $t \in [0,1]$ et Q_X a fonction quantile d'une variable aléatoire discrète.
1. $\mathbb{P}(X \leqslant u_0) \geqslant t \Rightarrow Q_X(t) \leqslant u_0$.
2. $\mathbb{P}(X \leqslant u_0) < t \Rightarrow Q_X(t) \geqslant u_0$.

Démonstration.
1. Par définition $Q_X(t)$ est le plus petit u tel que $\mathbb{P}(X \leqslant u) \geqslant t$ de sorte que si u_0 vérifie aussi cette propriété il est nécessairement supérieur à ce minimum.
2. Si $Q_X(t) < u_0$ alors (car la fonction de répartition $u \mapsto \mathbb{P}(X \leqslant u)$ est croissante) $\mathbb{P}(X \leqslant u_0) \geqslant \mathbb{P}(X \leqslant Q_X(t)) \geqslant t$ ce qui contredit $\mathbb{P}(X \leqslant u_0) < t$.

□

Théorème

Soient une loi de probabilité $\mathbb{P}_\vartheta$ dépendant d'un paramètre ϑ et T_n un estimateur de ϑ.
L'estimateur T_n est convergent si et seulement si

$$\forall 0 < \beta < \alpha,\ \forall \varepsilon > 0,\ \exists N > 0,\ \Big(n > N \Rightarrow \Big[Q_{T_n}(\beta); Q_{T_n}(1-\alpha+\beta)\Big] \subseteq [\vartheta - \varepsilon; \vartheta + \varepsilon]\Big)$$

Démonstration. On a les équivalences suivantes :

$$\begin{aligned}
\mathbb{P}_\vartheta(|T_n - \vartheta| > \varepsilon) \to 0 &\iff \mathbb{P}_\vartheta(|T_n - \vartheta| \leqslant \varepsilon) \to 1 \\
&\iff \mathbb{P}_\vartheta(\vartheta - \varepsilon \leqslant T_n \leqslant \vartheta + \varepsilon) \to 1 \\
&\iff \mathbb{P}_\vartheta(T_n \leqslant \vartheta + \varepsilon) - \mathbb{P}_\vartheta(T_n \leqslant \vartheta - \varepsilon) \to 1 \\
&\iff \big(\mathbb{P}_\vartheta(T_n \leqslant \vartheta - \varepsilon) \to 0\big) \wedge \big(\mathbb{P}_\vartheta(T_n \leqslant \vartheta + \varepsilon) \to 1\big)
\end{aligned}$$

Cette dernière implication s'observe par contraposée : si $\mathbb{P}_\vartheta(T_n \leqslant \vartheta - \varepsilon)$ tend vers $l \neq 0$ alors, puisque $\mathbb{P}_\vartheta(T_n \leqslant \vartheta + \varepsilon) \leqslant 1$, on a $\mathbb{P}_\vartheta(T_n \leqslant \vartheta + \varepsilon) - \mathbb{P}_\vartheta(T_n \leqslant \vartheta - \varepsilon) \to l' \leqslant 1 - l \neq 1$.

Premier cas. Supposons que T_n soit un estimateur convergent. C'est à dire que $\mathbb{P}_\vartheta(T_n \leqslant \vartheta - \varepsilon) \to 0$ et $\mathbb{P}_\vartheta(T_n \leqslant \vartheta + \varepsilon) \to 1$.

- Puisque $\mathbb{P}_\vartheta(T_n \leqslant \vartheta - \varepsilon) \to 0$, alors pour n suffisamment grand, puisque $\beta > 0$, on a $\mathbb{P}(T_n \leqslant \vartheta - \varepsilon) < \beta$ ce qui implique, d'après le précédent lemme $\boxed{Q_{T_n}(\beta) \geqslant \vartheta - \varepsilon}$
- De la même manière puisque $\alpha - \beta > 0$ le passage au complémentaire permet d'écrire les implications suivantes pour n suffisamment grand :

$$\begin{aligned}
\mathbb{P}_\vartheta(T_n \leqslant \vartheta + \varepsilon) \to 1 &\implies \mathbb{P}_\vartheta(T_n > \vartheta + \varepsilon) \to 0 \\
&\implies \mathbb{P}_\vartheta(T_n > \vartheta + \varepsilon) \leqslant \alpha - \beta \\
&\implies \mathbb{P}_\vartheta(T_n \leqslant \vartheta + \varepsilon) > 1 - (\alpha - \beta)
\end{aligned}$$

  ce qui prouve d'après le précédent lemme que $\boxed{Q_{T_n(1-\alpha+\beta)} \leqslant \vartheta + \varepsilon}$.

Ces deux inégalités permettent d'écrire

$$\vartheta - \varepsilon \leqslant Q_{T_n}(\beta) \leqslant Q_{T_n}(1-\alpha+\beta) \leqslant \vartheta + \varepsilon$$

ce qui prouve que $\Big[Q_{T_n}(\beta); Q_{T_n}(1-\alpha+\beta)\Big] \subseteq [\vartheta - \varepsilon; \vartheta + \varepsilon]$.

Second cas. Supposons que pour tout $0 < \alpha < \beta$ et pour n suffisamment grand on a $\Big[Q_{T_n}(\beta); Q_{T_n}(1-\alpha+\beta)\Big] \subseteq [\vartheta - \varepsilon; \vartheta + \varepsilon]$. Cette inclusion équivaut aux inégalités

$$\vartheta - \varepsilon \leqslant Q_{T_n}(\beta) \leqslant Q_{T_n}(1-\alpha+\beta) \leqslant \vartheta + \varepsilon$$

- Soit l la limite de $\mathbb{P}_\vartheta(T_n \leqslant \vartheta - \varepsilon)$. Si $l \neq 0$ alors pour n suffisamment grand $\mathbb{P}_\vartheta(T_n \leqslant \vartheta - \varepsilon) \geqslant \dfrac{l}{3}$ (pourquoi pas) ce qui implique d'après le lemme que $Q_{T_n}\left(\dfrac{l}{3}\right) \leqslant \vartheta - \varepsilon$ et contredit l'inégalité $\vartheta - \varepsilon \leqslant Q_{T_n}(\beta)$ pour $\beta = \dfrac{l}{3}$ (puisque cela doit être vraie pour tout $\beta > 0$ et que $l \neq 0$). On a donc nécessairement $\boxed{\mathbb{P}_\vartheta(T_n \leqslant \vartheta - \varepsilon) \to 0}$.
- Soit $\alpha = \dfrac{1}{k}$ et $\beta = \dfrac{1}{2k}$. Les inégalités impliquent que pour tout k et pour n suffisamment grand (qui peut varier avec k), $Q_{T_n}\left(1 - \dfrac{1}{2k}\right) \leqslant \vartheta + \varepsilon$. Puisque la fonction de répartition est croissante, on a

$$1 \geqslant \mathbb{P}_\vartheta(T_n \leqslant \vartheta + \varepsilon) \geqslant \mathbb{P}_\vartheta\left(T_n \leqslant Q_{T_n}\left(1 - \frac{1}{2k}\right)\right) \geqslant 1 - \frac{1}{2k}$$

Ainsi pour tout $k \in \mathbb{N}_{>0}$, il existe N_k suffisamment grand tel que pour tout $n > N_k$, $1 \geqslant \mathbb{P}_\vartheta(T_n \leqslant \vartheta + \varepsilon) \geqslant 1 - \dfrac{1}{2k}$. En passant à la limite lorsque k tend vers $+\infty$ alors n_k tend vers $+\infty$ (c'est un changement de variable), on trouve que $\boxed{\mathbb{P}_\vartheta(T_n \leqslant \vartheta + \varepsilon) \to 1}$

D'après l'équivalence introduite au début de la preuve, cela signifie que T_n est un estimateur convergent.

$\square$

7.2 Intervalles de confiances

L'idée est de raffiner la notion d'estimation en ne proposant pas une estimation ponctuelle mais plutôt un intervalle d'estimation. Le coeur de cette approche est de considérer les intervalles de dispersion comme des *a priori* et non comme des *a posteriori*.

Il ne faut pas confondre les intervalles de confiances et les intervalles de dispersion.

Intervalles de dispersion

Rappelons le théorème de la limite centrale qui est le coeur des intervalles d'estimation.

Théorème Limite centrale

Soit $(X_n)_{n\in\mathbb{N}_{>0}}$ une suite de variable aléatoire réelle, indépendantes et identiquement distribuée. Notons μ leur espérance et σ leur écart-type.

Soit $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Alors $(\overline{X_n})_{n\in\mathbb{N}_{>0}}$ converge la loi normale de moyenne μ et d'écart-type $\frac{\sigma}{\sqrt{n}}$.

$$\lim_{n\to+\infty} \overline{X_n} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Le sens à donnée à cette *limite en loi* est la suivante. Soit $\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{x^2}{2}}\,dx$ la fonction de répartition de la loi normale centrée réduite. Alors le théorème de le limite centrale s'énonce comme :

$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{\overline{X_n}-\mu}{\frac{\sigma}{\sqrt{n}}} \leqslant t\right) = \Phi(t)$$

Ce théorème est très impressionnant : les moyennes empirique de n'importe quelle loi converge vers une loi normale.

Considérons une urne avec une certaine proportion p de boule noire. On effectue n tirage et on observe une fréquence d'apparition des boules noires f. Du point de vu des probabilité, ce nombre est proche de p. Mais statistiquement il peut être un peu plus grand ou un peut plus petit.

Si n est suffisamment grand [9], on peut passer par la loi normale d'après le théorème de la limite centrale. On considère X la variable aléatoire de Bernouilli de paramètre p. Dans ce cas $\mathbb{E}(X) = p$ et $\mathbb{V}(X) = \sqrt{p(1-p)}$. Alors, pour n suffisamment grand, on peut écrire :

$$\begin{aligned}
\mathbb{P}\left(\frac{\overline{X_n}-p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}} \leqslant t\right) \simeq \Phi(t) \quad &\Longrightarrow \quad \mathbb{P}\left(\left|\frac{\overline{X_n}-p}{\sqrt{\frac{p(1-p)}{n}}}\right| \leqslant t\right) \simeq 2\Phi(t)-1 \\
&\Longrightarrow \quad \mathbb{P}\left(\left|\overline{X_n}-p\right| \leqslant t\sqrt{\frac{p(1-p)}{n}}\right) \simeq 2\Phi(t)-1 \\
&\Longrightarrow \quad \mathbb{P}\left(-t\sqrt{\frac{p(1-p)}{n}} \leqslant \overline{X_n}-p \leqslant t\sqrt{\frac{p(1-p)}{n}}\right) \simeq 2\Phi(t)-1 \\
&\Longrightarrow \quad \mathbb{P}\left(p-t\sqrt{\frac{p(1-p)}{n}} \leqslant \overline{X_n} \leqslant p+t\sqrt{\frac{p(1-p)}{n}}\right) \simeq 2\Phi(t)-1
\end{aligned}$$

On déduit que la fréquence f observe varie, se disperse, dans l'intervalle $\left[p-t\sqrt{\frac{p(1-p)}{n}}; p+t\sqrt{\frac{p(1-p)}{n}}\right]$.

9. Le calcul numérique, permet de dire que si $n \geqslant 30$ alors n *est suffisamment grand.*

- Si l'on veut être sure à 95% que f appartienne à cet intervalle, il faut choisir t tel que $2\Phi(t)-1 \simeq 0.95$ c'est à dire $\Phi(t) = 0.975$. La table de la loi normale donne $t \simeq 1.96$ de sorte que

$$\left[p - 1.96\sqrt{\frac{p(1-p)}{n}}; p + 1.96\sqrt{\frac{p(1-p)}{n}}\right]$$

est l'intervalle de fluctuation de niveau 95%.

- On peut légèrement simplifier l'intervalle précédent, en observant que puisque $0 \leqslant p \leqslant 1$ alors $p(1-p) \leqslant \frac{1}{4}$ et que naturellement $1.96 < 2$ alors

$$\left[p - 1.96\sqrt{\frac{p(1-p)}{n}}; p + 1.96\sqrt{\frac{p(1-p)}{n}}\right] \subseteq \left[p - 2\sqrt{\frac{1}{4n}}; p + 2\sqrt{\frac{1}{4n}}\right] = \left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$$

qui est souvent l'intervalle de fluctuation choisie.

- Si l'on veut être sure à 99.9% que f appartienne à cet intervalle, il faut choisir t tel que $2\Phi(t)-1 \simeq 0.999$ c'est à dire $\Phi(t) = 0.9995$. La table de la loi normale donne $t \simeq 3.29$ de sorte que

$$\left[p - 3.29\sqrt{\frac{p(1-p)}{n}}; p + 3.29\sqrt{\frac{p(1-p)}{n}}\right] \subseteq \left[p - \frac{2}{\sqrt{n}}; p + \frac{2}{\sqrt{n}}\right]$$

est l'intervalle de fluctuation de niveau 99.9%.

Intervalle de confiance de l'estimateur de la moyenne

On prend le problème *à l'envers*. On ne connait pas la proportion p de boule noire dans l'urne mais on dispose d'une statistique d'apparition f sur un échantillon de taille n. En raisonnant exactement comme précédemment mais en considérant une loi de Bernouilli de paramètre f connu et non p, on détermine de la même manière des intervalles. Mais ces intervalles estime p. Précisément, ils permettent d'encadrer la valeur de p avec une bonne probabilité.

Définition

Soient $(X_1, \dots, X_n)$ un échantillon d'une loi $\mathbb{P}$ d'espérance inconnue μ, d'écart-type connue σ et $0 \leqslant \alpha \leqslant 1$. L'**intervalle de confiance de niveau** $1-\alpha$ de μ est

$$\left[\overline{X_n} - z_\alpha \frac{\sigma}{\sqrt{n}}; \overline{X_n} + z_\alpha \frac{\sigma}{\sqrt{n}}\right]$$

où z_α est l'unique solution de l'équation $\Phi(z_\alpha) = 1 - \frac{\alpha}{2}$ où Φ désigne la fonction de répartition de la loi normale centrée réduite.

En utilisant la fonction quantile, on peut aussi écrire $z_\alpha = Q_{\mathcal{N}(0,1)}\left(1 - \frac{\alpha}{2}\right)$.

Un intervalle de confiance de niveau 95% est inclus dans $\left[f - \frac{2\sigma}{\sqrt{n}}; f + \frac{2\sigma}{\sqrt{n}}\right]$.

Imaginons que l'urne contient 500 boules [10] et qu'en en tire 10. Sur ces 10, on en observe 3 noires de sorte que $f = 0.3$. La proportion réelle, à 95% de boule noire se trouve alors dans l'intervalle $\left[0.3 - \frac{1}{\sqrt{10}}; 0.3 + \frac{1}{\sqrt{10}}\right] \simeq [0; 0.61]$... Autant dire inutilisable tant son amplitude est grande. Ceci est du au faible échantillon choisi.

Imaginons à présent que nous en avons tiré 100 et que 27 d'entre elles soient noires, c'est à dire $f = 0.27$ alors $\left[0.27 - \frac{1}{\sqrt{100}}; 0.27 + \frac{1}{\sqrt{100}}\right] = [0.17; 0.37]$ est un intervalle contenant p à 95% ce qui est déjà plus raisonnable.

Plaçons nous à présent dans un cadre un peu plus générale.

Définition

10. Il pourrait même en avoir 500 millions, ça ne change rien au calcul.

Définition

Soient une loi de probabilité $\mathbb{P}_\vartheta$ dépendant d'un paramètre ϑ et $T_n^{(1)}$, $T_n^{(2)}$ des estimateurs de ϑ. On dira que $[T_n^{(1)}; T_n^{(2)}]$ est un **intervalle de confiance de niveau $1-\alpha$ de ϑ** si

$$\mathbb{P}_\vartheta(T_n^{(1)} \leqslant \vartheta \leqslant T_n^{(2)}) \geqslant 1-\alpha$$

Si on réalise $T_n^{(1)}$ et $T_n^{(2)}$ on obtient des estimations particulière t_1 et t_2. On pourra alors affirmer que $\vartheta \in [t_1; t_2]$ avec une probabilité de $1-\alpha$.

Si on cherche par exemple à encadrer la moyenne d'une variable aléatoire d'écart-type σ, on considère les estimateurs $T_n^{(1)} = \overline{X_n} - z_\alpha \frac{\sigma}{\sqrt{n}}$ et $T_n^{(2)} = \overline{X_n} + z_\alpha \frac{\sigma}{\sqrt{n}}$ (qui sont bien des estimateurs convergent de la moyenne, asymptotiquement sans biais) qui font de $[T_n^{(1)}; T_n^{(2)}]$ un intervalle de confiance.

Les intervalles de dispersions permettent d'obtenir des intervalles de confiances.

Définition

Soit X une variable aléatoire réelle de loi $\mathbb{P}_\vartheta$ dépendant d'un paramètre réel ϑ. On note

$$\begin{array}{rccl} q_{X,(\gamma)}: \mathbb{R} & \longrightarrow & \mathbb{R} \\ \vartheta & \longmapsto & Q_X(\gamma) = \min\left\{u \in \mathbb{R} \middle| \mathbb{P}_\vartheta(X \leqslant \gamma) \geqslant t\right\} \end{array}$$

Lemme : Supposons que $\mathbb{P}_\vartheta$ est une loi continue. Alors les fonction $q_{X,(\gamma)}$ sont strictement croissante.

Démonstration. Admise □

On rappel que toute fonction strictement croissante réalise une bijection entre son domaine et son codomaine. Il existe donc une fonction réciproque $q^{-1}_{X,(\gamma)}$.

Théorème

Soient $\mathbb{P}_\vartheta$ une loi de probabilité continue dépendant d'un paramètre ϑ estimé par T_n et $0 < \alpha < 1$.

Pour tout $\beta \in]0; \alpha[$,

$$\left[q^{-1}_{T_n,(1-\alpha+\beta)}(T_n); q^{-1}_{T_n,(\beta)}(T_n)\right]$$

est un intervalle de confiance de niveau au moins $1-\alpha$.

Démonstration. Si $\vartheta \leqslant q^{-1}_{T_n,(\beta)}(T_n)$ alors $Q_{T_n}(\beta) = q_{T_n,(\beta)}(\vartheta) \leqslant T_n$. De même si $\vartheta \geqslant q^{-1}_{T_n,(1-\alpha+\beta)}(T_n)$ alors $Q_{T_n}(1-\alpha+\beta) = q_{T_n,(1-\alpha+\beta)}(\vartheta) \geqslant T_n$. En d'autre terme

$$\vartheta \in \left[q^{-1}_{T_n,(1-\alpha+\beta)}(T_n); q^{-1}_{T_n,(\beta)}(T_n)\right] \Longleftrightarrow T_n \in [Q_{T_n}(\beta); Q_{T_n}(1-\alpha+\beta)]$$

De sorte que $\mathbb{P}_\vartheta\left(\vartheta \in \left[q^{-1}_{T_n,(1-\alpha+\beta)}(T_n); q^{-1}_{T_n,(\beta)}(T_n)\right]\right) = \mathbb{P}_\vartheta\left(T_n \in [Q_{T_n}(\beta); Q_{T_n}(1-\alpha+\beta)]\right) = 1-\alpha$ □

Dans la pratique, on choisi β pour faire de $[Q_{T_n}(\beta); Q_{T_n}(1-\alpha+\beta)]$ un intervalle dispersion optimale ; propriété dont héritera l'intervalle de confiance $\left[q^{-1}_{T_n,(1-\alpha+\beta)}(T_n); q^{-1}_{T_n,(\beta)}(T_n)\right]$.

Considérons des estimateurs pour la moyenne et la variance.

Intervalles de confiance d'estimateur de la moyenne et de l'écart-type

Proposition

Soient $n \in \mathbb{N}_{>0}$ et $(X_1, \dots, X_n)$ un échantillon d'une loi $\mathbb{P}$.

1. $\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k$ est un estimateur convergent et sans biais de $\mathbb{E}(X_1)$.

2. $S_n = \frac{1}{n}\sum_{k=1}^{n} X_k^2$ est un estimateur convergent et sans biais de $\mathbb{E}\left(X_1^2\right)$.

3. $\overline{V_n}^{(1)} = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \overline{X_n}^2 = S_n - \overline{X_n}^2$ est un estimateur convergent et asymptotiquement sans biais de $\mathbb{V}(X_1)$.

4. $\overline{V_n}^{(2)} = \frac{n}{n-1}\sum_{k=1}^{n} X_k^2 - \overline{X_n}^2 = \frac{n}{n-1}\left(S_n - \overline{X_n}^2\right) = \frac{n}{n-1}\overline{V_n}^{(1)}$ est un estimateur convergent et sans biais de $\mathbb{V}(X_1)$.

Démonstration. Nous avons déjà démontrer le premier point en utilisant la loi faible des grand nombre. Ce même résultat montre que S_n est un estimateur convergent de $\mathbb{E}\left(X_1^2\right)$ de sorte que $\overline{V_n}^{(1)}$ et $\overline{V_n}^{(2)}$ sont bien des estimateurs convergent de $\mathbb{V}(X_1)$. Déterminons les biais.

$$\begin{aligned}
\mathbb{E}\left(\overline{X_n}\right) &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k)\\
&= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_1)\\
&= \mathbb{E}(X_1)
\end{aligned}$$

De sorte que $B(\overline{X_n}, \mathbb{E}(X_1)) = \mathbb{E}\left(\overline{X_n}\right) - \mathbb{E}(X_1) = 0$ et $\overline{X_n}$ est bien sans bien sans biais.

$$\begin{aligned}
\mathbb{E}(S_n) &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(X_k^2\right)\\
&= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(X_1^2\right)\\
&= \mathbb{E}\left(X_1^2\right)
\end{aligned}$$

De sorte que $B(S_n, \mathbb{E}\left(X_1^2\right)) = \mathbb{E}(S_n) - \mathbb{E}\left(X_1^2\right) = 0$ et S_n est bien sans bien sans biais.

$$\begin{aligned}
\mathbb{E}\left(\overline{V_n}^{(2)}\right) &= \frac{n}{n-1}\sum_{k=1}^{n}\mathbb{E}\left(\overline{V_n}^{(1)}\right)\\
&= \frac{n}{n-1}\frac{n-1}{n}\mathbb{V}(X_1)\\
&= \mathbb{V}(X_1)
\end{aligned}$$

De sorte que $B(\overline{V_n}^{(2)}, \mathbb{V}(X_1)) = \mathbb{E}\left(\overline{V_n}^{(2)}\right) - \mathbb{V}(X_1) = 0$ et $\overline{V_n^{(2)}}$ est bien sans bien sans biais.

$$\begin{aligned}
\mathbb{E}\left(\overline{V_n}^{(1)}\right) &= \mathbb{E}(S_n) - \mathbb{E}\left(\overline{X_n}^2\right)\\
&= \mathbb{E}(X_1^2) - \mathbb{E}\left(\frac{1}{n^2}\sum_{k=1}^{n} X_k\right)\\
&= \mathbb{E}(X_1^2) - \frac{1}{n^2}\mathbb{E}\left(\sum_{k=1}^{n} X_k^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} X_iX_j\right)\\
&= \mathbb{E}(X_1^2) - \frac{1}{n^2}\left(\sum_{k=1}^{n}\mathbb{E}(X_k^2) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{E}(X_i)\mathbb{E}(X_j\right.\\
&= \mathbb{E}(X_1^2) - \frac{1}{n^2}\left(\sum_{k=1}^{n}\mathbb{E}(X_1^2) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{E}(X_1)\mathbb{E}(X_1\right.\\
&= \mathbb{E}(X_1^2) - \frac{1}{n^2}\left(n\mathbb{E}(X_1^2) + 2\frac{n(n-1)}{2}\mathbb{E}(X_1)\mathbb{E}(X_1)\right)\\
&= \mathbb{E}(X_1^2) - \frac{1}{n}\left(\mathbb{E}(X_1^2) + (n-1)\mathbb{E}(X_1)^2\right)\\
&= \frac{1}{n}\left(n\mathbb{E}(X_1^2) - \mathbb{E}(X_1^2) - (n-1)\mathbb{E}(X_1)^2\right)\\
&= \frac{1}{n}\left((n-1)\mathbb{E}(X_1^2) - (n-1)\mathbb{E}(X_1)^2\right)\\
&= \frac{n-1}{n}\left(\mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2\right)\\
&= \frac{n-1}{n}\mathbb{V}(X_1)
\end{aligned}$$

De sorte que $B(\overline{V_n}^{(1)}, V(X_1)) = \mathbb{E}\left(\overline{V_n}^{(1)}\right) - \mathbb{V}(X_1) = -\frac{1}{n}\mathbb{V}(X_1)$ et $\overline{V_n}^{(1)}$ est asymptotiquement sans biais.

□

Lemme : Soient $n \in \mathbb{N}_{>0}$ et $(X_1, \ldots, X_n)$ un échantillon d'une loi normale $\mathcal{N}(\mu, \sigma)$.

1. La variable aléatoire $\frac{\sqrt{n}}{\sigma}\left(\overline{X_n} - \mu\right)$ suit la loi normale $\mathcal{N}(0, 1)$.
2. La variable aléatoire $\sqrt{\frac{n-1}{\overline{V_n}^{(1)}}}\left(\overline{X_n} - \mu\right)$ suit la loi de Student $\mathcal{T}_{n-1}$.

3. La variable aléatoire $\dfrac{n\overline{V_n}^{(1)}}{\sigma^2}$ suit la loi du khi deux χ^2_{n-1}.

Démonstration. Il s'agit principalement du théorème de la limite centrale. □

Théorème

On fixe $\alpha \in]0,1[$ et $0 < \beta < \alpha$.

Soient $n \in \mathbb{N}_{>0}$ et $(X_1, \ldots, X_n)$ un échantillon d'une loi normale $\mathcal{N}(\mu, \sigma)$.

1. Soit $z_1 = Q_{\mathcal{N}(0,1)}(\beta)$ et $z_2 = Q_{\mathcal{N}(0,1)}(1-\alpha+\beta)$. Si on connait la valeur de σ,

$$\left[\overline{X_n} - z_2\frac{\sigma}{\sqrt{n}}; \overline{X_n} - z_1\frac{\sigma}{\sqrt{n}}\right]$$

est un intervalle de confiance de niveau $1-\alpha$ de μ.

2. Soit $t_1 = Q_{\mathcal{T}_{n-1}}(\beta)$ et $t_2 = Q_{\mathcal{T}_{n-1}}(1-\alpha+\beta)$. Si on ne connait pas la valeur de σ,

$$\left[\overline{X_n} - t_2\sqrt{\frac{\overline{V_n}^{(1)}}{n-1}}; \overline{X_n} - t_1\sqrt{\frac{\overline{V_n}^{(1)}}{n-1}}\right]$$

est un intervalle de confiance de niveau $1-\alpha$ de μ.

3. Soit $u_1 = Q_{\chi^2_{n-1}}(\beta)$ et $u_2 = Q_{\chi^2_{n-1}}(1-\alpha+\beta)$.

$$\left[\frac{n\overline{V_n}^{(1)}}{u_2}; \frac{n\overline{V_n}^{(1)}}{u_1}\right]$$

est un intervalle de confiance de niveau $1-\alpha$ de σ^2.

Démonstration. On applique le même principe de preuve que le (affreux) théorème du précédent paragraphe et en appliquant le lemme, on a :

1.

$$\begin{aligned}
\frac{\sqrt{n}}{\sigma}\left(\overline{X_n}-\mu\right) \in [z_1; z_2] &\iff \overline{X_n}-\mu \in \left[z_1\frac{\sigma}{\sqrt{n}}; z_2\frac{\sigma}{\sqrt{n}}\right] \\
&\iff \mu - \overline{X_n} \in \left[-z_2\frac{\sigma}{\sqrt{n}}; -z_1\frac{\sigma}{\sqrt{n}}\right] \\
&\iff \mu \in \left[\overline{X_n} - z_2\frac{\sigma}{\sqrt{n}}; \overline{X_n} - z_1\frac{\sigma}{\sqrt{n}}\right]
\end{aligned}$$

2.

$$\begin{aligned}
\sqrt{\frac{n-1}{\overline{V_n}^{(1)}}}\left(\overline{X_n}-\mu\right) \in [t_1; t_2] &\iff \overline{X_n}-\mu \in \left[t_1\sqrt{\frac{\overline{V_n}^{(1)}}{n-1}}; t_2\sqrt{\frac{\overline{V_n}^{(1)}}{n-1}}\right] \\
&\iff \mu - \overline{X_n} \in \left[-t_2\sqrt{\frac{\overline{V_n}^{(1)}}{n-1}}; -t_1\sqrt{\frac{\overline{V_n}^{(1)}}{n-1}}\right] \\
&\iff \mu \in \left[\overline{X_n} - t_2\sqrt{\frac{\overline{V_n}^{(1)}}{n-1}}; \overline{X_n} - t_1\sqrt{\frac{\overline{V_n}^{(1)}}{n-1}}\right]
\end{aligned}$$

3.

$$\begin{aligned}
\frac{n\overline{V_n}^{(1)}}{\sigma^2} \in [u_1; u_2] &\iff \frac{\sigma^2}{n\overline{V_n}^{(1)}} \in \left[\frac{1}{u_2}; \frac{1}{u_1}\right] \\
&\iff \sigma^2 \in \left[\frac{n\overline{V_n}^{(1)}}{u_2}; \frac{n\overline{V_n}^{(1)}}{u_1}\right]
\end{aligned}$$

□

7.3 Loi faible des grands nombres

Lemme 7.3.1 : [Inégalité de Markov] Soient $a > 0$ et Z une variable aléatoire réelle presque surement à valeur positive.

$$\mathbb{P}(Z \geqslant a) \leqslant \frac{\mathbb{E}(Z)}{a}$$

Démonstration. Notons p la fonction de densité de Z. Comme Z est supposée presque surement à valeur positive, on peut supposer que pour tout $z \in \mathbb{R}$, $p(z) = p(z)\mathbb{1}_{\mathbb{R}_+}(z)$. Alors, par le relation de Chalses :

$$\begin{aligned}
\mathbb{E}(Z) &= \int_{\mathbb{R}} zp(z)\, dz \\
&= \int_{[0;+\infty[} zp(z)\, dz \\
&= \int_{]0;a]} zp(z)\, dz + \int_{]a;+\infty[} zp(z)\, dz \\
&\geqslant \int_{]a;+\infty[} zp(z)\, dz \\
&\geqslant \int_{]a;+\infty[} ap(z)\, dz \\
&= a\int_{]a;+\infty[} p(z)\, dz \\
&= a\mathbb{P}(Z \geqslant a)
\end{aligned}$$

□

Lemme 7.3.2 : [Inégalité de Bienaymé-Tchebychev] Soient X une variable aléatoire réelle de variance σ^2 et a un réel strictement positif.

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geqslant a\right) \leqslant \frac{\sigma^2}{a^2}$$

Démonstration. C'est l'inégalité de Markov pour $Z = |X - \mathbb{E}(X)|^2$ □

Ce lemme permet de démontrer le célèbre résultat suivant :

Théorème 7.3.3 Loi faible des grand nombre

Soient $(X_n)_{n\in\mathbb{N}_{>0}}$ une suite de variable aléatoire indépendante et identiquement distribuée de moyenne μ.

$$\forall \varepsilon > 0,\ \lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| \geqslant \varepsilon\right) = 0$$

Démonstration. On applique l'inégalité de Bienaymé-Tchebychev à $X = \frac{1}{n}\sum_{k=1}^{n} X_k$ (et $a = \varepsilon$) qui vérifie par hypothèse $E(X) = \mu$ et $V(X) = \frac{V(X_1)}{n}$:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| \geqslant \varepsilon\right) = \mathbb{P}\left(|X - \mathbb{E}(X)| \geqslant \varepsilon\right) \leqslant \frac{V(X)}{n\varepsilon^2}$$

qui tend bien vers 0 lorsque n tend vers l'infini. □

7.4 Tables

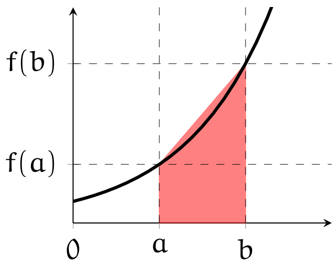
Principe de l'approximation

Pour déterminer les tables suivantes, on approche l'aire sous les fonctions de répartition par la *méthode des trapèzes* ou la *méthode de Simpson* (dites des *paraboles*).

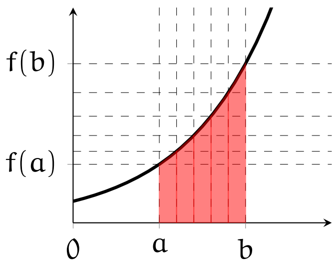
On découpe l'intervalle d'intégration en n morceaux et on fait la somme de l'approximation de l'intégrale sur de petits intervalles $[x_k; x_{k+1}]$ où $x_k = a + \frac{b-a}{n}k$.

Méthode des trapèzes. On approche l'aire sous une courbe par l'aire facile d'un trapèze.

Découpage en 1 trapèze



Découpage en 5 trapèzes



Formule des trapèzes

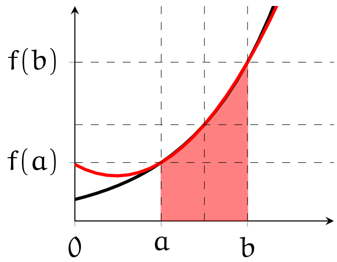
$$\int_a^b f(x)dx = \lim_{n\to+\infty} \sum_{k=0}^{n-1} \frac{b-a}{2n} \left(f(x_k) + f(x_{k+1})\right)$$

L'algorithme en `Python` d'approximation par un découpage en n trapèzes est :

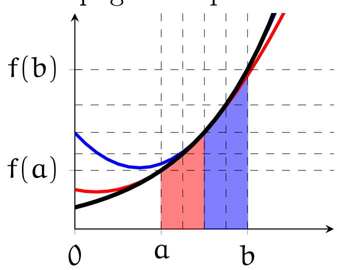
```
def ApproxIntTrapeze(f, a, b, n) :
    h=(b-a)/n
    res=[]
    m=a
    for i in range(n-1) :
        M=m+h
        res.append(h*(f(m)+f(M))/2)
        m=M
    return res
```

Méthode de Simpson. On approche l'aire sous la courbe par celle d'une parabole obtenue par les polynômes d'interpolation de Lagrange. Les trois point d'interpolation son les points aux extrémités et le point médian.

Découpage en 1 parabole



Découpage en 2 paraboles



Formule de Simpson

$$\int_a^b f(x)dx = \lim_{n\to+\infty} \sum_{k=0}^{n-1} \frac{b-a}{6n} \left(f(x_k) + 4f\left(\frac{x_k + x_{k+1}}{2}\right) + f(x_{k+1})\right)$$

L'algorithme en `Python` d'approximation par un découpage en n paraboles est :

```
def ApproxIntSimpson(f, a, b ) :
    h=(b-a)/n
    res=[]
    m=a
    for i in range(n-1) :
        M=m+h
        res.append((h/6)*(f(m)
                        +6*f((m+M)/2)+f(M)))
        m=M
    return res
```

Les valeurs de retour de ces fonctions sont des listes contenant la valeur approché de l'intégrale sur $[x_k; x_{k+1}]$. En particulier la somme de toutes les valeurs de ces listes approchent $\int_a^b f(x)\ dx$.

Construction des tables

Table de la loi normale. Si $X \sim \mathcal{N}(\mu, \sigma)$ alors $X = \sigma Z + \mu$ où $Z \sim \mathcal{N}(0, 1)$ est la loi normale centrée réduite.

Il suffit donc de déterminer les valeurs de $\mathbb{P}(Z \leqslant t)$ pour différente valeur de t. On observe que Z est symétrique, c'est à dire qu'il ne suffit de déterminer $\mathbb{P}(Z \leqslant t)$ que pour des valeurs positives de t. En effet, si t est négatif on a

$$\begin{aligned} \mathbb{P}(Z \leqslant t) &= \mathbb{P}(Z \geqslant -t) \\ &= 1 - \mathbb{P}(Z \leqslant -t) \end{aligned}$$

On rappel que la densité de Z est$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Le principe de l'algorithme est le suivant : on fixe une borne a représentant l'infini (on prendra 4 car $\mathbb{P}(Z > 4) < 0.1\%$ il n'y a donc (presque) plus d'aire après 4). On approche l'intégrale entre sur $[0; a]$ par la méthode des trapèzes. On ajoute case par case les approximations retournées par la fonction `ApproxIntTrapeze` jusqu'à ce que l'on atteingne la valeur t_1 en commençant à `0.5` (car $\mathbb{P}(Z \leqslant 0) = 0.5$). La fonction de répartition étant croissante il suffit d'ajouter les valeurs suivantes de la liste jusqu'à la valeur suivante t_2. Le paramètre `pa`, de précision d'approximation, indique le nombre de décimale que l'on souhaite voir apparaitre.

```
def normale(x) : return (1/sqrt(2*pi))*exp(-x**2/2)

def TableNormale(a, n, pa) :
    X=ApproxIntTrapeze(normale, 0 , a, n)

    case=dict()
    pos=0
    for i in range(a*10) :
        case[i]=dict()
        for j in range(10) :
            if(i==0 and j==0) : case[i][j]=0.5
            else :
                if(j==0) : case[i][j] = case[i-1][9]
                else : case[i][j] = case[i][j-1]
                while(pos/n*a<i/10+j/100 and pos<n) :
                    case[i][j]+=X[pos]
                    pos+=1

    print("Table de la loi normale")
    for j in range(10) :
        print("\t",j/100, end=' ')
    print("")
    for i in range(a*10) :
        print(i/10, end=' ')
        for j in range(10) :
            txt=str(round(case[i][j], pa))
            while(len(txt)<pa+2) : txt+='0'
            print("\t",txt, end=' ')
        print("")
TableNormale(4, 10**7, 3)
```

Table du χ^2. Dans les cas pratiques d'utilisation des lois de χ^2_n, à n degrés de liberté, on s'intéresse à trouver t tel que $\mathbb{P}(X \leqslant t) \simeq 0.9725$ ou 0.99.

On peut démontrer via le théorème de la limite centrale que pour n suffisamment grand χ^2_n s'approchent bien par $\mathcal{N}(n, \sqrt{2n})$ de sorte qu'il n'est pas nécessaire d'effectuer le calcul pour de trop grande valeur de n. On choisi de ne pas aller plus loin que 100.

La fonction de répartition de la loi du χ^2_n est $\gamma_n(x) = c_n \int_0^{\frac{x}{2}} t^{\frac{n}{2}-1} e^{-t}\, dt$ où c_n est la constante $\dfrac{1}{\int_0^{+\infty} t^{\frac{n}{2}-1} e^{-t}\, dt}$.

Le principe de l'algorithme est le suivant : pour chaque de degrés de liberté deg fixés et chaque valeur val, on approche l'intégrale et on cherche la valeur de t satisfaisant $\mathbb{P}(X \leqslant t) \simeq val$

```
def gamma(xmax, deg, n=10**(6)) :
    def f(t) : return t**(n-1) *exp(-t)
    return ApproxIntSimpson(f, 10**(-10), xmax, n)

def TableChi2DEG(deg, VAL, n=10**(6)) :

    gam=gamma(150, deg/2, n)
    gamTOT=sum(gam)

    res=dict()
    pos=0
    som=0
    t=0
    for val in VAL :
        while(som<val*gamTOT and pos<n-1) :
            som+=gam[pos]
            pos+=1
            t+=150/n
        res[val]=2*t
    return res

def TableChi2(n=10**(6), pa=3) :

    case=dict()
    VAL=[0.001, 0.005, 0.025, 0.05, 0.1, 0.9, 0.95, 0.975, 0.995, 0.999]

    for deg in range(1, 10) : case[deg]=TableChi2DEG(deg, VAL, n)

    print("Table du Chi2")
    for val in VAL :
        print("\t",val, end=’ ’)
    print("")
    for deg in range(1, 10) :
        print(deg, end=’ ’)
        for val in VAL :
            txt=str(round(case[deg][val], pa))
            print("\t",txt, end=’ ’)
        print("")
TableChi2(10**7, pa=2)
```

Table de Student. C'est le même principe que la loi du χ^2_n sachant que la densité est $p(x) = c_n \left(1 + \dfrac{x^2}{n}\right)^{-\frac{n+1}{2}}$ où la constante c_n fait de p une fonction de densité. De même, comme pour la loi normale centrée réduite, les lois de Student sont symétriques : $\mathbb{P}(X \leqslant -t) = 1 - \mathbb{P}(X \leqslant t)$.

Enfin pour n suffisamment grand la loi de Student à n degrés de liberté est approché par $\mathcal{N}\left(n, \sqrt{\dfrac{n}{n-2}}\right)$.

```
def TableStudentDEG(deg, VAL, n=10**(6)) :

    def f(x) : return (1+x**2/deg)**(-(deg+1)/2)
```

```
    X=ApproxIntSimpson(f, 0 , 100, n)
    studentTOT=2*sum(X)
    student=[s/studentTOT for s in X]

    res=dict()
    pos=0
    som=0.5
    t=0
    for val in VAL :
        while(som<val and pos<n-1) :
            som+=student[pos]
            pos+=1
            t+=100/n
        res[val]=t
    return res

def TableStudent(n=10**(6), pa=3) :

    case=dict()
    VAL=[0.6, 0.7, 0.8, 0.9, 0.95, 0.975, 0.99, 0.995, 0.999]

    for deg in range(1, 11) : case[deg]=TableStudentDEG(deg, VAL, 100, n)

    print("Table de Student")
    for val in VAL :
        print("\t",val, end=’ ’)
    print("")
    for deg in range(1, 11) :
        print(deg, end=’ ’)
        for val in VAL :
            txt=str(round(case[deg][val], pa))
            print("\t",txt, end=’ ’)
        print("")
TableStudent()
```

Loi normale centrée réduite

Dans le tableau, à l'intersection de la ligne i et de la colonne j approche (assez bien) $\mathbb{P}(Z \leqslant i+j)$ pour $Z \sim \mathcal{N}(0,1)$.

	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59484	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

Table des lois du χ^2

Si $X \sim \chi^2_n$ alors dans le tableau, la valeur t à l'intersection de la ligne n et de la colonne m vérifie (assez bien) $\mathbb{P}(X \leqslant t) = m$.

	0.001	0.005	0.025	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.995	0.999
1	0.0	0.0	0.001	0.003	0.015	0.062	0.145	0.271	0.45	0.703	1.069	1.636	2.699	3.834	5.017	7.872	10.819
2	0.002	0.01	0.051	0.103	0.211	0.446	0.713	1.022	1.386	1.833	2.408	3.219	4.605	5.991	7.378	10.597	13.816
3	0.024	0.072	0.216	0.352	0.584	1.005	1.424	1.869	2.366	2.946	3.665	4.642	6.251	7.815	9.348	12.838	16.266
4	0.091	0.207	0.484	0.711	1.064	1.649	2.195	2.753	3.357	4.045	4.878	5.989	7.779	9.488	11.143	14.86	18.467
5	0.21	0.412	0.831	1.145	1.61	2.343	3.0	3.656	4.351	5.132	6.064	7.289	9.236	11.07	12.833	16.75	20.515
6	0.381	0.676	1.237	1.635	2.204	3.07	3.828	4.57	5.348	6.211	7.231	8.558	10.645	12.592	14.449	18.548	22.458
7	0.598	0.989	1.69	2.167	2.833	3.822	4.671	5.493	6.346	7.283	8.383	9.803	12.017	14.067	16.013	20.278	24.322
8	0.857	1.344	2.18	2.733	3.49	4.594	5.527	6.423	7.344	8.351	9.524	11.03	13.362	15.507	17.535	21.955	26.124
9	1.152	1.735	2.7	3.325	4.168	5.38	6.393	7.357	8.343	9.414	10.656	12.242	14.684	16.919	19.023	23.589	27.877
10	1.479	2.156	3.247	3.94	4.865	6.179	7.267	8.295	9.342	10.473	11.781	13.442	15.987	18.307	20.483	25.188	29.588
11	1.834	2.603	3.816	4.575	5.578	6.989	8.148	9.237	10.341	11.53	12.899	14.631	17.275	19.675	21.92	26.757	31.264
12	2.214	3.074	4.404	5.226	6.304	7.807	9.034	10.182	11.34	12.584	14.011	15.812	18.549	21.026	23.337	28.3	32.909
13	2.617	3.565	5.009	5.892	7.042	8.634	9.926	11.129	12.34	13.636	15.119	16.985	19.812	22.362	24.736	29.819	34.528
14	3.041	4.075	5.629	6.571	7.79	9.467	10.821	12.078	13.339	14.685	16.222	18.151	21.064	23.685	26.119	31.319	36.123
15	3.483	4.601	6.262	7.261	8.547	10.307	11.721	13.03	14.339	15.733	17.322	19.311	22.307	24.996	27.488	32.801	37.697
16	3.942	5.142	6.908	7.962	9.312	11.152	12.624	13.983	15.339	16.78	18.418	20.465	23.542	26.296	28.845	34.267	39.252
17	4.416	5.697	7.564	8.672	10.085	12.002	13.531	14.937	16.338	17.824	19.511	21.615	24.769	27.587	30.191	35.718	40.79
18	4.905	6.265	8.231	9.39	10.865	12.857	14.44	15.893	17.338	18.868	20.601	22.76	25.989	28.869	31.526	37.156	42.312
19	5.407	6.844	8.907	10.117	11.651	13.716	15.352	16.85	18.338	19.91	21.689	23.9	27.204	30.144	32.852	38.582	43.82
20	5.921	7.434	9.591	10.851	12.443	14.578	16.266	17.809	19.337	20.951	22.775	25.038	28.412	31.41	34.17	39.997	45.315
21	6.447	8.034	10.283	11.591	13.24	15.445	17.182	18.768	20.337	21.992	23.858	26.171	29.615	32.671	35.479	41.401	46.797
22	6.983	8.643	10.982	12.338	14.042	16.314	18.101	19.729	21.337	23.031	24.939	27.301	30.813	33.924	36.781	42.796	48.268
23	7.529	9.26	11.689	13.091	14.848	17.187	19.021	20.69	22.337	24.069	26.018	28.429	32.007	35.172	38.076	44.181	49.728
24	8.085	9.886	12.401	13.848	15.659	18.062	19.943	21.652	23.337	25.106	27.096	29.553	33.196	36.415	39.364	45.559	51.179
25	8.649	10.52	13.12	14.611	16.473	18.94	20.867	22.616	24.337	26.143	28.172	30.675	34.382	37.652	40.646	46.928	52.62

	0.001	0.005	0.025	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.995	0.999
26	9.222	11.16	13.844	15.379	17.292	19.82	21.792	23.579	25.336	27.179	29.246	31.795	35.563	38.885	41.923	48.29	54.052
27	9.803	11.808	14.573	16.151	18.114	20.703	22.719	24.544	26.336	28.214	30.319	32.912	36.741	40.113	43.195	49.645	55.476
28	10.391	12.461	15.308	16.928	18.939	21.588	23.647	25.509	27.336	29.249	31.391	34.027	37.916	41.337	44.461	50.993	56.892
29	10.986	13.121	16.047	17.708	19.768	22.475	24.577	26.475	28.336	30.283	32.461	35.139	39.087	42.557	45.722	52.336	58.301
30	11.588	13.787	16.791	18.493	20.599	23.364	25.508	27.442	29.336	31.316	33.53	36.25	40.256	43.773	46.979	53.672	59.703
31	12.196	14.458	17.539	19.281	21.434	24.255	26.44	28.409	30.336	32.349	34.598	37.359	41.422	44.985	48.232	55.003	61.098
32	12.811	15.134	18.291	20.072	22.271	25.148	27.373	29.376	31.336	33.381	35.665	38.466	42.585	46.194	49.48	56.328	62.487
33	13.431	15.815	19.047	20.867	23.11	26.042	28.307	30.344	32.336	34.413	36.731	39.572	43.745	47.4	50.725	57.648	63.87
34	14.057	16.501	19.806	21.664	23.952	26.938	29.242	31.313	33.336	35.444	37.795	40.676	44.903	48.602	51.966	58.964	65.247
35	14.688	17.192	20.569	22.465	24.797	27.836	30.178	32.282	34.336	36.475	38.859	41.778	46.059	49.802	53.203	60.275	66.619
36	15.324	17.887	21.336	23.269	25.643	28.735	31.115	33.252	35.336	37.505	39.922	42.879	47.212	50.998	54.437	61.581	67.985
37	15.965	18.586	22.106	24.075	26.492	29.635	32.053	34.222	36.336	38.535	40.984	43.978	48.363	52.192	55.668	62.883	69.346
38	16.611	19.289	22.878	24.884	27.343	30.537	32.992	35.192	37.335	39.564	42.045	45.076	49.513	53.384	56.896	64.181	70.703
39	17.262	19.996	23.654	25.695	28.196	31.441	33.932	36.163	38.335	40.593	43.105	46.173	50.66	54.572	58.12	65.476	72.055
40	17.916	20.707	24.433	26.509	29.051	32.345	34.872	37.134	39.335	41.622	44.165	47.269	51.805	55.758	59.342	66.766	73.402
41	18.575	21.421	25.215	27.326	29.907	33.251	35.813	38.106	40.335	42.651	45.224	48.363	52.949	56.942	60.561	68.053	74.745
42	19.239	22.138	25.999	28.144	30.765	34.157	36.755	39.077	41.335	43.679	46.282	49.456	54.09	58.124	61.777	69.336	76.084
43	19.906	22.859	26.785	28.965	31.625	35.065	37.698	40.05	42.335	44.706	47.339	50.548	55.23	59.304	62.99	70.616	77.419
44	20.576	23.584	27.575	29.787	32.487	35.974	38.641	41.022	43.335	45.734	48.396	51.639	56.369	60.481	64.201	71.893	78.75
45	21.251	24.311	28.366	30.612	33.35	36.884	39.585	41.995	44.335	46.761	49.452	52.729	57.505	61.656	65.41	73.166	80.077
46	21.929	25.041	29.16	31.439	34.215	37.795	40.529	42.968	45.335	47.787	50.507	53.818	58.641	62.83	66.617	74.437	81.4
47	22.61	25.775	29.956	32.268	35.081	38.708	41.474	43.942	46.335	48.814	51.562	54.906	59.774	64.001	67.821	75.704	82.72
48	23.295	26.511	30.755	33.098	35.949	39.621	42.42	44.915	47.335	49.84	52.616	55.993	60.907	65.171	69.023	76.969	84.037
49	23.983	27.249	31.555	33.93	36.818	40.534	43.366	45.889	48.335	50.866	53.67	57.079	62.038	66.339	70.222	78.231	85.351
50	24.674	27.991	32.357	34.764	37.689	41.449	44.313	46.864	49.335	51.892	54.723	58.164	63.167	67.505	71.42	79.49	86.661

	0.001	0.005	0.025	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.995	0.999
51	25.368	28.735	33.162	35.6	38.56	42.365	45.261	47.838	50.335	52.917	55.775	59.248	64.295	68.669	72.616	80.747	87.968
52	26.065	29.481	33.968	36.437	39.433	43.281	46.209	48.813	51.335	53.942	56.827	60.332	65.422	69.832	73.81	82.001	89.272
53	26.765	30.23	34.776	37.276	40.308	44.199	47.157	49.788	52.335	54.967	57.879	61.414	66.548	70.993	75.002	83.253	90.573
54	27.468	30.981	35.586	38.116	41.183	45.117	48.106	50.764	53.335	55.992	58.93	62.496	67.673	72.153	76.192	84.502	91.872
55	28.173	31.735	36.398	38.958	42.06	46.036	49.055	51.739	54.335	57.016	59.98	63.577	68.796	73.311	77.38	85.749	93.168
56	28.881	32.49	37.212	39.801	42.937	46.955	50.005	52.715	55.335	58.04	61.031	64.658	69.919	74.468	78.567	86.994	94.461
57	29.592	33.248	38.027	40.646	43.816	47.876	50.956	53.691	56.335	59.064	62.08	65.737	71.04	75.624	79.752	88.236	95.751
58	30.305	34.008	38.844	41.492	44.696	48.797	51.906	54.667	57.335	60.088	63.129	66.816	72.16	76.778	80.936	89.477	97.039
59	31.02	34.77	39.662	42.339	45.577	49.718	52.858	55.643	58.335	61.112	64.178	67.894	73.279	77.931	82.117	90.715	98.324
60	31.738	35.535	40.482	43.188	46.459	50.641	53.809	56.62	59.335	62.135	65.227	68.972	74.397	79.082	83.298	91.952	99.607
61	32.459	36.301	41.303	44.038	47.342	51.564	54.761	57.597	60.335	63.158	66.274	70.049	75.514	80.232	84.476	93.186	100.888
62	33.181	37.068	42.126	44.889	48.226	52.487	55.714	58.574	61.335	64.181	67.322	71.125	76.63	81.381	85.654	94.419	102.166
63	33.906	37.838	42.95	45.741	49.111	53.412	56.666	59.551	62.335	65.204	68.369	72.201	77.745	82.529	86.83	95.649	103.442
64	34.633	38.61	43.776	46.595	49.996	54.337	57.62	60.528	63.335	66.226	69.416	73.276	78.86	83.675	88.004	96.878	104.716
65	35.362	39.383	44.603	47.45	50.883	55.262	58.573	61.506	64.335	67.249	70.462	74.351	79.973	84.821	89.177	98.105	105.988
66	36.093	40.158	45.431	48.305	51.77	56.188	59.527	62.484	65.335	68.271	71.508	75.424	81.085	85.965	90.349	99.33	107.258
67	36.826	40.935	46.261	49.162	52.659	57.115	60.481	63.461	66.335	69.293	72.554	76.498	82.197	87.108	91.519	100.554	108.526
68	37.561	41.713	47.092	50.02	53.548	58.042	61.436	64.44	67.335	70.315	73.6	77.571	83.308	88.25	92.689	101.776	109.791
69	38.298	42.494	47.924	50.879	54.438	58.97	62.391	65.418	68.335	71.337	74.645	78.643	84.418	89.391	93.856	102.996	111.055
70	39.036	43.275	48.758	51.739	55.329	59.898	63.346	66.396	69.334	72.358	75.689	79.715	85.527	90.531	95.023	104.215	112.317
71	39.777	44.058	49.592	52.6	56.221	60.827	64.302	67.375	70.334	73.38	76.734	80.786	86.635	91.67	96.189	105.432	113.577
72	40.519	44.843	50.428	53.462	57.113	61.756	65.258	68.353	71.334	74.401	77.778	81.857	87.743	92.808	97.353	106.648	114.835
73	41.264	45.629	51.265	54.325	58.006	62.686	66.214	69.332	72.334	75.422	78.822	82.927	88.85	93.945	98.516	107.862	116.092
74	42.01	46.417	52.103	55.189	58.9	63.616	67.17	70.311	73.334	76.443	79.865	83.997	89.956	95.081	99.678	109.074	117.346
75	42.757	47.206	52.942	56.054	59.795	64.547	68.127	71.29	74.334	77.464	80.908	85.066	91.061	96.217	100.839	110.286	118.599

	0.001	0.005	0.025	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.995	0.999
76	43.507	47.997	53.782	56.92	60.69	65.478	69.084	72.27	75.334	78.485	81.951	86.135	92.166	97.351	101.999	111.495	119.85
77	44.258	48.788	54.623	57.786	61.586	66.409	70.042	73.249	76.334	79.505	82.994	87.203	93.27	98.484	103.158	112.704	121.1
78	45.01	49.582	55.466	58.654	62.483	67.341	70.999	74.228	77.334	80.526	84.036	88.271	94.374	99.617	104.316	113.911	122.348
79	45.764	50.376	56.309	59.522	63.38	68.274	71.957	75.208	78.334	81.546	85.078	89.338	95.476	100.749	105.473	115.117	123.594
80	46.52	51.172	57.153	60.391	64.278	69.207	72.915	76.188	79.334	82.566	86.12	90.405	96.578	101.879	106.629	116.321	124.839
81	47.277	51.969	57.998	61.261	65.176	70.14	73.874	77.168	80.334	83.586	87.161	91.472	97.68	103.01	107.783	117.524	126.083
82	48.036	52.767	58.845	62.132	66.076	71.074	74.833	78.148	81.334	84.606	88.202	92.538	98.78	104.139	108.937	118.726	127.324
83	48.796	53.567	59.692	63.004	66.976	72.008	75.792	79.128	82.334	85.626	89.243	93.604	99.88	105.267	110.09	119.927	128.565
84	49.557	54.368	60.54	63.876	67.876	72.943	76.751	80.108	83.334	86.646	90.284	94.669	100.98	106.395	111.242	121.126	129.804
85	50.32	55.17	61.389	64.749	68.777	73.878	77.71	81.089	84.334	87.665	91.325	95.734	102.079	107.522	112.393	122.325	131.041
86	51.085	55.973	62.239	65.623	69.679	74.813	78.67	82.069	85.334	88.685	92.365	96.799	103.177	108.648	113.544	123.522	132.277
87	51.85	56.777	63.089	66.498	70.581	75.749	79.63	83.05	86.334	89.704	93.405	97.863	104.275	109.773	114.693	124.718	133.512
88	52.617	57.582	63.941	67.373	71.484	76.685	80.59	84.031	87.334	90.723	94.445	98.927	105.372	110.898	115.841	125.913	134.745
89	53.386	58.389	64.793	68.249	72.387	77.622	81.55	85.012	88.334	91.742	95.484	99.991	106.469	112.022	116.989	127.106	135.978
90	54.155	59.196	65.647	69.126	73.291	78.558	82.511	85.993	89.334	92.761	96.524	101.054	107.565	113.145	118.136	128.299	137.208
91	54.926	60.005	66.501	70.003	74.196	79.496	83.472	86.974	90.334	93.78	97.563	102.117	108.661	114.268	119.282	129.491	138.438
92	55.698	60.815	67.356	70.882	75.1	80.433	84.433	87.955	91.334	94.799	98.602	103.179	109.756	115.39	120.427	130.681	139.666
93	56.472	61.625	68.211	71.76	76.006	81.371	85.394	88.936	92.334	95.818	99.641	104.241	110.85	116.511	121.571	131.871	140.893
94	57.246	62.437	69.068	72.64	76.912	82.309	86.356	89.918	93.334	96.836	100.679	105.303	111.944	117.632	122.715	133.059	142.119
95	58.022	63.25	69.925	73.52	77.818	83.248	87.318	90.899	94.334	97.855	101.717	106.364	113.038	118.752	123.858	134.247	143.344
96	58.799	64.063	70.783	74.401	78.725	84.187	88.279	91.881	95.334	98.873	102.755	107.425	114.131	119.871	125.0	135.433	144.567
97	59.577	64.878	71.642	75.282	79.633	85.126	89.241	92.862	96.334	99.892	103.793	108.486	115.223	120.99	126.141	136.619	145.789
98	60.356	65.694	72.501	76.164	80.541	86.065	90.204	93.844	97.334	100.91	104.831	109.547	116.315	122.108	127.282	137.803	147.01
99	61.137	66.51	73.361	77.046	81.449	87.005	91.166	94.826	98.334	101.928	105.868	110.607	117.407	123.225	128.422	138.987	148.23
100	61.918	67.328	74.222	77.929	82.358	87.945	92.129	95.808	99.334	102.946	106.906	111.667	118.498	124.342	129.561	140.169	149.449

Table des lois de Student

Si $X \sim \mathcal{T}_n$ alors dans le tableau, la valeur t à l'intersection de la ligne n et de la colonne m vérifie (assez bien) $\mathbb{P}(X \leqslant t) = m$.

	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.995	0.999
1	0.3227	0.7205	1.3593	2.9967	5.969	11.3414	24.3062	39.1924	76.7861
2	0.2878	0.6152	1.0561	1.8713	2.8783	4.1854	6.5232	8.7779	14.453
3	0.2766	0.5842	0.978	1.6365	2.3502	3.1746	4.5147	5.7766	9.7132
4	0.2707	0.5686	0.9409	1.5331	2.1315	2.7756	3.7444	4.5982	7.1307
5	0.2672	0.5594	0.9195	1.4759	2.015	2.5705	3.3646	4.0314	5.8885
6	0.2648	0.5534	0.9057	1.4398	1.9432	2.4469	3.1426	3.7073	5.2069
7	0.2632	0.5491	0.896	1.4149	1.8946	2.3646	2.998	3.4995	4.7852
8	0.2619	0.5459	0.8889	1.3968	1.8595	2.306	2.8965	3.3554	4.5008
9	0.261	0.5435	0.8834	1.383	1.8331	2.2622	2.8214	3.2498	4.2968
10	0.2602	0.5415	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437
11	0.2596	0.5399	0.8755	1.3634	1.7959	2.201	2.7181	3.1058	4.0247
12	0.259	0.5386	0.8726	1.3562	1.7823	2.1788	2.681	3.0546	3.9296
13	0.2586	0.5375	0.8702	1.3502	1.771	2.1604	2.6503	3.0123	3.852
14	0.2582	0.5366	0.8681	1.345	1.7613	2.1448	2.6245	2.9768	3.7874
15	0.2579	0.5357	0.8662	1.3406	1.7531	2.1315	2.6025	2.9467	3.7328
16	0.2576	0.535	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862
17	0.2574	0.5344	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458
18	0.2571	0.5338	0.862	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105
19	0.2569	0.5333	0.861	1.3277	1.7291	2.093	2.5395	2.8609	3.5794
20	0.2567	0.5329	0.86	1.3253	1.7247	2.086	2.528	2.8453	3.5518
21	0.2566	0.5325	0.8591	1.3232	1.7208	2.0796	2.5177	2.8314	3.5272
22	0.2564	0.5321	0.8583	1.3212	1.7172	2.0739	2.5083	2.8188	3.505
23	0.2563	0.5318	0.8575	1.3195	1.7139	2.0687	2.4999	2.8073	3.485
24	0.2562	0.5314	0.8569	1.3178	1.7109	2.0639	2.4922	2.797	3.4668
25	0.2561	0.5312	0.8563	1.3164	1.7082	2.0595	2.4851	2.7875	3.4502
26	0.256	0.5309	0.8557	1.315	1.7056	2.0555	2.4786	2.7787	3.435
27	0.2559	0.5307	0.8552	1.3137	1.7033	2.0518	2.4727	2.7707	3.421
28	0.2558	0.5304	0.8547	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082
29	0.2557	0.5302	0.8542	1.3114	1.6991	2.0452	2.462	2.7564	3.3962
30	0.2556	0.53	0.8538	1.3104	1.6973	2.0423	2.4573	2.75	3.3852
31	0.2555	0.5299	0.8534	1.3095	1.6955	2.0395	2.4528	2.7441	3.3749
32	0.2555	0.5297	0.853	1.3086	1.6939	2.037	2.4487	2.7385	3.3653
33	0.2554	0.5295	0.8527	1.3078	1.6924	2.0345	2.4448	2.7333	3.3563
34	0.2553	0.5294	0.8523	1.307	1.6909	2.0323	2.4412	2.7284	3.348
35	0.2553	0.5292	0.852	1.3062	1.6896	2.0301	2.4377	2.7238	3.34
36	0.2552	0.5291	0.8517	1.3055	1.6883	2.0281	2.4345	2.7195	3.3326
37	0.2552	0.529	0.8514	1.3049	1.6871	2.0262	2.4314	2.7154	3.3256
38	0.2551	0.5288	0.8512	1.3042	1.686	2.0244	2.4286	2.7116	3.319
39	0.2551	0.5287	0.851	1.3036	1.6849	2.0227	2.4258	2.7079	3.3128
40	0.2551	0.5286	0.8507	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069

	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.995	0.999
41	0.255	0.5285	0.8505	1.3025	1.6829	2.0195	2.4208	2.7012	3.3013
42	0.255	0.5284	0.8503	1.3021	1.682	2.0181	2.4185	2.6981	3.296
43	0.2549	0.5283	0.8501	1.3016	1.6811	2.0167	2.4163	2.6951	3.2909
44	0.2549	0.5282	0.8499	1.3011	1.6802	2.0154	2.4141	2.6923	3.2861
45	0.2549	0.5281	0.8497	1.3007	1.6794	2.0141	2.4121	2.6896	3.2815
46	0.2548	0.5281	0.8495	1.3002	1.6787	2.0129	2.4102	2.687	3.2771
47	0.2548	0.528	0.8493	1.2998	1.6779	2.0118	2.4084	2.6846	3.2729
48	0.2548	0.5279	0.8492	1.2994	1.6772	2.0106	2.4066	2.6822	3.2689
49	0.2547	0.5278	0.849	1.2991	1.6766	2.0096	2.4049	2.68	3.2651
50	0.2547	0.5278	0.8489	1.2987	1.6759	2.0086	2.4033	2.6778	3.2614
51	0.2547	0.5277	0.8487	1.2984	1.6753	2.0076	2.4017	2.6757	3.2579
52	0.2547	0.5276	0.8486	1.2981	1.6747	2.0066	2.4002	2.6737	3.2545
53	0.2546	0.5276	0.8485	1.2977	1.6741	2.0058	2.3988	2.6718	3.2513
54	0.2546	0.5275	0.8483	1.2974	1.6736	2.0049	2.3974	2.67	3.2482
55	0.2546	0.5275	0.8482	1.2971	1.673	2.0041	2.3961	2.6682	3.2452
56	0.2546	0.5274	0.8481	1.2969	1.6725	2.0032	2.3948	2.6665	3.2423
57	0.2545	0.5274	0.848	1.2966	1.672	2.0025	2.3936	2.6649	3.2395
58	0.2545	0.5273	0.8479	1.2963	1.6716	2.0017	2.3924	2.6633	3.2368
59	0.2545	0.5272	0.8478	1.2961	1.6711	2.001	2.3912	2.6618	3.2342
60	0.2545	0.5272	0.8477	1.2958	1.6707	2.0003	2.3901	2.6603	3.2317
61	0.2545	0.5272	0.8476	1.2956	1.6702	1.9996	2.3891	2.6589	3.2293
62	0.2544	0.5271	0.8475	1.2954	1.6698	1.999	2.388	2.6575	3.227
63	0.2544	0.5271	0.8474	1.2951	1.6694	1.9983	2.387	2.6562	3.2247
64	0.2544	0.527	0.8473	1.2949	1.669	1.9977	2.386	2.6549	3.2225
65	0.2544	0.527	0.8472	1.2947	1.6686	1.9971	2.3851	2.6536	3.2204
66	0.2544	0.5269	0.8471	1.2945	1.6683	1.9966	2.3842	2.6524	3.2184
67	0.2544	0.5269	0.847	1.2943	1.6679	1.996	2.3833	2.6512	3.2164
68	0.2544	0.5269	0.8469	1.2941	1.6676	1.9955	2.3824	2.6501	3.2145
69	0.2543	0.5268	0.8469	1.294	1.6672	1.995	2.3816	2.649	3.2126
70	0.2543	0.5268	0.8468	1.2938	1.6669	1.9944	2.3808	2.6479	3.2108
71	0.2543	0.5268	0.8467	1.2936	1.6666	1.994	2.38	2.6469	3.209
72	0.2543	0.5267	0.8467	1.2934	1.6663	1.9935	2.3793	2.6459	3.2073
73	0.2543	0.5267	0.8466	1.2933	1.666	1.993	2.3785	2.6449	3.2057
74	0.2543	0.5267	0.8465	1.2931	1.6657	1.9925	2.3778	2.6439	3.2041
75	0.2543	0.5266	0.8464	1.2929	1.6654	1.9921	2.3771	2.643	3.2025
76	0.2542	0.5266	0.8464	1.2928	1.6652	1.9917	2.3764	2.6421	3.201
77	0.2542	0.5266	0.8463	1.2926	1.6649	1.9913	2.3758	2.6412	3.1995
78	0.2542	0.5266	0.8463	1.2925	1.6646	1.9909	2.3751	2.6404	3.198
79	0.2542	0.5265	0.8462	1.2924	1.6644	1.9905	2.3745	2.6395	3.1966
80	0.2542	0.5265	0.8461	1.2922	1.6641	1.9901	2.3739	2.6387	3.1953

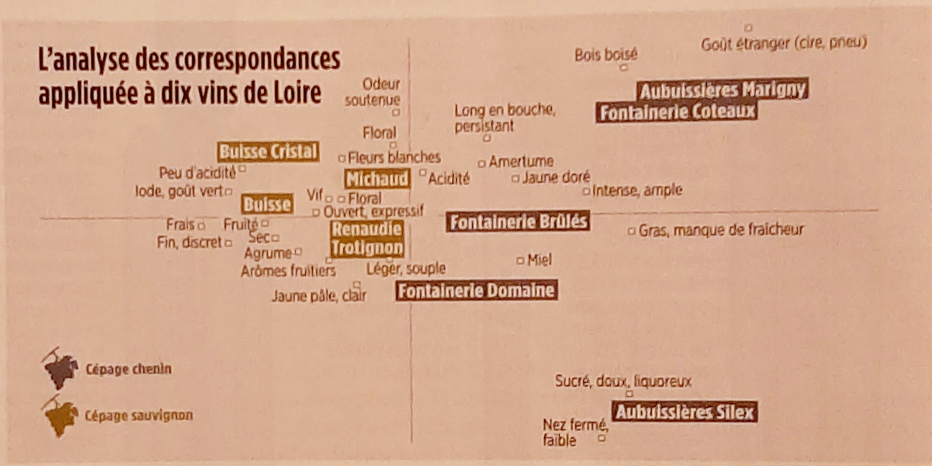
	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.995	0.999
81	0.2542	0.5265	0.8461	1.2921	1.6639	1.9897	2.3733	2.6379	3.1939
82	0.2542	0.5265	0.846	1.292	1.6637	1.9893	2.3727	2.6371	3.1926
83	0.2542	0.5264	0.846	1.2918	1.6634	1.989	2.3721	2.6364	3.1914
84	0.2542	0.5264	0.8459	1.2917	1.6632	1.9886	2.3716	2.6356	3.1901
85	0.2541	0.5264	0.8459	1.2916	1.663	1.9883	2.371	2.6349	3.1889
86	0.2541	0.5264	0.8458	1.2915	1.6628	1.988	2.3705	2.6342	3.1877
87	0.2541	0.5263	0.8458	1.2914	1.6626	1.9876	2.37	2.6335	3.1866
88	0.2541	0.5263	0.8457	1.2913	1.6624	1.9873	2.3695	2.6329	3.1854
89	0.2541	0.5263	0.8457	1.2911	1.6622	1.987	2.369	2.6322	3.1844
90	0.2541	0.5263	0.8457	1.291	1.662	1.9867	2.3685	2.6316	3.1833
91	0.2541	0.5262	0.8456	1.2909	1.6618	1.9864	2.368	2.6309	3.1822
92	0.2541	0.5262	0.8456	1.2908	1.6616	1.9861	2.3676	2.6303	3.1812
93	0.2541	0.5262	0.8455	1.2907	1.6614	1.9858	2.3671	2.6297	3.1802
94	0.2541	0.5262	0.8455	1.2906	1.6612	1.9855	2.3667	2.6292	3.1792
95	0.2541	0.5262	0.8454	1.2905	1.6611	1.9853	2.3663	2.6286	3.1783
96	0.2541	0.5261	0.8454	1.2904	1.6609	1.985	2.3658	2.628	3.1773
97	0.254	0.5261	0.8453	1.2903	1.6607	1.9847	2.3654	2.6275	3.1764
98	0.254	0.5261	0.8453	1.2903	1.6606	1.9845	2.365	2.6269	3.1755
99	0.254	0.5261	0.8453	1.2902	1.6604	1.9842	2.3646	2.6264	3.1746
100	0.254	0.5261	0.8452	1.2901	1.6602	1.984	2.3642	2.6259	3.1737
101	0.254	0.5261	0.8452	1.29	1.6601	1.9837	2.3638	2.6254	3.1729
102	0.254	0.5261	0.8452	1.2899	1.6599	1.9835	2.3635	2.6249	3.1721
103	0.254	0.526	0.8451	1.2898	1.6598	1.9833	2.3631	2.6244	3.1713
104	0.254	0.526	0.8451	1.2898	1.6596	1.983	2.3627	2.6239	3.1705
105	0.254	0.526	0.8451	1.2897	1.6595	1.9828	2.3624	2.6235	3.1697
106	0.254	0.526	0.845	1.2896	1.6594	1.9826	2.3621	2.623	3.1689
107	0.254	0.526	0.845	1.2895	1.6592	1.9824	2.3617	2.6226	3.1682
108	0.254	0.526	0.845	1.2895	1.6591	1.9822	2.3614	2.6221	3.1674
109	0.254	0.5259	0.8449	1.2894	1.659	1.982	2.3611	2.6217	3.1667
110	0.254	0.5259	0.8449	1.2893	1.6588	1.9818	2.3607	2.6213	3.166
111	0.254	0.5259	0.8449	1.2892	1.6587	1.9816	2.3604	2.6209	3.1653
112	0.254	0.5259	0.8449	1.2892	1.6586	1.9814	2.3601	2.6204	3.1646
113	0.254	0.5259	0.8448	1.2891	1.6585	1.9812	2.3598	2.62	3.1639
114	0.254	0.5259	0.8448	1.289	1.6583	1.981	2.3595	2.6197	3.1633
115	0.2539	0.5259	0.8448	1.289	1.6582	1.9808	2.3592	2.6193	3.1626
116	0.2539	0.5259	0.8447	1.2889	1.6581	1.9806	2.3589	2.6189	3.162
117	0.2539	0.5258	0.8447	1.2888	1.658	1.9805	2.3587	2.6185	3.1614
118	0.2539	0.5258	0.8447	1.2888	1.6579	1.9803	2.3584	2.6181	3.1607
119	0.2539	0.5258	0.8447	1.2887	1.6578	1.9801	2.3581	2.6178	3.1601
120	0.2539	0.5258	0.8446	1.2887	1.6577	1.9799	2.3578	2.6174	3.1595

7.5 A propos de l'analyse par correspondance

Extrait de *Sciences et Avenir*.

Les statistiques, un outil pour comprendre les mots du vin

Deux chercheurs ont utilisé une méthode mathématique pour étudier les corrélations entre les noms de vins et les adjectifs qui les caractérisent.



SYLVIE DAOUDAL

La méthode permet de représenter sur le même graphique les vins et les mots qui les décrivent, afin de visualiser la proximité de deux cuvées du même cépage.

FRUITÉ, BOISÉ, INTENSE... LES ŒNOLOGUES ne manquent pas de vocabulaire pour caractériser un vin. François Husson et Jérôme Pagès, du Laboratoire de mathématiques appliquées d'Agro-campus Ouest à Rennes, avec Jean-Paul Gouvazé et Christian Asselin, d'InterLoire (organisation interprofessionnelle des vins de Loire), ont voulu savoir si ces adjectifs correspondaient à un profil typique de vin ou s'ils étaient indifféremment employés pour des cépages très variés.

Dix vins de Loire — cinq cépages sauvignon, appellation touraine, cinq cépages chenin, appellation vouvray — ont été présentés lors d'une dégustation à l'aveugle à 12 œnologues. Ceux-ci ont eu la liberté de décrire le nectar avec leurs propres mots, les mathématiciens tenant compte au final uniquement des termes ayant été cités plus de trois fois. 30 mots, prononcés chacun entre 20 et 28 fois, ont ainsi été soumis à une méthode statistique appelée analyse des correspondances *(lire l'encadré)*, dont la spécificité est de placer sur le même tableau à la fois les variables (les vins) et les valeurs (les mots). Le résultat est présenté sous la forme d'un graphique *(voir ci-dessus)*.

On y distingue deux pôles bien définis par un nombre restreint d'adjectifs : le vouvray domaine des Aubuissières silex, caractérisé par des mots évoquant le sucré (il est effectivement le seul à contenir du sucre résiduel), et l'Aubuissières Marigny ainsi que le domaine de la Fontainerie Coteaux qualifiés de « boisés » : ces deux vins ont été élevés en fût. Un troisième pôle, constitué des cinq vins d'appellation Touraine, est quant à lui moins défini, ceux-ci étant souvent qualifiés de « frais » et « aromatique ». L'axe vertical sépare les sauvignons (à gauche, en vert) des chenins (à droite, en bleu) sur la base de la fraîcheur et de l'expression aromatique. Quant à l'axe horizontal, il différencie les chenins entre eux selon « goût sucré/goût boisé ». Que conclure ? « *Que les sauvignons sont homogènes et les chenins divers* », explique François Husson. Autrement dit, il n'y aurait qu'une seule façon de faire du sauvignon et de nombreuses façons de faire du chenin. ■ **Azar Khalatbari**

THÉORIE

L'analyse des correspondances

Cette méthode a été conçue à Rennes au début des années 1960 par Jean-Paul Benzécri, fondateur de l'École française d'analyse des données. Elle est couramment utilisée pour représenter le profil des différents produits et leur proximité. Elle consiste à mettre en correspondance deux nuages de points. Dans l'exemple des vins, dix points représentent les dix crus : plus les vins sont décrits par les mêmes adjectifs, plus leur position dans le nuage est proche. Le second nuage représente les adjectifs : deux adjectifs utilisés pour décrire les mêmes vins sont proches. *« La méthode de l'analyse des correspondances ramène l'ensemble des points de ces deux nuages au niveau d'un plan. »* Ainsi chaque vin se trouve du côté des mots avec lesquels il a été souvent associé et chaque mot à proximité du vin qu'il qualifie.