

Biostatistique

Loi normale et théorème central limite

Anicet Ebou, Institut National Polytechnique Félix Houphouët-Boigny,
ediman.ebou@inphb.ci

Plan

Loi normale	3
Loi normale centrée réduite	11
loi lognormale	19
Théorèmes limites	23

Loi normale

Loi normale

De nombreuses variables, en particulier les variables « biologiques », tendent à être distribuées normalement. Par exemple, si l'on mesure la taille de l'ensemble de la population féminine adulte d'une grande ville et que l'on représente la fréquence des individus en fonction de leur taille, la distribution ressemblera à une cloche symétrique, que l'on appelle la **distribution normale**.

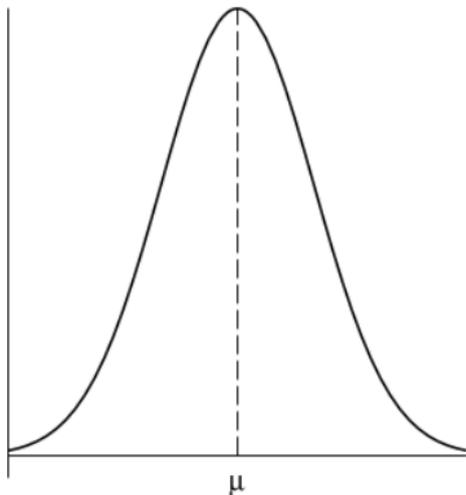
On dit qu'une variable aléatoire continue X suit une *loi normale* de paramètres μ et σ^2 si sa fonction de densité est

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

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

Propriétés de la loi normale

1. $\lim_{x \rightarrow \pm\infty} f_X(x) = 0$
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 $x = \mu \pm \sigma$



Propriétés de la loi normale (ii)

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

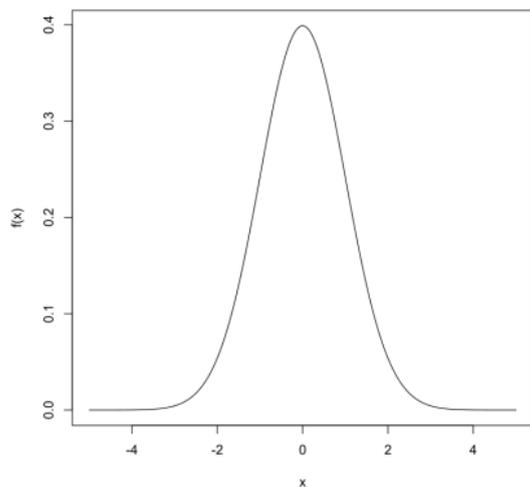
1. $P(X < \mu - x) = P(X > \mu + x)$
2. $F_X(\mu - x) = 1 - F_X(\mu + x)$

Moyenne et variance de la loi normale

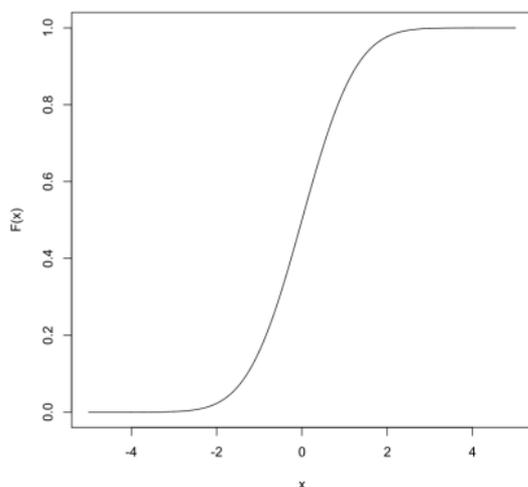
Si $X \sim N(\mu, \sigma^2)$ alors

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

fonction de densité de $X \sim N(0,1)$



fonction de répartition de $X \sim N(0,1)$



L'écart-type

L'écart-type est important parce que la moyenne d'une population normale plus ou moins 1 écart-type comprends 68,27% des valeurs de cette population.

De plus, 95% des valeurs dans une population normale sont comprises entre plus ou moins 1,96 écart-type de la moyenne. Cela est particulièrement important vu que les 5% de valeurs restantes seront en dehors de cet intervalle et par conséquent plus loin de la moyenne.

Ces deux statistiques sont tout ce dont on a besoin pour décrire la position et la forme d'une distribution normale. Ils peuvent aussi être utilisé pour déterminer la proportion de la population qui est plus ou moins grande qu'une certaine valeur.

L'écart-type (ii)

Exemple: Pour une population normalement distribuée dont la taille moyenne est de 170 cm et l'écart-type de 10, 95 % des individus de cette population auront une taille comprise entre $170 (1,96 \times 10)$ (c'est-à-dire entre 150,4 et 189,6 cm). Vous n'avez que 5 % de chances de trouver quelqu'un qui soit plus grand que 189,6 cm ou plus petit que 150,4 cm.

Calcul avec les logiciels

- Excel:
 - $f_X(x) = \text{LOI.NORMALE}(x, \mu, \sigma, 0)$
 - $F_X(x) = \text{LOI.NORMALE}(x, \mu, \sigma, 1)$
- R:
 - $f_X(x) = \text{dnorm}(x, \text{mean}=\mu, \text{sd}=\sigma)$
 - $F_X(x) = \text{pnorm}(x, \text{mean}=\mu, \text{sd}=\sigma)$

Loi normale centrée réduite

Statistique Z

Pour une distribution normale, la différence entre une valeur quelconque et la moyenne, divisée par l'écart-type, donne un rapport appelé statistique Z qui est également normalement distribué, avec une moyenne de zéro et un écart-type de 1,00.

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

Par conséquent, la valeur de la statistique Z indique le nombre d'écarts types par rapport à la moyenne.

En prenant l'exemple plus haut, une valeur de 189,6 cm est $\frac{189,6-170}{10} = 1,96$ écart-types par rapport à la moyenne.

Lorsque ce rapport est supérieur à +1,96 ou inférieur à -1,96, la probabilité d'obtenir cette valeur de X est inférieure à 5%.

Loi normale centrée réduite

Lorsque $\mu = 0$ et $\sigma^2 = 1$, la loi normale $N(0, 1)$ 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 $\Phi(z)$.

Loi normale centrée réduite (ii)

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

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

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

Méthodes de calcul

Si $Z \sim N(0, 1)$

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

Exercice 1:

1. Vérifier que $\mu = 0$ et $\sigma^2 = 1$ si $X \sim N(0, 1)$
2. Déterminer Q_1, Q_2 et Q_3 si $X \sim N(0, 1)$.
3. Si $Z \sim N(0, 1)$, calculer:
 1. $P(Z \leq 1, 25)$
 2. $P(Z \leq -0, 52)$

Méthodes de calcul (ii)

3. $P(Z > -1)$
4. Si $X \sim N(\mu = 100, \sigma^2 = 4)$, 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, \dots, X_n des variables aléatoires indépendantes avec $X_i \sim N(\mu_i, \sigma_i^2)$ pour tout i .

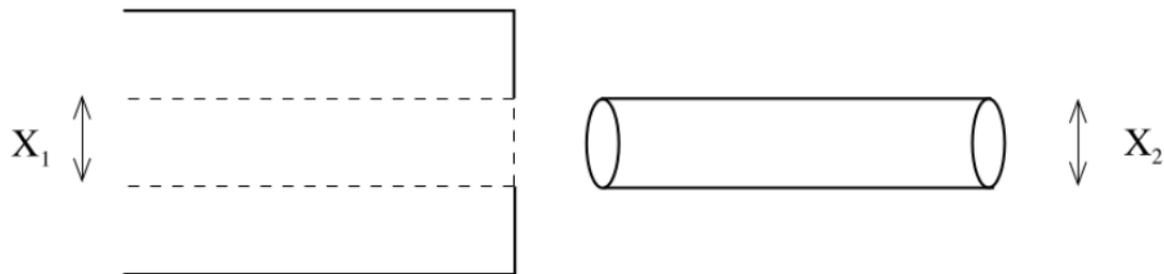
Soit $Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n$. Alors $Y \sim N(\mu, \sigma^2)$, où

$$\mu = a_0 + a_1\mu_1 + a_2\mu_2 + \dots + a_n\mu_n$$

$$\sigma^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \dots + a_n^2\sigma_n^2$$

Additivité (ii)

Exercice 2: Un assemblage consiste à insérer un arbre dans un palier selon le schéma ci-dessous.



Si $X_1 \sim N(1,5; 0,0016)$ et $X_2 \sim N(1,42; 0,0009)$ 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 ?

loi lognormale

loi lognormale

Une variable aléatoire X suit une *loi lognormale* de paramètres μ_Y et σ_Y^2 si $Y = \ln(X) \sim N(\mu_Y, \sigma_Y^2)$

C'est équivalent à définir $X = e^Y$.

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

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

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

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

loi lognormale: moyenne et variance

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

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

Propriétés: Soit X_1, X_2, \dots, X_n des variables aléatoire lognormales indépendantes de paramètres μ_{Y_i} , et $\sigma_{Y_i}^2$, pour $i = 1, 2, \dots, n$. Alors

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

suit une loi lognormale de paramètres

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

et

$$\sigma_Y^2 = \sum_{i=1}^n a_i^2 \sigma_{Y_i}^2$$

loi lognormale: moyenne et variance (ii)

Exercice 3: Soit $Y_1 = \ln(X_1) \sim N(4, 1)$, $Y_2 = \ln(X_2) \sim N(3; 0, 5)$, $Y_3 = \ln(X_3) \sim N(2; 0, 4)$, $Y_4 = \ln(X_4) \sim N(1; 0, 01)$ et $W = e^{1,5} (X_1^{2,5} X_2^{0,2} X_3^{0,7} X_4^{3,1})$

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

Théorèmes limites

Loi faible des grands nombres

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

$$S_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Alors la suite (S_n) tend vers μ en probabilité, c'est-à-dire que, pour tout $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|S_n - \mu| > \varepsilon) = 0$$

Le nom de loi faible des grands nombres est donné à tout théorème donnant une information sur la convergence en probabilité d'une moyenne de variables aléatoires.

Par exemple, on lance un dé non pipé, X_n vaut 1 si le n-ème lancer amène 5, et 0 sinon. Alors (S_n) tend vers 1/6 en probabilité.

Théorème central limite (TCL)

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires indépendantes, avec $\mathbb{E}(X_i) = \mu_i$ et $\mathbb{V}(X_i) = \sigma_i^2$ pour $i = 1, 2, \dots, n$. 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 $N(0, 1)$ si n est grand.

Cas particuliers:

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

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

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

Théorème central limite (TCL) (ii)

Autres formulations pour $n \rightarrow \infty$:

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

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

Exercice 4: On lance un dé 100 fois. Quelle est la probabilité que la somme des résultats soit entre 340 et 360 ?

Théorème central limite: quelle valaue 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 (par exemple loi uniforme), alors pour $n \geq 12$ l'approximation donnée par le théorème central limite est bonne.