Lois normales et autres lois dérivées

1 - Lois normales

a) - Définition

On dit qu'une variable aléatoire réelle X suit la **loi normale** (ou **gaussienne**) de paramètres μ et σ, notée $\mathcal{N}(\mu\,;\sigma)$,

si elle admet pour densité la fonction f définie sur $\mathbf{R}$ par $f(t) = \dfrac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$.

Si $\mu = 0$ et $\sigma = 1$, on dit que X suit la loi normale centrée, réduite.

Remarque

Les lois normales interviennent très souvent et en particulier lorsque le phénomène étudié est la résultante de nombreuses composantes aléatoires (<u>commerce</u> : fluctuation des ventes, <u>industrie</u> : diamètres de pièces usinées qui sont la résultante de la qualité des matières premières, du réglage de la machine, de l'usure de l'outil, de la température...).

b) - Propriétés

Si une variable aléatoire réelle X suit la loi normale de paramètres μ et σ, alors l'espérance mathématique de X est $\mathrm{E}(X) = \mu$. Sa variance est $\mathrm{V}(X) = \sigma^2$ et donc son écart-type est σ.

c) - Théorème 1

Soit X et Y deux variables aléatoires réelles indépendantes suivant les lois normales $\mathcal{N}(\mu_X\,;\sigma_X)$ et $\mathcal{N}(\mu_Y\,;\sigma_Y)$ respectivement et soit a et b deux réels.

$a\,X + b\,Y$ est une variable aléatoire suivant la loi normale $\mathcal{N}\left(a\,\mu_X + b\,\mu_Y\,;\sqrt{a^2\,\sigma_X{}^2 + b^2\,\sigma_Y{}^2}\right)$.

d) - Théorème 2

Si une variable aléatoire réelle X suit la loi normale de paramètres μ et σ, alors $\dfrac{\mathrm{X}-\mu}{\sigma}$ suit la loi normale centrée réduite.

Ce résultat est important car il déduit l'étude des lois normales de celle de la loi normale centrée, réduite.

e) - Théorème 3

Soit X et Y deux variables aléatoires réelles suivant des lois normales. X et Y sont indépendantes si et seulement si la covariance $\mathrm{Cov}(X\,,Y)$ de X et Y est nulle.

Remarque

La condition $\mathrm{Cov}(X\,,Y) = 0$ est nécessaire mais non suffisante dans le cas de variables aléatoires quelconques.

2 - Vecteurs gaussiens

a) - Définition

Soit $n \in \mathbb{N}^*$ et $X = (X_1\,, X_2\,, \ldots\,, X_n)$ un n-uplet de variables aléatoires.

X est un **vecteur aléatoire normal** ou **gaussien** si pour tout n-uplets de réels $(\lambda_1\,, \lambda_2\,, \ldots\,, \lambda_n)$, $\sum_{i=1}^{n} \lambda_i\,X_i$ est une variable aléatoire normale.

b) - Propriété

Soit $X = (X_1, X_2, \ldots, X_n)$ un vecteur aléatoire normal, chacune des X_i est une variable aléatoire normale.

c) - Théorème 4

Soit $X = (X_1, X_2, \ldots, X_n)$ un vecteur aléatoire normal, il est caractérisé par le n-uplet $m(X)$ des espérances des X_i et sa matrice de dispersion $D(X)$ de terme général $\mathrm{Cov}(X_i, X_j)$ pour $1 \le i \le n$ et $1 \le j \le n$.

d) - Théorème 5

Soit X un vecteur aléatoire normal et $\mathcal{A}$ une transformation affine de $\mathbb{R}^n$, alors le vecteur aléatoire $\mathcal{A}(X)$ est normal.

e) - Théorème 6

Soit X un vecteur aléatoire normal d'espérance $m(X) \in \mathbb{R}^n$ et de matrice de dispersion $D(X)$.

Soit φ une application linéaire de $\mathbb{R}^n$ de matrice M, alors le vecteur aléatoire $\varphi(X)$ est normal et a pour espérance $\varphi(m(X)) \in \mathbb{R}^n$ et pour matrice de dispersion $M\,D(X)\,{}^tM$.

f) - Théorème 7 : Indépendance de lois normales

Soit $n \in \mathbb{N}^*$. Pour que n variables aléatoires normales $X_1, X_2, \ldots, X_n$ soient indépendantes, il faut et il suffit que pour $1 \le i < j \le n$, $\mathrm{Cov}(X_i, X_j)$ soit nulle.

Donc il faut et suffit que la matrice de dispersion du vecteur aléatoire $(X_1, X_2, \ldots, X_n)$ soit diagonale.

Remarques

- La condition est toujours nécessaire.
- Un vecteur normal de variables aléatoires, indépendantes, centrées, réduites est caractérisé par un vecteur des espérances nul et une matrice de dispersion égale à une matrice Identité.

3 - Lois du Khi-deux (de Pearson)

a) - Définition

Soit $n \in \mathbb{N}^*$ et $X_1, X_2, \ldots, X_n$ n variables aléatoires indépendantes et identiquement distribuées selon la loi normale centrée réduite.

La variable $Y_n = \sum_{i=1}^{n} X_i^2$ suit la **loi du Khi-deux** à n degrés de liberté.

b) - Propriétés

L'espérance de Y_n est n et sa variance est $2n$.

c) - Théorème 8 de Fisher

Si n variables aléatoires indépendantes suivent des lois de χ^2, alors leur somme suit la loi de χ^2 avec comme nombre de degrés de liberté, la somme des nombres de degrés de liberté des n variables aléatoires.

4 - Lois de Student

a) - Définition

$n \in \mathbb{N}^*$. Soit les variables aléatoires X de loi normale centrée réduite et Y de loi du Khi-deux à n degrés de liberté.

Si X et Y sont indépendantes, $T = \dfrac{X}{\sqrt{\dfrac{Y}{n}}}$ suit la **loi de Student** à n degrés de liberté.

b) - Propriétés

L'espérance de T est 0 et sa variance est $\dfrac{n}{n-2}$ lorsque $n > 2$.

c) - Théorème 9

Soit $n \in \mathbb{N}^*$ et $X_1, X_2, ..., X_n$ n variables aléatoires indépendantes suivant la même loi, normale $\mathcal{N}(\mu, \sigma)$.

Posons $\overline{X} = \dfrac{1}{n}\sum_{k=1}^{n} X_k$ et $\hat{S}^2 = \dfrac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2$, variance corrigée d'échantillon.

alors $\dfrac{(n-1)\,\hat{S}^2}{\sigma^2} = \dfrac{\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2}{\sigma^2}$ suit la loi du Khi-deux à n–1 degré de liberté χ^2_{n-1}.

Démonstration

Posons $Y_k = \dfrac{X_k - \mu}{\sigma}$. Les Y_k sont des variables aléatoires indépendantes distribuées selon la loi normale centrée réduite (th. 2). On pose $\overline{Y} = \dfrac{1}{n}\sum_{k=1}^{n} Y_k$ alors $X_k = \sigma\, Y_k + \mu$ et $\overline{X} = \sigma\, \overline{Y} + \mu$.

On en déduit que $\dfrac{(n-1)\,\hat{S}^2}{\sigma^2} = \dfrac{1}{\sigma^2}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2 = \dfrac{1}{\sigma^2}\sum_{k=1}^{n}\left[(\sigma\, Y_k + \mu) - (\sigma \overline{Y} + \mu)\right]^2 = \sum_{k=1}^{n}\left(Y_k - \overline{Y}\right)^2$.

Or $\sum_{k=1}^{n}\left(Y_k - \overline{Y}\right)^2 = \sum_{k=1}^{n} Y_k^2 - 2\overline{Y}\sum_{k=1}^{n} Y_k + n\overline{Y}^2 = \sum_{k=1}^{n} Y_k^2 - 2\overline{Y}\, n\overline{Y} + n\overline{Y}^2 = \sum_{k=1}^{n} Y_k^2 - n\overline{Y}^2$.

Construisons à partir des Y_k les variables aléatoires gaussiennes $Z_1, Z_2, ..., Z_n$ obtenues par :

$\begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} = P \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$ où $P = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ & & M & \end{pmatrix}$ avec M matrice telle que P soit orthogonale (ses lignes (et ses colonnes) sont orthogonales et de norme 1), alors P est une matrice de changement de bases orthonormales (elle conserve la norme euclidienne) et $P^{-1} = {}^tP$, $P\,{}^tP = I_n$, la matrice identité d'ordre n.

La première ligne de P permet de dire que $Z_1 = \dfrac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_k = \sqrt{n}\,\overline{Y}$.

P étant orthogonale, on a $\sum_{k=1}^{n} Y_k^2 = \sum_{k=1}^{n} Z_k^2$.

Comme de plus les Y_k sont indépendantes et suivent la loi normale centrée réduite, les Z_k suivent la loi normale centrée réduite et sont indépendantes

En effet, la matrice de dispersion des Y_k est I_n, donc celle des Z_k est ${}^tP = P\, I_n\, {}^tP = I_n$ ce qui caractérise un vecteur normal de n variables aléatoires indépendantes de loi normale centrée réduite (th. 6 et 7).

Alors $\sum_{k=1}^{n} Y_k^2 - n\overline{Y}^2 = \sum_{k=1}^{n} Z_k^2 - Z_1^2 = \sum_{k=2}^{n} Z_k^2$ suit la loi du χ^2 à n–1 degrés de liberté en tant que somme de $n-1$ variables aléatoires indépendantes de loi normale centrée réduite, ainsi $\dfrac{(n-1)\,\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1}$.

d) - Théorème 10

Soit $n \in \mathbb{N}^*$ et $X_1, X_2, ..., X_n$ n variables aléatoires indépendantes suivant la même loi, normale $\mathcal{N}(\mu, \sigma)$.

Posons $\overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$ et $\hat{S}^2 = \frac{1}{n-1} \sum_{k=1}^{n} \left(X_k - \overline{X}\right)^2$, variance corrigée d'échantillon.

$\frac{\overline{X} - \mu}{\hat{S}/\sqrt{n}}$ suit la loi de Student à $n-1$ degrés de liberté

Démonstration

D'après la démonstration du théorème 9, $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\,\overline{Y} = Z_1$ et $\frac{(n-1)\,\hat{S}^2}{\sigma^2} = \sum_{k=2}^{n} Z_k^2$.

Comme les Z_k sont indépendantes, $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ et $\frac{(n-1)\,\hat{S}^2}{\sigma^2}$ sont indépendantes, ainsi $\frac{\overline{X} - \mu}{\hat{S}/\sqrt{n}} = \frac{\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{\frac{(n-1)\,\hat{S}^2}{\sigma^2}}{n-1}}}$ suit la loi de Student à $n-1$ degrés de liberté en tant que quotient de deux variables indépendantes, une suivant la loi normale centrée réduite et l'autre étant la racine carrée du quotient d'une variable de loi du Khi-deux à $n-1$ degrés de liberté par son nombre de degrés de liberté.

5 - Lois de Fisher-Snédécor

a) - Définition

n_1 et n_2 désignent des entiers naturels non nuls.

Soit Z_1 et Z_2 deux variables aléatoires indépendantes suivant des lois du Khi-deux à n_1 et n_2 degrés de liberté respectivement, alors $F = \frac{\frac{Z_1}{n_1}}{\frac{Z_2}{n_2}}$ suit la **loi de Fisher-Snédécor** à n_1 et n_2 degrés de liberté

b) - Théorème 11

n_1 et n_2 désignent des entiers naturels non nuls.

Soit $X_1, X_2, ..., X_{n_1}, Y_1, Y_2, ..., Y_{n_2}$ n_1+n_2 variables aléatoires indépendantes distribuées selon les lois normales $\mathcal{N}(\mu_X, \sigma_X)$ pour les X_i $(1 \leq i \leq n_1)$ et $\mathcal{N}(\mu_Y, \sigma_Y)$ pour les Y_j, $(1 \leq j \leq n_2)$ où μ_X et μ_Y sont des réels et σ_X et σ_Y des réels strictement positifs.

$F = \frac{\hat{S}_X^2}{\hat{S}_Y^2} \times \frac{\sigma_Y^2}{\sigma_X^2}$ suit la loi de Fisher-Snédécor à n_1-1 et n_2-1 degrés de liberté où $\hat{S}_X$ et $\hat{S}_Y$ sont les variables aléatoires égales aux écarts-types corrigés des X_i, $1 \leq i \leq n_1$, et des Y_j respectivement, $1 \leq j \leq n_2$.

Démonstration

La démonstration repose sur le résultat du théorème 9.

Exercices

Exercice 1

Sachant que la variable aléatoire X suit la loi normale $\mathcal{N}(0, 1)$, calculer :

a) $P[X < 3{,}1]$
b) $P[X < -2{,}7]$
c) $P[X > 2{,}2]$
d) $P[X > -1{,}5]$
e) $P[X \leq 1{,}79]$
f) $P[X \leq -0{,}54]$
g) $P[X \geq -1{,}55]$
h) $P[0{,}3 < X < 1{,}6]$
i) $P[-2 < X \leq 1{,}1]$
j) $P[-3 < X < 2]$
k) $P[|X| < 2]$
l) $P[|X| \geq 3]$

Exercice 2

Sachant que la variable aléatoire X suit la loi normale $\mathcal{N}(24 ; 6{,}5)$, calculer :

a) $P[X \leq 27]$ **b)** $P[X > 10]$ **c)** $P[18 < X \leq 20]$ **d)** $P[|X| > 27]$.

Exercice 3

La variable aléatoire X suit la loi normale $\mathcal{N}(0, 1)$. Déterminer le nombre réel a tel que :

a) $P[X \leq a] = 0{,}99$ **b)** $P[|X| < a] = 0{,}99$ **c)** $P[X > a] = 0{,}05$ **d)** $P[X < a] = 0{,}1$.

Exercice 4

Une étude de marché a permis d'admettre que le chiffre d'affaires mensuel d'une société est distribué selon la loi normale de moyenne 110 kilo-euros et d'écart-type 25 k€. Le seuil de rentabilité de l'entreprise est fixé à 85 k€. On choisit un mois au hasard.

1°) - Quelle est la probabilité de l'événement : "le chiffre d'affaires de l'entreprise est supérieur ou égal à 85 k€" ?

2°) - Déterminer le réel a tel que la probabilité que le chiffre d'affaires soit situé entre $110 - a$ et $110 + a$ soit 82 %.

Exercice 5

Dans une entreprise qui exploite un parc de 500 taxis, on admet que la distance en km parcourue en une journée par les taxis est distribuée selon la loi normale de moyenne 151 et d'écart-type 15. On choisit un taxi sur une journée au hasard

Déterminer la probabilité de l'événement : "la distance parcourue par le taxi durant la journée est comprise entre 119,5 km et 155,5 km".

Exercice 6

Une entreprise produit des rondelles. On admet que le diamètre en mm des rondelles est distribué selon la loi normale de moyenne 6 et d'écart-type 0,1. On considère qu'une rondelle est défectueuse si son diamètre est inférieur à 5,80 mm ou supérieur à 6,25 mm. On choisit une rondelle au hasard dans la production, quelle est la probabilité de l'événement "la rondelle choisie est défectueuse" ?

Exercice 7

Une entreprise fabrique des roulements à billes dont le diamètre est distribué suivant la loi normale de moyenne 0,614 cm et d'écart-type 0,0025 cm. Déterminer le pourcentage de roulements ayant un diamètre :

1°) - compris entre 0,610 et 0,618 cm ;

2°) - plus grand que 0,617 cm ;

3°) - plus petit que 0,608 cm ;

4°) - égal à 0,615 cm.

Exercice 8

1°) - Pour rendre visite à son ami Alexis, Benjamin va emprunter une navette maritime qui effectue une liaison entre son île de résidence et le continent. La variable aléatoire mesurant la durée du trajet suit la loi normale de moyenne 5 heures 20 minutes et d'écart-type 24 minutes. Quelle est la probabilité que le trajet dure plus de 5 heures ?

2°) - Alexis va chercher son ami au port et prendra sa voiture. La variable aléatoire mesurant la durée du trajet suit la loi normale de moyenne 31 minutes et d'écart-type 12 minutes. Quelle est la probabilité que le trajet dure entre 18 et 35 minutes ?

3°) - Benjamin quitte son île à 12 heures 30 et Alexis quitte son domicile à 17 heures 10.
En supposant que les durées des trajets des deux amis sont indépendantes, quelle est la probabilité qu'Alexis constate à 17 heures 50 que son ami n'est toujours pas arrivé au port ?
