La loi normale

Pierre Duchesne

January 2, 2017

Loi normale

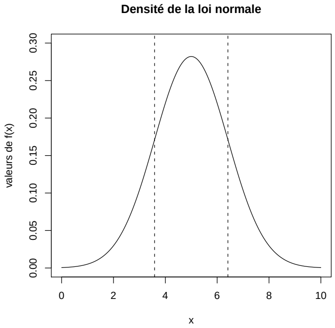
On dit qu'une v.a. X admet une loi normale de paramètres μ et σ^2, notée $\mathcal{N}(\mu, \sigma^2)$, où $\mu \in \mathbb{R}$ et $\sigma^2 \in \mathbb{R}^+$, si sa fonction de densité est:

$$f_X(x) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty.$$

Si $Z \sim \mathcal{N}(0,1)$ on dira que c'est une normale centrée réduite et parfois l'on notera la densité $\phi(z)$, $z \in \mathbb{R}$. De plus, si $Z \sim \mathcal{N}(0,1)$, on notera:

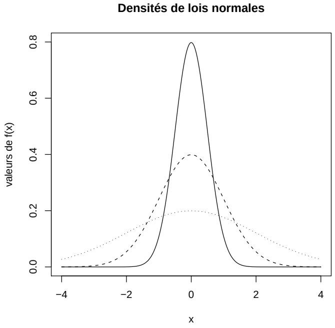
$$\begin{aligned} F_Z(z) &= P(Z \leq z), \\ &= \int_{-\infty}^{z} \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{x^2}{2}\right) dx, \\ &= \Phi(z). \end{aligned}$$

Exemple d'une loi normale



Le graphique précédent illustre une loi $\mathcal{N}(5, 2)$. Les points d'inflexions si situent à $5 \pm \sqrt{2}$. En général, si $X \sim \mathcal{N}(\mu, \sigma^2)$, les points d'inflexion se situent à $\mu \pm \sigma$.

Autres exemples de loi normale



Dans cet exemple, $\mu = 0$, et $\sigma = 0.5, 1, 2$.

Il n'existe pas de forme explicite pour $\Phi(z)$. On utilise des tables, ou encore mieux, l'ordinateur (en `R`, `dnorm()` donne la densité, `pnorm()` la f.r.).

On note que la distribution normale est symétrique par rapport à μ: $f_X(\mu - x) = f_X(x - \mu)$. Le maximum est à μ.

La loi normale joue un rôle central en statistique.

On parle également de la distribution gaussienne, en l'honneur de Carl Friedrich Gauss, qui étudiait les erreurs de mesure.

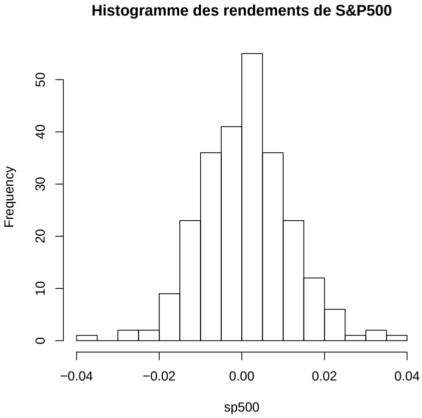
Interprétation: Une v.a. est normale si elle est la résultante d'un très grand nombre de causes, et qui prise une à une, ont un effet négligeable.

Exemple: Standard & Poors 500

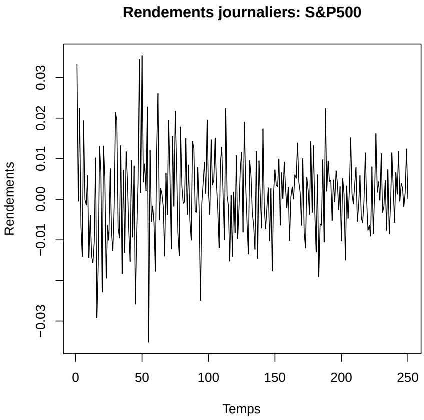
Le S & P 500 est un indice boursier basé sur 500 grandes sociétés cotées sur les bourses américaines. Standard & Poors est une société de notation financière qui gère cet indice.

Les rendements sont souvent compatibles avec l'hypothèse de normalité.

On dispose des rendements pour l'année 2003.


Histogramme des rendements de S&P500
Frequency
0
10
20
30
40
50
−0.04
−0.02
0.00
0.02
0.04
sp500

Les rendements sont souvent représentés sous la forme d'une série chronologique

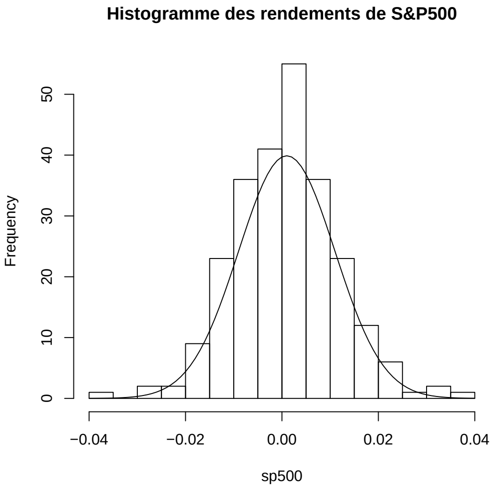


Calcul de statistiques descriptives:

$$\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t = 0.0009869522,$$

$$S^2 = \frac{1}{n-1}\sum_{t=1}^{n}(X_t - \bar{X}_n)^2 = 0.0001170825$$

En R, on peut utiliser les fonctions `mean()` et `var()` pour ce genre de calculs de statistiques descriptives.



Histogramme des rendements de S&P500
Frequency
0
10
20
30
40
50
−0.04
−0.02
0.00
0.02
0.04
sp500

Application: Calcul de la Valeur à Risque (VaR)

Une compagnie financière pourrait utiliser le modèle normal,

$$\mathcal{N}(\mu = 0.001, \sigma^2 = 0.001)$$

pour calculer la VaR.

La fonction `R` `qnorm()` permet de trouver les quantiles de la loi normale. Ici:

`qnorm(0.05, mean=0.0009869522,sd=0.01082047) = -0.0168`.

- Ainsi, utilisant un horizon temporel de un jour et un niveau de confiance de 95%, si on dénote V_0 la valeur courante de l'indice, la VaR est de $0.0168\,V_0$.
- Ainsi, si V_0 est 10$ millions, la VaR est de $168 000.
- Une interprétation est qu'avec un niveau de confiance de 95%, la perte ne devrait pas excéder $168 000 pour le jour suivant.
- Il ne faut pas s'étonner si la perte est plus grande un jour sur 20 journées actives.

Le théorème de De Moivre-Laplace

C'est un exemple de théorème limite.
Considérons une v.a. X, telle que $X \sim Bin(n, p)$. On suppose que $n \to \infty$ alors que p demeure fixé.
On sait que $X = \sum_{i=1}^{n} X_i$, où $X_i \sim Bernoulli(p)$.
Or

$$\frac{Var(X_i)}{Var(X)} = \frac{pq}{npq} = \frac{1}{n} \to 0,$$

et ce quelque soit i fixé.
Intuitivement, on s'attend à ce que la loi de X se rapproche d'une loi normale $N(\mu, \sigma^2)$ avec $\mu = E(X) = np$, $\sigma^2 = Var(X) = npq$ à mesure que n augmente.

L'énoncé du théorème de De Moivre - Laplace est le suivant: Soit $\{X_1, X_2, \ldots\}$ une suite de v.a. indépendantes *Bernoulli*(p), alors $\sum_{i=1}^{n} X_i$ est *Bin*(n, p), $n = 1, 2, \ldots$. Posons:

$$Z_n = \frac{\sum_{i=1}^{n} X_i - np}{(npq)^{1/2}}.$$

alors

$$\lim_{n \to \infty} P(Z_n \leq z) = \Phi(z).$$

Ainsi, à mesure que n augmente, la loi d'une *Bin*(n, p), correctement centrée et standardisée (on dit également réduite), s'approche de celle d'une $\mathcal{N}(0, 1)$.

Remarques:

1. L'approximation s'avère bonne dès que $np \geq 5$, $nq \geq 5$.
2. Le théorème de De Moivre - Laplace est un cas particulier d'un théorème beaucoup plus fondamental: le théorème limite central.
3. Dans l'énoncé du théorème de De Moivre - Laplace, la suite de v.a. $\{Z_n\}_{n=1}^{\infty}$ converge *en loi* (on dit également *en distribution*) vers une v.a. limite Z, disons, qui est dans le cas présent une $\mathcal{N}(0,1)$:

$$\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z).$$

   On sera donc amené à considérer la notion de convergence d'une suite de v.a. vers une v.a. limite.

Comparaison d'une $Bin(10, \frac{1}{2})$ avec une $\mathcal{N}(5, \frac{5}{2})$