MAT4681 - Statistique pour les sciences

Arthur Charpentier

04 - Moyenne, variance (et rappels de maths) # 2

été 2022

Variable aléatoire I

Une variable aléatoire X assigne à chaque état de la nature $\omega \in \Omega$ une valeur (réelle). L'évènement $\{\omega : X(\omega) = a\}$ sera noté $X = a$.

Fonction de masse

La fonction de masse est $f(a) = \mathbb{P}(X = a)$.

Fonction de répartition

La fonction de répartition est $F(a) = \mathbb{P}(X \leq a)$

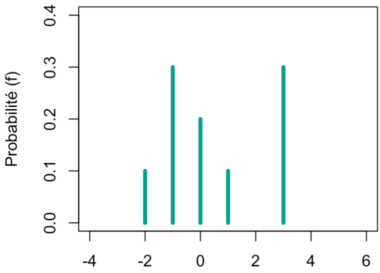
Notons que la fonction de répartition est croissante
(car $\{\omega : X(\omega) \leq a\} \subset \{\omega : X(\omega) \leq b\}$ si $a \leq b$)

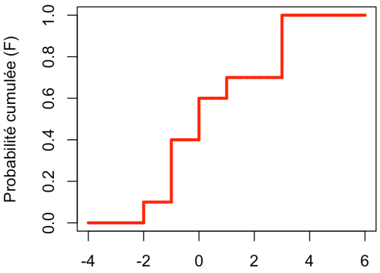
Exemple : Considérons X qui prend 5 valeurs

x	-2	-1	0	1	3
$f(x)$	0.1	0.3	0.2	0.1	0.3
$F(x)$	0.1	0.4	0.6	0.7	1.0

Variable aléatoire II

```
> x = c(-2, -1, 0, 1, 3)
> p = c(0.1, 0.3, 0.2, 0.1, 0.3)
> plot(x, p, type="h")
```





Une fonction de répartition $F : \mathbb{R} \to [0,1]$ vérifie

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

Variable aléatoire III

$$\begin{cases} f(x,y) = \mathbb{P}[\{X = x\} \cap \{Y = y\}] = \mathbb{P}[X = x, Y = y], \;\; \forall \; x, y \\ F(x,y) = \mathbb{P}[\{X \leq x\} \cap \{Y \leq y\}] = \mathbb{P}[X \leq x, Y \leq y], \;\; \forall \; x, y \end{cases}$$

Indépendance

X et Y sont indépendantes ($X \perp\!\!\!\perp Y$) si et seulement si

$$\begin{cases} f(x,y) &= f_X(x) \cdot f_Y(y), \;\; \forall \; x, y \\ F(x,y) &= F_X(x) \cdot F_Y(y), \;\; \forall \; x, y \end{cases}$$

Loi de Bernoulli, $x \in \{0, 1\}$

Exemple: $\boldsymbol{x} = \{1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0\}$

Loi de Bernoulli $\mathcal{B}(p)$, $x \in \{0, 1\}$

$$f(x) = \mathbb{P}(X = x) = \begin{cases} p & \text{si } x = 1, \\ 1 - p & \text{si } x = 0, \\ 0 & \text{sinon.} \end{cases}$$

où $p \in [0, 1]$, soit

$$\mathbb{P}(X = x) = p^x(1 - p)^{1-x}, x \in \{0, 1\}$$

La notation avec p est la plus usuelle, mais on peut aussi utiliser

$$\text{cote (ou } \textit{odds}\text{)} = \frac{\mathbb{P}(X = 1)}{\mathbb{P}(X = 0)} = \frac{p}{1 - p} = e^\theta, \text{ où } \theta \in \mathbb{R}$$

Loi binomiale, $x \in \{0, 1, \cdots, n\}$

Exemple: $\boldsymbol{x} = \{2, 1, 1, 0, 1, 2, 0, 0, 1, 1, 2, 2, 0, 1, 4, 2, 1, 1, 0, 1\}$

Loi binomiale $\mathcal{B}(n, p)$, $x \in \{0, 1, \cdots, n\}$

$$f(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \; x \in \{0, 1, \cdots, n\}$$

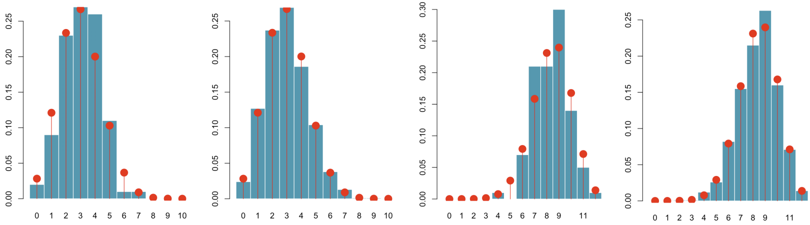
où $p \in [0, 1]$, et $n \in \mathbb{N}_*$.

Note: souvent, n est connu, alors que p est un paramètre inconnu...
Exemple: $f(5)$ avec $X \sim \mathcal{B}(14, 30\%)$ ($f(5) \approx 19.63\%$)

Loi de Bernoulli et loi binomiale

Si $Z_1, \cdots, Z_n \sim \mathcal{B}(p)$ indépendantes,

$$X = \sum_{k=1}^{n} Z_k \sim \mathcal{B}(n, p).$$



Aussi, si $X \sim \mathcal{B}(m, p)$ et $Y \sim \mathcal{B}(n, p)$, avec $X \perp\!\!\!\perp Y$, alors

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

mais si $X \sim \mathcal{B}(m, p)$ et $Y \sim \mathcal{B}(m, q)$, avec $X \perp\!\!\!\perp Y$, alors ... rien

Loi de Poisson, $x \in \{0, 1, \cdots, \}$

Exemple: $\boldsymbol{x} = \{5, 6, 11, 4, 9, 9, 8, 5, 7, 7, 6, 5, 4, 0, 2, 2, 1, 4, 2, 5\}$
Exemple: $\boldsymbol{x} = \{0, 0, 2, 2, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 1, 0, 0, 1, 0, 1\}$

Loi de Poisson $\mathcal{P}(\lambda)$, $x \in \{0, 1, \cdots\} = \mathbb{N}$

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

où $\lambda \in \mathbb{R}_+$.

Note: on peut montrer que $\exp(z) = \sum_{n=0}^{+\infty} \frac{z^n}{n!}$, $\forall z$, donc

$$\sum_{x=0}^{+\infty} \frac{\lambda^x}{x!} = e^{\lambda} \text{ et donc } \sum_{x=0}^{+\infty} \mathbb{P}(X = x) = \sum_{x=0}^{+\infty} e^{-\lambda}\frac{\lambda^x}{x!} = 1$$

Approximation pour la loi binomiale ★★★

Loi binomiale et loi de Poisson

si $pn \to \lambda$ lorsque $n \to \infty$,

$$\lim_{n\to+\infty} \binom{n}{k} p^k (1-p)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}$$

Dans la pratique, la théorie asymptotique est utilisé selon les cas comme une approximation plus ou moins bonne [...] *La raison de l'utilisation généralisée de l'infini est qu'il peut fournir des approximations exploitables dans des circonstances où les résultats exacts sont difficiles voire impossibles à obtenir. L'opération mathématique cruciale qui conduit à ces approximations est le passage à la limite, la limite étant l'état où la notion d'infini apparaît. Les limites intéressantes peuvent être nulles, finies, ou infinies. Les limites nulles ou finies fournissent habituellement des approximations recherchées: des éléments difficiles à évaluer dans un contexte réaliste et fini sont remplacés par leurs limites comme approximation*, Davidson & MacKinnon (1993).

Approximation pour la loi binomiale

Example 2 Un hôpital a 12000 patients agés, et on a estimé que la probabilité qu'un patient souffre d'un accident cardiaque pendant une journée était de 1/8000. L'hôpital possède seulement trois machines respiratoires nécessaires pour ces accidents cardiaques et se demande si son équipement sera suffisant pour une journée particulière

$X_i = \mathbf{1}(\text{patient } i \text{ souffre d'un accident cardiaque})$,
$X_i \sim \mathcal{B}(p)$ avec $p = 1/8000$
$X = \sum_{i=1}^{12000} X_i \sim \mathcal{B}(n,p)$ avec $n = 12000$. On veut $\mathbb{P}[X \leq 3]$?

```
> pbinom(3,12000,1/8000)
[1] 0.9343693133
```

Notons que $X \approx \mathcal{P}(\lambda)$ où $\lambda = pn = 1.5$

```
> ppois(3,1.5)
[1] 0.9343575456
```

Approximation pour la loi binomiale

```
> dbinom(0:2, size=10,prob=.2)
[1] 0.1073742 0.2684355 0.3019899
> dpois(0:2,2)
[1] 0.1353353 0.2706706 0.2706706
```

'Preuve' numérique

n	p	λ	$p_B(0)$	$p_P(0)$	$p_B(1)$	$p_P(1)$	$p_B(2)$	$p_P(2)$
10	0.2	2	0.107	0.135	0.268	0.271	0.302	0.271
20	0.1	2	0.122	0.135	0.270	0.271	0.285	0.271
50	0.04	2	0.13	0.135	0.271	0.271	0.276	0.271
100	0.02	2	0.133	0.135	0.271	0.271	0.273	0.271
200	0.01	2	0.134	0.135	0.271	0.271	0.272	0.271

Approximation pour la loi binomiale ★★★

Preuve

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \approx \lim_{n\to\infty} \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \lim_{n\to\infty} \frac{n^k + O\left(n^{k-1}\right)}{k!} \frac{\lambda^k}{n^k} \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \lim_{n\to\infty} \frac{\lambda^k}{k!} \left(1-\frac{\lambda}{n}\right)^{n-k}$$

Comme

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda} \text{ et } \lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{-k} = 1$$

on en déduit

$$\binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{\lambda^k e^{-\lambda}}{k!}.$$

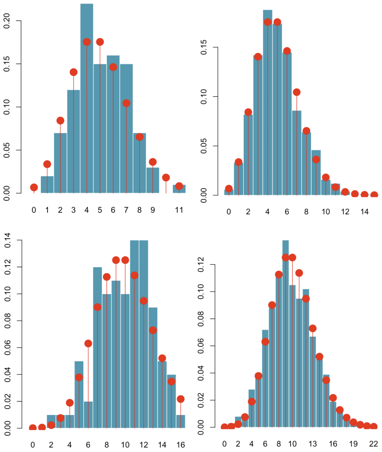
Loi de Poisson

The Poisson distribution $\mathcal{P}(\lambda)$, $\lambda > 0$, has distribution

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

Further, if $X_1 \sim \mathcal{P}(\lambda_1)$ and $X_2 \sim \mathcal{P}(\lambda_2)$ are independent, then $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$
A recursive equation can be obtained

$$\frac{\mathbb{P}(X = x + 1)}{\mathbb{P}(X = x)} = \frac{\lambda}{x + 1}$$

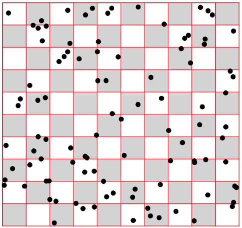


Note: $\mathbb{P}(N = 0) = e^{-\lambda}$, e.g. if $\lambda = 1$, $\mathbb{P}(N = 0) \simeq 36.788\%$ (and $\mathbb{P}(N > 0) \simeq 63.212\%$)

Application de la loi de Poisson

Considérons un échiquier 10×10, lançons $n = 100$ pierres, et comptons le nombre de pierre sur chaque case,

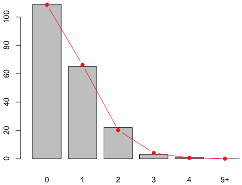
```
> data.frame(N,F=table(nb_cell),P=c
    (dpois(0:4,1),1-ppois(4,1)))
   N  F     P
1  0 36 36.78
2  1 39 36.78
3  2 16 18.39
4  3  7  6.13
5  4  2  1.53
6 5+  0  0.37
> 36+39+16+7+2
[1] 100
```



Application de la loi de Poisson

Von Bortkiewicz (1898) a étudié le nombre de morts par ruade de cheval dans l'armée prussienne de 1875 à 1894 dans 10 corps de cavalerie (soit 200 corps annuels) .

```
> data.frame(N,F=table(ruades),P=c(
    dpois(0:4,mean(ruades)),1-ppois
    (4,mean(ruades))))
    N    F        P
1   0  109  108.67
2   1   65   66.21
3   2   22   20.22
4   3    3    4.11
5   4    1    0.63
6  5+    0    0.08
> 109+65+22+3+1+0
[1] 200
```



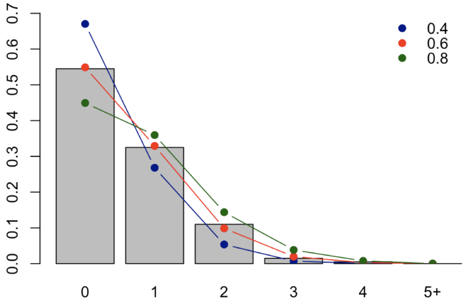
Application de la loi de Poisson

- choisir une loi de probabilité : loi de Poisson
- si la loi dépend d'un ou plusieurs paramètres, il faut trouver comment choisir ce paramètre [estimation]
- comparer les fréquences observées avec celles prédites par le modèle [prédiction]
- définir un seuil d'acceptabilité et tester si la différence est acceptables [test]

Ici on peut regarder les fréquences de plusieurs lois $\mathcal{P}(\lambda)$

```
> g = function(p) c(dpois(0:4,p),1-ppois(4,p))
> data.frame(N,F=table(ruades),Prob=table(ruades)/sum(
   N))
    N   F   Prob  g(.2)  g(.4)  g(.6)  g(.8)  g(1)
1   0 109  0.545  0.819  0.670  0.549  0.449 0.368
2   1  65  0.325  0.164  0.268  0.329  0.359 0.368
3   2  22  0.110  0.016  0.054  0.099  0.144 0.184
4   3   3  0.015  0.001  0.007  0.020  0.038 0.061
5   4   1  0.005  0.000  0.001  0.003  0.008 0.015
6  5+   0  0.000  0.000  0.000  0.000  0.001 0.004
```

Application de la loi de Poisson




```
>  n
[1] 109    65    22     3     1     0
> 200*g(.61)
[1] 108.6  66.2  20.2   4.2   0.6   0.0
```

Loi Geometrique, $x \in \mathbb{N}_*$ I

En lien avec la suite/série géométrique, $u_{n+1} = q \times u_n$ avec comme premier terme u_0, telle que $u_n = q^n \times u_0$,

$$S_n = \sum_{k=0}^{n} u_k = a \sum_{k=0}^{n} q^k = a \frac{1 - q^{n+1}}{1 - q} \to \frac{a}{1 - q}$$

lorsque $n \to \infty$, si $q \in [0, 1)$.

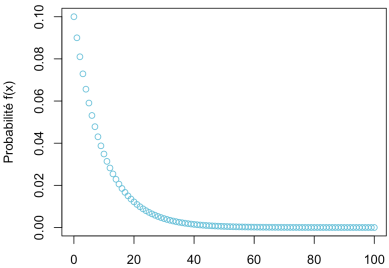
Loi Géométrique $\mathcal{G}(p)$ sur $\mathbb{N}_*$

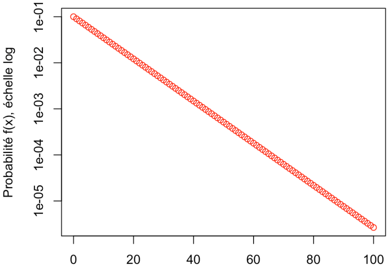
$$f(x) = \mathbb{P}(X = x) = p(1 - p)^{x-1} \text{ pour } x = 1, 2, \cdots$$

où $p \in [0, 1]$, avec $F(x) = \mathbb{P}(X \leq x) = 1 - p^x$.

Loi Geometrique, $x \in \mathbb{N}_*$ II

$$f(x) = p\,(1-p)^{x-1} \text{ ou } \log f(x) = \log p + (x-1)\log(1-p)$$





Les points $\{(x, \log f(x))\}$, sont alignés suivant une droite (de pente négative, $\log(1-p)$).

Loi Geometrique, $x \in \mathbb{N}_*$

Si $Z_1, \cdots, Z_n, \cdots \sim \mathcal{B}(p)$ indépendantes, ntons X la date du premier instant où '1' apparaît,

$$X = \inf_{n \in \mathbb{N}_*} \left\{ n : \; Z_n = 1 \right\} \sim \mathcal{G}(p)$$

Cette distribution vérigie

$$\frac{f(x+1)}{f(x)} = \frac{\mathbb{P}(X = x+1)}{\mathbb{P}(X = x)} = 1 - p \text{ (= constante) pour } x \geq 1$$

(autre) Loi Géométrique $\mathcal{G}(p)$ sur $\mathbb{N}$

$$f(x) = \mathbb{P}(X = x) = p(1-p)^x \text{ pour } x = 0, 1, 2, \cdots$$

Retour sur l'approximation de la loi de Poisson

Soient $Z_i \sim \mathcal{B}(p)$, indépendantes et notons

$$X = \inf_{n\in\mathbb{N}_*} \{n : \ Z_n = 1\} \text{ et } X_n = \sum_{i=1}^{n} Z_i$$

alors $X \sim \mathcal{G}(p)$ et $X_n \sim \mathcal{B}(n,p)$.
Si $n \cdot p \simeq \lambda$, alors $X_n \approx \mathcal{P}(\lambda)$

$$\begin{aligned}\mathbb{P}(X_n = 0) = \mathbb{P}(X > n) &= (1-p)^n \\ &\simeq \left(1 - \frac{\lambda}{n}\right)^n \\ &\approx e^{-\lambda}\end{aligned}$$

	Number of years without catastrophes				
	10	20	50	100	200
10	65.1%	40.1%	18.3%	9.6%	4.9%
20	87.8%	64.2%	33.2%	18.2%	9.5%
50	99.5%	92.3%	63.6%	39.5%	22.5%
100	99.9%	99.4%	86.7%	63.4%	39.5%
200	99.9%	99.9%	98.2%	86.6%	63.3%

Loi de Zipf, $x \in \{1, 2, \cdots, n\}$ I

Loi de Zipf $\mathcal{Z}(\alpha)$

Définie pour $x \in \{1, 2, \cdots, n\}$ par $f(x) \propto x^{-\alpha}$ où $\alpha > 0$.

Si on note $H_{n,s} = \displaystyle\sum_{k=1}^{n} \frac{1}{k^s}$ le n-ème nombre harmonique généralisé,

$$f(x) = \frac{x^{-\alpha}}{H_{n,\alpha}},\ x \in \{1, 2, \cdots, n\}$$

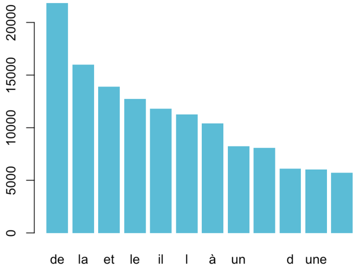
Log-log-Pareto plot

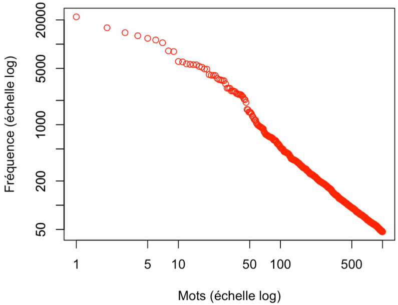
Si $f(x) = H_{n,\alpha}^{-1} x^{-\alpha}$, $\log f(x) = -\log H_{n,\alpha} - \alpha \log x$

Loi de Zipf, $x \in \{1, 2, \cdots, n\}$ II

Si $f(x) = H_{n,\alpha}^{-1} x^{-\alpha}$, $\log f(x) = -\log H_{n,\alpha} - \alpha \log x$. Aussi, les points $\{(\log x, \log f(x))\}$, sont alignés suivant une droite (de pente négative, α).

Application en linguistique, *Les Misérables* de Victor Hugo, avec (environ) 551,536 mots





Loi de Benford, $x \in \{1, 2, \cdots, 9\}$ I

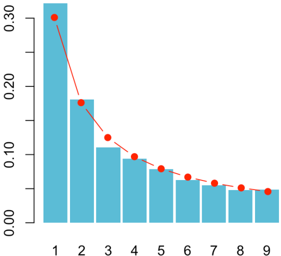
Loi de Benford

Si X prend les valeurs $\{1, 2, \cdots, 9\}$, et

$$f(x) = \log_{10}(x+1) - \log_{10}(x) = \log_{10}\left(1 + \frac{1}{x}\right)$$

x	1	2	3	4	5	6	7	8	9
$f(x)$	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%

```
> library(maps)
> data(world.cities)
> X = as.integer(substr(world.
    cities[,3],1,1))
> hist(X)
```



Espérance mathématique I

Espérance

Si X prend les valeurs $\{x_1, x_2, \cdots, x_k\}$, avec la fonction de masse f, son espérance est

$$\mathbb{E}[X] = \sum_{i=1}^{k} f(x_i) \cdot x_i = \mathbb{P}(X = x_1) \cdot x_1 + \cdots + \mathbb{P}(X = x_k) \cdot x_k$$

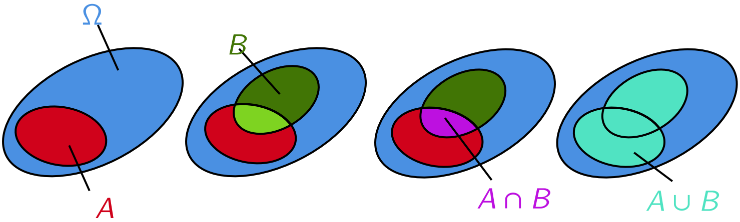
Espérance d'une somme

Quelles que soient les variables X et Y, et les constantes a, b, c

$$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$$

Espérance mathématique II

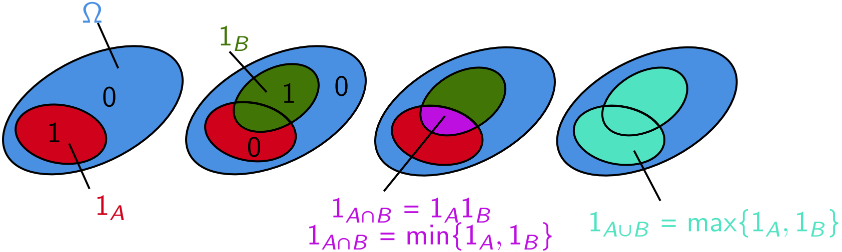
$$A \cup B = \{x : x \in A \text{ ou } x \in B\} \text{ et } A \cap B = \{x : x \in A \text{ et } x \in B\}$$



$$\mathbf{1}_A = \begin{cases} 1 \text{ si } A \text{ survient} \\ 0 \text{ si } A \text{ ne survient pas} \end{cases}$$

Espérance mathématique III

$$\mathbf{1}_A = \begin{cases} 1 \text{ si } A \text{ survient} \\ 0 \text{ si } A \text{ ne survient pas} \end{cases}$$



Alors $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}[A]$ et

$$\mathbb{E}[\mathbf{1}_{A\cup B}] = \mathbb{P}[A\cup B] = \mathbb{E}[\mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A\cap B}] = \mathbb{E}[\mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A\mathbf{1}_B]$$

Espérance mathématique IV

On lance un dé (à 6 faces), et on note X la face apparente
Que vaut $\mathbb{E}[X]$?

Réponse:

$$\mathbb{E}[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{7}{2} = 3.5$$

```
> sum(1:6)/6
[1] 3.5
```

On lance deux dés (à 6 faces), et on note X_1 et X_2 les faces apparentes. Que vaut $\mathbb{E}[X_1 + X_2]$?

Réponse 1: on regarde la loi de $X_1 + X_2$

Espérance mathématique V

$x_1 \backslash x_2$	1	2	3	4	5	6	
6	1/36	1/36	1/36	1/36	1/36	1/36	
5	1/36	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36	2/36
3	1/36	1/36	1/36	1/36	1/36	1/36	3/36
2	1/36	1/36	1/36	1/36	1/36	1/36	4/36
1	1/36	1/36	1/36	1/36	1/36	1/36	5/36
		1/36	2/36	3/36	4/36	5/36	6/36

```
> sum((2:12)*c(1:6,5:1)/36)
[1] 7
```

Réponse 2: l'espérance est linéaire

$$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = \frac{7}{2} + \frac{7}{2} = 7$$

Espérance mathématique VI

Considérons X qui prend 5 valeurs

x	-2	-1	0	1	3
$f(x)$	0.1	0.3	0.2	0.1	0.3

Que vaut $\mathbb{E}[X]$?

Réponse:

$$\mathbb{E}[X]=\sum_x x\cdot f(x)=(-2)\times\frac{1}{10}+(-1)\times\frac{3}{10}+0\times\frac{2}{10}+1\times\frac{1}{10}+3\times\frac{3}{10}$$

soit

$$\mathbb{E}[X]=\frac{-2-3+0+1+9}{10}=\frac{1}{2}$$

```
> x = c(-2, -1, 0, 1, 3)
> p = c(0.1, 0.3, 0.2, 0.1, 0.3)
> sum(x*p)
[1] 0.5
```

Espérance mathématique VII

Que vaut $\mathbb{E}[|X|]$?

Réponse:

$$\mathbb{E}[|X|] = \sum_x |x|\cdot f(x) = |-2|\times\frac{1}{10}+|-1|\times\frac{3}{10}+0\times\frac{2}{10}+1\times\frac{1}{10}+3\times\frac{3}{10}$$

soit

$$\mathbb{E}[|X|] = \frac{2+3+0+1+9}{10} = \frac{3}{2}$$

```
> x = c(-2, -1, 0, 1, 3)
> p = c(0.1, 0.3, 0.2, 0.1, 0.3)
> sum(abs(x)*p)
[1] 1.5
```

Notons que $\mathbb{E}[|X|] \neq |\mathbb{E}[X]|$.

Espérance mathématique VIII

Que vaut $\mathbb{E}[X^2]$?

Réponse:

$$\mathbb{E}[X^2]=\sum_x x^2\cdot f(x)=2^2\times\frac{1}{10}+1^2\times\frac{3}{10}+0\times\frac{2}{10}+1^2\times\frac{1}{10}+3^2\times\frac{3}{10}$$

soit

$$\mathbb{E}[X^2]=\frac{2^2+3^2+0^2+1^2+9^2}{10}=\frac{7}{2}$$

```
> x = c(-2, -1, 0, 1, 3)
> p = c(0.1, 0.3, 0.2, 0.1, 0.3)
> sum(x^2*p)
[1] 3.5
```

Notons que $\mathbb{E}[X^2]\neq\mathbb{E}[X]^2$.

Espérance mathématique IX

Espérance d'un produit, avec $X \perp\!\!\!\perp Y$

Si les variables X et Y sont indépendantes

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Preuve soit $Z = XY$,

$$\mathbb{P}[Z = z] = \sum_{(x,y):xy=z} \mathbb{P}[X = x, Y = y] = \sum_{(x,y):xy=z} \mathbb{P}[X = x] \cdot \mathbb{P}[Y = y]$$

car $X \perp\!\!\!\perp Y$, et donc

$$\mathbb{E}[Z] = \sum_{z} z \cdot \mathbb{P}[Z = z] = \sum_{z} \sum_{(x,y):xy=z} x\mathbb{P}[X = x] \cdot y\mathbb{P}[Y = y]$$

$$\mathbb{E}[Z] = \sum_{x} x\mathbb{P}[X = x] \cdot \sum_{y} y\mathbb{P}[Y = y] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Espérance mathématique X

Espérance de la moyenne

Si $X_1, \cdots, X_n$ sont des variables indépendantes de même espérance μ, si $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$, $\mathbb{E}(\overline{X}) = \mu$.

Preuve:

$$\mathbb{E}(\overline{X}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\mathbb{E}(X_i)}_{=\mu} = \frac{n\mu}{n} = \mu$$

Variance (et écart-type) I

Variance

La variance de X est définie par

$$\text{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big]$$

Variance

Si X prend les valeurs $\{x_1, x_2, \cdots, x_k\}$, avec la fonction de masse f, si $\mu = \mathbb{E}[X]$, sa variance est définie par

$$\text{Var}[X] = \sum_{i=1}^{k} f(x_i) \cdot \big(x_i - \mu\big)^2 \text{ où } \mu = \sum_{i=1}^{k} f(x_i) \cdot x_i$$

Variance (et écart-type) II

La variance est positive

Quelle que soit X, $\text{Var}[X] \geq 0$.

Preuve $\text{Var}[X] = \sum_{i=1}^{k} \underbrace{f(x_i)}_{\geq 0} \cdot \underbrace{(x_i - \mu)^2}_{\geq 0} \geq 0.$

Écart-type

L'écart-type de X est défini par

$$\sigma(X) = \sqrt{\text{Var}(X)}$$

Variance (et écart-type) III

Exemple : Considérons X qui prend 5 valeurs

x	-2	-1	0	1	3
$f(x)$	0.1	0.3	0.2	0.1	0.3

Quelle est la variance de X ?

On avait vu que

$$\mathbb{E}[X]=\sum_x x\cdot f(x)=(-2)\times\frac{1}{10}+(-1)\times\frac{3}{10}+0\times\frac{2}{10}+1\times\frac{1}{10}+3\times\frac{3}{10}$$

soit

$$\mathbb{E}[X]=\frac{-2-3+0+1+9}{10}=\frac{1}{2}$$

Aussi,

$$\text{Var}[X]=\sum_x\left(x-\frac{1}{2}\right)^2\cdot f(x)=\left(-2-\frac{1}{2}\right)^2\times\frac{1}{10}+\left(-1-\frac{1}{2}\right)^2\times\frac{3}{10}+\cdots+\left(3\right.$$

Variance (et écart-type) IV

soit

$$\text{Var}[X] = \left(\frac{-5}{2}\right)^2 \times \frac{1}{10} + \left(\frac{-3}{2}\right)^2 \times \frac{3}{10} + \cdots + \left(\frac{5}{2}\right)^2 \times \frac{3}{10} = \frac{13}{2}$$

```
> x = c(-2, -1, 0, 1, 3)
> p = c(0.1, 0.3, 0.2, 0.1, 0.3)
> sum((x-1/2)^2*p)
[1] 3.25
```

Variance - formule de König-Huygens

La variance de X vérifie

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Variance (et écart-type) V

Preuve: pour rappel $(a-b)^2 = a^2 - 2ab + b^2$,

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

Si $\text{Var}[X] = 0$, que peut-on dire de X ?

Si X prend les (vraiment) le valeurs $\{x_1, x_2, \cdots, x_k\}$, avec $f(x_i) > 0$,

$$\text{Var}[X] = \sum_{i=1}^{k} \underbrace{f(x_i)}_{>0} \cdot \underbrace{(x_i - \mu)^2}_{\geq 0} \geq 0$$

$\text{Var}[X] = 0$ si et seulement si $x_i - \mu = 0$, $\forall i$: X est constante.

Variance (et écart-type) VI

Variance

Soient $a, b \in \mathbb{R}$, $\text{Var}(a + bX) = b^2\text{Var}(X)$

Preuve:

$$
\begin{aligned}
\text{Var}(a + bX) = & \underbrace{\mathbb{E}\Big[(a + bX)^2\Big]}_{a^2+2ab\mathbb{E}[X]+b^2\mathbb{E}[X^2]} - \underbrace{\mathbb{E}[a + bX]^2}_{(a+b\mathbb{E}[X])^2} \\
& = \left(a^2 + 2ab\mathbb{E}[X] + b^2\mathbb{E}[X^2]\right) - \left(a^2 + 2ab\mathbb{E}[X] + b^2\mathbb{E}[X]^2\right. \\
& = b^2\left(\mathbb{E}[X^2] - \mathbb{E}[X]^2\right) = b^2\text{Var}(X)
\end{aligned}
$$

Variance (et écart-type) VII

Variance d'une somme, $X \perp\!\!\!\perp Y$

Si $X \perp\!\!\!\perp Y$, $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.

Preuve:

$$\text{Var}\big(X+Y\big) = \mathbb{E}\big[(X+Y)^2\big] - \mathbb{E}\big[(X+Y)\big]^2$$

$$\underbrace{\mathbb{E}(X^2) - \mathbb{E}(X)^2}_{=\text{Var}(X)} + \underbrace{\mathbb{E}(Y^2) - \mathbb{E}(Y)^2}_{=\text{Var}(Y)} + 2\underbrace{\Big(\mathbb{E}(XY) - \mathbb{E}(XY)\Big)}_{=0}$$

i.e. $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.

Variance (et écart-type) VIII

Variance de la moyenne

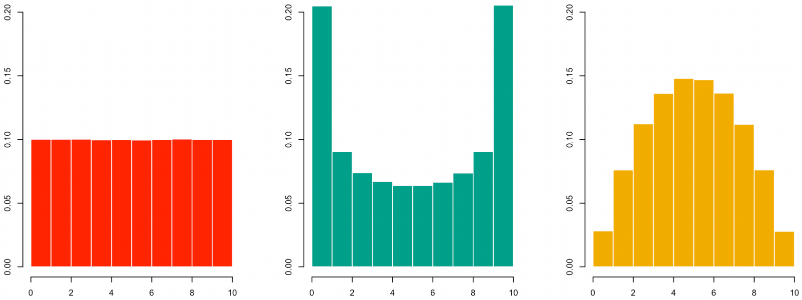
Si $X_1, \cdots, X_n$ sont des variables indépendantes de même variance σ^2, si $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$, $\text{Var}(\overline{X}) = \dfrac{\sigma^2}{n}$.

Preuve:

$$\text{Var}(\overline{X}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\underbrace{\text{Var}\left(X_i\right)}_{=\sigma^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Variance (et écart-type) IX

On considère les trois distributions suivantes (de même moyenne), associées respectivement aux variables X, Y, Z



Ordonnez les trois distributions en fonction de leur variance

Réponse: $\text{Var}[Z] < \text{Var}[X] < \text{Var}[Y]$.

Variable aléatoire continue I

Soient $X_1, \cdots, X_n, \cdots$ des variables de Bernoulli, $X_i \in \{0, 1\}$.

$$S_n = \sum_{i=1}^{n} X_i = X_1 + \cdots + X_n \in \{0, 1, \cdots, n\}$$

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{X_1 + \cdots + X_n}{n} \in \left\{0, \frac{1}{n}, \frac{2}{n}, \cdots, \frac{n-1}{n}, 1\right\}$$

Si "$\overline{X}_n \to X$" quand $n \to \infty$, alors $X \in [0, 1]$.

Si "$\sqrt{n}(\overline{X}_n - 1/2) \to X$" quand $n \to \infty$, alors $X \in \mathbb{R}$.

Examples pratiques : X durée de vie d'une ampoule électrique, X âge d'une personne quand elle décède

Variable aléatoire continue II

Fonction de répartition

La fonction de répartition est $F(a) = \mathbb{P}(X \leq a)$

Densité

La densité est $f(a) = \left.\frac{dF(x)}{dx}\right|_{x=a} = F'(a)$

Pour rappel,

$$f(a) = F'(a) = \lim_{h\to 0} \frac{F(a+h) - F(a)}{h} = \lim_{h\to 0} \frac{1}{h} \cdot \mathbb{P}(X \in [a, a+h])$$

ou

$$F(a) = \int_{-\infty}^{a} f(x)dx$$

Variable aléatoire continue III

$$\mathbb{P}(X \in [a,b]) = F(b) - F(a) = \int_a^b f(x)dx$$

Soit X une variable aléatoire à valeurs dans $[0,2]$ telle que $f(x) = \alpha x^2$ pour $x \in [0,2]$.
Que vaut α ? Que vaut $\mathbb{P}[X \in [1,2]]$?

X une variable aléatoire à valeurs dans $[0,2]$ donc

$$\mathbb{P}[X \in [0,2]] = 1 = \int_0^2 f(x)dx = \int_0^2 \alpha x^2 dx = \alpha \left[\frac{x^3}{3}\right]_0^2 = \frac{8\alpha}{3}$$

autrement dit $\alpha = \frac{3}{8}$. Aussi, la fonction de répartition est, pour $x \in [0,2]$,

$$F(x) = \int_0^x \frac{3}{8} t^2 dt = \frac{3}{8}\left[\frac{t^3}{3}\right]_0^x = \frac{x^3}{8}$$

Variable aléatoire continue IV

avec $F(x) = 0$ pour $x < 0$ et $F(X) = 1$ pour $x > 2$.
Enfin,

$$\mathbb{P}[X \in [1,2]] = F(2) - F(1) = \frac{2^3}{8} - \frac{1^3}{8} = \frac{7}{8}$$

On peut montrer que

$$\mathbb{P}[X \in [1,2]] = \int_1^2 \frac{3}{8} t^2 dt = \frac{3}{8}\left[\frac{t^3}{3}\right]_1^2 = \frac{7}{8}$$

Soit Y une variable aléatoire à valeurs dans $[0, b]$ telle que $F(x) = y^2/9$ pour $y \in [0, \beta]$.
Que vaut γ ?

Réponse Comme $F(\beta) = 1$, $\dfrac{\beta^2}{9} = 1$ autrement dit $\beta = 3$ (car $\beta > 0$).

Variable aléatoire continue V

Si X est continue,

$$\mathbb{P}[X = 0] = 0 \text{ et } \mathbb{P}[X = a] = 0, \ \ \forall a$$

aussi

$$\mathbb{P}\big[X \in [a, b]\big] = \mathbb{P}\big[X \in (a, b)\big]$$

Variable aléatoire uniforme I

Loi uniforme (sur $[0,1]$)

X suit une loi uniforme sur $[0,1]$ si $f(x) = \mathbf{1}_{[0,1]}(x)$, $x \in \mathbb{R}$.

On note $X \sim \mathcal{U}([0,1])$. La fonction de répartition est

$$F(x) = \mathbb{P}[X \leq x] = \begin{cases} 0 \text{ si } x \leq 0 \\ x \text{ si } x \in [0,1] \\ 1 \text{ si } x \geq 1 \end{cases}$$

Loi uniforme

Si X suit une loi uniforme sur $[0,1]$, pour tout $h \in (0,1)$,

$$\mathbb{P}\big(X \in [x, x+h]\big) = h, \text{ pour tout } 0 \leq x \leq 1-h.$$

Variable aléatoire uniforme II

Loi uniforme (sur $[a, b]$)

X suit une loi uniforme sur $[a, b]$ si $f(x) = \dfrac{1}{b-a}\mathbf{1}_{[a,b]}(x)$.

Transformation affine d'une variable uniforme

Si X suit une loi uniforme sur $[0, 1]$, pour tout a, b, $a > 0$, $Y = aX + b$ suit une loi uniforme sur $[b, b + a]$.

Preuve soit $y \in \mathbb{R}$,

$$F(y) = \mathbb{P}[Y \leq y] = \mathbb{P}[aX + b \leq y] = \mathbb{P}\left[X \leq \frac{y-b}{a}\right]$$

Variable aléatoire uniforme III

Si $y \leq b$, $\dfrac{y-b}{a} \leq 0$, $F(y) = 0$; si $y \geq b+a$, $\dfrac{y-b}{a} \geq 1$, $F(y) = 1$.

Sinon $F(y) = \mathbb{P}\left[X \leq \dfrac{y-b}{a}\right] = \dfrac{y-b}{a}$, $f(y) = F'(y) = \dfrac{1}{a}\mathbf{1}_{[b,b+a]}(y)$

Variable aléatoire exponentielle I

Loi exponentielle de paramètre λ (sur $\mathbb{R}_+$)

Pour $\lambda > 0$, $X \sim \mathcal{E}(\lambda)$ si

$$f(x) = \lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x) \text{ ou } F(x) = (1 - e^{-\lambda x})\mathbf{1}_{\mathbb{R}_+}(x)$$

Soit X une variable aléatoire $\mathcal{E}(0.1)$. Que vaut $\mathbb{P}[X \in [3,7]]$?

Réponse première approche

$$\mathbb{P}\big[X \in [3,7]\big] = \int_3^7 0.1e^{-0.1\times x}dx = \left[e^{-0.1\times x}\right]_3^7 = e^{-0.3} - e^{-0.7} \approx 24.4\%$$

Seconde approche

$$\mathbb{P}\big[X \in [3,7]\big] = F(7) - F(3) = [1 - e^{-0.7}] - [1 - e^{-0.3}] \approx 24.4\%$$

Variable aléatoire exponentielle II

Loi expontielle

Si X suit une loi exponentielle $X \sim \mathcal{E}(\lambda)$, si $h > 0$, $Y = hX$ suit une loi exponentielle $Y \sim \mathcal{E}(\lambda)$.

Preuve soit $y \geq 0$,

$$\mathbb{P}[Y \leq y] = \mathbb{P}[hX \leq y] = \mathbb{P}\left[X \leq \frac{y}{h}\right] = 1 - e^{-\lambda \frac{y}{h}} = 1 - e^{-\frac{\lambda}{h} y}$$

ce qui correspond à une loi exponentielle de paramètre λ / h.

Loi exponentielle

Si X suit une loi exponentielle $X \sim \mathcal{E}(1)$, $Y = e^{-X}$ suit une loi uniforme sur $[0, 1]$.

Variable aléatoire exponentielle III

Preuve soit $y \geq 0$,

$$\mathbb{P}[Y \leq y] = \mathbb{P}[e^{-X} \leq y] = \mathbb{P}[-X \leq \log(y)] = \mathbb{P}[X > -\log(y)] = e^{\log(y)}$$

ce qui correspond à une loi uniforme.

Loi exponentielle et loi géométrique

Si X suit une loi exponentielle $X \sim \mathcal{E}(\lambda)$, $Y = \lceil X \rceil$ suit une loi géométrique de paramètre $p = 1 - e^{-\lambda}$.

Preuve $n = \lceil x \rceil$ signifie $n - 1 < x \leq n$, donc

$$\mathbb{P}[Y = y] = \mathbb{P}\big(X \in [n-1, n]\big) = F(n) - F(n-1) = e^{\lambda(n-1)} - e^{\lambda n}$$

$$\mathbb{P}[Y = y] = \mathbb{P}[X \in [n-1, n]] = [e^{-\lambda}]^{n-1}\Big(1 - e^{-\lambda}\Big)$$

Variable aléatoire de Pareto I

Loi de Pareto

X suit une loi de Pareto,

$$\mathbb{P}[X \leq x] = 1 - x^{-\alpha}, \text{ pour } x \geq 1$$

pour $\alpha \in \mathbb{R}_+$.

Loi de Pareto

Si $X \sim \mathcal{P}(\alpha)$, $Y = \log X$ suit une loi exponentielle $\mathcal{E}(\alpha)$.

Preuve soit $y \geq 0$,

$$\mathbb{P}[Y \leq y] = \mathbb{P}[\log X \leq y] = \mathbb{P}\left[X \leq e^y\right] = 1-(e^y)^{-\alpha} = 1-\exp(-\alpha y)$$

ce qui correspond à une loi exponentielle.

Variable aléatoire gaussienne I

Loi normale, centrée réduite (sur $\mathbb{R}$)

$X \sim \mathcal{N}(0,1)$ admet pour densité, pour $x \in \mathbb{R}$,

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

Loi normale (sur $\mathbb{R}$)

$X \sim \mathcal{N}(\mu, \sigma^2)$ admet pour densité

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

Variable aléatoire gaussienne II

Si $X \sim \mathcal{N}(0,1)$

$$\Phi(x) = \mathbb{P}[X \leq x]$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt$$

$$\mathbb{P}[X \leq 1.645] = 95\%$$

$$\mathbb{P}[X \leq 1.96] = 97.5\%$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Variable aléatoire gaussienne III

Si $X \sim \mathcal{N}(0,1)$

$$\Phi(x) = \mathbb{P}[X \leq x]$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt$$

$$\mathbb{P}[X \leq 1.645] = 95\%$$

$$\mathbb{P}[X \leq 1.96] = 97.5\%$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Variable aléatoire gaussienne IV

Si $X \sim \mathcal{N}(0,1)$

$$\Phi(x) = \mathbb{P}[X \leq x]$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt$$

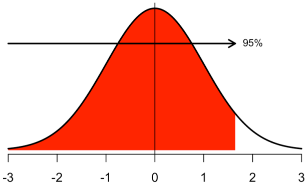
$$\mathbb{P}[X \leq 1.645] = 95\%$$

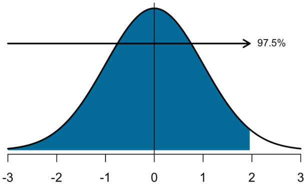
$$\mathbb{P}[X \leq 1.96] = 97.5\%$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

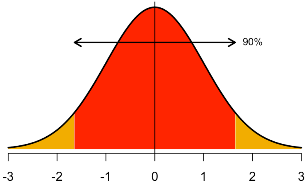
Variable aléatoire gaussienne V

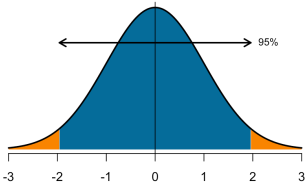
$$\mathbb{P}[X \leq 1.645] = 95\% \text{ et } \mathbb{P}[X \leq 1.96] = 97.5\%$$





$$\mathbb{P}[-1.645 \leq X \leq 1.645] = 90\% \text{ et } \mathbb{P}[-1.96 \leq X \leq 1.96] = 95\%$$

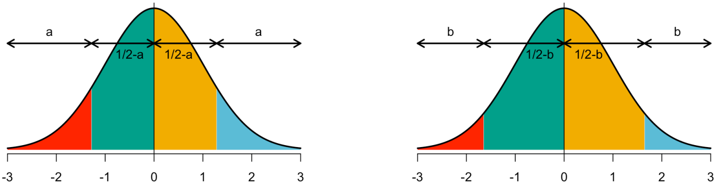




Variable aléatoire gaussienne VI

La loi normale $\mathcal{N}(0,1)$ est symétrique par rapport à 0

$$\mathbb{P}[X < -q] = \mathbb{P}[X > q]$$



Loi normale

Si $X \sim \mathcal{N}(\mu, \sigma^2)$, $Z = \dfrac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

Variable aléatoire gaussienne VII

Si $X \sim \mathcal{N}(75, 8^2)$, que vaut $\mathbb{P}[X \leq 79]$?

$$\mathbb{P}[X \leq 79] = \mathbb{P}\left[Z = \frac{X - 75}{8} \leq \frac{79 - 75}{8} = \frac{4}{8}\right] = \Phi(0.5)$$

```
> pnorm(.5)
[1] 0.6914625
> pnorm(79,75,8)
[1] 0.6914625
```

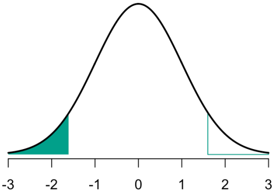
	$\boxed{0.00}$	0.01
0.0	0.5000	0.5040
0.1	0.5398	0.5438
0.2	0.5793	0.5832
0.3	0.6179	0.6217
0.4	0.6554	0.6591
$\boxed{0.5}$	$\boxed{0.6915}$	0.6950
0.6	0.7257	0.7291
0.7	0.7580	0.7611
0.8	0.7881	0.7910
0.9	0.8159	0.8186

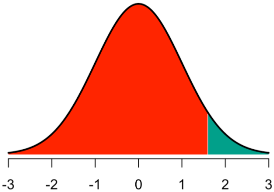
Variable aléatoire gaussienne VIII

Si $X \sim \mathcal{N}(11, 5^2)$, que vaut $\mathbb{P}[X \leq 3]$?

$$\mathbb{P}[X \leq 3] = \mathbb{P}\left[Z = \frac{X - 11}{5} \leq \frac{3 - 11}{5} = \frac{-8}{5}\right] = \Phi\,(-1.6) = 1 - \Phi\,(1.6)$$

```
> 1-pnorm(1.6)
[1] 0.05479929
> pnorm(-1.6)
[1] 0.05479929
> pnorm(3,11,5)
[1] 0.05479929
```





	0.00	0.01
0.0	0.5000	0.5040
0.1	0.5398	0.5438
0.2	0.5793	0.5832
0.3	0.6179	0.6217
0.4	0.6554	0.6591
0.5	0.6915	0.6950
0.6	0.7257	0.7291
0.7	0.7580	0.7611
0.8	0.7881	0.7910
0.9	0.8159	0.8186
1.0	0.8413	0.8438
1.1	0.8643	0.8665
1.2	0.8849	0.8869
1.3	0.9032	0.9049
1.4	0.9192	0.9207
1.5	0.9332	0.9345
1.6	0.9452	0.9463
1.7	0.9554	0.9564
1.8	0.9641	0.9649

Variable aléatoire gaussienne IX

Si $X \sim \mathcal{N}(61, 20^2)$, que vaut $\mathbb{P}[X > 20]$?

$$\mathbb{P}[X > 20] = \mathbb{P}\left[Z = \frac{X-61}{20} > \frac{20-61}{20} = \frac{-41}{20}\right] = 1 - \Phi(-2.05)$$

soit $\Phi(2.05)$.

```
> pnorm(2.05)
[1] 0.9798178
> 1-pnorm(20,61,20)
[1] 0.9798178
```

	0.00	0.05
0.0	0.5000	0.5199
0.1	0.5398	0.5596
0.2	0.5793	0.5987
0.3	0.6179	0.6368
0.4	0.6554	0.6736
0.5	0.6915	0.7088
0.6	0.7257	0.7422
0.7	0.7580	0.7734
0.8	0.7881	0.8023
0.9	0.8159	0.8289
1.0	0.8413	0.8531
1.1	0.8643	0.8749
1.2	0.8849	0.8944
1.3	0.9032	0.9115
1.4	0.9192	0.9265
1.5	0.9332	0.9394
1.6	0.9452	0.9505
1.7	0.9554	0.9599
1.8	0.9641	0.9678
1.9	0.9713	0.9744
2.0	0.9772	0.9798
2.1	0.9821	0.9842
2.2	0.9861	0.9878

Variable aléatoire gaussienne X

Si $X \sim \mathcal{N}(64, 5^2)$, que vaut $\mathbb{P}[X \in [63, 72]]$?

$$\mathbb{P}[X \leq 72] = \mathbb{P}\left[Z = \frac{X - 64}{5} \leq \frac{72 - 64}{5} = \frac{8}{5}\right] = \Phi(1.6)$$

$$\mathbb{P}[X \geq 63] = \mathbb{P}\left[Z = \frac{X - 64}{5} \geq \frac{63 - 64}{5} = \frac{-1}{5}\right] = \Phi(-.02)$$

```
> pnorm(-.2)
[1] 0.4207403
> pnorm(63,64,5)
[1] 0.4207403
> pnorm(1.6)
[1] 0.9452007
> pnorm(72,64,5)
[1] 0.9452007
> pnorm(1.6)-pnorm(-.2)
[1] 0.5244604
```

Variable aléatoire gaussienne XI

Si $X \sim \mathcal{N}(64, 5^2)$, que vaut $\mathbb{P}[X \in [63, 72]]$?

```
> pnorm(1.6)-pnorm(-.2)
[1] 0.5244604
```

Notons que

$$\mathbb{P}[X \in [63, 72]] = \int_{63}^{72} f(x)dx = \int_{-0.2}^{+1.6} \varphi(z)dz$$

```
> integrate(dnorm,-.2,1.6)
0.5244604 with absolute error < 5.8e-15
```

Variable aléatoire gaussienne XII

Si $X \sim \mathcal{N}(10, 1.5^2)$, que serait q tel que $\mathbb{P}[X \leq q] = 0.25$?

$$\mathbb{P}[X \leq q] = \mathbb{P}\left[Z = \frac{X-10}{1.5} \leq \frac{q-10}{1.5}\right] = 0.25$$

Or $\Phi(z) = 0.25$ signifie que que $z \approx -0.6745$

```
> qnorm(.25)
[1] -0.6744898
> pnorm(-0.6744898)
[1] 0.25
```

donc $\dfrac{q-10}{1.5} = -0.6745$, soit $q = 10 - -0.6745 \times 1.5 = 8.988$

```
> qnorm(.25,10,1.5)
[1] 8.988265
> pnorm(8.988265,10,1.5)
[1] 0.249999
```

Variable aléatoire gaussienne XIII

La taille d'un enfant de 13 ans suit une loi normale, de moyenne 156 cm, et d'écart-type 4.2 cm. Quelle est la probabilité qu'un garçon choisit au hasard mesure entre 150 cm et 160 cm ?

Autrement dit $X \sim \mathcal{N}(156, 4.2^2)$, et on veut $\mathbb{P}[X \in [150, 160]]$,

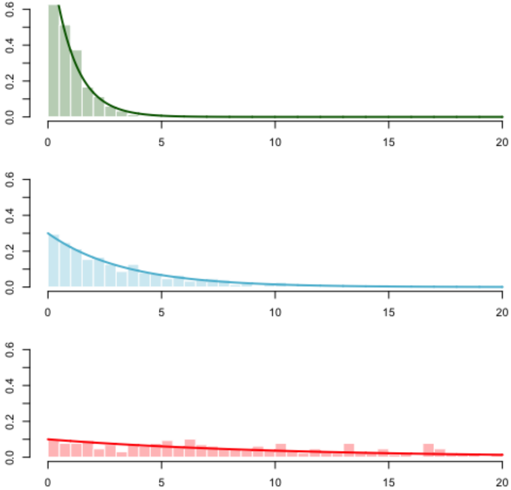
$$\mathbb{P}[150 \leq X \leq 160] = \mathbb{P}\left[\frac{150-156}{4.2} \leq \underbrace{\frac{X-156}{4.2}}_{Z \sim \mathcal{N}(0,1)} \leq \frac{160-156}{4.2}\right]$$

i.e. $\mathbb{P}[-1.43 \leq Z \leq 0.95]$ qui vaut environ 3/4,

```
> pnorm((160-156)/4.2)
[1] 0.8295481
> pnorm((150-156)/4.2)
[1] 0.07656373
> pnorm((160-156)/4.2) - pnorm((150-156)/4.2)
[1] 0.7529844
```

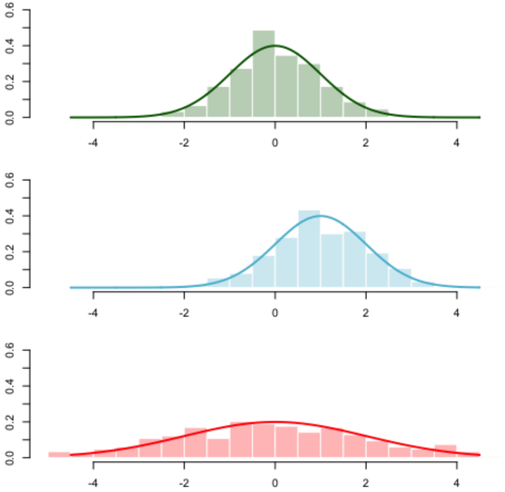
Variables aléatoires continues I

0.337	1.032	1.648
0.655	2.195	13.24
1.876	2.141	18.912
1.035	11.205	0.033
1.055	8.614	2.287
4.424	0.884	16.902
1.238	0.025	1.058
0.762	1.551	0.26
1.391	0.413	11.219
0.147	7.808	19.567
0.957	6.545	2.917
0.54	0.76	9.263
1.23	4.282	1.831
2.895	0.812	4.102
0.436	5.918	10.66
0.14	3.469	7.839
0.146	7.031	0.379
1.182	12.194	5.28
0.755	1.401	17.2



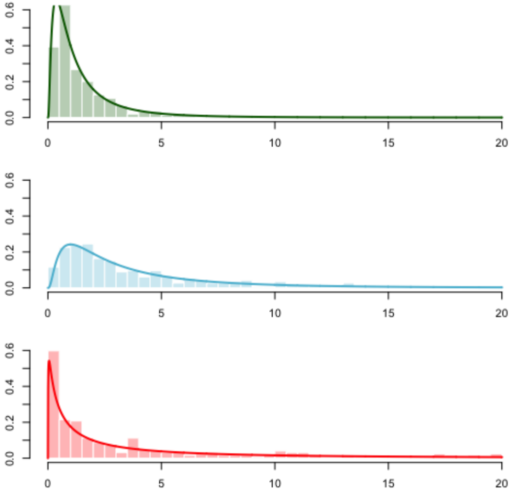
Variables aléatoires continues II

0.821	-0.294	-0.496
0.944	0.879	-3.585
-0.016	0.516	-0.597
-0.045	2.069	-0.847
1.125	0.964	5.351
-2.215	0.344	-2.798
-0.621	0.546	-0.77
0.39	1.099	0.665
1.512	2.044	0.329
-0.305	1.323	0.443
0.576	-0.025	4.003
0.738	1.567	-0.447
0.487	1.083	-2.993
-0.82	2.512	-2.273
0.33	2.654	-0.273
1.595	0.616	1.084
-0.836	2.971	1.057
0.184	-0.047	3.005
-0.626	1.894	-0.682



Variables aléatoires continues III

2.273	0.745	1.655
2.57	2.408	0.075
0.984	1.675	1.496
0.956	7.918	1.165
3.08	2.622	573.342
0.109	1.411	0.166
0.537	1.726	1.258
1.477	3.001	5.287
4.535	7.718	3.776
0.737	3.755	4.235
1.779	0.976	148.924
2.092	4.793	1.739
1.628	2.953	0.136
0.44	12.332	0.28
1.39	14.213	2.068
4.93	1.852	8.04
0.434	19.518	7.819
1.202	0.954	54.864
0.534	6.644	1.374



Variables aléatoires continues IV

Somme de variables gaussiennes, $X \perp\!\!\!\perp Y$

Si $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ et $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, avec $X \perp\!\!\!\perp Y$

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

propriété admise

Moyenne de variables gaussiennes, $X_i \perp\!\!\!\perp X_j$

Si $X_1, \cdots, X_n \sim \mathcal{N}(\mu, \sigma_X)$ sont indépendantes,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

propriété admise

Espérance et variance I

Espérance

Si X a pour densité f,

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

(si l'intégrale est convergente).

Soit X une variable aléatoire $\mathcal{U}([0,1])$. Que vaut $\mathbb{E}[X]$?

Réponse

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\mathbf{1}_{[0,1]}(x)dx = \int_0^1 xdx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1-0}{2} = \frac{1}{2}$$

Espérance et variance II

Soit X une variable aléatoire $\mathcal{E}(\lambda)$. Que vaut $\mathbb{E}[X]$?

Réponse

$$\mathbb{E}[X] = \int_0^\infty x\lambda e^{-\lambda x} dx = \left[\lambda x \cdot \frac{-1}{\lambda} e^{-\lambda x}\right]_0^\infty - \int_0^\infty \lambda \cdot \frac{-1}{\lambda} e^{-\lambda x} dx$$

$$\mathbb{E}[X] = \int_0^\infty e^{-\lambda x} dx = \left[\frac{-1}{\lambda} e^{-\lambda x}\right]_0^\infty = \frac{1}{\lambda}$$

Espérance et variance III

Variance

Si X a pour densité f,

$$\text{Var}[X] = \int_{-\infty}^{+\infty} \left(x - \mathbb{E}[X]\right)^2 f(x)dx$$

(si l'intégrale est convergente).

Soit X une variable aléatoire $\mathcal{U}([0,1])$. Que vaut $\text{Var}[X]$?

Réponse on a vu que $\mathbb{E}[X] = \dfrac{1}{2}$,

$$\text{Var}[X] = \int_0^1 \left(x - \frac{1}{2}\right)^2 dx = \left[\frac{1}{3}\left(x - \frac{1}{2}\right)^3\right]_0^1 = \frac{2}{3 \cdot 2^3} = \frac{1}{12}$$

Espérance et variance IV

Variable gaussienne

Si $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mathbb{E}[X] = \mu$ et $\text{Var}[X] = \sigma^2$.

propriété admise

Variable centrée et réduite

Si X a pour moyenne μ et pour variance σ^2,

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

est la version centrée et réduite de X.
Z est centrée car $\mathbb{E}[Z] = 0$ et réduite car $\text{Var}[Z] = 1$.

Approximation pour la loi binomiale ★★★

Théorème de de Moivre–Laplace

$$\binom{n}{k}p^k(1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi np(1-p)}}\, e^{-\frac{(k-np)^2}{2np(1-p)}}$$

Preuve Formule de Stirling $n! \simeq n^n e^{-n}\sqrt{2\pi n}$ lorsque $n \to \infty$

$$\begin{aligned}
\binom{n}{k}p^k q^{n-k} &= \frac{n!}{k!(n-k)!}p^k q^{n-k}, \text{ où } q = 1-p, \\
&\simeq \frac{n^n e^{-n}\sqrt{2\pi n}}{k^k e^{-k}\sqrt{2\pi k}(n-k)^{n-k}e^{-(n-k)}\sqrt{2\pi(n-k)}}p^k q^{n-k} \\
&= \sqrt{\frac{n}{2\pi k\,(n-k)}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}
\end{aligned}$$

$$\binom{n}{k}p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}$$

Approximation pour la loi binomiale ★★★

$$\binom{n}{k}p^kq^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}}\exp\left\{\log\left(\left(\frac{np}{k}\right)^k\right)+\log\left(\left(\frac{nq}{n-k}\right)^{n-k}\right)\right\}$$

$$= \frac{1}{\sqrt{2\pi npq}}\exp\left\{-k\log\left(\frac{k}{np}\right)+(k-n)\log\left(\frac{n-k}{nq}\right)\right\}$$

On pose $x = \dfrac{(k-np)}{\sqrt{npq}}$

$$= \frac{1}{\sqrt{2\pi npq}}\exp\left\{-k\log\left(\frac{np+x\sqrt{npq}}{np}\right)+(k-n)\log\left(\frac{n-np-x\sqrt{npq}}{nq}\right)\right.$$

$$= \frac{1}{\sqrt{2\pi npq}}\exp\left\{-k\log\left(1+x\sqrt{\frac{q}{np}}\right)+(k-n)\log\left(1-x\sqrt{\frac{p}{nq}}\right)\right\}$$

$$= \frac{1}{\sqrt{2\pi npq}}\exp\left\{-k\left(x\sqrt{\frac{q}{np}}-\frac{x^2q}{2np}+\cdots\right)+(k-n)\left(-x\sqrt{\frac{p}{nq}}-\frac{x^2p}{2nq}-\right.\right.$$

car $\log(1+x) \simeq x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \cdots$

Approximation pour la loi binomiale ★★★

$$= \frac{1}{\sqrt{2\pi npq}} \exp\left\{-k\left(x\sqrt{\frac{q}{np}} - \frac{x^2 q}{2np} + \cdots\right) + (k-n)\left(-x\sqrt{\frac{p}{nq}} - \frac{x^2 p}{2nq} -\right.\right.$$

$$= \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{1}{2}x^2 q - \frac{1}{2}x^2 p - \cdots\right\} = \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{1}{2}x^2\right\}$$

i.e.

$$\binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{(k-np)^2}{2npq}\right\}$$

ou, en posant $\mu = np$ et $\sigma^2 = npq$

$$\binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(k-\mu)^2}{2\sigma^2}\right\}$$

Approximation pour la loi binomiale

Example 3 On lance 10 pièces (bien équilibrées). Quelle est la probabilité d'obtenir entre 3 et 6 fois '*face*'

- valeur exacte (loi binomiale)

Soit X le nombre de '*face*' obtenus, $X \sim \mathcal{B}(10, 1/2)$,

$$\mathbb{P}[X=3]=\binom{10}{3}\frac{1}{2^3}\frac{1}{2^7}=\frac{15}{128},\ \mathbb{P}[X=4]=\binom{10}{4}\frac{1}{2^4}\frac{1}{2^6}=\frac{105}{512}$$

$$\mathbb{P}[X=5]=\binom{10}{5}\frac{1}{2^5}\frac{1}{2^5}=\frac{63}{256},\ \mathbb{P}[X=6]=\binom{10}{6}\frac{1}{2^6}\frac{1}{2^4}=\frac{105}{512}$$

donc

$$\mathbb{P}[X\in\{3,4,5,6\}]=\frac{15}{128}+\frac{105}{512}+\frac{63}{256}+\frac{105}{512}=\frac{99}{128}\sim 0.7734$$

```
> sum(dbinom(3:6,size = 10,prob = 1/2))
[1] 0.7734375
```

Approximation pour la loi binomiale

Example 3 On lance 10 pièces (bien équilibrées). Quelle est la probabilité d'obtenir entre 3 et 6 fois '*face*'

- valeur approchée (loi normale) - version 1

$\mu = np = 10/2 = 5$ et $\sigma = \sqrt{np(1-p)} = \sqrt{10/4} = \sqrt{5}/2 \sim 1.58$

```
> pnorm(6,5,sqrt(5)/2)-pnorm(3,5,sqrt(5)/2)
[1] 0.6335038
```

On peut centrer et réduire : $Y \sim \mathcal{N}(\mu, \sigma^2)$, $X = \dfrac{Y-\mu}{\sigma}$

$$\mathbb{P}[3 < Y \leq 6] = \mathbb{P}\left[\frac{3-5}{1.58} < \underbrace{\frac{Y-5}{1.58}}_{\sim\mathcal{N}(0,1)} \leq \frac{6-5}{1.58}\right] = \mathbb{P}[-1.788 < X \leq 0.894]$$

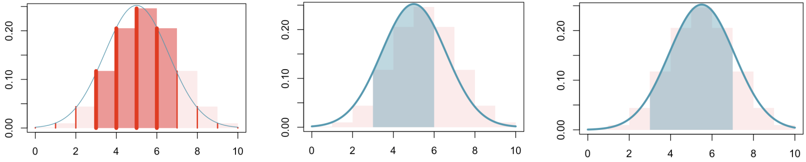
soit $\mathbb{P}[X \leq 0.894] - \mathbb{P}[X \leq -1.788]$

```
> pnorm((6-5)/sqrt(10)*2)-pnorm((3-5)/sqrt(10)*2)
[1] 0.6335038
```

Approximation pour la loi binomiale

Example 3 On lance 10 pièces (bien équilibrées). Quelle est la probabilité d'obtenir entre 3 et 6 fois '*face*'

- valeur approchée (loi normale) - version 2




```
> pnorm((6+.5-5)/sqrt(10/4))-pnorm((3-.5-5)/sqrt(10/4)
[1] 0.771686
```

- décallage de $+1/2$ (Berry-Essen)
- $X \in \{3, 4, 5, 6\}$: intervalle de longueur 4 (et pas 3)

Approximation pour la loi binomiale ★★★

L'inégalité de Berry-Essen fournit une borne de la différence entre les deux fonctions de répartition lorsque n est grand, pour X une variable aléatoire de loi binomiale $\mathcal{B}(n,p)$ et Y de loi normale $\mathcal{N}(0,1)$ de fonction de répartition Φ :

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{X-np}{\sqrt{npq}}\leq x\right)-\Phi(x)\right|\leq\frac{0.4748}{\sqrt{npq}}$$

Autrement dit,

$$\mathbb{P}(X\leq x)=\Phi\left(\frac{x-np+1/2}{\sqrt{npq}}\right)+\frac{\text{quelque chose}}{\sqrt{n}}$$

uniformément pour tout x, lorsque $n\to\infty$.

Approximation pour la loi binomiale I

Example 2 Un hôpital a 12000 patients agés, et on a estimé que la probabilité qu'un patient souffre d'un accident cardiaque pendant une journée était de 1/8000. L'hôpital possède seulement trois machines respiratoires nécessaires pour ces accidents cardiaques et se demande si son équipement sera suffisant pour une journée particulière

On avait utilisé

```
> pbinom(3,12000,1/8000)
[1] 0.9343693133
> ppois(3,1.5)
[1] 0.9343575456
```

On peut aussi considérer

```
> mu = 12000/8000
> s2 = 12000/8000*(1-1/8000)
> pnorm(3.5,mu,sqrt(s2))
[1] 0.9487755
```

Loi des grands nombres et théorème central limite I

Loi des grands nombres

Soient $X_1, X_2, \cdots$ des variables i.i.d. de même moyenne μ et de même variance σ^2. Soit

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + \cdots + X_n}{n}$$

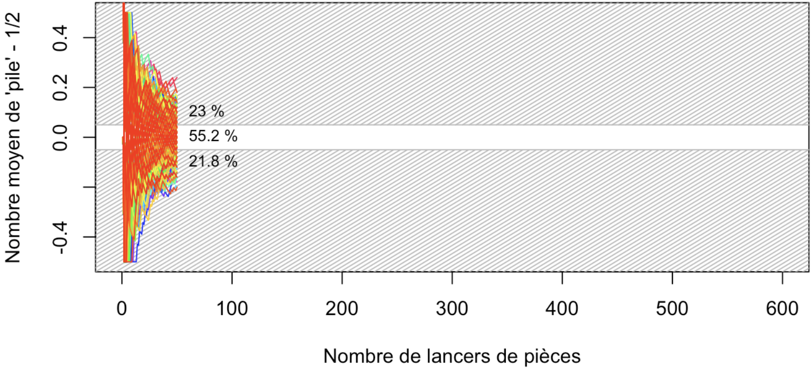
alors pour tout ε (aussi petit que possible),

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) = 1$$

Avec n suffisamment grand, $\overline{X}_n$ sera aussi proche que nous le souhaitons de μ, avec une probabilité proche de 1.

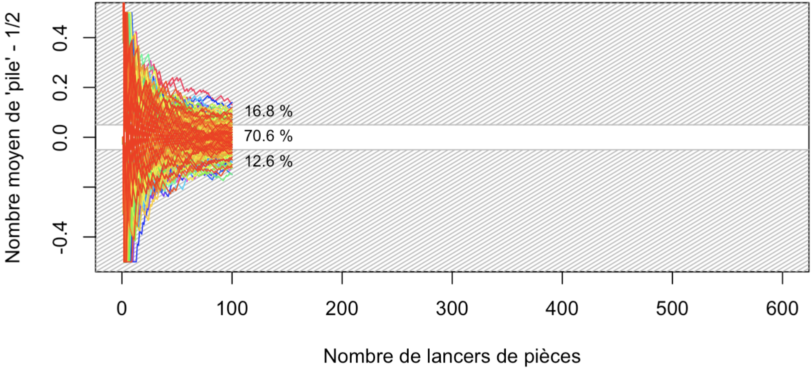
Convergence, loi des grands nombres

$$\mathbb{P}\left(\left|\overline{X}_{50} - \mu\right| > 5\%\right) \sim 45\% \; (n = 50 \text{ lancers})$$



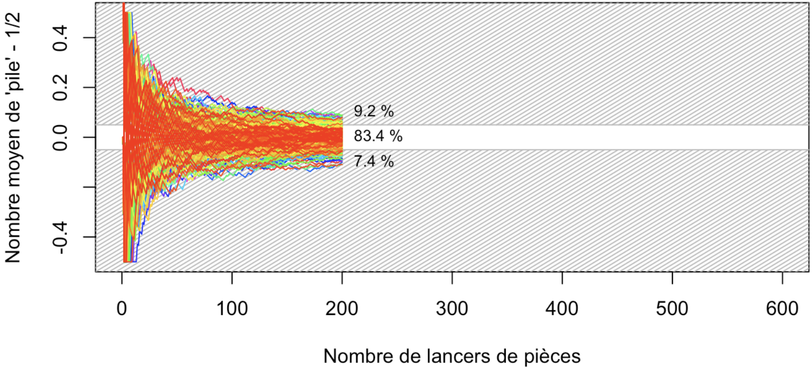
Convergence, loi des grands nombres

$$\mathbb{P}\left(\left|\overline{X}_{100} - \mu\right| > 5\%\right) \sim 30\% \; (n = 100 \text{ lancers})$$



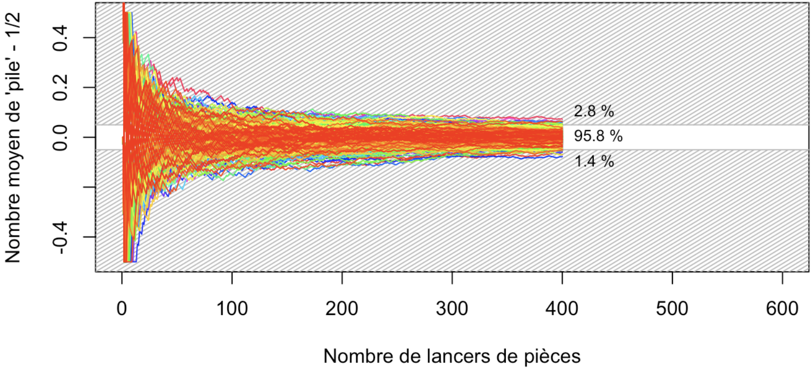
Convergence, loi des grands nombres

$$\mathbb{P}\left(\left|\overline{X}_{200} - \mu\right| > 5\%\right) \sim 17\% \ (n = 200 \text{ lancers})$$



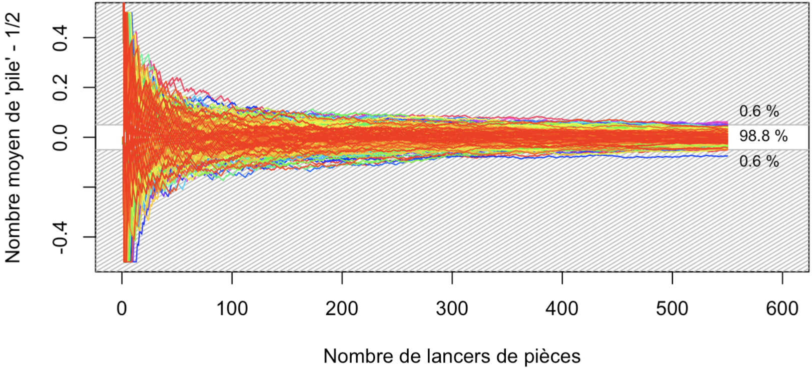
Convergence, loi des grands nombres

$$\mathbb{P}\left(\left|\overline{X}_{400} - \mu\right| > 5\%\right) \sim 6\% \ (n = 400 \text{ lancers})$$



Convergence, loi des grands nombres

$$\mathbb{P}\left(\left|\overline{X}_{500} - \mu\right| > 5\%\right) \sim 1\% \ (n = 500 \text{ lancers})$$



Loi des grands nombres et théorème central limite I

Théorème Central Limite

Soient $X_1, X_2, \cdots$ des variables i.i.d. de même moyenne μ et de même variance σ^2. Soit

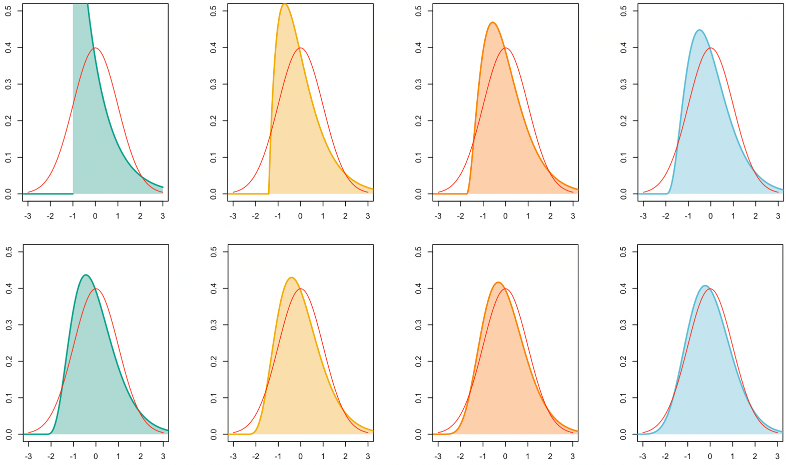
$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + \cdots + X_n}{n}$$

alors

$$\sqrt{n} \cdot \frac{\overline{X}_n - \mu}{\sigma} \approx \mathcal{N}(0,1).$$

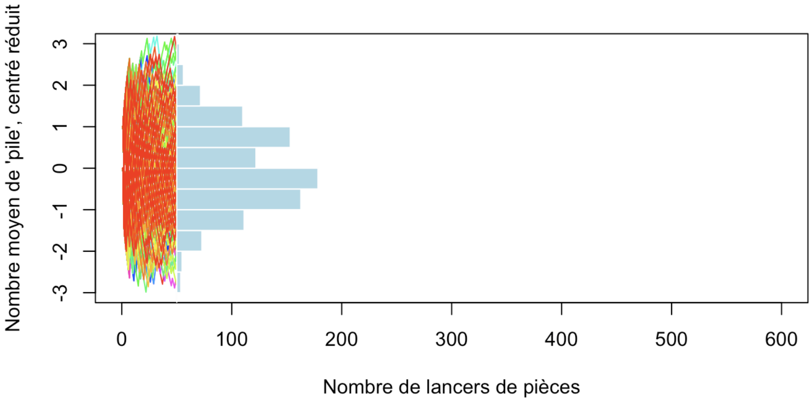
Loi des grands nombres et théorème central limite II

Exemple : moyenne standardisée de variables $\mathcal{E}(1)$ indépendantes ($n = 1, 2, 3, 4, 6, 7, 10, 20$)



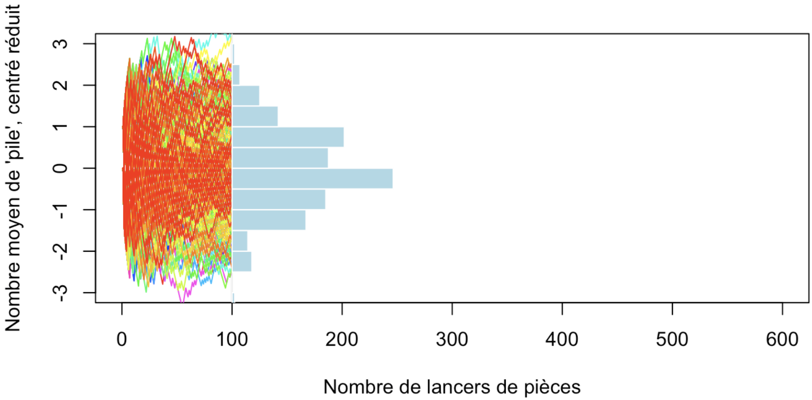
Convergence, théorème central limite

Histogramme de $\sqrt{50} \cdot \dfrac{\overline{X}_{50} - \mu}{\sigma}$ (n = 50 lancers)



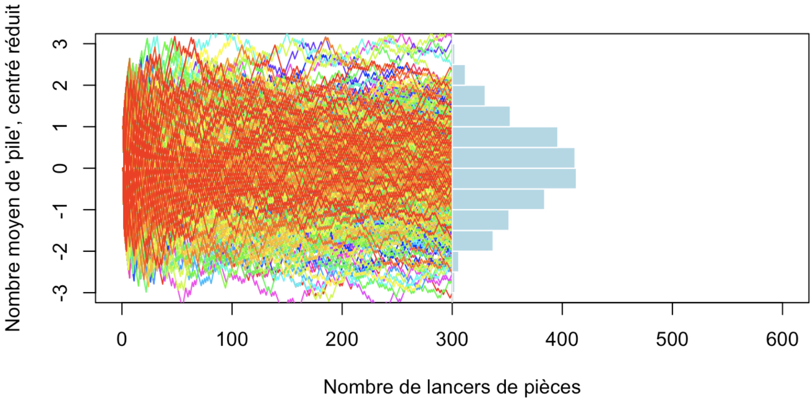
Convergence, théorème central limite

Histogramme de $\sqrt{100} \cdot \dfrac{\overline{X}_{100} - \mu}{\sigma}$ (n = 100 lancers)



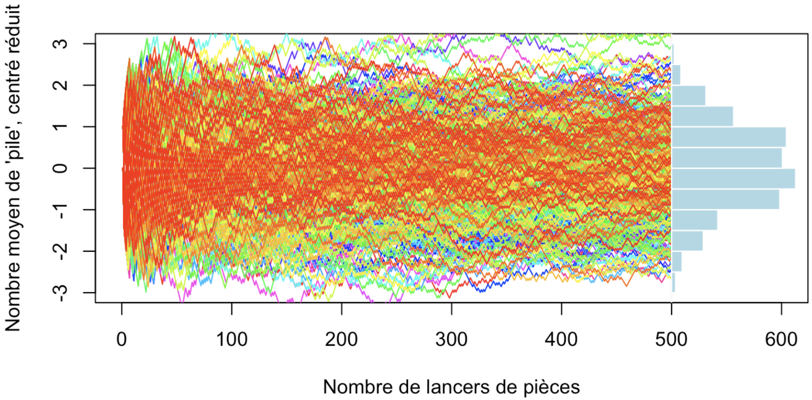
Convergence, théorème central limite

Histogramme de $\sqrt{300} \cdot \dfrac{\overline{X}_{300} - \mu}{\sigma}$ (n = 300 lancers)



Convergence, théorème central limite

Histogramme de $\sqrt{500} \cdot \dfrac{\overline{X}_{500} - \mu}{\sigma}$ ($n = 500$ lancers)

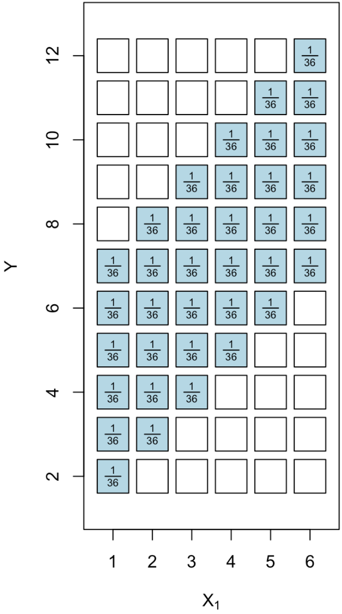


Distribution jointe

On lance 2 dés (6 face)
X_1 la valeur faciale du 1er dé
X_2 la valeur faciale du 2nd dé
Notons $Y = X_1 + X_2$.
$\mathbb{P}(X_1 = x_1, Y = y)$ est la loi jointe

- x_1 $(\in \{1, 2, \cdots, 6\})$
- y $(\in \{2, 3, \cdots, 12\})$

E.g. $\mathbb{P}(X_1 = 4, Y = 3) = 0$



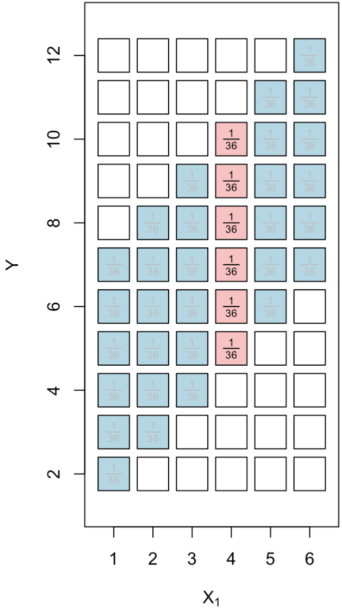
Conditionnement

$$\mathbb{P}(Y = y | X_1 = x_1) \overset{def}{=} \frac{\mathbb{P}(X_1 = x_1, Y = y)}{\mathbb{P}(X_1 = x_1)}$$

E.g. $x_1 = 4$, alors

$$\mathbb{P}(Y = y | X_1 = 4) = \frac{1}{6}, \ y \in \{5, 6, \cdots, 10\}$$

(uniforme sur $\{5, 6, \cdots, 10\}$)



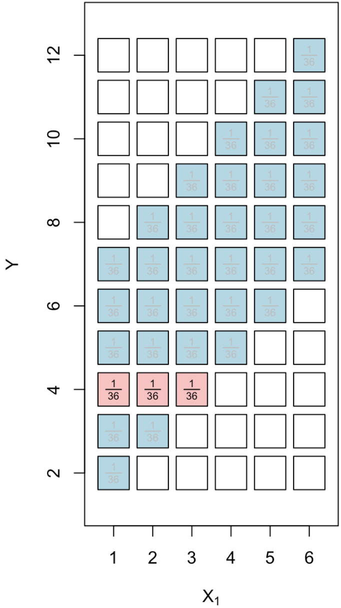
Conditionnement

$$\mathbb{P}(X_1 = x_1 | Y = y) \overset{def}{=} \frac{\mathbb{P}(X_1 = x_1, Y = y)}{\mathbb{P}(Y = y)}$$

E.g. $y = 4$, alors

$$\mathbb{P}(X_1 = x_1 | Y = 4) = \frac{1}{3},\ x_1 \in \{1, 2, 3\}$$

(uniforme sur $\{1, 2, 3\}$)



Couple de variables aléatoires I

La fonction de répartition du couple (X, Y) est

$$F(\boldsymbol{z}) = F(x, y) = \mathbb{P}(X \leq x, Y \leq y), \text{ pour tout } \boldsymbol{z} = (x, y) \in \mathbb{R} \times \mathbb{R}.$$

La densité de $\boldsymbol{Z} = (X, Y)$ est

$$f(\boldsymbol{z}) = f(x, y) = \begin{cases} \dfrac{\partial^2 F(x, y)}{\partial x \partial y} \text{ dans le cas continu, } \boldsymbol{z} = (x, y) \in \mathbb{R}^2 \\ \mathbb{P}(X = x, Y = y) \text{ dans le cas discret, } \boldsymbol{z} = (x, y) \end{cases}$$

Indépendance

X et Y sont indépendants (noté $X \perp\!\!\!\perp Y$) si et seulement si

$$\begin{cases} F(x, y) = F_X(x) \cdot F_Y(y), \ \forall\ x, y \in \mathbb{R} \\ f(x, y) = f_X(x) \cdot f_Y(y), \ \forall\ x, y \in \mathbb{R} \end{cases}$$

conséquence directe de la définition

Couple de variables aléatoires II

$f(x,y) = \mathbf{1}_{[0,1]^2}(x,y)$ désigne la loi uniforme sur $[0,1]\times[0,1]$,

$$f(x,y) = \mathbf{1}_{[0,1]^2}(x,y) = \mathbf{1}_{[0,1]}(x)\times\mathbf{1}_{[0,1]}(y) = f_X(x)\times f_Y(y)$$

alors $X \perp\!\!\!\perp Y$.

$\mathbb{E}[h(X,Y)]$

Soient X et Y des variables aléatoires,

$$\mathbb{E}[h(X,Y)] = \int_{\mathbb{R}}\int_{\mathbb{R}} h(x,y)\times f(x,y)dxdy$$

Couple de variables aléatoires III

Covariance

On appelle covariance d'un couple de variable aléatoire

$$\text{Cov}(X,Y) = \mathbb{E}\big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big] = \mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]$$

(à condition que $\mathbb{E}[X^2] < \infty$ et $\mathbb{E}[Y^2] < \infty$).

Covariance

Soient X, X_1, X_2 et Y des variables aléatoires,

- $\text{Cov}(X,Y) = \text{Cov}(Y,X)$,
- $\text{Cov}(X,X) = \text{Var}(X)$,
- $\text{Cov}(aX+b, cY+d) = ac\text{Cov}(X,Y)$,
- $\text{Cov}(X_1+X_2, Y) = \text{Cov}(X_1,Y) + \text{Cov}(X_2,Y)$,

Couple de variables aléatoires IV

Covariance

Soient X et Y des variables aléatoires,

$$X \perp\!\!\!\perp Y \Longrightarrow \text{Cov}(X,Y) = 0$$

Preuve Si $X \perp\!\!\!\perp Y$, $f(x,y) = f_X(x)f_Y(y)$, et

$$\mathbb{E}[XY] = \int_{\mathbb{R}}\int_{\mathbb{R}} xyf_X(x)f_Y(y)dxdy = \underbrace{\int_{\mathbb{R}} xf_X(x)dx}_{\mathbb{E}[X]} \cdot \underbrace{\int_{\mathbb{R}} yf_Y(y)dy}_{\mathbb{E}[Y]}$$

Covariance

Soient X et Y des variables aléatoires,

$$\text{Cov}(X,Y) = 0 \not\Longrightarrow X \perp\!\!\!\perp Y$$

Couple de variables aléatoires V

Example Soient X, Z trois variables $\mathcal{N}(0,1)$ indépendantes. Si $Y = rX + \sqrt{1-r^2}Z$,

$$\text{Cov}(X, Y^2) = 0, \;\; \text{Cov}(X, Y) = r$$

et X et Y^2 ne sont pas indépendantes
Example Soit (X, Y) uniforme sur $[0,1]\times[0,1]$,

$$f(x,y) = \mathbf{1}_{[0,1]^2}(x,y) = \mathbf{1}_{[0,1]}(x) \times \mathbf{1}_{[0,1]}(y) = f_X(x) \times f_Y(y)$$

alors $X \perp\!\!\!\perp Y$, et

$$\mathbb{E}[XY] = \int_0^1 \left(\int_0^1 xdx\right) ydy = \frac{1}{2}\int_0^1 ydy = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{2^2}$$

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{1}{2^2} - \frac{1}{2}\cdot\frac{1}{2} = 0.$$

Couple de variables aléatoires VI

Soient $X_1, \cdots, X_{12}$ des indicatrices d'obtention de 'face' en lançant 12 fois une pièce. Soient $Y_1 = X_1 + \cdots + X_7$ et $Y_2 = X_6 + \cdots + X_{12}$.

Que vaut $\text{Cov}(Y_1, Y_2)$?

Comme $X_i = \mathbf{1}(\text{'face' au } i\text{-ème lancer})$, $X_i \sim \mathcal{B}(1/2)$ donc $\mathbb{E}[X_i] = 1/2$ et $\text{Var}[X_i] = 1/4$.

$$\text{Cov}(Y_1, Y_2) = \sum_{i=1}^{7} \sum_{i=6}^{12} \text{Cov}(X_i, X_j)$$

avec pour $i \neq j$, $\text{Cov}(X_i, X_j) = 0$, aussi

$$\text{Cov}(Y_1, Y_2) = \text{Cov}(X_6, X_6) + \text{Cov}(X_7, X_7) = \text{Var}(X_6) + \text{Var}(X_7) = \frac{1}{2}.$$

Couple de variables aléatoires VII

Corrélation

On appelle corrélation d'un couple de variable aléatoire

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}$$

(à condition que $\mathbb{E}[X^2] < \infty$ et $\mathbb{E}[Y^2] < \infty$).

Corrélation $\in [-1, +1]$

Soient X et Y des variables aléatoires,

$$\text{Cor}(X, Y) \in [-1, +1].$$

propriété admise

Couple de variables aléatoires VIII

Corrélation ± 1

Soient X et Y des variables aléatoires,

$$\text{Cor}(X,Y)=\pm 1 \Longleftrightarrow Y=aX+b$$

propriété admise

Non linéarité de la variance

For all X and Y, $\forall a,b\in\mathbb{R}$.

$$\text{Var}(aX+bY)=a^2\text{Var}(X)+b^2\text{Var}[Y]+2ab\text{Cov}(X,Y).$$

Note $\text{Cor}(X,Y)$ est parfois appelée 'corrélation de Pearson'