

Loi Normale

1. Définition

Soit $\mu \in \mathbb{R}, \sigma \in \mathbb{R}_+^*$ et $\forall x \in \mathbb{R}, f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

La loi de probabilité sur $\mathbb{R}$ de densité f est appelée loi normale (ou loi gaussienne, ou loi de Laplace-Gauss) de paramètres (μ, σ) et notée $\mathcal{N}(\mu, \sigma)$.

2. Propriétés

1/ Si $X \sim \mathcal{N}(\mu, \sigma), (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ alors $E(X) = \mu$ et $V(X) = \sigma^2$

2/ Si $X \sim \mathcal{N}(\mu, \sigma), (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$, si $\alpha \in \mathbb{R}_+^*$ et $\beta \in \mathbb{R}$ alors : $\alpha X + \beta \sim \mathcal{N}(\alpha\mu + \beta, |\alpha|\sigma)$.

En particulier, si on pose $Z = \frac{X - \mu}{\sigma}$, alors Z est une v.a.r. centrée réduite de loi $\mathcal{N}(0,1)$ (loi normale centrée réduite).

La fonction de répartition de la loi $\mathcal{N}(0,1)$ vérifie :

$$\forall x \in \mathbb{R}, F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

Cette fonction est tabulée.

On a :

$$\forall (a, b) \in (\mathbb{R})^2, a < b, P(a < X \leq b) = P\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = F\left(\frac{b - \mu}{\sigma}\right) - F\left(\frac{a - \mu}{\sigma}\right).$$

3/ Si $X_i \sim \mathcal{N}(\mu_i, \sigma_i)$, avec $(\mu_i, \sigma_i) \in \mathbb{R} \times \mathbb{R}_+^*, i = 1, 2$ et X_1 et X_2 indépendantes en probabilité, alors $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

4/ Si $X_i \sim \mathcal{N}(\mu, \sigma)$, avec $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*, i = 1, \dots, n$ et $(X_i)_{i=1, \dots, n}$, indépendantes en probabilité,

alors $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ (c'est-à-dire $\sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma} \right) \sim \mathcal{N}(0,1)$).

5/ Théorèmes limites.

Soit X une v.a.r. d'ordre 2 d'espérance mathématique μ et de variance σ^2 et $(X_i)_{i \in \mathbb{N}^*}$ une suite de v.a.r. i.i.d. (indépendantes et identiquement distribuées) comme X .

Alors la v.a.r. $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$ a pour espérance mathématique μ et pour variance $\frac{\sigma^2}{n}$.

On a de plus la loi faible des grands nombres :

$$(\bar{X}_n) \xrightarrow{\text{Prob}} \mu$$

(c'est-à-dire $\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$),

et le théorème de la limite centrée :

$$\sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma} \right) \xrightarrow{\text{Loi}} \mathcal{N}(0,1).$$

(c'est-à-dire, en utilisant la f.d.r. F_{Z_n} de $Z_{n_i} := \sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma} \right), \forall x \in \mathbb{R} \quad \lim_{n \rightarrow \infty} F_{Z_n}(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$)

Dans le cas particulier d'un schéma de Bernoulli, $X \sim \mathcal{B}(p)$ alors $Y_n = \sum_{i=1}^n X_i \sim \mathcal{B}(n, p)$ et $\bar{X}_n$ (notée F_n) est la fréquence de succès sur n épreuves :

$$\sqrt{n} \frac{F_n - p}{\sqrt{p(1-p)}} = \frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow{\text{Loi}} \mathcal{N}(0,1)$$

(approximation d'une loi binomiale par une loi normale)

3. Applications

3.1. Statistique inférentielle ; échantillonnage, estimation, test d'hypothèse

1/ Echantillonnage

Quelle est la probabilité que, sur 6000 lancers de dé, le nombre de « 6 » s'écarte de 50 ou plus de 1000 ?

Solution

Soit X le nombre de « 6 » sur 6000 lancers, $X \sim \mathcal{B}\left(6000; \frac{1}{6}\right)$ et on a l'approximation

$X \sim \mathcal{N}(1000; 28.87)$ d'où (F désignant la f.d.r. de la loi $\mathcal{N}(0,1)$) :

$$P(|X - 1000| \geq 50) = 1 - P\left(\left|\frac{X - 1000}{28.87}\right| \leq 1.732\right) = 2 - 2F(1.732) = 0.0833$$

2/ Estimation par intervalle de confiance à 95% d'une proportion inconnue p dans une population

(par exemple, la proportion d'électeurs votant pour un candidat A).

On tire un échantillon aléatoire simple à probabilités égales et sans remise de taille n dans la population (on suppose que le taux de sondage est négligeable).

Soit f_n la proportion d'électeurs de l'échantillon votant pour A.

Donner une estimation de p par intervalle de confiance à 95%.

Solution

La proportion inconnue p appartient, avec un niveau de confiance de 95%, à l'intervalle :

$$\left[f_n - 1.96 \sqrt{\frac{f_n(1-f_n)}{n}} ; f_n + 1.96 \sqrt{\frac{f_n(1-f_n)}{n}} \right]$$

3/ Test d'hypothèse

La répartition des sexes à la naissance est-elle équilibrée ?

Sur 53040 naissances consécutives dans la ville de Grenoble, on observe 25946 naissances de filles.

Peut-on accepter l'hypothèse selon laquelle la probabilité p que naisse une fille est $1/2$?

Solution

On suppose les résultats des naissances indépendants les uns des autres. Si l'hypothèse « $p = 1/2$ » est vraie, alors le nombre X de filles sur 53040 naissances suit une loi binomiale de paramètres $(53040; 0.5)$, loi qui peut être approchée par une loi normale de paramètres $(26520; 115.15)$.

On a alors (avec $Z \sim \mathcal{N}(0,1)$) :

$$P(X \leq 25946) = P\left(\frac{X - 26520}{115.15} \leq \frac{25946 - 26520}{115.15}\right) = P(Z \leq -4.98).$$

La probabilité est quasiment nulle. On rejette l'hypothèse de répartition des sexes à la naissance équilibrée et on conclut que l'on a très probablement $p < 1/2$.

3.2. Les erreurs de mesure

Cf. article de 17 pages intitulé "Incertitudes des mesures de grandeur" de C. Robert et J. Treiner dans "Statistique", Commission de Réflexion sur l'Enseignement des Mathématiques, rapport adopté le 15 mars 2003, disponible sur le site de la Société Mathématique de France :

<http://smf.emath.fr/Enseignement/CommissionKahane/RapportStatistiqueProba/Statistique-annexe.pdf> , 53 pages.

3.3. Les données gaussiennes

On pourra se référer à l'annexe intitulée « A propos des données gaussiennes » p. 39 à 41 du document d'accompagnement des programmes de Première L (définition de la "plage de normalité") et aux exercices sur les courbes de croissance proposés dans de nombreux manuels.

4. Historique

Historiquement, la loi normale (ou de Laplace-Gauss) émerge au début du XIX^{ème} siècle :

- d'une part, dans le cadre de la théorie des erreurs de mesure, apparaît le triptyque "critère des moindres carrés", "résumé numérique par la moyenne arithmétique" et "modélisation de la loi des erreurs de mesure par la loi normale",

- d'autre part dans le théorème central limite.

Cf. le chapitre "Grande et petite histoire de la statistique" dans "Enseigner la statistique au lycée : des enjeux aux méthodes" de J.-L. Piednoir et Ph. Dutarte, IREM Paris-Nord, 2001 ou

"La théorie des erreurs (1750-1820), enjeux, problématiques, résultats" de M. Armatte dans "Histoires de probabilités et de statistiques", IREM Histoire des Mathématiques, Ellipses, 2004.

Cf. la planche de Galton (dans tous les ouvrages du secondaire)