Haute école de gestion de Genève
Geneva School of Business Administration

Distributions continues

Dr Sacha Varone

heg

Objectifs

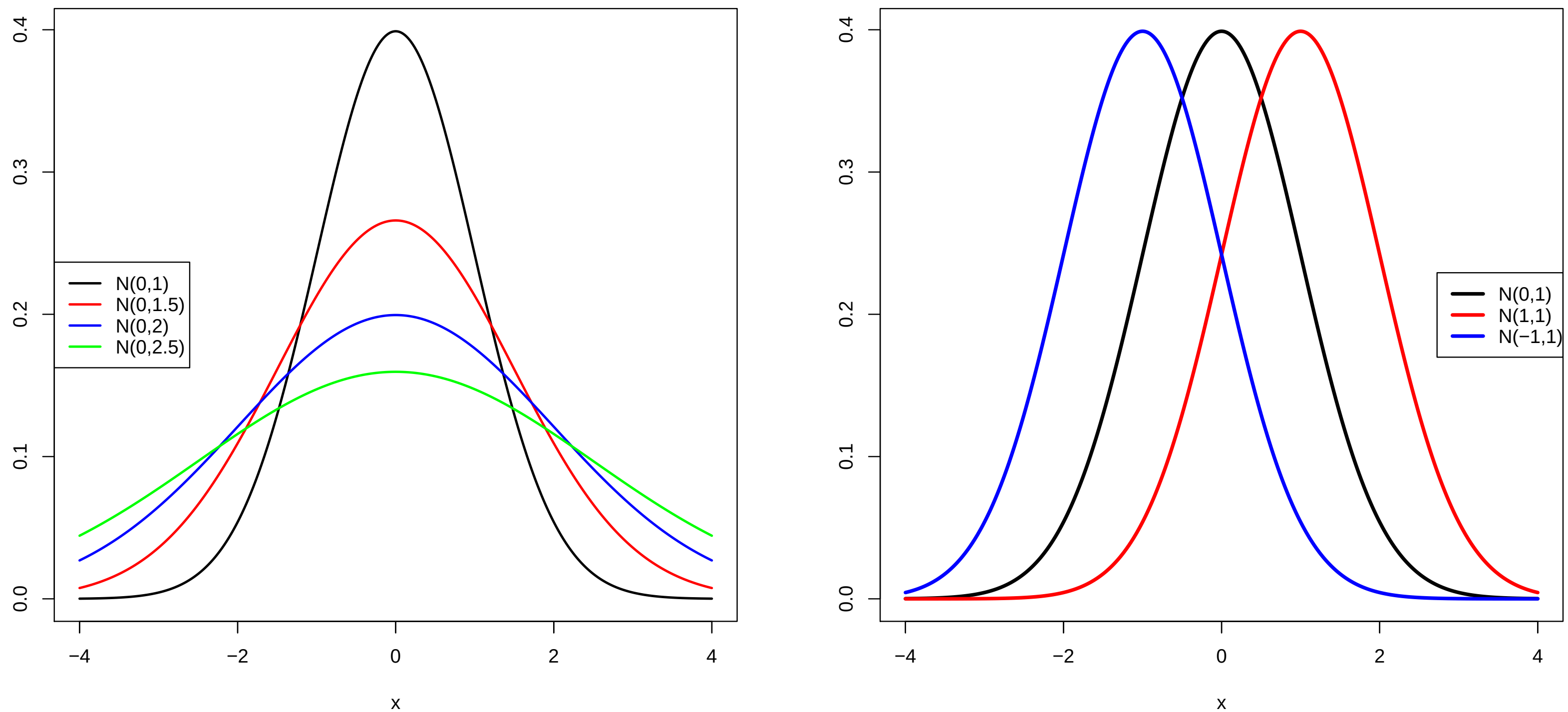
Savoir reconnaître et utiliser

- une loi normale $\mathcal{N}$
- une loi du χ^2
- une loi de Student $\mathcal{T}_n$

Loi normale (Rappel)

h e g

Famille de lois normales



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

σ^2 détermine la largeur de la courbe. Plus sa valeur est élevée, plus la courbe sera large et aplatie. μ détermine la position de la moyenne.

Probabilités

Probabilité $\rightarrow$ aire sous la courbe de densité $f(x)$.
Par symétrie $P(X \leq \mu) = P(X \geq \mu) = 0.5$
Donc $P(X \leq x) = 1 - P(X \geq x)$

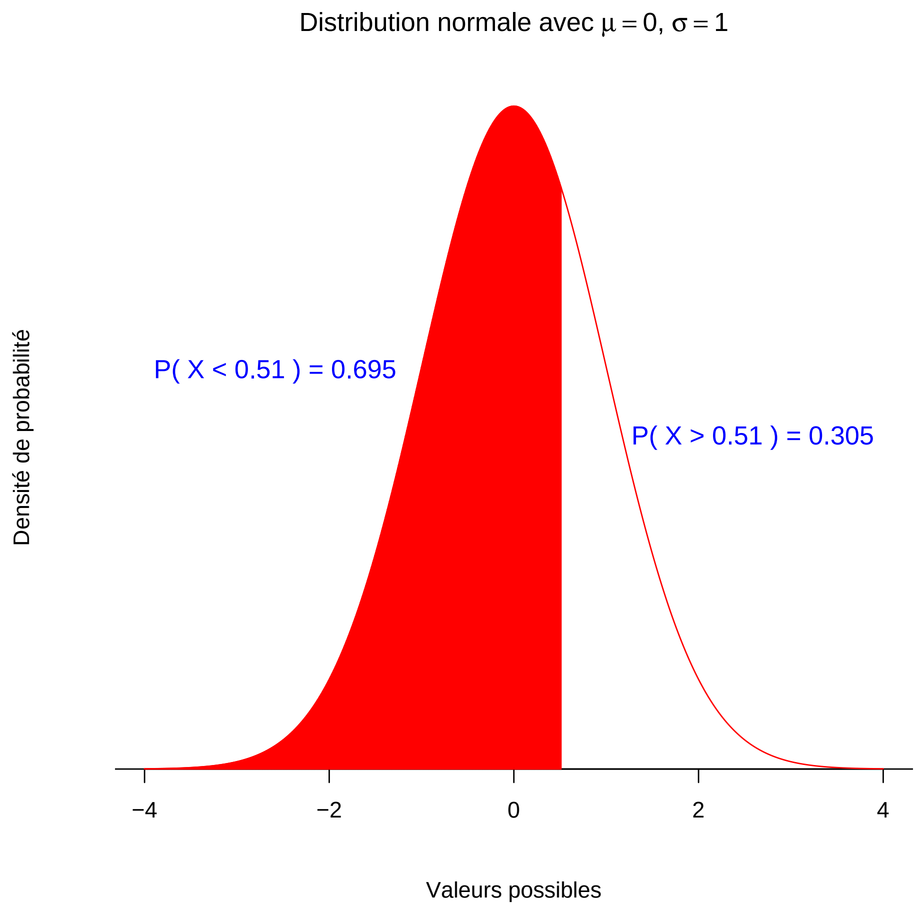


Table de la loi normale

La table de la loi normale donne les probabilités d'occurrence jusqu'à la z-valeur considérée. La ligne donne la valeur de Z jusqu'au dixième, et la colonne donne la valeur au centième.

z	0.00	0.01	0.02	0.03	0.04	...
0	0.5000	0.5040	0.5080	0.5120	0.5160	...
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	...
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	...
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	...
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	...
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	...
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	...
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	...
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	...
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	...
...	...	...	...	...	...	...

Normale centrée réduite

Une loi normale de moyenne nulle et d'écart type 1, écrite $\mathcal{N}(0,1)$, est dite *loi normale centrée réduite*. La fonction de densité est alors

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Transformation

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad \Rightarrow \quad Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1)$$

Les probabilités suivantes sont alors équivalentes

$$X \leq x \quad \Rightarrow \quad Z \leq \frac{x - \mu}{\sigma}$$

$$P(X \leq x) = P(Z \leq \frac{x - \mu}{\sigma})$$

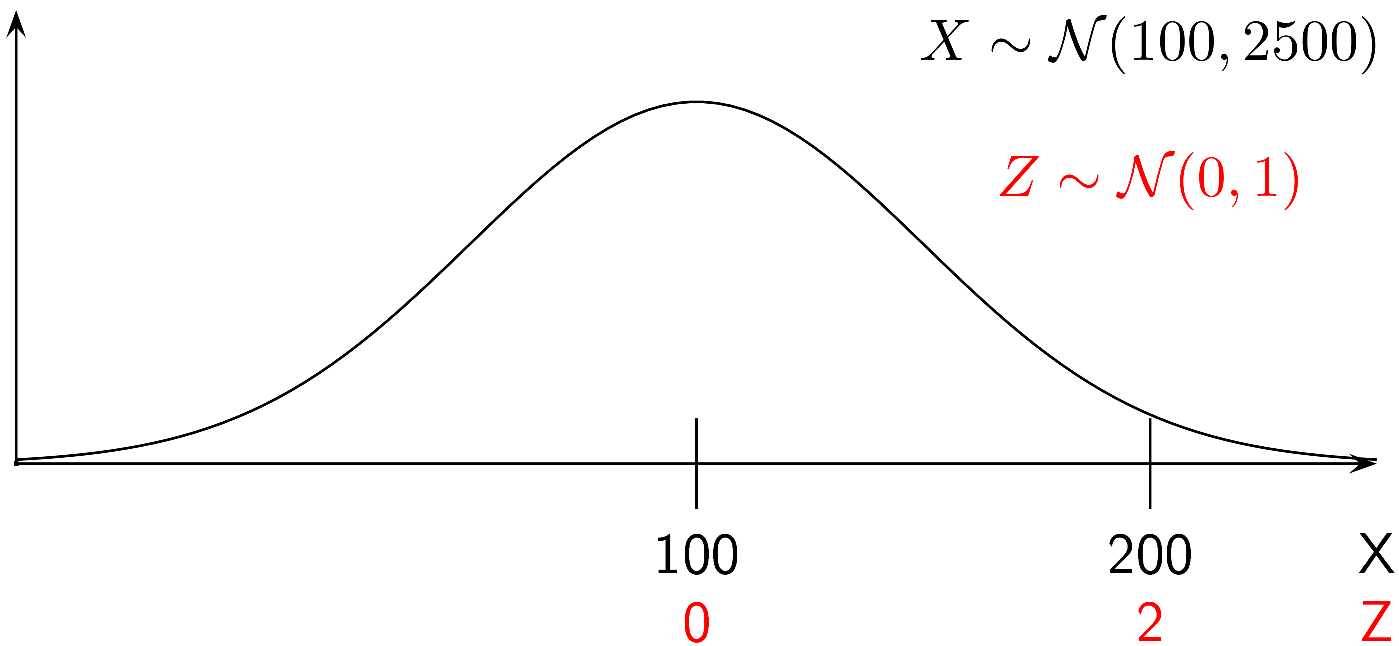
Inversement, on a

$$Z \sim \mathcal{N}(0,1) \quad \Rightarrow \quad X = \mu + Z\sigma \sim \mathcal{N}(\mu, \sigma^2)$$

Exemple

Soit X est une variable aléatoire suivant une loi normale de centre $\mu = 100$ et d'écart type $\sigma = 50$, *i.e.*

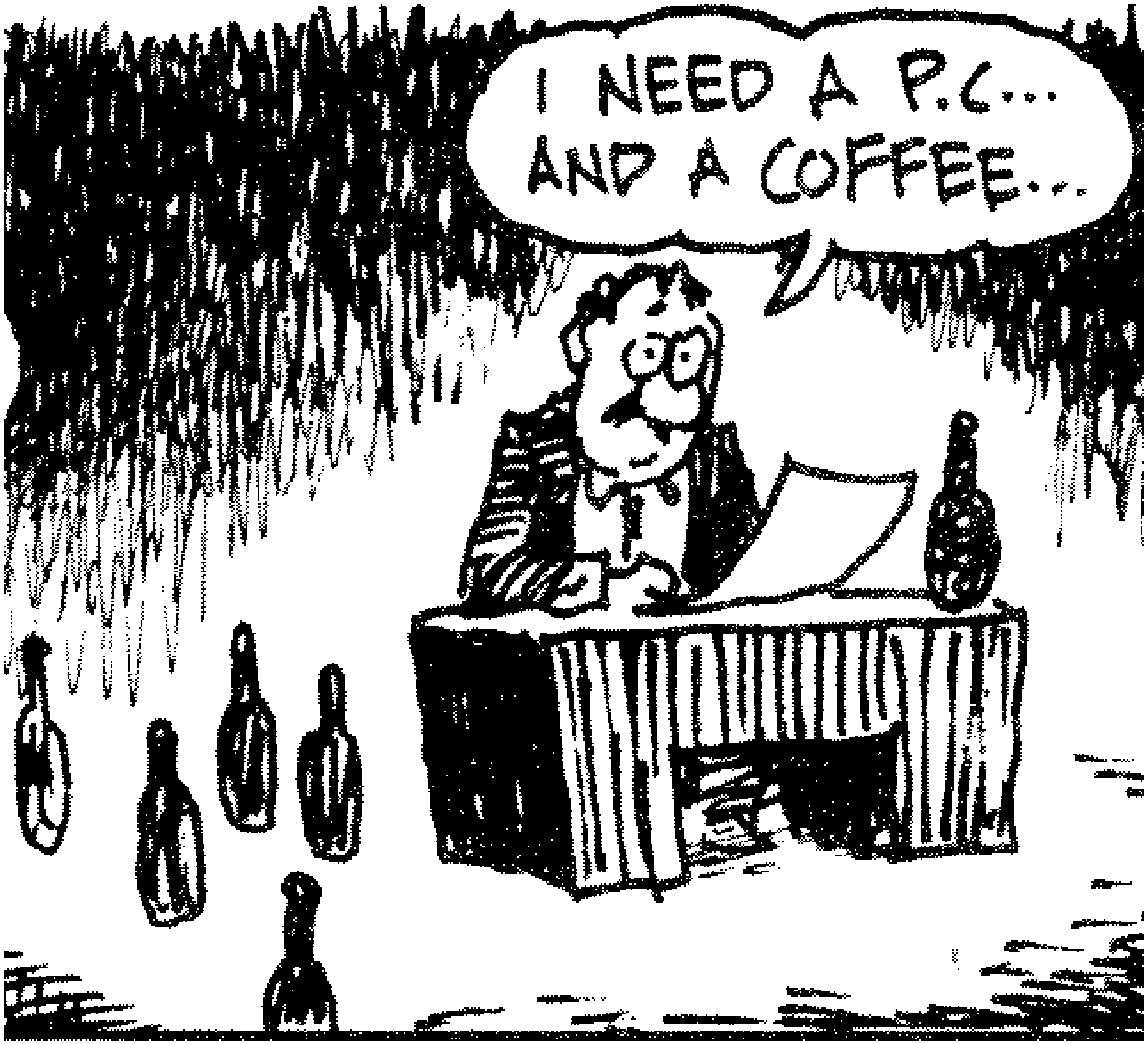
$$X \sim \mathcal{N}(100, 2500) \qquad Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$$



Si $X = 200$ alors $Z = \frac{200-100}{50} = 2$
Et donc $\mu + Z\sigma = 100 + 2 \cdot 50 = 200 = X$

Loi de Student $\mathcal{T}_n$

William Gosset

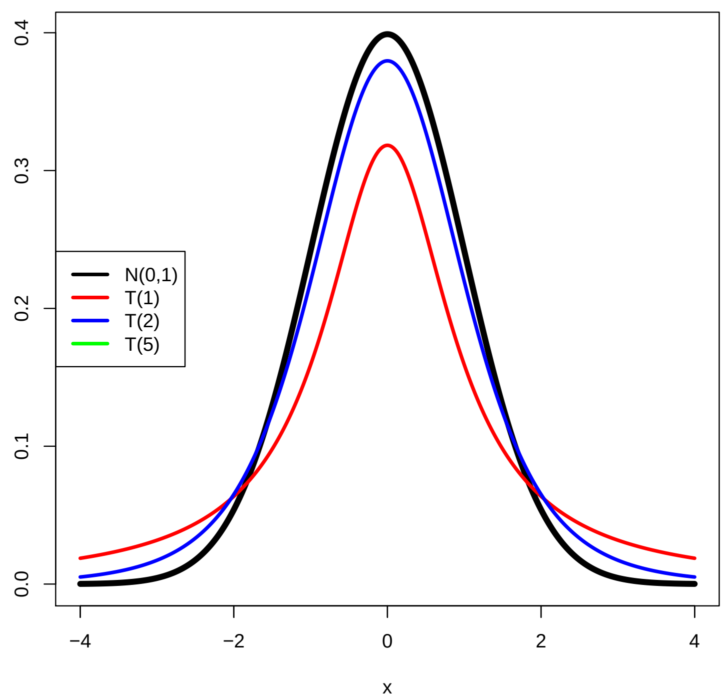


source : "The Cartoon Guide to Statistics", L. Gonick & W. Smith

$\mathcal{T}_n$

Loi de Student à n degrés de liberté

- Distribution de Student (t-distribution) = famille de distribution en forme de cloche et symétrique.
- Caractéristique : nombre de degrés de liberté



Propriétés

- $\mathrm{E}(\mathcal{T}_n) = 0\,, \quad n > 1$

  Espérance n'existe pas lorsque $n = 1$.

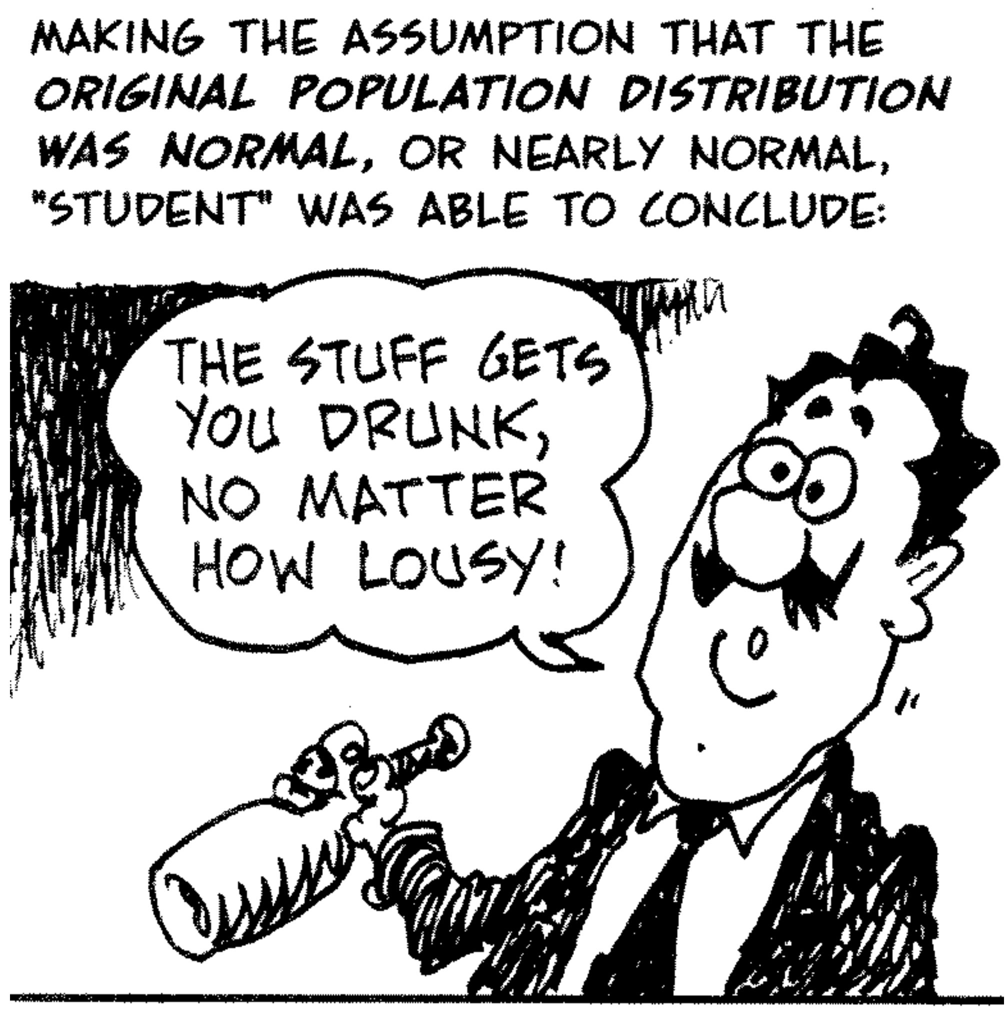
  Symétrie autour de 0.

- $\mathrm{Var}(\mathcal{T}_n) = \frac{n}{n-2}\,, \quad n > 2$

  Variance infinie pour $n \leq 2$

Remarque. Lorsque le nombre de degrés de liberté n tend vers l'infini, la loi de Student tend vers la loi normale $\mathcal{N}(0, 1)$.

William Gosset



source : "The Cartoon Guide to Statistics", L. Gonick & W. Smith

Table de Student

La distribution de Student est tabulée, tout comme la loi normale.

- **Ligne**
  nombre de degrés de liberté n
- **Colonne**
  une erreur de première espèce α.
- **Intersection ligne/colonne** $t_{\alpha,n}$

$$P(\mathcal{T}_n > t_{\alpha,n}) = \alpha \quad \text{et} \quad P(\mathcal{T}_n \leq t_{\alpha,n}) = p$$

La relation entre p et α est $p = 1 - \alpha$.

Exemple

$$P(\mathcal{T}_{10} \leq t_{\alpha,10}) = 0.95 \quad \Longrightarrow t_{0.05,10} = 1.8125$$

t	**Valeurs de α**					
	0.2	**0.1**	**0.05**	**0.025**	**0.01**	**0.005**
dl						
1	1.3764	3.0777	6.3138	12.7062	31.8205	63.6567
2	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.9785	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.9410	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.8960	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Théorème

Soit un échantillon aléatoire de taille n, de moyenne $\bar{x}$ et de variance s^2, issu d'une loi normale $\mathcal{N}(\mu, \sigma^2)$. Alors

$$\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \sim \mathcal{T}_{n-1}$$

Utilité : inférence sur la moyenne d'une population suivant une loi normale de moyenne μ et de variance σ^2 inconnue.

Loi du χ^2

Définition et propriétés

Soit n variables aléatoires normales centrées-réduites Z_i, indépendantes les unes des autres et identiquement distribuées : $Z_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $i = 1, 2, \ldots, n$. Alors la variable formée de la somme des carrés de ces variables

$$Q_n = \sum_{i=1}^{n} Z_i^2 \sim \chi^2$$

suit une *loi du* χ^2 *à* n *degrés de liberté*, ce que l'on note souvent $\chi^2(n)$ ou χ^2_n.

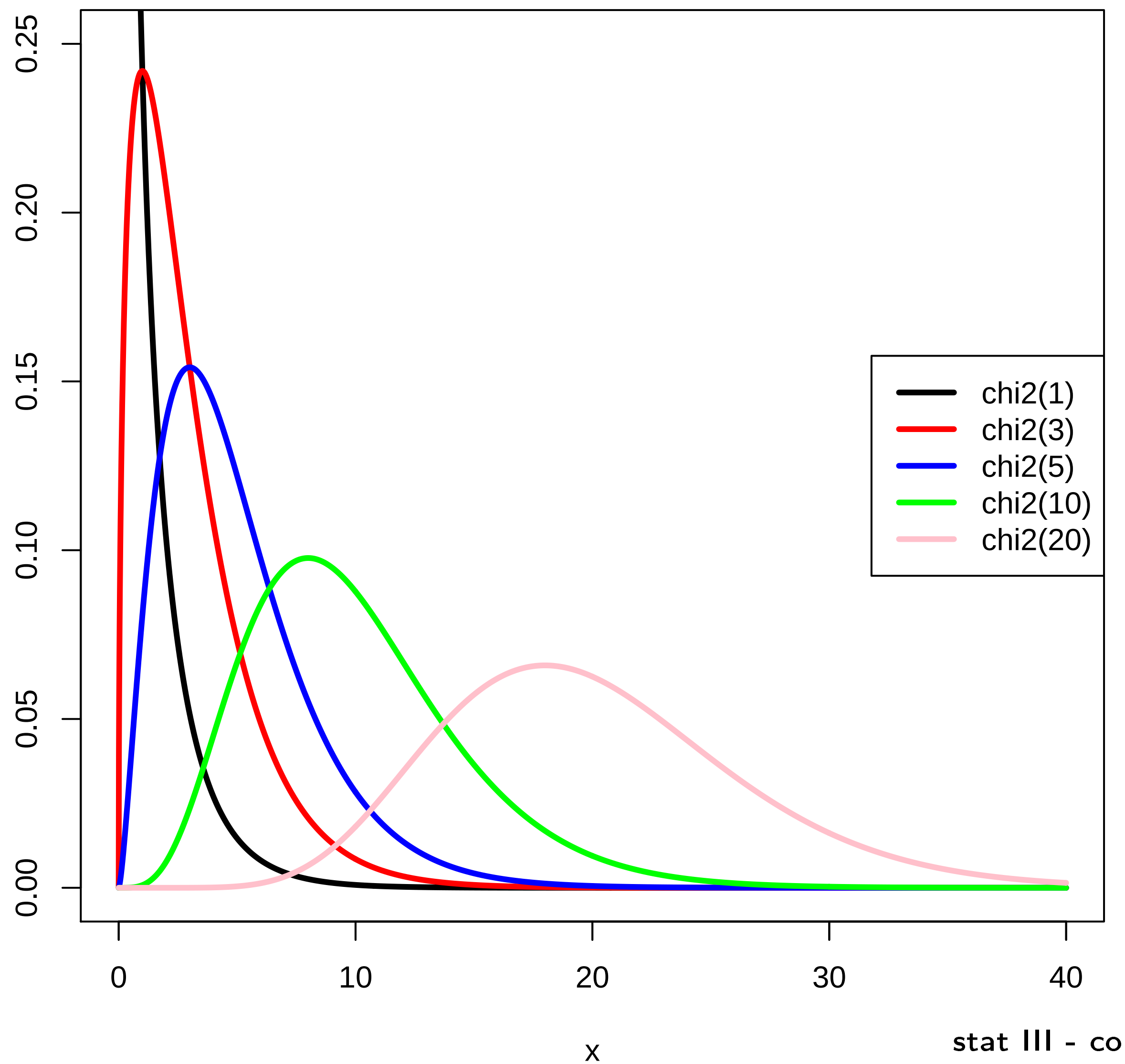
Remarque : les valeurs sont forcément positives.

Propriétés :

- Son espérance vaut $\mathrm{E}(Q_n) = n$
- Sa variance vaut $\mathrm{Var}(Q_n) = 2n$

h e g

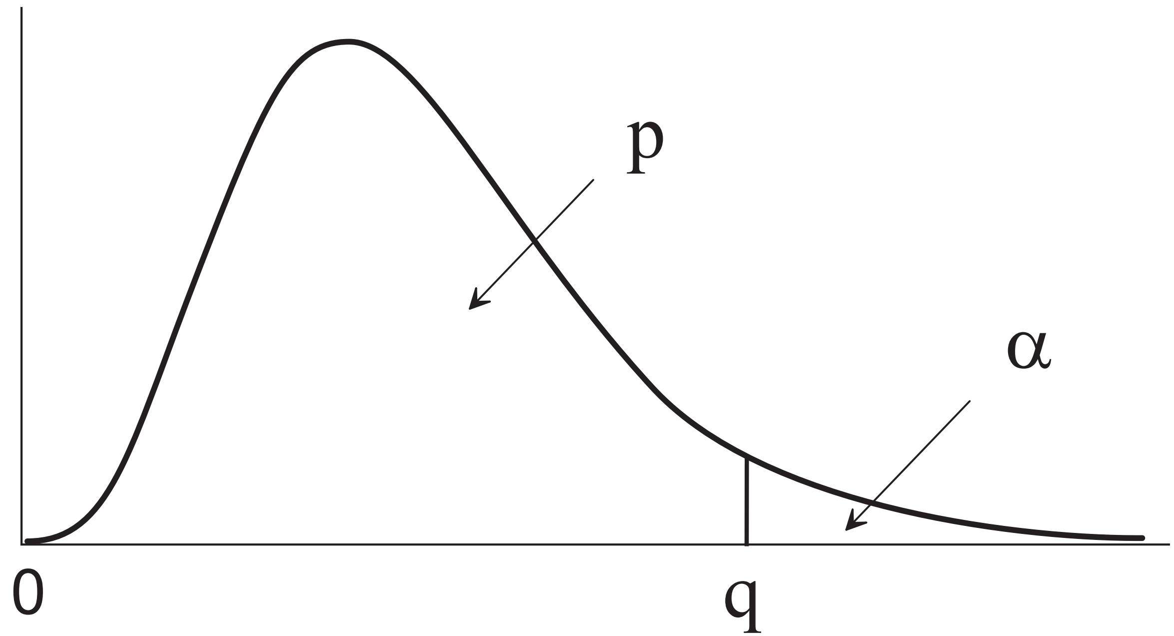
Illustration



Table

Soit $Q_n \sim \chi^2_n$
$P(Q_n \leq q_{\alpha,n}) = p$ et $P(Q_n > q_{\alpha,n}) = \alpha$ avec

$$p = 1 - \alpha$$



Exemple

$P(Q_7 \leq q_{\alpha,7}) = 0.95$
Alors $\alpha = 1 - 0.95 = 0.05$
Et donc $\quad q_{0.05,7} = 14.0671$

	Valeurs de α					
	0.99	**0.975**	**0.95**	**0.9**	**0.1**	**0.05**
dl						
1	0.0002	0.0010	0.0039	0.0158	2.7055	3.8415
2	0.0201	0.0506	0.1026	0.2107	4.6052	5.9915
3	0.1148	0.2158	0.3518	0.5844	6.2514	7.8147
4	0.2971	0.4844	0.7107	1.0636	7.7794	9.4877
5	0.5543	0.8312	1.1455	1.6103	9.2364	11.0705
6	0.8721	1.2373	1.6354	2.2041	10.6446	12.5916
7	1.2390	1.6899	2.1673	2.8331	12.0170	14.0671
8	1.6465	2.1797	2.7326	3.4895	13.3616	15.5073
9	2.0879	2.7004	3.3251	4.1682	14.6837	16.9190
10	2.5582	3.2470	3.9403	4.8652	15.9872	18.3070
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Propriété

La statistique χ^2 à $n-1$ degrés de liberté vaut

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

où

χ^2	$=$	variable chi-2 standard
s^2	$=$	variance de l'échantillon
σ^2	$=$	variance de la population
n	$=$	taille de l'échantillon

Utilité : inférence sur la variance d'une population