7. Loi normale et théorème central limite

MTH2302D

S. Le Digabel, École Polytechnique de Montréal

A2017

(v2)

Plan

Loi normale

On dit qu'une variable aléatoire continue X suit une *loi normale* de paramètres μ et σ^2 si sa fonction de densité est

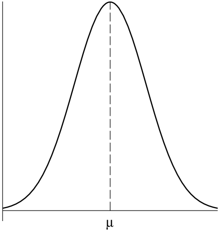
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \text{ pour tout } x\ .$$

On dénote ceci $X \sim \mathsf{N}(\mu, \sigma^2)$.

Loi normale : propriétés

Propriétés de f_X :

1. $\lim_{x \to \pm\infty} f_X(x) = 0.$
2. $f_X(\mu + x) = f_X(\mu - x)$ (symétrie par rapport à l'axe $x = \mu$).
3. f_X atteint son maximum en $x = \mu$ (μ est *le mode* de X).
4. Les points d'inflexion du graphe de f_X sont $x = \mu \pm \sigma$.



Loi normale : propriétés (suite)

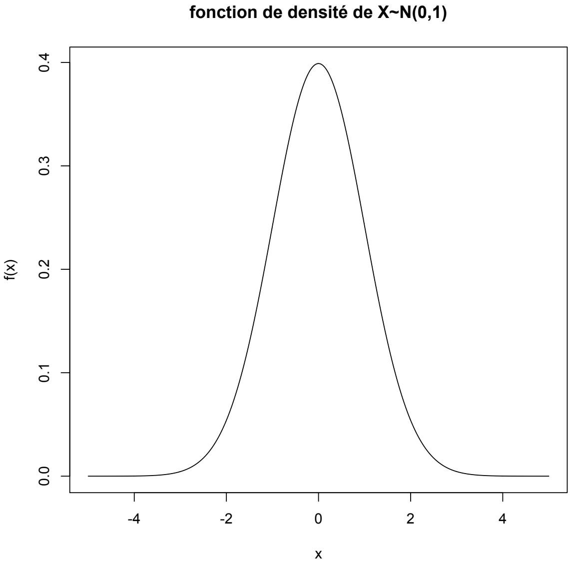
Si $X \sim \mathrm{N}(\mu, \sigma^2)$ alors

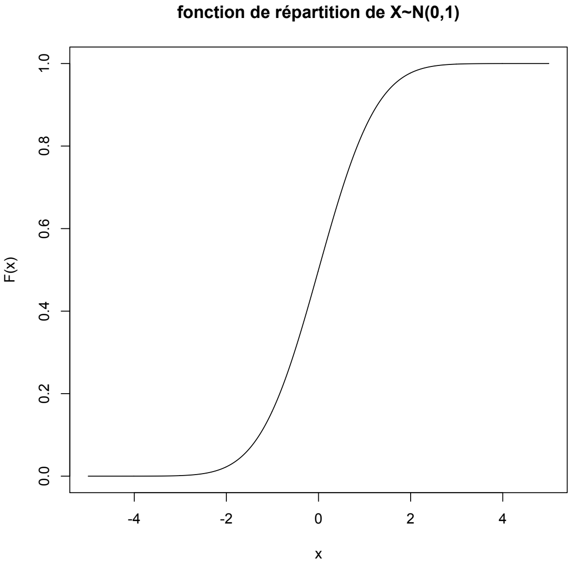
1. $P(X < \mu - x) = P(X > \mu + x)$.
2. $F_X(\mu - x) = 1 - F_X(\mu + x)$.

Moyenne et variance de la loi normale

Si $X \sim \mathsf{N}(\mu, \sigma^2)$ alors

1. $\mathsf{E}(X) = \mu$.
2. $\mathsf{V}(X) = \sigma^2$.


fonction de densité de X~N(0,1)
f(x)
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
x


fonction de répartition de X~N(0,1)
F(x)
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
x

Loi normale : calcul avec des logiciels

- Excel :
  $f_X(x) =$ LOI.NORMALE(x, μ, σ, 0).
  $F_X(x) =$ LOI.NORMALE(x, μ, σ, 1).

- R :
  $f_X(x) =$ dnorm(x, mean=μ, sd=σ).
  $F_X(x) =$ pnorm(x, mean=μ, sd=σ).

Loi normale centrée réduite

Lorsque $\mu = 0$ et $\sigma^2 = 1$, la loi normale $\mathrm{N}(0,1)$ est appelée *centrée réduite* et on la dénote par Z.

Sa fonction de densité est

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \ .$$

Sa fonction de répartition est

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt \ .$$

Puisque cette intégrale est difficile à évaluer, on a recours à une table de loi normale pour calculer $\Phi(z)$. Voir livre page 476 (2ème édition) / page 512 (3ème édition) ou sur le site web du cours.

Loi normale centrée réduite (suite)

Si $X \sim \mathsf{N}(\mu, \sigma^2)$ alors

$$Z = \frac{X - \mu}{\sigma} \sim \mathsf{N}(0, 1).$$

On peut donc ramener toute loi normale à une loi centrée réduite.

Méthodes de calcul

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

- Si $b \geq 0$ alors $P(Z \leq b) = \Phi(b)$.
- Si $b < 0$ alors
  $\Phi(b) = P(Z \leq b) = 1 - P(Z \leq -b) = 1 - \Phi(-b)$.
- $P(Z \geq b) = 1 - P(Z \leq b) = 1 - \Phi(b)$.
- $P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a) = \Phi(b) - \Phi(a)$.

Si $X \sim \mathrm{N}(\mu, \sigma^2)$, alors $P(X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right)$ et

$$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Exemples

- **Exemple 1 :** Vérifier que $\mu = 0$ et $\sigma^2 = 1$ si $X \sim \mathsf{N}(0,1)$.
- **Exemple 2 :** Déterminer Q_1, Q_2 et Q_3 si $X \sim \mathsf{N}(0,1)$.

Exemple 3

Si $Z \sim \mathsf{N}(0, 1)$, calculer

1. $P(Z \leq 1.25)$.
2. $P(Z \leq -0.52)$.
3. $P(Z > -1)$.
4. Si $X \sim \mathsf{N}(\mu = 100, \sigma^2 = 4)$, calculer $P(98 < X \leq 104)$.
5. Si $P(Z \leq b) = 0.6628$, déterminer b.
6. Si $P(Z \leq b) = 0.3446$, déterminer b.

Additivité

Soit $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes avec

$$X_i \sim \mathsf{N}(\mu_i, \sigma_i^2) \text{ pour tout } i.$$

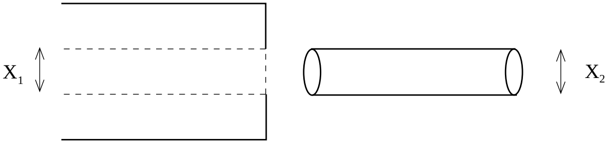
Soit $Y = a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$.

Alors $Y \sim \mathsf{N}(\mu, \sigma^2)$, où

$$\begin{aligned} \mu &= a_0 + a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n \ , \\ \sigma^2 &= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2 \ . \end{aligned}$$

Exemple 4

Un assemblage consiste à insérer un arbre dans un palier selon le schéma ci-dessous.



Si $X_1 \sim \mathrm{N}(1.5, 0.0016)$ et $X_2 \sim \mathrm{N}(1.48, 0.0009)$ sont les deux diamètres, le jeu entre les deux éléments est $Y = X_1 - X_2$. Les v.a. X_1 et X_2 sont indépendantes.

L'assemblage échoue si $X_1 < X_2$.

Dans quel pourcentage de cas l'assemblage échoue-t-il ?

Approx. d'une loi binomiale par une loi normale

Soit $X \sim \mathsf{B}(n,p)$ une variable aléatoire suivant une loi binomiale. Alors X est la somme de variables de Bernoulli indépendantes de paramètre p.

Si n est grand alors X suit approximativement une loi normale $\mathsf{N}(\mu = np, \sigma^2 = np(1-p))$.

Cette approximation est bonne si

- $np > 5$ lorsque $p \leq \frac{1}{2}$.
- $n(1-p) > 5$ lorsque $p > \frac{1}{2}$.

Approx. d'une binomiale par une normale (suite)

Puisque $X \sim \mathrm{B}(n, p)$ est une variable discrète, on cherche à calculer des probabilités comme $P(X = x)$.

Or ceci n'a pas de sens pour une v.a. continue et on doit corriger la valeur cherchée pour pouvoir utiliser l'approximation de X par une loi normale.

Par exemple

Valeur cherchée	Valeur corrigée
$P(X = x)$	$P(x - \frac{1}{2} \leq X \leq x + \frac{1}{2})$
$P(a \leq X \leq b)$	$P(a - \frac{1}{2} \leq X \leq b + \frac{1}{2})$

Cette correction est appelée *correction pour la continuité*.

Exemple 5

On lance une pièce 200 fois. Quelle est la probabilité d'obtenir au moins 110 piles ?

Loi lognormale

Une variable aléatoire X suit une *loi lognormale* de paramètres μ_Y et σ_Y^2 si

$$Y = \ln(X) \sim \mathsf{N}(\mu_Y, \sigma_Y^2) \ .$$

C'est équivalent à définir $X = \exp(Y)$.

Loi lognormale : fonction de densité

La fonction de densité d'une variable aléatoire lognormale X de paramètres μ_Y, σ_Y^2 est

$$f_X(x) = \begin{cases} \dfrac{1}{x\sigma_Y\sqrt{2\pi}} \exp\left(-\dfrac{(\ln(x)-\mu_Y)^2}{2\sigma_Y^2}\right) & \text{si } x > 0\ , \\ 0 & \text{sinon.} \end{cases}$$

Loi lognormale : fonction de répartition

La fonction de répartition d'une variable aléatoire lognormale X de paramètres μ_Y, σ_Y^2 est

$$F_X(x) = \begin{cases} \Phi\left(\dfrac{\ln(x) - \mu_Y}{\sigma_Y}\right) & \text{si } x > 0 \ , \\ 0 & \text{sinon.} \end{cases}$$

Loi lognormale : moyenne et variance

Soit X une variable aléatoire lognormale de paramètres μ_Y, σ_Y^2.

Alors :

1. $\mathsf{E}(X) = \exp(\mu_Y + \frac{1}{2}\sigma_Y^2)$.
2. $\mathsf{V}(X) = \exp(2\mu_Y + \sigma_Y^2)\big(\exp(\sigma_Y^2) - 1\big)$.

Loi lognormale : propriétés

Soit $X_1, X_2, \ldots, X_n$ des variables aléatoires lognormales indépendantes de paramètres μ_{Y_i}, et $\sigma^2_{Y_i}$ pour $i = 1, 2, \ldots, n$. Alors

$$W = bX_1^{a_1} X_2^{a_2} \cdots X_n^{a_n}$$

suit une loi lognormale de paramètres

$$\mu_Y = \ln(b) + \sum_{i=1}^{n} a_i \mu_{Y_i}$$

et

$$\sigma^2_Y = \sum_{i=1}^{n} a_i^2 \sigma^2_{Y_i}.$$

Exemple 6

Soit

$$\begin{aligned} Y_1 &= \ln(X_1) \sim \mathsf{N}(4,1) \\ Y_2 &= \ln(X_2) \sim \mathsf{N}(3,0.5) \\ Y_3 &= \ln(X_3) \sim \mathsf{N}(2,0.4) \\ Y_4 &= \ln(X_4) \sim \mathsf{N}(1,0.01) \end{aligned}$$

et

$$W = e^{1.5}\left(X_1^{2.5} X_2^{0.2} X_3^{0.7} X_4^{3.1}\right).$$

Calculer $P(2 \times 10^4 \leq W \leq 6 \times 10^5)$.

Loi des grands nombres

Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes ayant la même distribution, avec $\mathsf{E}(X_i) = \mu$ pour $i = 1, 2, \ldots, n$.

Alors pour tout $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left[\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right] = 0\ .$$

Théorème central limite (TCL)

Soit $X_1, X_2, \ldots, X_n$ une suite de variables aléatoires indépendantes, avec $\mathsf{E}(X_i) = \mu_i$ et $\mathsf{V}(X_i) = \sigma_i^2$ pour $i = 1, 2, \ldots, n$.

Alors la variable aléatoire

$$Z = \frac{\sum_{i=1}^{n} (X_i - \mu_i)}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$$

suit approximativement une loi normale $\mathsf{N}(0, 1)$ si n est grand.

TCL : cas particulier

Soit $X_1, X_2, \ldots, X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées, avec $\mathsf{E}(X_i) = \mu$ et $\mathsf{V}(X_i) = \sigma^2$ pour $i = 1, 2, \ldots, n$.

Alors la variable aléatoire

$$Z = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$$

suit approximativement une loi normale $\mathsf{N}(0,1)$ si n est grand.

Autres formulations (pour n $\to \infty$) :

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2), \text{ ou}$$

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

Exemple 7

On lance un dé 100 fois. Quelle est la probabilité que la somme des résultats soit entre 340 et 360 ?

Théorème central limite : quelle valeur de n est assez grande ?

- Si les lois des X_i sont proches d'une loi normale alors pour $n \geq 4$ l'approximation donnée par le théorème central limite est bonne.
- Si les lois des X_i sont moyennement proches d'une loi normale (p. ex. loi uniforme) alors pour $n \geq 12$ l'approximation donnée par le théorème central limite est bonne.
- Si les lois des X_i ne sont pas proches d'une loi normale alors pour $n \geq 100$ l'approximation donnée par le théorème central limite sera bonne (par exemple fonction de densité très asymétrique).