Probabilités I

(‘S110016 CR’)

GSEM, Université de Genève

Semestre de printemps 2020

Dr. Enrico Chavez

Transparents de cours — Version du 20.02.2020

Table des matières

1. Introduction

Qu'est-ce que la statistique?

- De nos jours, les données et l'analyse de données jouent un rôle de plus en plus prépondérant dans les entreprises, sociétés et organisations.

- Il est largement admis que l'avantage concurrentiel sur le marché reviendra aux entreprises qui sauront utiliser les '*big data*' (ou mégadonnées, parfois appelées données massives) et l'analyse des données ('*analytics*') pour identifier les tendances en matière de consommation suffisamment tôt, percer les préférences des consommateurs et rendre leur propre fonctionnement plus efficace.

‘Analytics will define the difference between the losers and winners going forward.’

Tim McGuire, 2013

Source: Interview avec Tim McGuire, directeur chez McKinsey, en mars 2013 (`youtu.be/Sc5FFY-IVDQ`).

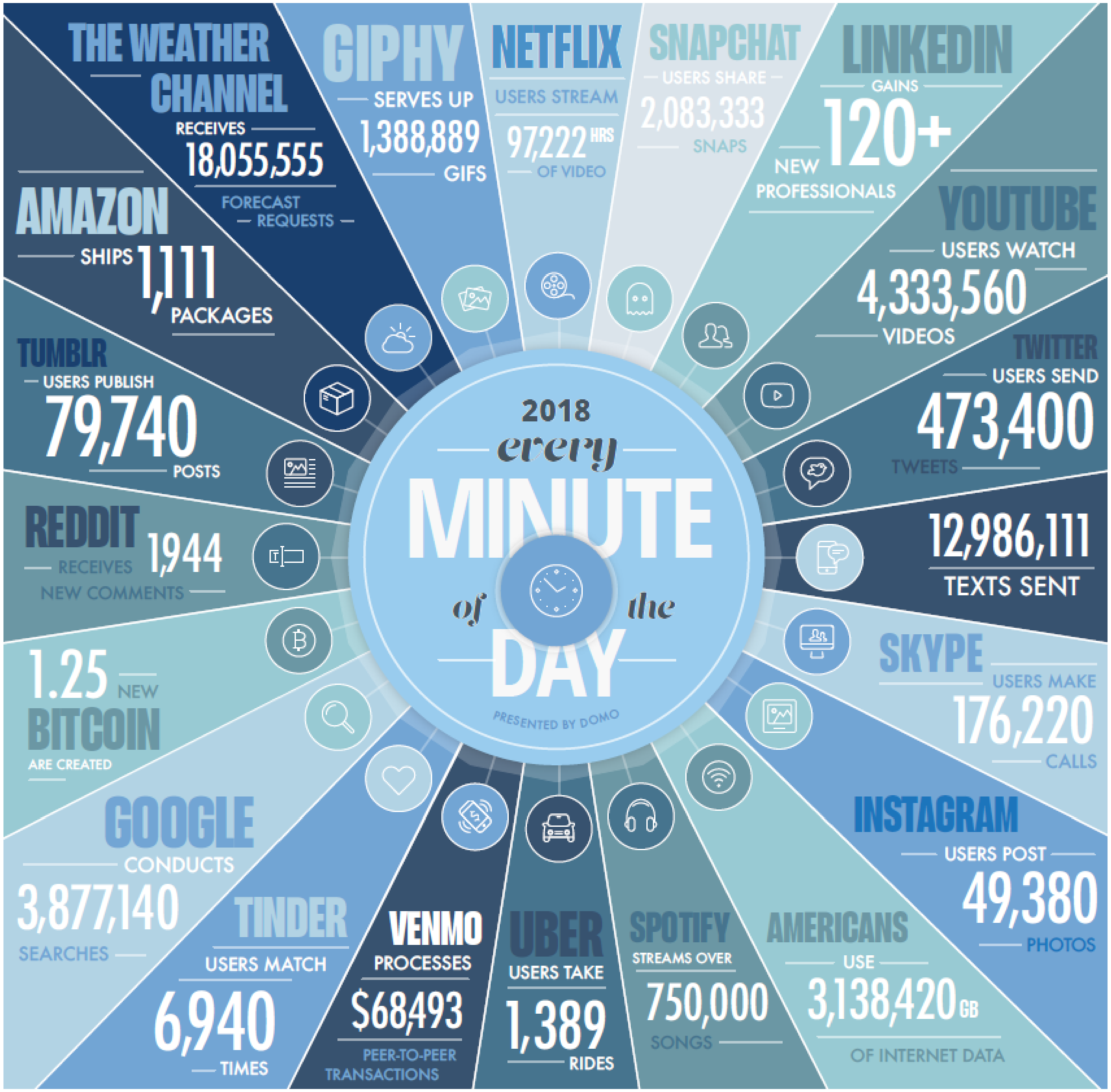
- On assiste à un flux considérable de données issues de saisies automatisées.

  $\rightsquigarrow$ Pour l'illustrer, *WallMart*, une grande chaîne de distribution américaine, enregistre chaque semaine plus de 250 millions de transactions à partir de ses $11'000$ points de vente.

- Cependant, une étude de la *International Data Corporation* (IDC) en 2014 montrait que 5% des données générées en 2013 ont été réellement analysés.

- Le volume des (mega)données stockées ('*big data*') aujourd'hui est en pleine expansion et explosion!

$\rightsquigarrow$ Selon l'étude de la IDC en 2014, les données numériques créées dans le monde augmentera par un facteur de 10, passant de $4'400$ milliards de gigabytes en 2013 à $44'000$ milliards d'ici à 2020 (*i.e.* plus de $5'200$ gigabytes pour chaque homme, femme ou enfant en 2020), donc en doublant *grosso modo* tous les deux ans.



Source: 'Data Never Sleeps 6.0' (`www.domo.com/learn/data-never-sleeps-6/`), juin 2018.

• Face à cette inflation de données, on ressent donc un besoin croissant en analyse et en étude de données.

$\rightsquigarrow$ Ceci devient même une nécessité pour le manager actuel.

$\rightsquigarrow$ Le manager actuel doit disposer d'outils et de méthodes performants d'aide à la décision.

$\rightsquigarrow$ La statistique permet à un manager de décrypter, de critiquer et de comprendre les données qui lui sont soumis.

'I keep saying the sexy job in the next ten years will be statisticians.'

Hal Varian, 2009

Source: Interview dans le 'The McKinsey Quarterly' de janvier 2009 avec l'économiste en chef de Google.

‘And with ongoing advances in high-performance computing and the explosion of data, ... I would venture to say that statistician could be the sexy job of the century.’

James (Jim) Goodnight, 2010

Source: 'Point fort' du journal de l'UNIGE du 5 au 26 septembre 2013 (numéro 78); voir `www.unige.ch/communication/lejournal/journal78.html`.

Tribune de Genève | Samedi-dimanche 6-7 décembre 2014

Emploi

Entre le pire et le meilleur

Pour son classement, CareerCast (spécialiste du recrutement) a pris en compte cadre de travail, salaire, stress et perspectives d'emploi.

Top 3 des pires et meilleurs jobs en 2014

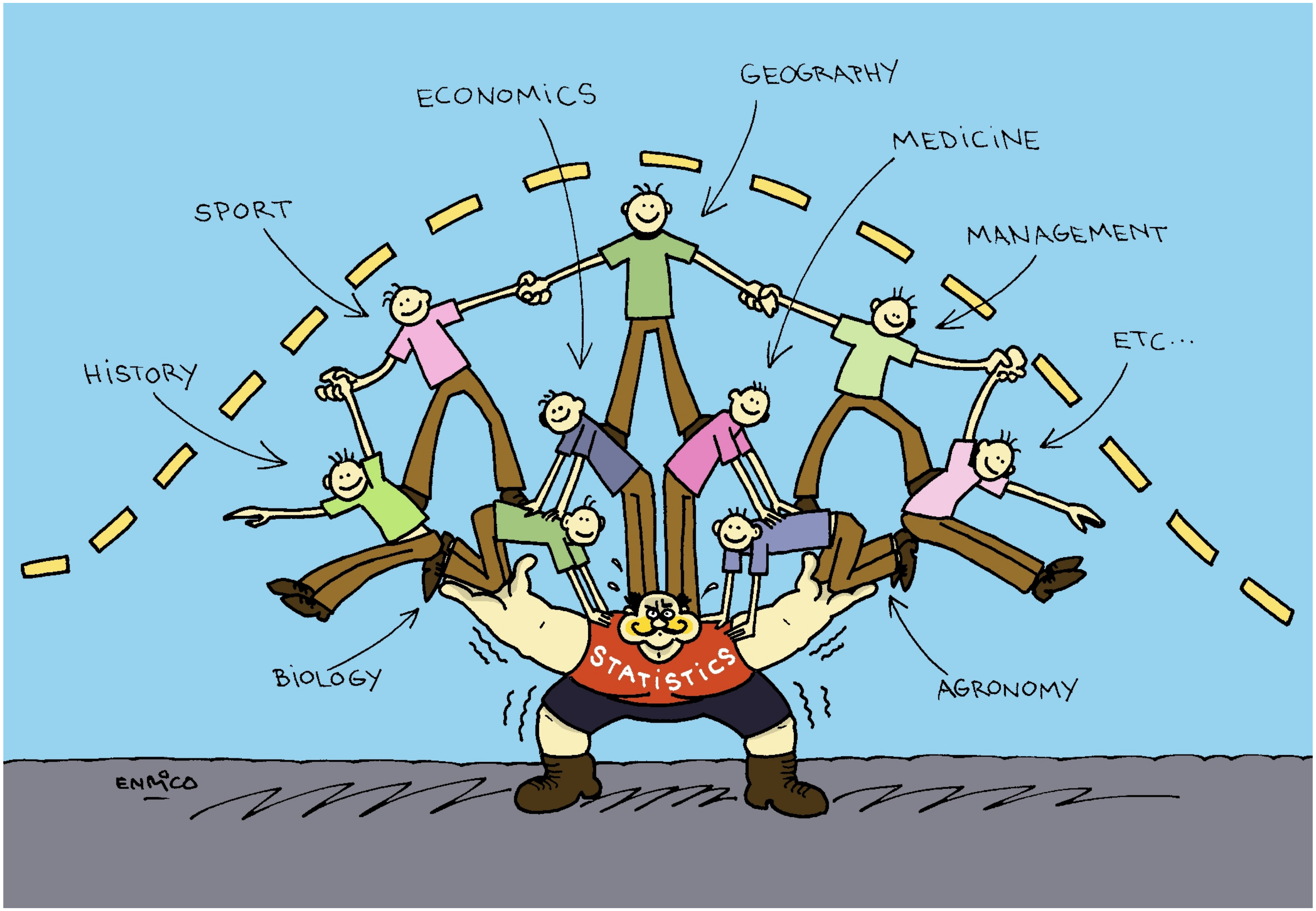
Les pires	Les meilleurs
1. Bûcheron	1. Mathématicien
2. Journaliste de presse écrite	2. Professeur d'université
3. Soldat	3. Statisticien

SOURCE: CAREERCAST


THE BEST AND THE WORST JOBS
WORST JOBS OF 2016
Newspaper reporter
Logger
Broadcaster
Discjockey
Enlisted military service personnel
Pest control worker
Retail salesperson
Advertising salesperson
Taxi driver
Firefighter
BEST JOBS OF 2016
Data scientist
Statistician
Information security analyst
Audiologist
Diagnostic medical sonographer
Mathematician
Software engineer
Computer systems analyst
Speech pathologist
Actuary
Source: www.careercast.com
TOI FOR MORE INFOGRAPHICS DOWNLOAD TIMES OF INDIA APP

Exemples d'applications de la statistique

- Analyser le comportement du consommateur dans un système de vente par correspondance;
- Prévoir les ventes dans une grande chaîne de distribution et anticiper au mieux les tendances du marché;
- Constituer une segmentation (typologie) des clients d'une banque pour cibler des opérations de marketing ou des attributions de crédit;
- Rechercher, spécifier puis cibler les niches de marché les plus profitables (banque) ou au contraire les plus risquées (assurance);
- Découvrir des structures et des relations se trouvant dans de grandes bases de données complexes.


ECONOMICS
GEOGRAPHY
MEDICINE
SPORT
MANAGEMENT
HISTORY
ETC...
STATISTICS
BIOLOGY
AGRONOMY
ENRICO

‘Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.’

Samuel S. Wilks, 1951

Traduction: ‘La pensée statistique sera un jour aussi nécessaire à l’humanité que lire et écrire.’

‘*Statistical thinking*’ — La pensée statistique

- La pensée statistique est une façon de penser terre-à-terre pour prendre des décisions face à l’incertitude.

- Le but de la pensée statistique est de transformer les données en information exploitable, tout en considérant le système sur lequel les données ont été collectées et leur variabilité.

- La *American Society for Quality, Statistics Division* a définit en 1996 la pensée statistique comme une philosophie d'apprentissage et d'action basée sur les principes fondamentaux suivants:
  - tout travail se produit dans un système de processus interconnectés;
  - la variabilité (donnant lieu à l'incertitude) existe dans tous les processus; et
  - comprendre et réduire la variabilité (indésirable) sont des clés de succès.

- La variabilité est partout. Juste à penser au temps qu'il faut pour se rendre de la maison à l'université. Tous les jours, le temps de transport est différent.

$\leadsto$ Pour un manager actuel, un leadership efficace exige la pensée statistique!

‘In this world there is nothing certain but death and taxes.’

Benjamin Franklin

Traduction: ‘Dans ce monde, rien n’est certain sauf la mort et les impôts.’

Concepts de base de la statistique

- Les concepts de base de la statistique sont

la variabilité

et

l'incertitude.

▷ $\boxed{\text{La variabilité}}$: Des variations dues à de petits, parfois grands changements, d'ailleurs souvent imprévisibles, se manifestent dans toutes les situations.

Exemples:

a) Les êtres humains sont différents les uns des autres et physiologiquement leur organisme se modifie de jour en jour, voire même de minute en minute. En effet, la pression du sang, la taille et le poids d'un individu varient constamment;

b) Dans son travail, le chercheur est confronté à des erreurs expérimentales et des erreurs de mesure dont il doit en tenir compte lors de l'analyse des données mesurées.

▷ $\boxed{\text{L'incertitude}}$: Une grande partie de notre vie est faite d'incertitude que l'on souhaite quantifier. Ne se pose-t-on pas souvent la question: 'Que va-t-il se passer?' ou encore 'Que va-t-il m'advenir demain, dans un mois, dans une année, dans dix ans?'.

$\rightsquigarrow$ Par la statistique,

on cherche à tirer des conclusions sur des données soumises à la variabilité tout en contrôlant le niveau d'incertitude.

Cette démarche se résume à exploiter les données, à y découvrir des informations quantifiables et pertinentes, des tendances.

La statistique devient un outil précieux d'aide à la décision!

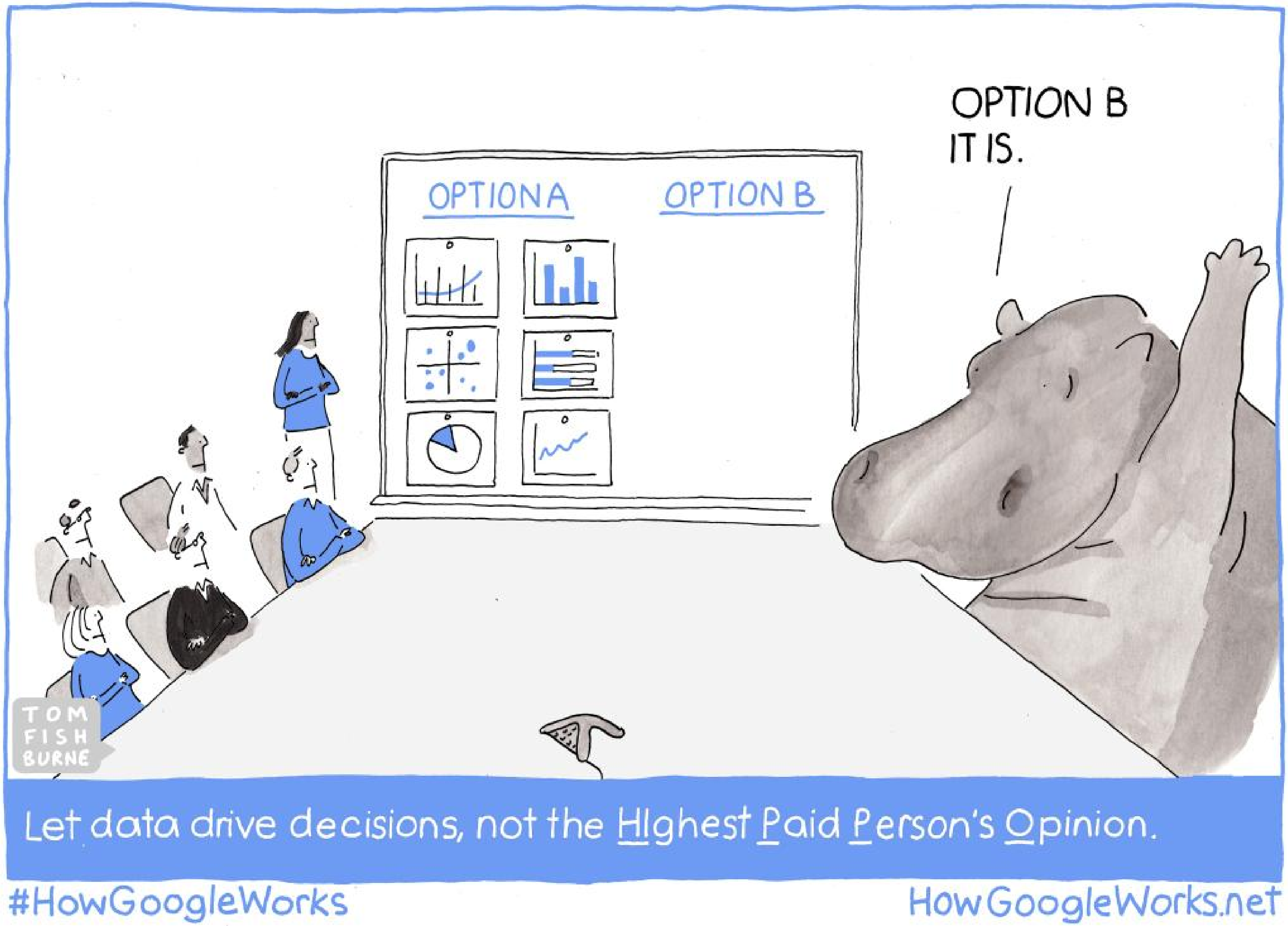
Qu'est-ce que la statistique?

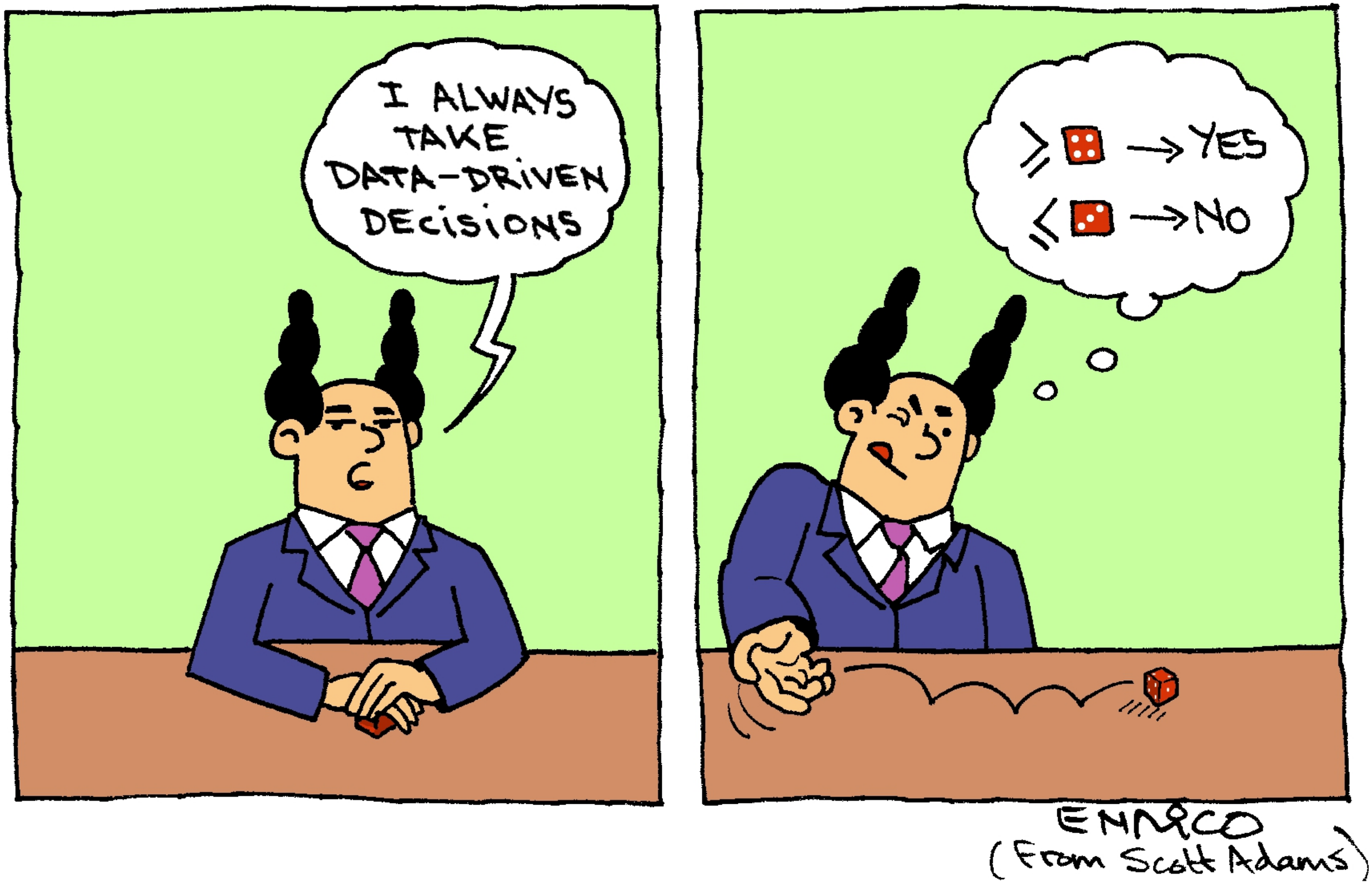
◇ La statistique est la science de l'apprentissage des données (ou de donner un sens à des données) et de mesurer, contrôler et communiquer l'incertitude.

⇝ L'incertitude est mesurée par le calcul de probabilités.

⇝ La probabilité est la monnaie ('*currency*') ou la grammaire ('*grammar*') de la statistique!

⇝ La statistique est concerné par l'étude de la prise de décisions fondée sur les données face à l'incertitude (⇝ '*data-driven decision making in the face of uncertainty*').


OPTION B
IT IS.
OPTION A
OPTION B
TOM FISH BURNE
Let data drive decisions, not the HIghest Paid Person's Opinion.
#HowGoogleWorks
HowGoogleWorks.net


I ALWAYS TAKE DATA-DRIVEN DECISIONS
≥ → YES
≤ → NO
ENRICO
(From Scott Adams)

‘We are in the era of big data, and big data needs statisticians to make sense of it.’

Eric Schmidt et Jonathan Rosenberg, 2014

Source: Eric Schmidt, président du conseil d’administration et ancien PDG de Google, et Jonathan Rosenberg, ancien vice-président senior produits de Google, dans leur livre ‘How Google Works’ (2014, New York, NY: Grand Central Publishing — `HowGoogleWorks.net`*).*

‘Les questions les plus importantes de la vie ne sont en effet, pour la plupart, que des problèmes de probabilité.’

Pierre-Simon Laplace, 1820

2. Calcul élémentaire de probabilités

• La théorie des probabilités permet de décrire et modéliser des phénomènes aléatoires.

$\rightsquigarrow$ Les actions qui peuvent amener des résultats aléatoires sont dites des **expériences aléatoires**.

$\rightsquigarrow$ Plus précisément, une expérience est dite **aléatoire** s'il est impossible de prévoir son résultat.

$\rightsquigarrow$ En principe, on admet qu'une expérience aléatoire peut être répétée (indéfiniment) dans des conditions identiques; son résultat peut donc varier d'une réalisation à l'autre.

Exemples:

- Résultat d'un jet d'une pièce de monnaie ou d'un dé
- Résultat du tirage d'une carte
- Résultat du tirage du loto
- Occurence d'un accident
- Résultats sportifs
- Nombre de fusibles défectueux dans un lot de vingt unités
- Nombre de personnes entrant dans un établissement donné (banque, poste)
- Temps d'attente entre l'arrivée de deux messages électroniques
- Trafic aérien (nombre d'arrivées et de départs) à l'aéroport de Cointrin durant une période de pointe

2.1 Formalisation

Une probabilité est une mesure de vraisemblance.

Exemple: La probabilité d'obtenir pile lorsque qu'une pièce de monnaie est lancée.

Objectif: Formaliser la manière avec laquelle les probabilités sont calculées.

Définitions:

Expérience: Action dont les résultats possibles sont connus, sans savoir lequel se réalise.

$\rightsquigarrow$ Expérience (ou épreuve) aléatoire (ou stochastique).

Réalisations: Résultats possibles d'expériences $e_1, e_2, \ldots, e_N$.

$\rightsquigarrow$ N réalisations possibles (de même vraisemblance).

Ensemble fondamental (ou univers): $S = \{e_1, e_2, \ldots, e_N\}$

Exemple: $S = \{e_1, e_2, e_3, e_4, e_5, e_6\}$

Evénement: Ensemble de réalisations

$A, B, C, \ldots$

Exemple: $A = \{e_2, e_4, e_6\}$

$\rightsquigarrow$ En d'autres termes, un événement est un sous-ensemble de S.

Probabilité: Mesure de vraisemblance d'une réalisation, d'un événement, ...

Exemples: $P(e_3), P(A), P(B), \ldots$

<u>But:</u> Calculer des probabilités associées à des événements.

‘La probabilité d’un événement est le rapport du nombre des cas qui lui sont favorables au nombre de tous les cas possibles, lorsque rien ne porte à croire que l’un de ces cas doit arriver plutôt que les autres, ce qui les rend, pour nous, également possibles.’

Pierre-Simon Laplace, 1814

Règles de base:

1. S'il y a N_A réalisations de même vraisemblance (= même probabilité) dans l'événement A

$$\Rightarrow P(A) = \frac{N_A}{N} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}$$

2. $0 \leq P(A) \leq 1$

$\rightsquigarrow$ Une probabilité est toujours comprise entre 0 et 1!

3. $\displaystyle\sum_{i=1}^{N} P(e_i) = 1$

$\rightsquigarrow$ La somme des probabilités de toutes les réalisations possibles vaut 1.

4. $\overline{A} = S \setminus A = A^c$ (événement **complémentaire**)

$$\Rightarrow P(\overline{A}) = 1 - P(A)$$

5. $A = \emptyset$ (événement **impossible**)

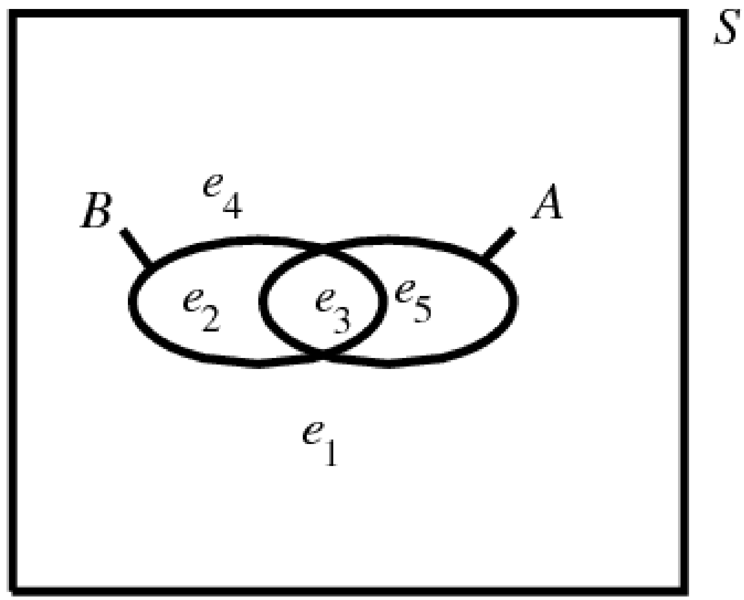
$$\Rightarrow P(A) = 0$$

6. $A = S$ (événement **certain**)

$$\Rightarrow P(A) = 1$$

Outil graphique: Diagramme de Venn

Ce diagramme permet de représenter des ensembles.



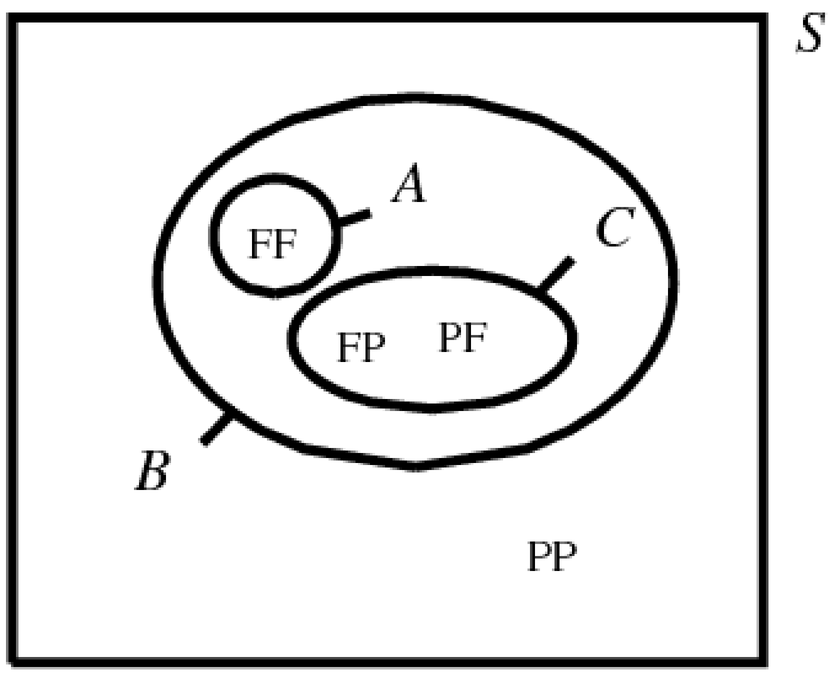
	A	$\overline{A}$
B	e_3	e_2
$\overline{B}$	e_5	e_1, e_4

EXEMPLES

1. Deux pièces sont lancées. On s'intéresse aux différentes configurations de pile (P) et face (F).

$$
\begin{aligned}
S &= \{\text{PP, PF, FP, FF}\} \\
N &= 4 \\
A &= \text{`Deux faces'} = \{\text{FF}\} \\
P(A) &= 1/4 = 0.25 \\
B &= \text{`Au moins une face'} = \{\text{PF, FP, FF}\} \\
P(B) &= 3/4 = 0.75
\end{aligned}
$$

$$
\begin{aligned}
C &= \text{`Différents'} = \{\text{PF, FP}\} \\
P(C) &= 2/4 = 0.5
\end{aligned}
$$



2. Deux dés sont jetés. Nous sommes intéressés aux différentes paires possibles.

$$
\begin{aligned}
S &= \{(1,1), (1,2), (1,3), \ldots, (2,1), (2,2), (2,3), \\
&\qquad \ldots, (6,1), (6,2), (6,3), \ldots, (6,6)\} \\
N &= 36 \\
A &= \text{‘Obtenir au moins un 6’} \\
&= \{(1,6), (2,6), (3,6), (4,6), (5,6), \\
&\qquad (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\} \\
P(A) &= 11/36
\end{aligned}
$$

$$
\begin{aligned}
B &= \text{‘La différence est de deux’} \\
&= \{(1,3),\ (2,4),\ (3,1),\ (3,5),\ (4,2), \\
&\qquad (4,6),\ (5,3),\ (6,4)\} \\
\\
P(B) &= 8/36 = 2/9
\end{aligned}
$$

2.2 Le diagramme d'arbre

Représentation **graphique** de l'**ensemble fondamental** lorsque les réalisations sont elles-même composées d'une **succession de réalisations, indépendantes ou non.**

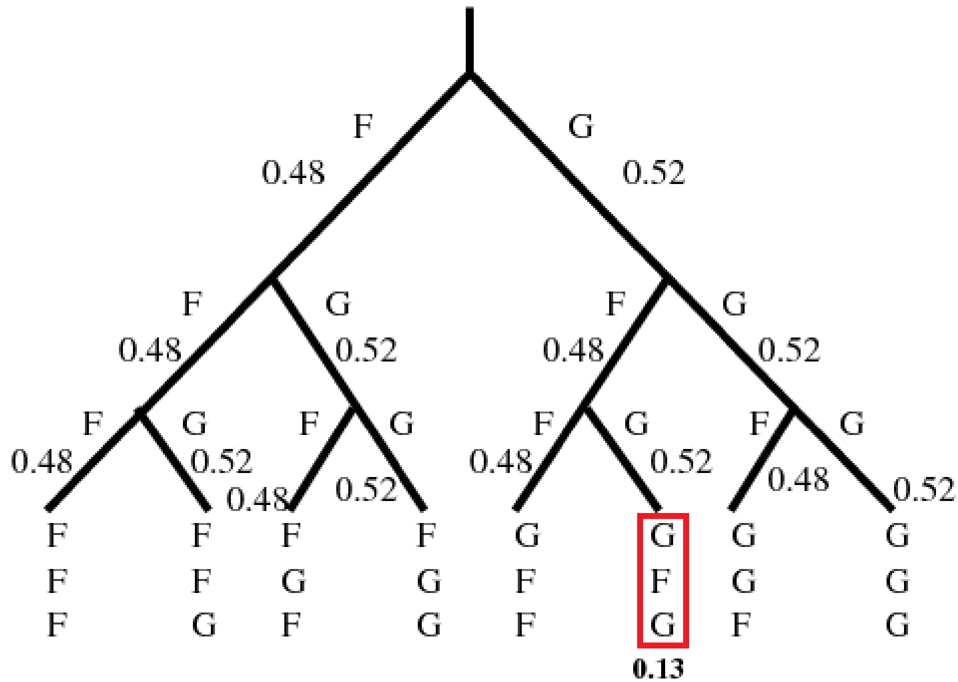
Exemples: Le sexe de trois enfants successifs, les combinaisons de nombres sur les faces de trois dés.

Règle de base:

La probabilité associée à une réalisation, elle même composée d'une succession de réalisations indépendantes, est égale au produit des probabilités associées aux différentes réalisations indépendantes.

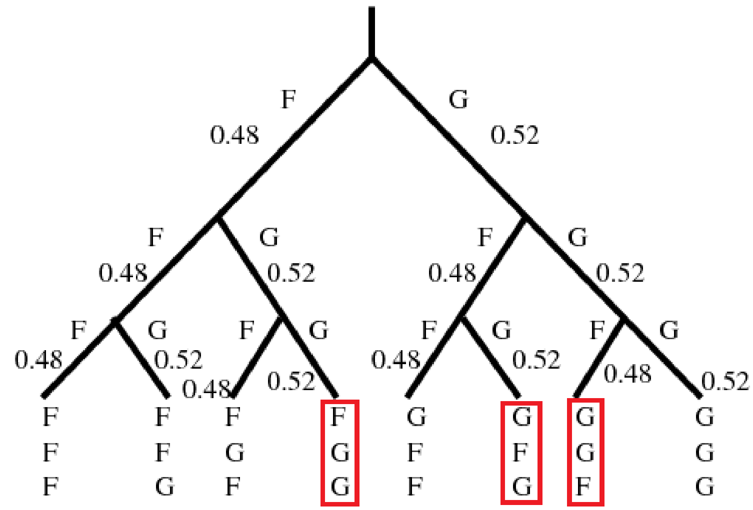
EXEMPLE

Le sexe de trois enfants. Soit G = 'Garçon' et F = 'Fille', avec $P(G) = 0.52$ et $P(F) = 0.48$.



Une réalisation GFG a pour probabilité

$$P(GFG) = P(G)P(F)P(G) = 0.52 \cdot 0.48 \cdot 0.52 \approx 0.13.$$



$$
\begin{aligned}
A &= \text{`Deux garçons exactement'} \\
P(A) &= P(FGG) + {\color{red}P(GFG)} + P(GGF) \\
&\approx (0.48 \cdot 0.52 \cdot 0.52) \\
&\quad + {\color{red}0.13} \\
&\quad + (0.52 \cdot 0.52 \cdot 0.48) \\
&\approx 0.39
\end{aligned}
$$

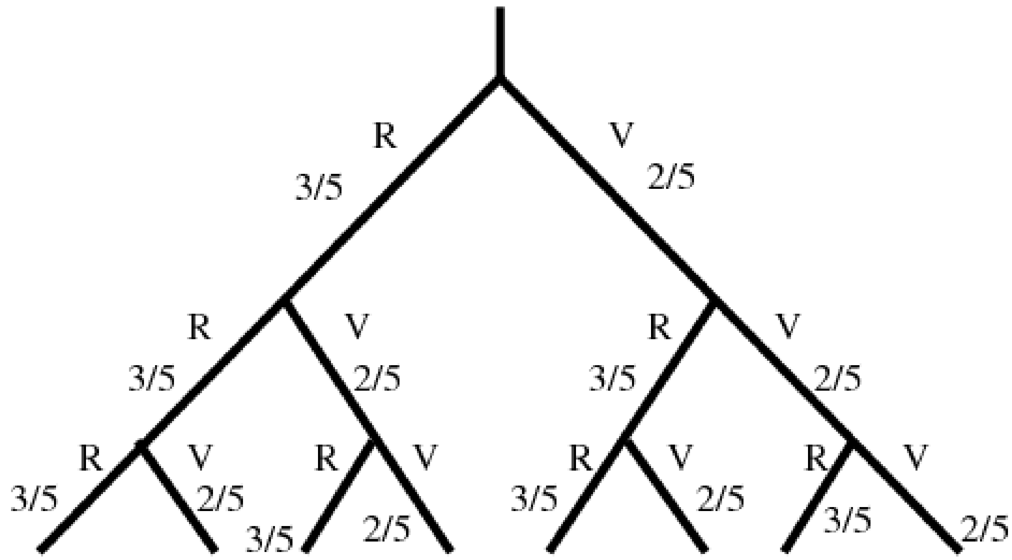
Remarque: Dans le cas de succession de réalisations **dépendantes**, le modèle de l'arbre peut encore être utilisé, mais les **probabilités changent** d'un embranchement à l'autre.

EXEMPLE

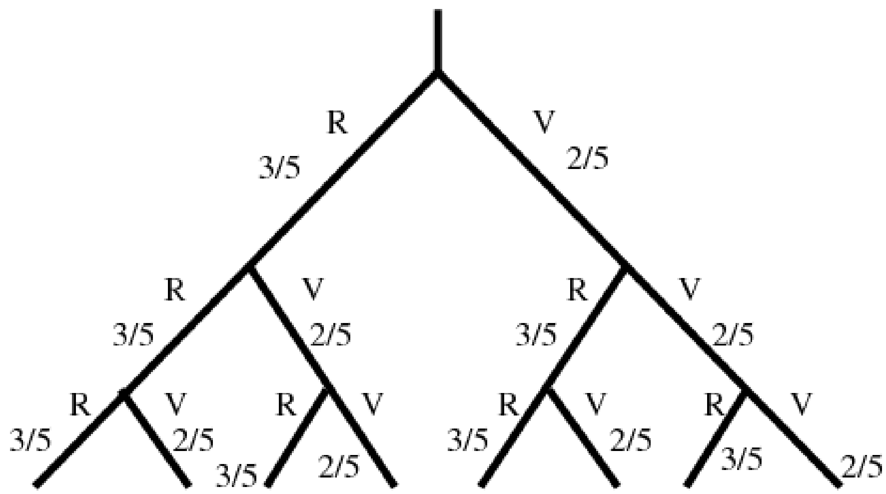
Le problème de l'urne... avec **3 boules rouges** et **2 boules vertes**.

On s'intéresse aux couleurs de trois boules tirées successivement de l'urne **avec remise** ou **sans remise**.

Avec remise (indépendance):

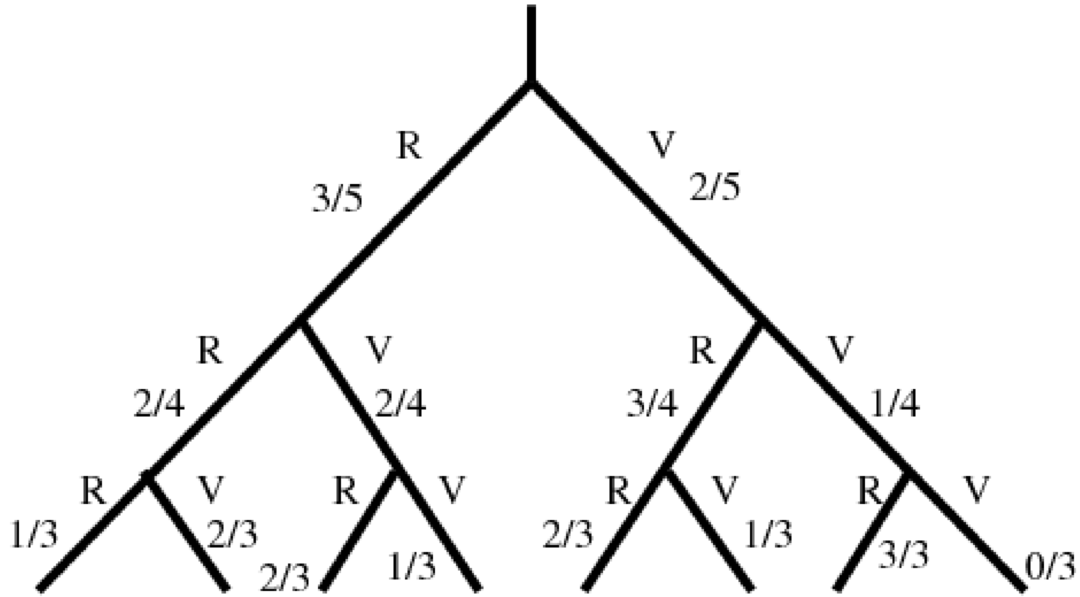


$$P(RVV) = 3/5 \cdot 2/5 \cdot 2/5 = 12/125 = 0.096$$

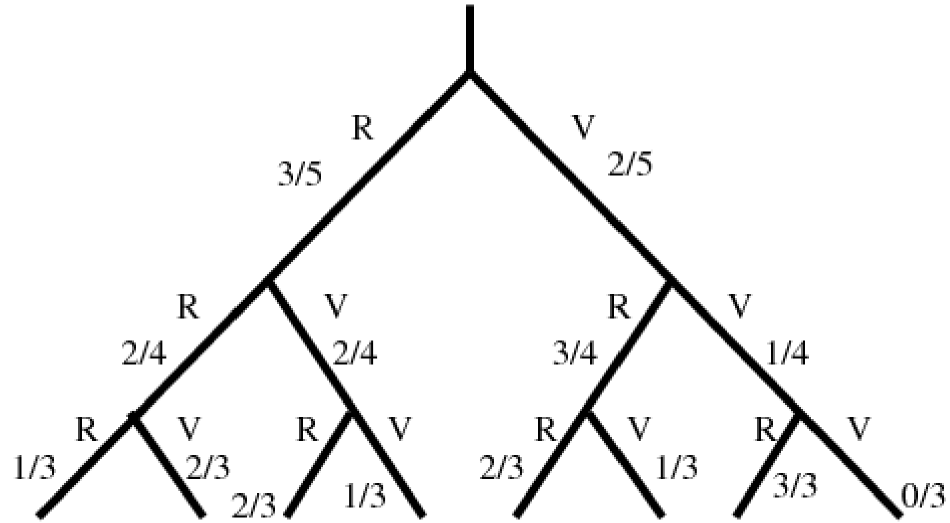


$$
\begin{aligned}
A &= \text{‘Deux boules vertes exactement’} \\
P(A) &= P(RVV) + P(VRV) + P(VVR) \\
&= 0.096 \\
&\quad + (2/5 \cdot 3/5 \cdot 2/5) \\
&\quad + (2/5 \cdot 2/5 \cdot 3/5) \\
&= 0.288
\end{aligned}
$$

Sans remise (dépendance):



$$P(RVV) = 3/5 \cdot 2/4 \cdot 1/3 = 1/10 = 0.1$$



$$
\begin{aligned}
A &= \text{‘Deux boules vertes exactement’} \\
P(A) &= P(RVV) + P(VRV) + P(VVR) \\
&= 0.1 \\
&\quad + (2/5 \cdot 3/4 \cdot 1/3) \\
&\quad + (2/5 \cdot 1/4 \cdot 3/3) \\
&= 0.3
\end{aligned}
$$


DR. RANDOM
PROFESSOR OF PROBABILITY
IN
OUT
PLEASE USE THIS COIN
Sumanta
sumanta.baruah@gmail.com

Application: Problème de Monty Hall

• Le **problème de Monty Hall** est un casse-tête probabiliste librement inspiré du jeu télévisé américain '*Let's Make a Deal*'.

⇝ Il est simple dans son énoncé mais non intuitif dans sa résolution et c'est pourquoi on parle parfois à son sujet de **paradoxe de Monty Hall**.

⇝ Il porte le nom de celui qui a présenté ce jeu aux États-Unis pendant treize ans, Monty Hall.

⇝ Voir la page

`fr.wikipedia.org/wiki/Problème_de_Monty_Hall`

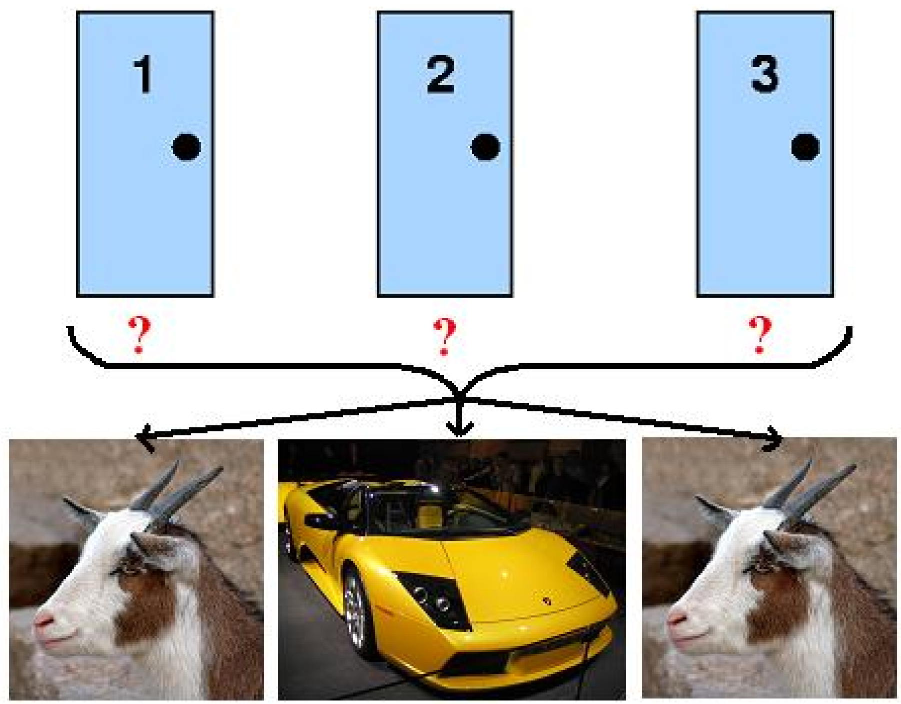
pour plus de détails et

`bit.ly/MontyHallProb`

pour des vidéos *YouTube*.

Énoncé

• Le jeu oppose un présentateur à un candidat (le joueur). Ce joueur est placé devant trois portes fermées. Derrière l'une d'elles se trouve une voiture (ou tout autre prix magnifique) et derrière chacune des deux autres se trouve une chèvre (ou tout autre prix sans importance).



$\rightsquigarrow$ Le candidat doit tout d'abord désigner une porte. Puis le présentateur ouvre une porte qui n'est ni celle choisie par le candidat, ni celle cachant la voiture (le **présentateur sait quelle est la bonne porte dès le début**!).

$\rightsquigarrow$ Le candidat a alors le droit ou bien d'ouvrir la porte qu'il a choisie initialement, ou bien d'ouvrir la troisième porte (et donc modifier son choix initial).

- Les questions qui se posent au candidat sont:
  - Augmente-t-il ses chances de gagner la voiture en changeant son choix initial?
  - Est-ce que la probabilité de gagner en changeant de porte est plus grande que la probabilité de gagner sans changer de porte?
  - Quelle est donc la meilleure stratégie: Faire un nouveau choix ou rester avec le choix initial? Les chances de gain vont-elles augmenter, diminuer ou bien resteront-elles les mêmes?

La solution: Les hypothèses importantes

• **Les trois portes ont**, tant qu'aucune n'est ouverte, **la même probabilité d'être la porte gagnante**.

$\rightsquigarrow$ Après avoir choisi la porte numéro 3, par exemple, le candidat a une chance sur trois $(1/3)$ de tomber directement sur la voiture avec deux chances sur trois $(2/3)$ que la voiture soit parmi les deux portes restantes:

$\rightsquigarrow$ Puisqu'il n'y a qu'une seule voiture, il y a 100% de chance qu'il y ait une chèvre derrière au moins une des portes 1 ou 2.

- Un choix doit ensuite être fait entre les conditions suivantes:
  - **Le présentateur ne peut ouvrir qu'une porte**, et celle-ci ne peut être ni la porte choisie par le joueur, ni la porte gagnante (il connaît l'emplacement de cette dernière, ce qui lui permet de répondre à cette condition **sans risque d'erreur**).
  - Quand le présentateur a le choix entre deux portes à ouvrir (toutes deux perdantes), **il choisit au hasard entre les deux, avec équiprobabilité**, *i.e.* $1/2$.

$\rightsquigarrow$ Le candidat choisi la porte 3, par exemple:

$\rightsquigarrow$ Le présentateur ouvre maintenant la porte 1.

$\rightsquigarrow$ Bien sûr le présentateur n'ouvre jamais une porte donnant sur la voiture, donc sans surprise la porte 1 donne sur une chèvre ce qui a pour effet de transférer la probabilité de $2/3$ d'avoir une voiture sur les portes 1 et 2 (comme expliqué précédemment) sur la porte 2 uniquement.

La solution

• De manière encore plus simple, on peut reformuler en disant que si après le choix initial du candidat il était envisageable que la voiture se trouve derrière les portes 1 et 2 (avec une probabilité de $2/3$), ce n'est plus le cas après l'ouverture de la porte 1 par le présentateur: Seule la porte 2 est encore susceptible de cacher la voiture (et par conséquent, toujours avec une probabilité de $2/3$).

$\rightsquigarrow$ Ainsi, **le candidat double ses chances de gagner la voiture en modifiant systématiquement son choix initial!**

$\rightsquigarrow$ Le diagramme suivant montre le même raisonnement d'une manière plus complète et plus formalisée:

Le candidat doit choisir d'ouvrir une des trois portes :

Probabilités après le choix initial du candidat :

Le candidat a choisi la porte derrière laquelle est cachée :

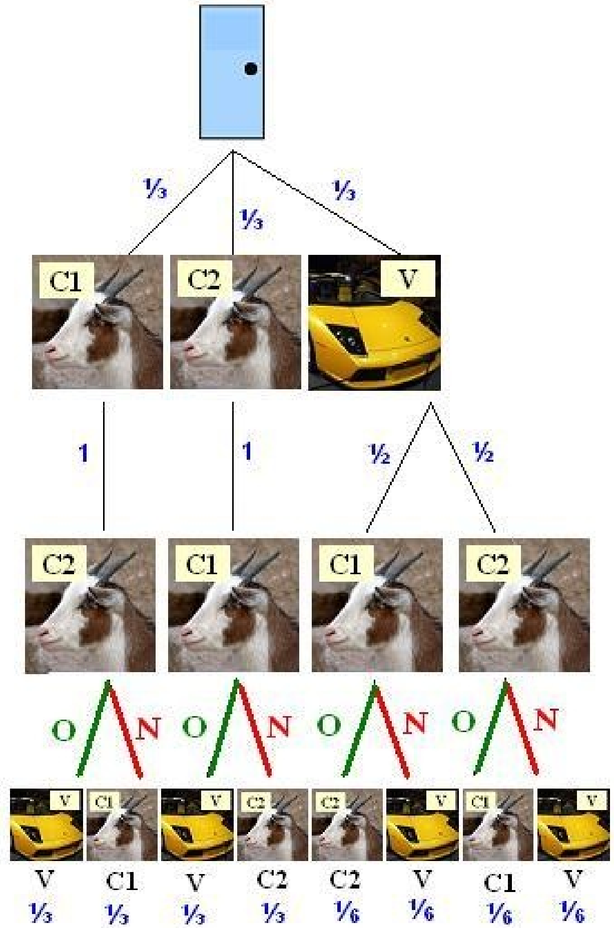
Probabilités intermédiaires résultant du choix du présentateur :

Le présentateur ouvre alors la porte donnant sur :

Le candidat change t-il par rapport à son choix initial ?

Au final, le candidat obtient :

Probabilités après le choix définitif du candidat :



Résultats :

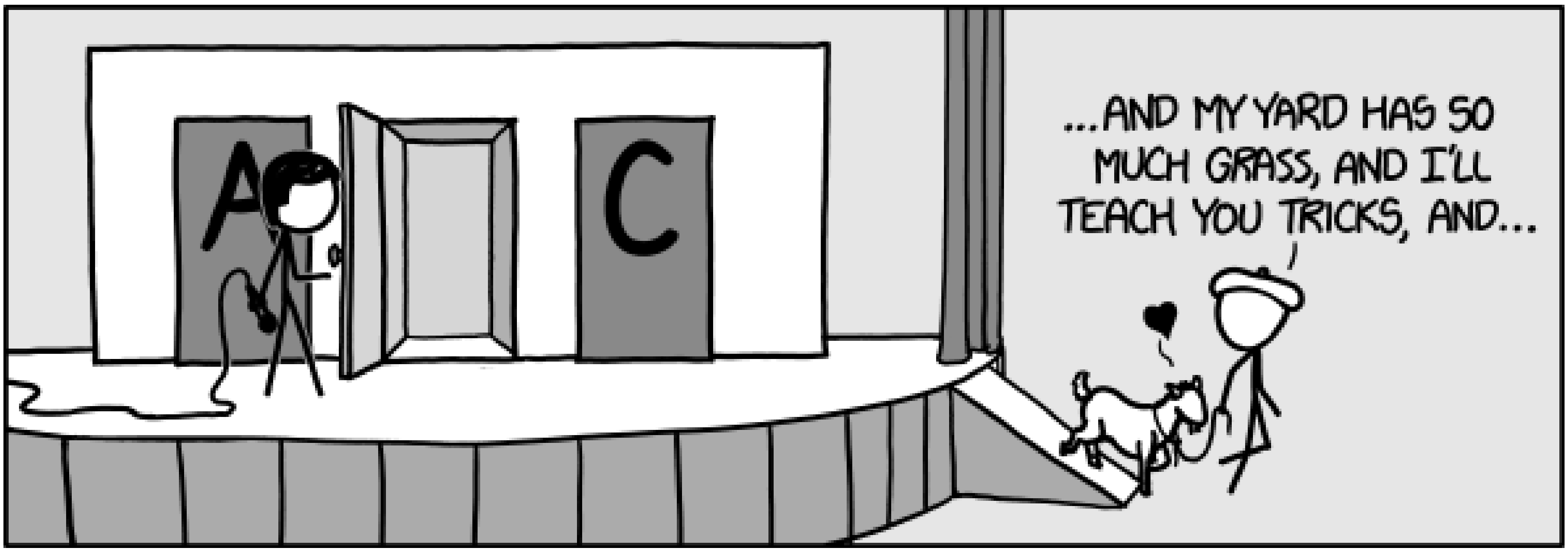
Avec changement	Sans changement
- Voiture : 1/3+1/3= 2/3	- Voiture : 1/6+1/6= 1/3
- Chèvre 1 : 1/6	- Chèvre 1 : 1/3
- Chèvre 2 : 1/6	- Chèvre 2 : 1/3

• De nombreuses **variantes** ont été proposées, modifiant les paramètres (par exemple, un nombre élevé de portes, les portes n'ont pas à l'origine une probabilité égale de cacher la voiture, changements des règles de l'ouverture de la porte par le présentateur).

$\rightsquigarrow$ Voir la page

`fr.wikipedia.org/wiki/Problème_de_Monty_Hall`

pour plus de détails.


A
C
...AND MY YARD HAS SO MUCH GRASS, AND I'LL TEACH YOU TRICKS, AND...

2.3 Rappels d'analyse combinatoire

Permutations: De combien de manières peut-on arranger n objets?

En première position:	n possibilités
En deuxième position:	$(n-1)$ possibilités
En troisième position:	$(n-2)$ possibilités
...	
En avant-dernière position:	2 possibilités
En dernière position:	1 possibilité

Au total: $n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1 = n!$ possibilités.

Nombre de permutations de n objets:

$$\boxed{n!}$$

Exemple: De combien de manières Aline, Brigitte et Carmen peuvent-elles s'asseoir sur 3 sièges, de gauche à droite?

Réalisations possibles: (A, B, C), (A, C, B), (B, A, C), (B, C, A), (C, A, B), (C, B, A)

Total: $N = 6 = 3 \cdot 2 \cdot 1 = 3!$

Combinaisons: De combien de manières peut-on choisir k objets parmi n?

▷ De combien de manières peut-on arranger k objets parmi n?

En première position: n possibilités

En deuxième position: $(n-1)$ possibilités

...

En k-ième position: $(n-k+1)$ possibilités

Au total: $n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = \dfrac{n!}{(n-k)!}$ possibilités.

▷ On a $k!$ manières de permuter les k objets qui ont été sélectionnés.

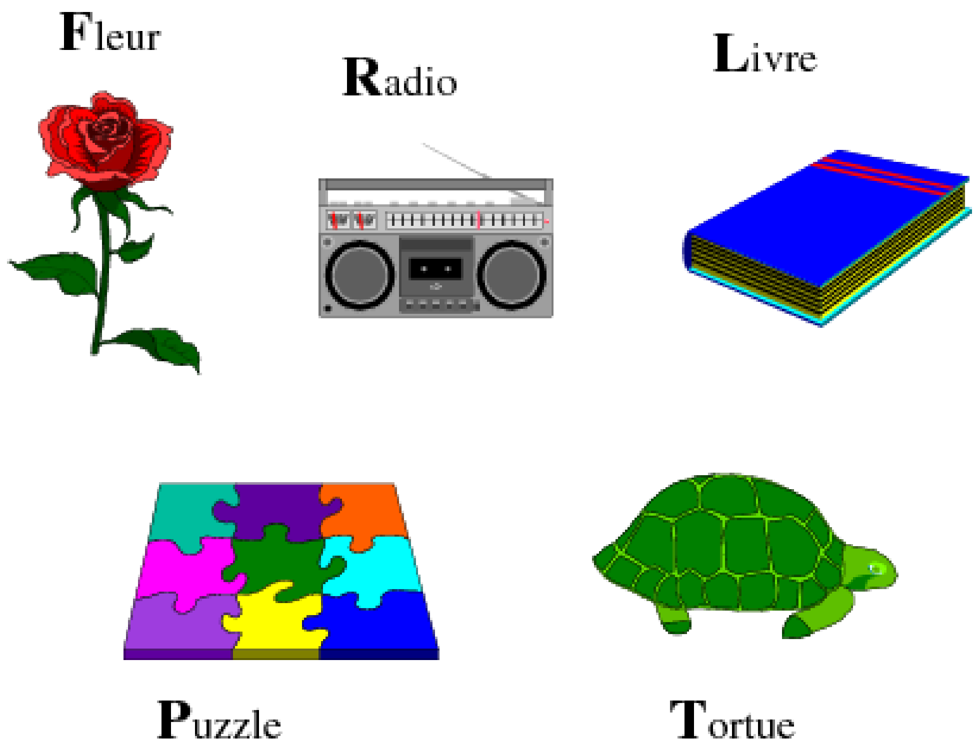
▷ Le nombre de tirages possibles sans tenir compte de l'ordre des objets est alors:

$$\frac{n!/(n-k)!}{k!} = \frac{n!}{k!(n-k)!}$$

Nombre de combinaisons de k objets parmi n:

$$\boxed{\binom{n}{k} = \frac{n!}{k!(n-k)!} = C_n^k}$$

Exemple: De combien de manières peut-on choisir 3 cadeaux parmi les 5 à disposition?



Réalisations possibles:

(F, R, L)	(F, R, P)	(F, R, T)	(F, L, P)
(F, L, T)	(F, P, T)	(R, L, P)	(R, L, T)
(R, P, T)	(L, P, T)		

Total: $N = 10$

Ou:

$$\binom{5}{3} = \frac{5!}{3!2!} = \frac{5 \cdot 4 \cdot 3 \cdot 2}{3 \cdot 2 \cdot 2} = \frac{5 \cdot 4}{2} = 10 = N$$

Triangle de Pascal

Etant donné que $\binom{n}{0} = \binom{n}{n} = 1$, on peut déterminer toutes les combinaisons $\binom{n}{k}$ en complétant le triangle ligne par ligne:

	k=0	1	2	3	4	5	6	7	8	9	10	11	12
n=0	1												
1	1	1											
2	1	2	1										
3	1	3	3	1									
4	1	4	6	4	1								
5	1	5	10	10	5	1							
6	1	6	15	20	15	6	1						
7	1	7	21	35	35	21	7	1					
8	1	8	28	56	70	56	28	8	1				
9	1	9	36	84	126	126	84	36	9	1			
10	1	10	45	120	210	252	210	120	45	10	1		
11	1	11	55	165	330	462	462	330	165	55	11	1	
12	1	12	66	220	495	792	924	792	495	220	66	12	1

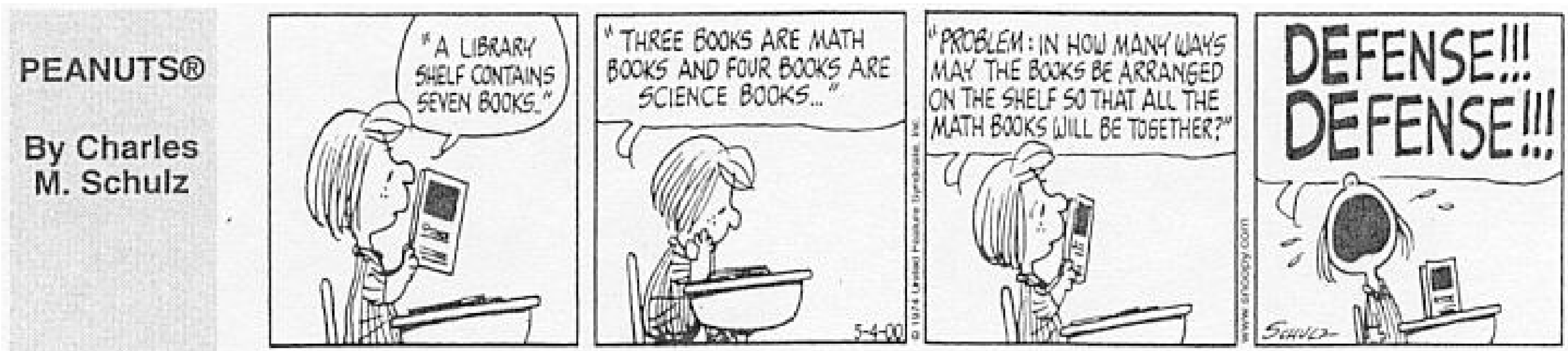
Exemple: On effectue un tirage (sans remise) au hasard de 3 cadeaux parmi les 5 de la page 68. Quelle est la probabilité d'obtenir une fleur?

de la page 68.

Soit l'événement $A = \{\text{une fleur}\}$.

On a

$$P(A) = \frac{\binom{1}{1} \cdot \binom{4}{2}}{\binom{5}{3}} = \frac{6}{10} = \frac{3}{5}.$$



L'analyse combinatoire est parfois bien compliquée!

Application: Paradoxe des anniversaires

• Le **paradoxe des anniversaires** est à l'origine une estimation probabiliste du nombre de personnes que l'on doit réunir pour avoir une chance sur deux (50%) que deux personnes de ce groupe aient leur anniversaire le même jour de l'année.

$\rightsquigarrow$ Il se trouve que ce nombre est environ égal à 23, ce qui choque un peu l'intuition. À partir d'un groupe de 57 personnes, la probabilité est supérieure à 99%.

• Cependant, il ne s'agit pas d'un paradoxe dans le sens de contradiction logique; c'est un paradoxe, dans le sens où c'est une vérité mathématique qui contredit l'intuition: La plupart des gens estiment que cette probabilité est très inférieure à 50%.

$\rightsquigarrow$ Voir la page

`fr.wikipedia.org/wiki/Paradoxe_des_anniversaires`

pour plus de détails.

Démonstration

• Par souci de simplicité, on suppose que toutes les années sont non-bissextiles. Prendre en compte le 29 février changerait peu les résultats, mais rendrait les calculs très délicats pour les applications où les personnes ne sont pas nées la même année.

• Le plus simple pour obtenir le résultat annoncé est de **calculer la probabilité que chaque personne ait un jour anniversaire différent de celui des autres**: L'inverse de ce que l'on cherche.

$\rightsquigarrow$ L'événement 'un jour anniversaire différent par personne' $(\overline{B})$ est le complémentaire de 'au moins deux identiques' (B).

$\rightsquigarrow$ On cherche

$$P(B) = 1 - P(\overline{B}).$$

$\rightsquigarrow$ Il y a n personnes, pour chacune il y a 365 jours possibles, donc au total si on ne se fixe aucune contrainte, il a 365^n possibilités.

$\rightsquigarrow$ Si maintenant on veut des jours différents, nous obtenons un arrangement de n parmi 365, soit

$$(365-0)(365-1)\cdots(365-n+1)=\frac{365!}{(365-n)!}.$$

$\rightsquigarrow$ On a donc

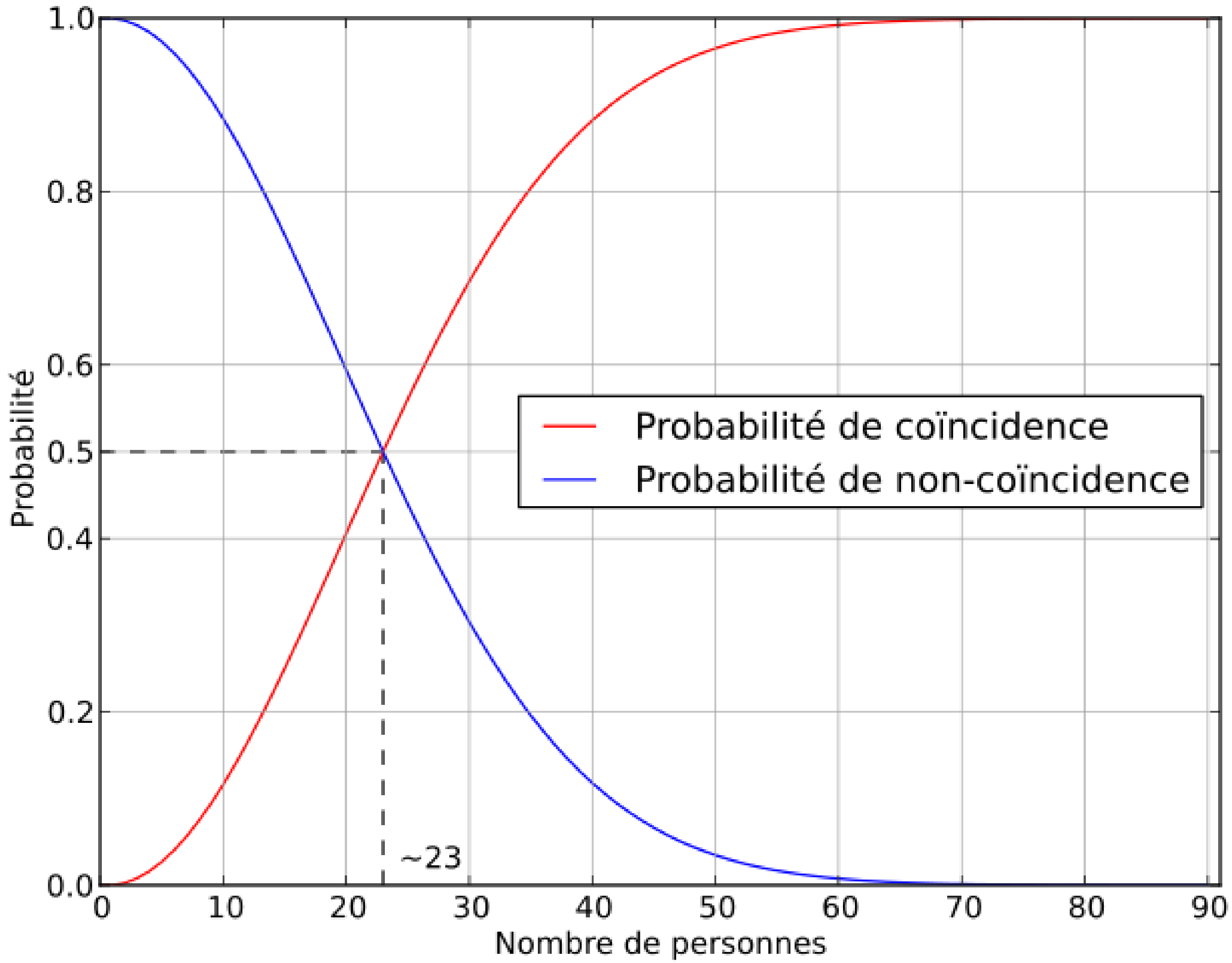
$$P(\overline{B})=\frac{365!}{(365-n)!}\cdot\frac{1}{365^n}.$$

$\rightsquigarrow$ Par conséquent, la probabilité recherchée est

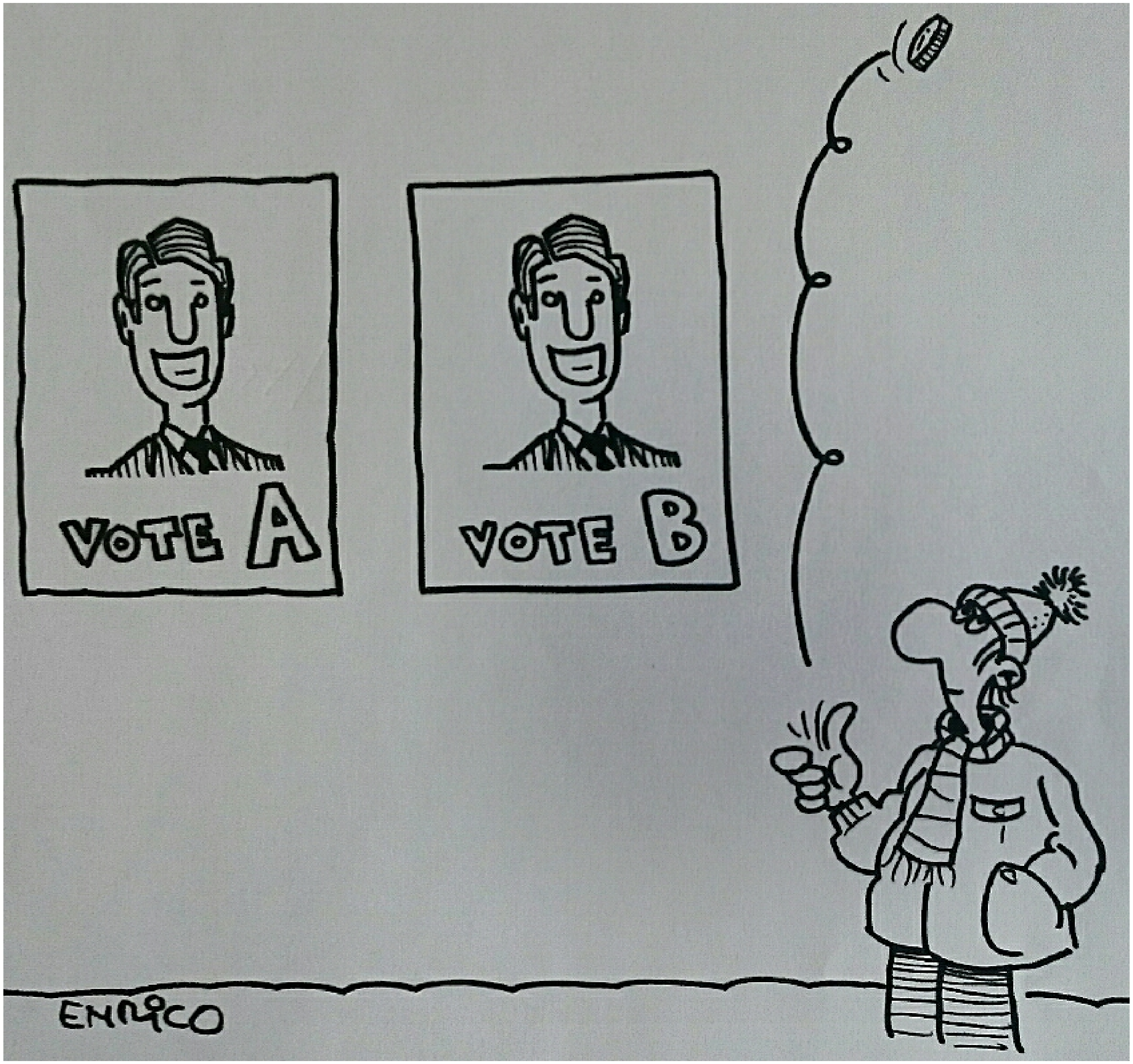
$$P(B)=1-P(\overline{B})=1-\frac{365!}{(365-n)!}\cdot\frac{1}{365^n}.$$

n, nombre de personnes	$P(B)$, probabilité de coïncidence
5	2.71%
10	11.69%
15	25.29%
20	41.14%
25	56.87%
30	70.63%
40	89.12%
50	97.04%
60	99.41%
80	99.99%
100	99.99997%
200	$99.9999999999999999999999998\%$
300	$(1 - 7.10^{-73}) \cdot 100\%$
350	$(1 - 3.10^{-131}) \cdot 100\%$
$>$ 365	$\approx 100\%$

$\rightsquigarrow$ On trouve 50.73% pour 23 personnes.



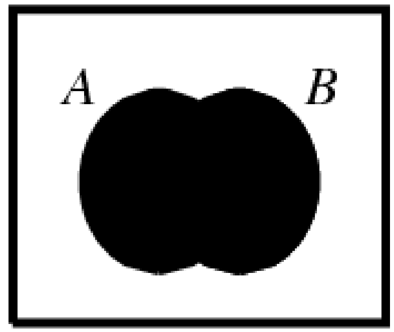
Source: `fr.wikipedia.org/wiki/Paradoxe_des_anniversaires` *(soumis à la licence CC-BY-SA 3.0).*


VOTE A
VOTE B
ENRICO

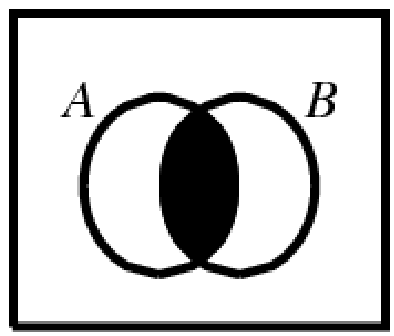
2.4 Evénements composés

On s'intéresse aux probabilités associées à des événements composés.

A ou B:
$A \cup B$



A et B:
$A \cap B$



Probabilité d'union:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Probabilité d'intersection:

$$P(A \cap B) = P(A) + P(B) - P(A \cup B)$$

EXEMPLE

2 dés sont jetés.

$$N = 36$$

$$A = \text{‘la somme est 8’}$$
$$B = \text{‘la valeur est la même’}$$

$$A = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$$
$$B = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$$

$$\begin{aligned}
A \cup B &= \{(1,1), (2,2), (2,6), (3,3), (3,5), (4,4) \\
&\qquad (5,3), (5,5), (6,2), (6,6)\} \\
A \cap B &= \{(4,4)\} \\
P(A) &= 5/36 \\
P(B) &= 6/36 = 1/6 \\
P(A \cup B) &= 10/36 = 5/18 \\
P(A \cap B) &= 1/36 \\
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\
&= 5/36 + 6/36 - 1/36 \\
&= 10/36 = 5/18
\end{aligned}$$

Evénements mutuellement exclusifs:

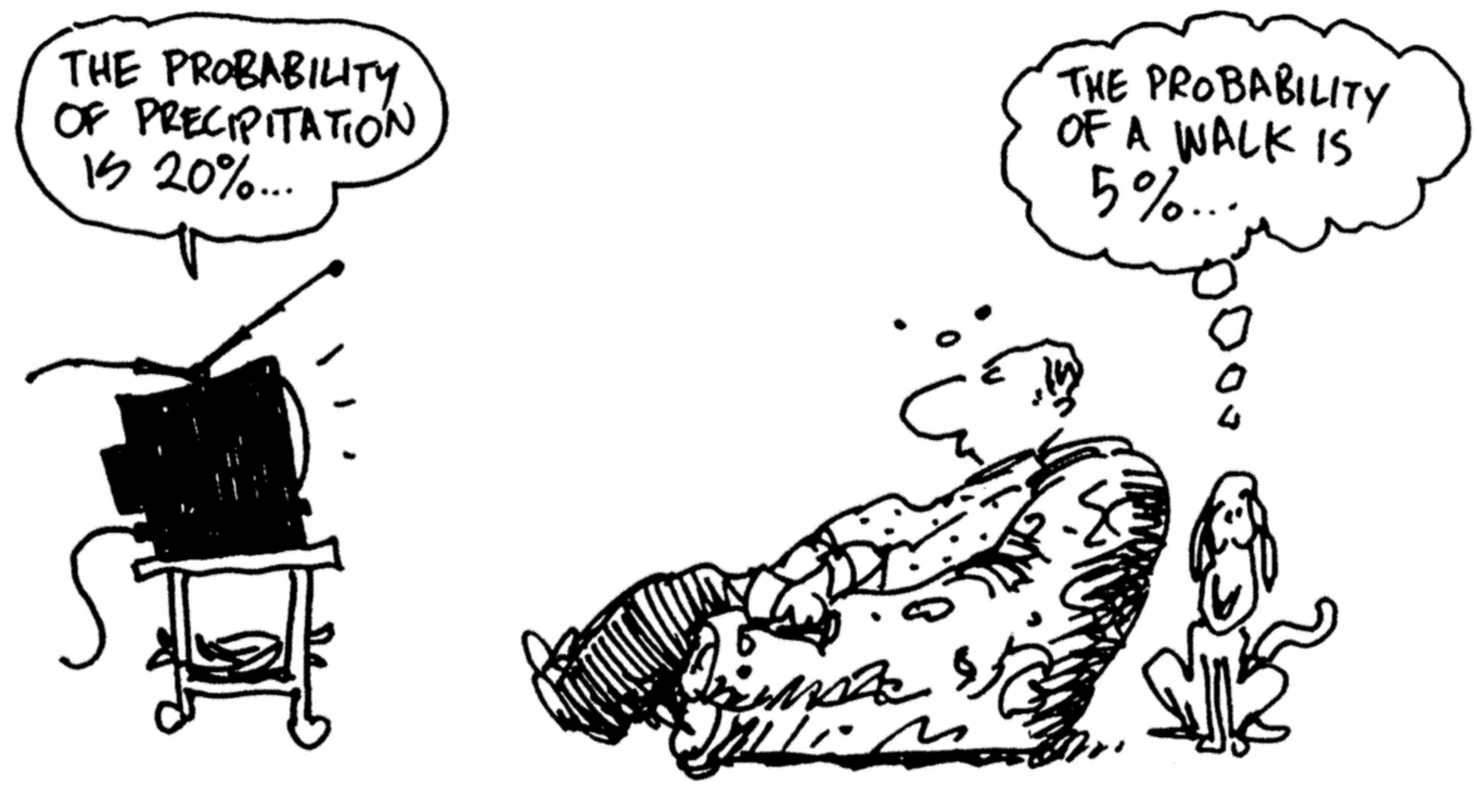
▷ Deux événements A et B sont incompatibles ou mutuellement exclusifs s'ils ne pourront pas se réaliser en même temps, *i.e.* si $A \cap B = \emptyset$. Autrement dit si l'événement A et B est impossible.

$$\boxed{\begin{array}{c} A \cap B = \emptyset \\ P(A \cap B) = 0 \\ P(A \cup B) = P(A) + P(B) \end{array}}$$

Distributivité:

$$\boxed{A \cap (B \cup C) = (A \cap B) \cup (A \cap C)}$$

$$\boxed{A \cup (B \cap C) = (A \cup B) \cap (A \cup C)}$$


THE PROBABILITY OF PRECIPITATION IS 20%...
THE PROBABILITY OF A WALK IS 5%...

2.5 Probabilité conditionnelle

On s'intéresse à la probabilité d'un événement A étant donné un autre événement B.

$$\Leftrightarrow$$

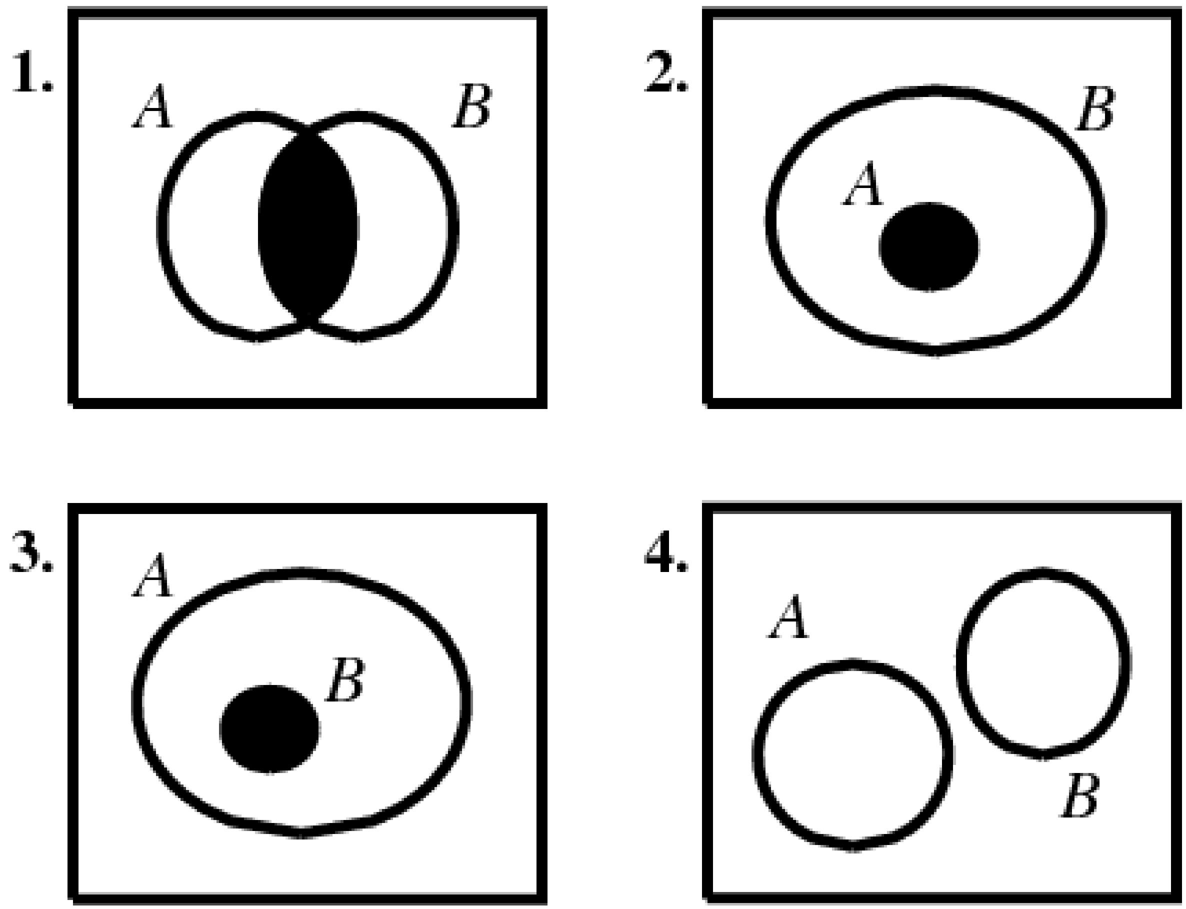
On réduit l'espace des réalisations aux réalisations contenues dans B.

Notation: A étant donné B (ou A sachant B) est noté $A \mid B$

Probabilités:

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

avec $P(B) \neq 0$ (voir figure 1 de la page 88).


1.
A
B
2.
B
A
3.
A
B
4.
A
B

Cas particuliers (voir figures 2 à 4 de la page 88):

2. $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A)}{P(B)}$, car $A \cap B = A$;

3. $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$, car $A \cap B = B$;

4. $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = 0$, car $A \cap B = \emptyset$.

Propriétés:

$$
\begin{aligned}
P(A \cap B) &= P(A \mid B) \cdot P(B) \\
P(A \cap B) &= P(B \mid A) \cdot P(A) \\
P(A \mid B) &= P(B \mid A) \cdot \frac{P(A)}{P(B)} \\
P(B \mid A) &= P(A \mid B) \cdot \frac{P(B)}{P(A)}
\end{aligned}
$$

EXEMPLE

On considère un patient qui peut suivre au choix deux traitements, notés A et B.

La probabilité qu'il suive le traitement A est égale à 75%. On sait par ailleurs que la probabilité de succès du traitement A est égale à 80%, tandis que celle du traitement B est égale à 90%.

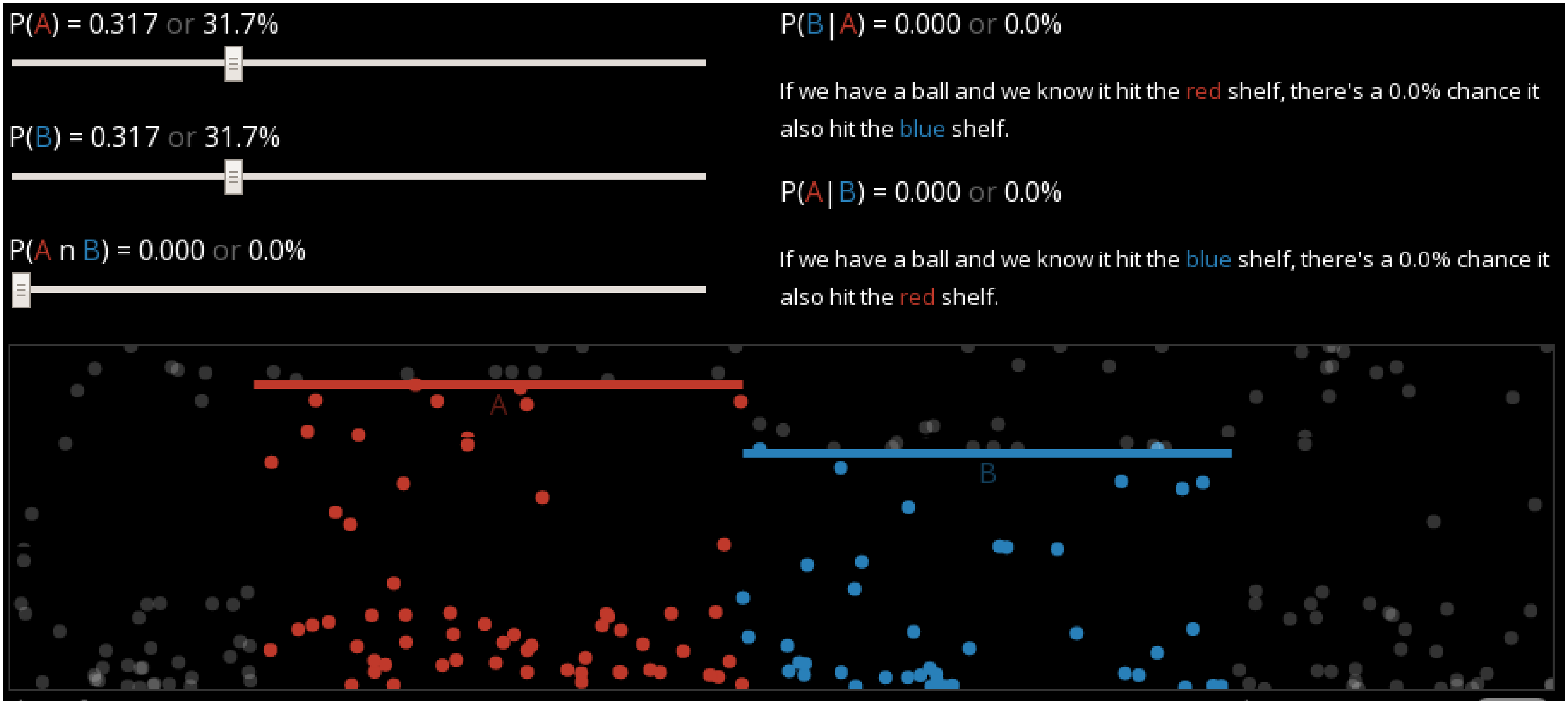
On admet que la probabilité que ce patient soit guéri, événement noté G, est égale à 82.5%.

Quelle est la probabilité que le patient ait suivi le traitement A sachant qu'il est guéri?

On a

$$P(A \mid G) = \frac{P(A \cap G)}{P(G)} = \frac{P(G \mid A) \cdot P(A)}{P(G)} = \frac{0.8 \cdot 0.75}{0.825} = 0.7273.$$

‘A visual explanation of conditional probability’



Source: `setosa.io/conditional/`.


I THINK IT WILL RAIN TODAY
HOW DO YOU KNOW?
BECAUSE YOU HAVE AN UMBRELLA
ENRICO

Formule des probabilités totales (version simple)

$$\begin{aligned} B &= B \cap S = B \cap (A \cup \overline{A}) \\ &= (B \cap A) \cup (B \cap \overline{A}) \end{aligned}$$

où S est l'ensemble fondamental.

Alors $P(B) = P(B \cap A) + P(B \cap \overline{A}) - 0$

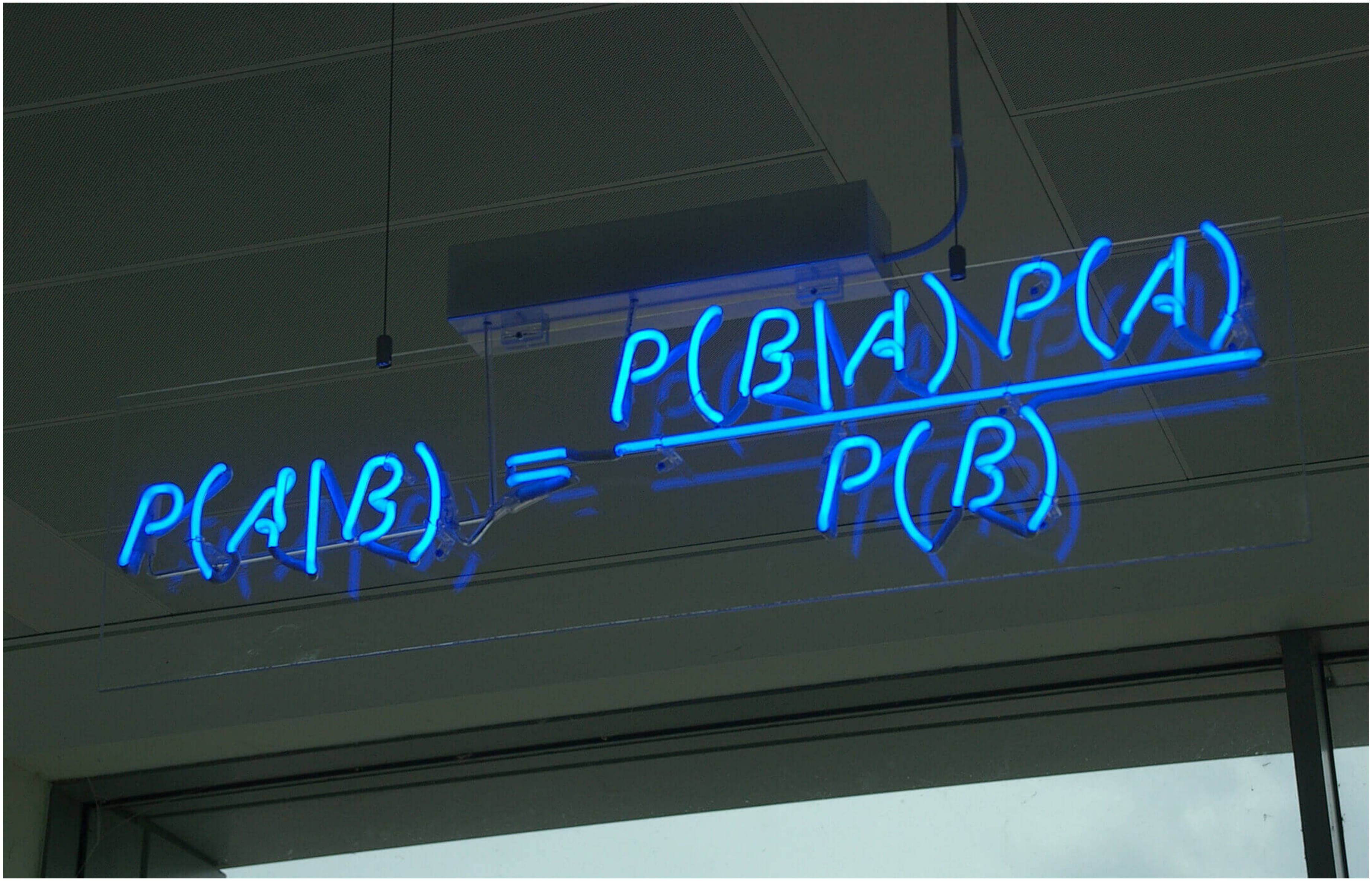
et donc

$$\boxed{P(B) = P(B \mid A) \cdot P(A) + P(B \mid \overline{A}) \cdot P(\overline{A})}$$

Théorème de Bayes (version simple)

Puisque $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$, on a

$$\boxed{P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A})}}$$


P(A|B) = P(B|A)P(A) / P(B)

EXEMPLE

	F Femmes	H Hommes	Total
A: ‘mathématiques’	4	14	18
B: ‘économétrie’	17	41	58
C: ‘physique’	4	25	29
D: ‘histoire’	28	11	39
Total	53	91	144

$$
\begin{aligned}
P(A) &= 18/144 \\
P(F) &= 53/144 \\
\\
P(A \cap F) &= 4/144 \\
P(A \cup F) &= (4 + 14 + 17 + 4 + 28)/144 \\
&= 18/144 + 53/144 - 4/144 \\
&= 67/144 \\
\\
P(A \mid F) &= 4/53 = \frac{4/144}{53/144} \\
P(F \mid A) &= 4/18
\end{aligned}
$$

$$
\begin{aligned}
P(\overline{A}) &= (144-18)/144 = 126/144 \\
P(F \mid \overline{A}) &= \frac{P(F \cap \overline{A})}{P(\overline{A})} = \frac{(17+4+28)/144}{126/144} = 49/126 \\
P(F) &= P(F \mid A) \cdot P(A) + P(F \mid \overline{A}) \cdot P(\overline{A}) \\
&= (4/18)(18/144) + (49/126)(126/144) \\
&= 4/144 + 49/144 = 53/144
\end{aligned}
$$

EXEMPLE

Analyse des résultats d'un audit.

Bill (B) analyse 60% des audits. Georges (G) analyse le reste. La probabilité d'erreur de lecture (E) de Bill est de 5% et celle de Georges de 3%.

1. Quelle est la probabilité d'erreur de lecture?
2. On a trouvé une erreur. Quelle est la probabilité qu'elle ait été commise par Bill?

On a

$$P(B) = 0.6$$

$$P(G) = 0.4$$

$$P(E \mid B) = 0.05$$

$$P(E \mid G) = 0.03$$

1.

$$
\begin{aligned}
P(E) &= P(E \mid B)P(B) + P(E \mid \overline{B})P(\overline{B}) \\
&= P(E \mid B)P(B) + P(E \mid G)P(G) \\
&= 0.05 \cdot 0.6 + 0.03 \cdot 0.4 = 0.042
\end{aligned}
$$

2.

$$
\begin{aligned}
P(B \mid E) &= P(E \mid B) \cdot \frac{P(B)}{P(E)} \\
&= 0.05 \cdot \frac{0.6}{0.042} \approx 0.71
\end{aligned}
$$

Analyse des résultats d'un audit, un autre regard.

Sur $1'000$ audits, 600 seront analysés par Bill (et donc 400 par Georges). Il y aura $600 \cdot 0.05 = 30$ erreurs de lecture chez Bill et $400 \cdot 0.03 = 12$ erreurs de lecture chez Georges, pour un total de 42 erreurs de lecture sur $1'000$ audits (4.2% ou probabilité de 0.042).

Sur $1'000$ audits on aura donc 42 erreurs de lecture, dont 30 commises par Bill. La probabilité que ce soit Bill qui ait commis l'erreur que l'on a trouvée est donc de $30/42 \approx 0.71$.

EXEMPLE

Un étudiant répond à une question à choix multiple.

De deux choses l'une: Soit il connaît la réponse, soit il la devine.

Soit p la probabilité que l'étudiant connaisse la réponse et donc $1 - p$ celle qu'il la devine.

On admet que l'étudiant qui devine répondra correctement avec probabilité $1/m$ où m est le nombre de réponses possibles.

Quelle est la probabilité qu'un étudiant connaisse la réponse à une question s'il y a répondu correctement?

Soient les événements 'C: l'étudiant répond correctement à la question' et 'R: il connaît vraiment la réponse'. On cherche donc $P(R \mid C)$.

On a

$P(R) = p$ et $P(\overline{R}) = 1 - p$;

$P(C \mid R) = 1$ et $P(C \mid \overline{R}) = 1/m$.

Alors

$$\begin{aligned} P(R \mid C) &= \frac{P(C \mid R)P(R)}{P(C \mid R)P(R) + P(C \mid \overline{R})P(\overline{R})} \\ &= \frac{1 \cdot p}{1 \cdot p + (1/m)(1-p)}. \end{aligned}$$

En prenant par exemple $m = 5$ et $p = 1/2$, la probabilité qu'un étudiant connaisse la réponse à une question sachant qu'il a répondu correctement sera ainsi $5/6 \approx 83.3\%$.

Et si par exemple $m = 5$ et $p = 1/5$, on obtient $5/9 \approx 55.5\%$.

EXEMPLE

Pourquoi la surveillance globale d'internet ne pourra jamais empêcher le terrorisme?

Prenons un cas fictif (avec des données de l'année 2014; voir `goo.gl/QbxoIr`), mais qui illustre très bien le propos.

En 2014, le Canada a une population d'environ 35 millions d'individus. Au Canada, il y a environ 85% de la population qui utilise internet.

Cela signifie donc que le système de surveillance automatisé d'internet s'appliquerait à 29'750'000 individus.

En 2014, le gouvernement canadien suspectait environ 80 personnes d'être impliquées dans des activités terroristes.

Admettons que ce chiffre est hautement conservateur et qu'il y a environ $1'000$ terroristes actifs au Canada — c'est impossible, mais disons que nous sommes dans un pays hautement 'à risque', cela signifierait qu'environ 0.0035% de la population serait impliquée dans des activités terroristes, donc 99.9965% ne le sont pas.

Admettons aussi que le Canada bénéficie d'un système de surveillance global des activités de ses citoyens sur internet.

Ce système serait d'une redoutable efficacité. Il serait capable d'identifier les terroristes dans 90% des cas — ce qui est bien au-delà de tout ce qui se fait actuellement, et il n'identifierait des terroristes à tort ('faux positif') que seulement dans 0.5% des cas.

Quelles sont les chances qu'un citoyen détecté par le système de surveillance globale en soit véritablement un terroriste ('vrai positif')?

Soient les événements 'T: être terroriste' et 'D: être détecté par le système de surveillance globale'. On cherche donc $P(T \mid D)$.

On a

$$P(T) = 0.0035\% \quad \text{et} \quad P(\overline{T}) = 99.9965\%;$$
$$P(D \mid T) = 90\% \quad \text{et} \quad P(D \mid \overline{T}) = 0.5\%.$$

Alors

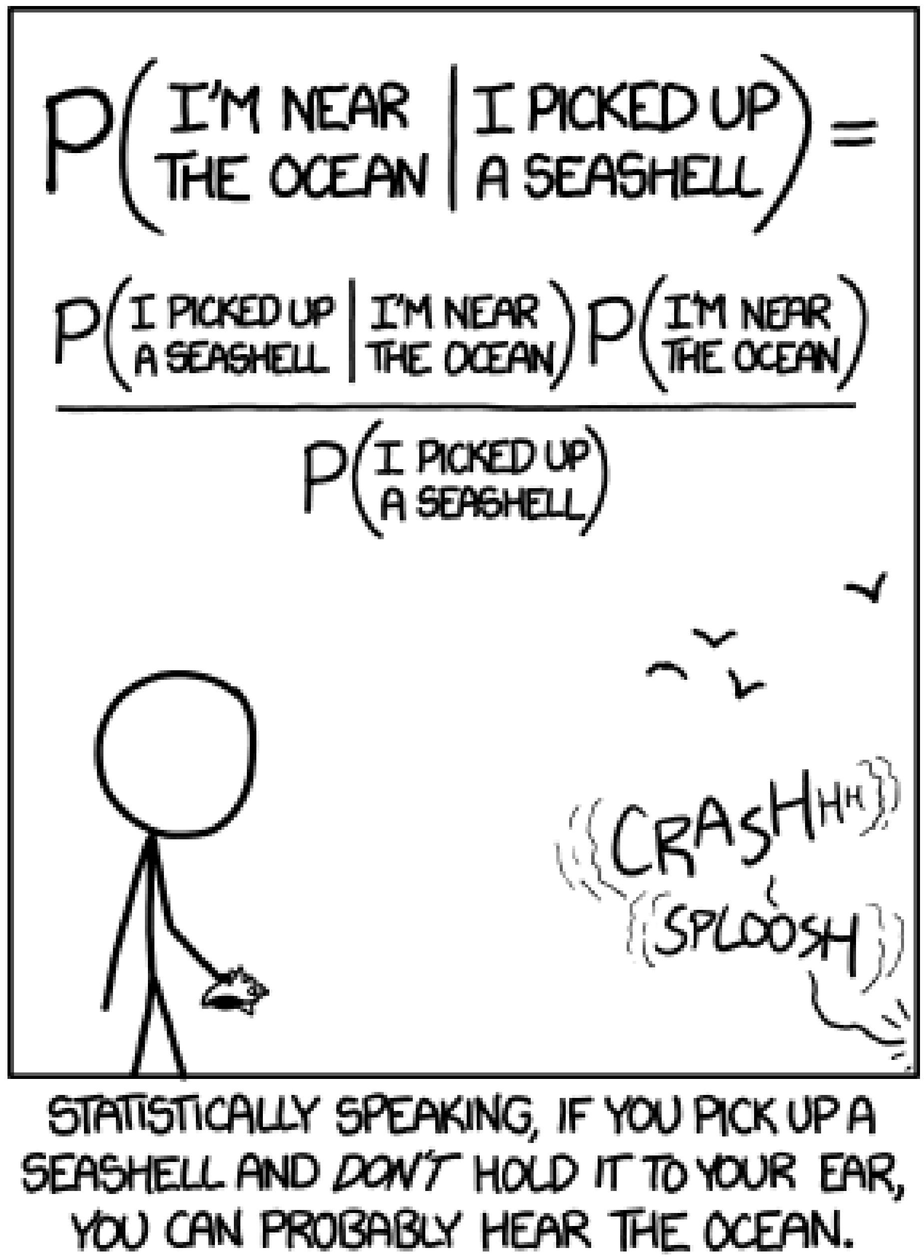
$$\begin{aligned} P(T \mid D) &= \frac{P(D \mid T)P(T)}{P(D \mid T)P(T) + P(D \mid \overline{T})P(\overline{T})} \\ &= \frac{90\% \cdot 0.0035\%}{90\% \cdot 0.0035\% + 0.5\% \cdot 99.9965\%} \approx 0.63\%. \end{aligned}$$

Donc, si le système de surveillance global retourne une détection ‘positive’, dans les faits, les probabilités que le citoyen détecté soit effectivement un terroriste ne sont que de 0.63%, plutôt que de 90% ce qui correspond à la soi-disant validité du système de surveillance globale.

Un autre regard sur la question: Si vous avez $1'000'000$ personnes devant vous, vous aurez environ 35 personnes qui seront considérées comme terroristes, et le système sera en mesure de les détecter dans 90% des cas.

Des $999'965$ personnes restantes, environ 0.5% seront détectées comme des terroristes sans l’être, soit environ $5'000$ personnes.

Ainsi, avec un exemple basé sur une population plus petite, cela signifie que vous aurez un ratio de 35 pour $5'000$, soit environ 0.70%, ce qui est proche du calcul qu’on vient de faire; la variabilité pouvant s’expliquer par la population qui est moindre.


P(I'M NEAR THE OCEAN | I PICKED UP A SEASHELL) =
P(I PICKED UP A SEASHELL | I'M NEAR THE OCEAN) P(I'M NEAR THE OCEAN)
P(I PICKED UP A SEASHELL)
CRASHHHH
SPLOOSH
STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND DON'T HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

Formule (ou théorème) des probabilités totales (version généralisée)

Soit un ensemble fondamental S. On appelle partition de S, ou encore système complet d'événements, tout ensemble $A_1, A_2, \ldots, A_q$ de sous-ensembles disjoints de S dont la réunion est S, *i.e.* $A_1 \cup A_2 \cup \cdots \cup A_q = S$.

Soit B un événement. Alors,

$$\begin{aligned} P(B) &= P(B \,|\, A_1) \cdot P(A_1) + P(B \,|\, A_2) \cdot P(A_2) \\ &\quad + \cdots + P(B \,|\, A_q) \cdot P(A_q). \end{aligned}$$

EXEMPLE

Trois machines M_1, M_2 et M_3 fabriquent des pièces dans les proportions respectives 25%, 35% et 40%. On sait que 5% des pièces produites par M_1 sont défectueuses, pour M_2 on a 4% et pour M_3 on a 2%.

$\rightsquigarrow$ Soit A l'événement 'la pièce choisie aléatoirement est défectueuse'.

$\rightsquigarrow$ Trouver $P(A)$.

$\rightsquigarrow$ Tous les événements sont disjoints, donc

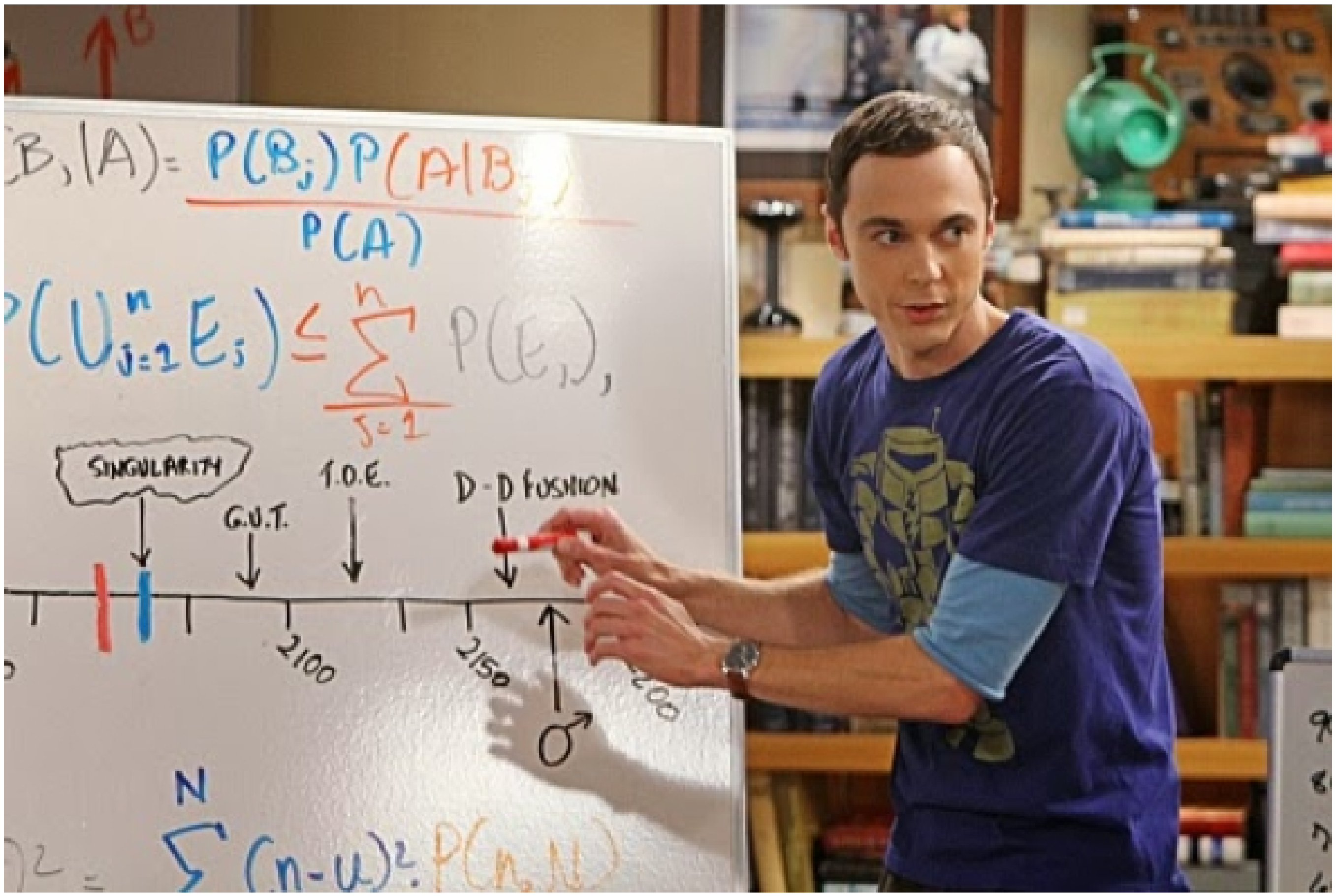
$$
\begin{aligned}
P(A) &= P(D \cap M_1) + P(D \cap M_2) + P(D \cap M_3) \\
&= P(D \mid M_1) \cdot P(M_1) + P(D \mid M_2) \cdot P(M_2) + P(D \mid M_3) \cdot P(M_3) \\
&= 0.05 \cdot 0.25 + 0.04 \cdot 0.35 + 0.02 \cdot 0.40 \\
&= 0.0125 + 0.014 + 0.008 = 0.0345.
\end{aligned}
$$

Théorème de Bayes (version généralisée)

Soient une partition $A_1, A_2, \ldots, A_q$ et un événement B d'un ensemble fondamental S, avec $P(B) \neq 0$. Pour tout indice k, $1 \leq k \leq q$, on aura,

$$
\begin{aligned}
P(A_k \,|\, B) &= \frac{P(A_k \cap B)}{P(B)} \\
&= \frac{P(B \,|\, A_k) \cdot P(A_k)}{P(B \,|\, A_1) \cdot P(A_1) + \cdots + P(B \,|\, A_q) \cdot P(A_q)}.
\end{aligned}
$$

Théorème de Bayes et la sitcom ‘*The Big Bang Theory*’



Source: ‘Bayes’s Theorem: What’s the Big Deal?’ (`goo.gl/2RUzwg`*).*

Remarque

- La formule de Bayes est davantage qu'une simple 'formule mathématique'.

$\rightsquigarrow$ Interprétons les événements A_k comme étant les états dans lesquels peut se trouver un système et B le résultat d'une expérience réalisée dans le but de déterminer dans lequel des états A_k le système se trouve. Alors,

- $P(A_k)$: Probabilité **a priori**; opinion de l'observateur **avant** l'expérience;

- $P(B \mid A_k)$: Probabilité d'observer B **si** le système se trouve dans l'état A_k;

- $P(A_k \mid B)$: Probabilité **a posteriori**; opinion de l'observateur après l'expérience sachant que B s'est réalisé.

La formule de Bayes apparaît comme une **règle méthodologique**.

$\rightsquigarrow$ En effet, elle précise comment on doit modifier son opinion en tenant compte d'une expérience.

Application: Filtre Bayesien anti-spam

• Dans le milieu informatique, le courrier électronique (courriel) publicitaire (spam) est un véritable fléau.

$\rightsquigarrow$ Le nombre de messages spam est quotidiennement en hausse.

• Actuellement pour combattre le problème grandissant du courriel publicitaire, l'outil le plus efficace est sans doute le **filtre Bayesien**.

$\rightsquigarrow$ Grâce à ce filtre, dont le fonctionnement repose sur le théorème de Bayes, on parvient maintenant à un taux de détection de spam de plus de 98%!

$\rightsquigarrow$ Mais comment fonctionne le filtre Bayesien pour détecter aussi bien le courriel publicitaire?

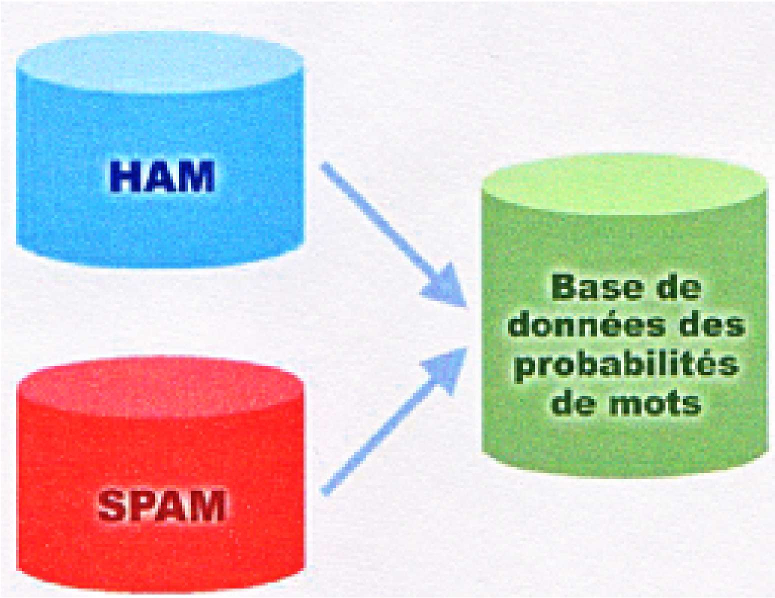
• Avant que le courriel puisse être filtré, l'utilisateur a besoin de générer une base de données de mots et signes particuliers comme 'hypothèque', 'viagra', 'sexe', '$\$$', 'miracle', rassemblés à partir d'exemples de courriel spam (courriel non désiré) et ham (courriel légitime).

• Par simplification, notons ces mots ou signes $1, 2, \ldots, n$. Pour chacun de ces mots ou signes, on détermine les probabilités suivantes:

p_i : Probabilité qu'un mot choisi au hasard dans un courriel est le mot i en sachant que le message est un spam;

q_i : Probabilité qu'un mot choisi au hasard dans un courriel est le mot i en sachant que le message n'est pas un spam.

$\rightsquigarrow$ Les probabilités p_i et q_i des mots i $(i = 1, \ldots, n)$ sont calculées directement à partir d'exemples de courriel spam et ham propres à la compagnie, à l'établissement scolaire ou à l'organisation.



• Ecrivons les probabilités p_i et q_i à l'aide des événements:

M_i : 'le mot choisi au hasard dans le courriel est le mot i';

Sp : 'le courriel est un spam'.

Ainsi,

$$p_i = P(M_i|Sp) \quad \text{et} \quad q_i = P(M_i|\overline{Sp}).$$

$\rightsquigarrow$ Probabilités p_i et q_i calculées pour les mots et signes choisis, le filtre est prêt à l'emploi.

$\rightsquigarrow$ À l'arrivée d'un nouveau courriel, celui-ci est analysé et décomposé en mots et les mots les plus importants sont isolés. Le filtre calcule alors la probabilité que ce message soit un spam ou non.

$\rightsquigarrow$ Si la probabilité dépasse un certain seuil (par exemple 0.9), le message est considéré comme spam.

• Pour illustrer le fonctionnement du filtre Bayesien, supposons que la proportion de messages spam vaut 0.9, *i.e.* $P(Sp) = 0.9$ et donc $P(\overline{Sp}) = 1 - 0.9 = 0.1$, et supposons que pour le mot 'hypothèque' noté 1, $p_1 = 0.05$ et $q_1 = 0.001$.

$\rightsquigarrow$ Pour ce mot, on a alors

$$p_1 = P(M_1|Sp) = 0.05 \quad \text{et} \quad q_1 = P(M_1|\overline{Sp}) = 0.001.$$

$\rightsquigarrow$ Un nouveau courriel vient d'arriver et le mot 'hypothèque' y apparaît. En appliquant le théorème de Bayes, la probabilité que le courriel électronique soit un spam est

$$P(Sp|M_1) = \frac{P(M_1|Sp) \cdot P(Sp)}{P(M_1|Sp) \cdot P(Sp) + P(M_1|\overline{Sp}) \cdot P(\overline{Sp})} = 0.9978.$$

Remarques

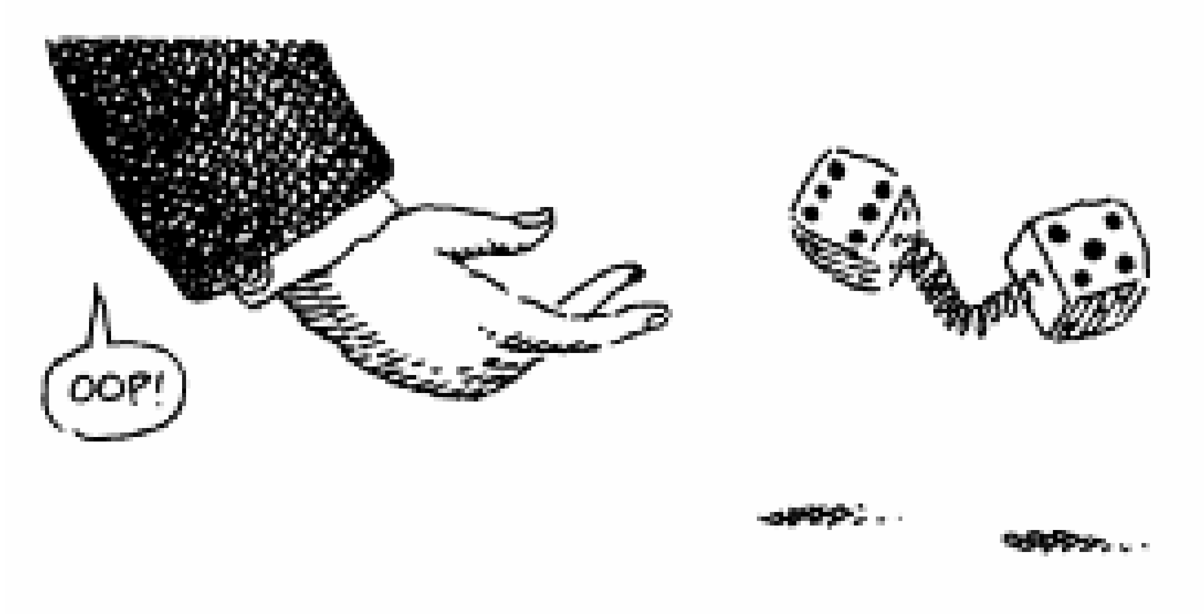
- Comme le choix des mots et signes et l'analyse de courriel ham et spam sont propres à l'utilisateur, le filtre Bayesien correspondra parfaitement aux besoins de celui-ci;

- La liste des mots clés se différencient selon les activités de l'établissement. Par exemple, le mot 'hypothèque' est considéré comme suspect pour un établissement scolaire mais pas du tout pour une banque. Ainsi, le filtre Bayesien apprend les habitudes de messagerie des établissements;

- La période d'apprentissage qui comprend l'analyse du courriel ham et spam et le calcul des probabilités dure quelques jours;

- On peut néanmoins avoir de fausses alertes en particulier au début de la mise en activité du filtre;

- Le filtre Bayesien
  - est une approche intelligente étant donné qu'elle reconnaît non seulement les mots clés qui identifient le spam mais aussi ceux qui dénotent un message valide. Elle va au-delà d'une simple vérification de la présence de mots clés et d'un classement direct d'un message dans la catégorie spam;
  - s'adapte et évolue continuellement en apprenant à partir de nouveaux spam et de nouveaux courriels. Par exemple lorsque les 'spammeurs' ont utilisé 'f-r-e-e' au lieu de 'free', ils sont parvenus à passer à travers le filtre jusqu'à ce que le mot soit intégré dans la base de données;
  - fonctionne pour n'importe quelle langue et n'importe quel langage.



Indépendance ou dépendance?

2.6 Indépendance

A est indépendant de B

$$\Leftrightarrow$$

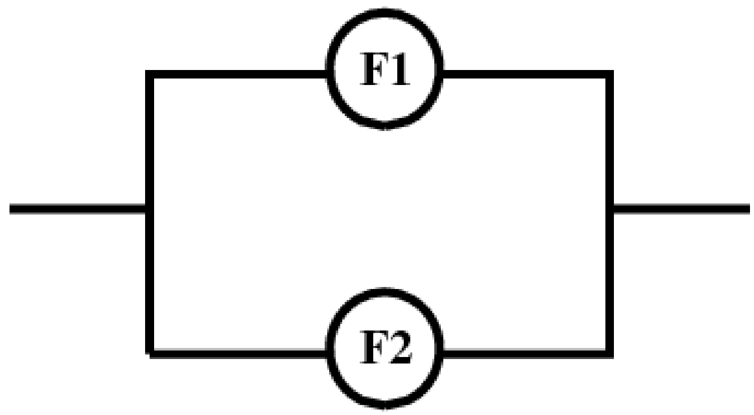
$$P(A \mid B) = P(A)$$
$$P(B \mid A) = P(B)$$

L'indépendence de deux événements A et B implique que

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

EXEMPLE

Fonctionnement d'une machine (simple):



$A =$ 'F1 fonctionne' et $B =$ 'F2 fonctionne' sont **indépendants**. On a $P(A) = 0.7$ et $P(B) = 0.8$.

Quelle est la probabilité que la machine fonctionne?

$$
\begin{aligned}
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\
&= P(A) + P(B) - P(A) \cdot P(B) \\
&= 0.7 + 0.8 - 0.7 \cdot 0.8 \\
&= 1.5 - 0.56 \\
&= 0.94
\end{aligned}
$$

Remarques

a) L'indépendance est supposée dans de nombreuses applications comme par exemple dans la répétition d'un lancer de dé, dans le fonctionnement d'un dispositif d'alarmes dans une banque.

$\rightsquigarrow$ La notion d'indépendance est fondamentale dans le calcul des probabilités;

b) La règle de multiplication peut être généralisé à plus de deux événements: un ensemble de n événements $A_1, A_2, \ldots, A_n$ est dit (totalement ou mutuellement) indépendant si pour tout sous-ensemble $A_1, A_2, \ldots, A_r$, $r \leq n$, on a

$$P(A_1 \cap A_2 \cap \ldots \cap A_r) = P(A_1) \cdot P(A_2) \cdot \cdots \cdot P(A_r);$$

c) L'indépendance de deux événements en probabilités peut être évidente. Par exemple, une pièce de monnaie équilibrée est jetée deux fois de suite. On définit les événements:

A: 'face apparaît au premier jet';

B: 'pile apparaît au second jet'.

Cependant, l'indépendance n'est pas toujours ressentie de manière intuitive aussi évidente. Par exemple, considérons les événements:

C: 'le même côté de la pièce sort deux fois de suite';

D: 'le nombre de face est strictement inférieur à 2'.

Sont-ils indépendants? La réponse n'est pas évidente;

d) Si les événements A et B sont incompatibles, *i.e.* $A \cap B = \emptyset$, ils ne sont pas indépendants;

e) Si A et B sont indépendants, A et $\overline{B}$ le sont aussi. Autrement dit, lorsque A est indépendant de B, la probabilité que A survienne n'est influencée ni par l'information que B est réalisé ni par celle que B ne l'est pas;

f)

- $\diamond$ A et B incompatibles $\Rightarrow P(A \cup B) = P(A) + P(B)$;
- $\diamond$ A et B indépendants $\Rightarrow P(A \cap B) = P(A) \cdot P(B)$.


ALEA IACTA EST !
$P = \frac{VI!/(VI-IV)!}{VI^{IV}}$
ENRICO

3. Distributions de probabilité discrètes

3.1 Généralités sur les variables aléatoires

• Il arrive fréquemment que lors d'une expérience aléatoire, on s'intéresse davantage à une **fonction du résultat** plutôt qu'au résultat en lui-même.

$\rightsquigarrow$ Illustrons cette idée à l'aide des exemples suivants:

1. Jet de deux dés: On s'intéresse à la somme obtenue, 7 par exemple, plutôt qu'au fait de savoir si c'est le couple $\{1, 6\}$ qui est apparu ou $\{2, 5\}$, $\{3, 4\}$, $\{4, 3\}$, $\{5, 2\}$ ou $\{6, 1\}$;
2. Jet d'une pièce de monnaie: On s'intéresse au nombre de fois où pile est apparu plutôt qu'à la séquence détaillée des piles ou faces.

$\rightsquigarrow$ Ces grandeurs auxquelles on s'intéresse sont en fait des fonctions réelles définies sur l'ensemble fondamental et sont appelées **variables aléatoires**.

• Soit un 'espace probabilisé' d'ensemble fondamental (*i.e.* ensemble des issues) Ω et de mesure de probabilité P.

$\rightsquigarrow$ Une variable aléatoire réelle X est une application de Ω dans $S \subset \mathbb{R}$:

$$\begin{aligned} X \quad : \quad \Omega &\longrightarrow S \subset \mathbb{R} \\ \omega &\longmapsto X(\omega). \end{aligned}$$

$\rightsquigarrow$ L'ensemble S des valeurs prises par la variable aléatoire (*i.e.* l'espace des réalisations) peut être:

- $\diamond$ discret $\rightsquigarrow$ variable aléatoire discrète;
- $\diamond$ continu $\rightsquigarrow$ variable aléatoire continue.

$\rightsquigarrow$ Une variable aléatoire discrète ne prend qu'un nombre fini ou dénombrable de valeurs.

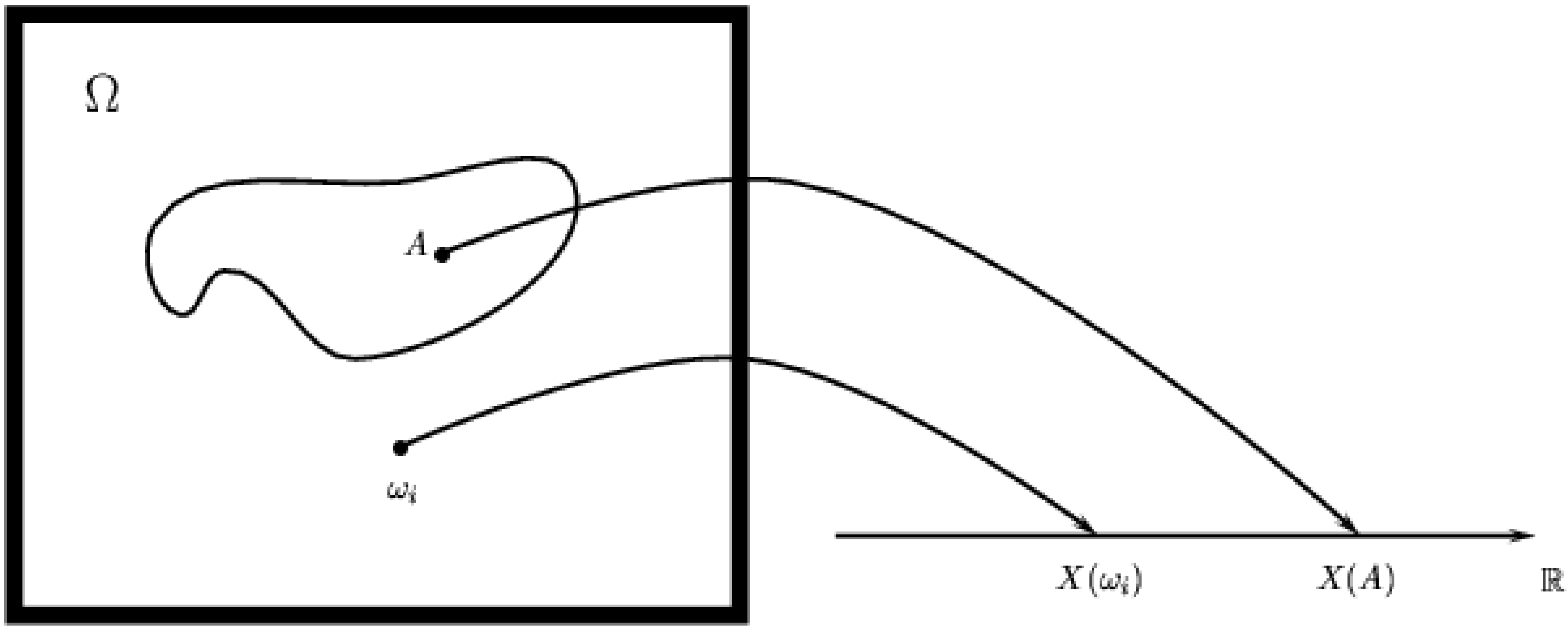


Illustration d'une variable aléatoire réelle

(avec un ensemble fondamental Ω, un événement élémentaire $\omega_i = e_i$ et un événement A).

EXEMPLES

- Nombre de pièces non conformes dans un échantillon de taille 10 prélevé au hasard dans la production: $S = \{0, 1, \ldots, 10\}$;
- Nombre de piles pour n parties de pile ou face: $S = \{0, 1, \ldots, n\}$;
- Nombre d'appels téléphoniques parvenant à un standard durant un intervalle de temps fixé: $S = \mathbb{N}$;
- Proportion de réponses 'oui' à une question posée dans un sondage: $S = \{x \in \mathbb{Q},\ 0 \leq x \leq 1\}$;
- Temps d'attente en minutes à un guichet: $S = \{x \in \mathbb{R},\ x > 0\}$.

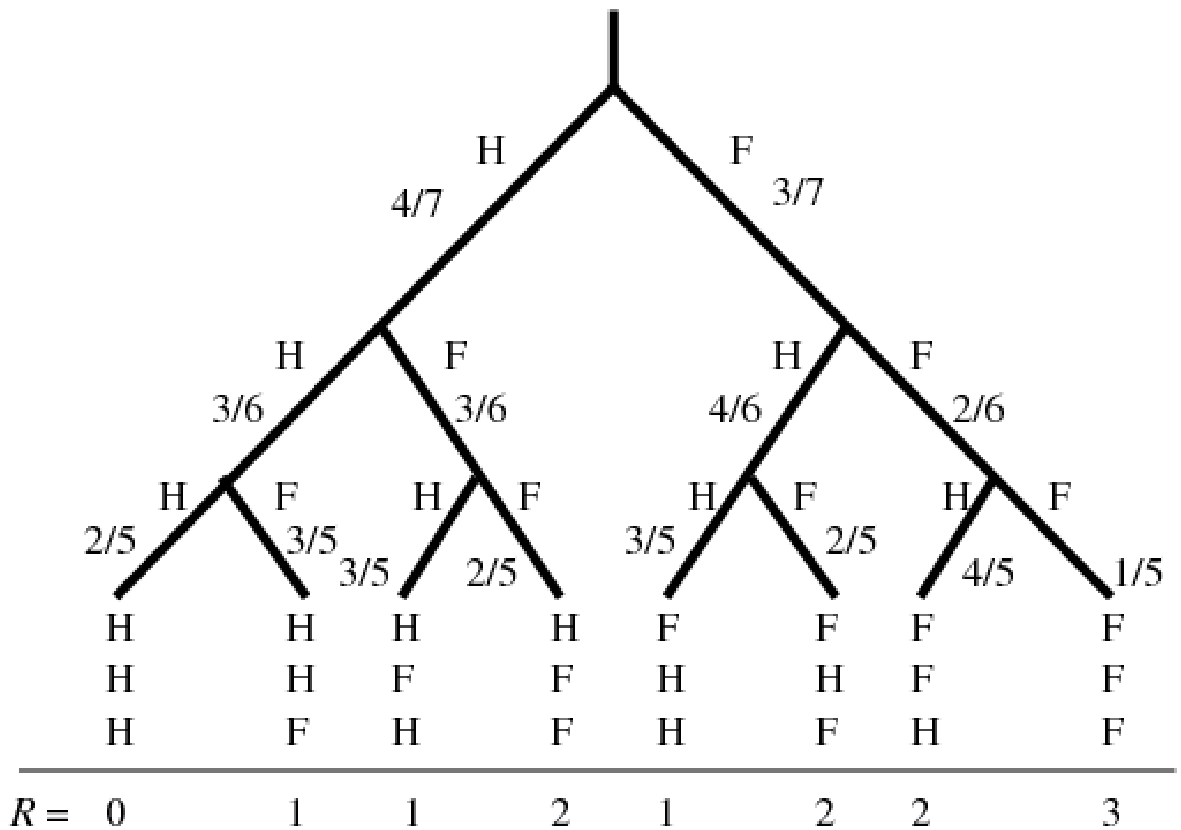
3.2 Définitions

Variable aléatoire discrète:	R
Réalisation:	r
Espace des réalisations:	S
Probabilité:	$P(R = r) = p_r$
Distribution de probabilité:	$S \to [0, 1]$ $r \mapsto p_r$
Représentation:	Tables (des probabilités), graphiques, formules

EXEMPLE

Un comité composé de 3 personnes est tiré au sort parmi 4 hommes et 3 femmes.

Soit R le nombre de femmes du comité, $S = \{0, 1, 2, 3\}$.



Distribution de probabilité:

r	0	1	2	3	Total
p_r	4/35	18/35	12/35	1/35	1

Fonction de répartition $F_R(r)$:

$$\boxed{F_R(r) = P(R \leq r)}$$

pour $-\infty \leq r \leq +\infty$.

C'est une fonction de probabilités cumulées, étendue à l'ensemble des nombres réels, et prend ses valeurs dans $[0,\ 1]$.

Propriétés:

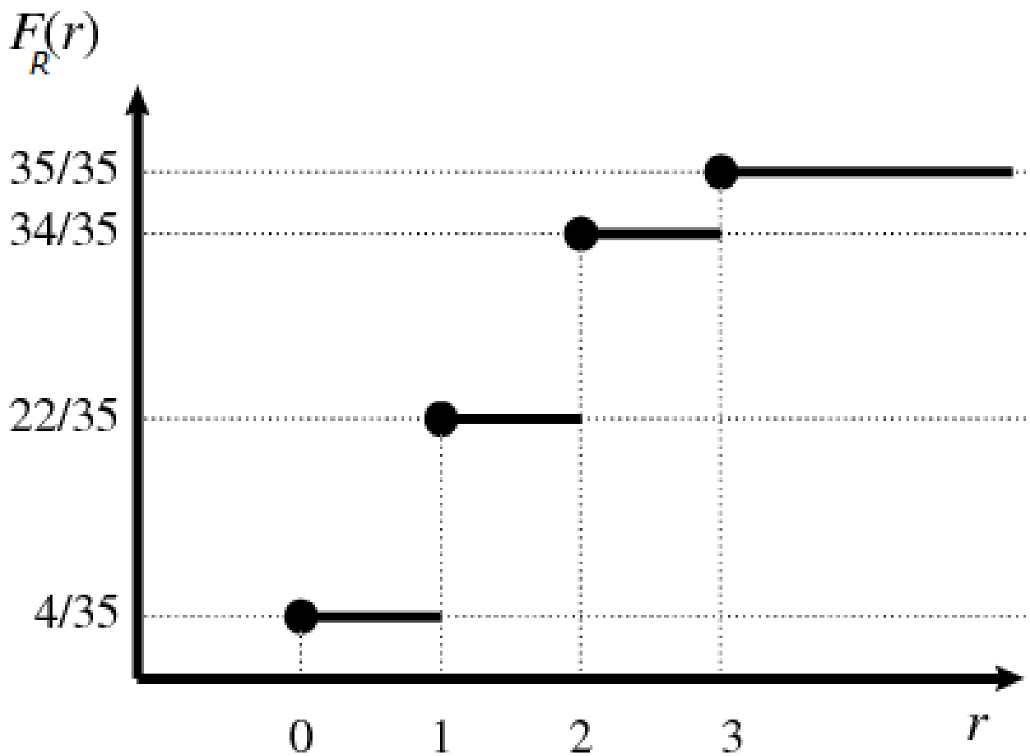
- $F_R(r)$ est une fonction croissante;
- $\lim_{r \to -\infty} F_R(r) = 0$;
- $\lim_{r \to +\infty} F_R(r) = 1$.

EXEMPLE

r	0	1	2	3	Total
p_r	4/35	18/35	12/35	1/35	1
$P(R \leq r)$	4/35	22/35	34/35	35/35	

$$F_R(r) = \begin{cases} 0 & r < 0 \\ 4/35 & 0 \leq r < 1 \\ 22/35 & 1 \leq r < 2 \\ 34/35 & 2 \leq r < 3 \\ 1 & r \geq 3 \end{cases}$$

Graphiquement: Fonction en escalier:



Remarque: La fonction de répartition $F_R(r)$ étant une fonction de probabilités cumulées, il est possible de l'inverser (dans le cas d'une variable aléatoire discrète, elle ne peut être inversée au sens propre!) afin de déterminer la valeur de r qui correspond à une certaine probabilité cumulée $\alpha = P(R \leq r)$, avec $0 \leq \alpha \leq 1$.

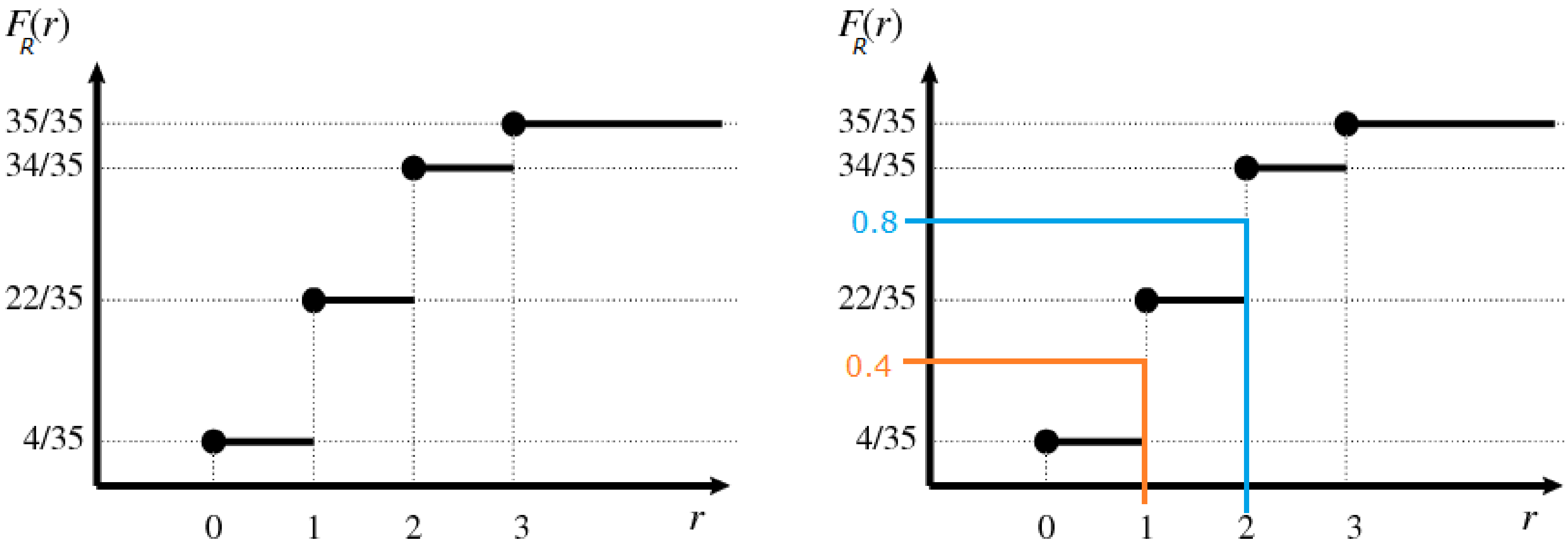
Fonction de répartition inverse $F_R^{-1}(\alpha)$ ou **quantile d'ordre** α, noté $Q(\alpha)$: La plus petite réalisation de R associée à une probabilité cumulée supérieure ou égale à α:

$$\boxed{F_R\left[F_R^{-1}(\alpha)\right] = P\left[R \leq F_R^{-1}(\alpha)\right] \geq \alpha}$$

pour $0 \leq \alpha \leq 1$.

Par construction, un quantile d'une distribution discrète est une réalisation de la variable R.

EXEMPLE



La quantile d'ordre $\alpha = 0.4$ correspond à la plus petite réalisation a de R (*i.e.* soit $0, 1, 2$ ou 3) telle que la probabilité cumulée $F_R(a) = P(R \leq a)$ soit supérieure ou égale à $0.4 = 14/35$. Donc $a = 1 = Q(0.4)$, *i.e.* $F_R^{-1}(0.4) = 1$.

Et la quantile d'ordre $\alpha = 0.8 = 28/35$ correspond à $Q(0.8) = 2$, *i.e.* $F_R^{-1}(0.8) = 2$.

▷ **Interprétation:** Si le quantile d'ordre $\alpha = 0.05$ est égal à $Q(0.05) = F_R^{-1}(0.05) = 10$, cela signifie qu'il y a 5% de chances que les réalisations de la variable aléatoire discrète R soient inférieures ou égales à 10.

Remarque: Le quantile d'ordre $\alpha = 0.5$ est appelé la **médiane**, noté par $Q_2 = Q(0.5)$. Les quantiles à 25% et 75% sont respectivement notés $Q_1 = Q(0.25)$ et $Q_3 = Q(0.75)$. La distance $Q_3 - Q_1$ est appelée **écart interquartile**. Cet écart est une mesure de la dispersion des réalisations de la variable aléatoire, *i.e.* de sa distribution.

3.3 Espérance et variance

μ : $\boxed{\textbf{Espérance de } R}$, dénotée par $\boxed{\mathrm{E}(R)}$

: **'Moyenne (vraie) de R'**

: **'Centre de la distribution'**

σ^2 : $\boxed{\textbf{Variance de } R}$, dénotée par $\boxed{\mathrm{var}(R)}$

: **'Déviation quadratique moyenne (vraie) de R'**

: **'Déviation quadratique espérée de R'**

: **'Dispersion de la distribution autour du centre'**

▷ Interprétations tirées de la physique: 'espérance $\equiv$ centre de gravité d'un ensemble de masses' et 'variance $\equiv$ moment d'inertie relative au centre de masse'.

$$\mu = \mathsf{E}(R) = \sum_{r \in S} r \cdot p_r$$

$$\sigma^2 = \mathsf{var}(R) = \mathsf{E}[(R-\mu)^2] = \sum_{r \in S} (r-\mu)^2 \cdot p_r$$

$\boxed{\rightsquigarrow \mu \text{ et } \sigma^2 \text{ sont des paramètres.}}$

Propriété de var(R):

$$\mathsf{var}(R) = \mathsf{E}(R^2) - [\mathsf{E}(R)]^2$$

et donc $\mathsf{var}(R) = \sum_{r \in S} r^2 \cdot p_r - \mu^2$.

Remarque: $\sigma = \sqrt{\sigma^2}$ est l'écart-type.

EXEMPLE

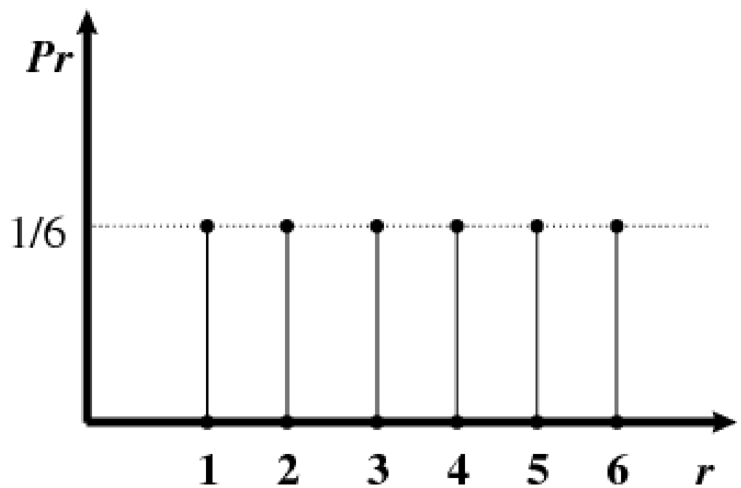
Un dé est jeté une fois.

$R =$ valeur sur la face supérieure du dé

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

Distribution de probabilité:

r	1	2	3	4	5	6	Total
p_r	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1



Donc $P(R = r) = p_r = 1/6$, $r = 1, \ldots, 6$.

Espérance:

$$\begin{aligned}
\mu &= \sum_{r=1}^{6} r \cdot p_r \\
&= 1 \cdot 1/6 + 2 \cdot 1/6 + 3 \cdot 1/6 + 4 \cdot 1/6 + 5 \cdot 1/6 + 6 \cdot 1/6 \\
&= 21/6 = 7/2 = 3.5
\end{aligned}$$

Variance:

$$\begin{aligned}
\sigma^2 &= \sum_{r=1}^{6} r^2 \cdot p_r - \mu^2 \\
&= 1 \cdot 1/6 + 4 \cdot 1/6 + 9 \cdot 1/6 + 16 \cdot 1/6 + 25 \cdot 1/6 + 36 \cdot 1/6 - 3.5^2 \\
&= 2.91\overline{6}
\end{aligned}$$

EXEMPLE

Nombre de femmes dans un comité de 3 personnes (voir page 137):

r	0	1	2	3	Total
p_r	4/35	18/35	12/35	1/35	1

$$\begin{aligned}\mu &= \sum_{r=0}^{4} r \cdot p_r \\ &= 0 \cdot 4/35 + 1 \cdot 18/35 + 2 \cdot 12/35 + 3 \cdot 1/35 \\ &= 45/35 = 9/7 \approx 1.2857\end{aligned}$$

On a $\mathsf{E}(R^2) = 0 \cdot 4/35 + 1 \cdot 18/35 + 4 \cdot 12/35 + 9 \cdot 1/35$
$= 75/35 = 15/7$

et donc $\sigma^2 = 15/7 - 81/49 = 24/49 \approx 0.4898$.

Propriétés

Pour une constante a on a

$$\mathsf{E}(a) = a \;\; \text{et} \;\; \mathsf{var}(a) = 0$$

Pour une constante a et une variable aléatoire R

$$\mathsf{E}(aR) = a\mathsf{E}(R) \;\; \text{et} \;\; \mathsf{var}(aR) = a^2\mathsf{var}(R)$$

Aussi, pour une transformation affine de R:

$$H = a + bR$$

avec $\mathsf{E}(R) = \mu_R$ et $\mathsf{var}(R) = \sigma_R^2$, on a que

$$\begin{aligned} \mathsf{E}(H) &= \mu_H = a + b\mu_R \\ \mathsf{var}(H) &= \sigma_H^2 = b^2\sigma_R^2 \end{aligned}$$

Remarque: **Ces propriétés sont aussi valables pour les variables continues.**

Combinaison linéaire de deux variables aléatoires

On a

$$\mathsf{E}(aR + bT) \quad = \quad a\mu_R + b\mu_T$$

avec $\mathsf{E}(T) = \mu_T$. En particulier:

$$\mathsf{E}(R + T) \quad = \quad \mu_R + \mu_T$$

$$\mathsf{E}(R - T) \quad = \quad \mu_R - \mu_T$$

Si de plus les variables sont **indépendantes**, alors

$$\mathsf{var}(aR + bT) \quad = \quad a^2\sigma_R^2 + b^2\sigma_T^2$$

avec $\mathsf{var}(T) = \sigma_T^2$.

En particulier:

$$\begin{aligned}\mathrm{var}(R+T) &= \sigma_R^2 + \sigma_T^2 \\ \mathrm{var}(R-T) &= \sigma_R^2 \boxed{+} \sigma_T^2 \qquad \boxed{!!!}\end{aligned}$$

Attention:

$$\sigma_{R+T} \neq \sigma_R + \sigma_T$$

$$\sigma_{R+T} = \sqrt{\sigma_R^2 + \sigma_T^2}$$

Remarque: **Ces propriétés sont aussi valables pour les variables continues.**


IS IT CLEAR TO EVERYONE?
QED
ENRICO 08

3.4 La distribution binomiale

La distribution **binomiale** est la loi de probabilité du nombre de succès parmi n tirages indépendants d'une expérience.

Exemples de variables aléatoires discrètes ayant une distribution binomiale:

- Le nombre de pile lorsqu'une pièce est lancée plusieurs fois;
- Le nombre de 5 lorsque un dé est jeté plusieurs fois;
- Le nombre de filles parmi trois enfants;
- Le nombre de pièces défectueuses dans un échantillon.

Caractéristiques d'une variable aléatoire binomiale:

1. On suppose qu'il y a n **tirages** d'une **expérience** (*i.e.* n expériences), par exemple n lancers d'une pièce de monnaie;
2. A chaque répétition de l'expérience, un événement se passe ou ne se passe pas. On a un <u>succès</u> (par exemple 'pile') ou un <u>échec</u> (par exemple 'face');
3. Les tirages sont indépendants;
4. La probabilité de succès p reste la même à chaque tirage (par exemple $p = 0.5$ pour 'pile').

Le nombre de succès durant n tirages indépendants est appelé une variable aléatoire binomiale.

Notation: R est une variable aléatoire binomiale:

$$\boxed{R \sim B(n, p)}$$

Espace des réalisations: $S = \{0, 1, 2, 3, 4, 5, \ldots, n\}$

Distribution de probabilité: $P(R = r) = ?$

EXEMPLE

Un dé est jeté 4 fois.

$R =$ nombre de 1 sur la face supérieure avec

$$\begin{cases} p = 1/6 & \text{probabilité de ‘succès’ (S)} \\ 1 - p = 5/6 & \text{probabilité d’‘échec’ (E)} \end{cases}$$

Réalisations:

						R
EEEE						0
SEEE	ESEE	EESE	EEES			1
SSEE	SESE	SEES	ESSE	ESES	EESS	2
SSSE	SSES	SESS	ESSS			3
SSSS						4

Probabilités:

$$
\begin{aligned}
P(R=0) \quad &= P(\mathsf{EEEE}) = (5/6)^4 \\
P(R=1) \quad &= P(\mathsf{SEEE}) + \cdots + P(\mathsf{EEES}) \\
&= 4 \cdot (1/6) \cdot (5/6)^3 = \binom{4}{1} \cdot (1/6)^1 \cdot (5/6)^3 \\
P(R=2) \quad &= P(\mathsf{SSEE}) + \cdots + P(\mathsf{EESS}) \\
&= 6 \cdot (1/6)^2 \cdot (5/6)^2 = \binom{4}{2} \cdot (1/6)^2 \cdot (5/6)^2 \\
P(R=3) \quad &= P(\mathsf{SSSE}) + \cdots + P(\mathsf{ESSS}) \\
&= 4 \cdot (1/6)^3 \cdot (5/6) = \binom{4}{3} \cdot (1/6)^3 \cdot (5/6)^1 \\
P(R=4) \quad &= P(\mathsf{SSSS}) = (1/6)^4
\end{aligned}
$$

Distribution de probabilité: $R \sim B(n, p)$

$$\boxed{\begin{aligned} p_r &= P(R = r) \\ &= \binom{n}{r} p^r (1-p)^{n-r} \end{aligned}}$$

Rappels: $\displaystyle\binom{n}{r} = \frac{n!}{r!(n-r)!}$ et on a

$$\sum_{r=0}^{n} p_r = \sum_{r=0}^{n} \binom{n}{r} p^r (1-p)^{n-r} = [p + (1-p)]^n = 1.$$

Remarque: Lorsque $p = 0.5$ on a

$$p_r = \binom{n}{r} (0.5)^r (1-0.5)^{n-r} = \binom{n}{r} (0.5)^n.$$

EXEMPLE

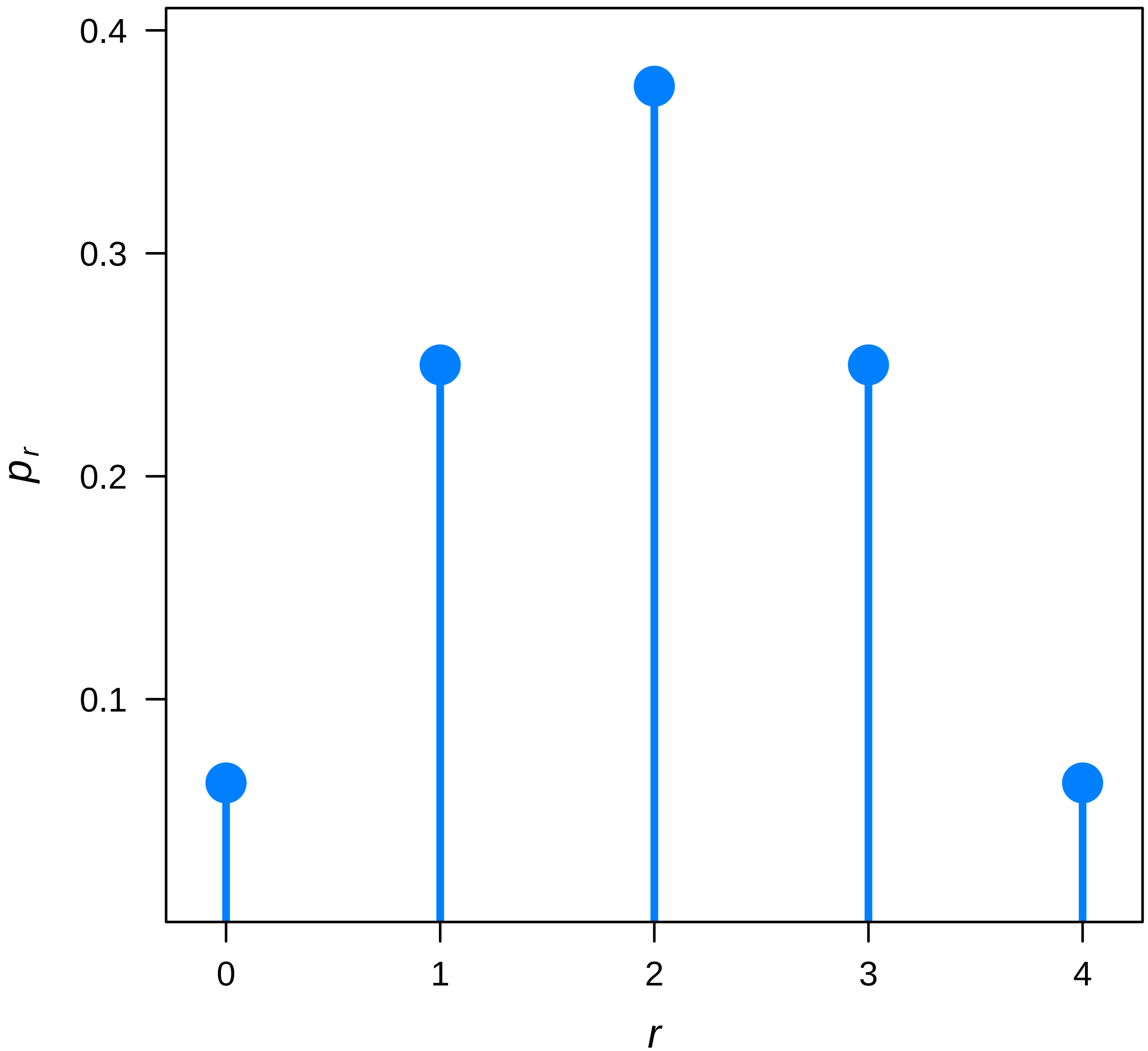
On lance une pièce équilibrée 4 fois.

$R =$ nombre de pile

$R \sim B(n = 4, p = 0.5)$

$$\begin{array}{rclcl}
P(R=0) & = & \binom{4}{0}(0.5)^4 & = & 1/16 \\
P(R=1) & = & \binom{4}{1}(0.5)^4 & = & 4/16 \\
P(R=2) & = & \binom{4}{2}(0.5)^4 & = & 6/16 \\
P(R=3) & = & \binom{4}{3}(0.5)^4 & = & 4/16 \\
P(R=4) & = & \binom{4}{4}(0.5)^4 & = & 1/16 \\
 & & \text{Total:} & & 1
\end{array}$$

$B(4, 0.5)$



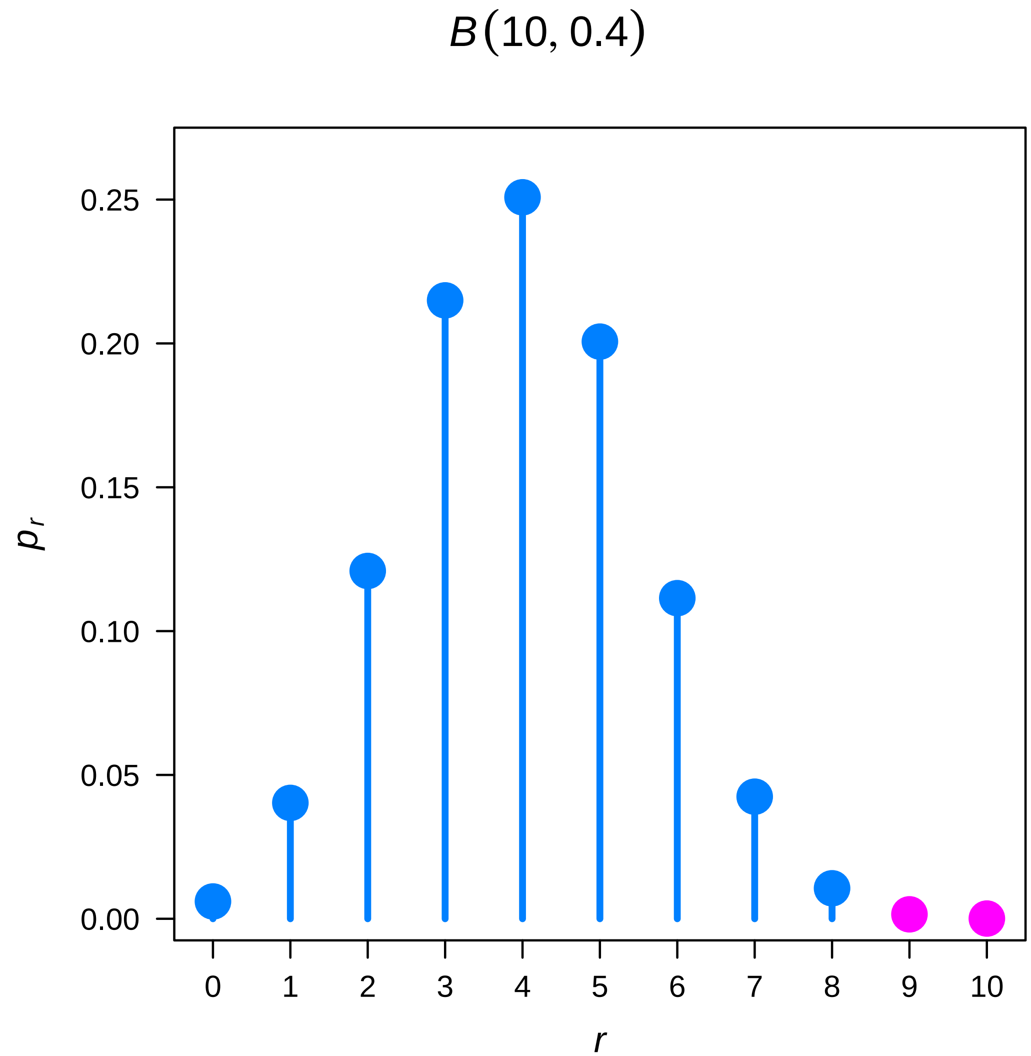
EXEMPLE

La probabilité pour un vendeur de gagner au moins 20 francs par heure est de 40%. Quelle est la probabilité d'avoir plus de 8 vendeurs sur 10 gagnant au moins 20 francs par heure?

Réponse: Soit R le nombre de vendeurs gagnant au moins 20 francs par heure. On a $R \sim B(10, 0.4)$ et donc

$$
\begin{aligned}
P(R > 8) &= P(R = 9) + P(R = 10) \\
&= \binom{10}{9} 0.4^9 (1 - 0.4)^{10-9} + \binom{10}{10} 0.4^{10} (1 - 0.4)^{10-10} \\
&= 0.0017.
\end{aligned}
$$

Illustration:



Cas particulier: $n = 1$

La distribution $B(1, p)$, *i.e.* nombre de succès sur un tirage, est appelée **distribution de $\boxed{\textbf{Bernoulli}}$**.

Espace des réalisations: $S = \{0, 1\}$

La loi de probabilité de $R \sim B(1, p)$ est donnée par:

r	0	1	Total
p_r	$1 - p$	p	1

$\rightsquigarrow$ Espérance et variance d'une Bernoulli:

$$\begin{aligned}
\mathsf{E}(R) &= 0 \cdot (1-p) + 1 \cdot p \\
&= p
\end{aligned}$$

$$\begin{aligned}
\mathsf{var}(R) &= (0-p)^2(1-p) + (1-p)^2 p \\
&= p^2(1-p) + (1-p)^2 p \\
&= p(1-p)(p+1-p) \\
&= p(1-p)
\end{aligned}$$

Espérance et variance d'une binomiale:

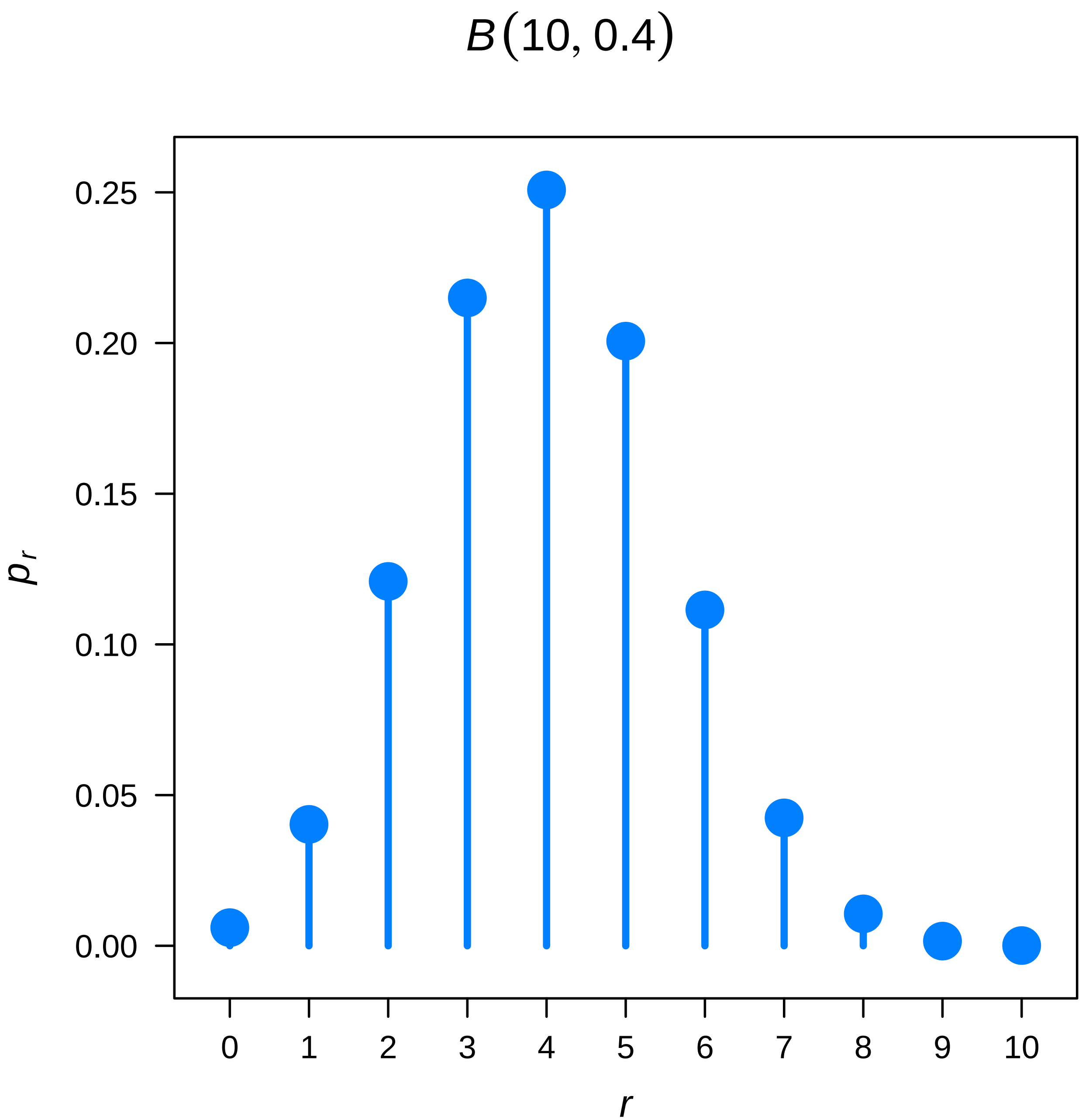
$$R \sim B(n,p)$$

$$\begin{aligned} \mathsf{E}(R) &= n \cdot p \\ \mathsf{var}(R) &= n \cdot p \cdot (1-p) \end{aligned}$$

Et l'écart-type est $\sqrt{\mathsf{var}(R)} = \sqrt{n \cdot p \cdot (1-p)}$.

Exemple: On a 10 vendeurs et la probabilité pour un vendeur de gagner au moins 20 francs par heure est de 40%.

Soit R le nombre de vendeurs gagnant au moins 20 francs par heure. On a $R \sim B(10, 0.4)$ et donc, en moyenne, il y a $10 \cdot 0.4 = 4$ vendeurs gagnant au moins 20 francs par heure sur 10, et l'écart-type est $\sqrt{10 \cdot 0.4 \cdot (1-0.4)} = \sqrt{2.4} \approx 1.55$.


$B(10, 0.4)$
0.25
0.20
0.15
0.10
0.05
0.00
p_r
0
1
2
3
4
5
6
7
8
9
10
r

En effet, R peut s'écrire comme la somme de n variables aléatoires indépendantes:

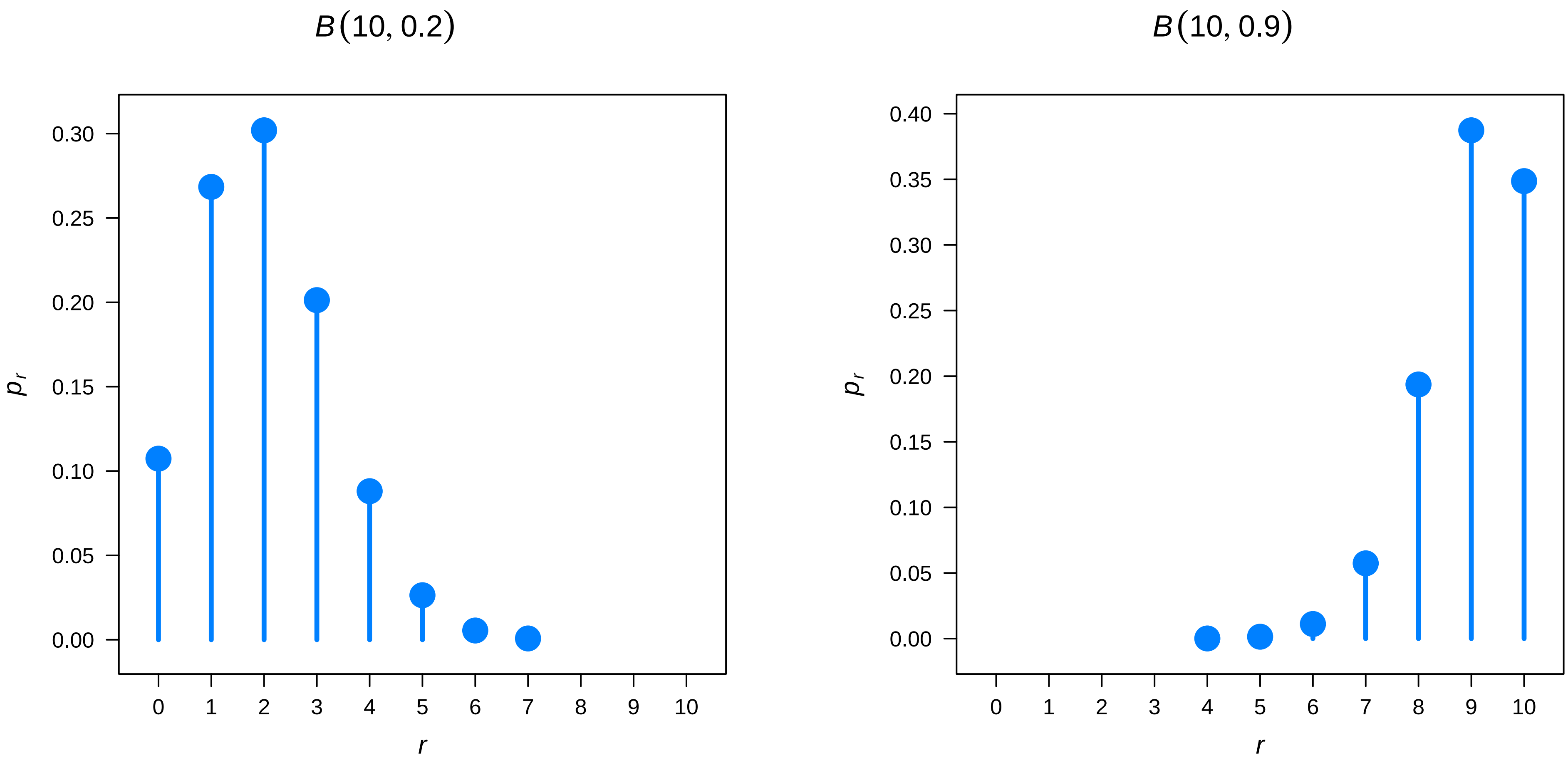
$$R = \sum_{i=1}^{n} R_i$$

où R_i, nombre de succès au i-ème tirage, suit une loi de Bernoulli.

Par conséquent, avec les propriétés de l'espérance et de la variance d'une somme de variables aléatoires indépendantes, on a:

$$\begin{aligned}
\mathsf{E}(R) \quad &= \mathsf{E}\left(\sum_{i=1}^{n} R_i\right) = \sum_{i=1}^{n} \mathsf{E}(R_i) = \sum_{i=1}^{n} p = np \\
\mathsf{var}(R) \quad &= \mathsf{var}\left(\sum_{i=1}^{n} R_i\right) = \sum_{i=1}^{n} \mathsf{var}(R_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)
\end{aligned}$$

Autres illustrations:



EXEMPLE

5% des pièces produites par une machine sont défectueuses.

1. Dans un échantillon de 10 pièces, quelle est la probabilité que le nombre de pièces défectueuses soit: (i) Moins de deux; (ii) Exactement deux; (iii) Plus de deux?
2. Dans un échantillon de 20 pièces, combien de pièces défectueuses s'attend-on à trouver?

Réponse: $R =$ nombre de pièces défectueuses.

1. $R \sim B(10, 0.05)$

$$\begin{aligned}
\text{(i) } P(R < 2) &= P(R = 0) + P(R = 1) \\
&= 0.95^{10} + \binom{10}{1} \cdot 0.05 \cdot 0.95^{9} \\
&\approx 0.5987 + 0.3151 = 0.9138
\end{aligned}$$

$$\text{(ii) } P(R=2) = \binom{10}{2} \cdot 0.05^2 \cdot 0.95^8 \approx 0.0746$$

$$\begin{aligned}\text{(iii) } P(R>2) &= 1 - P(R \leq 2) \\ &\approx 1 - 0.5987 - 0.3151 - 0.0746 \\ &= 0.0116\end{aligned}$$

2. $R \sim B(20, 0.05)$

   Nombre attendu (en moyenne), *i.e.* espérance:

$$\mathsf{E}(R) = n \cdot p = 20 \cdot 0.05 = 1$$

Lecture des probabilités dans une table:

Lorsque n est grand et l'on veut calculer $P(R > r)$ ou $P(R < r)$.

Dans les tables on peut lire $P(R = r)$ ou $P(R \leq r)$ pour des valeurs de n et p.

Exemple: $R \sim B(25, 0.2)$

$$\begin{array}{llcl}
\text{(i)} & P(R \leq 11) & \approx & 0.998 \\[1em]
\text{(ii)} & P(R \geq 12) & = & 1 - P(R \leq 11) \\
 & & \approx & 1 - 0.998 = 0.002 \\[1em]
\text{(iii)} & P(R > 7) & = & 1 - P(R \leq 7) \\
 & & \approx & 1 - 0.891 = 0.109 \\[1em]
\text{(iv)} & P(R > 29) & = & 0
\end{array}$$

EXEMPLE

Dans une région particulière il y a un taux de chômage de 5%. 20 personnes de cette région ont été tirées au hasard.

Soit $R =$ le nombre de chômeurs dans l'échantillon.

Quelle est la probabilité que:

1. Il y ait exactement 3 chômeurs dans l'échantillon?
2. Il y ait plus de 3 chômeurs dans l'échantillon?

Réponse:

- ▷ Le tirage de chaque personne est un essai indépendant (ou peut être considéré comme tel);
- ▷ Un 'succès' correspond à tirer un chômeur;
- ▷ La probabilité de succès: $p = 0.05$;
- ▷ Le nombre d'essais: $n = 20$.

Alors $R \sim B(20, 0.05)$.

1.

$$\begin{aligned} P(R=3) &= \binom{20}{3} \cdot 0.05^3 \cdot 0.95^{17} \\ &= \frac{20!}{3!17!} \cdot 0.05^3 \cdot 0.95^{17} \\ &= \frac{20 \cdot 19 \cdot 18}{3 \cdot 2} \cdot 0.05^3 \cdot 0.95^{17} \\ &\approx 0.0596 \approx 6\% \end{aligned}$$

ou à l'aide des tables:

$$\begin{aligned} P(R=3) &= P(R \leq 3) - P(R \leq 2) \\ &\approx 0.984 - 0.925 = 0.059 \end{aligned}$$

2.

$$\begin{aligned} P(R > 3) &= 1 - P(R \leq 3) \\ &= 1 - [P(R = 3) + P(R = 2) + P(R = 1) + P(R = 0)] \\ &= 1 - P(R = 3) - P(R = 2) - P(R = 1) - P(R = 0) \end{aligned}$$

$$\begin{aligned} P(R = 2) &= \binom{20}{2} \cdot 0.05^2 \cdot 0.95^{18} \\ &= \frac{20 \cdot 19}{2} \cdot 0.05^2 \cdot 0.95^{18} \approx 0.1887 \\ P(R = 1) &= \binom{20}{1} \cdot 0.05 \cdot 0.95^{19} \\ &= 20 \cdot 0.05 \cdot 0.95^{19} \approx 0.3773 \\ P(R = 0) &= 0.95^{20} \approx 0.3585 \end{aligned}$$

$$
\begin{aligned}
P(R > 3) &\approx 1 - 0.0596 - 0.1887 - 0.3773 - 0.3585 \\
&= 0.0159 \approx 1.6\%
\end{aligned}
$$

ou à l'aide des tables:

$$
\begin{aligned}
P(R > 3) &= 1 - P(R \leq 3) \\
&\approx 1 - 0.984 = 0.016 = 1.6\%
\end{aligned}
$$


ENRICO 12

3.5 La distribution de Poisson

R	$=$	Nombre d'événements arrivant au hasard dans une unité particulière d'une mesure continue (*e.g.* temps, distance) ou d'un très grand nombre d'unités.
R	$=$	**Variable aléatoire de Poisson**

Soit λ le nombre moyen d'événements par unité.

Notation: $\boxed{R \sim \text{Poisson}(\lambda)}$

Exemples:

- ▷ Arrêts d'une machine au cours du temps;
- ▷ Trous le long d'une route;
- ▷ Nombre de pièces défectueuses dans de très grands échantillons;
- ▷ Faux numéros téléphoniques composés en 1 jour;
- ▷ Clients pénétrant dans une banque donnée en l'espace d'une semaine.

Espace des réalisations: $S = \{0, 1, 2, 3, 4, 5, \ldots\}$

Distribution de probabilité: $R \sim \text{Poisson}(\lambda)$

$$\boxed{p_r = P(R = r) = \exp(-\lambda)\frac{\lambda^r}{r!} = e^{-\lambda}\frac{\lambda^r}{r!}}$$

pour $r = 0, 1, 2, \ldots, \infty$.

On a bien

$$\sum_{r=0}^{\infty} p_r = \exp(-\lambda) \cdot \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} = \exp(-\lambda) \cdot \exp(\lambda) = 1.$$

Espérance:

$$\begin{aligned}
\mathsf{E}(R) &= \sum_{r=0}^{\infty} r \cdot \exp(-\lambda) \frac{\lambda^r}{r!} = \sum_{r=1}^{\infty} r \cdot \exp(-\lambda) \frac{\lambda^r}{r!} \\
&= \lambda \exp(-\lambda) \sum_{r=1}^{\infty} \frac{\lambda^{r-1}}{(r-1)!} = \underbrace{\lambda \exp(-\lambda) \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{=1} = \lambda
\end{aligned}$$

Variance:

$$\mathsf{var}(R) = \lambda$$

L'espérance et la variance d'une variable aléatoire de Poisson sont donc toutes les deux égales à son paramètre λ.

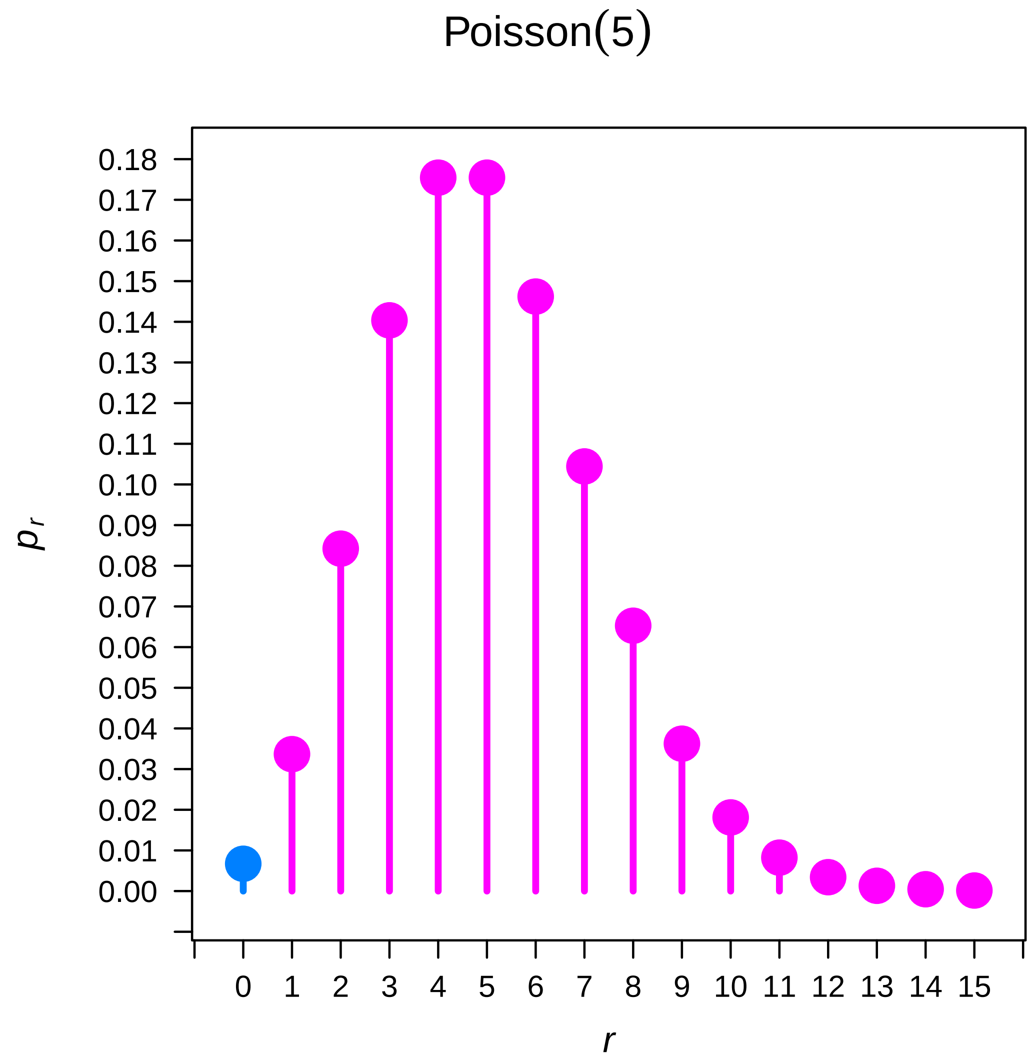
EXEMPLE

Le nombre moyen de journaux vendus par Alfred est de 5 par minute. Quelle est la probabilité qu'Alfred en vende au moins 1 en une minute?

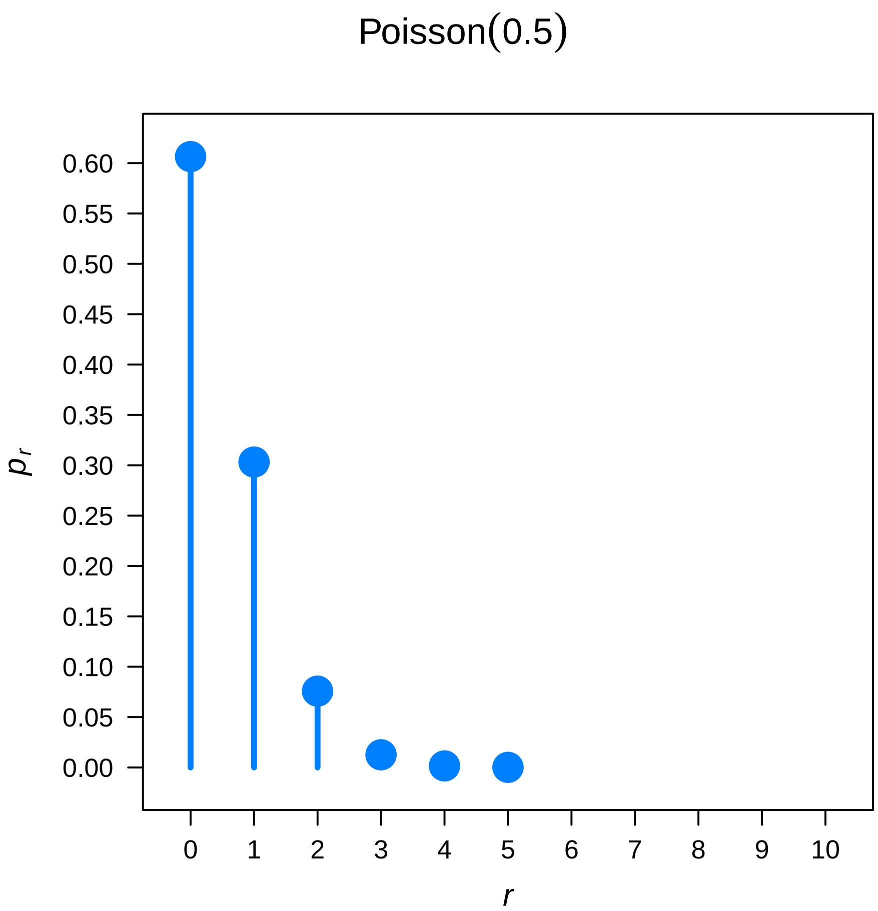
Réponse: Soit R le nombre de journaux vendus par Alfred en une minute. On a $R \sim$ Poisson$(\lambda = 5)$ et donc

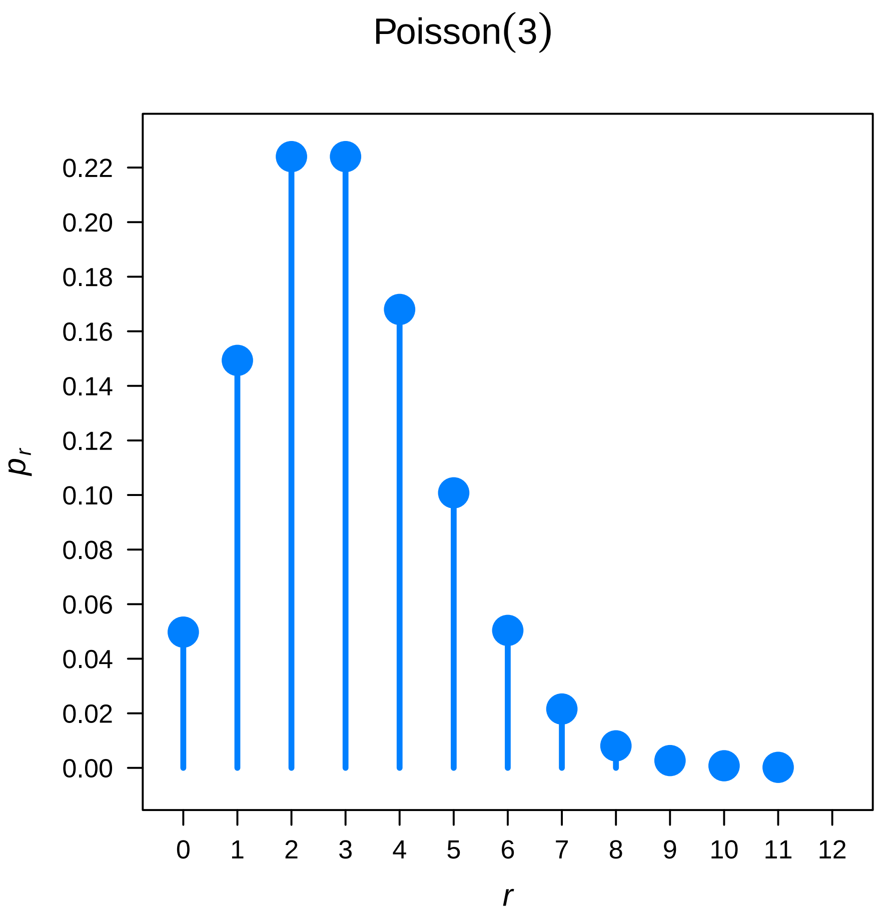
$$\begin{aligned} P(R \geq 1) &= 1 - P(R = 0) \\ &= 1 - \exp(-5)\frac{5^0}{0!} \\ &= 1 - \exp(-5) = 1 - e^{-5} \\ &\approx 1 - 0.0067 = 0.9933. \end{aligned}$$

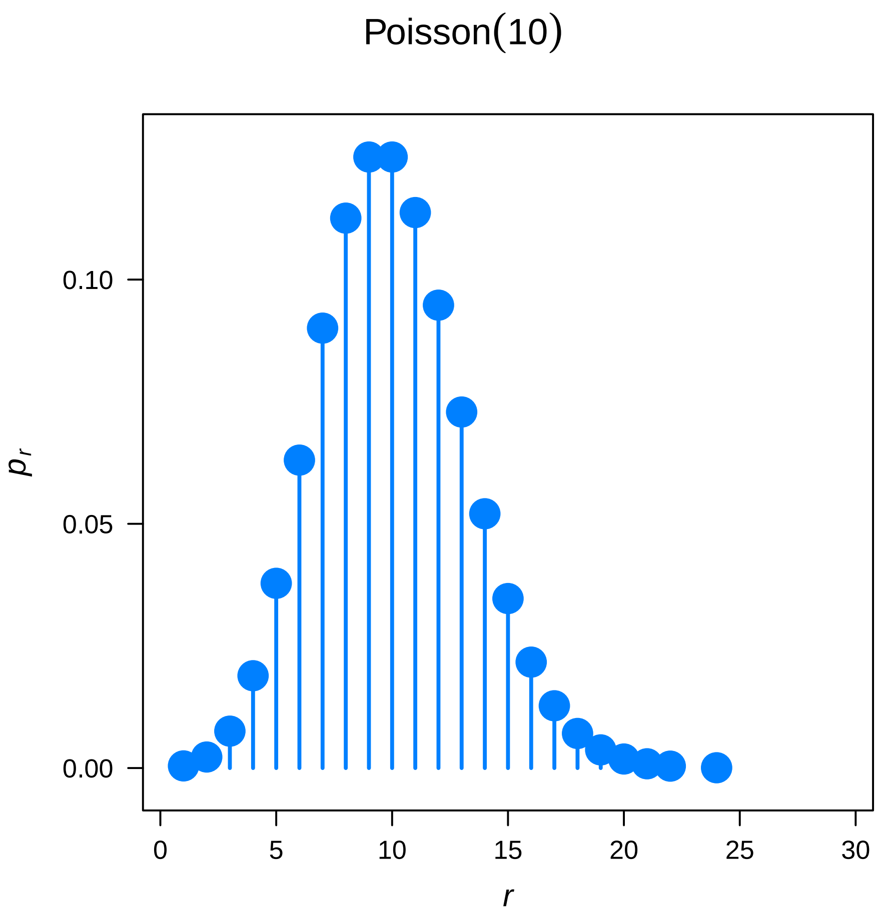
Illustration:

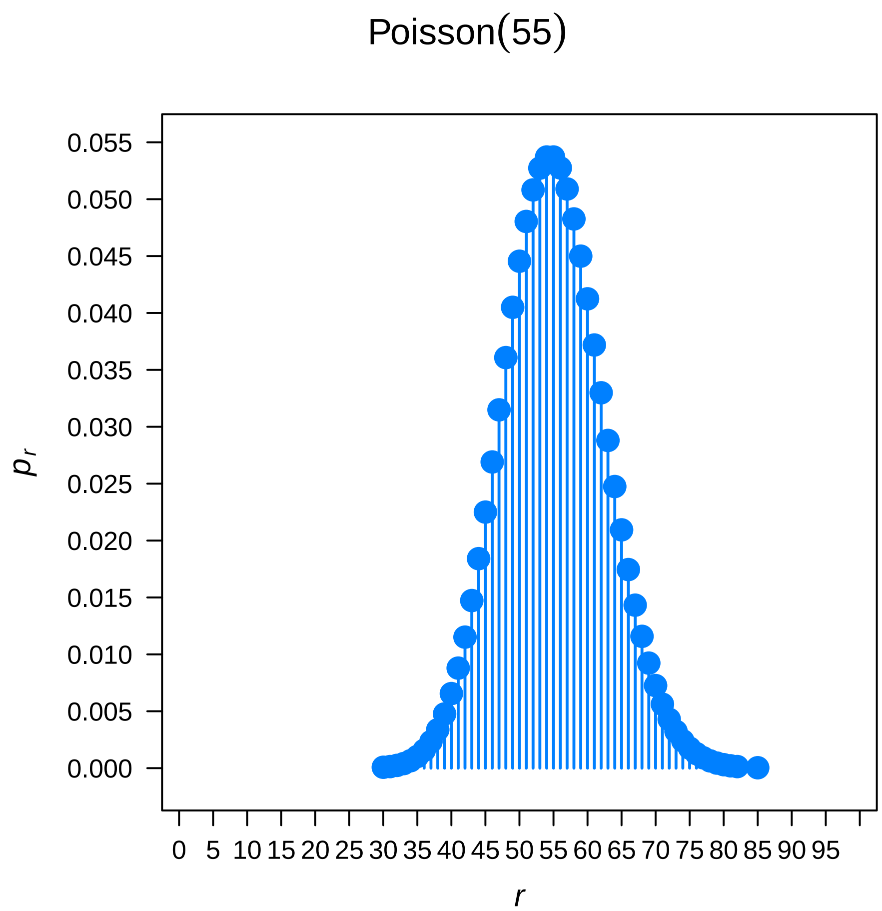


Autres illustrations de distributions de Poisson:






Poisson(10)
0.10
0.05
0.00
p_r
0
5
10
15
20
25
30
r


Poisson(55)
0.055
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
p_r
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
r

EXEMPLE

Une machine qui fonctionne du lundi au vendredi compris tombe en panne une fois par semaine en moyenne.

Quelle est la probabilité que la machine ne tombe pas en panne cette semaine?

Réponse: Soit R le nombre de pannes par semaine.

On a que $R \sim \text{Poisson}(\lambda)$ avec $\lambda = 1$.

Donc $P(R = 0) = \exp(-1)\dfrac{1^0}{0!} \approx 0.37$.

Somme de variables de Poisson

Soient $R_1 = $ le nombre d'événements dans l'intervalle I_1

$R_2 = $ le nombre d'événements dans l'intervalle I_2

tel que $I_1 \cap I_2 = \emptyset$

$R_1 \sim \text{Poisson}(\lambda_1)$

$R_2 \sim \text{Poisson}(\lambda_2)$

alors $R_1 + R_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$

dans $I_1 \cup I_2$.

EXEMPLE

Le nombre moyen de clients par heure arrivant dans un magasin est de 3 entre 8h et 10h et de 5 entre 10h et 12h.

Quelle est la probabilité que 10 clients exactement arrivent au magasin en une matinée?

Réponse:

$$\begin{aligned} R_1 &= \text{Nombre de clients arrivant entre 8h et 10h} \\ R_1 &\sim \text{Poisson}(3+3), \textit{ i.e. } \text{Poisson}(6) \end{aligned}$$

$$\begin{aligned} R_2 &= \text{Nombre de clients arrivant entre 10h et 12h} \\ R_2 &\sim \text{Poisson}(5+5), \textit{ i.e. } \text{Poisson}(10) \end{aligned}$$

$$\begin{aligned} T &= \text{Nombre de clients arrivant entre 8h et 12h} \\ T &= R_1 + R_2 \\ T &\sim \text{Poisson}(6+10), \textit{ i.e. } \text{Poisson}(16) \end{aligned}$$

Donc

$$P(T=10) = \exp(-16)\frac{16^{10}}{10!} \approx 0.0341.$$

Approximation de la binomiale:

Soit $R \sim B(n,p)$

et 1) n est grand;

2) p est petit.

Alors on a approximativement

$$R \sim \text{Poisson}(\lambda = np).$$

EXEMPLE

Un échantillon contient 200 ampoules. On sait qu'une sur 50 ne fonctionne pas.

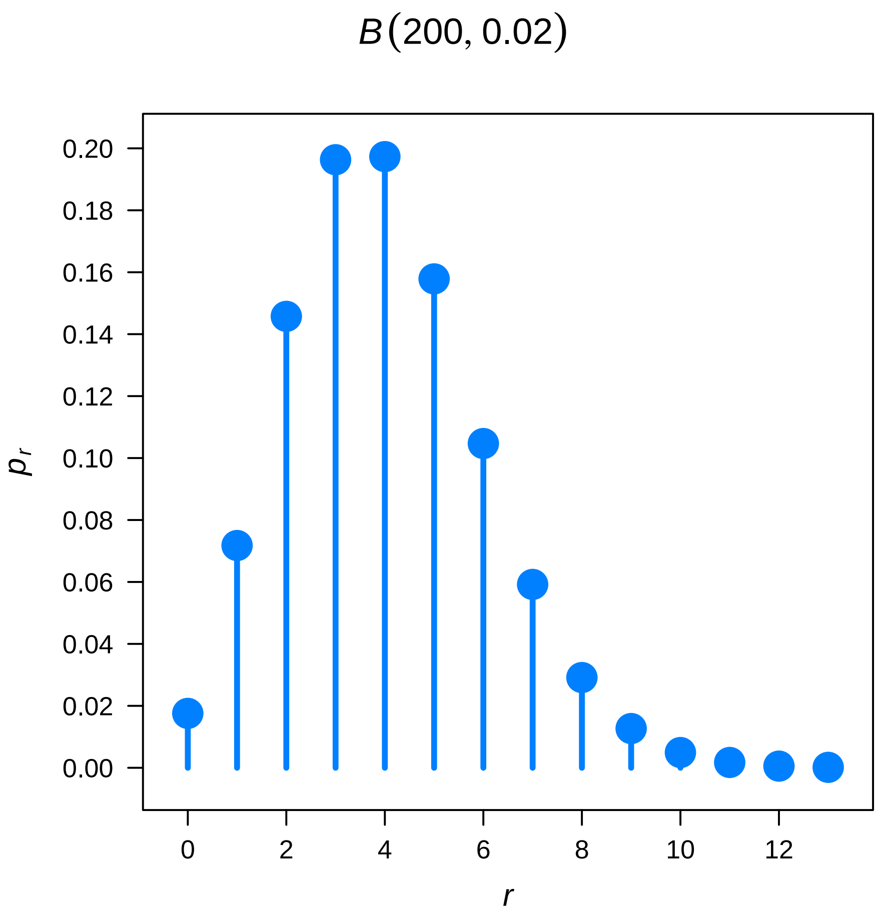
Quelle est la distribution de $R =$ nombre d'ampoules défectueuses dans l'échantillon?

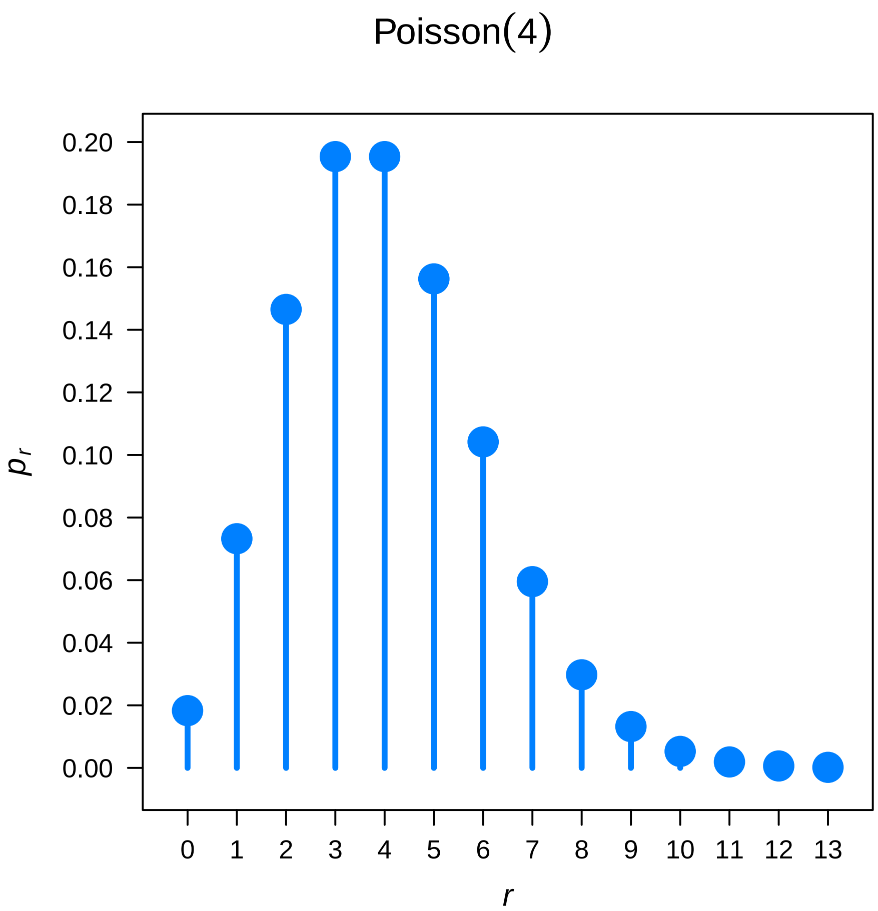
Réponse: On a $R \sim B(200, 0.02)$, mais n est grand et p est petit.

Donc $R \sim$ Poisson(λ) avec $\lambda = 200 \cdot 0.02 = 4$.

Remarque: La distribution de Poisson est plus facile à utiliser.

Illustration:





Lecture des probabilités dans une table:

Comment trouver $P(R > r)$ lorsque r est grand?

Dans les tables on peut lire $P(R \leq r)$ ou $P(R = r)$ pour un λ (ou 'α') donné.

Exemple: $R \sim$ Poisson(5)

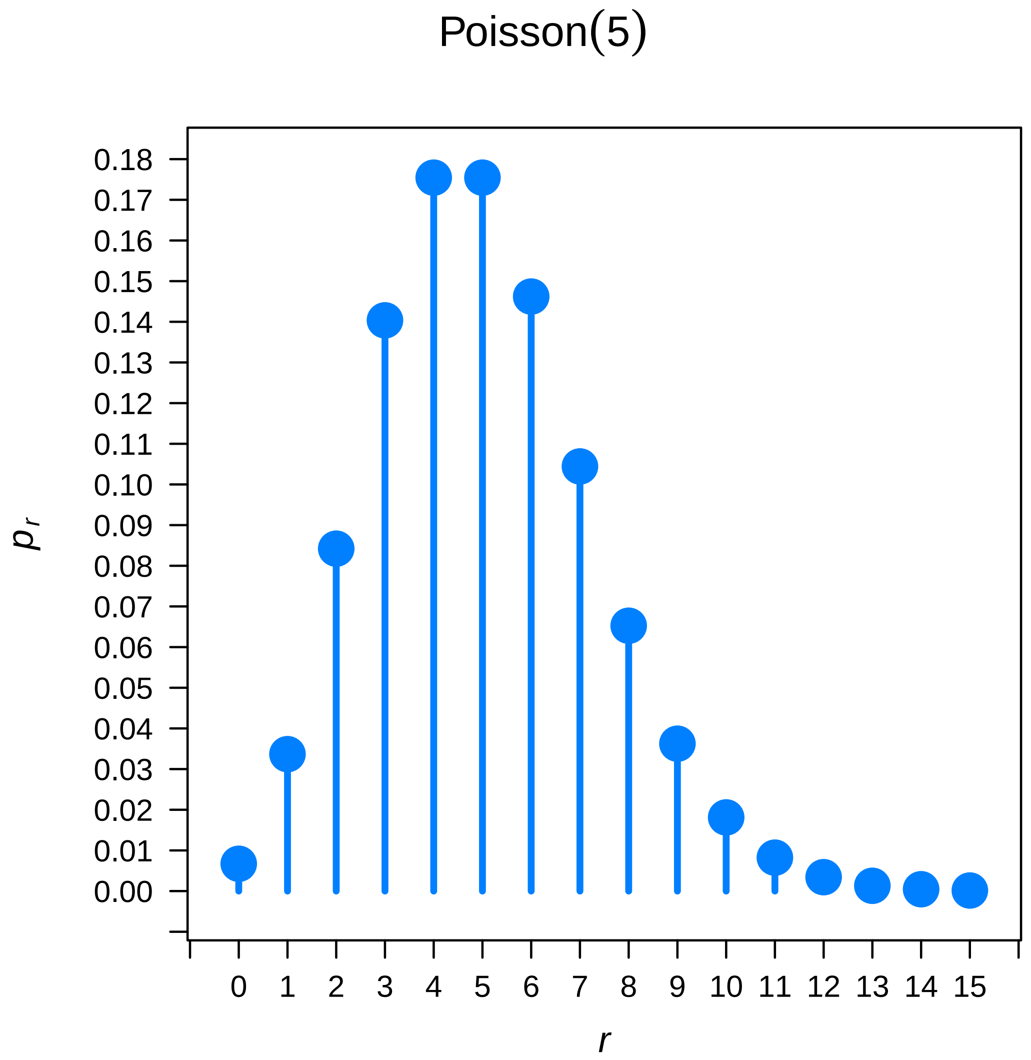
(i) $$P(R \geq 15) = 1 - P(R \leq 14) \approx 1 - 0.9998 = 0.0002$$

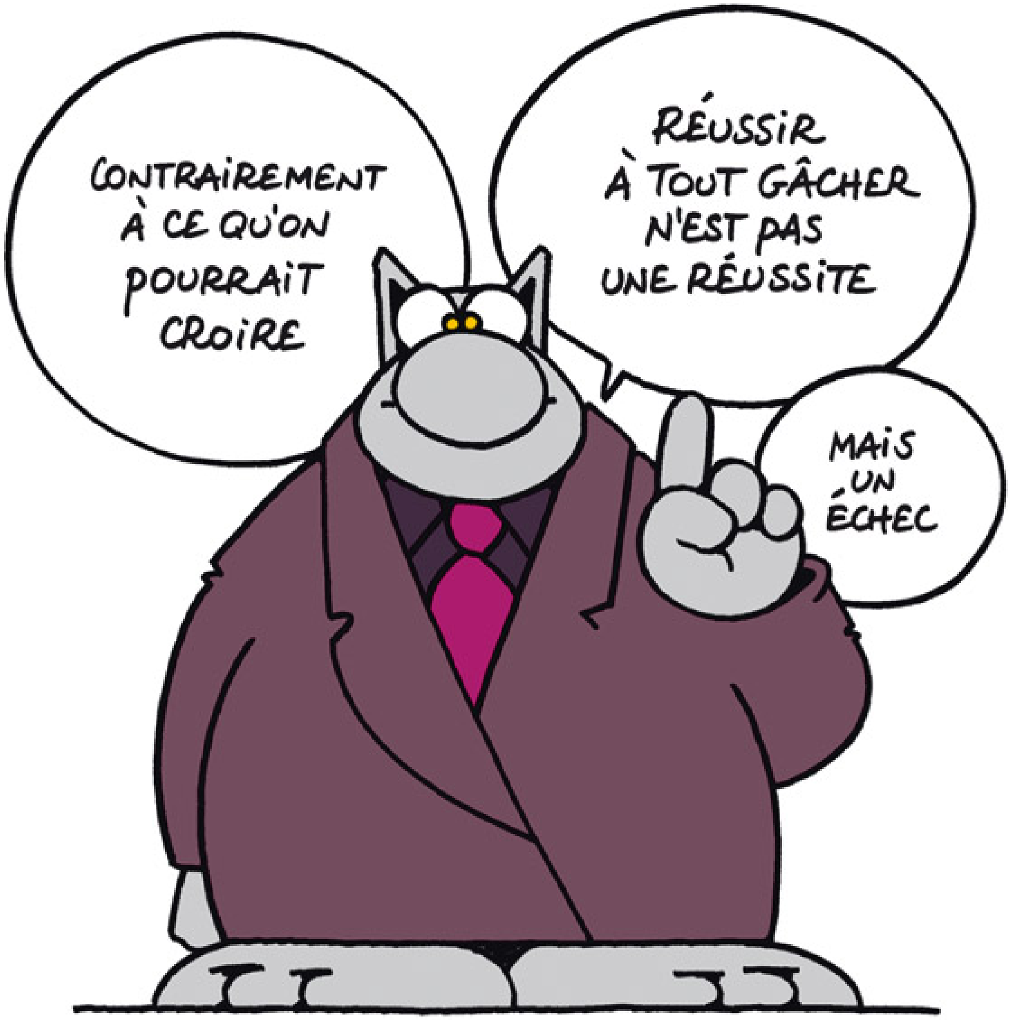
(ii) $$P(R > 3) = 1 - P(R \leq 3) \approx 1 - 0.2650 = 0.7350$$

(iii) $$P(R > 30) = 1 - P(R \leq 30) \approx 1 - 1.0000 = 0.0000$$

(iv) $$P(R \leq 9) \approx 0.9682$$

Illustration:




CONTRAIREMENT À CE QU'ON POURRAIT CROIRE
RÉUSSIR À TOUT GÂCHER N'EST PAS UNE RÉUSSITE
MAIS UN ÉCHEC

3.6 La distribution géometrique

On exécute une série d'épreuves indépendantes (*i.e.* 'avec remise') ayant chacune la probabilité p d'être une succès, $0 < p < 1$, jusqu'à obtenir le premier succès.

Le **nombre d'epreuves nécessaires jusqu'à au premier succès** est appelé une variable **aléatoire géométrique**.

Notation: R est une variable aléatoire géométrique:

$$\boxed{R \sim \text{Geom}(p)}$$

Espace des réalisations: $S = \{1, 2, 3, 4, 5, \ldots\}$

Distribution de probabilité:

$$\boxed{p_r = P(R = r) = (1-p)^{r-1}p}$$

pour $r = 1, 2, \ldots, \infty$.

En effet, pour que R prenne r pour valeur, il faut et suffit que les $r-1$ premières épreuves soient des échecs tandis que la r-ième devra être un succès.

Espérance et variance: $\mathrm{E}(R) = \dfrac{1}{p}$ et $\mathrm{var}(R) = \dfrac{1-p}{p^2}$.

EXEMPLE

On considère une roue de roulette comprenant 38 cases numérotées 0, 00 et de 1 à 36. Albert parie régulièrement sur la sortie des numéros 1 à 12.

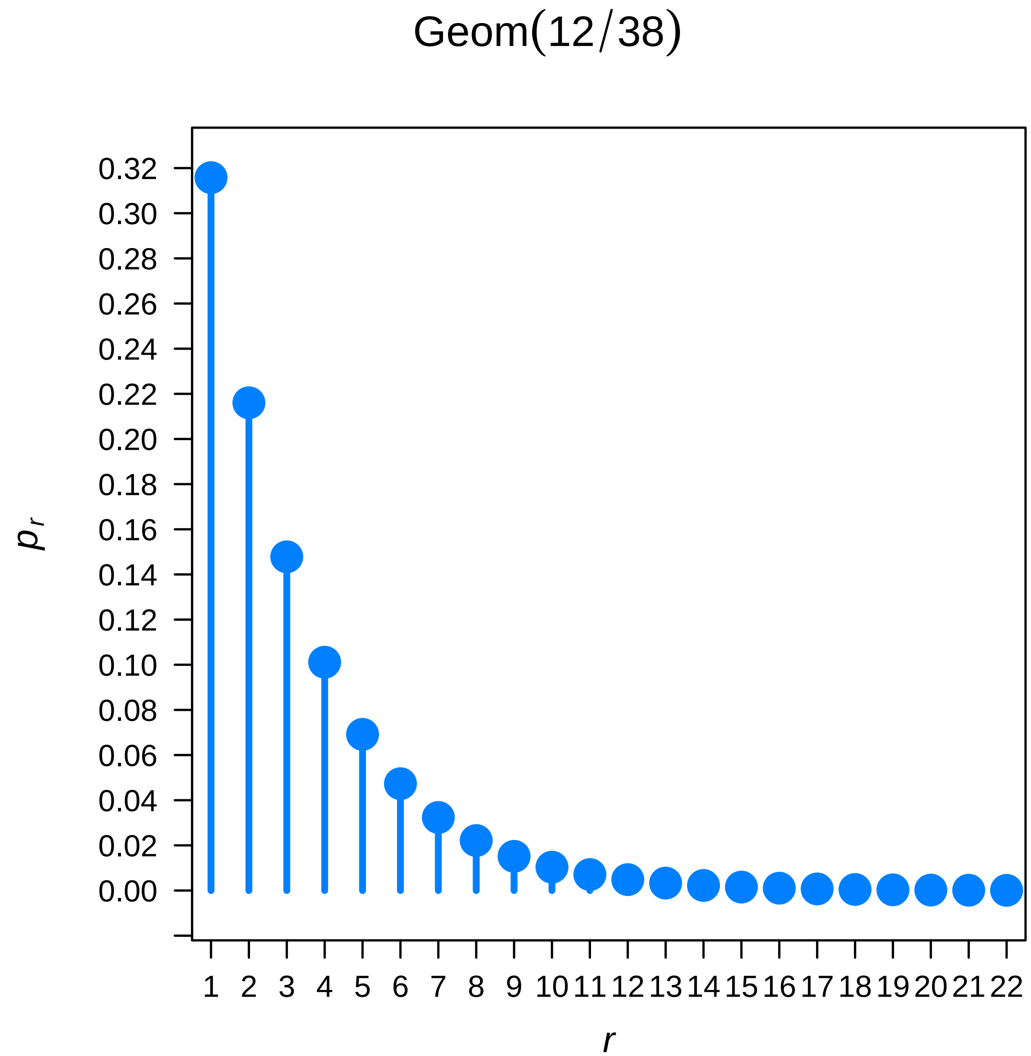
Quelle est la probabilité que son premier gain survienne lors du 4-ième tirage?

Réponse: Soit $R =$ le nombre de tirages nécessaires jusqu'à au premier gain, *i.e.* succès, avec une probabilité $p = 12/(36+2) = 12/38$ d'être une succès.

Alors $R \sim \text{Geom}(12/38)$. Donc

$$P(R=4) = (1-p)^{4-1}p = (1-12/38)^3 12/38 \approx 0.1012.$$

Illustration:



‘Le loto, c’est un impôt sur les gens qui ne comprennent pas le calcul de probabilités.’

Anonyme

3.7 La distribution binomiale négative

On exécute une série d'épreuves indépendantes (*i.e.* 'avec remise') ayant chacune une probabilité p de donner un succès, $0 < p < 1$, jusqu'à obtenir un total de s succès.

Le **nombre d'epreuves nécessaires pour atteindre s succès** est appelé une variable **aléatoire binomiale négative**.

Notation: R est une variable aléatoire binomiale négative:

$$\boxed{R \sim \mathrm{NB}(s, p)}$$

Espace des réalisations: $S = \{s, s+1, \ldots\}$

Distribution de probabilité:

$$\boxed{p_r = P(R = r) = \binom{r-1}{s-1} p^s (1-p)^{r-s}}$$

pour $r = s, s+1, \ldots, \infty$.

En effet, pour obtenir un s-ième succès lors de la r-ième épreuve il a fallu $s-1$ succès lors des $r-1$ premières épreuves et il faut que la r-ième épreuve soit un succès.

Espérance et variance: $\mathrm{E}(R) = \dfrac{s}{p}$ et $\mathrm{var}(R) = \dfrac{s(1-p)}{p^2}$.

Remarque: Une variable géométrique $\mathrm{Geom}(p)$ est une $\mathrm{NB}(1,p)$, *i.e.* avec $s = 1$.

EXEMPLE

Calculer l'espérance et la variance du nombre de jets d'un dé nécessaires pour obtenir 4 fois la valeur 1.

Réponse: Soit R la variable aléatoire étudiée. Alors $R \sim \mathsf{NB}(s, p)$ avec $s = 4$ et $p = 1/6$, *i.e.* $R \sim \mathsf{NB}(4, 1/6)$.

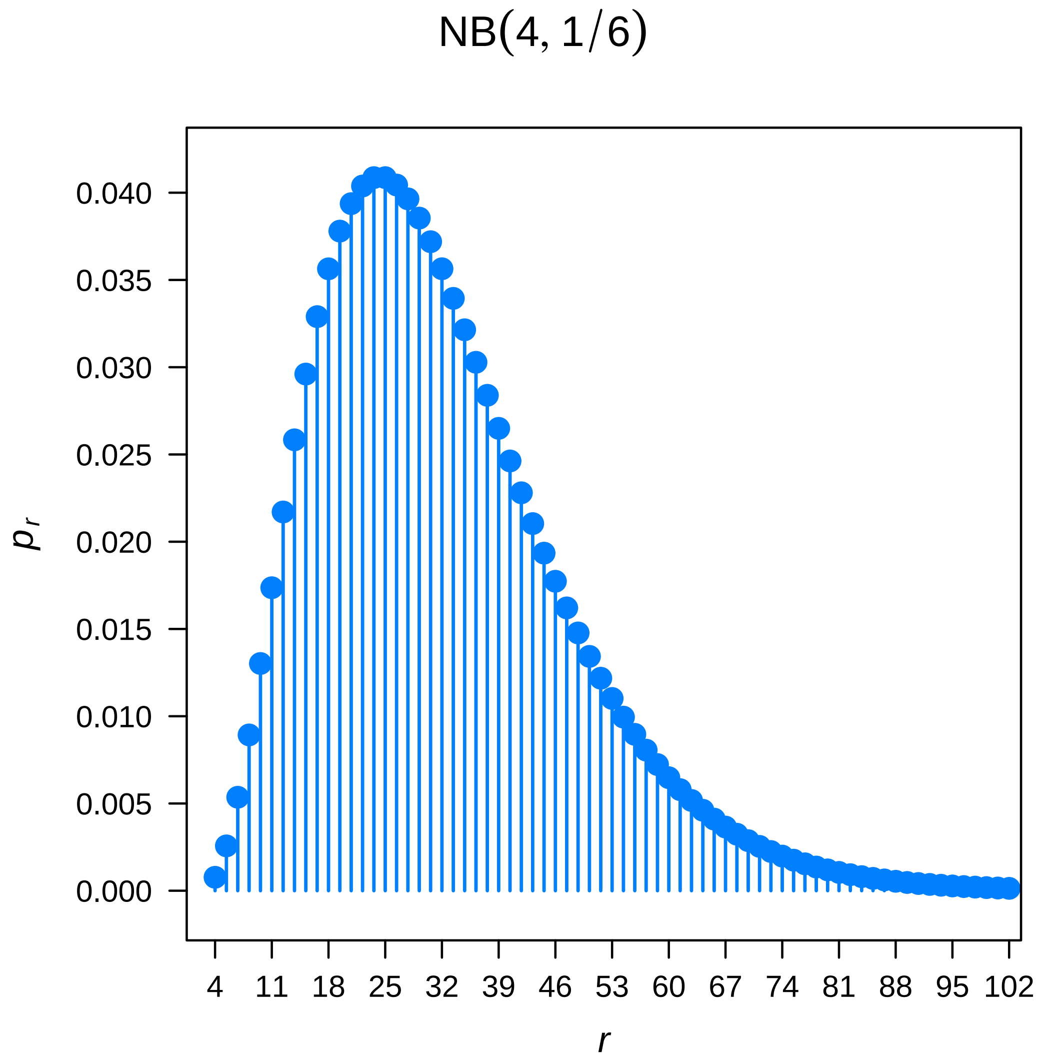
Donc

$$\mathsf{E}(R) = \frac{s}{p} = \frac{4}{1/6} = 24$$

et

$$\mathsf{var}(X) = \frac{s(1-p)}{p^2} = \frac{4 \cdot \frac{5}{6}}{\left(\frac{1}{6}\right)^2} = 120.$$

Illustration:



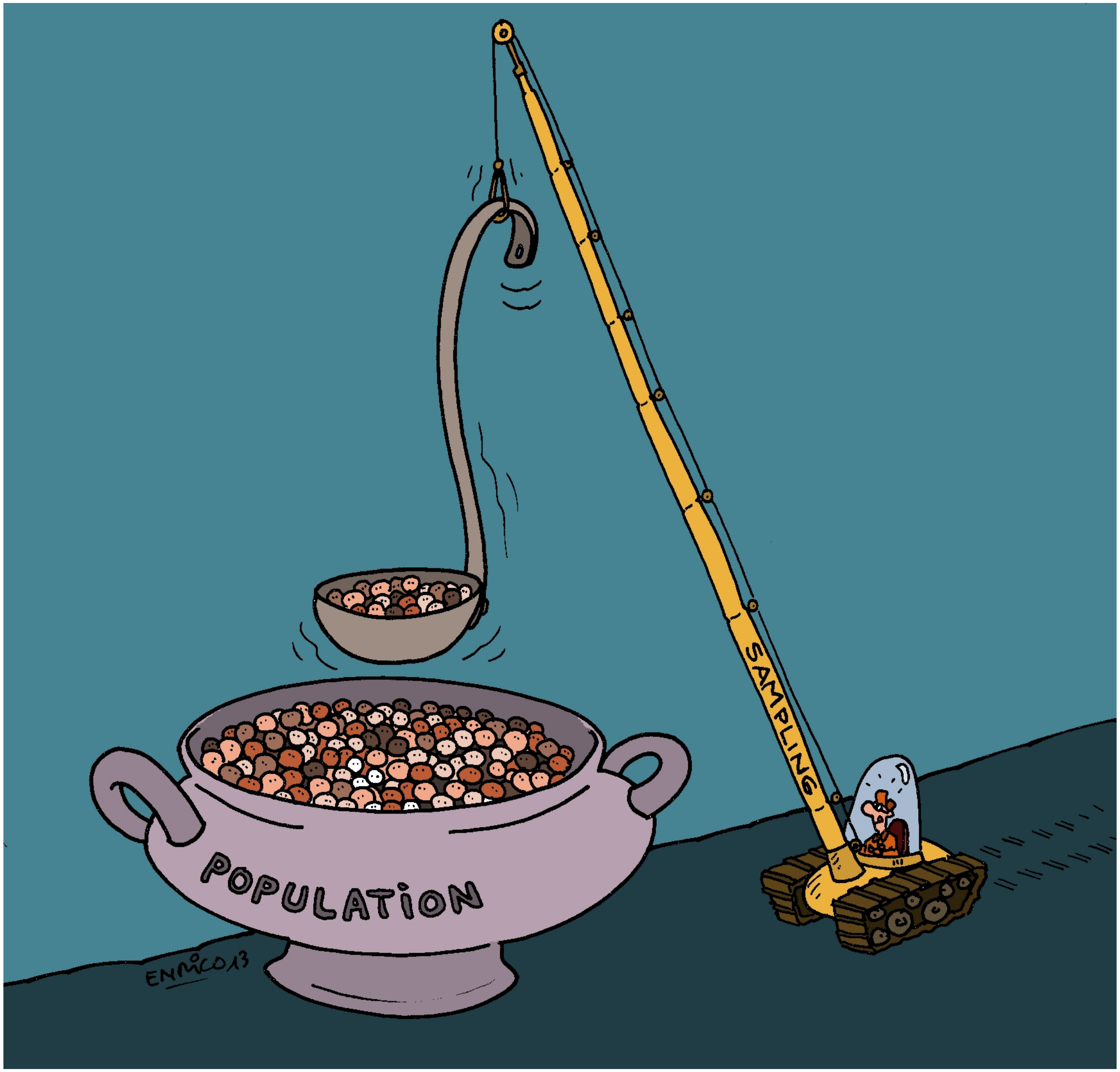
EXEMPLE

Quelle est la probabilité qu'un téléopérateur doive passer exactement 20 appels pour trouver 3 nouveaux clients, sachant que la probabilité qu'une personne sollicitée devienne cliente est de 0.119?

Réponse: Soit R le nombre d'appels pour trouver 3 nouveaux clients. Alors $R \sim \text{NB}(s, p)$ avec $s = 3$ et $p = 0.119$, *i.e.* $R \sim \text{NB}(3, 0.119)$.

Donc

$$P(R = 20) = \binom{20-1}{3-1} 0.119^3 (1-0.119)^{20-3} = \binom{19}{2} 0.119^3 (1-0.119)^{17} \approx 0.0334.$$


SAMPLING
POPULATION
ENRICO13

3.8 La distribution hypergéometrique

On tire <u>sans remise</u> un échantillon de n boules d'une urne en contenant N, dont m blanches ('succès') et $N - m$ sont noires.

Désignons par R le nombre de boules blanches tirées ('succès').

On aura

$$P(R = r) = \frac{\binom{m}{r}\binom{N-m}{n-r}}{\binom{N}{n}}$$

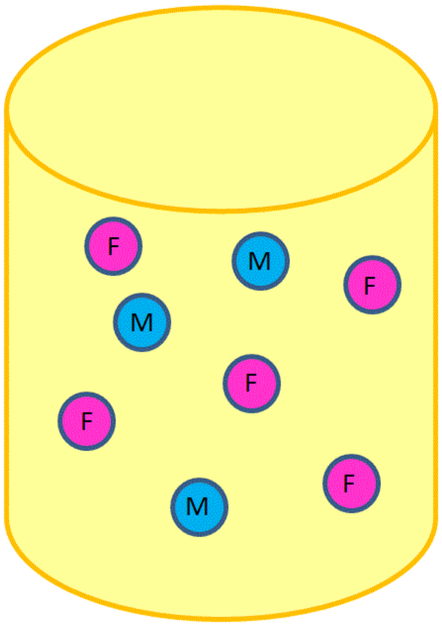
pour $r = 0, 1, \ldots, n$.

S'il existe certaines valeurs de n, N et m pour lesquelles la loi d'une variable aléatoire vérifie ceci, cette variable est dite **variable aléatoire hypergéometrique**.

Espérance et variance:

$$\mathsf{E}(R) = \frac{nm}{N} \quad \text{et} \quad \mathsf{var}(R) = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right].$$

Exemple: Un groupe de 8 étudiants contient 5 femmes et 3 hommes. Pour participer dans une étude psychologique 3 étudiants sont choisis aléatoirement. Quelle est la probabilité qu'il ait exactement 2 femmes?

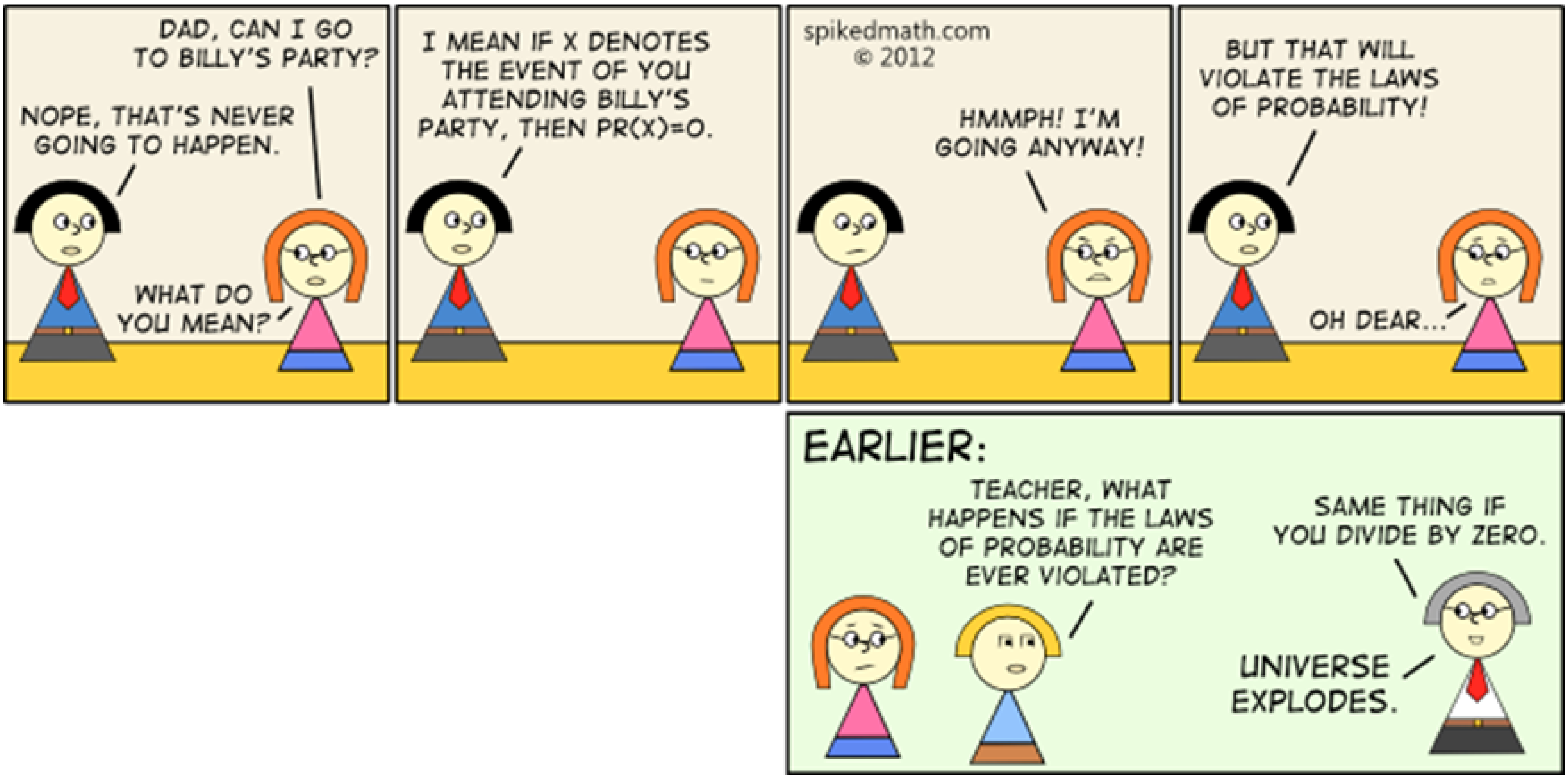


Réponse: Il y a $n = 3$ tirages et considérons une femme choisie lors d'un tirage comme un succès.

Ainsi on a $N = 8$, $m = 5$ et $r = 2$. Donc

$$\begin{aligned} P(R=2) &= \frac{\binom{5}{2}\binom{8-5}{3-2}}{\binom{8}{3}} \\ &= \frac{\frac{5!}{2!3!} \cdot \frac{3!}{1!2!}}{\frac{8!}{5!3!}} = 0.53571. \end{aligned}$$

Autre exemple: Nombre d'étudiants ayant leur anniversaire ce mois-ci parmi un échantillon de 10 personnes, pris sur une classe de 30 étudiants.


DAD, CAN I GO TO BILLY'S PARTY?
NOPE, THAT'S NEVER GOING TO HAPPEN.
WHAT DO YOU MEAN?
I MEAN IF X DENOTES THE EVENT OF YOU ATTENDING BILLY'S PARTY, THEN PR(X)=0.
spikedmath.com © 2012
HMMPH! I'M GOING ANYWAY!
BUT THAT WILL VIOLATE THE LAWS OF PROBABILITY!
OH DEAR...
EARLIER:
TEACHER, WHAT HAPPENS IF THE LAWS OF PROBABILITY ARE EVER VIOLATED?
UNIVERSE EXPLODES.
SAME THING IF YOU DIVIDE BY ZERO.

4. Distributions de probabilité continues

4.1 Définitions

• On appelle **variable aléatoire continue** une variable aléatoire qui peut prendre toutes les valeurs d'un intervalle (*i.e.* un intervalle borné, une demi-droite ou $\mathbb{R}$ tout entier).

$\rightsquigarrow$ Pour une variable aléatoire continue X la **fonction de densité** (de probabilité) f_X de X est définie par

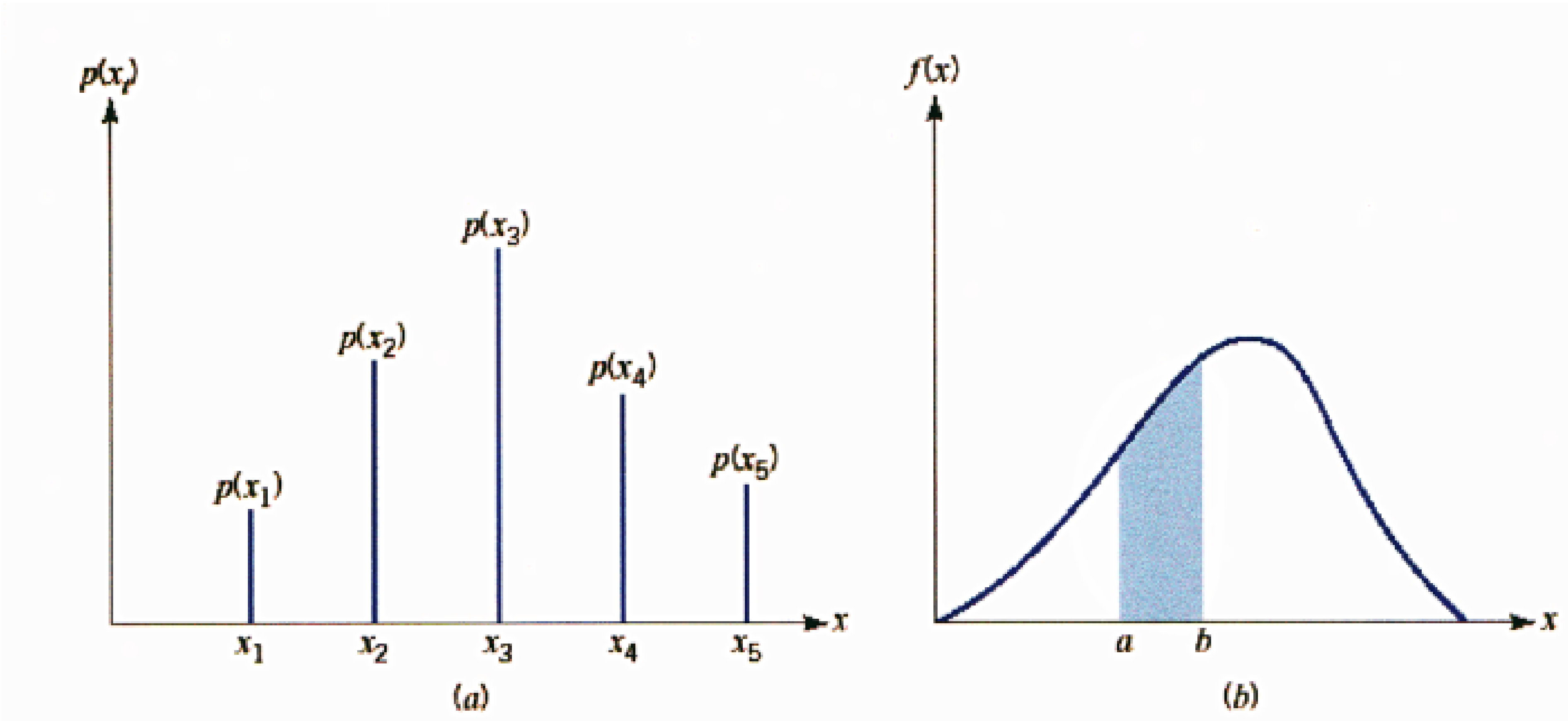
$$P(X \in A) = \int_A f_X(x)\,dx$$

où A est un ensemble de nombres réels (souvent un intervalle). Si aucune ambiguïté n'est à craindre, f_X est souvent notée f.

Une **distribution de probabilité continue**
est une fonction qui associe une
probabilité à un intervalle de valeurs.

$\rightsquigarrow$ La fonction de densité nous indique de quelle manière la masse de probabilité est distribuée sur un intervalle (fini ou infini).

$\rightsquigarrow$ Pour cette raison, on peut la considérer comme étant la notion correspondante dans le cas continu de la distribution de probabilité.



Histogramme d'une **distribution de probabilité discrète** *(gauche) avec* $p(x_r) = p_r,\ r = 1, \ldots, 5$, *et* **fonction de densité** *(droite).*

▷ **Propriétés** de la fonction de densité f_X:

i) $P(X \in]-\infty, \infty[) = \int_{-\infty}^{\infty} f_X(x)\, dx = 1$;

ii) $P(a < X \leq b) = \int_a^b f_X(x)\, dx$;

iii) $f_X(x) \geq 0$, $\forall x \in \mathbb{R}$.

⇝ Par conséquent, $P(X = a) = 0$, *i.e.* **tous les points ont une probabilité nulle**!

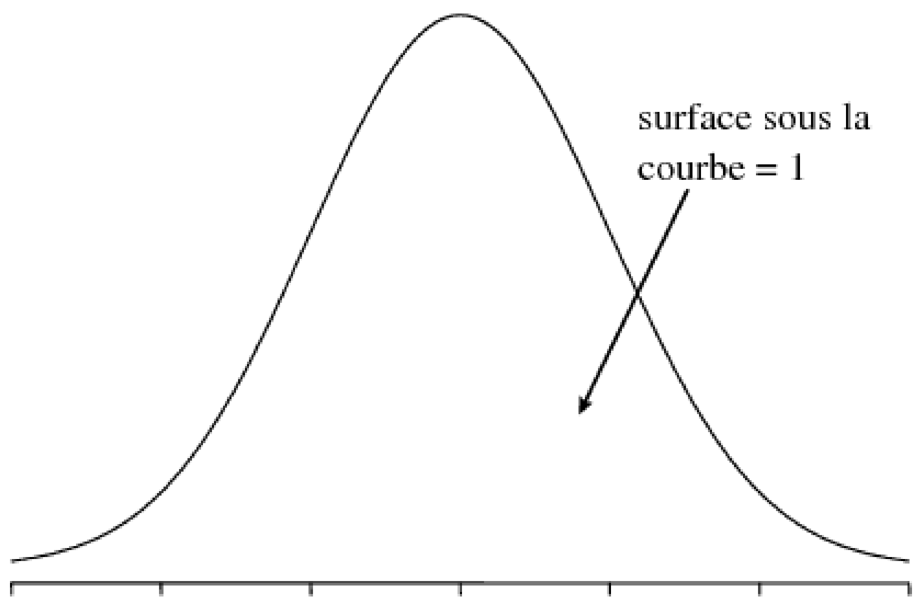


Illustration de la propriété i) d'une fonction de densité.

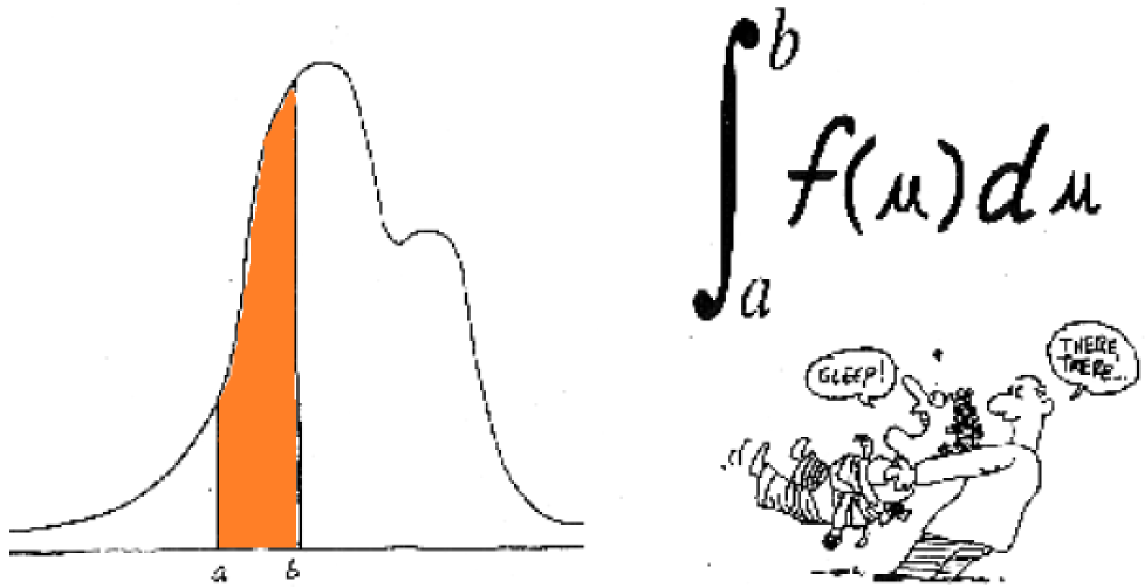


Illustration de la propriété ii) d'une fonction de densité.

Rappels du calcul d'intégrales:

$$\begin{aligned}\int_a^b f(x)dx &= F(x)\Big|_{x=a}^{x=b} \\ &= F(b) - F(a)\end{aligned}$$

où $F(x)$ est une primitive de $f(x)$.

Quelques primitives utiles:

$f(x)$	$F(x)$
a	ax
x^a	$\frac{1}{a+1}x^{a+1}$
e^x	e^x
e^{ax}	$\frac{1}{a}e^{ax}$
$\dfrac{1}{x}$	$\ln\|x\|$

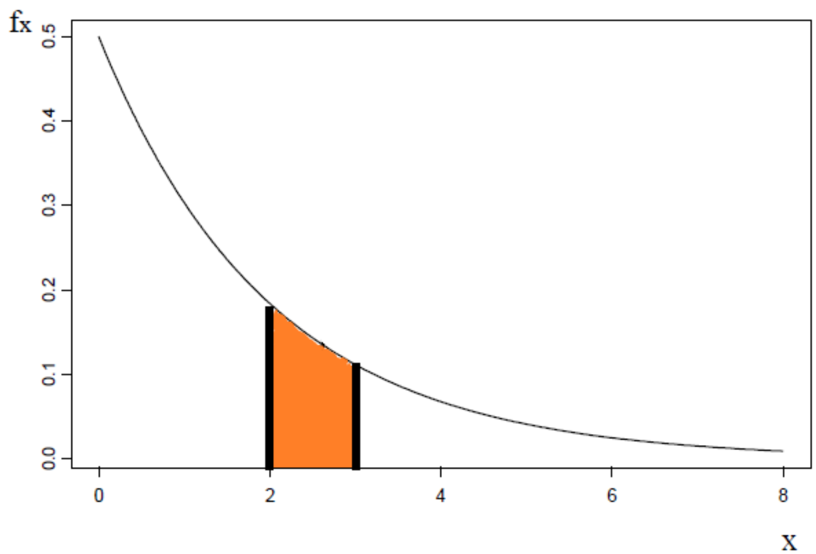
Remarque: $e^x = \exp(x)$.

EXEMPLE

Le temps d'attente X en minutes à un guichet a une fonction de densité (dite 'exponentielle'; voir plus tard)

$$f_X(x) = \begin{cases} 0.5\exp(-0.5x) & x > 0 \\ 0 & \text{sinon} \end{cases}$$

Quelle est la probabilité d'attendre entre 2 et 3 minutes?



Réponse:

$$
\begin{aligned}
P(2 < X \leq 3) &= \int_2^3 f_X(x)dx \\
&= \int_2^3 0.5\exp(-0.5x)dx \\
&= 0.5\ \int_2^3 \exp(-0.5x)dx \\
&= 0.5\left[-2\exp(-0.5x)\right]\Big|_{x=2}^{x=3} \\
&= -\exp(-1.5) + \exp(-1) \\
&\approx 0.1447
\end{aligned}
$$

EXEMPLE

X a pour densité $f_X(x) = c \cdot (1 - x^2)$ sur l'intervalle $[-1, 1]$ et 0 ailleurs.

Quelle est la valeur de c?

Réponse: Il faut que $\int_{-\infty}^{\infty} f_X(x)\, dx = \int_{-1}^{1} f_X(x) dx = 1$.

$$
\begin{aligned}
\int_{-1}^{1} c \cdot (1 - x^2) dx &= c \cdot \int_{-1}^{1} (1 - x^2) dx \\
&= c \cdot \Big(\int_{-1}^{1} 1 dx - \int_{-1}^{1} x^2 dx \Big) \\
&= c \cdot \big(x - x^3/3 \big) \Big|_{x=-1}^{x=1} \\
&= c \cdot [1 - 1/3 - (-1 + 1/3)] \\
&= c \cdot 4/3
\end{aligned}
$$

Donc

$$c \cdot 4/3 = 1 \quad \Rightarrow \quad c = 3/4$$

et

$$f_X(x) = \frac{3}{4}(1 - x^2),$$

pour $x \in [-1, 1]$ et 0 ailleurs.

• La **fonction de répartition** d'une variable aléatoire continue X est analogue à celle d'une variable aléatoire discrète.

$\rightsquigarrow$ La fonction de répartition F_X d'une variable aléatoire continue X est définie par

$$\boxed{F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)\, du.}$$

$\rightsquigarrow$ C'est une fonction de probabilités cumulées.

▷ **Propriétés:**

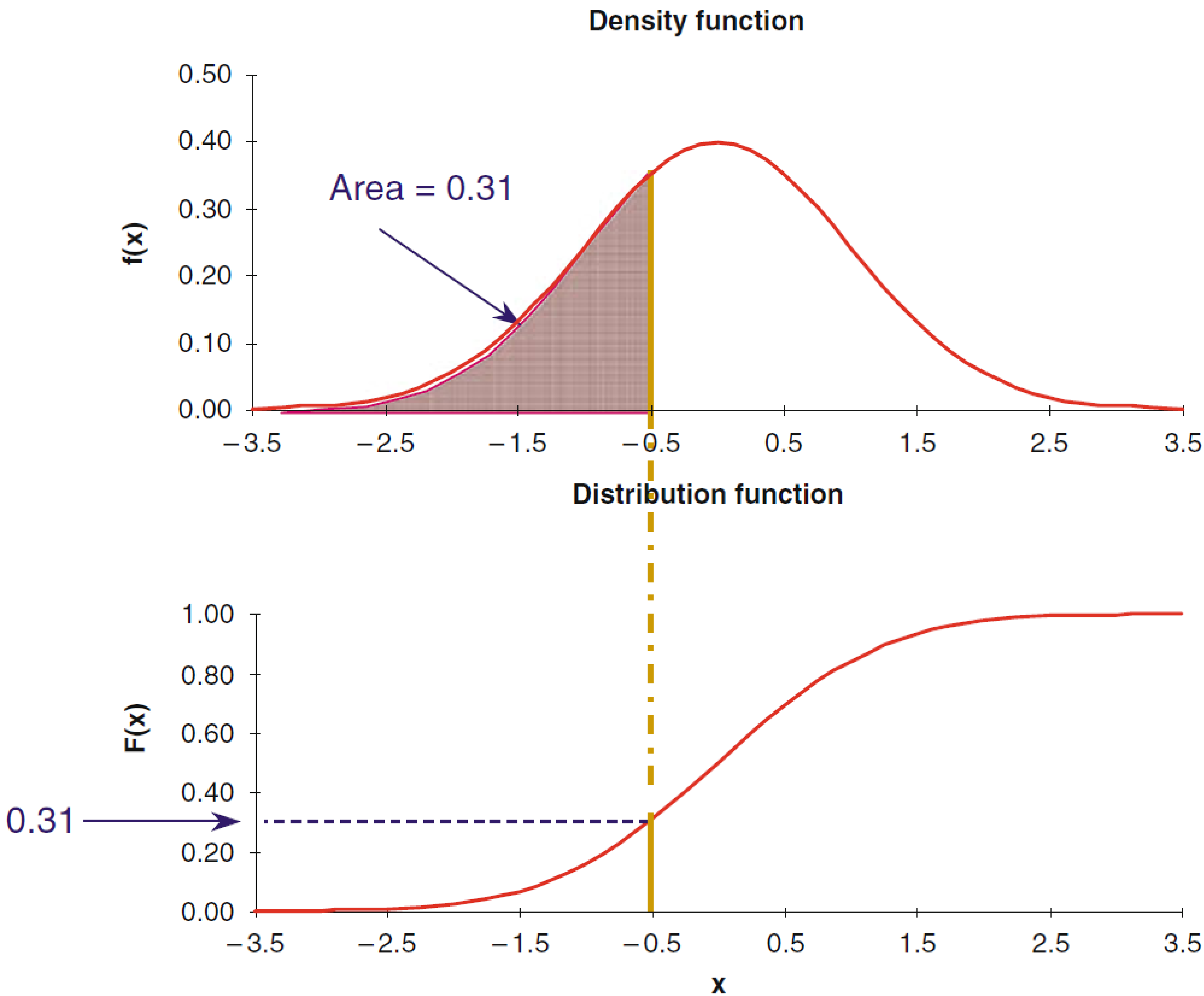
i) F_X prend ses valeurs dans $[0,\, 1]$;

ii) F_X est une fonction croissante;

iii) $\lim_{x \to -\infty} F_X(x) = 0$ et $\lim_{x \to +\infty} F_X(x) = 1$;

iv) F_X est une **fonction continue** dans l'intervalle $[0,\, 1]$;

v) $F_X(x)$ est la primitive de $f_X(x)$, *i.e.* $f_X(x) = \frac{\partial}{\partial x} F_X(x)$;


Density function
0.50
0.40
0.30
0.20
0.10
0.00
f(x)
Area = 0.31
−3.5
−2.5
−1.5
−0.5
0.5
1.5
2.5
3.5
Distribution function
1.00
0.80
0.60
0.40
0.20
0.00
F(x)
0.31
x

vi) Relation entre la probabilité d'intervalle et la fonction de répartition:

$$\begin{aligned}
P(a < X \leq b) &= \int_a^b f_X(u)\,du \\
&= \underbrace{\int_{-\infty}^b f_X(u)\,du}_{F_X(b)} - \underbrace{\int_{-\infty}^a f_X(u)\,du}_{F_X(a)} \\
&= F_X(b) - F_X(a).
\end{aligned}$$

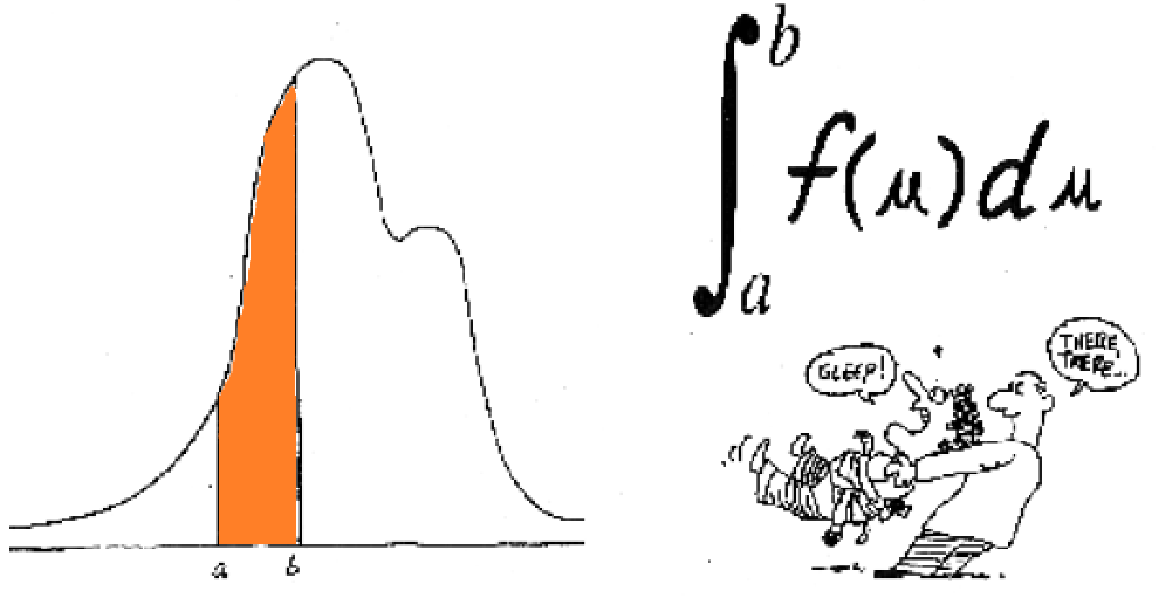


Illustration de la propriété vi) d'une fonction de répartition.

Remarques:

a) La fonction de répartition d'une variable aléatoire continue est une probabilité d'intervalle particulière: Celle de la demi-droite $]-\infty, x]$;

b) La propriété vi) reste valable pour une variable aléatoire discrète;

c) Connaître la fonction de répartition et connaître la fonction de densité est mathématiquement équivalent. Dans la plupart des cas, la distribution de probabilité d'une variable aléatoire continue est définie par la fonction de densité qui, en général, est plus simple à exprimer.

EXEMPLE

On a (voir page 224)

$$f_X(x) = \begin{cases} 0.5\exp(-0.5x) & x > 0 \\ 0 & \text{sinon} \end{cases}$$

Trouver la fonction de répartition.

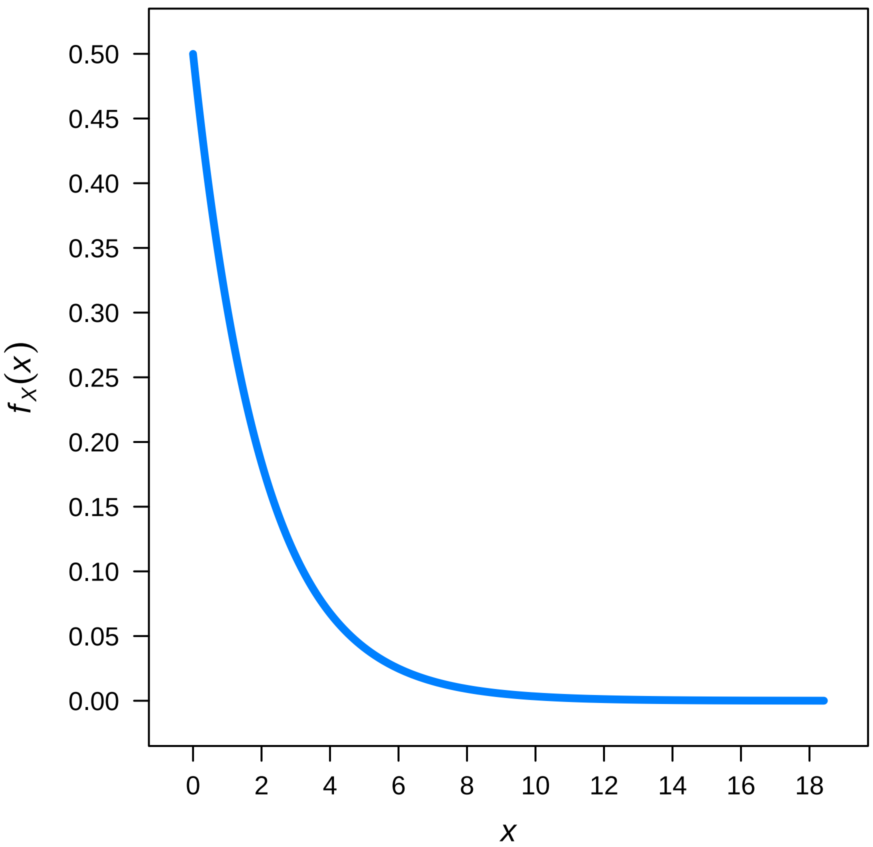
Pour $x > 0$, on a

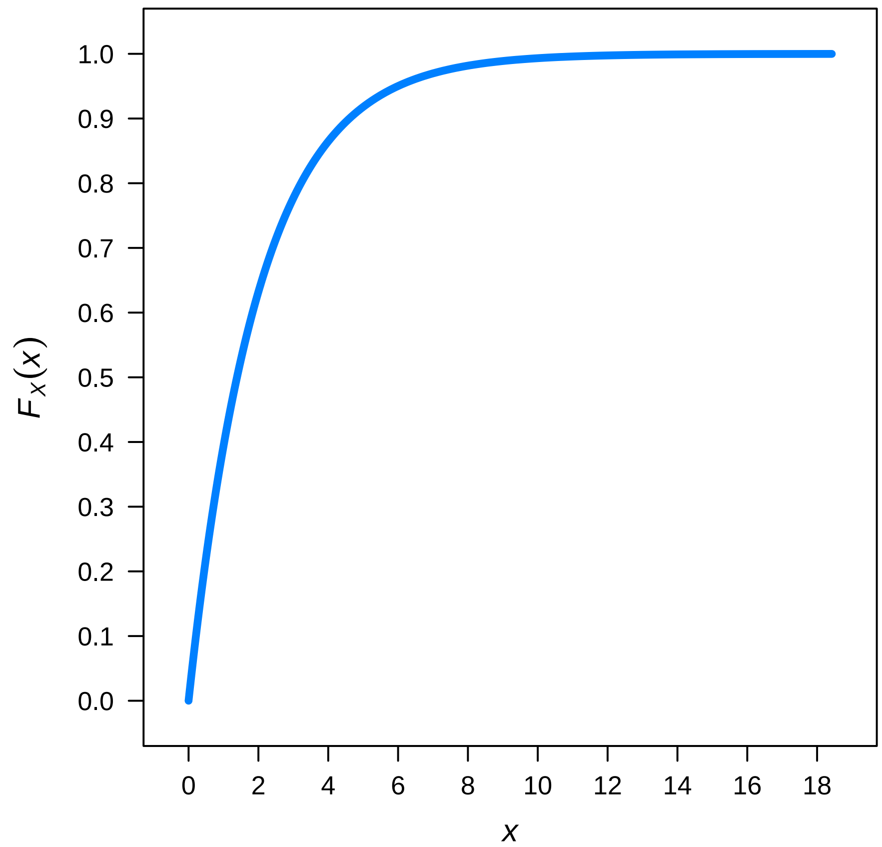
$$\begin{aligned} F_X(x) &= \int_0^x f_X(u)du \\ &= 0.5\left[-2\exp(-0.5u)\right]\Big|_{u=0}^{u=x} \\ &= 0.5\left[-2\exp(-0.5x) + 2\exp(0)\right] \\ &= 1 - \exp(-0.5x) \end{aligned}$$

et finalement

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \exp(-0.5x) & x > 0 \end{cases}$$

Illustrations:





Remarque: Tout comme dans le cas de variables discrètes, il est possible d'inverser la fonction de répartition $F_X(x)$ afin de déterminer la valeur de x qui correspond à une certaine probabilité cumulée $\alpha = P(X \leq x)$, avec $\alpha \in [0, 1]$.

Quantile d'ordre α, noté $F_X^{-1}(\alpha)$ ou $Q(\alpha)$: La réalisation de X correspondante à une probabilité cumulée égale à α:

$$\boxed{P\left[X \leq F_X^{-1}(\alpha)\right] = F_X\left[F_X^{-1}(\alpha)\right] = \alpha}$$

pour $\alpha \in [0, 1]$.

Tout comme pour le cas de variables discrètes, il convient de noter qu'un quantile est une réalisation de la variable aléatoire X.

EXEMPLE

On a (voir page 235)

$$F_X(x) = 1 - \exp(-0.5x), \text{ pour } x > 0.$$

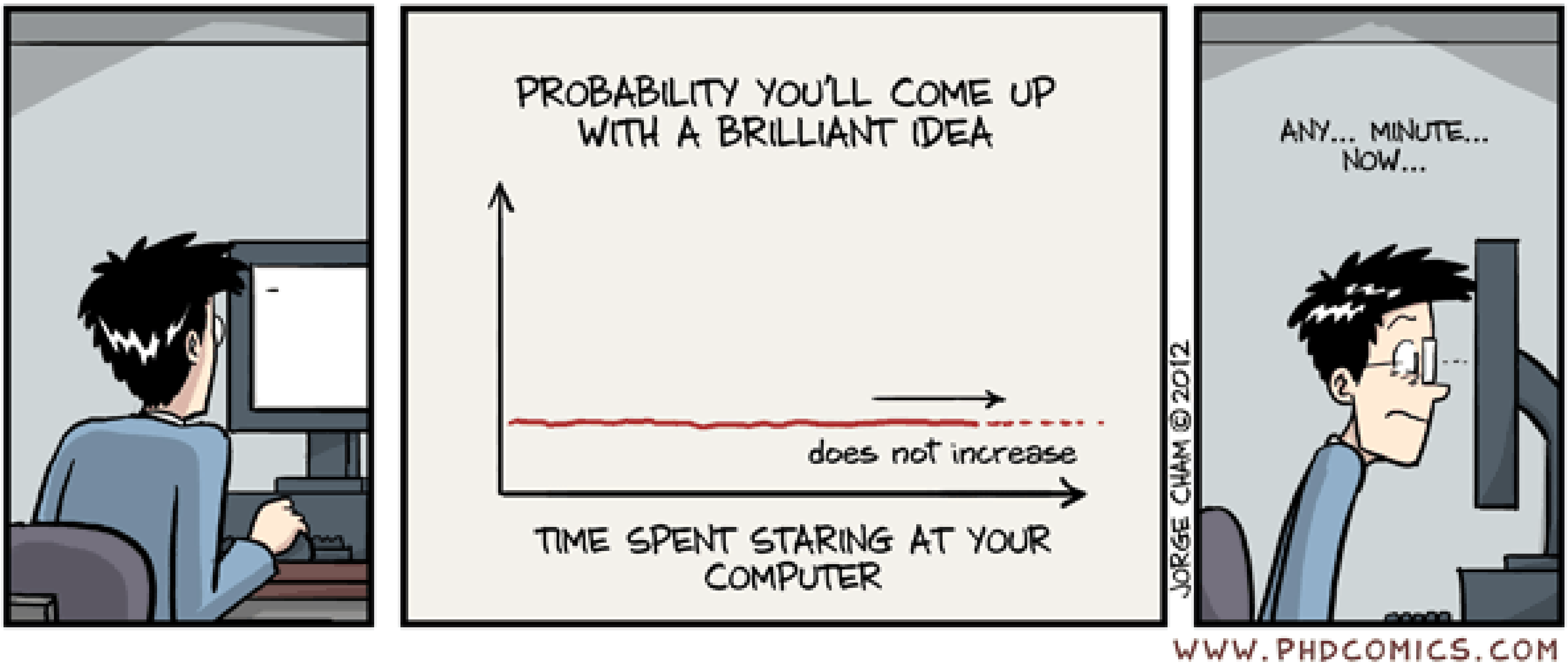
Posons $\alpha = F_X(x)$ et inversons la fonction $F_X(x)$:

$$F_X^{-1}(\alpha) = x = -\frac{\ln(1-\alpha)}{0.5} \text{ pour } \alpha \in [0,1].$$

Par exemple, le quantile d'ordre $\alpha = 0.05$ est

$$Q(0.05) = F_X^{-1}(0.05) = -\frac{\ln(1-0.05)}{0.5} \approx 0.1026.$$

Donc il y a 5% de chances que les réalisations de la variable aléatoire X soient inférieures ou égales à 0.1026, *i.e.* $P(X \leq 0.1026) = 0.05$.


PROBABILITY YOU'LL COME UP WITH A BRILLIANT IDEA
does not increase
TIME SPENT STARING AT YOUR COMPUTER
JORGE CHAM © 2012
ANY... MINUTE... NOW...
WWW.PHDCOMICS.COM

4.2 La distribution uniforme

La distribution **uniforme** sur l'intervalle $[a, b]$ associe la même probabilité aux intervalles de valeurs de même longueur:

$$f_X(x) = \begin{cases} c & a \leq x \leq b \\ 0 & \text{sinon} \end{cases}$$

Que vaut c?

On a

$$\int_a^b c dx = c \cdot x \Big|_{x=a}^{x=b} = c(b-a)$$

et donc $c = \frac{1}{b-a}$.

Donc

$$\boxed{f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & \text{sinon} \end{cases}}$$

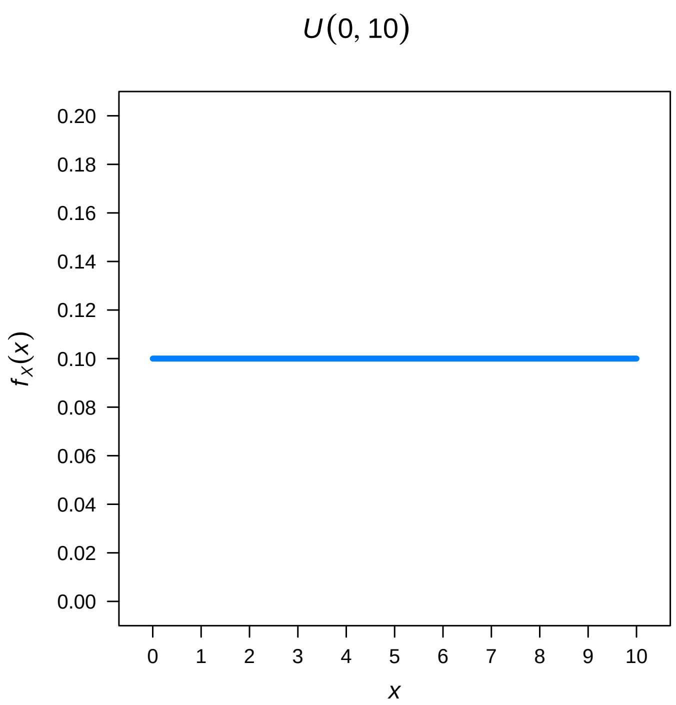
Notation: X est une variable aléatoire uniforme:

$$\boxed{X \sim U(a, b)}.$$

EXEMPLE

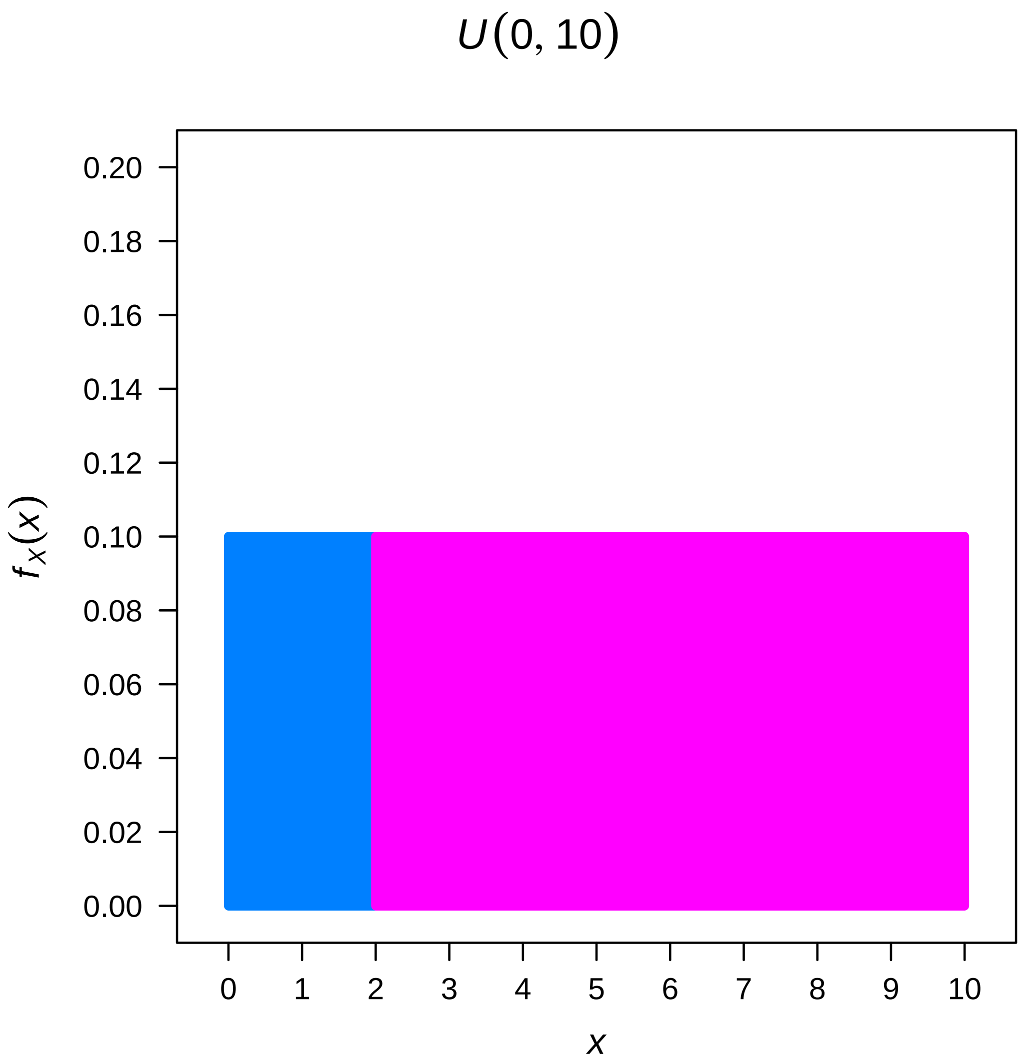
X est distribuée uniformément sur $[0, 10]$, *i.e.* $X \sim U(0, 10)$. Alors

$$f_X(x) = \begin{cases} \frac{1}{10-0} = 0.1 & 0 \leq x \leq 10 \\ 0 & \text{sinon} \end{cases}$$



$$
\begin{aligned}
P(0 \leq X \leq 1) \quad &= \int_0^1 0.1dx && = 0.1 \cdot x\Big|_{x=0}^{x=1} \\
&= 0.1(1-0) && = 0.1 \\
\\
P(0 \leq X \leq 2) \quad &= 2 \cdot 0.1 && = 0.2 \\
\\
P(2 \leq X \leq 4) \quad &= P(0 \leq X \leq 2) && = 0.2 \\
\\
P(X \geq 2) \quad &= P(2 \leq X \leq 10) && = 8 \cdot 0.1 = 0.8
\end{aligned}
$$

Illustration pour $P(X \geq 2) = 0.8$:



EXEMPLE

X est uniformément distribuée sur $[a, b]$, *i.e.* $X \sim U(a, b)$. Trouver la fonction de répartition $F_X(x)$.

Réponse: On a

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{sinon} \end{cases}$$

Alors, pour $a \leq x \leq b$

$$\begin{aligned} F_X(x) &= \int_a^x f_X(u)du = \int_a^x \frac{1}{b-a} du \\ &= \frac{1}{b-a} \, u \Big|_{u=a}^{u=x} = \frac{x-a}{b-a} \end{aligned}$$

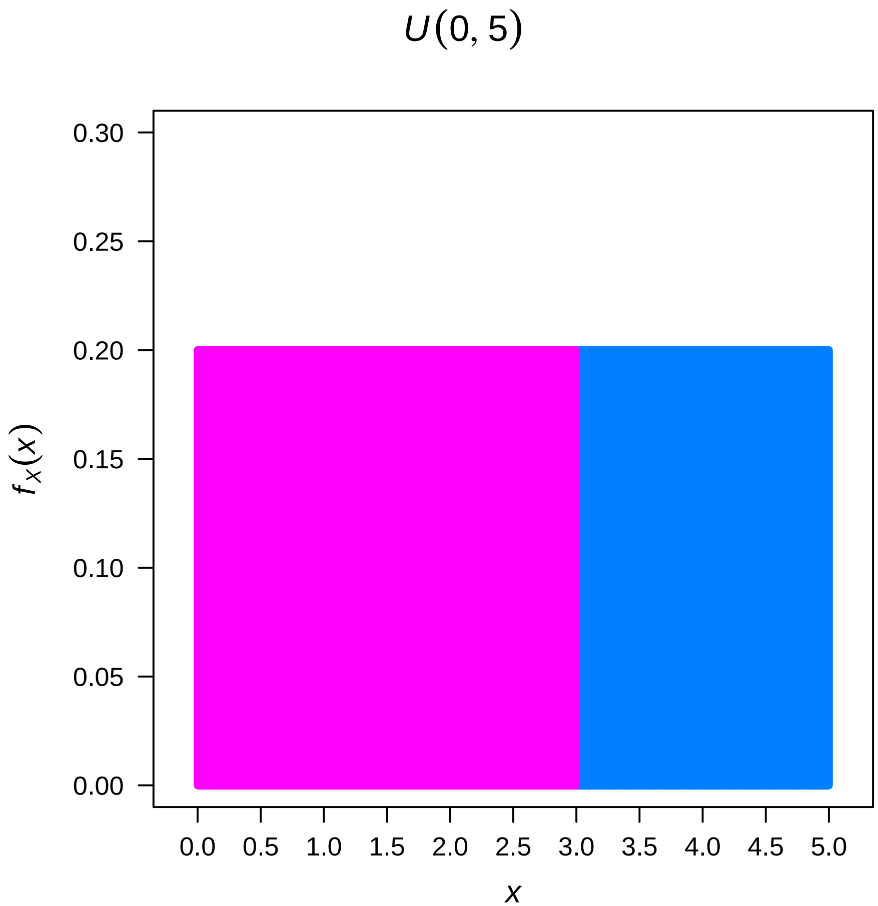
et finalement

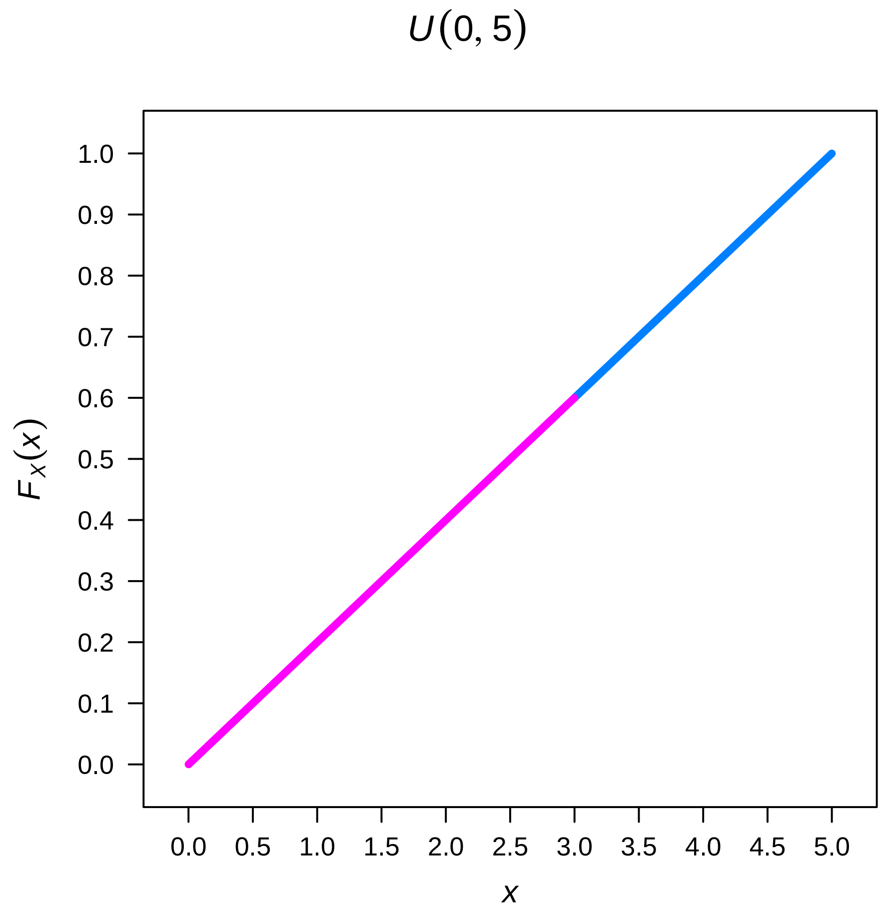
$$\boxed{F_X(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}}$$

Exemple: Si $a = 0$ et $b = 5$, *i.e.* $X \sim U(0,5)$, alors

$$P(X \le 3) = F_X(3) = \frac{3-0}{5-0} = 3/5 = 0.6.$$

Illustrations pour $P(X \leq 3) = F_X(3) = 0.6$:





$6^2 = ?$
36 !
ENRICO
MED

4.3 Espérance et variance

$$\boxed{\mu = \mathsf{E}(X) = \int_{-\infty}^{+\infty} x\, f_X(x)dx}$$

$$\boxed{\begin{aligned} \sigma^2 &= \mathsf{var}(X) = \mathsf{E}[(X-\mu)^2] \\ &= \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)dx \end{aligned}}$$

Propriété:

$$\mathsf{var}(X) = \mathsf{E}(X^2) - [\mathsf{E}(X)]^2 = \int_{-\infty}^{+\infty} x^2 f_X(x)dx - \mu^2$$

EXEMPLE

X est uniformément distribuée sur $[a, b]$, *i.e.* $X \sim U(a, b)$. Quelle est son espérance? Sa variance?

Réponse:

$$\begin{aligned}
\mu = \mathsf{E}(X) &= \int_a^b x \cdot \frac{1}{b-a} dx = \frac{1}{b-a} \left(0.5x^2\right)\Big|_{x=a}^{x=b} \\
&= 0.5 \frac{b^2 - a^2}{b-a} = 0.5 \frac{(b-a)(b+a)}{(b-a)} \\
&= 0.5(b+a) = \frac{b+a}{2}
\end{aligned}$$

$$
\begin{aligned}
\sigma^2 = \mathsf{var}(X) &= \int_a^b x^2 \cdot \frac{1}{b-a} dx - \mu^2 \\
&= \frac{1}{b-a} \left(\frac{1}{3} x^3 \right) \Bigg|_{x=a}^{x=b} - \mu^2 \\
&= \frac{1}{3} \frac{b^3 - a^3}{b-a} - \mu^2 \\
&= \frac{1}{3} \frac{(b-a)(b^2+ab+a^2)}{(b-a)} - \mu^2 \\
&= \frac{1}{3}(b^2+ab+a^2) - \frac{1}{4}(b+a)^2 \\
&= \frac{1}{12}(b^2 - 2ab + a^2) = \frac{1}{12}(b-a)^2
\end{aligned}
$$

Donc si $X \sim U(a,b)$ on a $\mathsf{E}(X) = \dfrac{b+a}{2}$ et $\mathsf{var}(X) = \dfrac{(b-a)^2}{12}$.

EXEMPLE

$$f_X(x) = \begin{cases} \frac{3}{4}(1-x^2) & -1 \leq x \leq 1 \\ 0 & \textsf{sinon} \end{cases}$$

$$\begin{aligned} \mu = \mathsf{E}(X) &= \int_{-1}^{1} x \cdot \frac{3}{4}(1-x^2)dx = \int_{-1}^{1} \frac{3}{4}(x-x^3)dx \\ &= \frac{3}{4}\left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Bigg|_{x=-1}^{x=1} \\ &= \frac{3}{4}\left[\left(\frac{1}{2}-\frac{1}{4}\right) - \left(\frac{1}{2}-\frac{1}{4}\right)\right] = 0 \end{aligned}$$

$$
\begin{aligned}
\sigma^2 = \mathsf{var}(X) &= \int_{-1}^{1} x^2 \cdot \frac{3}{4}(1-x^2)dx - 0^2 \\
&= \int_{-1}^{1} \frac{3}{4}(x^2 - x^4)dx \\
&= \frac{3}{4}\left(\frac{x^3}{3} - \frac{x^5}{5}\right)\Bigg|_{x=-1}^{x=1} \\
&= \frac{3}{4}\left[\left(\frac{1}{3} - \frac{1}{5}\right) - \left(-\frac{1}{3} + \frac{1}{5}\right)\right] \\
&= \frac{3}{4}\left(\frac{2}{3} - \frac{2}{5}\right) = \frac{3}{4}\frac{4}{15} = \frac{1}{5} = 0.2
\end{aligned}
$$

Rappels des propriétés

Si $Y = a + bX$ alors on a que

$$\begin{aligned} \mu_Y &= a + b\mu_X \\ \sigma_Y^2 &= b^2\sigma_X^2 \end{aligned}$$

Combinaison linéaire de deux variables X et Y:

$$\mathsf{E}(aX + bY) = a\mu_X + b\mu_Y$$

En particulier:

$$\begin{aligned} \mathsf{E}(X + Y) &= \mu_X + \mu_Y \\ \mathsf{E}(X - Y) &= \mu_X - \mu_Y \end{aligned}$$

Si de plus les variables sont indépendantes, alors

$$\text{var}(aX + bY) \quad = \quad a^2\sigma_X^2 + b^2\sigma_Y^2$$

En particulier:

$$\begin{aligned} \text{var}(X + Y) &= \sigma_X^2 + \sigma_Y^2 \\ \text{var}(X - Y) &= \sigma_X^2 + \sigma_Y^2 \end{aligned}$$

Remarque: Ces propriétés sont les mêmes que pour les variables discrètes.

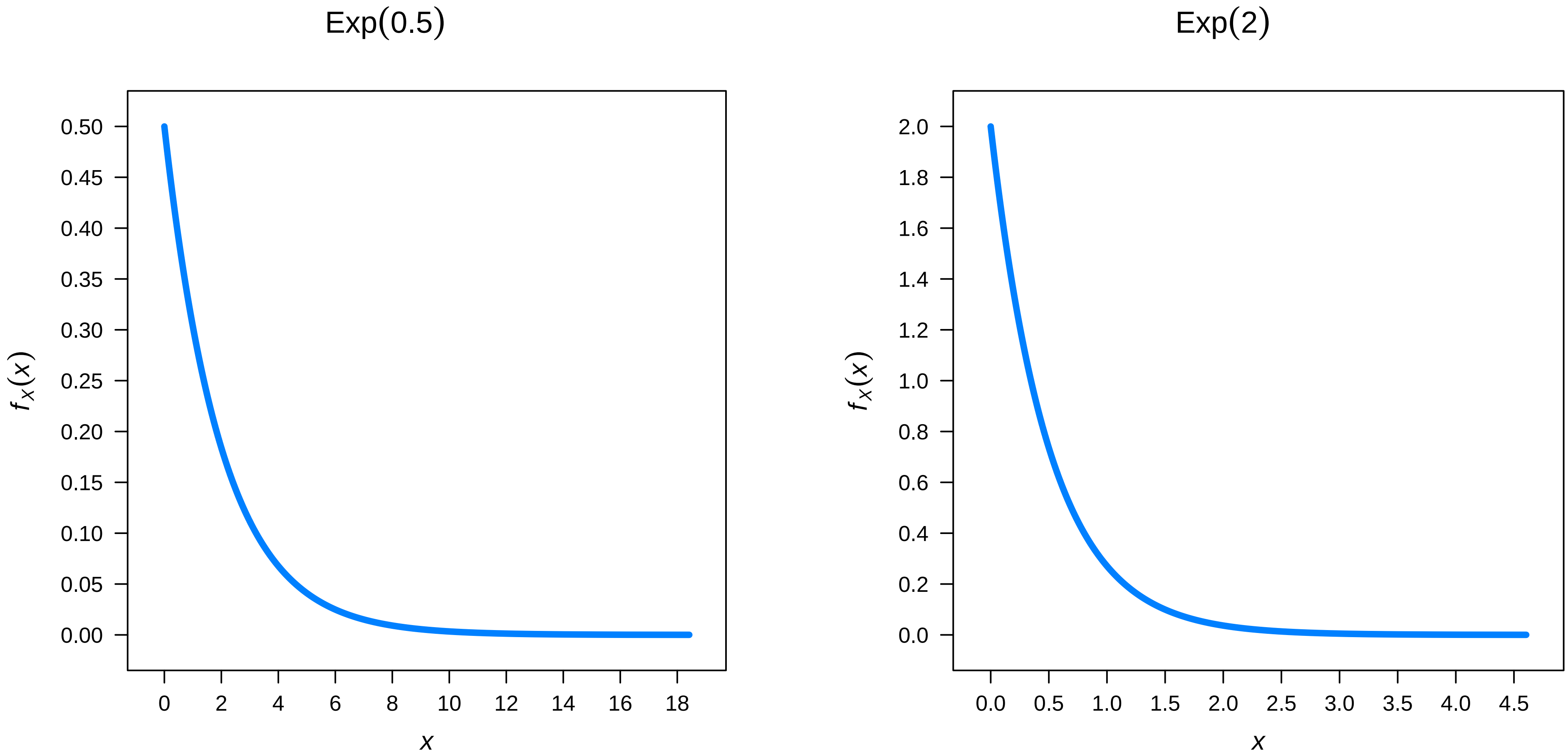
4.4 La distribution exponentielle

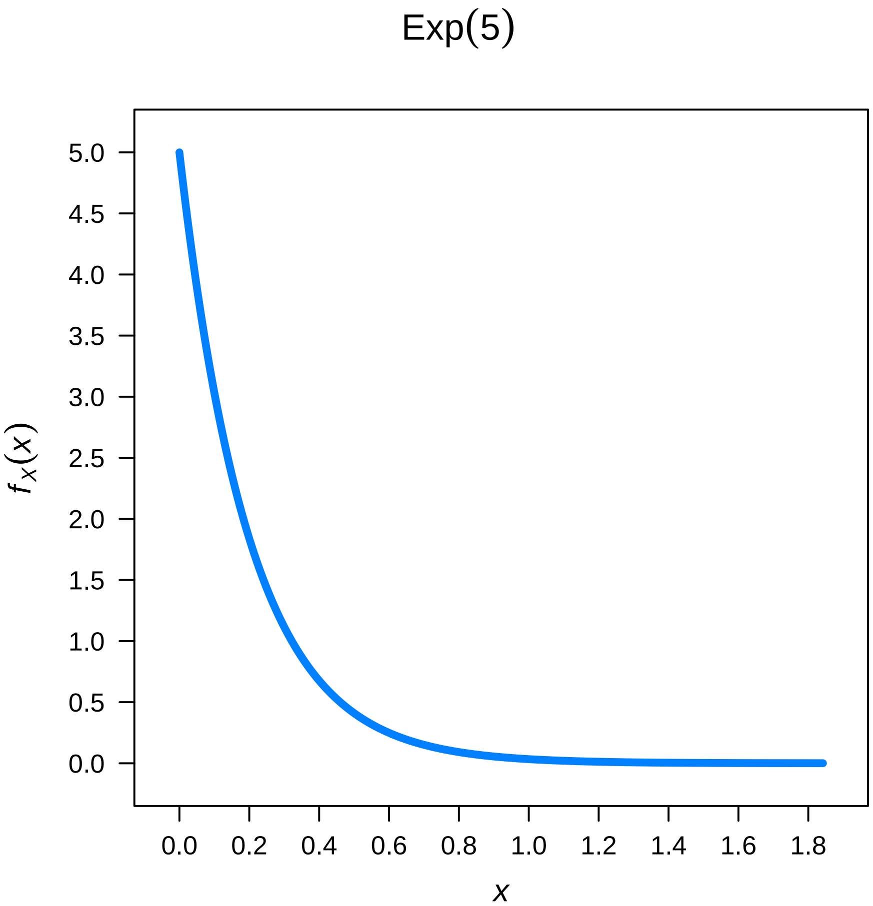
Soit X une variable aléatoire continue. Les caractéristiques et propriétés d'une variable aléatoire **exponentielle** sont:

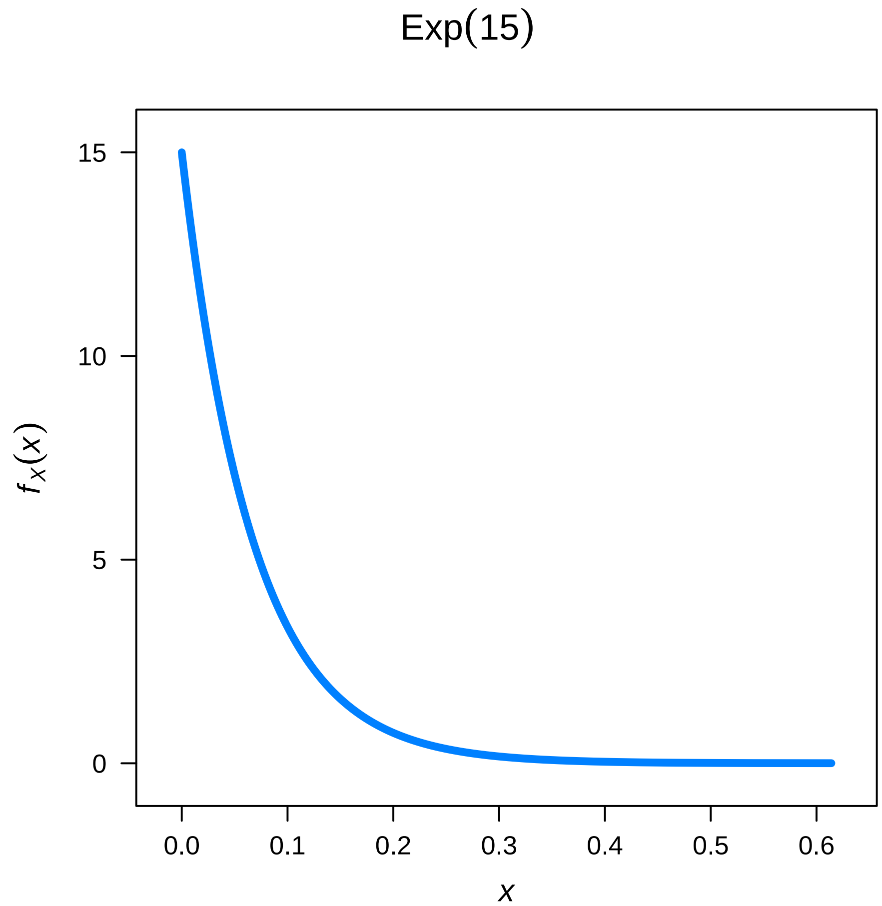
- ▷ X est définie sur $(0; \infty)$;
- ▷ Fonction de densité: $f_X(x) = \lambda \exp(-\lambda x)$ pour $\lambda > 0$;
- ▷ Fonction de répartition: $F_X(x) = 1 - \exp(-\lambda x)$;
- ▷ $\mathsf{E}(X) = 1/\lambda$;
- ▷ $\mathsf{var}(X) = 1/\lambda^2$.

$\rightsquigarrow$ **Notation:** $\boxed{X \sim \mathrm{Exp}(\lambda)}$.

Illustrations:




Exp(5)
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
$f_X(x)$
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
x


Exp(15)
15
10
5
0
$f_X(x)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
x

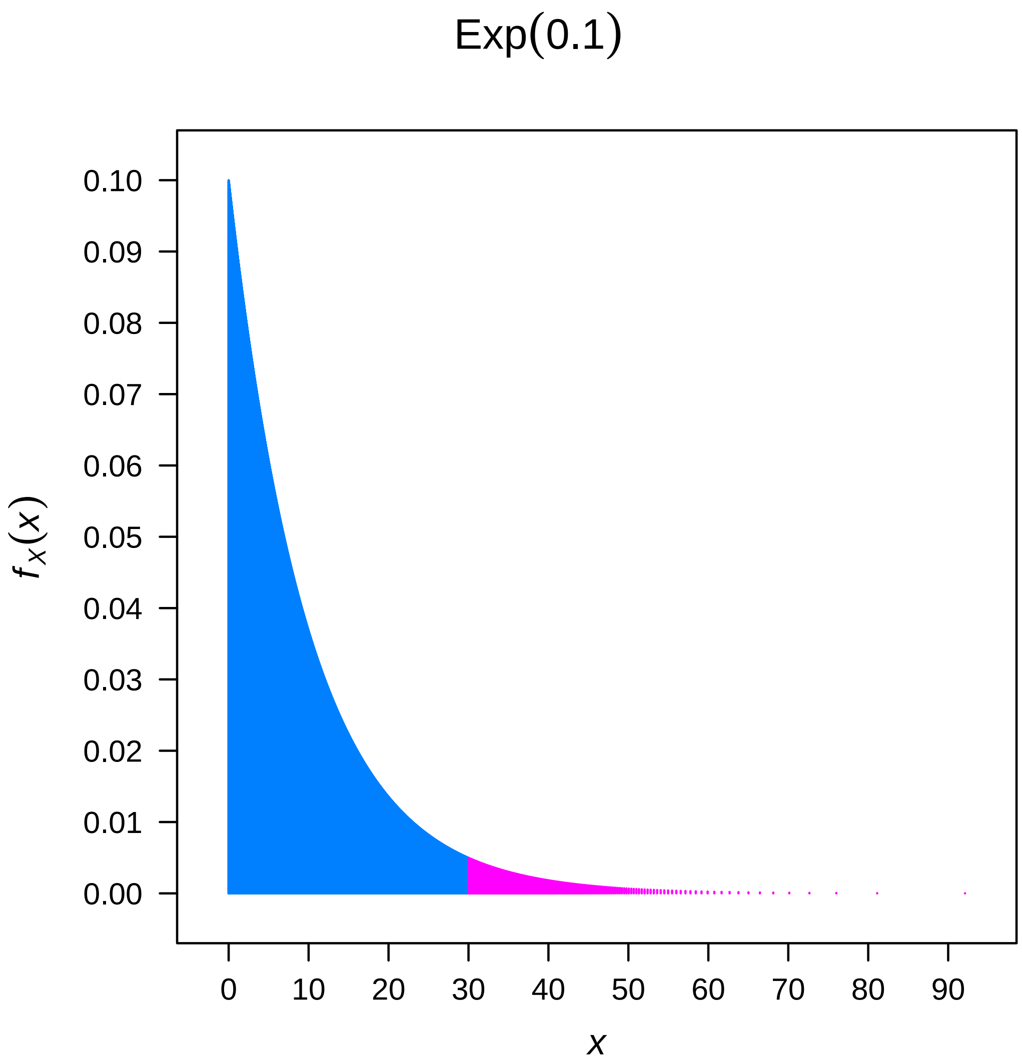
EXEMPLE

Le temps d'attente moyen entre deux clients est pour Alfred de 10 secondes. En supposant que le temps d'attente (en secondes) entre deux clients suit une distribution exponentielle, quelle est la probabilité qu'Alfred doive attendre plus de 30 secondes entre deux clients?

Réponse: X est le temps d'attente entre deux clients, $X \sim \mathsf{Exp}(\lambda)$. Le temps d'attente moyen est de 10 secondes. Donc $\mathsf{E}(X) = 1/\lambda = 10$ et $\lambda = 0.1$. On cherche

$$
\begin{aligned}
P(X > 30) &= 1 - P(X < 30) \\
&= 1 - \int_0^{30} 0.1 \exp(-0.1x) dx = 1 + \Big[\exp(-0.1x) \Big] \Big|_0^{30} \\
&= 1 + \Big[\exp(-3) - \exp(0) \Big] = 0.04979.
\end{aligned}
$$

Illustration pour $P(X > 30) = 0.04979$:



4.5 La distribution gamma

Une variable aléatoire X dont la fonction de densité est

$$f_X(x) = \begin{cases} \dfrac{\lambda(\lambda x)^{t-1}\exp(-\lambda x)}{\Gamma(t)} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

avec la 'fonction gamma' définie par

$$\Gamma(t) = \int_0^\infty \exp(-y)\, y^{t-1} dy$$

(avec $\Gamma(1) = 1$) est appelée une variable aléatoire de distribution **gamma** de paramètres (t, λ), pour $t > 0$ et $\lambda > 0$.

$\rightsquigarrow$ **Notation:** $\boxed{X \sim \Gamma(t, \lambda)}$.

▷ **Propriétés:** $\mathsf{E}(X) = t/\lambda$ et $\mathrm{var}(X) = t/\lambda^2$.

Remarques:

a) La distribution gamma est une généralisation de la distribution exponentielle. En effet, on remarque que la distribution exponentielle de paramètre λ est une distribution gamma de paramètres $t = 1$ et λ, autrement écrit,

$$\text{Exp}(\lambda) = \Gamma(1, \lambda);$$

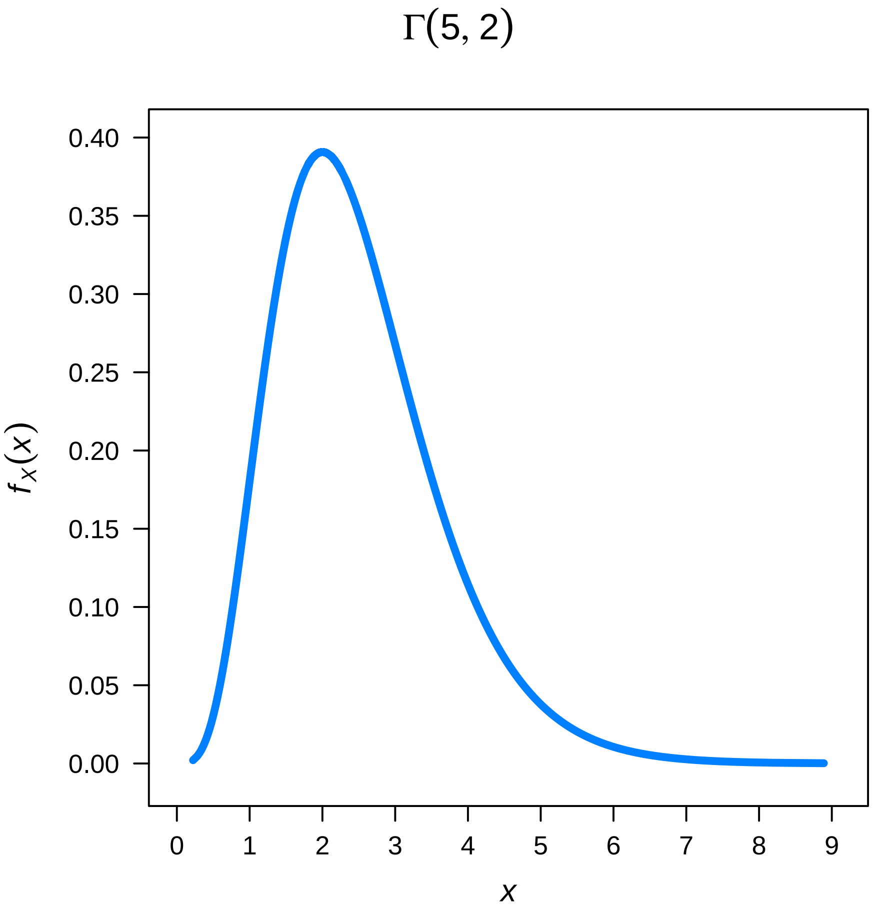
b) Une distribution gamma de paramètres $t = n/2$ (n étant ici un entier positif) et $\lambda = 1/2$ est aussi appelée distribution χ^2_n (lire **chi-carré** à n degrés de liberté), autrement écrit,

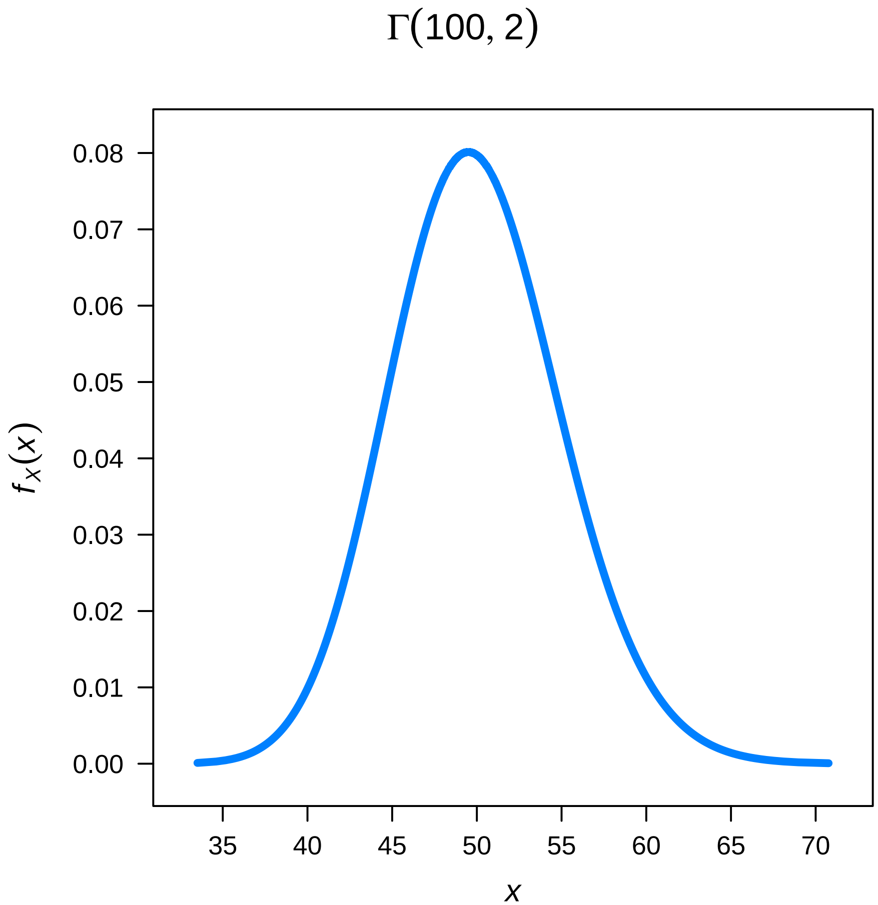
$$\Gamma(n/2, 1/2) = \chi^2_n;$$

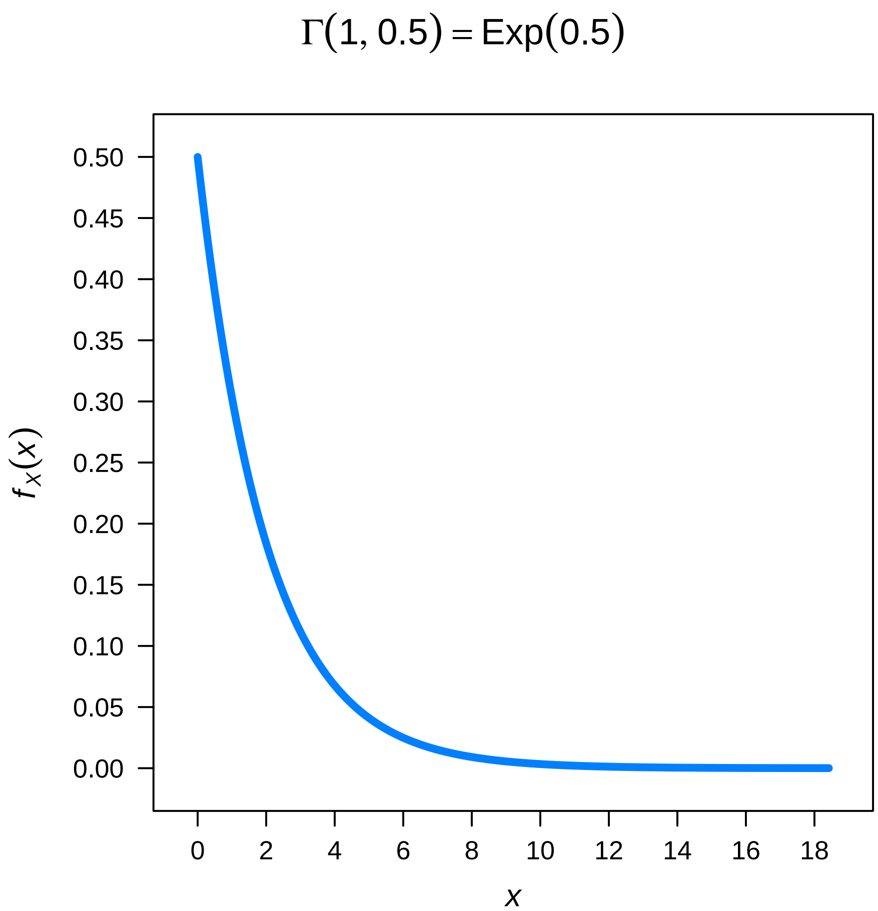
c) Exemple d'occurrence de la distribution gamma: On a de bonnes raisons de supposer que la durée de vie d'un organisme (ou d'un appareil ou d'un atome radioactif) suit une distribution exponentielle; l'inconvénient, en faisant cette hypothèse, est que l'on suppose, de façon implicite, l'absence de vieillissement de cet organisme (ou de cet appareil ou de cet atome radioactif).

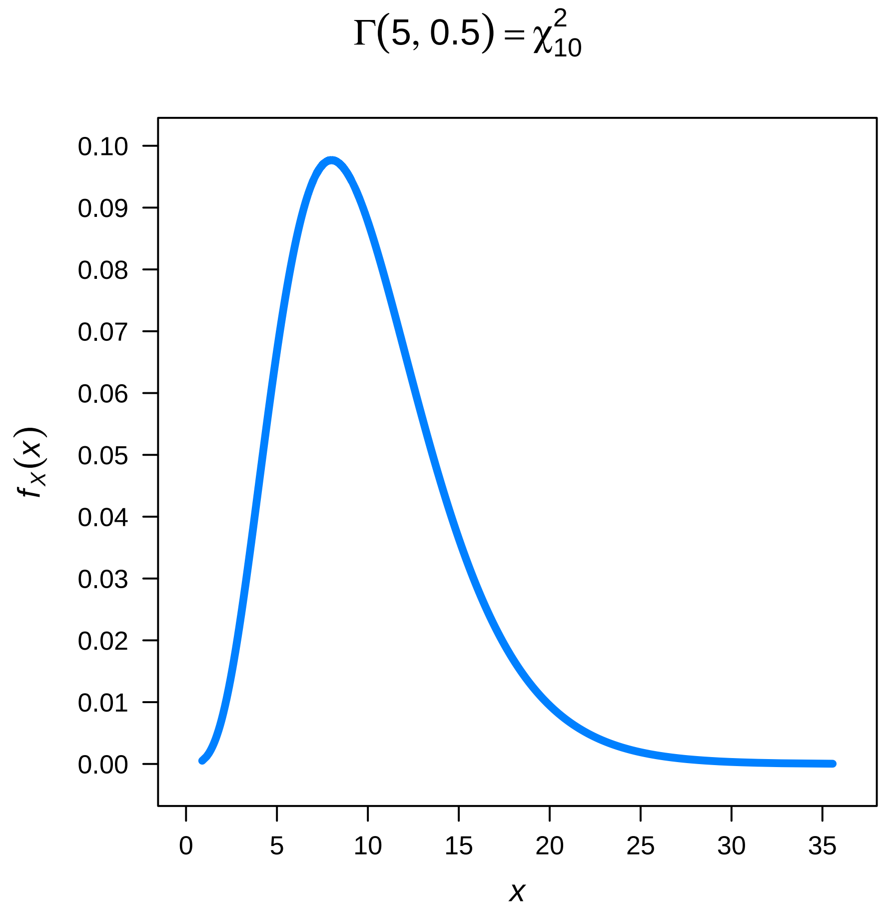
$\rightsquigarrow$ Une hypothèse plus réaliste consiste à supposer que la durée de vie suit une distribution gamma $\Gamma(t, \lambda)$, avec t légèrement plus grand que 1 et $\lambda > 0$.

Illustrations:






$\Gamma(1, 0.5) = \text{Exp}(0.5)$
$f_X(x)$
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
2
4
6
8
10
12
14
16
18
x


$\Gamma(5, 0.5) = \chi^2_{10}$
$f_X(x)$
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
5
10
15
20
25
30
35
x

4.6 La distribution de Weibull

Une variable aléatoire X dont la fonction de répartition est

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq \nu \\ 1 - \exp\left[-\left(\dfrac{x-\nu}{\alpha}\right)^{\beta}\right] & \text{si } x > \nu \end{cases}$$

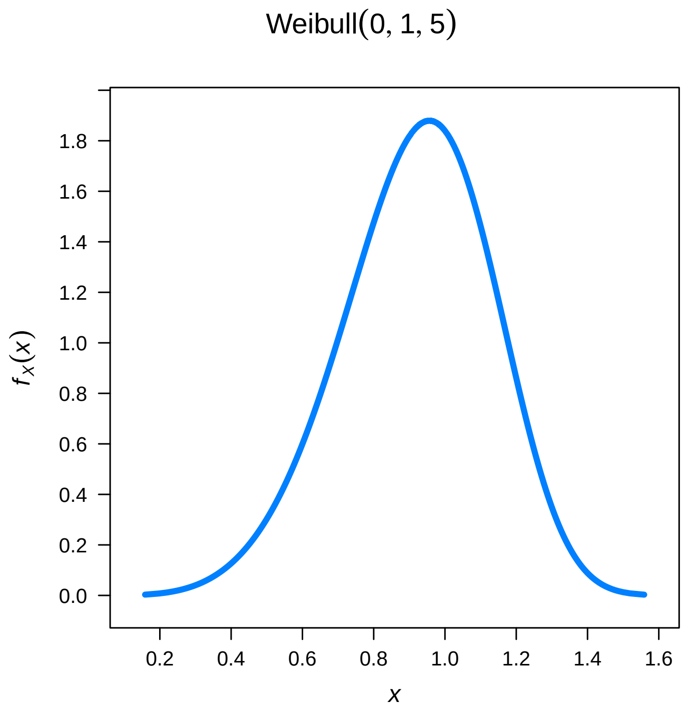
est appelée une variable aléatoire de distribution **Weibull** de paramètres (ν, α, β), pour $\alpha > 0$ et $\beta > 0$.

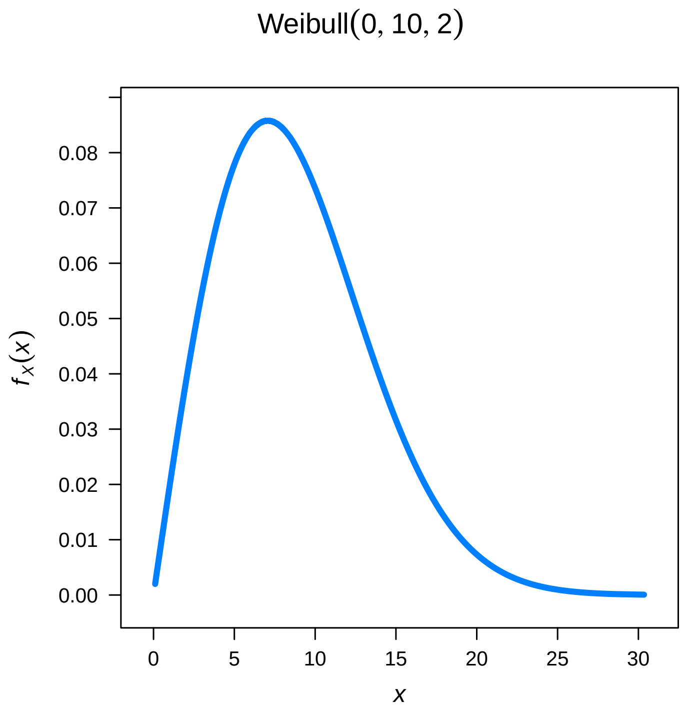
$\rightsquigarrow$ **Notation:** $\boxed{X \sim \text{Weibull}(\nu, \alpha, \beta)}$.

▷ Remarque: Si $X \sim \text{Weibull}(0, \alpha, 1)$ alors $X \sim \text{Exp}(1/\alpha)$.

Remarque: La distribution Weibull a été originellement utilisé pour représenter les effets de l'usure de pièces (par exemple, pour modéliser la durée de vie de pièces), mais son usage s'est beaucoup étendu (par exemple, pour estimer le potentiel éolien d'un site en modélisant la probabilité qu'un vent souffle à certaine vitesse sur ce site).

Illustrations:





4.7 La distribution Pareto

Soit la variable aléatoire X qui suit une distribution de **Pareto** de paramètres (k, β), avec k un réel positif, alors la fonction de densité vérifie

$$f_X(x) = k\frac{\beta^k}{x^{k+1}} \quad \text{pour } \beta < x < \infty,$$

pour $k > 0$ et $\beta > 0$.

$\rightsquigarrow$ **Notation:** $\boxed{X \sim \text{Pareto}(k, \beta)}$.

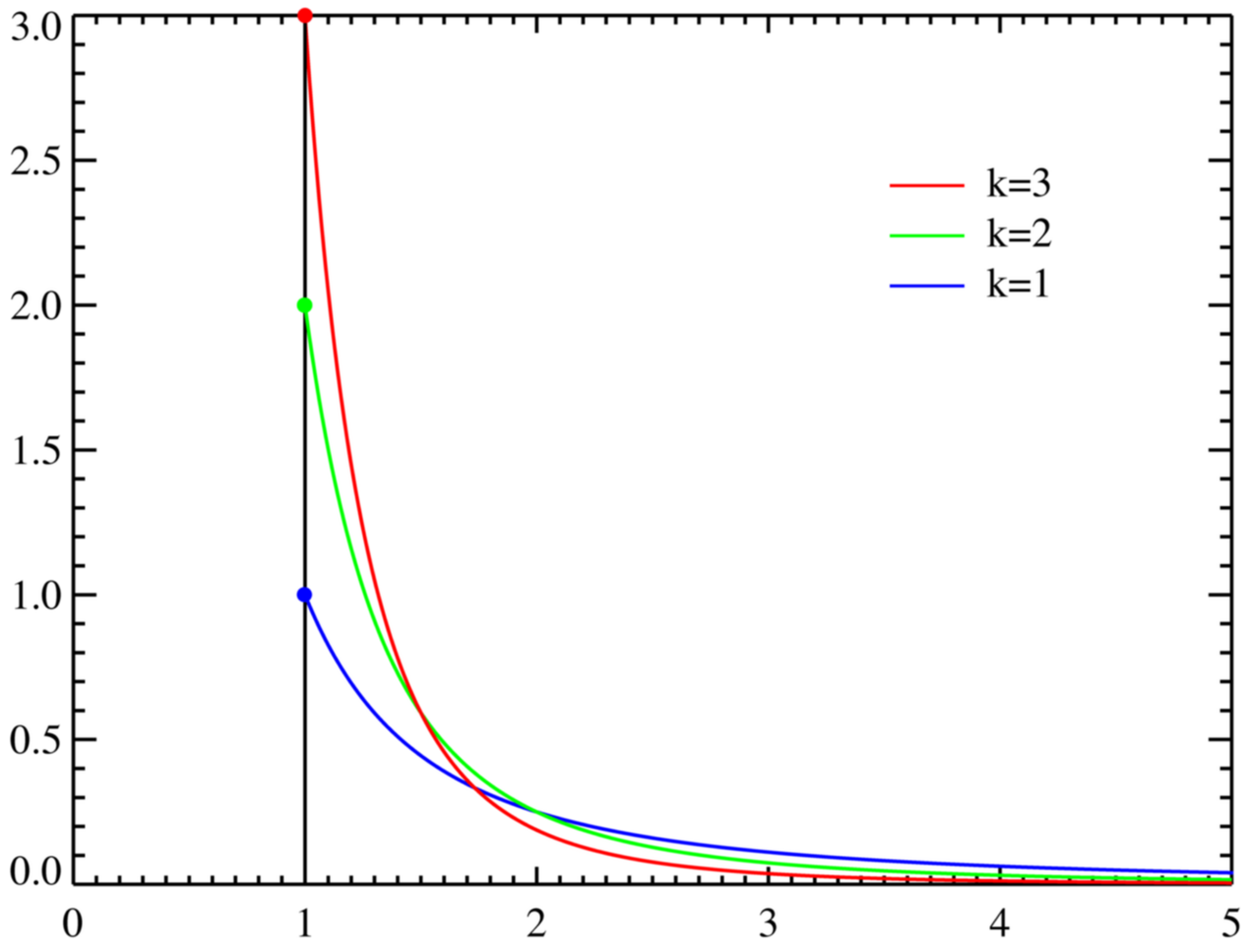
Remarques:

a) Le paramètre k est souvent nommé indice de Pareto;

b) La distribution de Pareto permet de donner une base théorique au 'principe des 80–20', aussi appelé **principe de Pareto**, pour signifier qu'environ 80% des effets sont le produit de 20% des causes.

   Exemples: 80% des réclamations proviennent de 20% des clients; 20% des buveurs consomment 80% des boissons; nous portons 20% de nos vêtements qui se trouvent dans notre placard; lors de réunions, 20% du temps est consacré à prendre 80% des décisions.

Illustrations de fonctions de densité avec $\beta = 1$ (l'axe horizontal symbolise x):



Source: `fr.wikipedia.org/wiki/Loi_de_Pareto_(probabilités)`

4.8 La distribution bêta

On dit qu'une variable aléatoire X, définie sur $[0, 1]$, suit une distribution **bêta** de paramètres (α, β) si sa fonction de densité est

$$f_X(x) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du} & \text{si } 0 < x < 1 \\ 0 & \text{sinon} \end{cases}$$

pour $\alpha > 0$ et $\beta > 0$.

$\rightsquigarrow$ **Notation:** $\boxed{X \sim \text{Beta}(\alpha, \beta)}$.

Remarques:

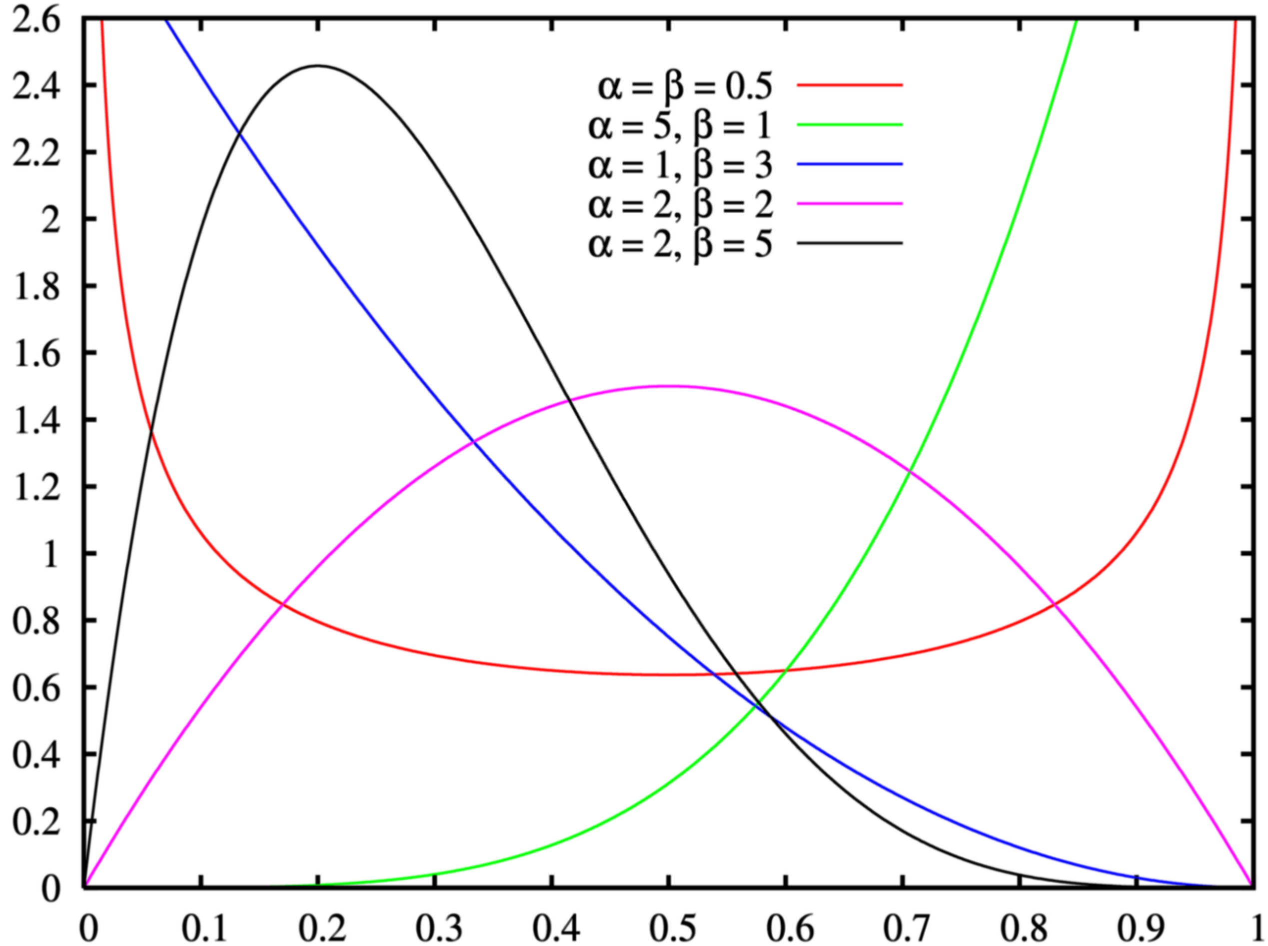
a) La distribution Beta$(1,1)$ est identique à la distribution uniforme $U(0,1)$;

b) Si X et Y sont indépendamment distribuées selon $\Gamma(\alpha,\lambda)$ et $\Gamma(\beta,\lambda)$, alors la variable aléatoire $\dfrac{X}{X+Y}$ est distribuée selon une distribution Beta(α,β);

c) Si $X \sim U(0,1)$ alors $X^2 \sim$ Beta$(1/2,1)$;

d) La distribution Beta$(1/2,1/2)$ est appelée distribution **arc sinus**.

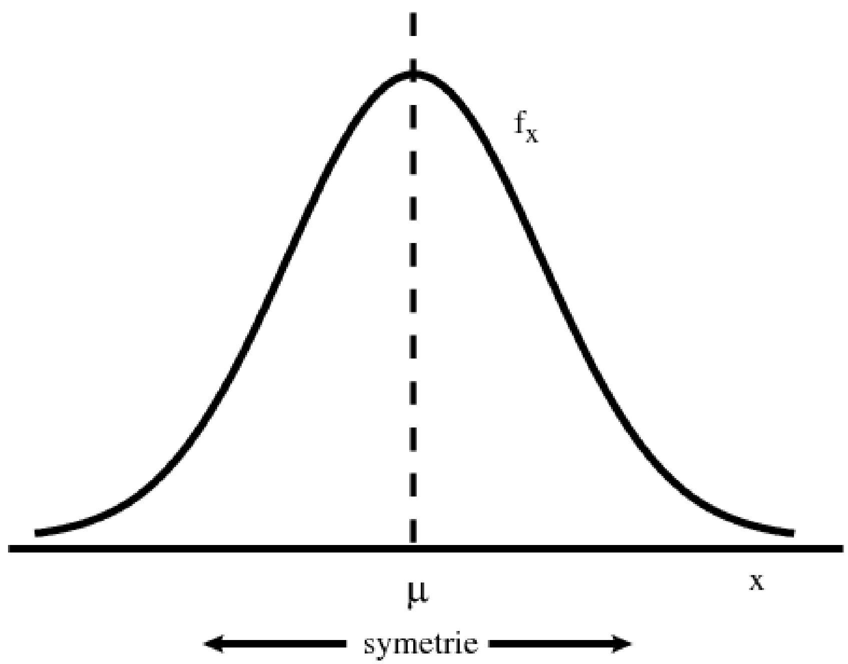
Illustrations de fonctions de densité Beta(α, β) (l'axe horizontal symbolise x):



Source: `fr.wikipedia.org/wiki/Loi_bêta` *(soumis à la licence CC-BY-SA 3.0).*

4.9 La distribution normale

Distribution **normale** = Loi normale = Loi de Gauss = Loi de Laplace–Gauss



▷ La distribution normale joue un rôle central en statistique!

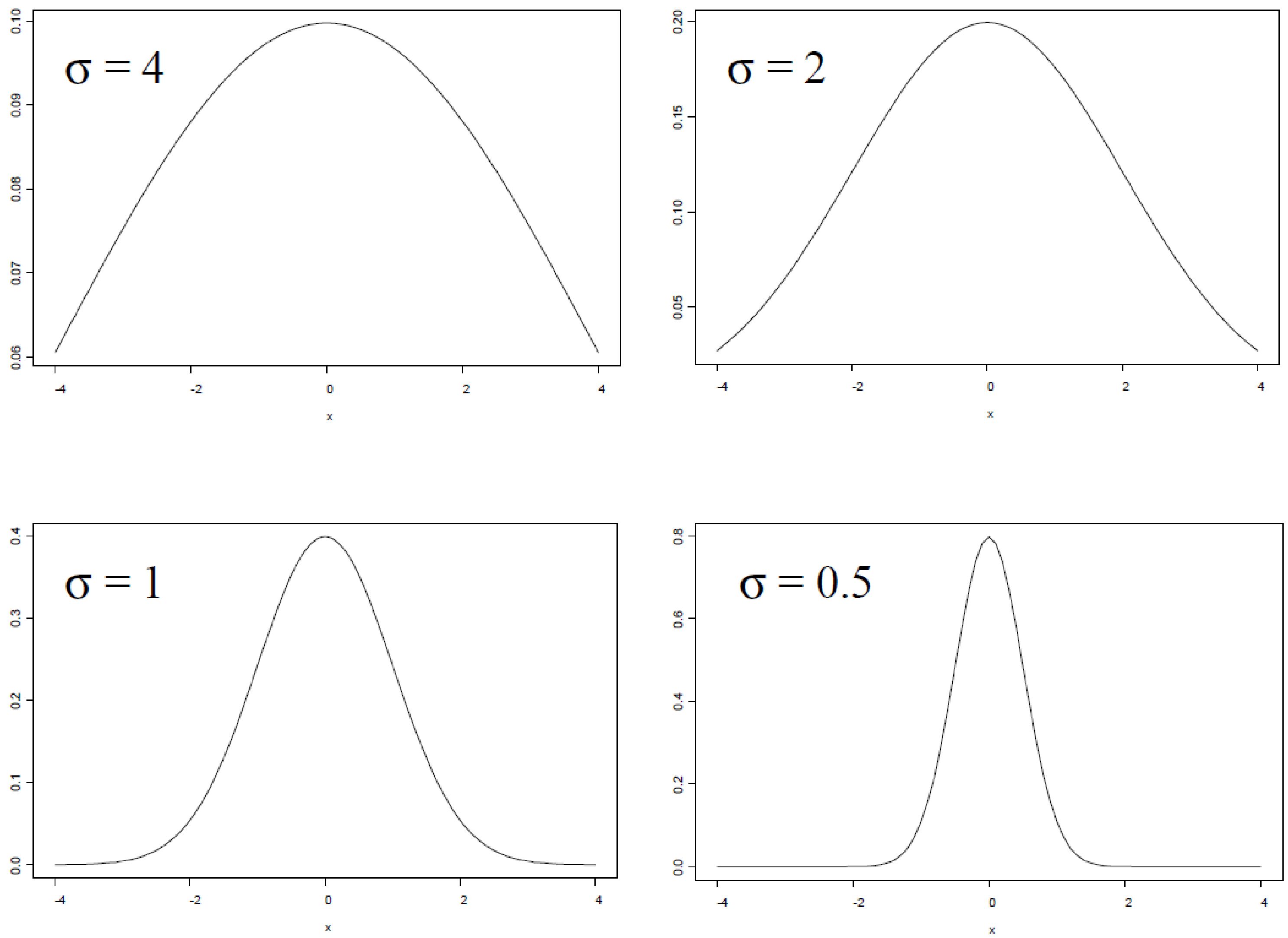
La distribution normale est caractérisée
par son espérance μ et sa variance σ^2.

$\rightsquigarrow$ **Notation:** On note $\boxed{X \sim N(\mu, \sigma^2)}$ avec $\mu = \mathrm{E}(X)$ et $\sigma^2 = \mathrm{var}(X)$.

Caractéristiques:

- ▷ Fonction de densité en forme de cloche;
- ▷ Symétrie autour de l'espérance, *i.e.* la moyenne, μ;
- ▷ Étendue entre $-\infty$ et $+\infty$;
- ▷ Concentration de la probabilité autour de μ;
- ▷ Probabilité qui décroît de façon exponentielle;
- ▷ Vitesse de décroissance selon l'écart-type σ.

Illustrations de fonctions de densité $N(\mu, \sigma^2)$ avec $\mu = 0$:



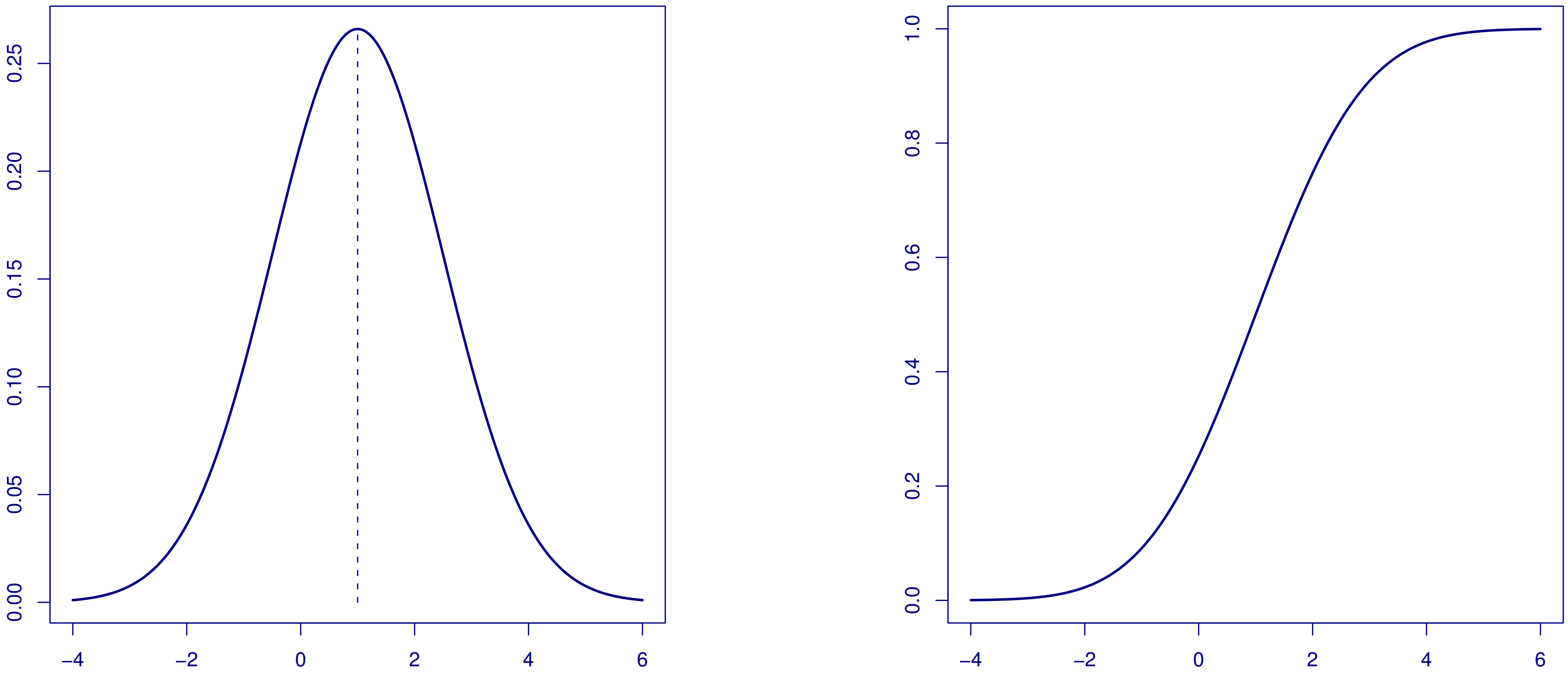
Fonction de densité:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Calcul de probabilités:

$$P(a < X \leq b) = \int_a^b f_X(x)dx$$

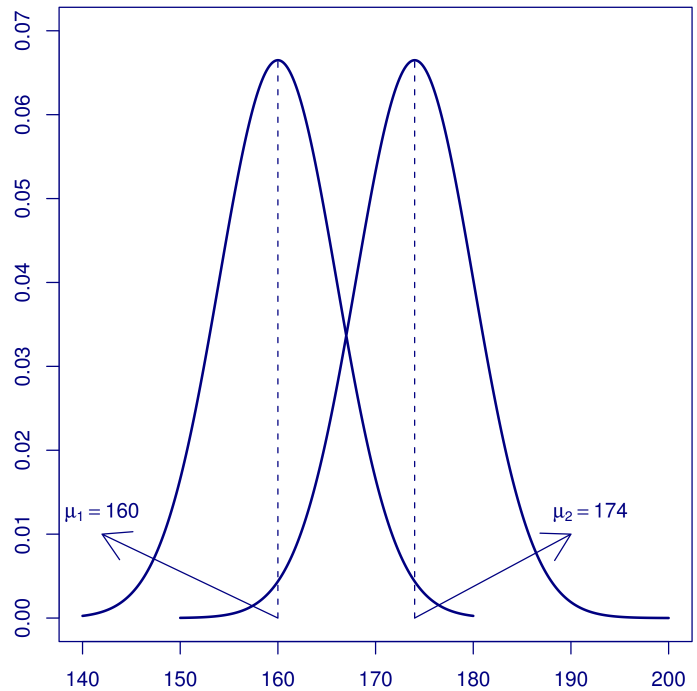
Remarque: La fonction de densité n'admet pas de primitive: Donc 'intégration numérique' ou utilisation d'une table.



Fonction de densité *(gauche) et* **fonction de répartition** *(droite) d'une variable aléatoire issue d'une distribution normale* $N(\mu, \sigma^2)$ *avec* $\mu = 1$, $\sigma^2 = 2.25$.

Remarques

a) Que se passe-t-il si on modifie l'espérance μ de X?

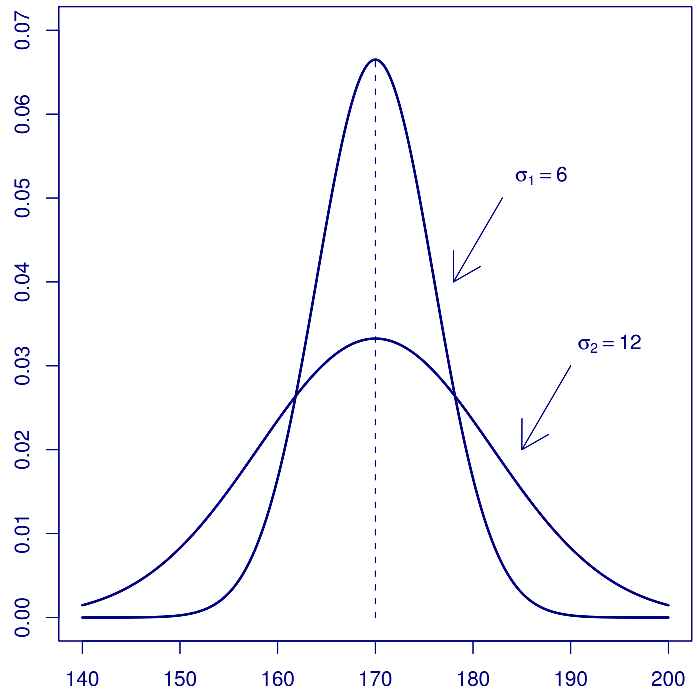


$$\sigma_1^2 = \sigma_2^2 = 36.$$

Modifier la valeur de l'espérance μ de X revient à déplacer la courbe de la fonction de densité le long de l'axe des abscisses.

$\rightsquigarrow$ La grandeur μ peut être considérée comme un paramètre de position de la distribution de X.

b) Que se passe-t-il si on augmente l'écart-type σ de X?

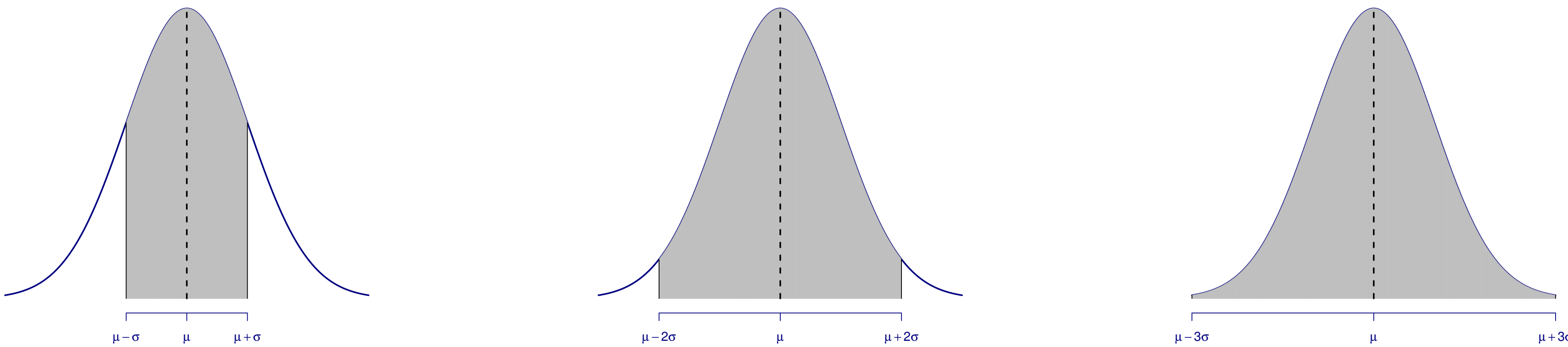


$$\mu_1 = \mu_2 = 170.$$

Augmenter la valeur de l'écart-type σ de X revient à étirer et aplatir la courbe de la fonction de densité.

$\rightsquigarrow$ La grandeur σ peut être considérée comme un paramètre d'échelle de la distribution de X.

c) Existe-t-il une relation entre la probabilité et l'écart-type σ?



$\rightsquigarrow$ Les aires des surfaces hachurées sous les graphes des fonctions de densité (*i.e.* définies par les intervalles centrés en μ et de rayons σ, $2\,\sigma$ et $3\,\sigma$) valent respectivement 0.683, 0.954 et 0.997.

La règle '68 – 95 – 99.7'

Si X est une variable aléatoire de distribution normale $N(\mu, \sigma^2)$ sa réalisation a une probabilité approximative de

- ▷ 68% de se trouver dans l'intervalle $[\mu - \sigma,\ \mu + \sigma]$;
- ▷ 95% de se trouver dans l'intervalle $[\mu - 2\,\sigma,\ \mu + 2\,\sigma]$;
- ▷ 99.7% de se trouver dans l'intervalle $[\mu - 3\,\sigma,\ \mu + 3\,\sigma]$.

- Supposons que X soit une variable aléatoire de distribution normale $N(\mu, \sigma^2)$.

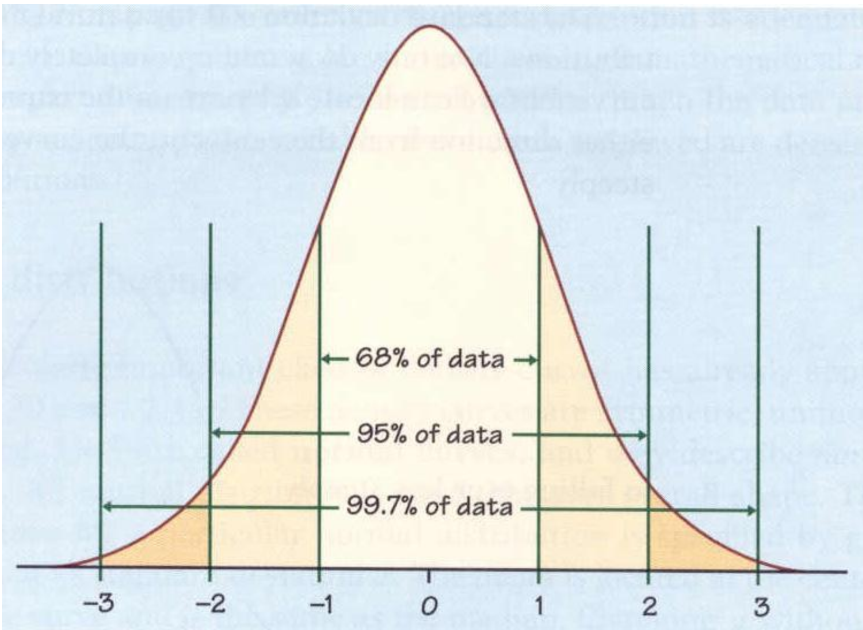
$\rightsquigarrow$ Pour toutes les valeurs prises par l'espérance μ et la variance σ^2, les probabilités que la réalisation de X appartienne aux intervalles centrés en μ et de rayons σ, $2\,\sigma$ et $3\,\sigma$ sont toutes constantes, à savoir approximativement 68%, 95% et 99.7%.

$\rightsquigarrow$ Ces probabilités constantes nous conduisent à introduire une distribution normale particulière à partir de laquelle une probabilité d'intervalle d'une variable aléatoire normale quelconque pourra être calculée.

Distribution normale standard

La distribution normale de référence $\boxed{N(0,1)}$, un cas particulier de la distribution normale pour lequel $\mu = 0$ et $\sigma^2 = 1$, est appelée distribution **normale standard**, distribution **normale centrée réduite** ou Z-**distribution**.

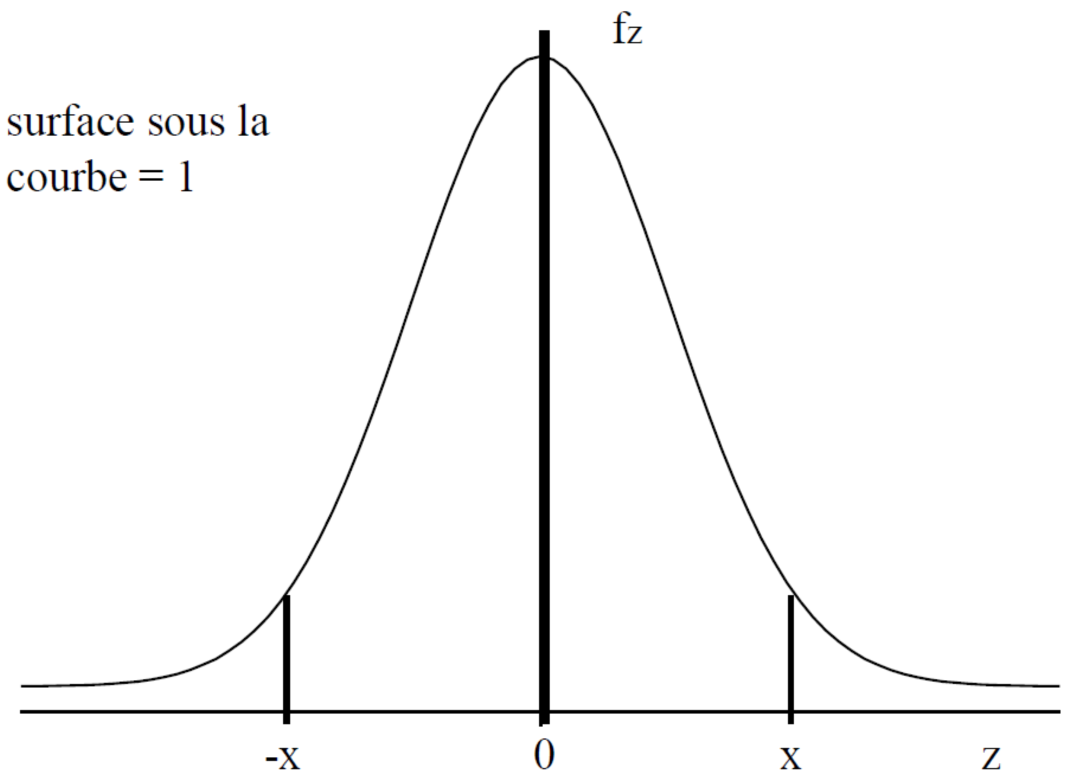
▷ Propriété:



Notation: $\boxed{Z \sim N(0,1)}$

Fonction de densité:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

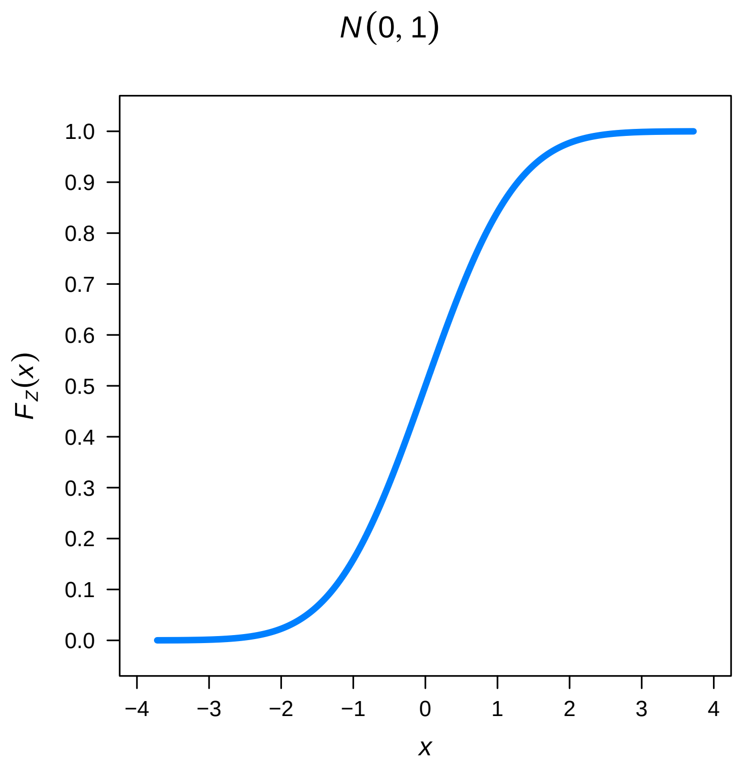


Propriétés de $Z \sim N(0, 1)$:

- $f_Z(z) > 0$;
- $\int_{-\infty}^{+\infty} f_Z(z) dz = 1$;
- $P(Z > x) = 1 - P(Z \leq x)$ (complémentarité);
- $P(Z \leq -x) = P(Z \geq x)$ (symétrie).

Fonction de répartition:

$$\begin{aligned} F_Z(x) &= \int_{-\infty}^{x} f_Z(z)dz \\ &= P(Z \leq x) = \Phi(x) \end{aligned}$$



Les **tables de la distribution normale** se rapportent à la fonction de répartition de la **distribution normale standard**.

$\rightsquigarrow$ Dans les tables on trouve

$$\textbf{soit} \quad F_Z(x) = P(Z \leq x) = \Phi(x)$$

$$\textbf{soit} \quad 1 - F_Z(x) = P(Z \geq x) = 1 - \Phi(x)$$

pour des valeurs positives de x.

$\rightsquigarrow$ Pour les valeurs négatives de x, il faut utiliser les propriétés de Z!

EXEMPLES

$$
\begin{aligned}
P(Z \leq 1) &\approx 0.8413 \\
P(Z \leq 1.96) &\approx 0.9750 \\
P(Z \geq 1.96) &= 1 - P(Z \leq 1.96) \approx 1 - 0.9750 = 0.0250 \\
P(Z \geq -1) &= P(Z \leq 1) \approx 0.8413 \\
P(Z \leq -1.5) &= P(Z \geq 1.5) \\
&= 1 - P(Z \leq 1.5) \approx 1 - 0.9332 = 0.0668
\end{aligned}
$$

Probabilités pour intervalles finis:

$$\boxed{P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)}$$

mais aussi

$$\boxed{P(a \leq Z \leq b) = P(Z \geq a) - P(Z \geq b)}$$

EXEMPLES

$$
\begin{aligned}
P(0.64 \le Z \le 1.96) &= P(Z \le 1.96) - P(Z \le 0.64) \\
&\approx 0.9750 - 0.7389 = 0.2361 \\
P(-0.64 \le Z \le 1.96) &= P(Z \le 1.96) - P(Z \le -0.64) \\
&= P(Z \le 1.96) - [1 - P(Z \le 0.64)] \\
&\approx 0.9750 - (1 - 0.7389) = 0.7139 \\
P(-1.96 \le Z \le -0.64) &= P(0.64 \le Z \le 1.96) \\
&\approx 0.2361
\end{aligned}
$$

Trouver la valeur de x étant donné une probabilité, *i.e.* **le quantile**:

Exemples:

▷ $P(Z \leq a) = 0.9192$

$a \approx 1.40 \quad \rightsquigarrow \quad a = Q(0.9192) \approx 1.40$

▷ $P(Z \geq a) = 0.2420$

$1 - P(Z \geq a) = 1 - 0.2420$

$P(Z \leq a) = 0.7580$

$a \approx 0.70$

▷ $P(Z \leq a) = 0.4325$

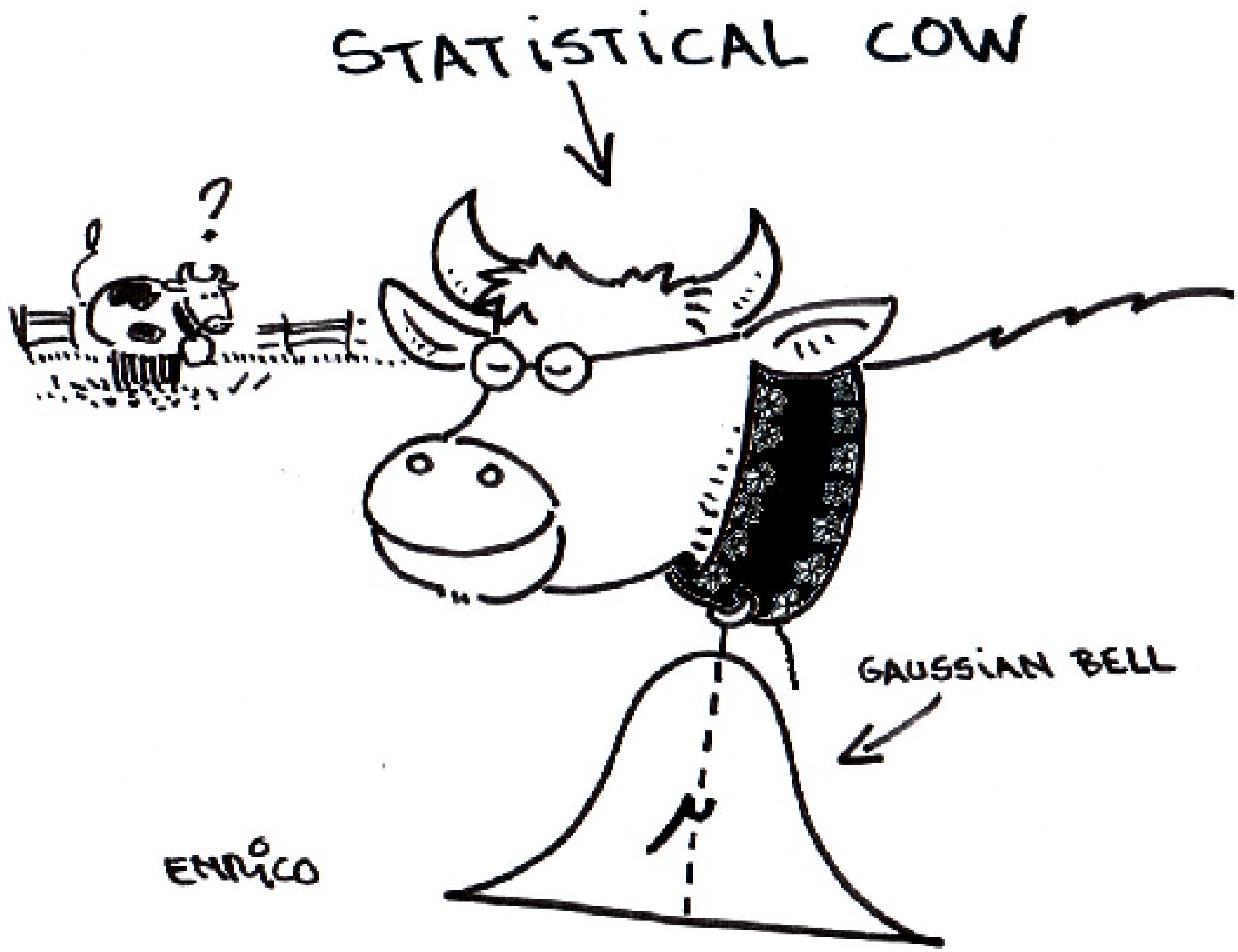
$P(Z \geq -a) = 0.4325$

$1 - P(Z \geq -a) = 1 - 0.4325$

$P(Z \leq -a) = 0.5675$

$-a \approx 0.17$

$a \approx -0.17 \qquad \rightsquigarrow \qquad a = Q(0.4325) \approx -0.17$


STATISTICAL COW
?
GAUSSIAN BELL
ENRICO

Standardisation

Soit X une variable aléatoire normale quelconque d'espérance μ et de variance σ^2:

$$\boxed{X \sim N(\mu, \sigma^2)}$$

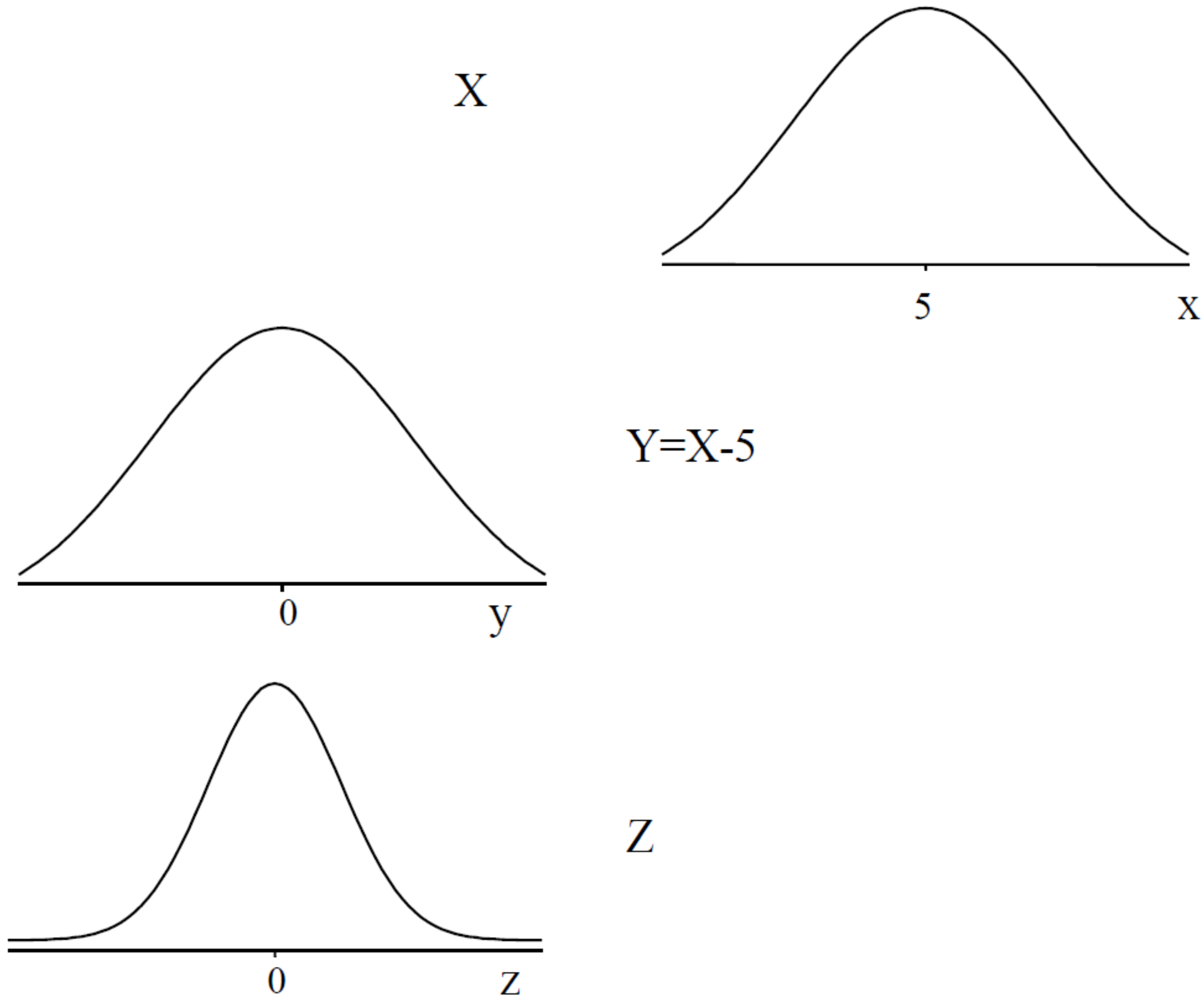
Standardisation: La variable aléatoire

$$\boxed{Z = \frac{X - \mu}{\sigma}}$$

suit une distribution normale standard:

$$\boxed{Z \sim N(0, 1)}$$

Illustration:



Calcul des probabilités:

Soit $X \sim N(\mu, \sigma^2)$

Alors
$$
\begin{aligned}
P(X \leq x) &= P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \\
&= P\left(Z \leq \frac{x-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{x-\mu}{\sigma}\right)
\end{aligned}
$$

EXEMPLES

$X \sim N(3,4)$, donc $\mu = 3$ et $\sigma = 2$.

$$
\begin{aligned}
P(X \leq 6) \quad &= \quad P\left(\tfrac{X-3}{2} \leq \tfrac{6-3}{2}\right) \\
&= \quad P(Z \leq 1.5) \\
&\approx \quad 0.9332
\end{aligned}
$$

$$
\begin{aligned}
P(X \leq 1.5) \quad &= \quad P\left(\tfrac{X-3}{2} \leq \tfrac{1.5-3}{2}\right) \\
&= \quad P(Z \leq -0.75) \; = \; P(Z \geq 0.75) \\
&= \quad 1 - P(Z \leq 0.75) \\
&\approx \quad 1 - 0.7734 = 0.2266
\end{aligned}
$$

$$
\begin{aligned}
P(1 < X \leq 4) &= P\left(\tfrac{1-3}{2} < \tfrac{X-3}{2} \leq \tfrac{4-3}{2}\right) \\
&= P(-1 < Z \leq 0.5) \\
&= P(Z \leq 0.5) - P(Z \leq -1) \\
&= P(Z \leq 0.5) - [1 - P(Z \leq 1)] \\
&\approx 0.6915 - (1 - 0.8413) = 0.5328
\end{aligned}
$$

EXEMPLE

$X \sim N(5, 16)$, donc $\mu = 5$ et $\sigma = 4$. Trouver a tel que $P(X \geq a) = 0.1446$:

$$\begin{aligned}
P\left(\tfrac{X-5}{4} \geq \tfrac{a-5}{4}\right) &= 0.1446 \quad \Rightarrow \\
P\left(Z \geq \tfrac{a-5}{4}\right) &= 0.1446 \quad \Rightarrow \\
1 - P\left(Z \geq \tfrac{a-5}{4}\right) &= 1 - 0.1446 \quad \Rightarrow \\
P\left(Z < \tfrac{a-5}{4}\right) &= 0.8554 \quad \Rightarrow \\
\tfrac{a-5}{4} &\approx 1.06
\end{aligned}$$

Donc $a \approx 4 \cdot 1.06 + 5 = 9.24$.

EXEMPLE

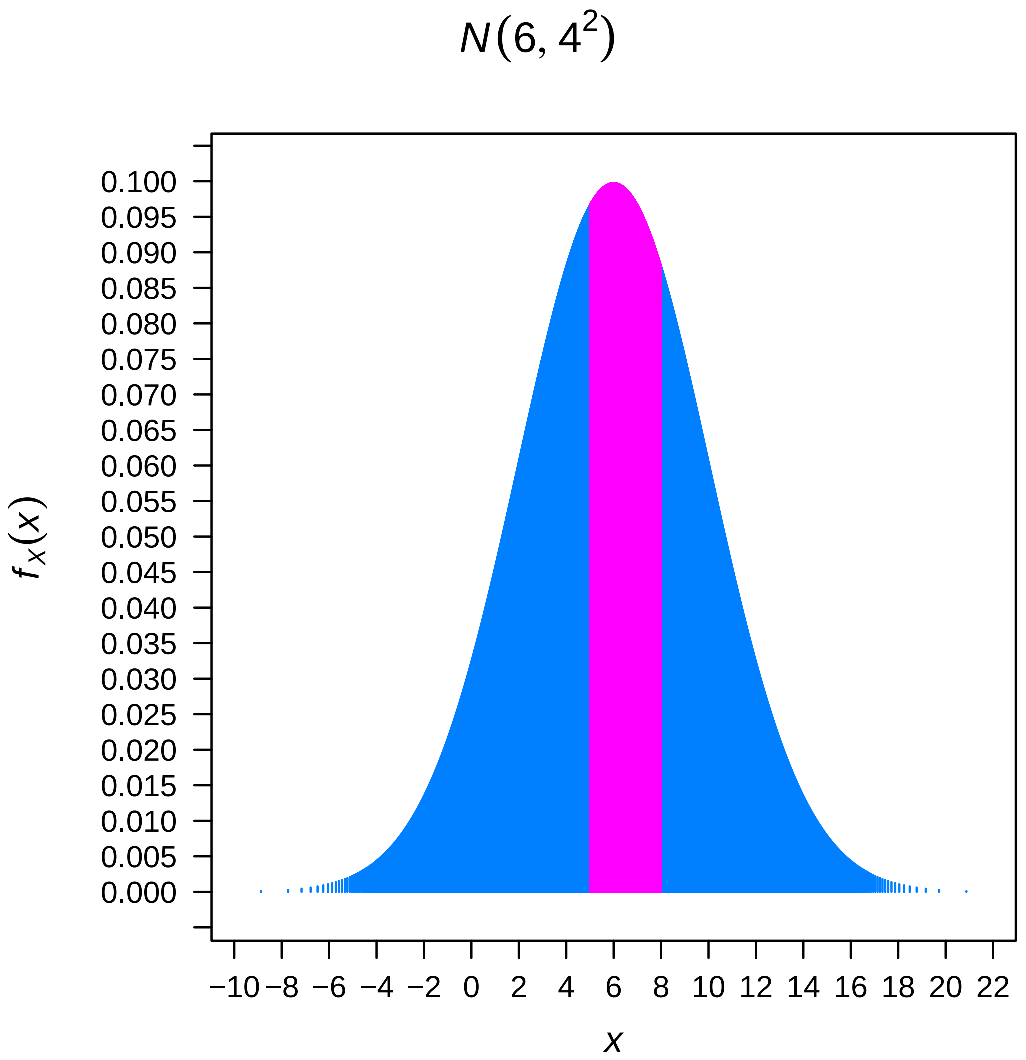
De l'expérience les réviseurs d'une société pensent que les revenues bruts suivent une distribution normale de revenu moyen espéré de 6 mille francs et d'écart-type 4 mille francs. Quelle est la probabilité que le revenu brut se trouve entre 5 et 8 mille francs?

Réponse: Soit X le revenu brut. On a que $X \sim N(\mu = 6,\, \sigma^2 = 4^2)$. Donc

$$\begin{aligned} P(5 < X < 8) &= P\left(\frac{5-6}{4} < \frac{X-6}{4} < \frac{8-6}{4}\right) \\ &= P(-0.25 < Z < 0.5) = \Phi(0.5) - \Phi(-0.25) \\ &= 0.6915 - [1 - \Phi(0.25)] \\ &= 0.6915 - [1 - 0.5987] = 0.2902 \end{aligned}$$

où $\Phi(x) = P(Z \leq x)$.

Illustration:



EXEMPLE

Le coût trimestriel en essence d'une voiture est distribué selon une distribution normale autour d'une moyenne de 850 francs avec un écart-type de 113 francs.

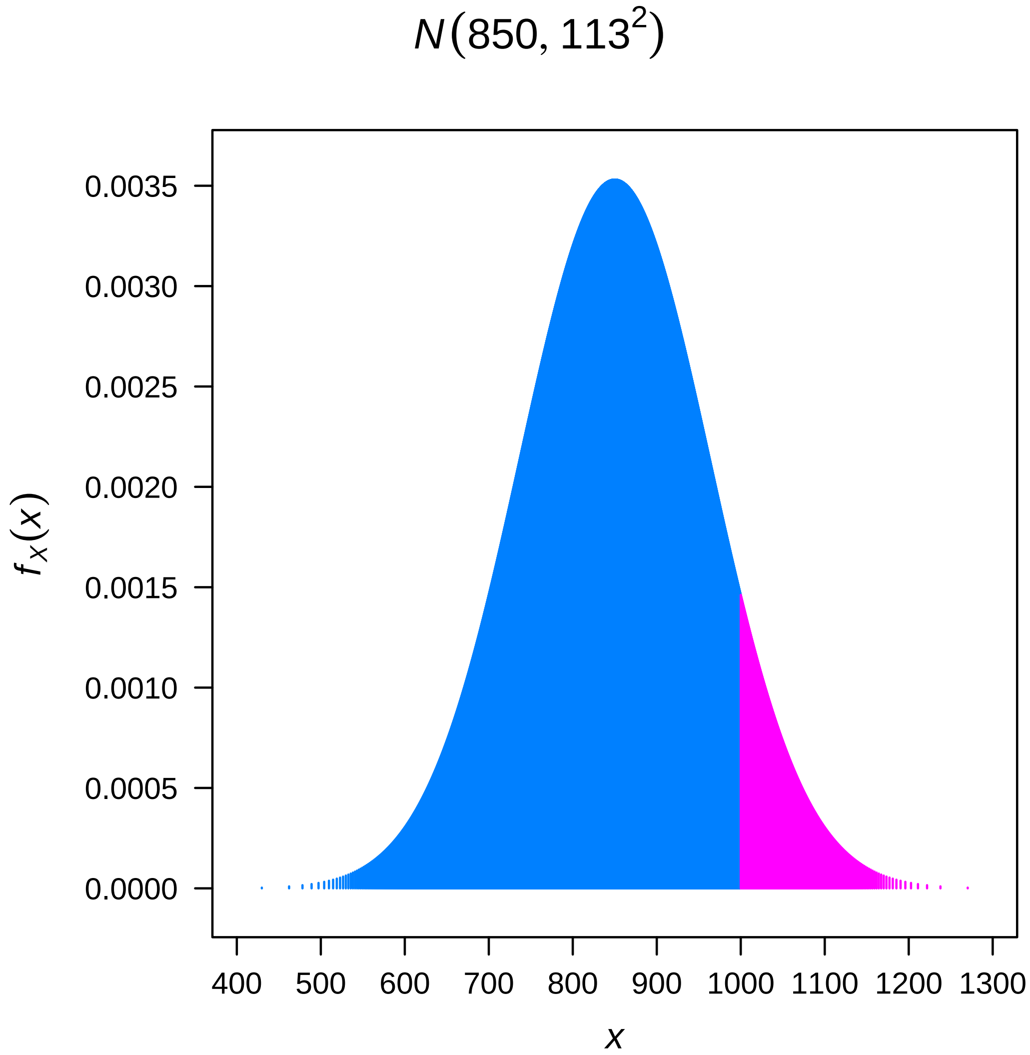
1. Quelle est la probabilité que le coût trimestriel dépasse 1'000 francs?
2. Quel est l'intervalle de coût (le plus petit) contenant 95% des coûts possibles?

Réponse:

1. $X =$ coût trimestriel, $X \sim N(850, 113^2)$.

$$\begin{aligned} P(X > 1'000) &= P\left(\frac{X - 850}{113} > \frac{1'000 - 850}{113}\right) \\ &\approx P(Z > 1.33) = 1 - P(Z \leq 1.33) \\ &\approx 1 - 0.9082 = 0.0918 \approx 9.2\% \end{aligned}$$

Illustration:



2. Le plus petit intervalle est celui qui est symétrique autour de la moyenne μ. Donc on cherche a tel que

$$
\begin{aligned}
P(\mu - a \leq X \leq \mu + a) = 0.95 \quad &\Rightarrow \\
P\left(\tfrac{-a}{\sigma} \leq Z \leq \tfrac{a}{\sigma}\right) = 0.95 \quad &\Rightarrow \\
P\left(Z \leq \tfrac{a}{\sigma}\right) - P\left(Z \leq \tfrac{-a}{\sigma}\right) = 0.95 \quad &\Rightarrow \\
P\left(Z \leq \tfrac{a}{\sigma}\right) - \left[1 - P\left(Z \leq \tfrac{a}{\sigma}\right)\right] = 0.95 \quad &\Rightarrow \\
2 \cdot P\left(Z \leq \tfrac{a}{\sigma}\right) - 1 = 0.95 \quad &\Rightarrow \\
P\left(Z \leq \tfrac{a}{\sigma}\right) = (1 + 0.95)/2 = 0.975&
\end{aligned}
$$

Donc $\frac{a}{\sigma} \approx 1.96 \quad \Rightarrow \quad a \approx 1.96\sigma$.

L'intervalle contenant 95% de la probabilité est donné en général par

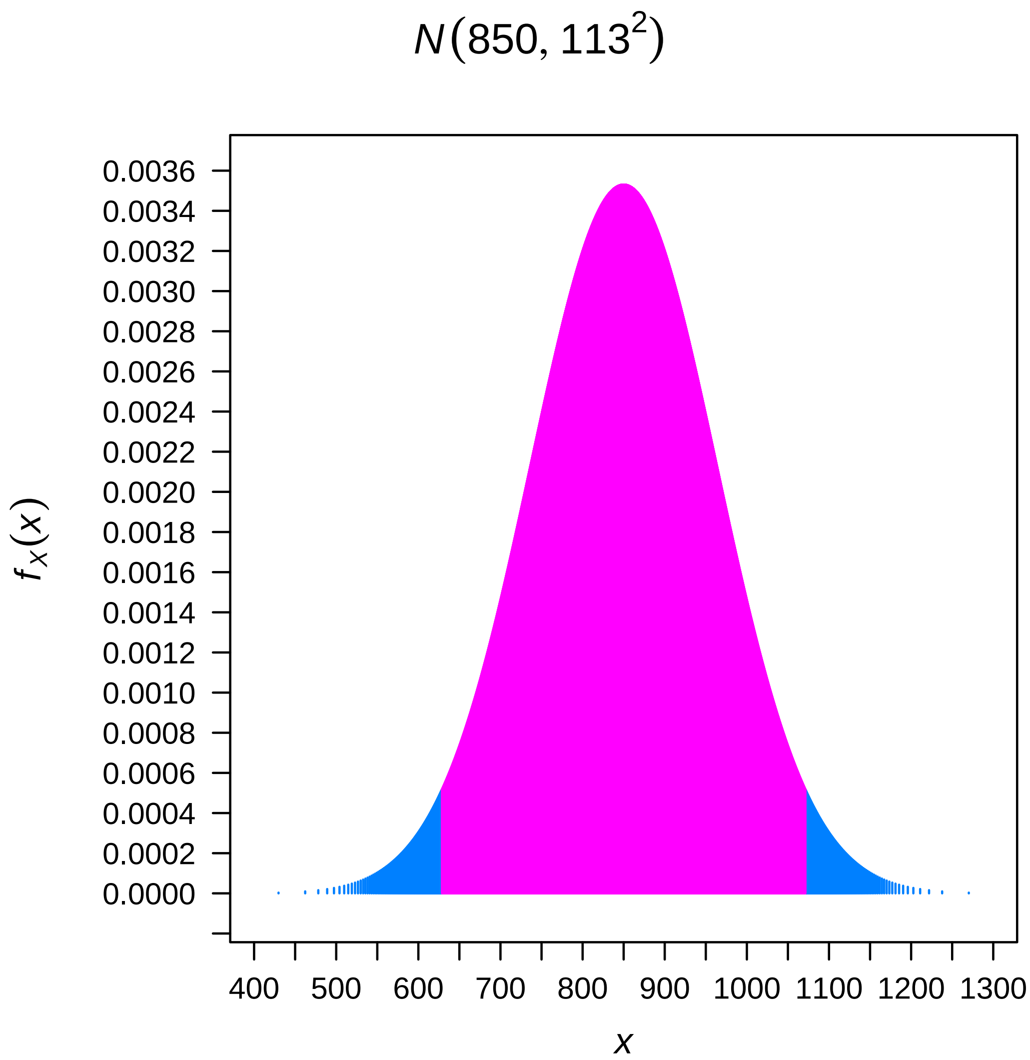
$$\boxed{[\mu - 1.96\sigma,\ \mu + 1.96\sigma]}$$

ou

$$\boxed{\mu \pm 1.96\sigma}$$

Donc dans l'exemple: $850 \pm 1.96 \cdot 113 = [628.52, 1'071.48]$.

Illustration:



Propriété des variables normales

La somme de variables normales indépendantes est aussi une variable normale.

Pour déterminer l'espérance et la variance:

μ : les espérances s'additionnent et se soustraient;

σ^2 : les variances s'additionnent toujours!

EXEMPLE

Bernard et José jouent au golf. La distribution des points de Bernard est $N(92, 36)$ et celle de José est $N(96, 9)$. En sachant que les résultats de José et de Bernard sont indépendants, calculez la probabilité que José ait moins de points que Bernard?

Réponse: On a $B \sim N(92, 36)$ et $J \sim N(96, 9)$. On cherche

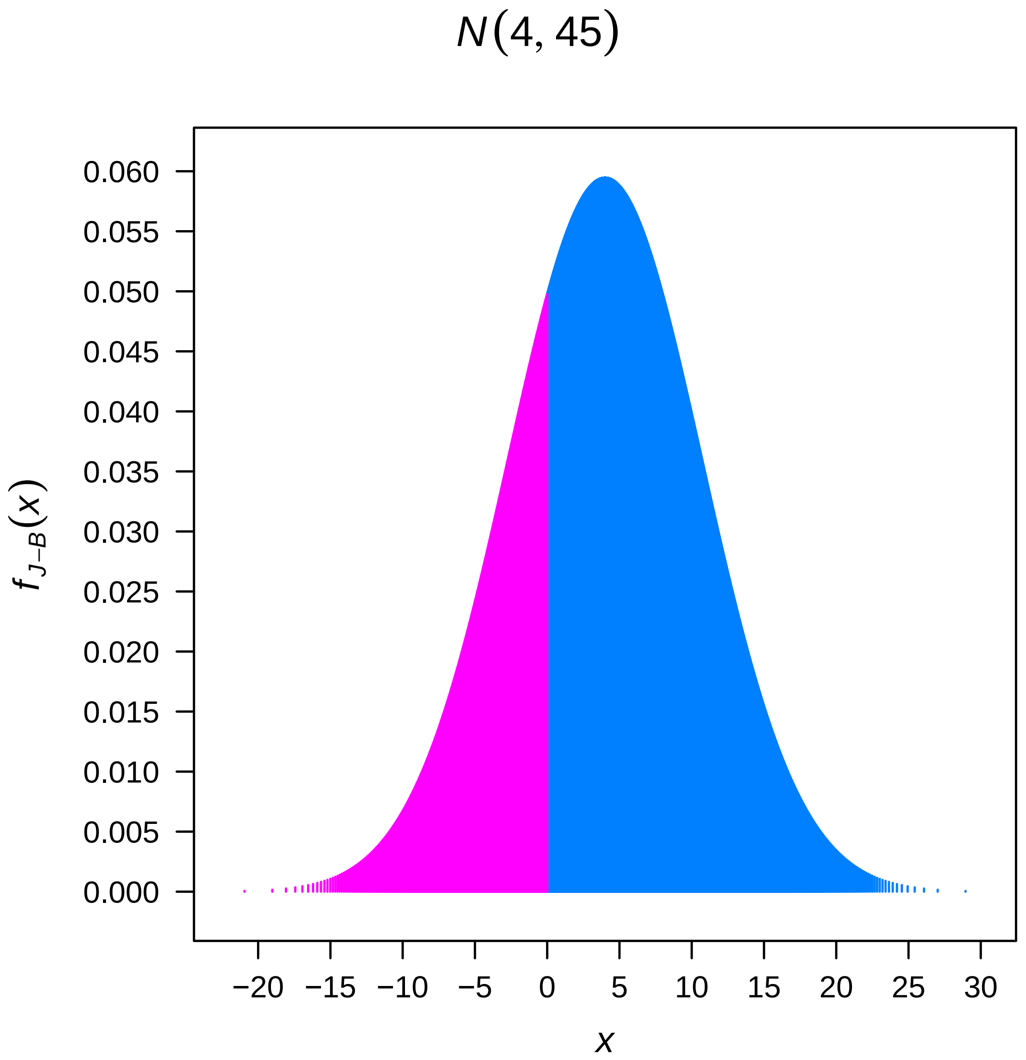
$$P(J < B) = P(J - B < 0).$$

Or $J - B \sim N(96 - 92, 9 + 36)$ ou $J - B \sim N(4, 45)$.

Donc

$$\begin{aligned} P(J - B < 0) &= P\left(\tfrac{J-B-4}{\sqrt{45}} < \tfrac{0-4}{\sqrt{45}}\right) \\ &\approx P(Z < -0.60) = P(Z > 0.60) \\ &= 1 - P(Z \leq 0.60) \approx 1 - 0.7257 = 0.2743. \end{aligned}$$

Illustration:



EXEMPLE

Le poids de caisses de papier suit une distribution normale avec $\mu = 150$kg et $\sigma = 12$kg. 20 caisses indépendantes sont mises dans un ascenseur qui ne fonctionne pas si sa charge dépasse 3'100kg. Quelle est la probabilité que l'ascenseur fonctionne?

Réponse: Soit $W_i =$ le poids de la caisse i avec $W_i \sim N(150, 12^2)$.

Soit $T = W_1 + \cdots + W_{20} = \sum_{i=1}^{20} W_i$ le poids des 20 caisses. On a

$$T \sim N(20 \cdot 150, 20 \cdot 12^2) \quad \text{ou} \quad T \sim N(3'000, 2'880).$$

Donc

$$P(T \leq 3'100) = P\left(\frac{T - 3'000}{\sqrt{2'880}} \leq \frac{3'100 - 3'000}{\sqrt{2'880}}\right) \approx P(Z \leq 1.86) \approx 0.9686.$$

EXEMPLE

Avec ses économies, Alfred décide d'investir en bourse. Il opte pour deux actions A et B. On suppose que la valeur dans 3 mois de l'action A (X_A) est $N(68, 6)$ et celle de l'action B (X_B) est $N(14.5, 2)$ et indépendante de X_A. Alfred étant plutôt prudent, il achète 10 actions A et 25 actions B qui valent aujourd'hui respectivement 64 francs et 13 francs. Quelle est la probabilité que le rendement (trimestriel) de sont portefeuille soit d'au moins 10%? Et celle qu'il soit de moins de 2%?

Réponse: $X = 10X_A + 25X_B$ valeur du portefeuille dans 3 mois, $X \sim N(\mu, \sigma^2)$ avec

$$\begin{aligned} \mu &= \mathsf{E}(X) = 10 \cdot 68 + 25 \cdot 14.5 = 1'042.5, \\ \sigma^2 &= \mathsf{var}(X) = 10^2 \cdot 6 + 25^2 \cdot 2 = 1'850. \end{aligned}$$

On obtient un rendement de 10% si la valeur du portefeuille dans 3 mois est 1.1 fois celle d'aujourd'hui qui vaut $10 \cdot 64 + 25 \cdot 13 = 965$.

▷ On cherche donc

$$\begin{aligned} P(X > 1.1 \cdot 965) &= P(X > 1'061.5) \\ &= 1 - P(X \leq 1'061.5) \\ &= 1 - P\left(Z \leq \frac{1'061.5 - 1'042.5}{\sqrt{1'850}}\right) \\ &= 1 - \Phi(0.44) = 1 - 0.6700 = 0.33. \end{aligned}$$

▷ On cherche aussi

$$\begin{aligned} P(X < 1.02 \cdot 965) &= P(X < 984.3) \\ &= P\left(Z < \frac{984.3 - 1'042.5}{\sqrt{1'850}}\right) \\ &= \Phi(-1.35) = 1 - \Phi(1.35) = 1 - 0.9115 = 0.0885. \end{aligned}$$

Approximations normales

Soit $R \sim B(n, p)$.

Si n est ‘grand’ (et p n’est pas ‘proche’ de 0 ou de 1), on a approximativement

$$R \sim N(\mu, \sigma^2) \quad \text{avec} \quad \mu = np \quad \text{et} \quad \sigma^2 = np(1-p).$$

EXEMPLE

Une pièce est lancée 300 fois. Quelle est la probabilité d'obtenir moins de 160 piles?

Réponse:

$R =$ nombre de piles sur $n = 300$ lancers.

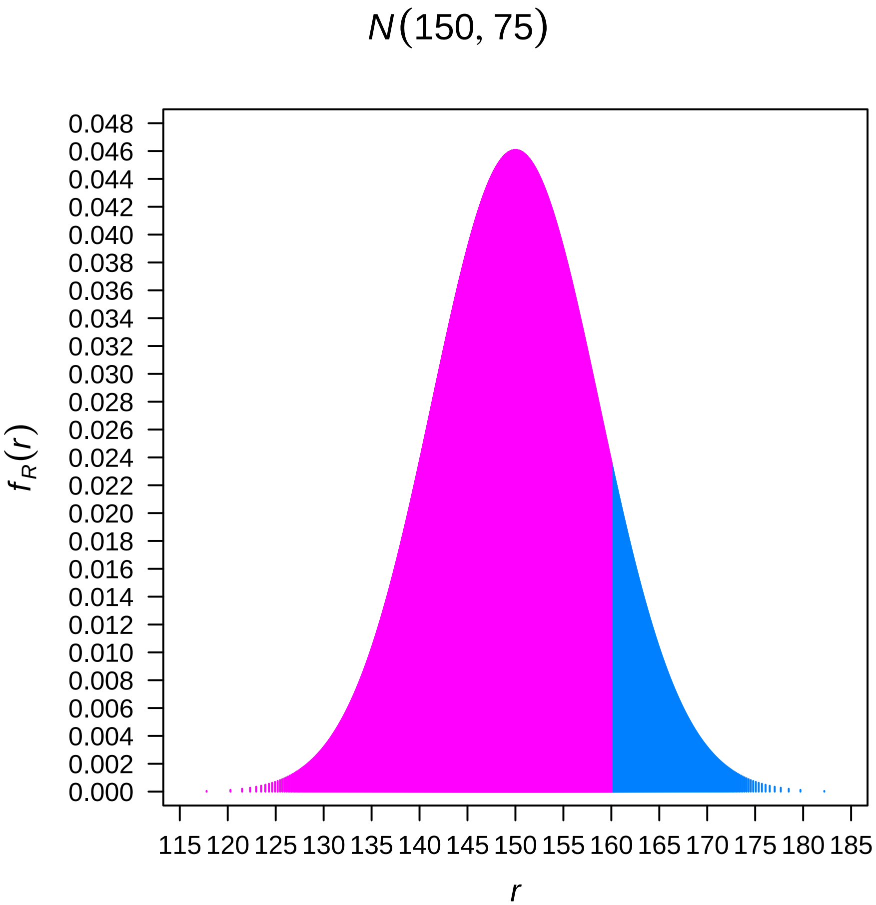
$R \sim B(300, 0.5)$.

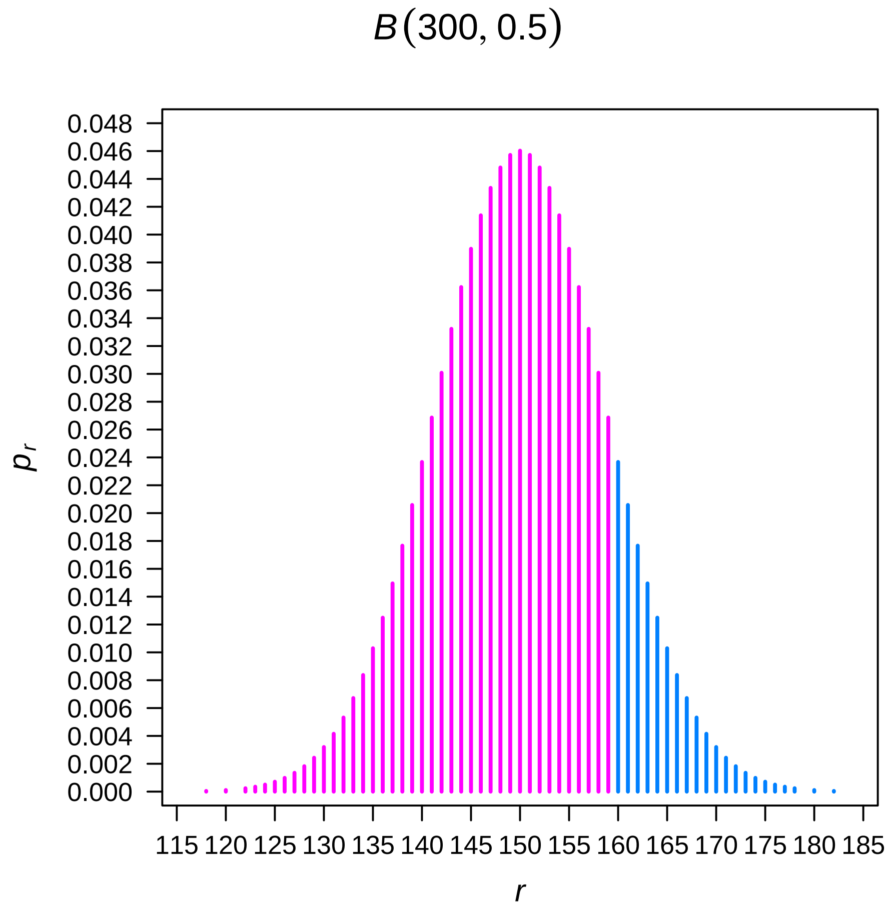
Mais n est grand et $p = 0.5$, alors on a approximativement $R \sim N(150, 75)$.

Donc

$$P(R < 160) = P\left(\frac{R-150}{\sqrt{75}} < \frac{160-150}{\sqrt{75}}\right) \approx P(Z < 1.15) \approx 0.8749.$$

Illustrations:





Correction de continuité

Si R est discrète $\Rightarrow$ $P(R < 60) = P(R \leq 59)$

Si R est continue $\Rightarrow$ $P(R < 60) \neq P(R \leq 59)$

Règles:

Discrète	Continue
$P(R < a)$	$P(R \leq a - 0.5)$
$P(R \leq a)$	$P(R \leq a + 0.5)$
$P(R > a)$	$P(R > a + 0.5)$
$P(R \geq a)$	$P(R > a - 0.5)$

EXEMPLE

$$R \sim B(100, 0.5) \quad \Rightarrow \quad R \sim N(50, 25) \ \text{ approximativement}$$

$$P(R < 60) \quad \Rightarrow \quad P(R \leq 59.5) = P(Z \leq 1.9) \approx 0.9713$$

Remarque: La correction n'est pas nécessaire lorsque $n > 200$.


ARE YOU NORMAL?
ENRICO

4.10 La distribution log-normale

Soit X une variable aléatoire normale quelconque d'espérance μ et de variance σ^2:

$$X \sim N(\mu, \sigma^2).$$

Alors la variable aléatoire

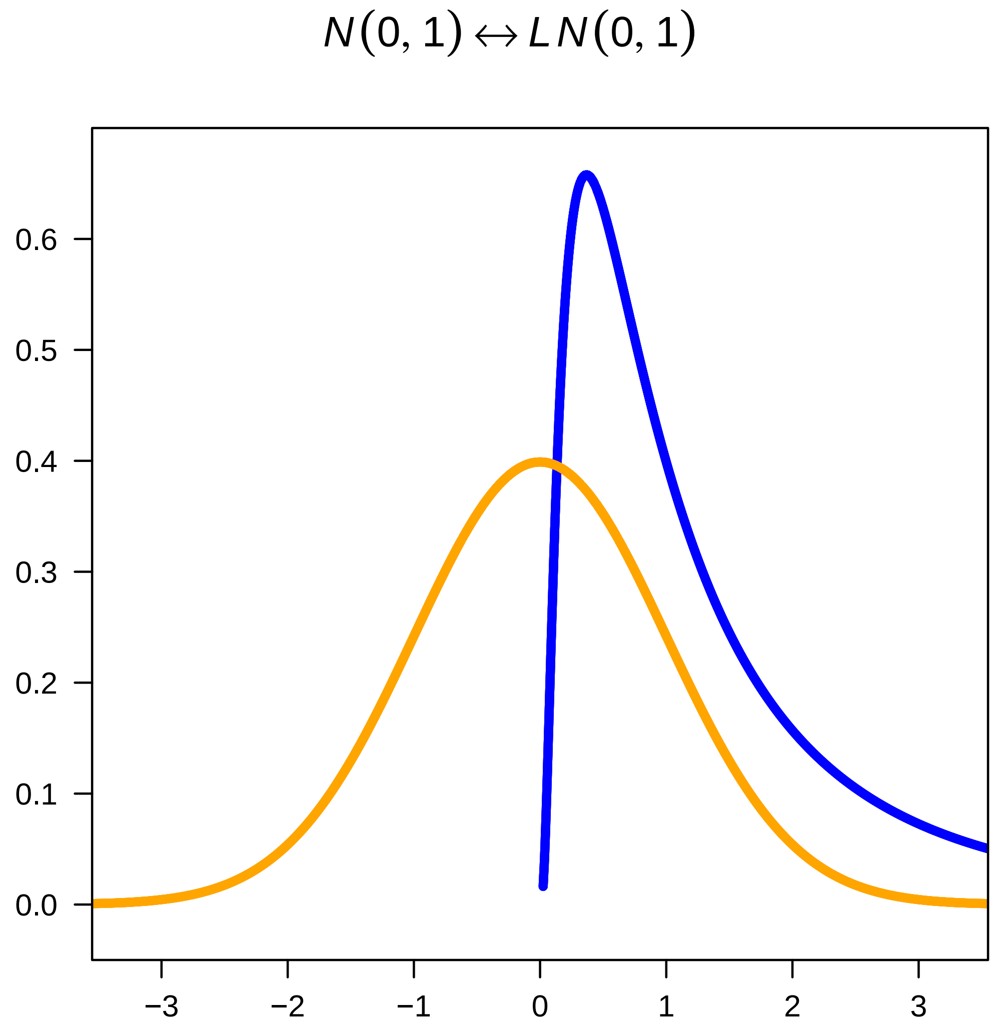
$$\boxed{Y = \exp(X)}$$

suit une distribution dite **log-normale** de paramètres μ et σ^2, puisque $\ln(Y)$ a une distribution normale.

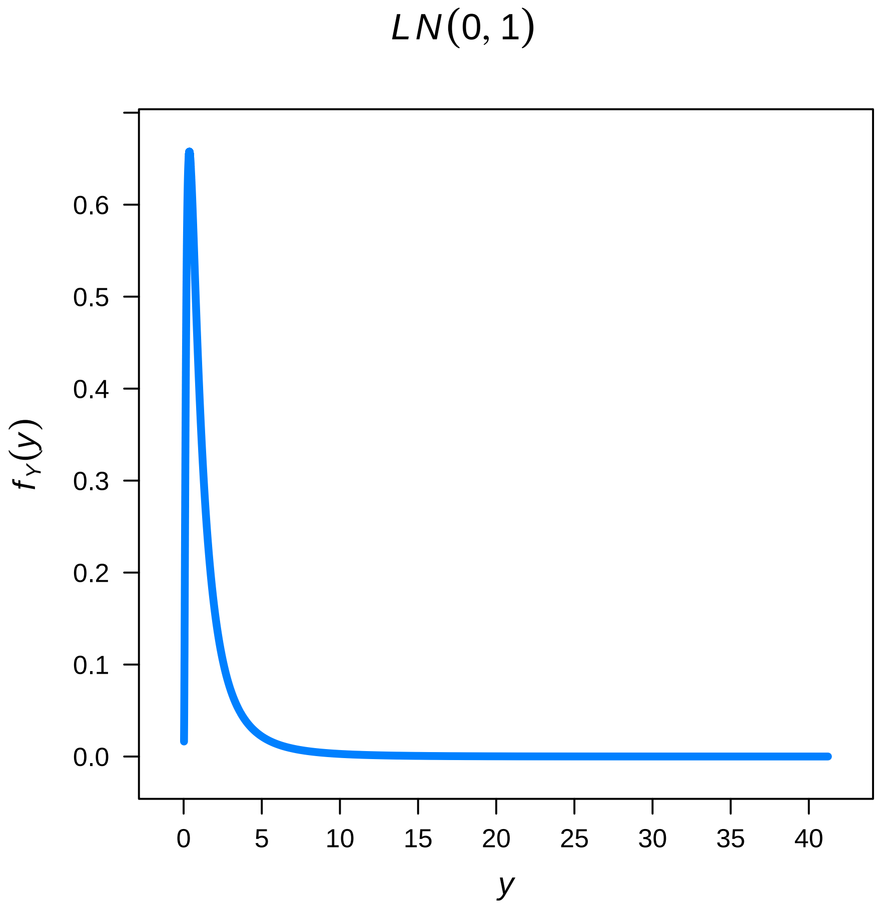
$\rightsquigarrow$ **Notation:** $\boxed{Y \sim LN(\mu, \sigma^2)}$.

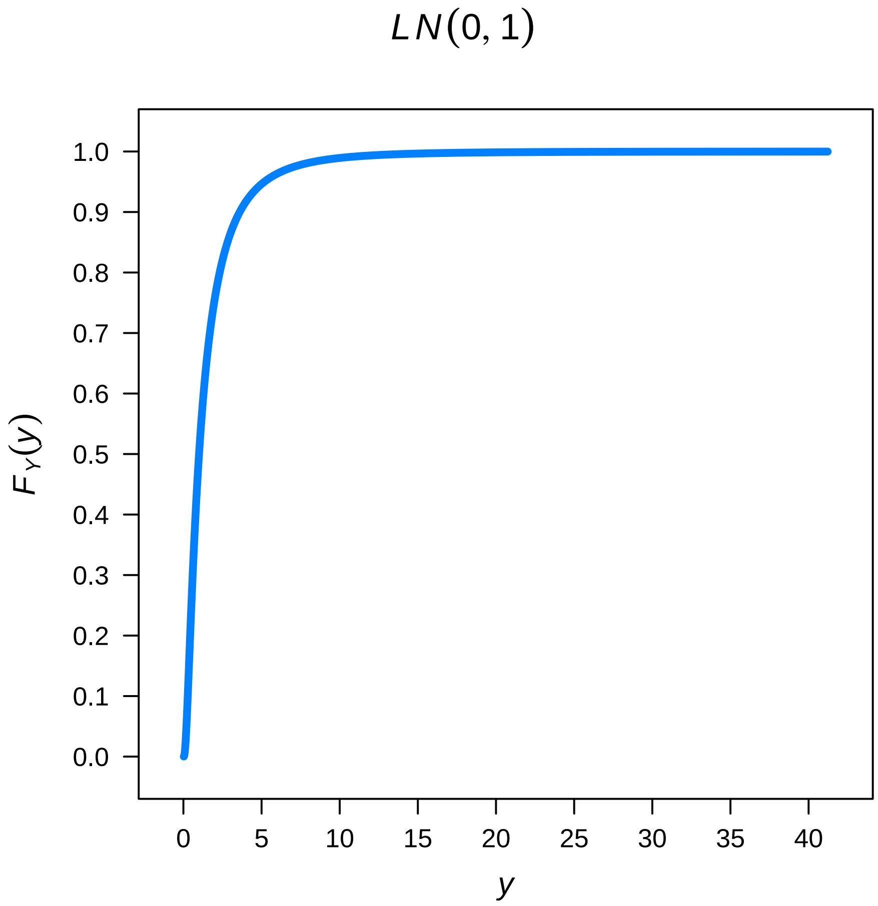
$\triangleright$ **Propriétés:** $\mathsf{E}(Y) = \exp(\mu + \sigma^2/2)$ et $\mathsf{var}(Y) = \exp(2\mu + \sigma^2)\left[\exp(\sigma^2) - 1\right]$.

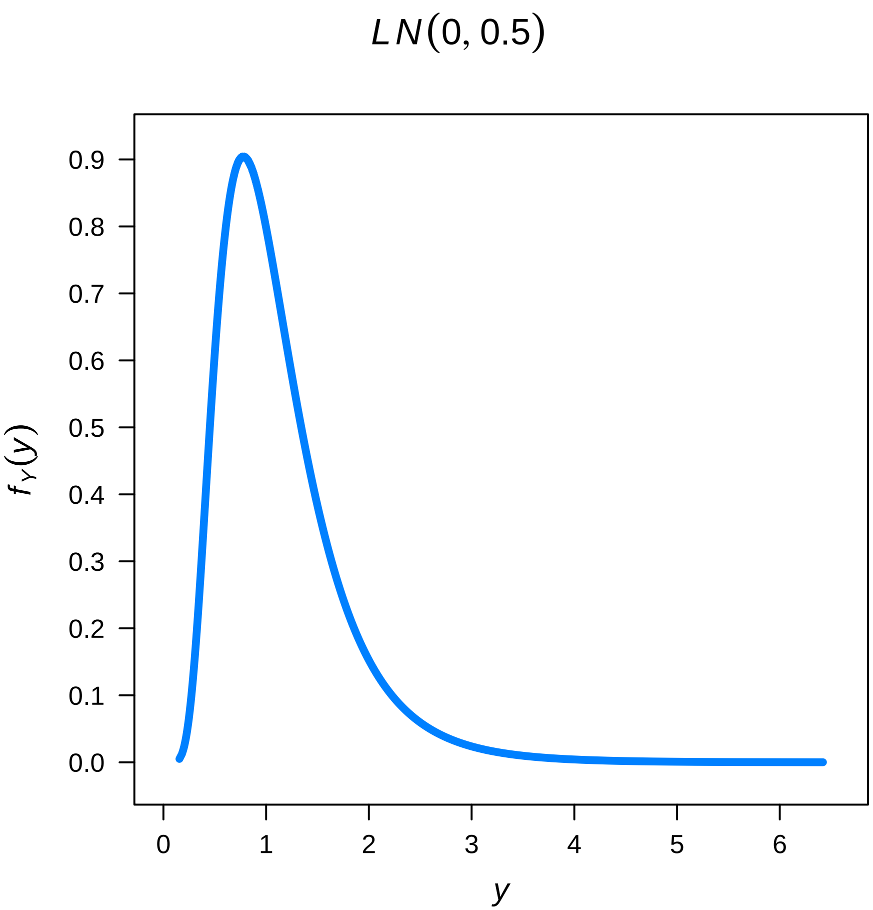
Illustration:

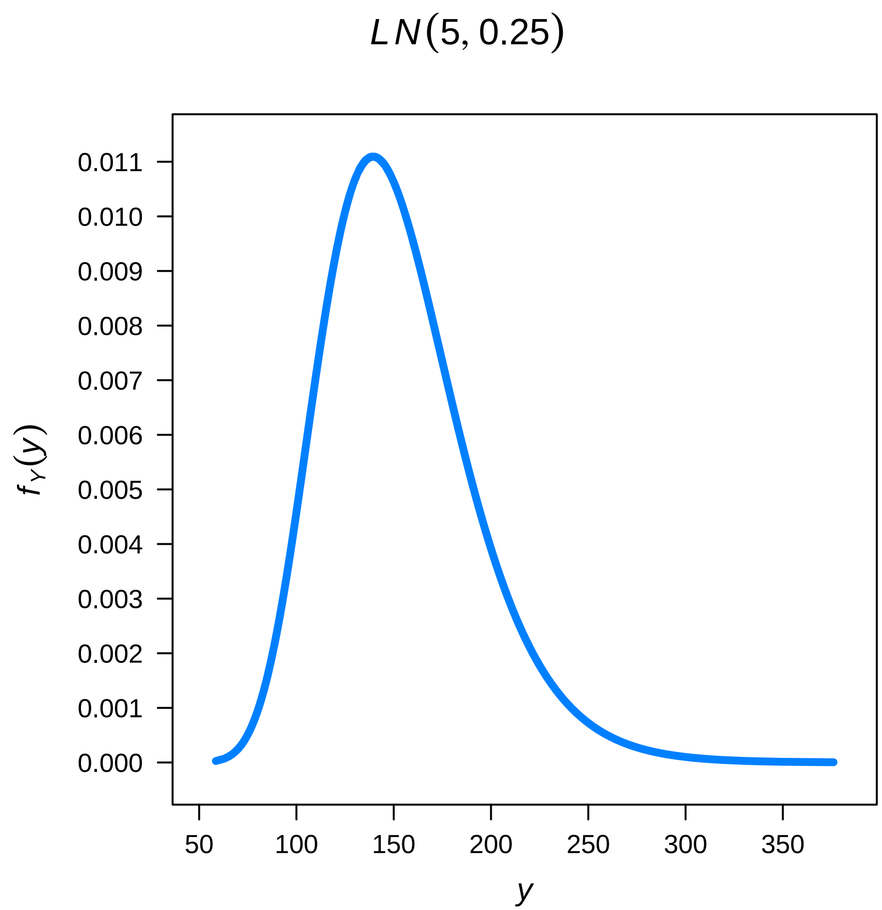


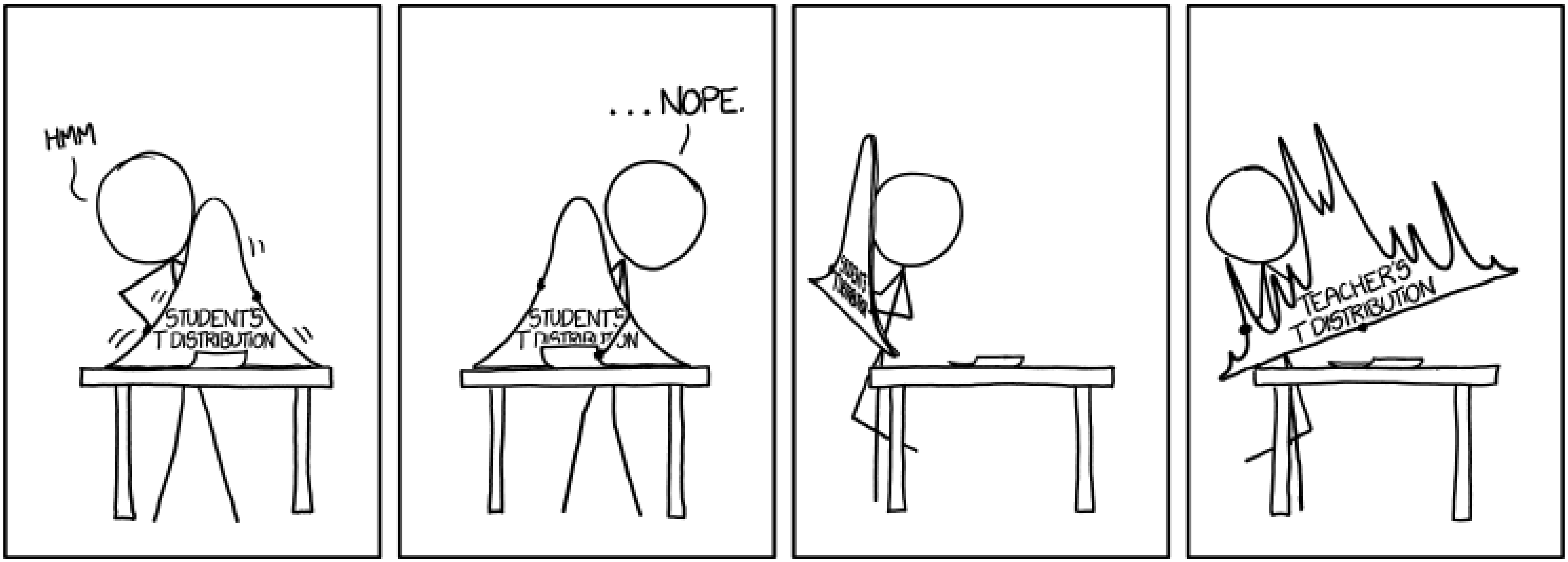
Autres illustrations:






$LN(0, 0.5)$
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
$f_Y(y)$
0
1
2
3
4
5
6
y


$LN(5, 0.25)$
0.011
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
$f_Y(y)$
50
100
150
200
250
300
350
y


HMM
STUDENT'S
T DISTRIBUTION
...NOPE.
STUDENT'S
TEACHER'S
T DISTRIBUTION

4.11 La distribution de Student

Soient Z et Y deux variables aléatoires indépendantes telles que $Z \sim N(0,1)$ et $Y \sim \chi^2_\nu$ (*i.e.* distribution chi-carré à ν dégrés de liberté, avec ν un entier non nul, voir remarque b) de la page 263) alors

$$\boxed{T = \frac{Z}{\sqrt{Y/\nu}}}$$

suit une distribution dite de **Student** (ou distribution t) de paramètre ν, appelé le nombre de dégrés de liberté.

⇝ **Notation:** $\boxed{T \sim t_\nu}$.

▷ **Propriétés:** $\mathrm{E}(T) = 0$, pour $\nu > 1$, et $\mathrm{var}(T) = \nu/(\nu - 2)$, pour $\nu > 2$.

Remarques:

a) La distribution de Student a été publié en 1908 par William Gosset pendant qu'il travaillait à la brasserie Guinness à Dublin. Il lui était interdit de publier sous son propre nom, c'est pour cette raison qu'il publia sous le pseudonyme de Student;

$\leadsto$ Plaque au '*Guinness Storehouse*' à Dublin commémorant Gosset:

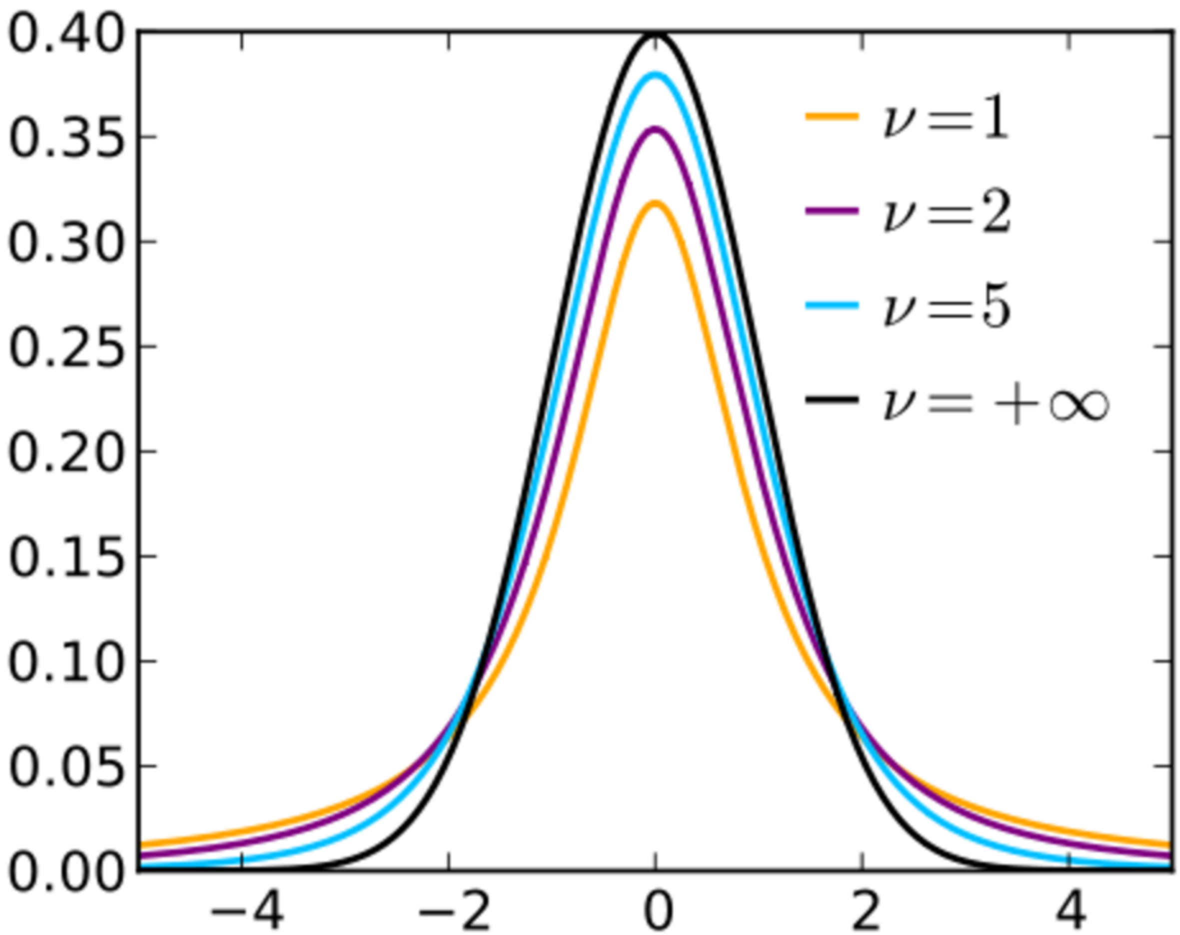
b) Plus le dégré de liberté ν augmente, plus la fonction de densité de la distribution de Student t_ν tend à se rapprocher de celle de la distribution normale standard $N(0,1)$;

c) Soient n variables aléatoires $Z_1, \ldots, Z_n$ normales standard et indépendantes: La variable $Y = \sum_{i=1}^{n} Z_i^2$ suit une distribution chi-carré à n degrés de liberté, *i.e.* χ_n^2;

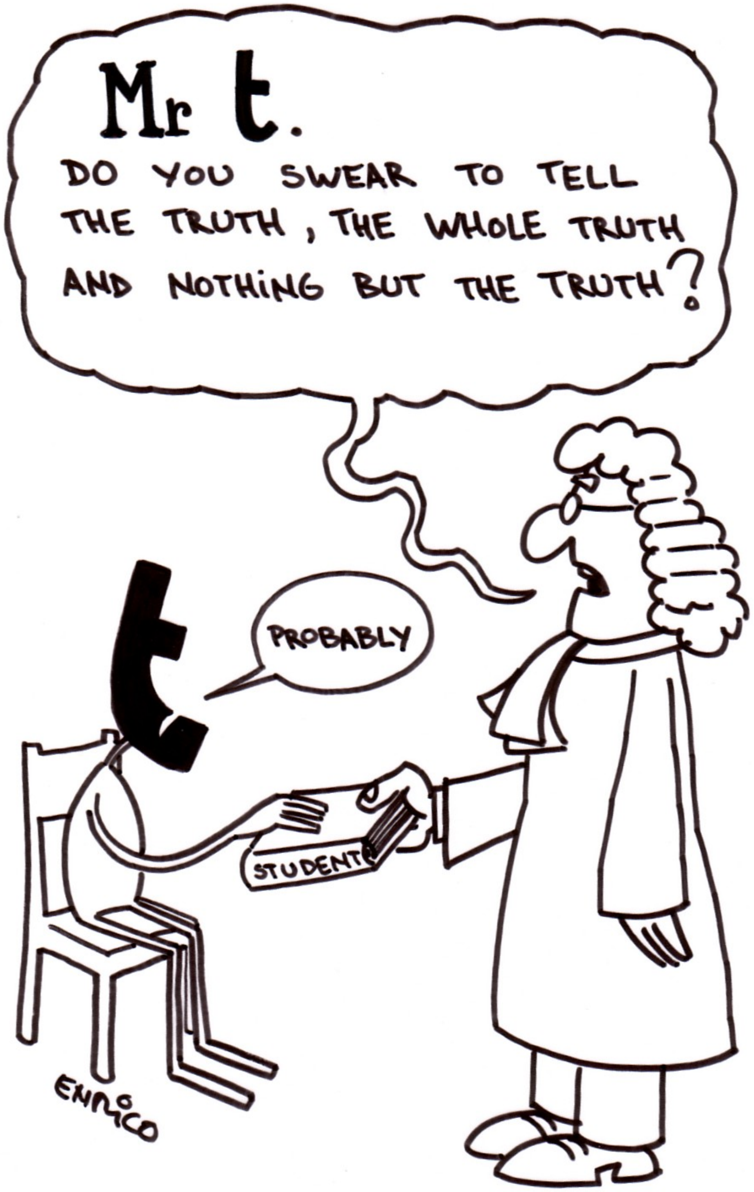
d) Si $Y \sim \chi_n^2$ on a $\mathsf{E}(Y) = n$ et $\mathsf{var}(Y) = 2n$;

e) Si $Y_1 \sim \chi_n^2$ et $Y_2 \sim \chi_m^2$ et indépendantes alors $Y_1 + Y_2 \sim \chi_{n+m}^2$.

Illustrations de fonctions de densité t_ν (l'axe horizontal symbolise t):



Source: `en.wikipedia.org/wiki/Student's_t-distribution` *(soumis à la licence CC-BY-SA 3.0).*


Mr t.
DO YOU SWEAR TO TELL THE TRUTH, THE WHOLE TRUTH AND NOTHING BUT THE TRUTH?
PROBABLY
STUDENT
ENRICO

4.12 La distribution Fisher–Snedecor

Soient Y_1 et Y_2 deux variables aléatoires indépendantes telles que $Y_1 \sim \chi^2_{d_1}$ et $Y_2 \sim \chi^2_{d_2}$ alors

$$\boxed{F = \frac{Y_1/d_1}{Y_2/d_2}}$$

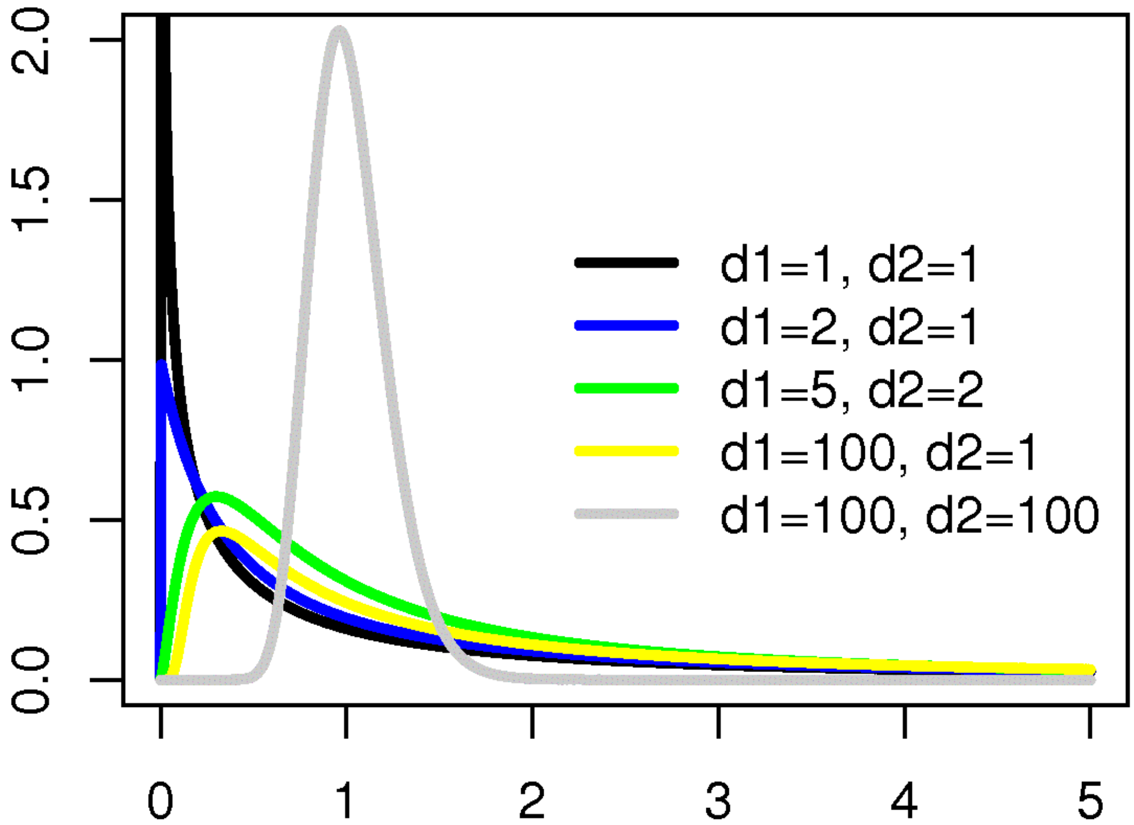
suit une distribution dite de **Fisher–Snedecor** (ou distribution de Fisher ou distribution F) de paramètres (d_1, d_2), appelés les nombres de dégrés de liberté.

$\rightsquigarrow$ **Notation:** $\boxed{F \sim F_{d_1,d_2}}$.

$\triangleright$ **Propriétés:** $\mathsf{E}(F) = d_2/(d_2 - 2)$, pour $d_2 > 2$, et

$$\mathsf{var}(F) = \frac{2d_2^2(d_1 + d_2 - 2)}{d_1(d_2 - 2)^2(d_2 - 4)}, \text{ pour } d_2 > 4.$$

Illustrations de fonctions de densité F_{d_1,d_2}:



Source: `fr.wikipedia.org/wiki/Loi_de_Fisher` *(soumis à la licence CC-BY-SA 3.0).*

4.13 Distribution d'une fonction de variable aléatoire

• Il arrive souvent que la distribution d'une variable aléatoire X soit connue mais que l'on s'intéresse plutôt à celle d'une fonction de cette variable $g(X) = Y$.

$\rightsquigarrow$ Pour y parvenir, il faut exprimer l'événement '$g(X) \leq y$' sous forme d'une condition où X appartient à un certain ensemble.

$\triangleright$ Remarque: Dans ce cours on va s'intéresser qu'au cas continue!

EXEMPLE

Pour toute variable aléatoire continue X de densité f_X, on obtient la fonction de répartition de $Y = X^2$ de la manière suivante: Pour $y \geq 0$,

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(X^2 \leq y) \\
&= P\left(-\sqrt{y} \leq X \leq \sqrt{y}\right) \\
&= F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right).
\end{aligned}$$

Une dérivation livre la densité, pour $y \geq 0$,

$$f_Y(y) = \frac{1}{2\sqrt{y}} \left[f_X\left(\sqrt{y}\right) + f_X\left(-\sqrt{y}\right) \right].$$

EXEMPLE

Soit X de densité f_X. La densité de $Y = |X|$ peut être calculée ainsi: Pour $y \geq 0$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(|X| \leq y) \\ &= P\left(-y \leq X \leq y\right) \\ &= F_X\left(y\right) - F_X\left(-y\right), \end{aligned}$$

et par dérivation on obtient, pour $y \geq 0$,

$$f_Y(y) = f_X\left(y\right) + f_X\left(-y\right).$$

Théorème (sans démonstration)

Soit X une variable aléatoire continue de densité f_X. Soit g une fonction strictement monotone (croissante ou décroissante) et dérivable, donc continue.

La densité de la variable aléatoire $Y = g(X)$ est alors

$$f_Y(y) = \begin{cases} \left|\frac{\partial}{\partial y} g^{-1}(y)\right| \cdot f_X\left[g^{-1}(y)\right] & \text{si } y = g(x) \text{ pour un } x \text{ quelconque} \\ 0 & \text{si } y \neq g(x) \text{ pour tout } x \end{cases}$$

où $g^{-1}(y)$ est défini comme étant égal à x tel que $g(x) = y$.

EXEMPLE

Soit X une variable aléatoire continue non négative de densité f_X et soit $Y = X^n$. Trouver la densité f_Y de Y.

Réponse: Si $y = g(x) = x^n$ alors

$$x = g^{-1}(y) = y^{1/n}$$

et

$$\frac{\partial}{\partial y} g^{-1}(y) = \frac{1}{n}\, y^{\frac{1}{n}-1}.$$

Donc, d'après le théorème précédent, on obtient

$$f_Y(y) = \frac{1}{n}\, y^{\frac{1}{n}-1} \cdot f_X\left(y^{1/n}\right).$$

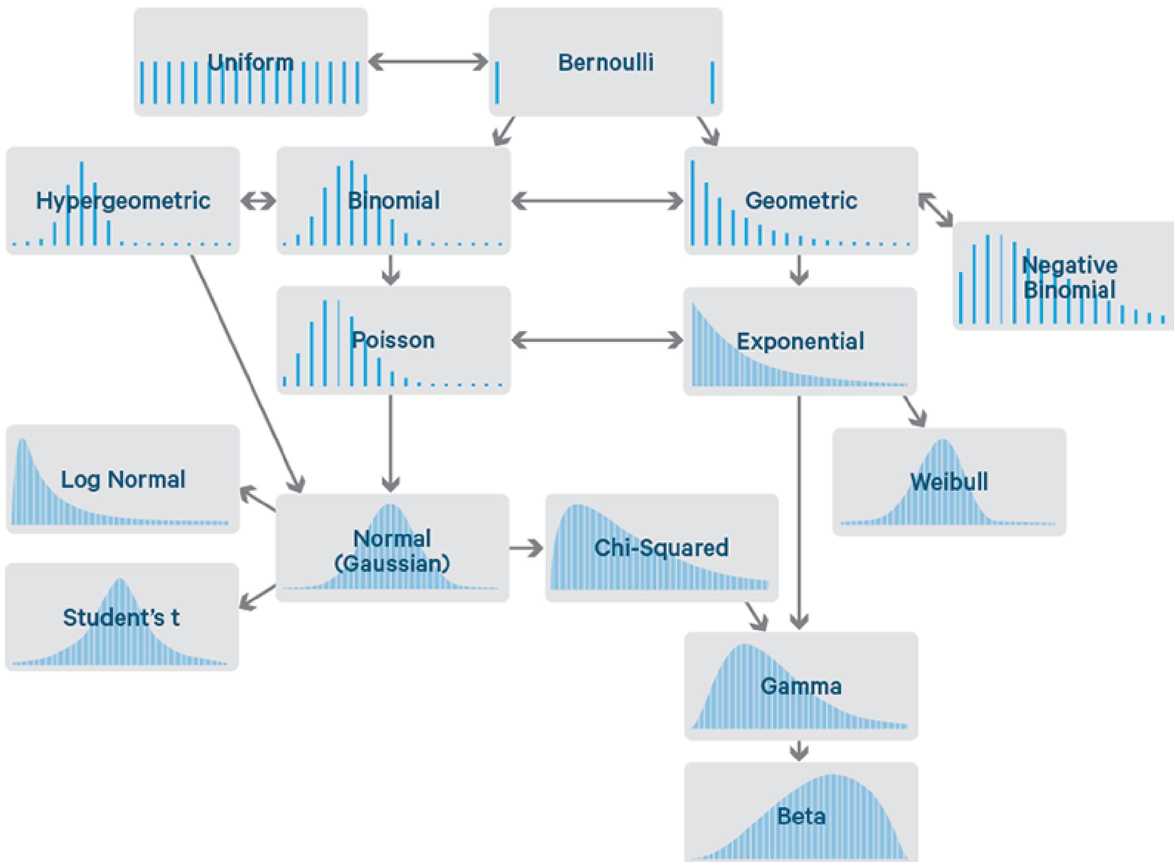
Si $n = 2$, cela donne

$$f_Y(y) = \frac{1}{2}\, y^{\frac{1}{2}-1} \cdot f_X\left(y^{1/2}\right) = \frac{1}{2\sqrt{y}}\, f_X\left(\sqrt{y}\right)$$

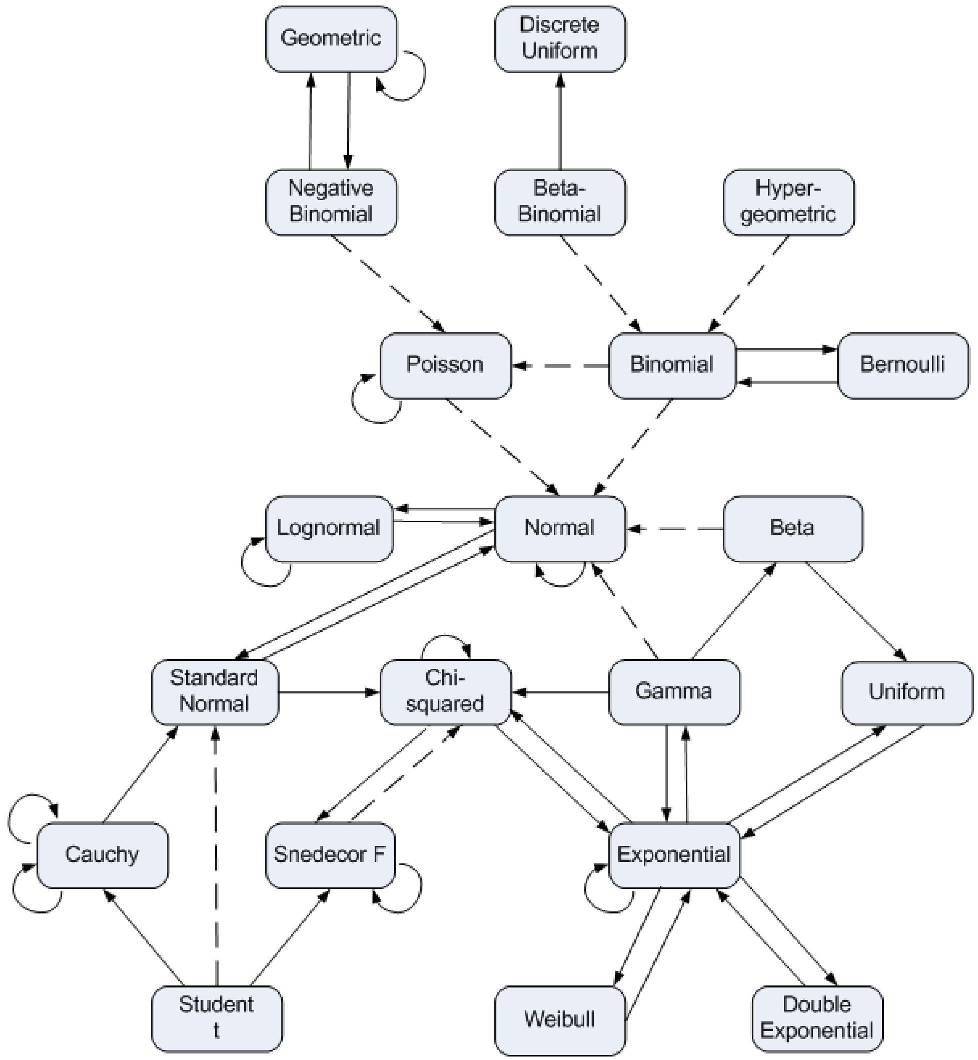
ce qui correspond, puisque $X \geq 0$, au résultat de la page 341.

5. Relations entre distributions de probabilité

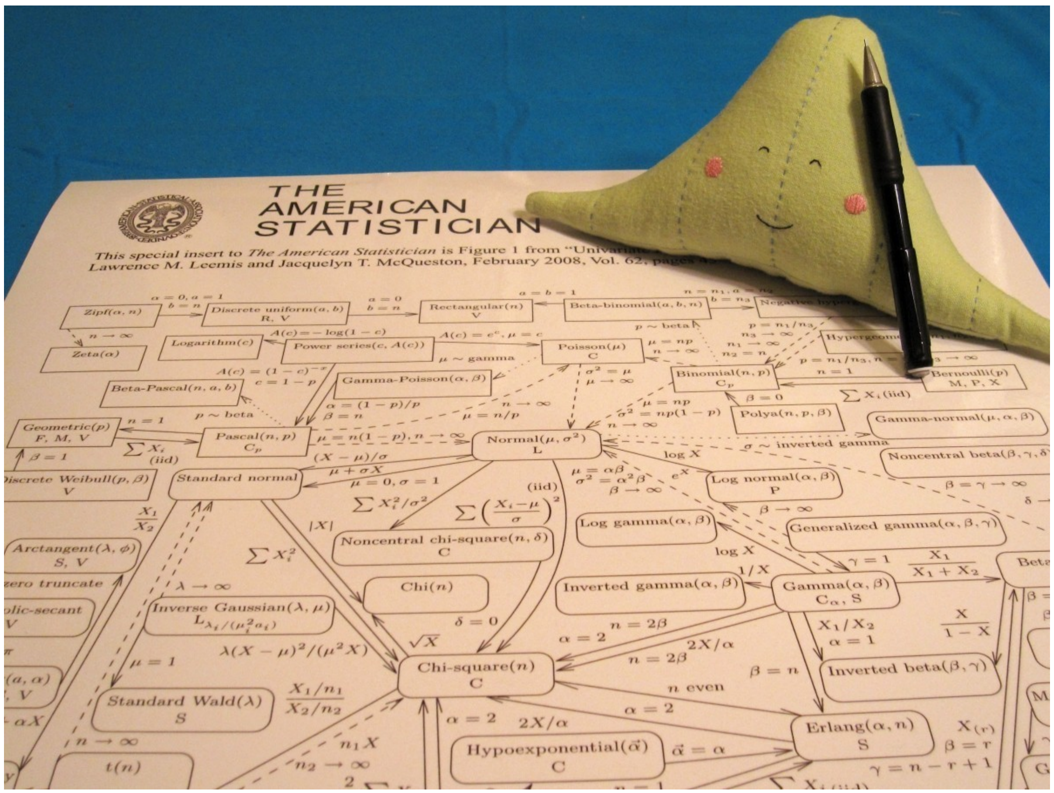
▷ '*Common probability distributions: the data scientist's crib sheet*' (goo.gl/NJRlXn):



▷ '*Diagram of distribution relationships*' (version interative sur goo.gl/qqIzzS):



▷ '*Univariate distribution relationships*' (version interative sur $\boxed{\text{goo.gl/XOBIkh}}$):



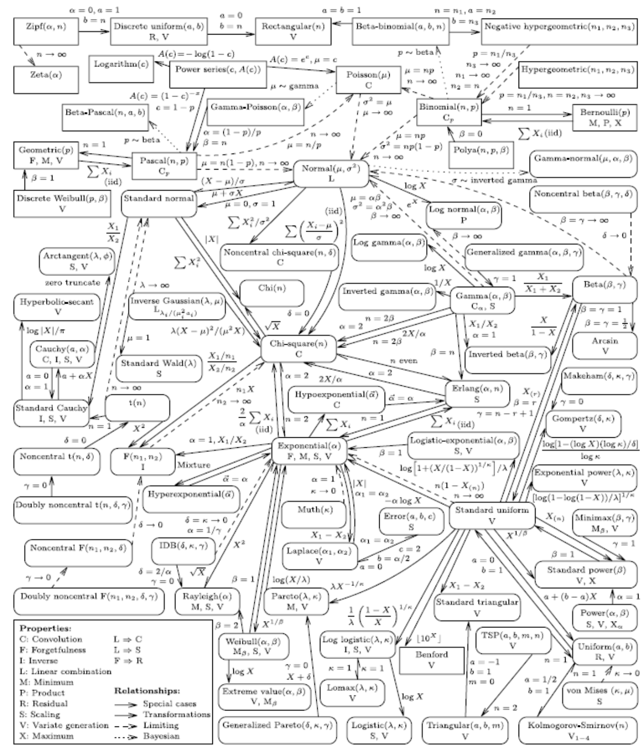


Figure 1. Univariate distribution relationships.

6. Théorèmes limites

• Les théorèmes limites constituent les **résultats théoriques** les plus importantes des probabilités.

⇝ Parmi eux, les principaux sont répertoriés sous deux dénominations:

▷ $\boxed{\text{Lois des grands nombres}}$ d'une part, et

▷ $\boxed{\text{Théorèmes centraux limites}}$ d'autre part.

⇝ Le théorème central limite est l'un des plus remarquables résultats de la théorie des probabilités!

6.1 Inégalités de Markov et de Tchebychev

◇ Nous allons commencer par établir l'inégalité appelée **inégalité de Markov**:

Soit X une variable aléatoire à valeurs non négatives. Pour tout réel $a > 0$ on a

$$P(X \geq a) \leq \frac{\mathsf{E}(X)}{a}.$$

Démonstration: Pour $a > 0$, soit

$$I = \begin{cases} 1 & \text{si } X \geq a \\ 0 & \text{sinon} \end{cases}$$

et noter que, puisque $X > 0$,

$$I \leq \frac{X}{a}.$$

Ainsi on obtient

$$\mathsf{E}(I) \leq \frac{\mathsf{E}(X)}{a}$$

ce qui, comme

$$\mathsf{E}(I) = 1 \cdot P(X \geq a) + 0 \cdot P(X < a) = P(X \geq a),$$

prouve le résultat.

◇ L'inégalité suivante — appelée **inégalité de Tchebychev** — est un corollaire de l'inégalité de Markov:

Soit X une variable aléatoire d'espérance μ et de variance σ^2 finies. Pour tout réel $k > 0$ on a

$$P(\mid X - \mu \mid \geq k) \leq \frac{\sigma^2}{k^2}.$$

Démonstration: On peut appliquer l'inégalité de Markov avec $a = k^2$ à la variable $(X-\mu)^2$ puisque celle-ci est à valeurs non négatives.

On obtient

$$P\left((X-\mu)^2 \geq k^2\right) \leq \frac{\mathsf{E}\left[(X-\mu)^2\right]}{k^2}.$$

Mais comme $(X-\mu)^2 \geq k^2$ équivaut à $\mid X-\mu \mid \geq k$, on peut finalement écrire

$$P(\mid X-\mu \mid \geq k) \leq \frac{\mathsf{E}\left[(X-\mu)^2\right]}{k^2} = \frac{\sigma^2}{k^2}$$

ce qui achève la démonstration.

Remarque: L'importance des inégalités de Markov et de Tchebychev réside en ce qu'elles **permettent de borner la valeur de certaines probabilités** là où seule l'espérance de la distribution est connue, plus éventuellement sa variance.

$\rightsquigarrow$ Il est évident que, si la distribution elle-même est connue, on ne recourra pas à des bornes, puisque la valeur exacte des ces probabilités est calculable.

EXEMPLE

On suppose que le nombre de pièces sortant d'une usine donnée en l'espace d'une semaine est une variable aléatoire d'espérance 50.

1. Peut-on estimer la probabilité que la production de la semaine prochaine soit plus grande ou égale à 75 pièces?
2. On sait de plus que la variance de la production hebdomadaire est de 25. Peut-on estimer la probabilité que la production de la semaine à venir soit comprise entre 40 et 60?

Réponse: Désignons par X le nombre de pièces produites en une semaine.

1. L'inégalité de Markov donne

$$P(X \geq 75) \leq \frac{\mathsf{E}(X)}{75} = \frac{50}{75} = \frac{2}{3}.$$

2. Comme 'entre 40 et 60' correspond à '50 ± 10', l'inégalité de Tchebychev donne

$$P(\mid X - 50 \mid \geq 10) \leq \frac{\sigma^2}{10^2} = \frac{25}{100} = \frac{1}{4}$$

et donc

$$P(\mid X - 50 \mid < 10) = 1 - P(\mid X - 50 \mid \geq 10) \geq 1 - \frac{1}{4} = \frac{3}{4}.$$

La probabilité que la production de la semaine à venir se situe entre 40 et 60 pièces est donc au moins 0.75.

Remarque: L'inégalité de Tchebychev étant valable pour n'importe quelle distribution de la variable X, il ne faut pas s'attendre à ce que la borne qu'elle fournit soit très proche de la probabilité exacte dans la majorité des cas.

EXEMPLE

Soit $X \sim U(0, 10)$. Ainsi $\mathrm{E}(X) = \frac{10+0}{2} = 5$ et $\mathrm{var}(X) = \frac{(10-0)^2}{12} = \frac{25}{3}$.

L'inégalité de Tchebychev pour $k = 4$ donne

$$P(\mid X - 5 \mid \geq \ 4) \ \leq \ \frac{\frac{25}{3}}{4^2} = \frac{25}{3 \cdot 16} \approx 0.52$$

alors que le résultat exact (sans démonstration) est

$$P(\mid X - 5 \mid \geq \ 4) = 0.20.$$

On voit bien que si l'inégalité de Tchebychev est fondée, la borne qu'elle fournit est ici loin d'être proche de la probabilité exacte.

6.2 Loi faible des grands nombres (version simple)

Approche expérimentale

Considérons l'expérience de jeter une pièce de monnaie $10'000$ fois et observons le nombre de pile obtenues.

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de Bernoulli, alors

$$X_i = \begin{cases} 1 & \text{si le } i\text{-ème jet donne pile} \\ 0 & \text{si le } i\text{-ème jet donne face} \end{cases}$$

Donc $S_n = X_1 + \cdots + X_n$ représente le nombre de pile sur n jets (essais).

Si p est la probabilité d'obtenir pile ('succès'), alors on a que

$$S_n \sim B(n, p).$$

La proportion de pile sur n jets n'est rien d'autre que

$$\frac{S_n}{n} = \overline{X}.$$

Donc

$$\begin{aligned}
\mathsf{E}(\overline{X}) &= \frac{1}{n} \cdot \mathsf{E}(S_n) = \frac{1}{n} \cdot np = p = \mu, \\
\mathsf{var}(\overline{X}) &= \frac{1}{n^2} \cdot \mathsf{var}(S_n) = \frac{1}{n^2} \cdot np(1-p) = \frac{p(1-p)}{n} = \sigma^2.
\end{aligned}$$

Par l'inégalité de Tchebychev on obtient pour $\varepsilon > 0$ que

$$P(|\,\overline{X} - \mu\,| \geq \varepsilon) = P(|\,\overline{X} - p\,| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} = \frac{p(1-p)}{n \cdot \varepsilon^2}.$$

Lorsque $n \to \infty$ cette borne (supérieure) converge vers 0.

Donc lorsque n est 'grand', la proportion de pile $\overline{X}$ est proche de la probabilité d'obtenir pile p.

Loi faible des grands nombres (version simple)

Soient $X_1, \ldots, X_n$ des variables aléatoires **indépendantes et identiquement distribuées** (**iid** — Notation: $\overset{\text{iid}}{\sim}$) avec $\mathrm{E}(X_i) = \mu$ finie, pour $i = 1, \ldots, n$.

Soit encore

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

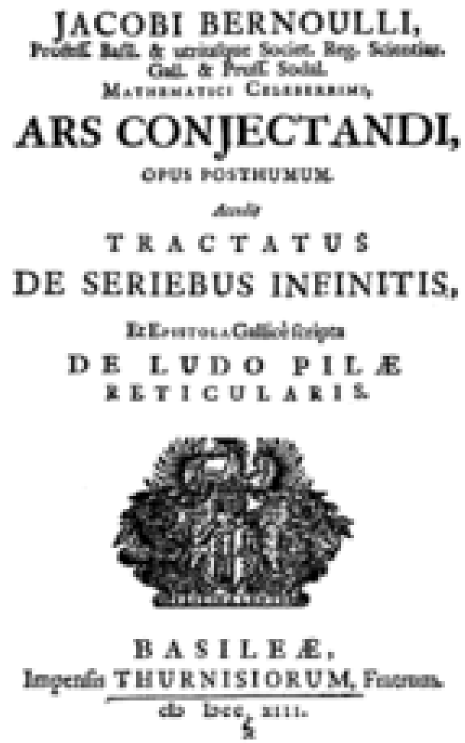
Alors pour tout $\varepsilon > 0$

$$P\left(\left| \overline{X} - \mu \right| \geq \varepsilon\right) \to 0$$

lorsque $n \to \infty$.

▷ Remarque: Donc $\overline{X}$ est proche de μ pour n 'grand'.

Remarque: La loi faible des grands nombres fut établie pour la première fois par Jacob Bernoulli pour le cas particulier où les X_i ne prennent pour valeur que 0 et 1 (et sont donc des variables Bernoulli). Son énoncé de ce théorème figure dans son ouvrage *Ars Conjectandi*, publié en 1713 par son neveu Nicolas Bernoulli, huit ans après sa mort.



EXEMPLE

Une roulette (à Las Vegas) comporte 38 cases, dont 18 cases noires, 18 cases rouges, une case 0 et une case 00. Si on mise 1 dollar sur noir, on va recevoir 2 dollars si la boule s'arrête dans une case noire et 0 sinon.

Soit X le gain d'un joueur qui mise 1 dollar sur noir. Puisque

$$P(X = +1) = \frac{18}{38} \quad \text{et} \quad P(X = -1) = \frac{20}{38},$$

on voit que

$$\mathsf{E}(X) = 1 \cdot \frac{18}{38} + (-1) \cdot \frac{20}{38} = -\frac{1}{19} \approx -0.0526.$$

Ceci signifie qu'on va perdre en moyenne 5.26 cents par jeu.

Imaginons qu'au cours de la soirée, il y a $1'000$ mises. Soit C_i le gain du casino lors de la i-ème mise. Puisque le casino gagne ce que les joueurs perdent,

$$\mathsf{E}(C_i) = \frac{1}{19} \approx +0.0526.$$

Le gain total du casino lors de ces $1'000$ mises est alors

$$S_{1'000} = C_1 + \cdots + C_{1'000}.$$

Le gain moyen du casino à chaque mise est

$$\frac{C_1 + \cdots + C_{1'000}}{1'000} = \frac{S_{1'000}}{1'000}.$$

Puisque $n = 1'000$ est relativement grand, il est très probable, d'après la loi faible des grands nombres, que $\frac{S_{1'000}}{1'000}$ soit voisin de 0.0526, donc $S_{1'000} \approx 52.6$ dollars.

La conclusion de cet exemple est que les bénéfices du casino sont systématiques, même s'ils ne paraissent pas spectaculaires!


MATHEMATICA
$\frac{1}{n}(x_1+\ldots+x_n) \longrightarrow E(X)$
HELVETIA 80
n
BURKARD WALTENSPÜL
1994
COURVOISIER

6.3 Loi forte des grands nombres (version simple)

Soient $X_1, \ldots, X_n$ des variables aléatoires iid d'espérance commune finie μ.

Alors, **avec probabilité 1**

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n} \to \mu$$

lorsque $n \to \infty$.

▷ Remarque: Il est donc certain que $\overline{X}$ soit proche de μ pour n 'grand'.

Remarques:

a) La loi forte des grands nombres est sans doute le **résultat le plus célèbre en théorie des probabilités**;

b) Il établit que la moyenne d'une suite de variables aléatoires iid tendra avec probabilité 1 vers l'espérance de cette distribution commune;

c) En d'autres termes la loi forte des grands nombres signifie que

$$P\left(\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu\right) = 1.$$

Comparaison de la loi faible et de la loi forte

▷ La loi faible des grands nombres assure que pour tout 'grande' valeur de n, disons $n^\star$ par exemple, $(X_1 + \cdots + X_{n^\star})/n^\star$ est probablement très proche (voisine) de μ.

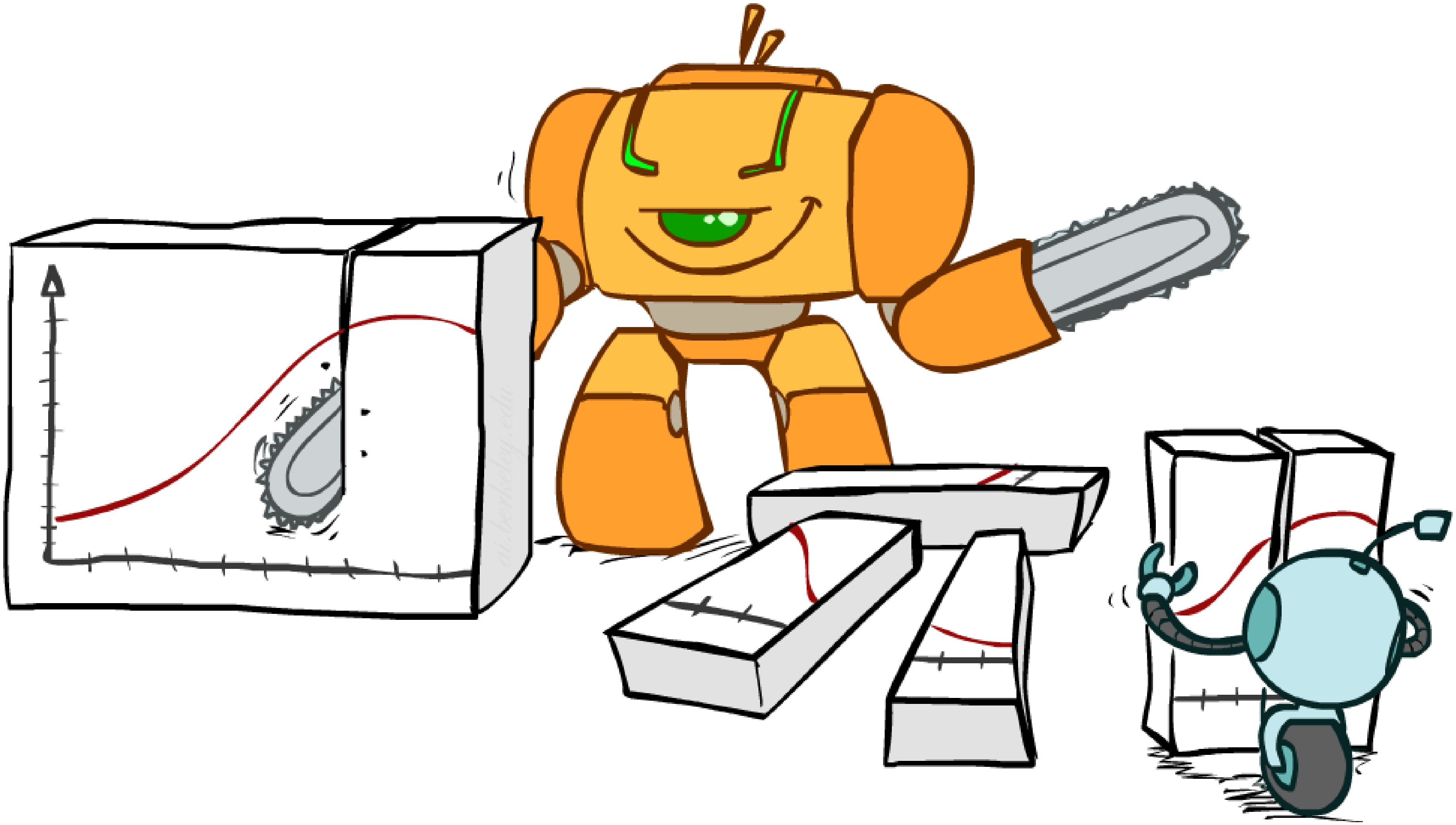
⇝ Elle n'assure pas cependant que $(X_1 + \cdots + X_n)/n = \overline{X}$ devra rester dans un voisinage étroit de μ pour toutes les valeurs de n supérieures à $n^\star$.

⇝ Elle laisse donc la porte ouverte à une situation où de larges écarts entre $\overline{X}$ et μ peuvent se produire pour une infinité d'événements, infinité dont la probabilité collective est très petite cependant.

▷ La loi forte des grands nombres exclut cette situation. Elle assure en particulier qu'avec probabilité 1 et pour toute valeur $\varepsilon > 0$,

$$| \, \overline{X} - \mu \, |$$

ne sera supérieure à ε qu'un nombre fini de fois.


ai.berkeley.edu

6.4 Théorème central limite (version simple)

Soient $X_1, \ldots, X_n$ des variables aléatoires iid avec $\mathsf{E}(X_i) = \mu$ et $\mathsf{var}(X_i) = \sigma^2$, pour tout $i = 1, \ldots, n$.

Soit $\overline{X}$ la moyenne des X_i. Que valent $\mathsf{E}(\overline{X})$ et $\mathsf{var}(\overline{X})$?

On a

$$\mathsf{E}(\overline{X}) = \mathsf{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}(X_i) = \frac{1}{n} \cdot n\,\mu = \mu,$$

et comme les X_i sont indépendantes,

$$\mathsf{var}(\overline{X}) = \mathsf{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathsf{var}(X_i) = \frac{1}{n^2} \cdot n\,\sigma^2 = \frac{\sigma^2}{n}.$$

▷ Remarque: Si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ on a $\overline{X} \sim N(\mu, \sigma^2/n)$.

Théorème central limite (version simple)

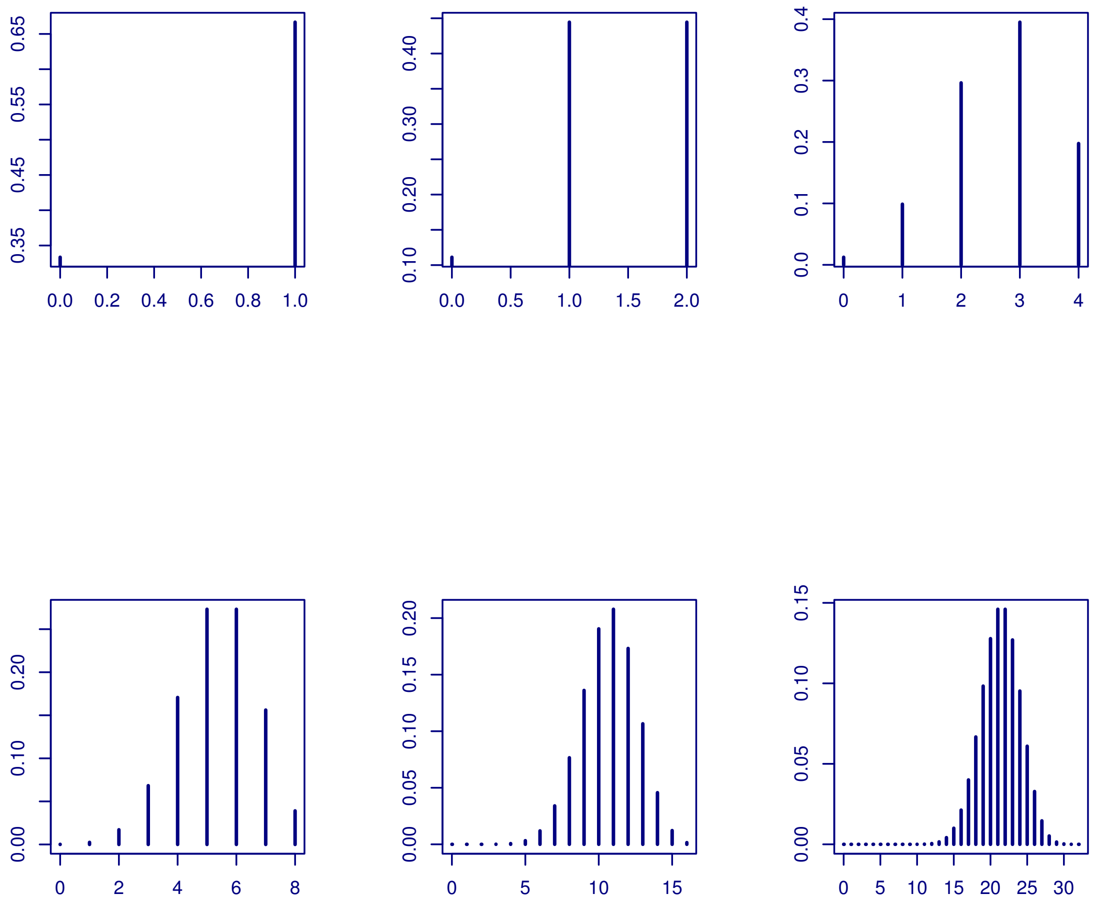
Soient $X_1, \ldots, X_n$ des variables aléatoires iid avec $\mathsf{E}(X_i) = \mu$ et $\mathsf{var}(X_i) = \sigma^2$ finies, pour tout $i = 1, \ldots, n$, et soit

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

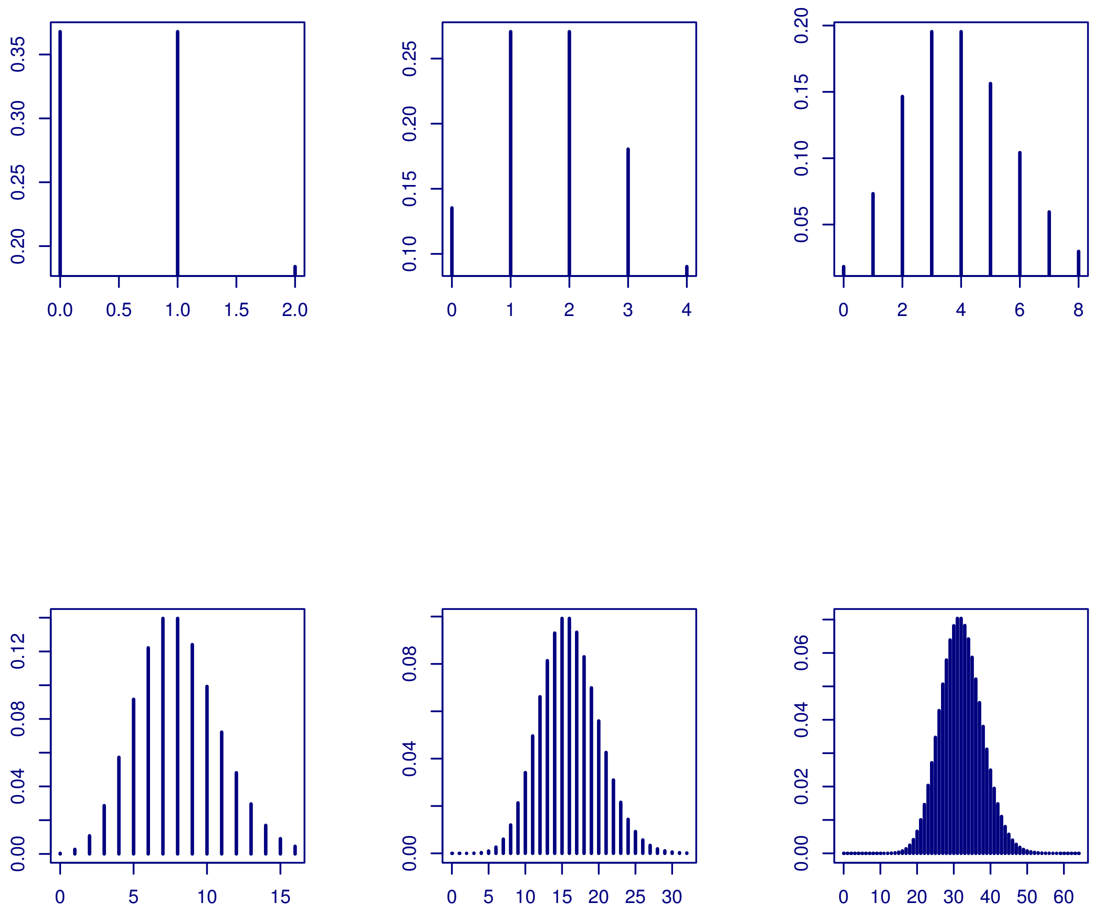
leur somme.

Alors lorsque $n \to \infty$ la distribution de S_n converge vers $N(n\mu, n\sigma^2)$, et donc la distribution de $\overline{X} = S_n/n$ converge vers $N(\mu, \sigma^2/n)$.

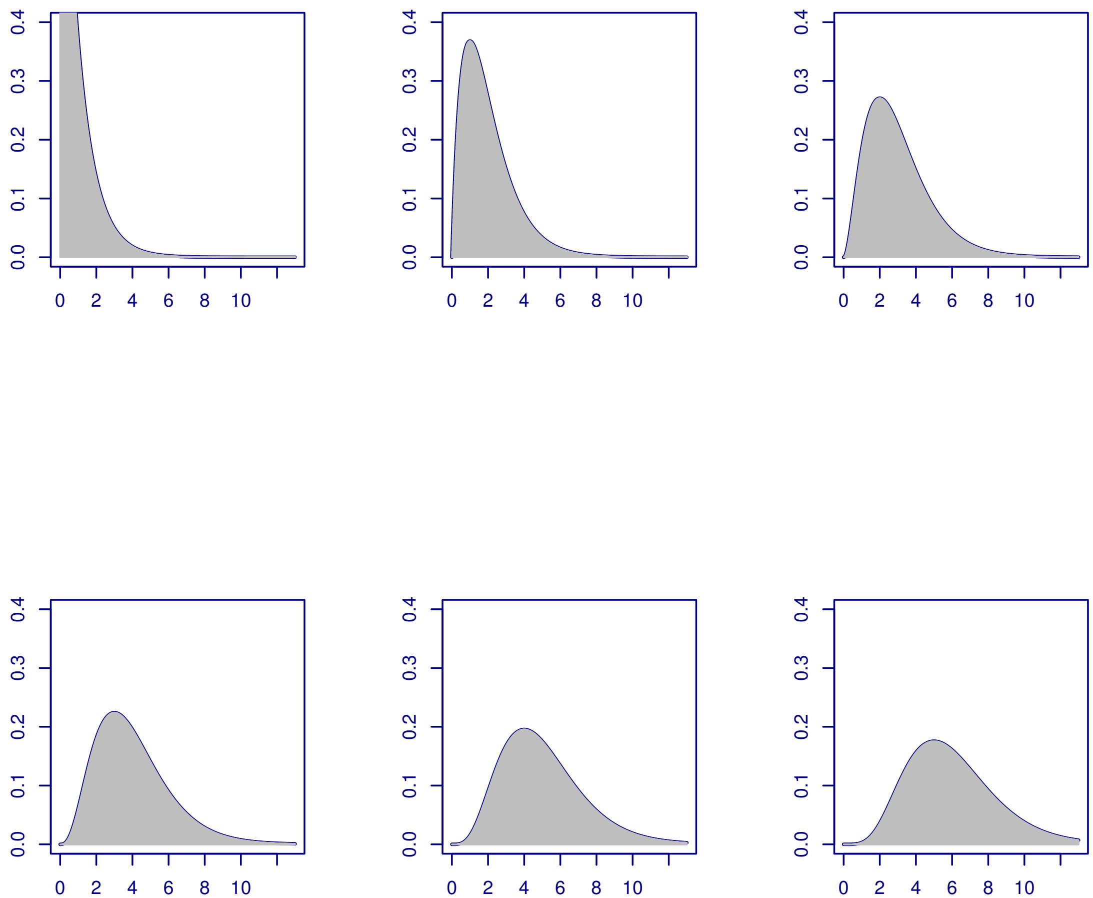
▷ Remarque: Si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ la distribution de $\overline{X} = S_n/n$ est exactement $N(\mu, \sigma^2/n)$ et celle de S_n exactement $N(n\mu, n\sigma^2)$.



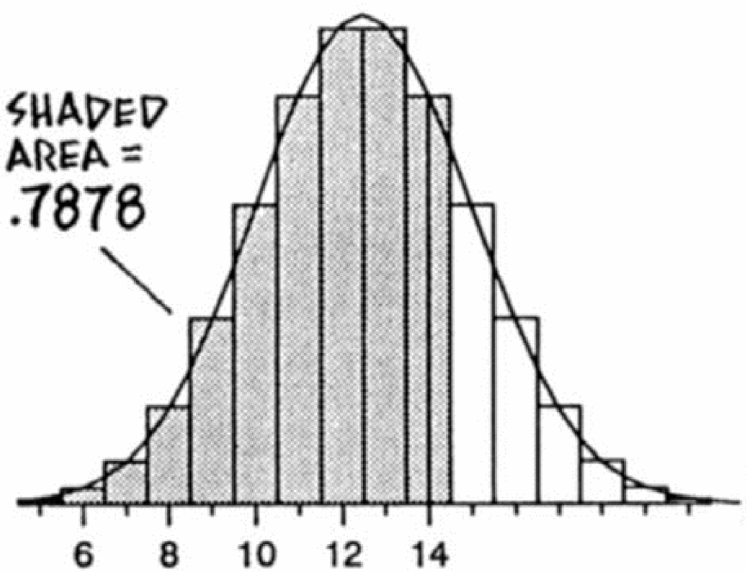
Distributions de la somme de 1, 2, 4, 8, 16 *et* 32 *variables aléatoires indépendantes de Bernoulli avec* $p = 2/3$, *i.e.* $B(1, 2/3)$.

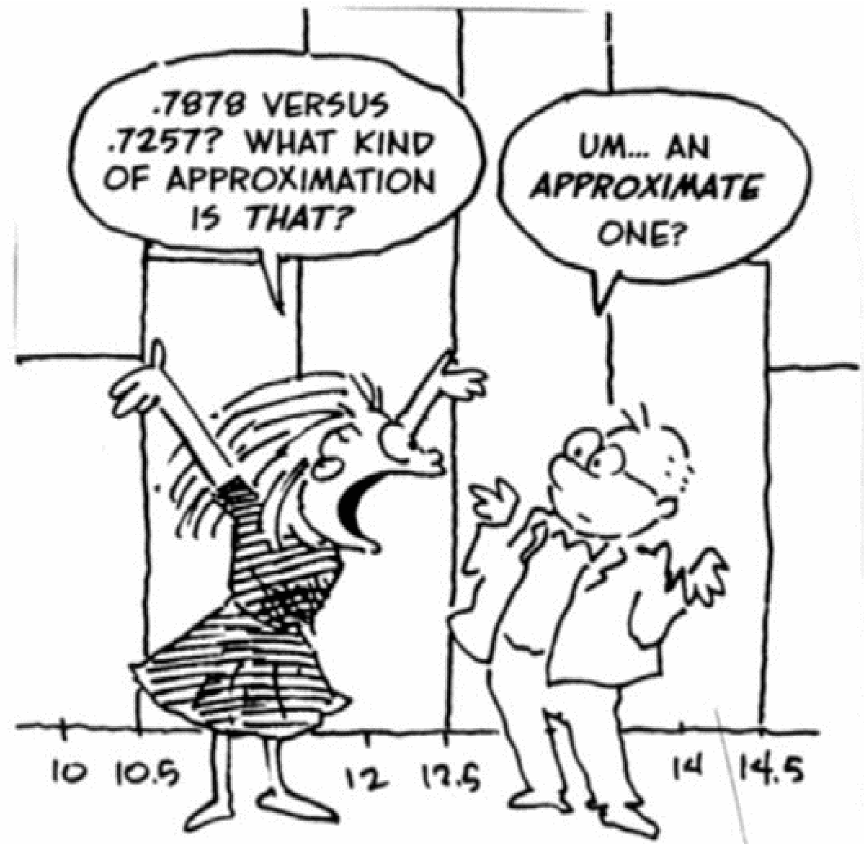


Distributions de la somme de 1, 2, 4, 8, 16 *et* 32 *variables aléatoires indépendantes de Poisson avec* $\lambda = 1$, *i.e.* $\mathrm{Poisson}(1)$.



Distributions de la somme de 1, 2, 3, 4, 5 *et* 6 *variables aléatoires indépendantes exponentielles avec* $\lambda = 1$, *i.e.* $\mathrm{Exp}(1)$.


SHADED AREA = .7878
6 8 10 12 14


.7878 VERSUS .7257? WHAT KIND OF APPROXIMATION IS THAT?
UM... AN APPROXIMATE ONE?
10 10.5 12 12.5 14 14.5

EXEMPLE

On lance 10 dés équilibrés. On cherche la probabilité que la somme des dix résultats soit comprise entre 30 et 40.

Réponse: On note X_i le résultat montré par le i-ème dé, $i = 1, \ldots, 10$.

Comme

$$\mathsf{E}(X_i) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = \frac{21}{6} = \frac{7}{2} = 3.5 = \mu$$

et

$$\begin{aligned} \mathrm{var}(X_i) &= \mathsf{E}(X_i^2) - \mathsf{E}(X_i)^2 \\ &= \left(1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + \cdots + 36 \cdot \frac{1}{6}\right) - \left(\frac{7}{2}\right)^2 = \frac{35}{12} = \sigma^2 \end{aligned}$$

(voir aussi page 149).

Soit

$$S_{10} = \sum_{i=1}^{10} X_i$$

la somme des $n = 10$ résultats montrés.

Ainsi on obtient par l'application du théorème central limite que la distribution de S_{10} est approximativement normale avec espérance $n\mu = 10 \cdot 3.5 = 35$ et variance $n\sigma^2 = 10 \cdot \frac{35}{12} = \frac{350}{12}$.

Donc

$$\begin{aligned} P(30 \leq S_{10} \leq 40) &= P\left(\frac{30-35}{\sqrt{\frac{350}{12}}} \leq \frac{S_{10}-35}{\sqrt{\frac{350}{12}}} \leq \frac{40-35}{\sqrt{\frac{350}{12}}}\right) \\ &\approx 2 \cdot \Phi(\sqrt{6/7}) - 1 \approx 2 \cdot \Phi(0.93) - 1 = 2 \cdot 0.8238 - 1 \\ &\approx 0.65. \end{aligned}$$

EXEMPLE

Devant l'augmentation des problèmes de poids dans la population européenne, une nouvelle étude est mandatée pour mesurer la relation entre celui-ci et la quantité de calories ingérées par habitant. Des études antérieures montrent qu'un Européen consomme en moyenne $2'700$ calories par jour avec un écart-type de 800.

On dispose d'un échantillon de 500 Européens et on cherche la probabilité que la moyenne des calories consommées par jour par les Européens dans l'échantillon soit supérieure à $2'750$.

Réponse: Soit X_i la variable aléatoire représentant le nombre de calories consommées par jour par l'individu i, $i = 1, \ldots, n = 500$, et

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

la moyenne des calories consommées par les individus dans l'échantillon.

On a $\mathsf{E}(X_i) = 2'700 = \mu$ et $\sqrt{\mathsf{var}(X_i)} = 800 = \sigma$.

En utilisant le théorème central limite, la distribution de $\overline{X}_n$ est approximativement normale avec espérance $\mu = 2'700$ et variance $\sigma^2/n = 800^2/500 = 1'280$.

Donc

$$\begin{aligned} P\left(\overline{X}_n > 2'750\right) &= P\left(\frac{\overline{X}_n - 2'700}{\sqrt{1'280}} > \frac{2'750 - 2'700}{\sqrt{1'280}}\right) \\ &\approx P(Z > 1.4) = 1 - P(Z \leq 1.4) \approx 0.09, \end{aligned}$$

où $Z \sim N(0, 1)$.

7. Variables aléatoires simultanées

• Soient deux variables aléatoires U et V définies respectivement dans leurs ensembles fondamentaux S_U et S_V.

$\rightsquigarrow$ On peut s'intéresser à des probabilités associées à des valeurs (ou intervalles de valeurs) de U et V.

$\rightsquigarrow$ Les distributions de $\boxed{\text{probabilités simultanées ou conjointes}}$ sont données par des **tableaux** (cas discret) ou des **formules de densité conjointe** (cas continue).

▷ Remarque: Dans ce cours on va s'intéresser qu'au cas discret!

EXEMPLE

Pour la vente des téléphones, il y a trois centres commerciaux intéressants (A, B et C). On vous dit que la répartition des commerciaux (séparés en deux types — expérimentés et novices) se fait selon le tableau suivant:

	Centre commercial (V)			
(U)	A	B	C	Total
Expérimentés (E)	0.26	0.30	0.04	0.60
Novices (N)	0.14	0.20	0.06	0.40
Total	0.40	0.50	0.10	1.00

$\rightsquigarrow$ Pour un commercial pris au hasard, quelle est la probabilité qu'il soit expérimenté et travaille au centre commercial A?

$$P(U = E, V = A) = 0.26$$

$\rightsquigarrow$ Pour un commercial pris au hasard, quelle est la probabilité qu'il soit expérimenté?

$$\begin{aligned} P(U = E) &= P(U = E, V = A) \\ &\quad + P(U = E, V = B) \\ &\quad + P(U = E, V = C) \;=\; 0.60 \end{aligned}$$

Cas discret

- En général:

(U)	$\ldots$	v_j (V)	$\ldots$	Total
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
u_i	$\ldots$	p_{ij}	$\ldots$	$p_{i\cdot}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
Total	$\ldots$	$p_{\cdot j}$	$\ldots$	1

$\rightsquigarrow$ On a $p_{ij} = P(U = u_i, V = v_j)$.

$\rightsquigarrow$ C'est la **probabilité simultanée ou conjointe**.

• **Probabilités marginales** (on ne s'intéresse qu'à une seule variable à la fois):

$$P(U = u_i) = \sum_j p_{ij} = p_{i\cdot}$$

$$P(V = v_j) = \sum_i p_{ij} = p_{\cdot j}$$

EXEMPLE

Pour un commercial pris au hasard, quelle est la probabilité qu'il soit expérimenté?

$$P(U = E) = P(U = E, V = A) + P(U = E, V = B) + P(U = E, V = C) = 0.60$$

- **Probabilités conditionnelles**:

$$P(U = u_i \,|\, V = v_j) \;=\; \frac{P(U = u_i, V = v_j)}{P(V = v_j)} \;=\; \frac{p_{ij}}{p_{\cdot j}}$$

EXEMPLE

Quelle est la probabilité qu'un commercial soit expérimenté s'il se trouve au centre commercial A?

$$P(U = E \,|\, V = A) \;=\; \frac{P(U = E, V = A)}{P(V = A)} \;=\; \frac{0.26}{0.40} = 0.65$$

On a que

$$
\begin{aligned}
P(U = u_i \,|\, V = v_j) &= \frac{P(U = u_i, V = v_j)}{P(V = v_j)} \\
&= \frac{P(U = u_i, V = v_j)}{P(U = u_i)} \cdot \frac{P(U = u_i)}{P(V = v_j)} \\
&= \frac{P(V = v_j, U = u_i)}{P(U = u_i)} \cdot \frac{P(U = u_i)}{P(V = v_j)} \\
&= P(V = v_j \,|\, U = u_i) \cdot \frac{P(U = u_i)}{P(V = v_j)}
\end{aligned}
$$

et

$$\begin{aligned} P(V=v_j) &= \sum_i P(V=v_j, U=u_i) \\ &= \sum_i P(V=v_j \mid U=u_i)P(U=u_i) \end{aligned}$$

$\rightsquigarrow$ On obtient le **théorème de Bayes**:

$$P(U=u_i \mid V=v_j) = \frac{P(V=v_j \mid U=u_i)P(U=u_i)}{\sum_i P(V=v_j \mid U=u_i)P(U=u_i)}$$

EXEMPLE

On peut en fait répartir les commerciaux en trois catégories, à savoir les novices (20%), les expérimentés (50%) et les chanceux (30%).

La probabilité qu'un commercial ait un rendement de plus de 20 francs par heure est pour chacun des types de commercial de respectivement 0.25, 0.55 et 0.90.

– Quelle est la proportion de commerciaux ayant un rendement de plus de 20 francs par heure?

– Si un commercial a un rendement de plus de 20 francs par heure, quelle est la probabilité qu'il appartienne au groupe des chanceux?

Réponse: Soient

- U = le type de commercial, trois modalités: 'novice' (u_1), 'expérimenté' (u_2), 'chanceux' (u_3); et
- V = le rendement horaire, deux modalités: 'plus de 20 francs par heure' (v_1) ou 'ne pas plus de 20 francs par heure' (v_2).

On a

- $P(U = u_1) = 0.2$;
- $P(U = u_2) = 0.5$;
- $P(U = u_3) = 0.3$;
- $P(V = v_1 \mid U = u_1) = 0.25$;
- $P(V = v_1 \mid U = u_2) = 0.55$;
- $P(V = v_1 \mid U = u_3) = 0.90$.

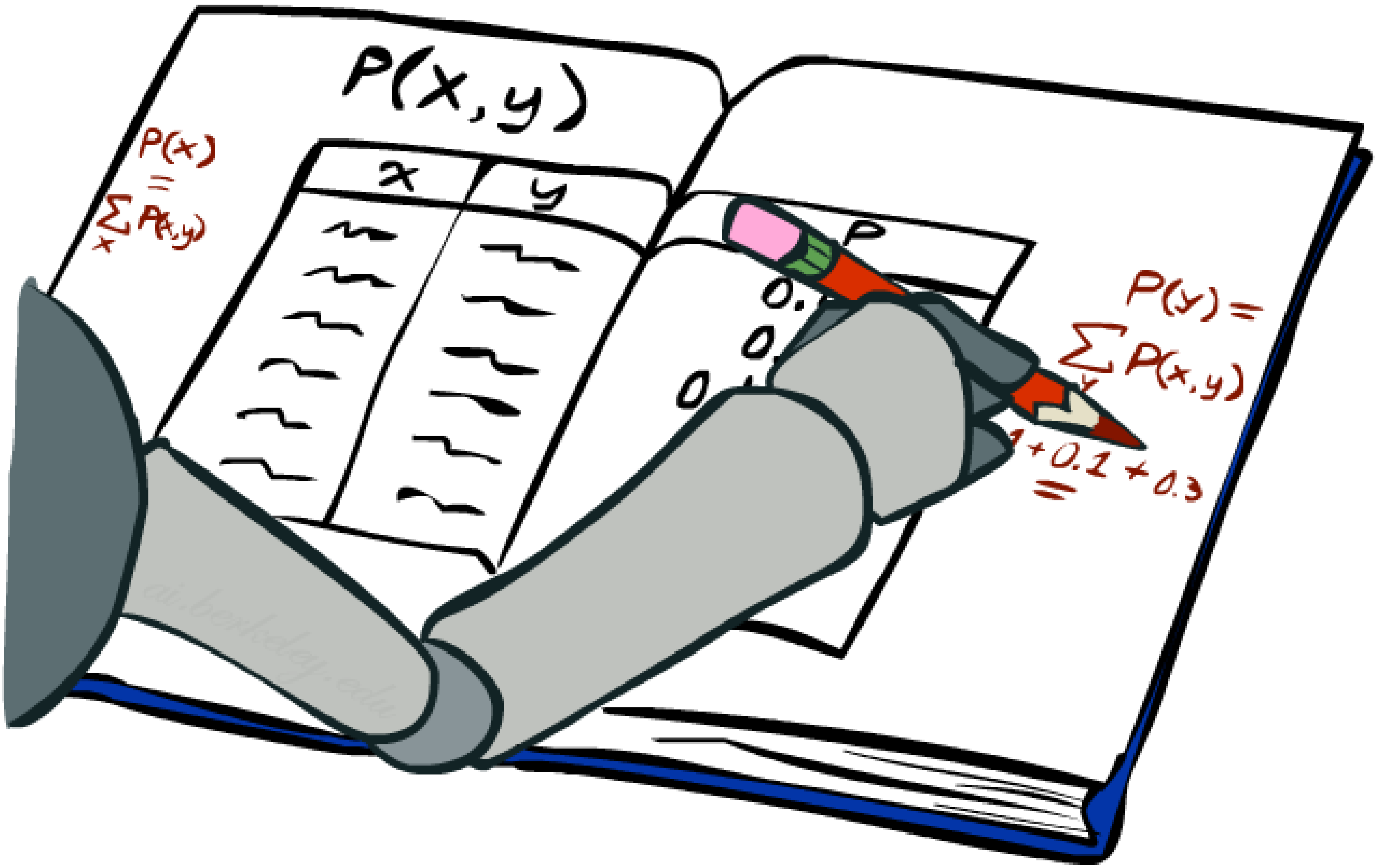
$\rightsquigarrow$ On cherche $P(V = v_1)$ et $P(U = u_3 \mid V = v_1)$.

On a

$$\begin{aligned} P(V = v_1) &= \sum_{i=1}^{3} P(V = v_1 \mid U = u_i) P(U = u_i) \\ &= 0.25 \cdot 0.20 + 0.55 \cdot 0.50 + 0.90 \cdot 0.30 = 0.595 \end{aligned}$$

et

$$\begin{aligned} P(U = u_3 \mid V = v_1) &= \frac{P(V = v_1 \mid U = u_3) P(U = u_3)}{\sum_{i=1}^{3} P(V = v_1 \mid U = u_i) P(U = u_i)} \\ &= \frac{0.90 \cdot 0.30}{0.595} = 0.454. \end{aligned}$$


P(x,y)
x
y
P
P(x) = Σ_x P(x,y)
P(y) = Σ_y P(x,y)
+0.1 + 0.3
=
ai.berkeley.edu

Espérances et variances

Pour les variables discrètes (non nominales), on peut aussi calculer des espérances et des variances.

EXEMPLE

(U)	(V) $v_1 = 0$	$v_2 = 1$	$v_3 = 2$	Total
$u_1 = 1$	0.10	0.20	0.05	0.35
$u_2 = 2$	0.10	0.25	0.10	0.45
$u_3 = 3$	0.05	0.10	0.05	0.20
Total	0.25	0.55	0.20	1.00

- **Espérance marginale**:

$$
\begin{aligned}
\mathsf{E}(U) &= 1 \cdot P(U=1) + 2 \cdot P(U=2) + 3 \cdot P(U=3) \\
&= 1 \cdot \sum_{j=1}^{3} p_{1j} + 2 \cdot \sum_{j=1}^{3} p_{2j} + 3 \cdot \sum_{j=1}^{3} p_{3j} \\
&= \sum_{i=1}^{3} u_i \sum_{j=1}^{3} p_{ij} \\
&= \sum_i \sum_j u_i p_{ij} = 1.85
\end{aligned}
$$

De même on obtient $\mathsf{E}(V) = \sum_i \sum_j v_j p_{ij} = 0.95$.

• **Variance marginale**:

$$\begin{aligned} \mathrm{var}(U) &= \sum_i \sum_j \left[u_i - \mathsf{E}(U)\right]^2 p_{ij} \\ &= \sum_i \sum_j u_i^2 p_{ij} - \left[\mathsf{E}(U)\right]^2 \end{aligned}$$

EXEMPLE

On obtient $\mathrm{var}(U) = 3.95 - 1.85^2 = 0.5275$ et $\mathrm{var}(V) = 1.35 - 0.95^2 = 0.4475$ (toujours positives!).

Propriété d'indépendance et covariance

$$\boxed{U \text{ et } V \text{ sont } \textbf{indépendantes} \iff p_{ij} = p_{i\cdot} \cdot p_{\cdot j} \text{ pour tout } i \text{ et } j.}$$

- **Covariance** (mesure de dépendance linéaire):

$$\begin{aligned} \mathsf{cov}(U, V) &= \mathsf{E}\big\{[U - \mathsf{E}(U)]\,[V - \mathsf{E}(V)]\,\big\} \\ &= \mathsf{E}(UV) - \mathsf{E}(U)\mathsf{E}(V) \end{aligned}$$

EXEMPLE (voir page 394)

UV	0	1	2	3	4	6	Total
p_i	0.25	0.2	0.3	0.1	0.1	0.05	1.00

On obtient $\mathsf{E}(UV) = 0 \cdot 0.25 + 1 \cdot 0.2 + \cdots + 6 \cdot 0.05 = 1.8.$

Autrement:

$$\begin{aligned}\mathsf{E}(UV) &= \sum_i \sum_j u_i v_j p_{ij} \\ &= 0 \cdot 1 \cdot 0.10 + 0 \cdot 2 \cdot 0.10 + \cdots + 1 \cdot 1 \cdot 0.20 + \cdots + 2 \cdot 3 \cdot 0.05 \\ &= 1.8.\end{aligned}$$

Donc $\mathsf{cov}(U, V) = \mathsf{E}(UV) - \mathsf{E}(U)\mathsf{E}(V) = 1.8 - 1.85 \cdot 0.95 = 0.0425.$

EXEMPLE

Quatre amis dont Nicolas et Manjola achètent des billets pour le 'Festival de Jazz de Montreux'.

Les billets reçus sont pour les sièges 5 à 8 d'une rangée particulière.

On suppose que les places sont réparties au hasard entre les quatre amis.

Soit U le numéro du siège occupé par Nicolas et V celui occupé par Manjola.

1. Déterminer la distribution simultanée ou conjointe de U et V.

2. Calculer $\mathsf{E}(U)$, $\mathsf{E}(V)$, $\mathsf{var}(U)$, $\mathsf{var}(V)$ et $\mathsf{cov}(U, V)$.

3. Calculer l'espérance du nombre de sièges qui séparent Nicolas et Manjola.

Réponse:

1. En faisant une énumération complète des $4! = 24$ dispositions possibles des quatre amis sur les sièges $5, 6, 7$ et 8, on obtient le tableau suivant des probabilités simultanées ou conjointes:

(U) \ (V)	$v_1 = 5$	$v_2 = 6$	$v_3 = 7$	$v_4 = 8$	Total
$u_1 = 5$	0	1/12	1/12	1/12	1/4
$u_2 = 6$	1/12	0	1/12	1/12	1/4
$u_3 = 7$	1/12	1/12	0	1/12	1/4
$u_4 = 8$	1/12	1/12	1/12	0	1/4
Total	1/4	1/4	1/4	1/4	1

2. Par symétrie: $\mathsf{E}(U) = \mathsf{E}(V)$ et $\mathsf{var}(U) = \mathsf{var}(V)$. En utilisant les définitions, on obtient les résultats suivants:

$$\mathsf{E}(U) = 5 \cdot \frac{1}{4} + 6 \cdot \frac{1}{4} + 7 \cdot \frac{1}{4} + 8 \cdot \frac{1}{4} = 6.5$$

et

$$\mathsf{var}(U) = \mathsf{E}(U^2) - [\mathsf{E}(U)]^2 = (5^2 + 6^2 + 7^2 + 8^2) \cdot \frac{1}{4} - 6.5^2 = 1.25$$

et

$$\begin{aligned}
\mathsf{cov}(U, V) &= \mathsf{E}(UV) - \mathsf{E}(U)\mathsf{E}(V) \\
&= 2 \cdot (5 \cdot 6 + 5 \cdot 7 + 5 \cdot 8 + 6 \cdot 7 + 6 \cdot 8 + 7 \cdot 8) \cdot \frac{1}{12} - 6.5^2 \\
&= -0.41\overline{6}.
\end{aligned}$$

3. Soit W la variable aléatoire qui représente le nombre de sièges qui séparent Nicolas et Manjola.

   W prend la valeur 0 avec la probabilité

$$\frac{1}{12}+\frac{1}{12}+\frac{1}{12}+\frac{1}{12}+\frac{1}{12}+\frac{1}{12}=6\cdot\frac{1}{12}=\frac{6}{12}=\frac{1}{2},$$

   et les valeurs $1, 2$ avec les probabilités $4/12 = 1/3$ et $2/12 = 1/6$ respectivement.

   Donc

$$\mathsf{E}(W)=0\cdot\frac{1}{2}+1\cdot\frac{1}{3}+2\cdot\frac{1}{6}=\frac{2}{3}.$$

- **Propriétés**:

$$\boxed{\begin{array}{c} U \text{ et } V \text{ sont indépendantes} \Rightarrow \text{cov}(U,V) = 0. \\ \\ \\ \text{cov}(U,V) \neq 0 \;\Rightarrow\; U \text{ et } V \text{ sont dépendantes.} \end{array}}$$

Preuve: Indépendance $\Longleftrightarrow p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$ pour tout i et j. Donc

$$\begin{aligned} \text{E}(UV) &= \sum_i \sum_j u_i v_j p_{ij} = \sum_i \sum_j u_i v_j p_{i\cdot} p_{\cdot j} \\ &= \sum_i u_i p_{i\cdot} \sum_j v_j p_{\cdot j} = \text{E}(U)\text{E}(V) \end{aligned}$$

et donc $\text{cov}(U,V) = \text{E}(UV) - \text{E}(U)\text{E}(V) = 0$.

• Propriétés de combinaisons de variables aléatoires (**cas général**):

$$\mathsf{E}\left(aU + bV\right) = a\mathsf{E}(U) + b\mathsf{E}(V)$$

$$\mathsf{var}\left(aU + bV\right) = a^2\mathsf{var}(U) + b^2\mathsf{var}(V) + 2ab \cdot \mathsf{cov}(U, V)$$

Remarque: Si les variables sont indépendantes, alors

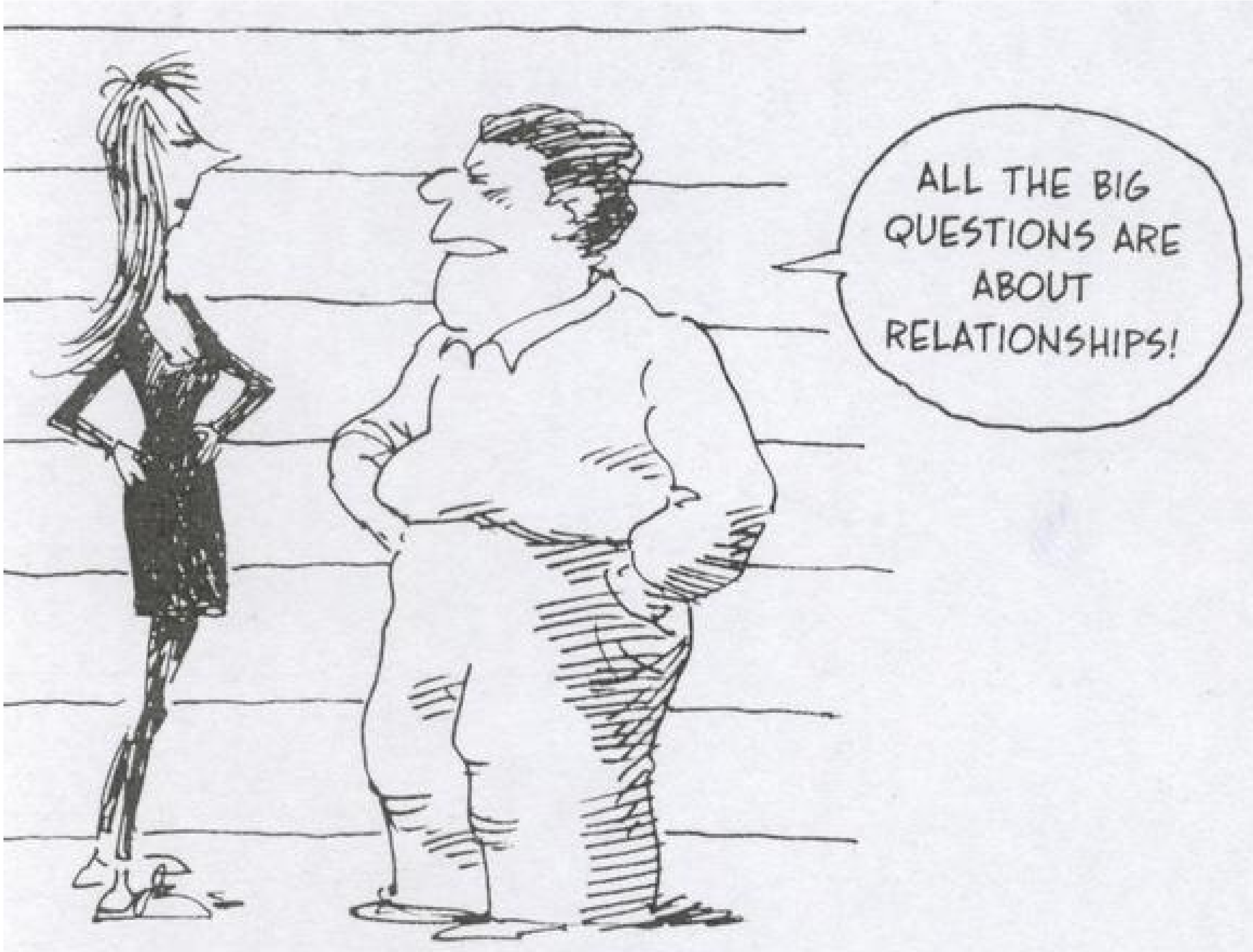
$$\mathsf{var}\left(aU + bV\right) = a^2\mathsf{var}(U) + b^2\mathsf{var}(V)$$

comme $\mathsf{cov}(U, V) = 0$.

- **Cas particulier:**

$$\mathsf{E}\,(U+V)=\mathsf{E}(U)+\mathsf{E}(V)$$

$$\mathsf{var}\,(U+V)=\mathsf{var}(U)+\mathsf{var}(V)+2\cdot\mathsf{cov}(U,V)$$


ALL THE BIG QUESTIONS ARE ABOUT RELATIONSHIPS!

Coefficient de corrélation linéaire

On préfère prendre une mesure standardisée de la covariance (dépendance linéaire entre U et V) donnée par le **coefficient de corrélation linéaire**:

$$\rho_{UV} = \frac{\text{cov}(U, V)}{\sqrt{\text{var}(U)\text{var}(V)}},$$

$$-1 \leq \rho_{UV} \leq 1.$$

Cette mesure de dépendance linéaire (et seulement linéaire!) ne dépend pas des échelles de mesure des variables.

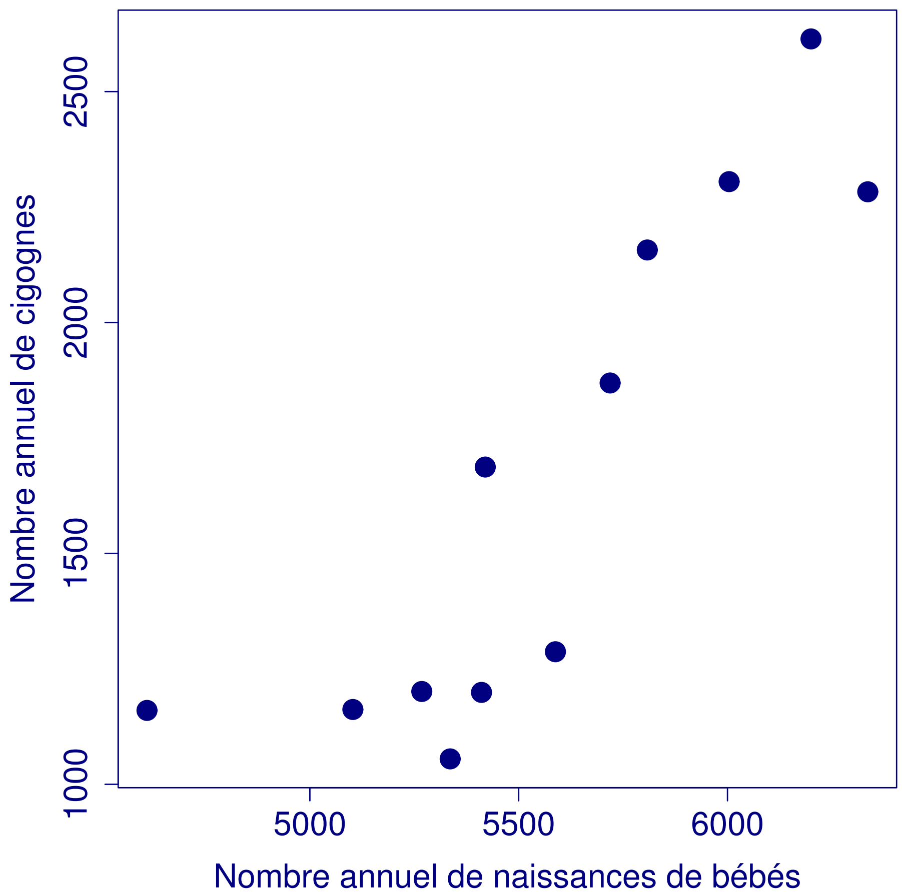
$$\boxed{\begin{array}{c} U \text{ et } V \text{ sont indépendantes} \\ \Rightarrow \\ \mathrm{cov}(U,V)=0 \\ \Rightarrow \\ \rho_{UV}=0. \end{array}}$$

▷ **Attention**: En général, la **réciproque n'est pas vraie**, *i.e.* $\rho_{UV} = 0$ n'implique pas que U et V sont indépendantes!

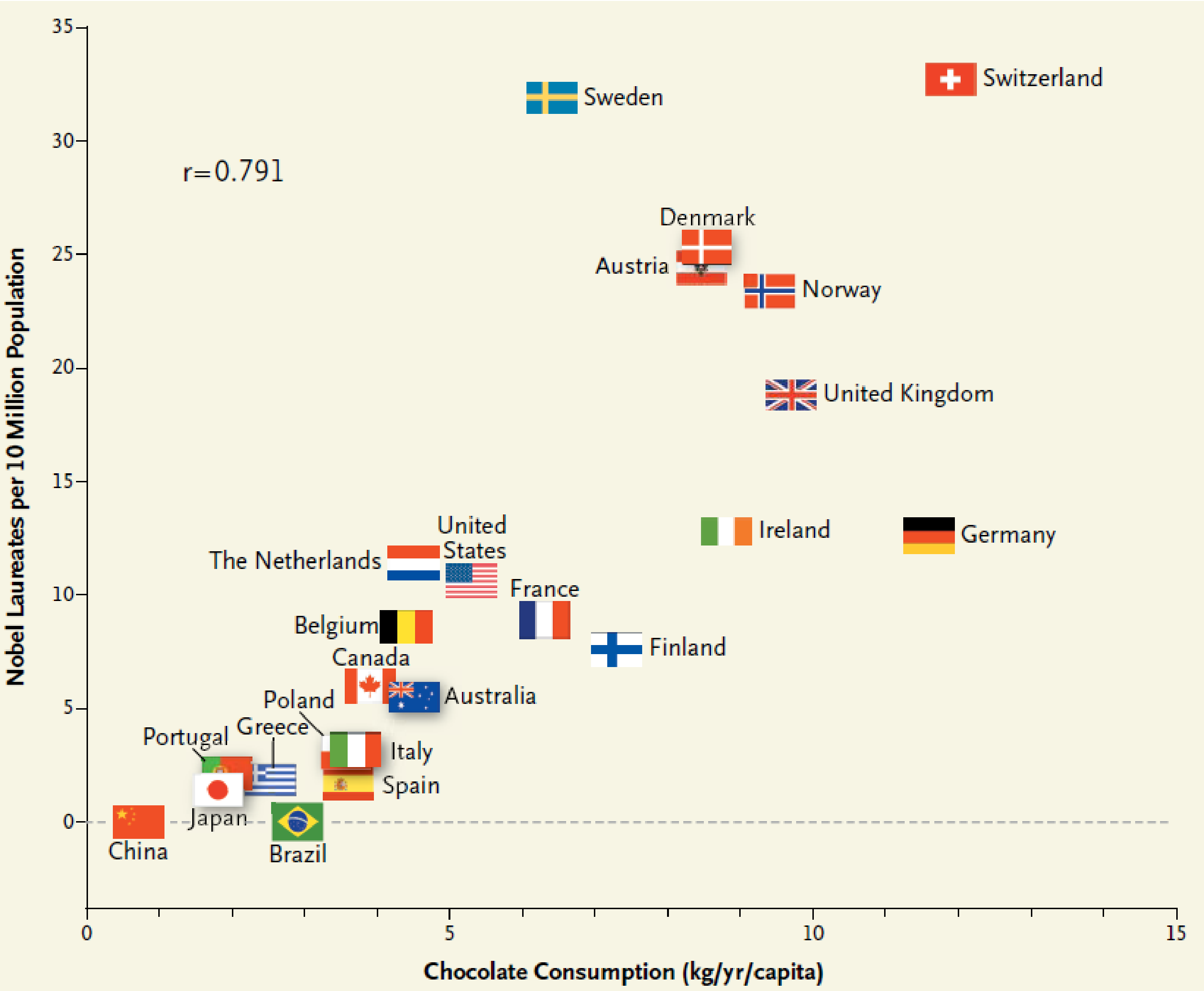
$$\boxed{\text{Corrélation} \neq \text{Causalité!}}$$

La confusion entre corrélation et causalité est appelée '**effet cigogne**'!

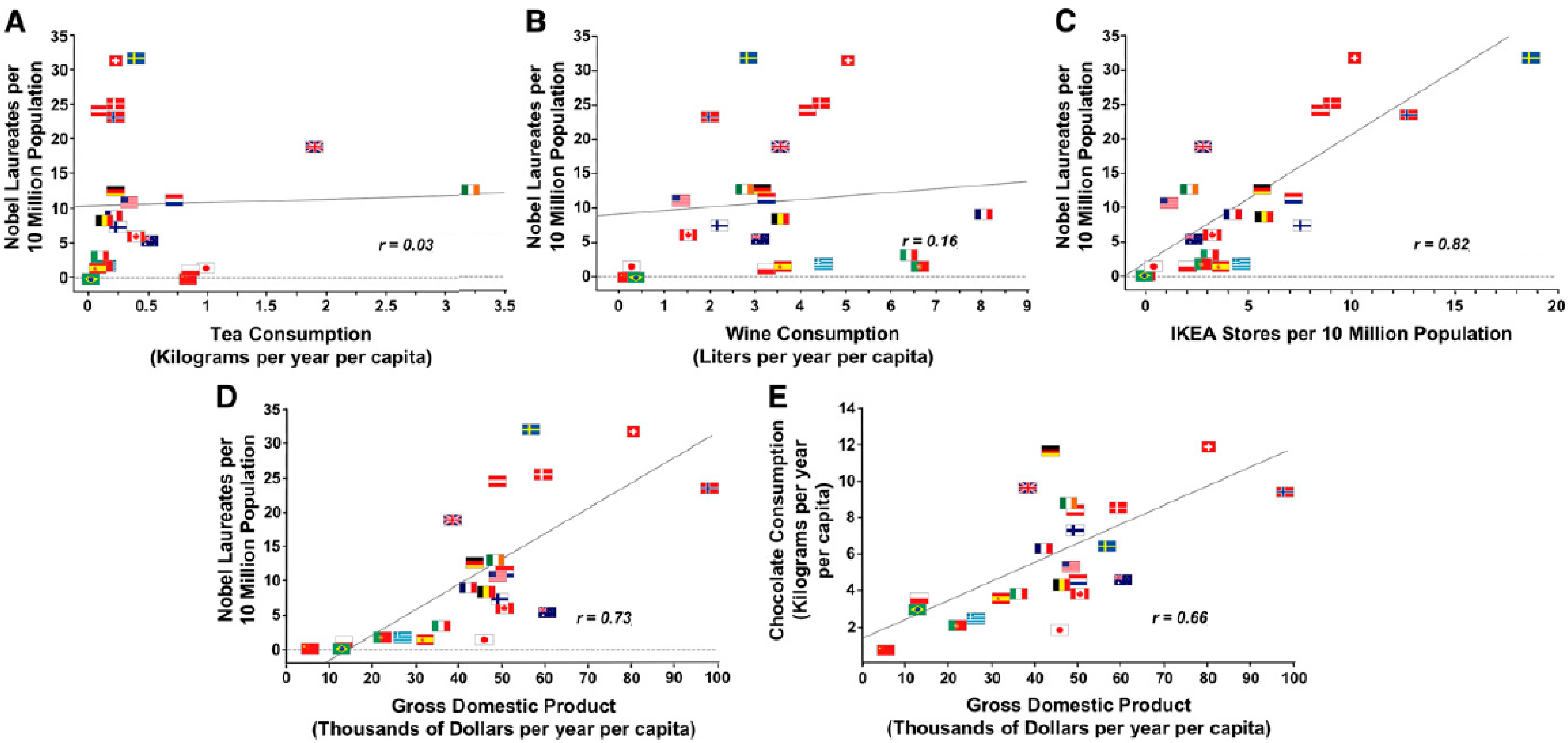
Relation entre le nombre annuel de naissances de bébés (U) et le nombre annuel de cigognes (V) à Copenhage de 1946 à 1957 avec $\rho_{UV} = 0.861(!)$:



Plus un pays mange de chocolat, plus il a de prix Nobel...

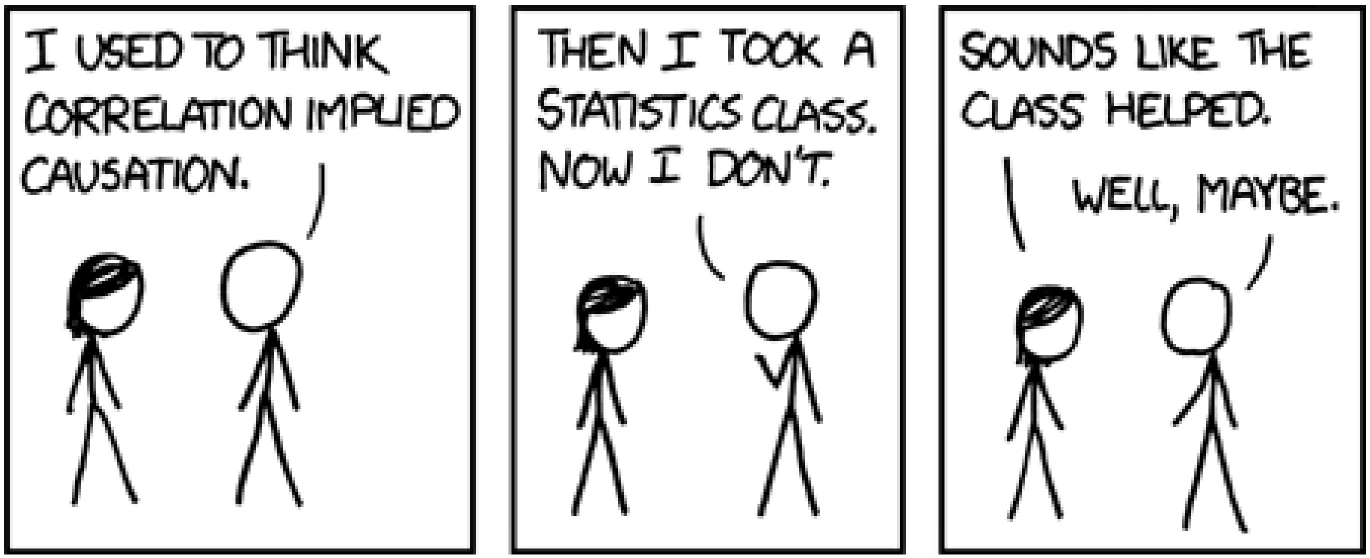


Source: Messerli, F. H. (2012). Chocolate consumption, cognitive function, and Nobel laureates. The New England Journal of Medicine. 367, 1562–1564.



Source: Maurage, P., Heeren, A. & Pesenti, M. (2013). Does chocolate consumption really boost Nobel Award chances?

The peril of over-interpreting correlations in health studies. *The Journal of Nutrition*. 143, 931–933.


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

8. Eléments de simulation

8.1 Introduction à la simulation

• Comment peut-on déterminer la **probabilité de gagner une partie de solitaire** ou de 'patience' (*i.e.* n'importe quelle version connue du jeu utilisant un paquet de 52 cartes standards et basée sur une 'stratégie fixe').

$\rightsquigarrow$ Une méthode consiste à admettre l'hypothèse raisonnable d'équiprobabilité des $52!$ permutations possibles des cartes, puis à déterminer combien parmi celles-ci sont gagnantes.

$\rightsquigarrow$ Il ne semble malheureusement pas aisé de mettre au point un critère systématique discriminant les permutations gagnantes: $52!$ est un nombre fort grand!

$\rightsquigarrow$ Il semble qu'il n'y ait pas d'autre moyen de savoir si une permutation mène à une réussite que de jouer la partie.

$\rightsquigarrow$ Cette approche ne conviendra pas!

$\rightsquigarrow$ Il apparaît à ce point que la détermination de la probabilité de gagner échappe au traitement mathématique.

• Cependant les choses ne s'arrêtent pas là car les probabilités sont du domaine des sciences appliquées comme de celui des mathématiques; or dans toutes les sciences appliquées l'**expérimentation est fort utile**.

• Dans le cas de notre partie de solitaire, par exemple, l'expérimentation revient à **exécuter un grand nombre de parties** ou mieux encore à **programmer un ordinateur** pour qu'il le fasse.

$\rightsquigarrow$ Après l'exécution de, disons, n parties on pourra poser

$$X_i = \begin{cases} 1 & \text{si le } i\text{-ème jeu est une victoire} \\ 0 & \text{sinon} \end{cases}$$

$\rightsquigarrow$ Les variables X_i, $i = 1, \ldots, n$, seront alors des variables aléatoires de Bernoulli pour lesquelles

$$\mathsf{E}(X_i) = P(\text{gagner une partie}) = \mu.$$

$\rightsquigarrow$ D'après la loi forte des grands nombres, nous saurons que

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{\text{Nombre de parties gagnées}}{\text{Nombre de parties jouées}}$$

tendra, lorsque $n \to \infty$, avec probabilité 1 vers μ, *i.e.* la probabilité de gagner une partie.

$\rightsquigarrow$ Ou encore, on peut dire qu'après un grand nombre de parties on peut utiliser la proportion de parties gagnées sur le nombre total de parties pour obtenir une **estimation de la probabilité cherchée**.

La méthode consistant à déterminer des probabilités de manière empirique à travers l'expérimentation est appelée **simulation**.


WARPED By Mike Cavna
250 B.C.:
WHY ARCHIMEDES COULD NEVER QUITE WIN THE LOTTERY...
DANG—THAT LAST NUMBER NEVER SEEMS TO GET PICKED!...
LOTTERY TICKET

8.2 Procédé de simulation

• Dans le but d'utiliser un ordinateur pour mener une étude de simulation, on doit pouvoir générer les valeurs d'une variable aléatoire uniforme $U(0,1)$.

$\leadsto$ Ces valeurs sont appelées des **nombres aléatoires**.

$\leadsto$ À partir d'un générateur de la distribution uniforme $U(0,1)$, on peut en principe simuler n'importe quelle distribution par la 'méthode de la fonction inverse'.

$\leadsto$ Une des questions que vous pouvez vous poser lorsque vous voyez défiler une séquence de nombres générée aléatoirement sur un ordinateur est 'Puis-je prévoir le nombre suivant?' ou autrement dit 'La séquence est-elle vraiment aléatoire?'.

• Créer une séquence aléatoire sur ordinateur est aujourd'hui encore impossible.

$\leadsto$ Le principe est plutôt d'**imiter l'aléatoire**.

$\rightsquigarrow$ Ce qui est certain, c'est que l'ordinateur ne sais pas lancer une pièce de monnaie, encore moins un dé.

$\rightsquigarrow$ En fait, l'ordinateur utilise ce que l'on appelle un **générateur de nombres pseudo aléatoires**: Une fonction prédéfinie qui produit une suite de **nombres pseudo-aléatoires**.

$\rightsquigarrow$ C'est une suite de nombres qui, pratiquement, est semblable à un échantillon issu d'une distribution uniforme $U(0, 1)$.

$\rightsquigarrow$ Le choix d'une telle fonction sort du cadre de ce cours!

Exemples dans le logiciel statistique libre R

▷ 1 réalisation d'une variable $B(10, 0.5)$:

```
> rbinom(1, 10, 0.5)
[1] 4

> rbinom(1, 10, 0.5)
[1] 6

> rbinom(1, 10, 0.5)
[1] 5
```


▷ 5 réalisations d'une variable $B(10, 0.5)$:

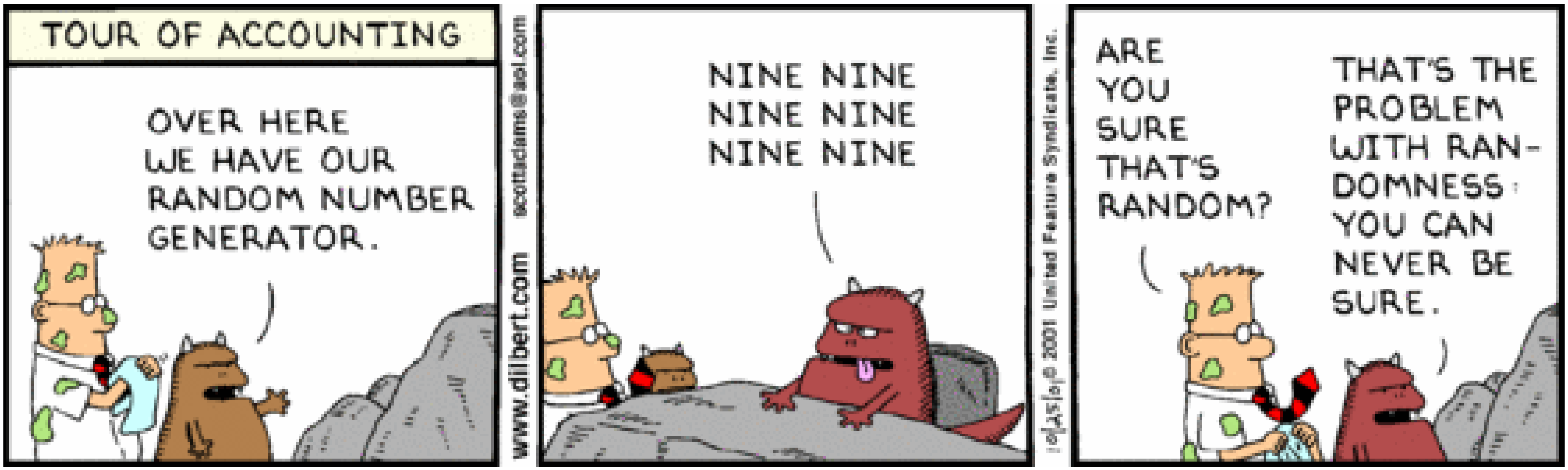
```
> rbinom(5, 10, 0.5)
[1] 3 6 5 4 5

> rbinom(5, 10, 0.5)
[1] 5 3 4 5 6
```


▷ 1 réalisation d'une variable $U(0,1)$:

```
> runif(1, 0, 1)
[1] 0.5858003
```

▷ 5 réalisations d'une variable $U(0,1)$:

```
> runif(5, 0, 1)
[1] 0.2002145 0.6852186 0.9168758 0.2843995 0.1046501
```

TOUR OF ACCOUNTING
OVER HERE WE HAVE OUR RANDOM NUMBER GENERATOR.
www.dilbert.com scottadams@aol.com
NINE NINE NINE NINE NINE NINE
© 2001 United Feature Syndicate, Inc.
ARE YOU SURE THAT'S RANDOM?
THAT'S THE PROBLEM WITH RANDOMNESS: YOU CAN NEVER BE SURE.

8.3 Exemples d'application de simulation

8.3.1 Position d'une particule

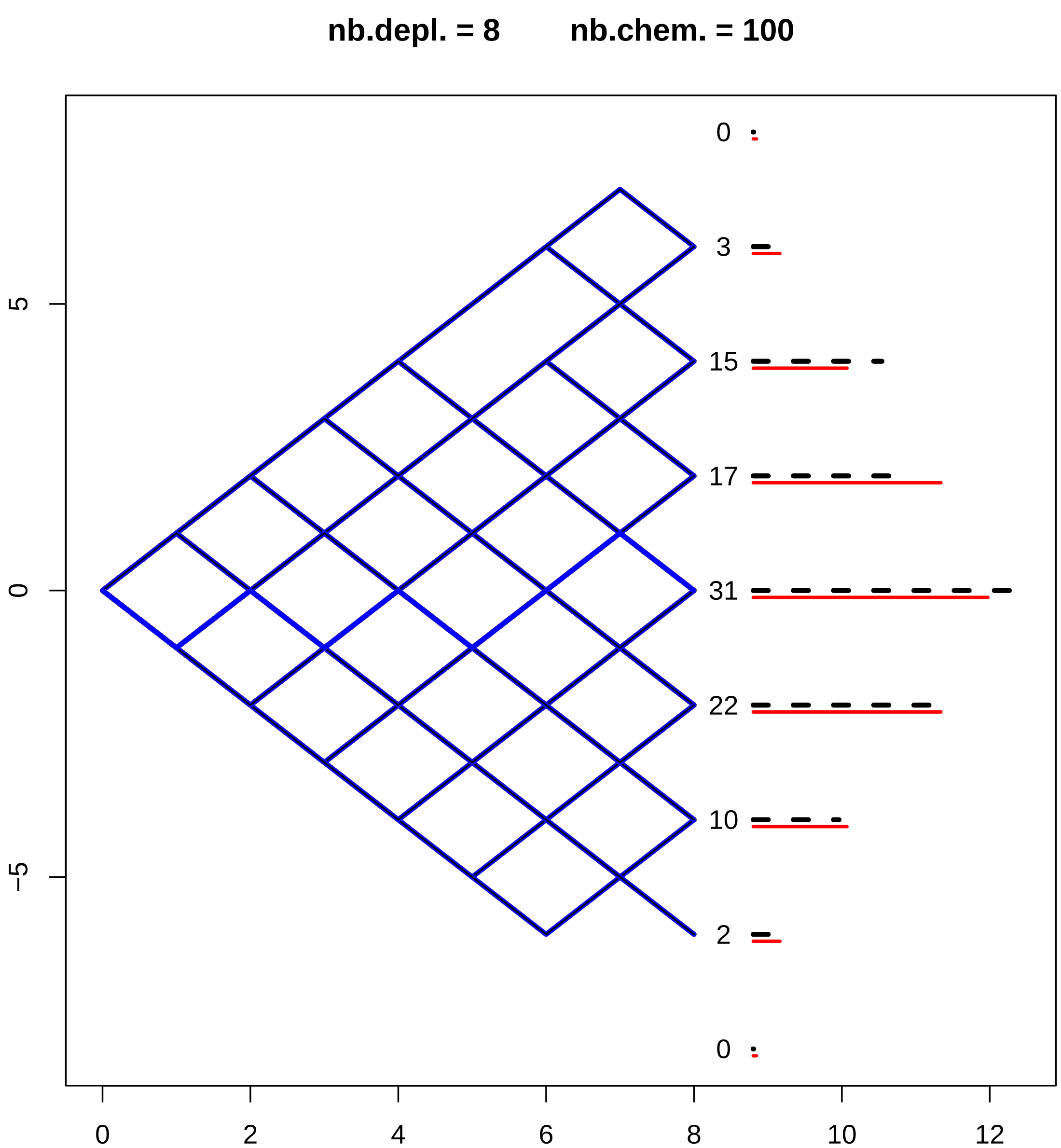
• On considère l'exemple d'une particule qui se trouve sur une droite et qui commence son parcours au point 0.

$\leadsto$ Lorsqu'elle se trouve en un point donné, elle se déplace vers la gauche ou vers la droite d'une manière équiprobable.

$\leadsto$ Soit X_n la position de la particule après n déplacements.

$\leadsto$ On peut démontrer rigoureusement que la variable aléatoire $Y_n = (n + X_n)/2$ est distribuée selon une distribution binomiale $B(n, 1/2)$. Ce n'est pas notre but ici!

$\leadsto$ Nous essaierons cependant d'étudier expérimentalement le comportement de X_n en comparant les valeurs observées aux valeurs théoriques.


nb.depl. = 8 nb.chem. = 100
0
3
15
17
31
22
10
2
0
5
0
−5
0
2
4
6
8
10
12

$\rightsquigarrow$ On constate, entre autre, les choses suivantes:

- La particule se promène presque partout;
- On est toujours très près des valeurs théoriques;
- On n'est jamais exactement sur les valeurs théoriques, mais toujours autour. Ceci est dû au fait que nos variables sont aléatoires. Et comme les trajectoires sont des fonctions d'un nombre fini de variables aléatoires, elles sont également aléatoires;
- On peut prévoir le comportement de la particule assez précisément.

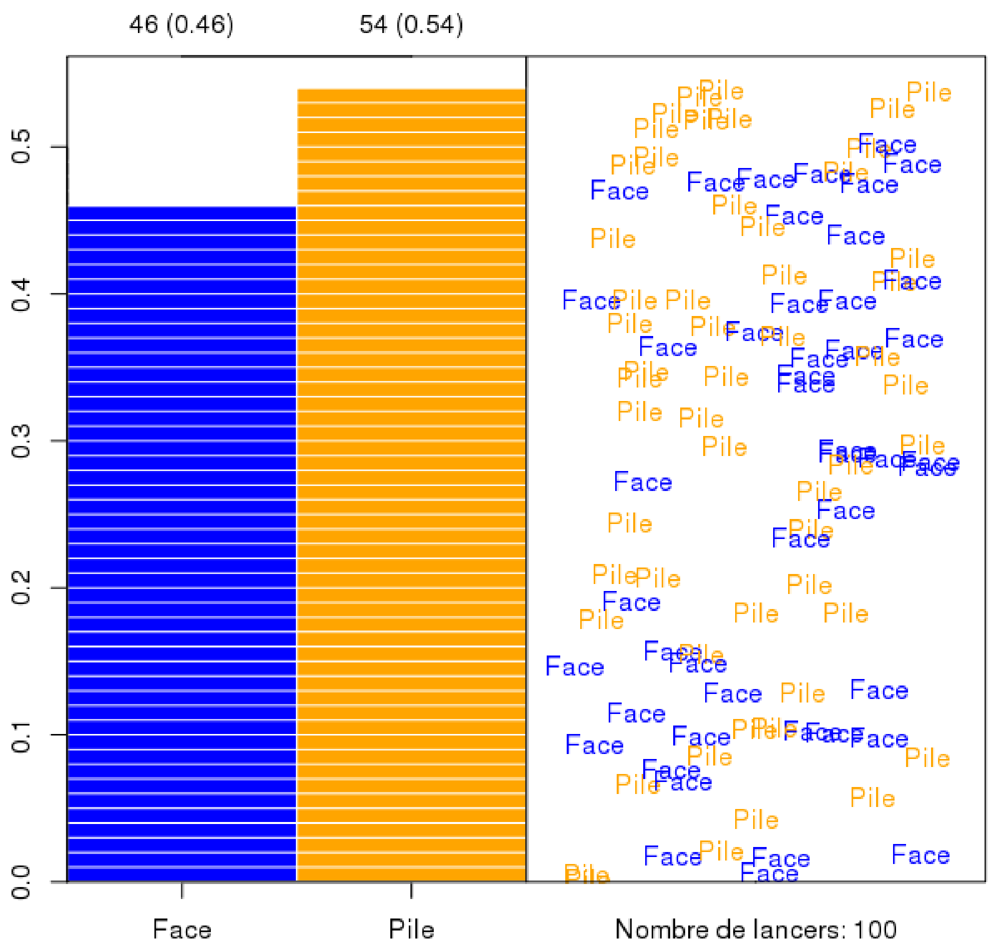

ENRICO 12

8.3.2 Lancers d'une pièce de monnaie

• Le 'pile ou face' est un jeu de hasard se jouant avec une pièce de monnaie.

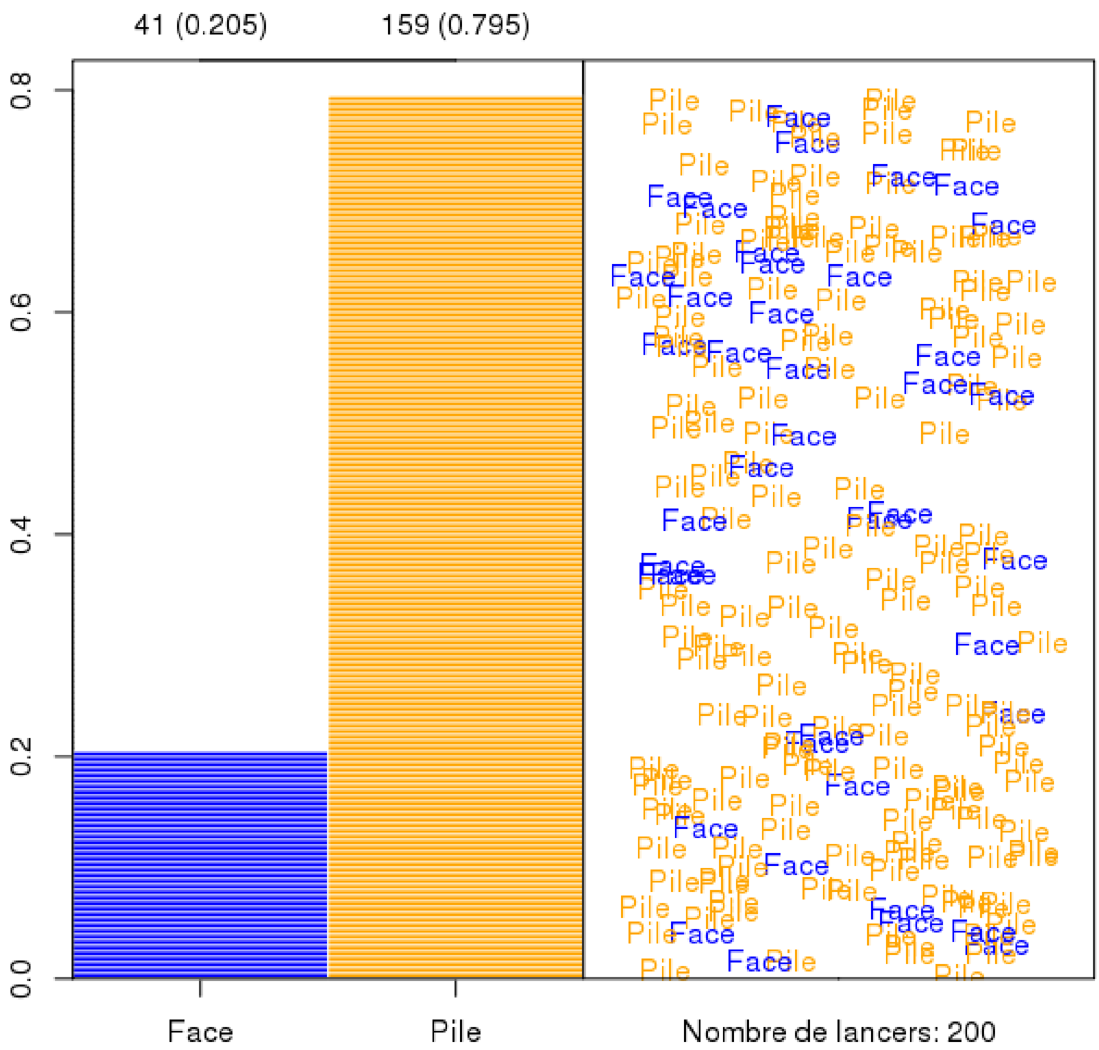
⇝ Le principe du jeu est de lancer en l'air une pièce et de parier sur le côté sorti.

▷ En utilisant une pièce équilibrée, *i.e.* avec une probabilité pour 'pile' de $p = 0.5$:

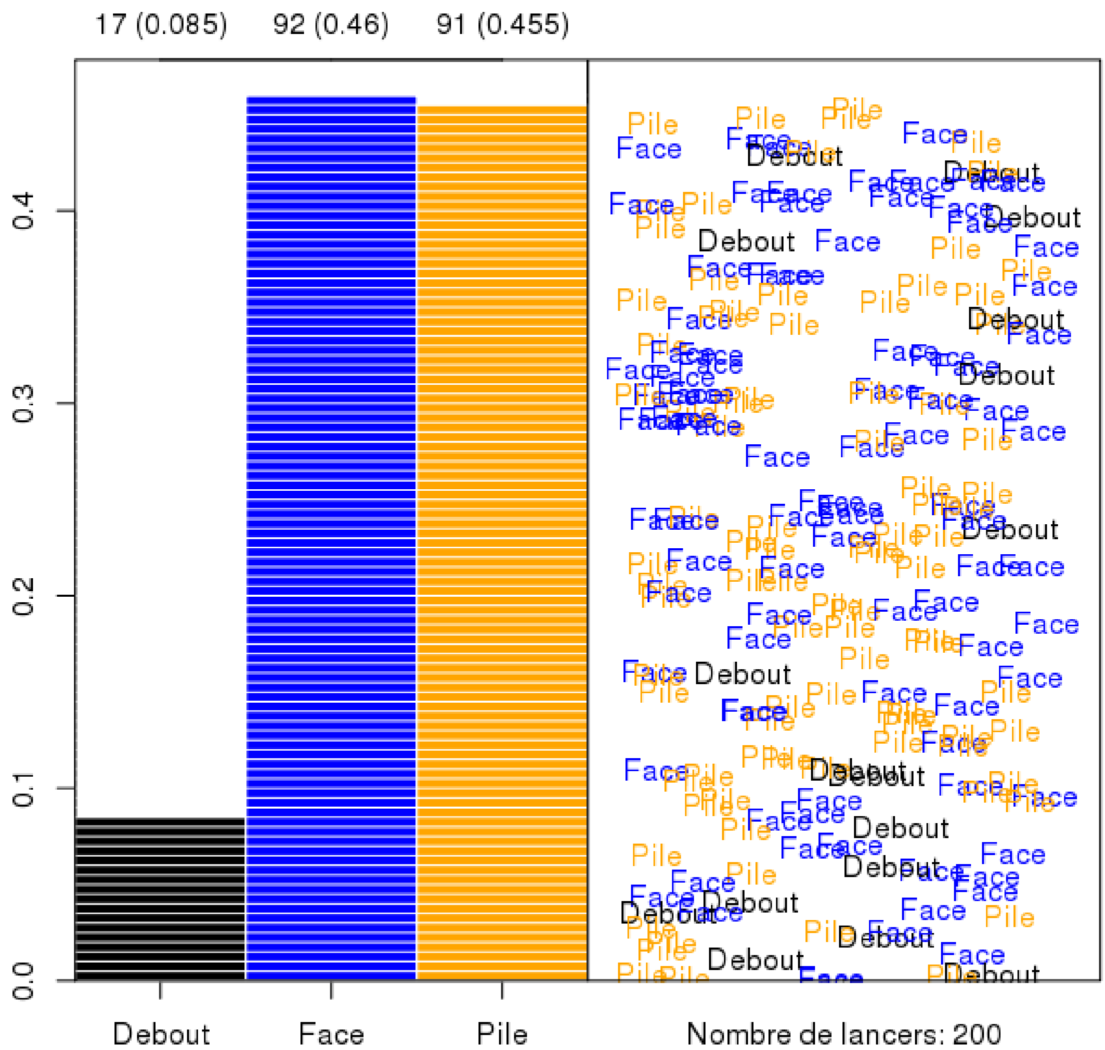


⇝ On n'est jamais exactement sur la valeur théorique, mais toujours autour. Ceci est dû au fait que les résultats sont aléatoires.

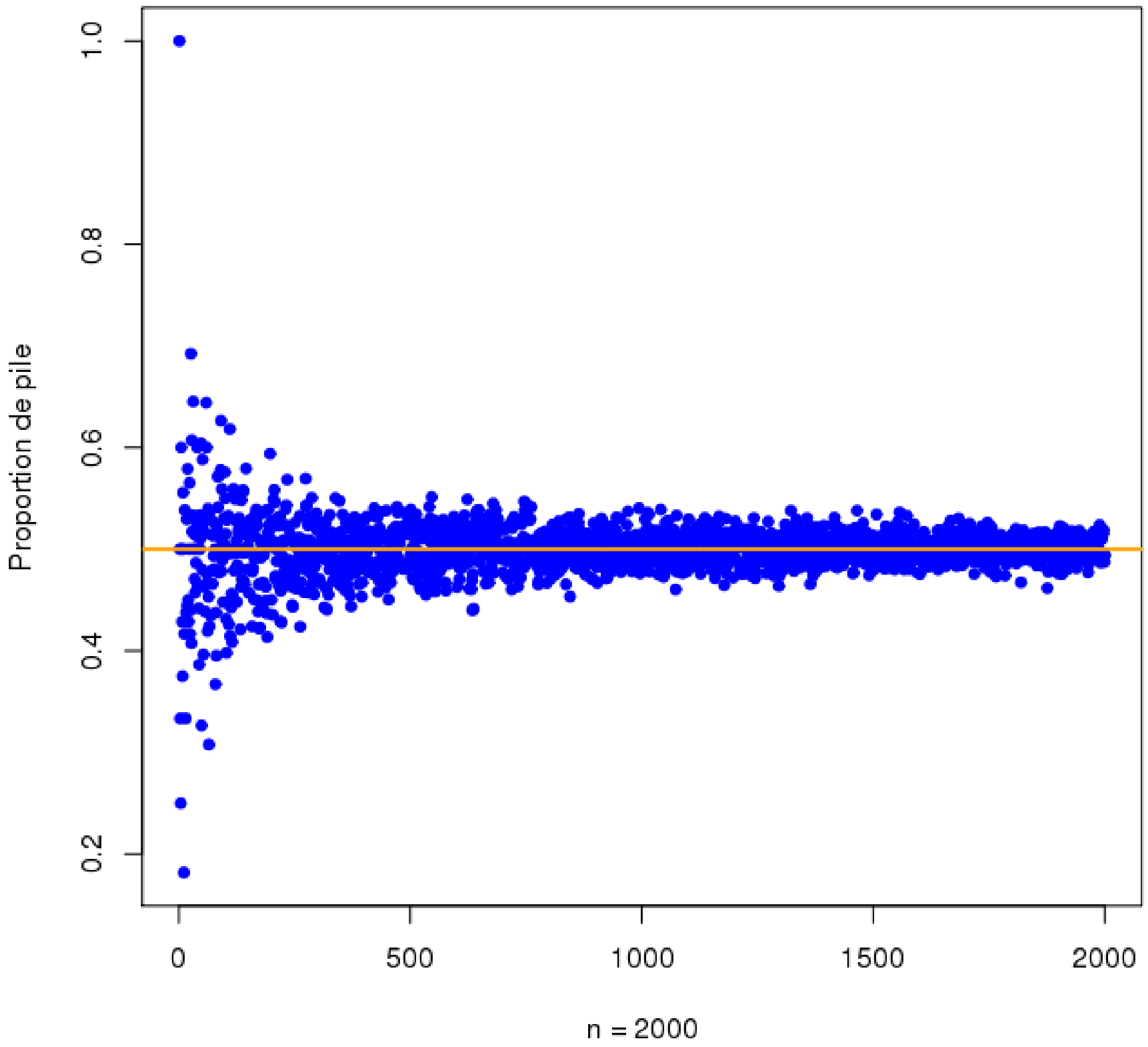
▷ En utilisant une pièce non équilibrée, *e.g.* avec une probabilité pour 'pile' de $p = 0.8$:

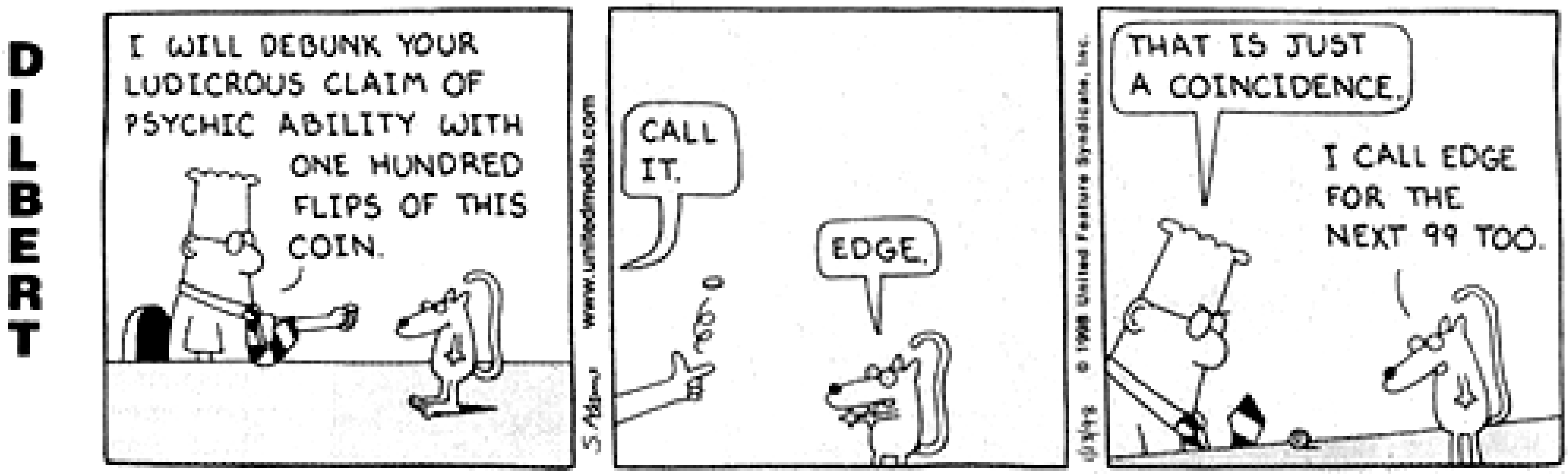


▷ En utilisant une pièce avec une probabilité pour 'pile' ou 'face' de 0.45 et avec une probabilité de rester 'debout' de 0.1:



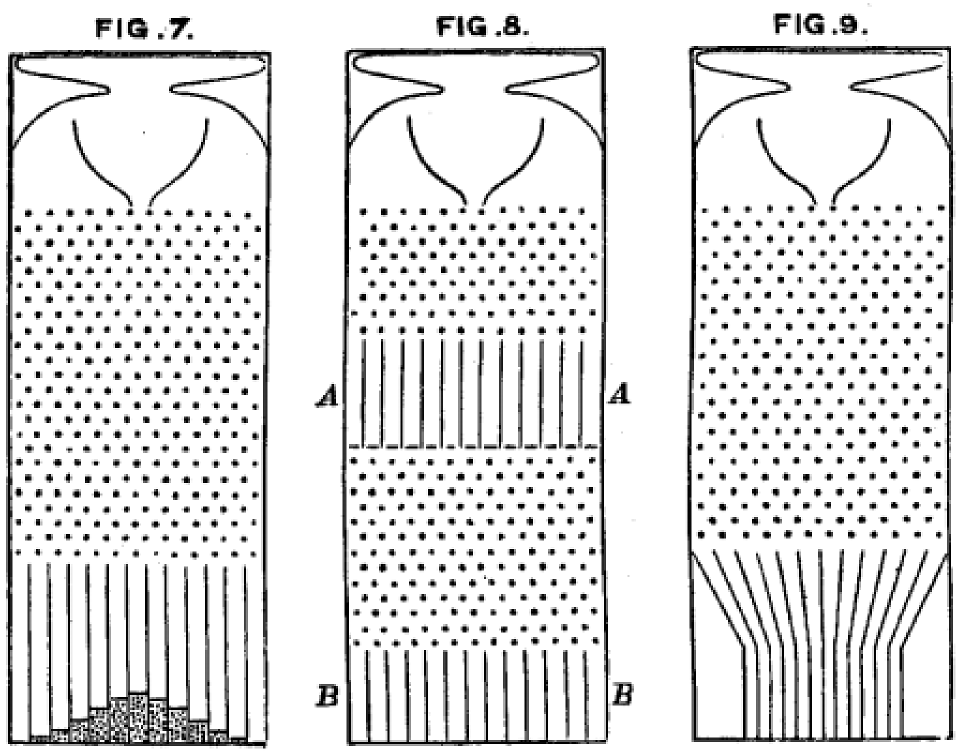
▷ Illustration de la loi des grands nombres en utilisant une pièce équilibrée:

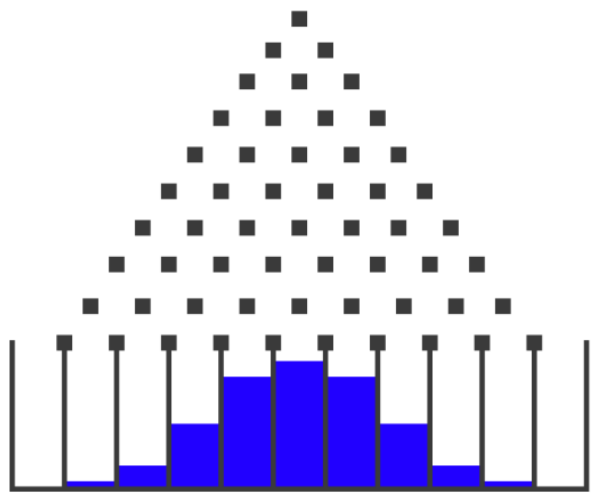



DILBERT
I WILL DEBUNK YOUR LUDICROUS CLAIM OF PSYCHIC ABILITY WITH ONE HUNDRED FLIPS OF THIS COIN.
www.unitedmedia.com
CALL IT.
EDGE.
THAT IS JUST A COINCIDENCE.
I CALL EDGE FOR THE NEXT 99 TOO.

8.3.3 Planche de Galton

• Une planche de Galton (ou crible de Galton ou '*bean machine*' ou '*quincunx*' ou '*Galton box*') est un dispositif inventé en 1894 par Sir Francis Galton qui illustre la convergence d'une distribution binomiale vers une distribution normale.

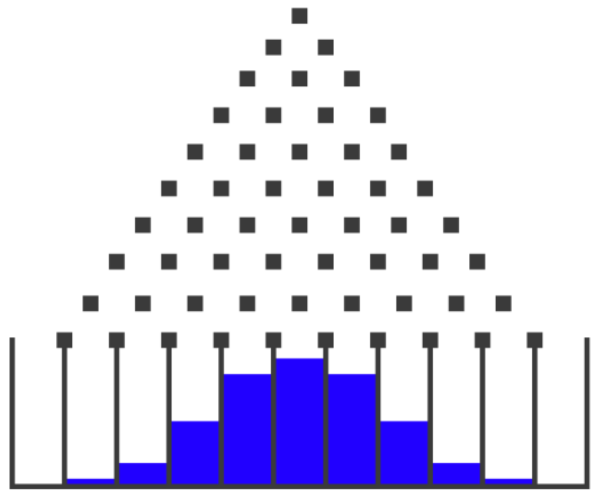




$\rightsquigarrow$ Des billes tombent verticalement sur un assemblage de clous placés en quinconce sur des lignes horizontales et équidistants de leurs voisins immédiats.

$\rightsquigarrow$ Le diamètre des billes est égal à la distance entre les clous. Chaque fois qu'une bille tape un clou, elle a la même probabilité de continuer sa chute à gauche ou à droite.

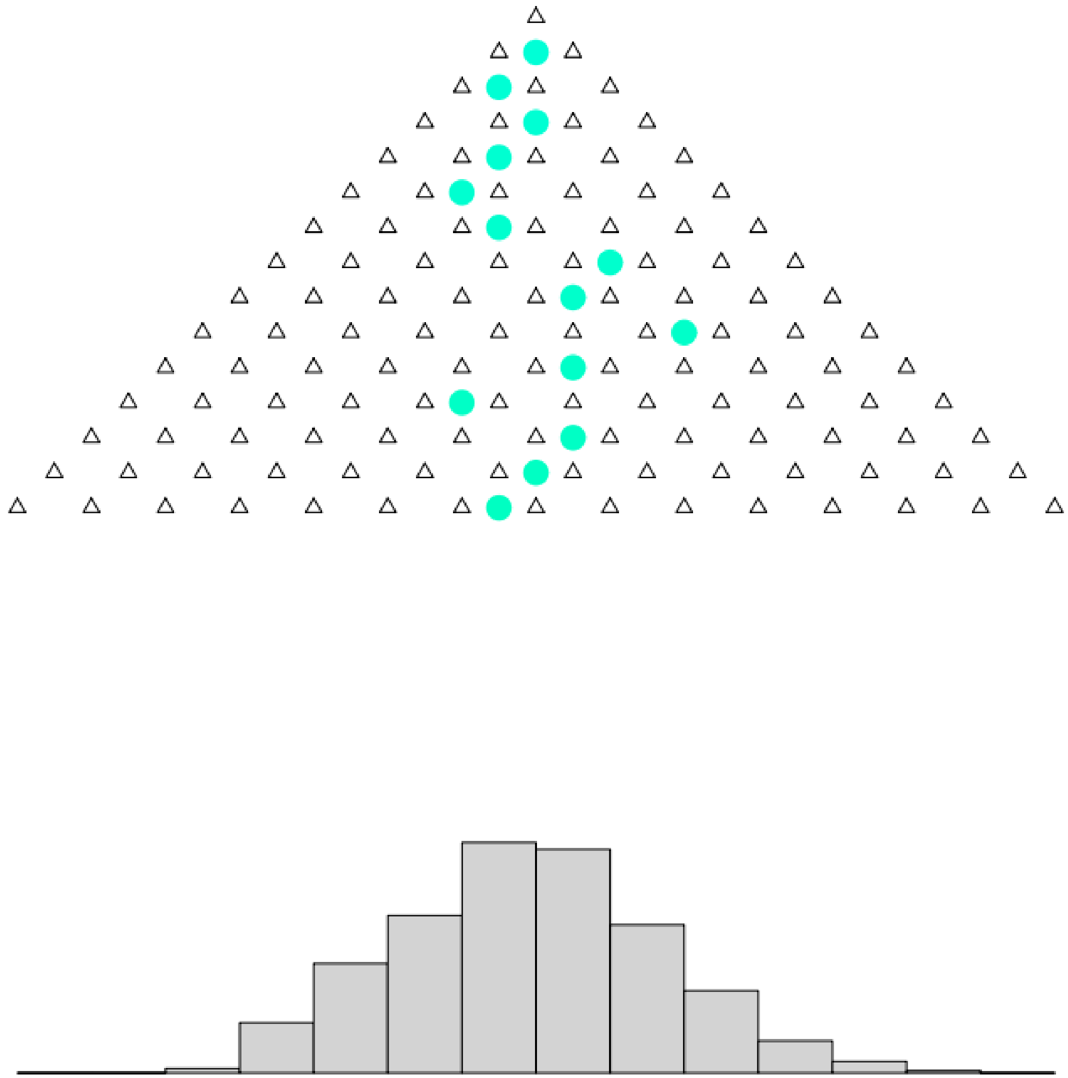
$\rightsquigarrow$ En bas du crible se trouvent des compartiments dans lesquels tombent les billes.



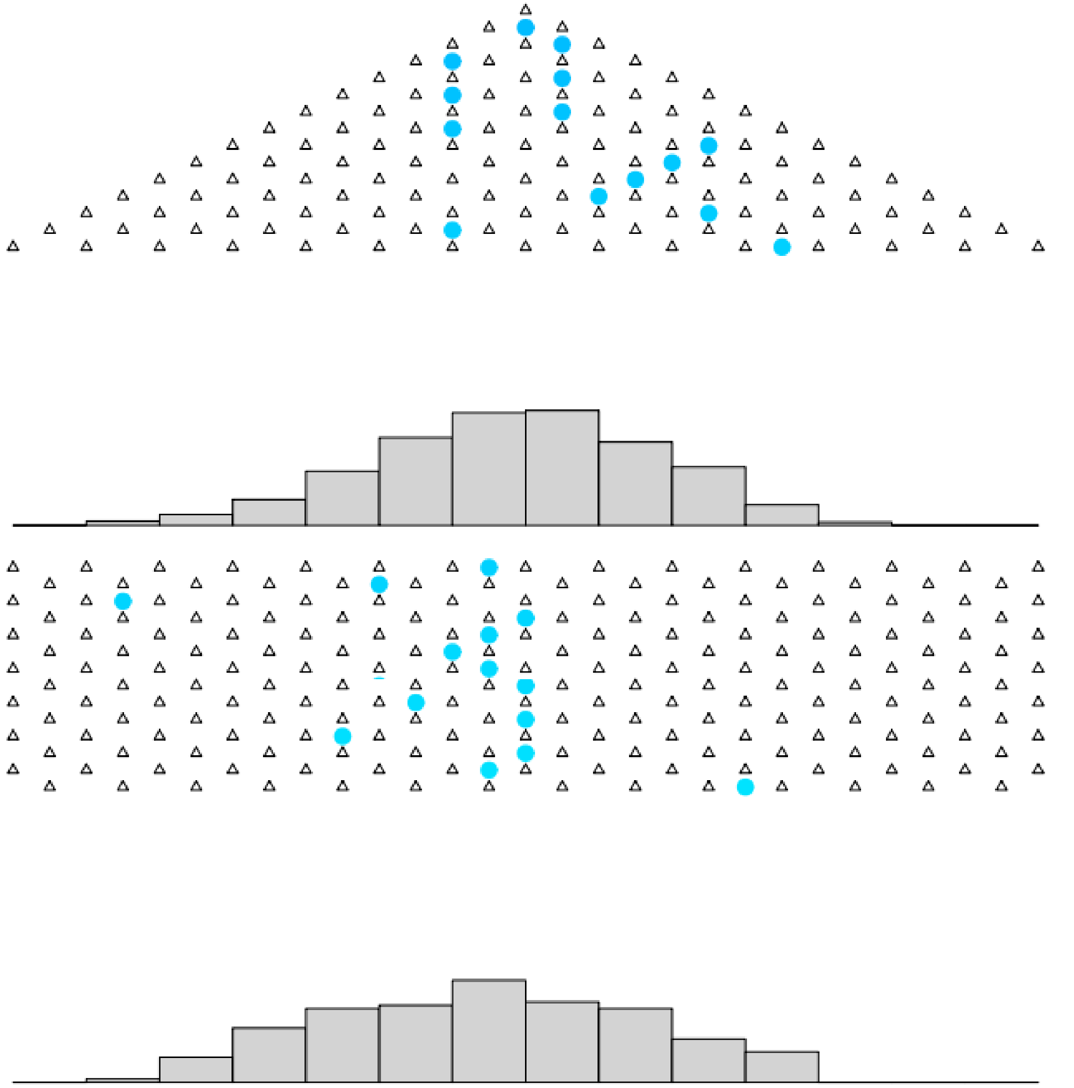
⇝ Si on réalise l'expérience un grand nombre de fois, les billes viennent s'accumuler dans les compartiments et forment ainsi un 'histogramme' (dont la hauteur d'un bâton est proportionnelle au nombre de billes s'y trouvant).

⇝ Ainsi chaque compartiment correspond à un résultat possible d'une expérience binomiale (en tant qu'une expérience de Bernoulli répétée) et on peut remarquer que la répartition des billes dans les compartiments approche la forme d'une courbe d'une distribution normale, ceci étant d'autant plus vrai que le nombre de rangées augmente.

▷ Planche de Galton avec 1 rangée:



▷ Planche de Galton avec 2 rangées:



‘On voit, par cet Essai, que la théorie des probabilités n’est, au fond, que le bon sens réduit au calcul; elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d’instinct, sans qu’ils puissent souvent s’en rendre compte.’

Pierre-Simon de Laplace, 1814

‘Il est remarquable qu’une science qui a commencé par la considération des jeux se soit élevée aux plus importants objets des connaissances humaines.’

Pierre-Simon de Laplace, 1812