

Loi de probabilité diffuse avec densité continue

Construction générale

Ω est une partie de \mathbb{R} ou de \mathbb{R}^n ; la densité p est une fonction numérique positive, continue par morceaux sur Ω t.q $\int_{\Omega} p = 1$

Alors pour toute partie A de Ω , mesurable on pose :

$$\mathbf{Prob}(A) = \int_A p$$

On constate donc qu'on est obligé de restreindre la tribu F (qui contient les parties pour lesquelles on veut calculer la probabilité) à ne contenir que des parties mesurables, cette restriction théorique ne pose aucun problème en pratique, car dans tous les cas concrets rencontrés, les parties A utiles ont une forme très simple, ce qui assure qu'elles sont mesurables.

Dans toute ce qui suit, cette restriction théorique, est supposée vérifiée.

On remarque que pour passer des probabilités discrètes (déjà vues) aux probabilités diffuses, il suffit formellement de remplacer

$p(x_n)$ par $p(x)$
et

$\sum_{x_n \in A}$ par $\int_{x \in A}$

Exemple : Loi uniforme sur Ω borné

La loi uniforme sur un ensemble borné Ω traduit la notion intuitive de "hasard"

donc on prend une densité constante

$$p = 1/(\text{mesure de } \Omega)$$

pour assurer la condition $\int_{\Omega} p = 1$

On a donc une formule que l'on peut énoncer (de manière analogue au cas discret) $\text{Prob}(A) = \text{mesure des cas favorables} / \text{mesure des cas possibles}$

Remarque

On pourrait aussi envisager des probabilités qui contiennent une partie discrète et une partie diffuse .

w_n étant une suite de points de Ω ,

$p(w_n)$ une suite de réels positifs

et f une fonction positive continue sur Ω

t.q

$$\sum_n p(w_n) + \int_{\Omega} f = 1$$

on définit $\text{Prob}(A)$ par

$$P_{\text{discrète}}(A) + P_{\text{diffuse}}(A)$$

Fonction de répartition d'une loi de probabilité sur une partie de \mathbb{R}

Ω étant une partie de \mathbb{R} , la fonction de répartition est la fonction

$$x \rightarrow F(x) = \text{Prob}(\{w \in \Omega \text{ t.q } w \leq x\})$$

F est une fonction croissante de 0 à 1 (les valeurs 0 et 1 pouvant être des valeurs asymptotiques)

Si la loi de probabilité est diffuse avec densité continue f alors F est une primitive de f .

Si la loi de probabilité contient une partie discrète et une partie diffuse (comme dans la remarque) alors en chaque point w_n , F présente une discontinuité avec un saut de hauteur $p(w_n)$, en dehors de ces points, F est une primitive de f .

Remarque :

L'utilisation de la fonction de répartition F est le moyen pratique pour chercher la densité d'une loi de probabilité.

Variables aléatoires diffuses

Une variable aléatoire sera appelée diffuse (avec densité continue) si sa loi de probabilité est diffuse (avec densité continue) .

Par la suite “avec densité continue” sera sous-entendu et on dira simplement : diffuse

La fonction de répartition F d'une variable aléatoire X à valeurs réelles est la fonction de répartition de sa loi de probabilité

Ainsi

$$\mathbf{F(x) = Prob(X \leq x)}$$

$$= \int_{t < x} f(t) dt$$

pour une loi diffuse de densité f

Loi conjointe d'un couple X,Y de variables aléatoires diffuses

Pour un couple X, Y ; la densité est une fonction de 2 variables : $h(x,y)$

alors $\text{Prob}(X \in A \text{ et } Y \in B) =$

$$\int_{A \times B} h(x,y) \, dx \, dy$$

Ainsi la loi marginale de Y que l'on récupère est

$\text{Prob}(Y \in B) =$

$$\int_B \left(\int_{\Omega} h(x,y) \, dx \right) \, dy$$

Ω étant l'espace des valeurs de X .

La densité de cette loi marginale de Y est donc

$$y \rightarrow \int_{\Omega} h(x,y) \, dx$$

Idem pour la loi marginale de X et généralisation immédiate au cas d'un n-uplet

Cas de variables aléatoires diffuses indépendantes

Propriété : X et Y sont indépendantes ssi la densité $h(x,y)$ de la loi conjointe du couple X , Y est le produit de la densité de la loi marginale de X , par la densité de la loi marginale de Y .

$$\text{Prob}(X \in A \text{ et } Y \in B) =$$

$$\int_{A \times B} h(x,y) \, dx \, dy$$

X et Y sont indépendantes

$$\text{ssi } \text{Prob}(X \in A \text{ et } Y \in B) =$$

$$\text{Prob}(X \in A) \cdot \text{Prob}(Y \in B) =$$

$$\int_A f(x) \, dx \int_B g(y) \, dy =$$

$$\int_{A \times B} f(x) g(y) \, dx \, dy$$

où on a noté f la densité de X ,et g celle de Y

Densité de X conditionnellement à Y = y₀

$h(x,y)$ étant la densité du couple (X,Y)

formule de Bayes

$$\begin{aligned} \text{Prob}(X \in A | Y=y_0) &= \\ \text{Prob}(X \in A \text{ et } Y=y_0) / \text{Prob}(Y=y_0) &= \\ &= 0/0 \end{aligned}$$

Il faut traduire plus finement l'assertion : sachant que $Y=y_0$?

notons I_ε pour $[y_0 - \varepsilon, y_0 + \varepsilon]$

$$\begin{aligned} \text{Prob}(X \in A | Y=y_0) &= \\ \lim_{\varepsilon \rightarrow 0} \text{Prob}(X \in A | Y \in I_\varepsilon) &= \\ \text{Prob}(X \in A \text{ et } Y \in I_\varepsilon) / \text{Prob}(Y \in I_\varepsilon) & \end{aligned}$$

$$= \lim_{\varepsilon \rightarrow 0}$$

$$\frac{\iint_{y \in I_\varepsilon, x \in A} h(x,y) dx dy}{\iint_{y \in I_\varepsilon, x} h(x,y) dx dy} =$$

$$\frac{2\varepsilon \int_{x \in A} h(x, y_0) dx}{2\varepsilon \int_x h(x, y_0) dx} =$$

$$\int_{x \in A} h(x, y_0) dx$$

$$\int_{x \in A} h(x, y_0) dx$$

$$\int_{x \in A} h(x, y_0) dx$$

donc la densité de X
conditionnellement à $Y = y_0$
est : $k h(x, y_0)$ où k est le coefficient de normalisation

Espérance mathématique d'une variable aléatoire diffuse

Dans le cas d'une variables aléatoire diffuse X , de densité f ; l'Espérance mathématique de X est définie par :

$$E(X) = \int x f(x) dx$$

où naturellement on suppose que l'intégrale est absolument convergente
cette hypothèse qui s'énonce simplement : $E(X)$ existe ,
sera toujours sous entendue dans la suite .

Formule de transfert

X est une variable aléatoire diffuse (de densité f) à valeurs dans Ω , et φ étant une fonction injective par morceau sur Ω on considère la variable aléatoire $\varphi(X)$

formule de transfert

$$\mathbf{E}(\varphi(\mathbf{X})) = \int \varphi(\mathbf{x}) \mathbf{f}(\mathbf{x}) \, d\mathbf{x}$$

Propriétés

De manière analogue au cas discret on obtient :

La **Linéarité** de E

et

Si X et Y sont indépendantes alors

$$E(XY) = E(X) E(Y)$$

Variance, Covariance, Coefficient de corrélation

Les notions de Variance , Covariance , Coefficient de corrélation, sont définies directement avec les mêmes formules que dans le cas discret, à partir de l'Espérance (qui a les mêmes propriétés que dans le cas discret)

On a les mêmes propriétés et les mêmes interprétations heuristiques.

$$\mathbf{Var(X)} = \mathbf{E(X^2)} - (\mathbf{E(X)})^2 = \mathbf{E((X - E(X))^2)}$$

$$\sigma(X) = \sigma_X = (\mathbf{Var(X)})^{1/2} .$$

$$\mathbf{C_{X,Y}} = \mathbf{E(XY)} - \mathbf{E(X) E(Y)} = \mathbf{E((X - E(X)) (Y - E(Y)))}$$

$$\rho_{X,Y} = \mathbf{C_{X,Y}} / (\sigma_X \cdot \sigma_Y)$$

$$\mathbf{V^*} = \mathbf{1/(N-1)} \sum_i (\mathbf{X_i} - \bar{\mathbf{X}})^2$$

$$\mathbf{V^*} = \mathbf{N/(N-1)} \overline{(\mathbf{X} - \bar{\mathbf{X}})^2} = \mathbf{N/(N-1)} (\overline{\mathbf{X}^2} - \bar{\mathbf{X}}^2)$$

$$\mathbf{E(V^*)} = \mathbf{Var(X)}$$

Inégalité de Tchebycheff

Théorème : $b > 0$ et Y une variable aléatoire (t.q $E(Y^2)$ existe) ; on note σ son écart type alors
 $\text{Prob}(|Y - E(Y)| > b \sigma) \leq 1/b^2$

Théorème des grands nombres :

Soit X une variable aléatoire

t.q $E(X)$ existe,

X_1, X_2, \dots, X_N désigne un échantillon de X de taille N

Alors

$\forall \varepsilon > 0 \text{ Prob}(|\bar{X} - E(X)| > \varepsilon)$

converge vers 0 lorsque N tend vers ∞ .

$\bar{X} \xrightarrow{cv_prob} E(X)$

De même :

$V^* \xrightarrow{cv_prob} \text{Var}(X)$

Y étant une variable aléatoire, d'espérance $= E(Y)$
souvent notée aussi m , et d'écart type σ

$$X = (Y - m) / \sigma$$

est dite variable aléatoire réduite
associée à Y ,

elle est telle que

$$E(X) = 0 \text{ et } \sigma(X) = 1$$

Exemples de variables aléatoires diffuses

Loi exponentielle

On rencontre cette loi dans les problèmes de durée de vie .

Si un appareil est mis en marche à l'instant t_0 , X désigne l'instant où il tombe en panne .

Cherchons la loi de X :

Notons F la fonction de répartition ,

$$F(t) = \text{Prob}(X \leq t) .$$

$$\text{Prob}(X \leq t+dt \mid X > t) =$$

$$a(t) dt + dt \varepsilon(dt) \quad (\text{développement de Taylor})$$

$$= \text{Prob}(X \leq t+dt \cap X > t) / \text{Prob}(X > t)$$

$$(F(t+dt) - F(t)) / (1 - F(t))$$

d'où

$$(F(t+dt) - F(t)) / dt =$$

$$(a(t) + \varepsilon(t)) \cdot (1 - F(t))$$

on suppose une hypothèse de non vieillissement : $a(t) = a$

d'où, en faisant tendre dt vers 0 on obtient :

$$F'(t) = a (1 - F(t))$$

en résolvant cette équation différentielle

on obtient

$$F(t) = 1 + c \exp(-at)$$

et en considérant la condition initiale

$$F(t_0) = 0 \quad \text{on trouve que}$$

$$c = - \exp(at_0)$$

d'où le résultat

$$F(t) = 1 - \exp(-a(t-t_0))$$

pour $t > t_0$

donc la densité est

$$f(t) = a \exp(-a(t-t_0))$$

pour $t > t_0$

on dit qu'il s'agit de la loi exponentielle de paramètre a et de début t_0 .

et on calcule que :

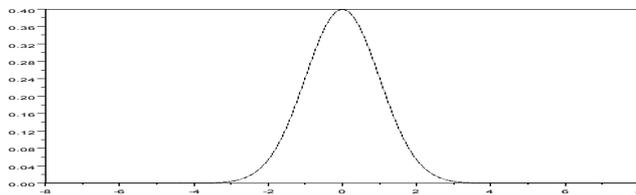
$$E(X) = 1/a + t_0 ;$$

$$\text{Var}(X) = (1/a)^2$$

Loi de Gauss ou Loi Normale

C'est la Loi la plus importante pour les applications Statistiques .

En étudiant de nombreuses variables aléatoires réduites, on s'aperçoit que leur densité a souvent la forme indiquée par le dessin suivant :



il s'agit donc de trouver une fonction densité qui ait cette forme et telle que

l'espérance déduite = 0

et la variance =1

c'est la fonction

$$f_{0,1}(x) = (2\pi)^{-1/2} \exp(-1/2 x^2)$$

qui convient, on l'appelle densité de GAUSS ,
on dit que X suit la loi de GAUSS réduite
encore appelée GAUSS(0,1).

Si X suit GAUSS(0,1)

alors on dit que

$Y = m + \sigma X$ suit GAUSS(m,σ) pour $\sigma > 0$

$E(Y) = m$ et écart type(Y) = σ

Si X suit la loi GAUSS(0,1)

la variable aléatoire $Y = m + \sigma X$

(où $\sigma > 0$)

suit la loi de densité

$$f(y) = 1/\sigma f_{0,1}((y-m)/\sigma)$$

c'est la densité de la loi GAUSS(m , σ)

Conséquence pratique : pour faire des calculs de probabilité concernant Y qui suit la loi Gauss(m, σ)

on pose $X = (Y - m) / \sigma$

X suit la loi Gauss(0,1) .

MAIS $F(x) = \text{Prob}(X \leq x)$ ne peut se calculer analytiquement et nécessite un calcul numérique, les résultats sont consignés dans des tables .

Voici quelques résultats importants en statistiques

$$\text{Pob}(X < 1.65) = 0.95$$

$$\text{Pob}(X > 1.65) = 0.05$$

$$\text{Pob}(X < -1.65) = 0.05$$

$$\text{Pob}(X > -1.65) = 0.95$$

$$\text{Pob}(|X| < 1.96) = 0.95$$

$$\text{Pob}(|X| > 1.96) = 0.05$$

Théorème : Si Y_1 suit la loi Gauss(m_1, σ_1) et Y_2 suit la loi Gauss(m_2, σ_2) et si Y_1 et Y_2 sont indépendantes alors $Y_1 + Y_2$ suit une loi Gauss(m, σ)

avec $m = m_1 + m_2$ et $\sigma^2 = \sigma_1^2 + \sigma_2^2$

Il s'ensuit que toute combinaison linéaire de variables aléatoires indépendantes qui suivent des lois de Gauss, suit une loi de Gauss

Cas particulier important :

soit X_1, X_2, \dots, X_N un échantillon de X qui suit la Gauss(m, σ)

alors

\bar{X} suit la loi Gauss($m, \sigma/\sqrt{N}$)

N uplet Gaussien :

X_1, X_2, \dots, X_N étant des variables aléatoires de Gauss(
 m_i, σ_i)

si elles sont indépendantes la loi conjointe du N-uplet
est

$$h(x_1, \dots, x_n) = \prod_i f_{m_i, \sigma_i}(x_i)$$

Mais si on doit traduire une certaine dépendance, celle
ci est caractérisée par des coefficients de corrélation

$-1 < r_{i,j} < 1$ que l'on doit prendre à priori.

avec $r_{i,i} = 1$ pour tout i

on appelle matrice des covariances, la matrice C dont
les coefficients sont

$$C_{i,j} = r_{i,j} \sigma_i \sigma_j$$

la problématique est alors de trouver une fonction
densité $h(x_1, \dots, x_n)$ telle que :

les loi marginales qui s'en déduisent sont les lois de
Gauss(m_i, σ_i)

et les coefficients de corrélations qui s'en déduisent
sont les $r_{i,j}$ pris à priori.

On note x_m
la matrice unicolonne
de composantes $x_i - m_i$;
 x_m^t désigne la matrice transposée

$$h(x_1, \dots, x_n) = \\ (2\pi)^{-N/2} \det(C)^{-1/2} * \\ \exp(-1/2 (x_m^t C^{-1} x_m))$$

est la densité cherchée

explicitons la formule de h, en dimension 2 :

$$h(x_1, x_2) = \frac{1}{(2\pi\sigma_1\sigma_2(1-r^2)^{1/2})} \cdot \exp \left[-\frac{1}{2} (1-r^2)^{-1} \left\{ \left(\frac{x_1-m_1}{\sigma_1} \right)^2 - 2r \frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-m_2}{\sigma_2} \right)^2 \right\} \right]$$

on remarque que dans ce cas, si r le coefficient de corrélation entre X_1 et X_2 est nul, alors la densité h peut s'écrire comme le produit d'une fonction de x_1 et d'une fonction de x_2 , ce qui montre que les variables aléatoires X_1 et X_2 sont indépendantes .

densité conditionnelle

(X,Y) étant un couple Gaussien de densité h, la densité conditionnelle de

X sachant que $Y = y_0$

est $h(x, y_0) = \text{coef} *$

$$\exp\left[-\frac{1}{2}(1-r^2)^{-1} \left\{ \left(\frac{x-m_x}{\sigma_x}\right)^2 - 2r\frac{(x-m_x)(y_0-m_y)}{\sigma_x\sigma_y} \right\}\right]$$

posons $\sigma = (1-r^2)^{1/2} \sigma_x$ et

$$m = m_x + r (y_0 - m_y) \sigma_x / \sigma_y$$

alors

$$h(x, y_0) = \text{coef} * \exp\left[-\frac{1}{2} \sigma^{-2} (x - m)^2\right]$$

c'est la densité de la loi Gauss(m, σ)

avec $\sigma = (1 - r^2)^{1/2} \sigma_x$

et $m = m_x + r (y_0 - m_y) \sigma_x / \sigma_y$

cohérence : faire $r=0$ puis $r=1$

Théorème Central Limite :

X étant une variable aléatoire diffuse de densité f telle que $E(X^3)$ existe et $E(X) = 0$

et X_1, X_2, \dots, X_N un échantillon de modèle X , de taille N on considère la suite (indicée par N) de variables aléatoires

$$Y_N = 1/N^{1/2} \sum_{k=1}^{k=N} X_k .$$

alors la suite des fonctions de répartition des Y_N converge simplement vers la fonction de répartition de la loi Gauss(0, σ_X)

ce qui signifie que pour N suffisamment grand, dans les calculs de probabilités, on pourra remplacer la loi de Y_N par la loi Gauss(0, σ_X)

on notera symboliquement ceci

$$Y_N \sim \text{Gauss}(0, \sigma_X)$$

Extension de ce type d'approximation :

Si une variable aléatoire Y se présente sous la forme :

$$Y = \sum_{k=1}^{k=N} X_k$$

avec N suffisamment grand (X étant non nécessairement centrée, $E(X)=m$) alors d'après ce qui a été dit précédemment :

$$Z = 1/N^{1/2} \left(Y - \sum_{k=1}^{k=N} E(X_k) \right) =$$

$$1/N^{1/2} \sum_{k=1}^{k=N} (X_k - E(X_k))$$

$\sim \text{Gauss}(0, \sigma)$

$$Y = Nm + N^{1/2} Z$$

d'où $Y \sim \text{Gauss}(Nm, N^{1/2} \sigma)$

L'utilisation heuristique (en statistiques) de ce type d'approximation, fait que l'on cherchera une approximation de la loi d'une variable aléatoire Y (qui est la somme d'un grand nombre de variables aléatoires indépendantes) sous la forme d'une loi de Gauss .

Comment vous expliquez vous, la remarque du début ?

Exemples :

$$\bar{X} \sim \text{Gauss}$$

$$N^{1/2} (\bar{X} - E(X)) / \sigma \sim \text{Gauss}(0,1)$$

$$N^{1/2} (\bar{X} - E(X)) / \sigma^* \sim \text{Gauss}(0, 1)$$

on rappelle que $\sigma^* = (V^*)^{1/2}$

$$\text{avec } V^* = 1/(N-1) \sum_i (X_i - \bar{X})^2$$

Loi du Chi-deux

X_1, X_2, \dots, X_n étant un échantillon de X qui suit la loi Gauss(0,1)

Considérons la variable aléatoire

$$U_n = \sum_{k=1}^{k=n} X_k^2$$

On dit que U_n suit la loi du Chi-deux à n degrés de liberté.

$$E(U_n) = n$$
$$\text{et } \text{Var}(U_n) = 2n$$

d'après le Théorème central limite :

$$U_n \sim \text{Gauss}(n, (2n)^{1/2})$$

Ce qui explique, que les tables du Chi-deux s'arrêtent à 30 degrés de liberté.

Pour $n \leq 30$ les calculs de probabilités concernant U_n nécessitent un calcul d'intégrale qui ne peut se faire analytiquement, cependant comme il s'agit d'un calcul d'intégrale en dimension n , on ne peut non plus, procéder directement avec un calcul numérique, on commence par faire un changement de variables qui

permet de se ramener au calcul d'une intégrale en dimension 1.

Principe du calcul

la fonction de répartition

$$F(u) = \text{Prob}(U_n \leq u) =$$

$$\frac{1}{(2\pi)^n} \int_A \exp(-1/2\|X\|^2) dx_1 \dots dx_n$$

où A désigne la partie

$$\{X = x_1, \dots, x_n \text{ t.q. } \|X\|^2 \leq u\}$$

on effectue un passage en coordonnées sphériques (en dimension n)

$$x_1 = r \cos(\theta_1) \cos(\theta_2) \dots \cos(\theta_{n-2}) \cos(\theta_{n-1})$$

$$x_2 = r \cos(\theta_1) \cos(\theta_2) \dots \cos(\theta_{n-2}) \sin(\theta_{n-1})$$

$$x_3 = r \cos(\theta_1) \cos(\theta_2) \dots \sin(\theta_{n-2})$$

.....

.....

$$x_{n-1} = r \cos(\theta_1) \sin(\theta_2)$$

$$x_n = r \sin(\theta_1)$$

avec r positif et

$$\theta_i \in [-\pi/2, \pi/2[\text{ pour } i=1,2, \dots, n-2$$

$$\text{et } \theta_{n-1} \in [0, 2\pi[$$

alors

$$|\det(\text{Jacobiennes})| =$$

$$r^{n-1} \cos^{n-2}(\theta_1) \cos^{n-3}(\theta_2) \dots \cos(\theta_{n-2})$$

le coefficient $1/(2\pi)^n$ avec les intégrations angulaires

donnent $C_n = 1/(2^m (m-1)!)$ pour $n = 2m$ et

$$C_n = 2/(2\pi)^{1/2} 3.5.7 \dots n-2$$

pour $n = 2m + 1$

il reste alors à calculer numériquement l'intégrale

$$\int_A r^{n-1} \exp(-1/2 r^2) dr$$

où A est le segment $(0, u^{1/2})$

Loi de Student

$X_0, X_1, X_2, \dots, X_n$ étant un échantillon de X qui suit la loi Gauss(0,1)

Considérons la variable aléatoire

$$St_n = X_0 / \left(\frac{1}{n} \sum_{k=1}^{k=n} X_k^2 \right)^{1/2}$$

On dit que St_n suit la loi de Student à n degrés de liberté .

Comme dans le cas de la loi du Chi-deux ; les calculs donnant la fonction de répartition donnent lieu à une intégrale , en dimension multiple , un calcul préliminaire (utilisant aussi , le même passage en coordonnées sphériques) permet de se ramener en dimension 1 , l'intégrale monodimensionnelle ainsi obtenue est alors numériquement calculée ; les résultats sont consignés dans des tables

Comme dans le cas de la loi du Chi-deux , les tables s'arrêtent à partir de $n = 30$.

Pour $n > 30$ on utilise le principe d'approximation issu du théorème central limite

$$St_n \sim \text{Gauss}(0,1)$$

Car $V^* = \text{cv_prob} \text{ Var}(X) = 1$

$$\text{et } \left(\frac{1}{n} \sum_{k=1}^{k=n} X_k^2 \right) = (n-1)/n V^*$$

donc $St_n \sim X_0$