Supposons un générateur de variables aléatoires iid uniformément distribuées sur [0 ; 1], Comment utiliser ces variables uniformes pour générer des variables aléatoires de distributions différentes ?

THE
NORMAL
LAW OF ERROR
STANDS OUT IN THE
EXPERIENCE OF MANKIND
AS ONE OF THE BROADEST
GENERALIZATIONS OF NATURAL
PHILOSOPHY ? IT SERVES AS THE
GUIDING INSTRUMENT IN RESEARCHES
IN THE PHYSICAL AND SOCIAL SCIENCES AND
IN MEDICINE AGRICULTURE AND ENGINEERING ?
IT IS AN INDISPENSABLE TOOL FOR THE ANALYSIS AND THE
INTERPRETATION OF THE BASIC DATA OBTAINED BY OBSERVATION AND EXPERIMENT
William Youden (1962)

Sommaire.

Simulation par la méthode d'inversion

Soit X est une variable alatoire avec une distribution uniforme dans $[0, 1]$, $F(x)$ sa fonction de répartition :

$$F(x) \ := \ P(X \leq x).$$

Rappels :

- F est croissante, continue à droite et limitée à gauche en tout point de $\mathbb{R}$.
- L'ensemble de ses discontinuités est au plus dénombrable.
- F a pour limite 0 en $-\infty$ et 1 en $+\infty$.

Dans le cas particulier où F est continue et strictement croissante sur tout $\mathbb{R}$, elle réalise une bijection de $\mathbb{R}$ sur $]0, 1[$ et admet donc une inverse

$$F^{-1} \ :]0, 1[\to \mathbb{R}$$

Si U est une variable aléatoire de loi uniforme sur $]0, 1[$, alors $Y \ : = F^{-1}(U)$ a même loi que X.
Preuve : calcul de la fonction de répartition de Y :

$$\forall x \in \mathbb{R}, \ P(Y \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x).$$

Dans cette suite d'égalités, la deuxième repose sur la croissance de F et de F^{-1} qui légitime l'équivalence $F^{-1}(U) \leq x \iff U \leq F(x)$. La dernière égalité est due au calcul de la fonction de répartition de la loi uniforme sur $]0, 1[$ (qui coincide sur $[0, 1]$ avec l'identité) et au fait que $F(x) \in [0, 1]$.

Simulation par la méthode d'inversion

Définition

La fonction inverse $x = F^{-1}(y)$ est définie pour toute valeur de y, $y \in]0,1[$ telle que $F^{-1}(y)$ soit la plus petite valeur de x satisfaisant $F(x) \geq y$. C'est-à-dire :

$$\forall y \in]0,1[\qquad F^{-1}(y) = inf\{x \in \mathbb{R};\ F_X(x) \geq y\}.$$

- Proposition

Si U est une v.a.r. de loi uniforme sur $]0,1[$ et X une v. a. r. de fonction de répartition F alors $F^{-1}(U)$ a la même loi que X, c'est-à-dire admet F pour fonction de répartition.

Pratiquement

Pour obtenir la distribution x d'une variable aléatoire X, on génére une suite de valeurs u de la variable aléatoire U distribuées sur $[0,1]$, on calcule $x = F^{-1}(u)$.

Lois uniformes

on dispose d'un générateur de nombres aléatoires dans l'intervalle $[0,\ 1]$.

- Distribution uniforme sur $[a,\ b]$: $\mathcal{U}([a,b])$

Une variable aléatoire X suit une loi **rectangulaire** (loi uniforme sur un segment $[a,b]$) si étant donné 2 nombres a et b, sa densité de probabilité est :

$$\begin{cases} f(x) = \frac{1}{b-a} & si\ a \leq x \leq b \\ f(x) = 0 & si\ x < a\ ou\ x > b \end{cases} . \quad \text{où} \quad f(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$$

$$F \quad \text{fonction de répartition} \begin{cases} F(x) = 0 & si\ x \leq a \\ F(x) = \frac{x-a}{b-a} & si\ x \in [a,b[\\ F(x) = 1 & si\ x \in [b,\infty[\end{cases} .$$

On peut générer une suite de telles variables à partir d'une suite de variables ayant une loi continue uniforme par la combinaison d'une homothétie de rapport $(b-a)$ et d'une translation d'amplitude a :

$$\mathbf{F^{-1}(u) = y = (b-a)u + a}.$$

Cette suite de nombres représentera une réalisation d'une suite de variables aléatoires $X_1,\ X_2,\ X_3,\ ....,\ X_N$ indépendantes et de loi uniforme sur $[a,\ b]$.

Loi exponentielle

- Distribution exponentielle : $\mathcal{E}(x)$,

Une variable aléatoire X suit une loi **exponentielle** de paramètre λ ($\lambda > 0$) si sa densité de probabilité est :

$$\mathbf{f(x) = \lambda e^{-\lambda x} \, 1\!\mathrm{I}_{\{x>0\}}(x).}$$

Sa fonction de répartition est

$$F(x) = 1 - e^{-\lambda x} \, 1\!\mathrm{I}_{\{x>0\}}(x) \qquad [E(X) = \frac{1}{\lambda}, \; V(X) = \frac{1}{\lambda^2}].$$

On peut générer une suite de telles variables à partir d'une suite de variables ayant une loi continue uniforme sur $[0, 1]$ par la résolution de l'équation :

$$\int_0^{y_i} \lambda e^{-\lambda t} \, dt = u.$$

Soit si u est une variable obtenue à partir d'une variable aléatoire de loi uniforme, on obtient donc une variable y de loi exponentielle par :

$$\mathbf{F^{-1}(u) = y = -\frac{1}{\lambda} \ln(1 - u).}$$

On peut remplacer $1 - u$ par u puisqu'ils suivent la même distribution et générer le nombre $y = -\frac{1}{\lambda} lnu$.

Loi de Weibull

- Distribution de Weibull : $\mathcal{W}(\alpha,\ \beta)$,

Une variable aléatoire X suit une loi de **Weibull** de paramètre $\alpha,\ \beta$ ($\alpha\ >\ 0\ \beta\ >\ 0$) si sa densité de probabilité est :

$$\mathbf{f}(\mathbf{x}|\alpha;\beta)\ =\ \alpha\,\beta\,\mathbf{x}^{\alpha-1}\ \mathbf{e}^{-\beta\mathbf{x}^{\alpha}}\ \mathbb{1}_{\mathbb{R}+}(\mathbf{x}).$$

Sa fonction de répartition :

$$\alpha\ >\ 0\ \beta\ >\ 0,\qquad F(x) = 1 - e^{-\beta x^{\alpha}}\ \mathbb{1}_{\mathbb{R}+}(x).$$

Soit si u est une variable obtenue à partir d'une variable aléatoire de loi uniforme, on obtient donc une variable y de loi de Weibull par :

$$\mathbf{F}^{-1}(\mathbf{u})\ =\ \mathbf{y}\ =\ \left(-\frac{1}{\beta}\ln(\mathbf{u})\right)^{\frac{1}{\alpha}}.$$

Loi de Cauchy

- Distribution de cauchy : $\mathcal{C}(0, 1)$,

$$F(x) = \frac{1}{n} arctan(x) + \frac{1}{2n} \quad \text{and} \quad \mathbf{F_X^{-1}(u) = tan(\pi(u - 1/2))}$$

Malheureusement, pour de nombreuses fonctions de distribution, nous ne disposons pas d'une expression facile utiliser pour l'inverse de F.

Sommaire.

Principe

- La méthode de rejet peut s'utiliser pour la génération de n'importe quel type de variable aléatoire.

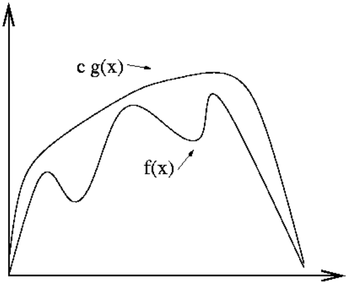
- Elle consiste à générer des données suivant une distribution de fonction de densité proche de celle désirée et ensuite d'éliminer une certaine proportion de ces données de manière à se ramener à des données qui suivent la distribution attendue.

- Soit $f(x)$ la densité de la V.A. à générer. On suppose que l'on sait simuler un vecteur aléatoire X de densité g et trouver une constante c $(c \geq 1)$ telle que $f(x) \ \leq \ c\ g(x)$ pour toute valeur de X.

- On suppose que $g(X)$ est telle que la fonction de répartition associée $G(x)$ est analytiquement continue ainsi que son inverse $G^{-1}(x)$.

Algorithme

1. Générer U_1 de distribution $\mathcal{U}[0, 1]$,
2. Calculer $X = G^{-1}(U_1)$
3. Générer U_2 de distribution $\mathcal{U}[0, 1]$,
4. si $c\ g(X)\ U_2\ \leq\ f(X)$ alors on garde X comme donnée générée sinon on la "jette" et on recommence l'étape 1.



Sur la figure, on tire tout d'abord un X de distribution $g(x)$. On tire ensuite une variable uniforme U_2 entre 0 et 1 qui va permettre de décider si on garde le x tiré. La probabilité de le garder dépend de la distance entre $f(x)$ et $cg(x)$ au point X. Plus la distance est grande, plus la probabilité de rejeter la donnée est grande. En fait tous les points tirés tels que $cg(X)\ U_2\ \geq\ f(X)$ sont rejetés, ce qui équivaut à corriger l'erreur d'approximation de f(x) par $cg(x)$.

Probabilité d'acceptation :

$$\begin{aligned}
P[X\text{accepté}] &= P\left[U \leq \frac{f(X)}{c\,g(X)}\right] \;=\; E\left[\mathbf{I}_{\{\mathbf{U}\leq \frac{\mathbf{f(X)}}{\mathbf{c\,g(X)}}\}}\right] \\
&= E\left[E\left[\mathbf{I}_{\{\mathbf{U}\leq \frac{\mathbf{f(X)}}{\mathbf{c\,g(X)}}\}}\right]|\mathbf{X}\right] \\
&= E\left[\frac{f(X)}{c\,g(X)}\right] \\
&= \int \frac{f(x)}{c\,g(x)}\; g(x)dx = \frac{1}{c}.
\end{aligned}$$

Loi de X :

$$\begin{aligned}
P[X < x | X\text{accepté}] &= \frac{P[X < x, X\text{accepté}]}{1/c} \\
&= c\;P\left[X < x, U < \frac{f(X)}{c\,g(X)}\right] \\
&= c\;E\left[\mathbf{I}_{\{\mathbf{X<x},\,\mathbf{U}\leq \frac{\mathbf{f(X)}}{\mathbf{c\,g(X)}}\}}\right] \\
&= c\;E\left[E\left[\mathbf{I}_{\{\mathbf{X<x},\,\mathbf{U}\leq \frac{\mathbf{f(X)}}{\mathbf{c\,g(X)}}\}}\right]|\mathbf{X}\right] \\
&= c\;E\left[I_{\{X<x\}}\frac{f(X)}{c\,g(X)}\right]
\end{aligned}$$

Sommaire.

Rappel : Distribution normale

La fonction densité de probabilité d'une distribution normale est :

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ \mathbb{1}_{[-\infty,\ +\infty]}(x)$$

- μ Espérance mathématique de la distribution (médiane et mode),
- σ^2 variance, σ écart-type.

La fonction cumulative (répartition) (CDF) de la distribution normale standard $F(x)$ est :

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt.$$

Cas particulier de la loi centrée réduite $\mu = 0$ and $\sigma = 1$ de PDF :

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \quad e^{-\frac{1}{2}x^2}$$

La fonction cumulative CDF de la loi normale, notée généralement Φ est l' intégrale :

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

Fonction cumulative (CDF)

En statistique, on utilise souvent la fonction d'erreur, ou erf(x), définie comme la probabilité qu'une variable alátoire à distribution normale de moyenne 0 et de variance 1/2 se situant dans l'intervalle $[-x,\ x]$; telle que

$$\mathrm{erf}(x) \;=\; \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2}\, dt$$

Les deux fonctions ϕ et erf sont étroitement liées, à savoir

$$\Phi(x) \;=\; \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$$

Pour une distribution normale générique f avec une moyenne μ et un écart-type σ, la fonction de distribution cumulée est la suivante

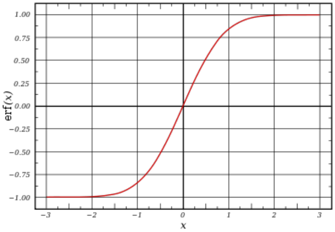
$$F(x) \;=\; \Phi\left(\frac{x-\mu}{\sigma}\right) \;=\; \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

Méthode inverse

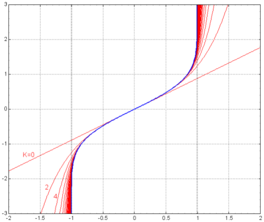
Hypothèse : on dispose d'un générateur de nombres aléatoires dans l'intervalle $[0,\ 1]$.

Remarque : La fonction de répartition de la variable qui utilise la loi normale $\mathcal{N}(0,1)$ peut se calculer avec la fonction **erf** sa fonction inverse avec erf^{-1} :

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = \frac{1}{2}\left(1 + erf\left(\frac{x}{\sqrt{2}}\right)\right) \qquad \blacksquare^{-1}(\mathbf{u}) = \mathbf{y} = \sqrt{\mathbf{2}}\, \mathbf{erf}^{-1}\left(\mathbf{2u} - \mathbf{1}\right).$$



Fonction erreur



Fonction erreur réciproque (source : wikipedia)

Théorème Central Limite

Une application immédiate du théorème central limite donne une méthode très simple de génération de variables aléatoires normales.

• On sait que si $X_1, X_2, ...X_n$ sont n variables aléatoires i.i.d. d'espérance μ et de variance σ^2, en notant S leur somme, on a :

$$\frac{S - n\,\mu}{\sqrt{n}\,\sigma} \overset{(loi)}{\rightarrow} \mathcal{N}(0,1).$$

La variable somme centrée réduite tend vers une loi normale centrée réduite lorsque $n \rightarrow \infty$. En particulier si les variables X_i ont une loi uniforme, on a vu que l'on avait $E(X_i) = \mu = 1/2$ et $\sigma_{X_i} = 1/\sqrt{12}$.

• Partant de n variables aléatoires X_i de loi uniforme $\mathcal{U}_{[0,\ 1]}$, on peut générer une variable aléatoire Y ayant une loi normale de valeur moyenne $\overline{Y}$ et d'écart-type σ_Y par la relation :

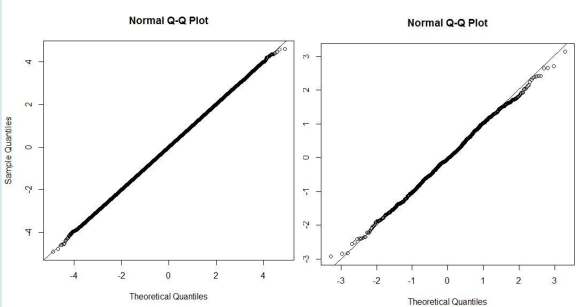
$$Y = \frac{\sum_{i=1}^{n}(X_i) - n/2}{\sqrt{n/12}}$$

Pratiquement, on simule 12 variables aléatoires indépendantes uniformes sur $[0,\ 1]$ et on en fait la somme. La variance de cette somme vaut 1 et il faut ensuite la centrer en lui retranchant son espérance, c'est à dire 6 :

$$X = \sum_{i=1}^{12} X_i \ - \ 6.$$

La loi de cette variable est très proche de la loi normale $\mathcal{N}(0,\ 1)$.
Application :

!!!! cette méthode est une approximation d'autant plus mauvaise qu'on s'éloigne de la moyenne. Multiplier le nombre de nombres à sommer améliore la qualité de l'approximation, mais naturellement au détriment de la vitesse. Le choix de 12 nombres n'est, en fait, qu'un joli compromis.



Méthode de Box et Muller

- 2n-échantillon de la loi $\mathcal{N}(0,1)$, à partir d'un générateur de la loi uniforme.

Si U_1 et U_2 indépendantes, de loi uniforme sur $]0,1[$, le couple

$$Y_0 = \sqrt{-2\ln U_1}\cos(2\pi U_2)\ ;\ Y_1 = \sqrt{-2\ln U_1}\sin(2\pi U_2).$$

est formé de variables gaussiennes indépendantes centrées de variance 1.
La simulation de deux variables indépendantes de loi uniforme conduit à la simulation de deux variables aléatoires gaussiennes standard indépendantes.

- Pour simuler une V. A. de moyenne μ et de variance σ^2, on pose

$$X = \mu + \sigma\ Y$$

où Y suit une loi normale centrée réduite.

- Preuve

A partir des variables aléatoires Y_1 et Y_2, les quantités y_1, y_2 et x_1, x_2 s'écrivent indifféremment

$$\left\{\begin{array}{lcl} y_1 & = & (-2Lnx_1)^{\frac{1}{2}}\ cos(2\pi x_2) \\ y_2 & = & (-2Lnx_1)^{\frac{1}{2}}\ sin(2\pi x_2) \end{array}\right. \qquad \left\{\begin{array}{lcl} x_1 & = & exp(-\frac{1}{2}\ (y_1^2 + y_2^2)) \\ x_2 & = & \frac{1}{2\pi}\ arctang(\frac{y_2}{y_1}) \end{array}\right.$$

Changement de variables $\Rightarrow$ calcul du Jacobien :

$$\iint\limits_V g(x_1,x_2)\ \mathrm{d}x_1\ \mathrm{d}x_2 = \iint\limits_U g\left(F\left(y_1,y_2\right)\right)\left|\det J_F(y_1,y_2)\right|\ \mathrm{d}y_1\ \mathrm{d}y_2$$

on montre que

$$J(y_1,y_2) \;=\; \left|\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)}\right| \;=\; \left|\begin{array}{cc} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{array}\right| \left(= \; -[\frac{1}{\sqrt{2\pi}}\; e^{-\frac{1}{2}y_1^2}][\frac{1}{\sqrt{2\pi}}\; e^{-\frac{1}{2}y_2^2}]\right).$$

$$J(y_1,y_2) = \left|\begin{array}{cc} -y_1\; exp(-\frac{1}{2}\;(y_1^2+y_2^2)) & -y_2\; exp(-\frac{1}{2}\;(y_1^2+y_2^2)) \\ -\;\frac{1}{2\pi}\;\frac{y_2}{y_1^2}\;\frac{y_1^2}{y_1^2+y_2^2} & \frac{1}{2\pi}\;\frac{1}{y_1}\;\frac{y_1^2}{y_1^2+y_2^2} \end{array}\right|$$

$$J(y_1,y_2) = -\frac{1}{2\pi}\; exp(-\frac{1}{2}\;(y_1^2+y_2^2))\;\frac{y_1^2}{y_1^2+y_2^2} \;-\; (\frac{1}{2\pi})\; exp(-\frac{1}{2}\;(y_1^2+y_2^2))\;(y_2\;\frac{y_2}{y_1^2})\;\frac{y_1^2}{y_1^2+y_2^2}$$

$$\begin{aligned} J(y_1,y_2) \;&=\; -\frac{1}{2\pi}\; exp\Big(-\frac{1}{2}\;(y_1^2+y_2^2)\Big)\;[\frac{y_1^2}{y_1^2+y_2^2} \;+\; \frac{y_2^2}{y_1^2+y_2^2}] \\ &=\; -\frac{1}{2\pi}\; exp\Big(-\frac{1}{2}\;(y_1^2+y_2^2)\Big)\;[\frac{y_1^2+y_2^2}{y_1^2+y_2^2}] \\ &=\; -\frac{1}{2\pi}\; exp\Big(-\frac{1}{2}\;(y_1^2+y_2^2)\Big) \end{aligned}$$

On vérifie alors que chaque variable y (indépendante) est distribuée suivant une loi normale.

Méthode polaire de Marsaglia

Pour éviter d'utiliser les fonctions trigonométriques des transformées de Box Muller, considérons les variables aléatoires , W_1, W_2 qui sont uniformément distribuées sur $[-1, 1]$ telles que $W_1^2 + W_2^2 < 1$ puissent être générées de la façon suivante

$$W_1 = 2U_1 - 1, \quad W_2 = 2U_2 - 1 \text{ where } 0 \le W_1^2 + W_2^2 < 1$$

On substitue

$$cos(2\pi U_1) \text{ with } \frac{W_1}{\sqrt{W_1^2 + W_2^2}}$$

$$sin(2\pi U_1) \text{ with } \frac{W_2}{\sqrt{W_1^2 + W_2^2}}$$

En plus, $W_1^2 + W_2^2$ est uniformément distribué sur $[0, 1)$, $-2ln(U_2)$ est remplacé par $\sqrt{-2ln(W_1^2 + W_2^2)}$.
Alors, X and Y sont générées par :

$$X = \sqrt{-2ln(W_1^2 + W_2^2)}\ \frac{W_1}{\sqrt{W_1^2 + W_2^2}}$$

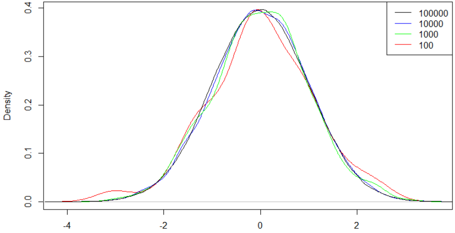
$$Y = \sqrt{-2ln(W_1^2 + W_2^2)}\ \frac{W_2}{\sqrt{W_1^2 + W_2^2}}$$

Méthode du rejet

Exemples :

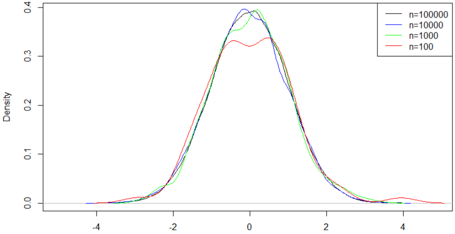
- Enveloppe de Cauchy

$$\forall x \in \mathbb{R}, \ \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \leq \sqrt{\frac{2\pi}{e}} \frac{1}{\pi} \frac{1}{1+x^2}.$$

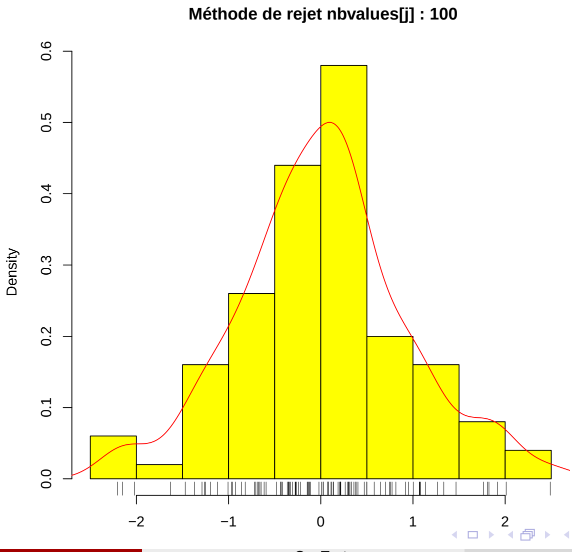


- Enveloppe de Laplace

$$\forall x \in \mathbb{R}, \ \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \leq 2\sqrt{\frac{e}{2\pi}} \frac{1}{2} e^{-|x|}.$$



Méthode du rejet



Loi normale multidimensionnelle

Une variable aléatoire X suit une loi normale multidimensionnelle si sa densité de probabilité est :

$$f(X) = \frac{1}{(2\pi)^{n/2}\,|C|^{1/2}}\ exp[-\frac{1}{2}\,(X-\mu)^t\ C^{-1}\ (X-\mu)]$$

$$\text{avec}\quad X \begin{pmatrix} X_1 \\ X_2 \\ . \\ . \\ X_n \end{pmatrix} \quad \mu \begin{pmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ \mu_n \end{pmatrix}$$

On pose $C^{-1} = \Sigma$ l'inverse de la matrice de variance-covariance
Σ est diagonalisable sous la forme :

$$\Sigma = U\,D_\lambda\,U^{-1} \quad \text{où} \quad \left\{ \begin{array}{ll} U & \textit{matrice orthogonale } (U^{-1} = U') \\ D_\lambda & \textit{matrice diagonale à termes positifs} \end{array} \right.$$

$$f(X) = \frac{1}{(2\pi)^{n/2}\,|C|^{1/2}}\ exp[-\frac{1}{2}\,(X-\mu)^t\ U\,D_\lambda\,U^{-1}\ (X-\mu)].$$

On effectue le changement de variable

$$Y = U^{-1}\,(X - \mu)$$

tel que $\quad Y^t = (U^{-1}\,(X-\mu))^t \;=\; (X-\mu)^t\,(U^{-1})^t$

La propriété $(U^{-1})^t = U$

est vérifiée car U est une matrice orthogonale.

$$f(Y) \;=\; \frac{1}{(2\pi)^{n/2}\,|C|^{1/2}}\;exp[-\frac{1}{2}\,Y^t\,D_\lambda\,Y]$$

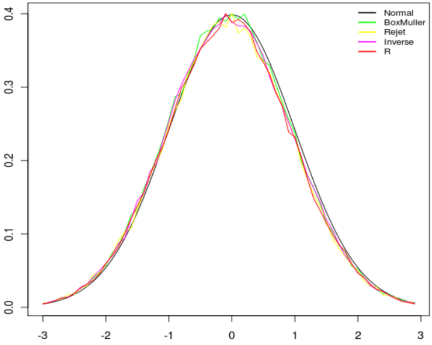
La loi de distribution $f(Y)$ concerne des variables aléatoires indépendantes. On se ramène au cas monodimensionnel en simulant successivement les variables de manière indépendante.

Le processus de simulation peut se résumer en trois étapes :

1. Génération de N observations Z de la variable centrée réduite $\mathcal{N}(0,1)$
2. Le passage à la loi normale $\mathcal{N}(0,\lambda_i)$ se fait par le calcul de $Y = D_\lambda\,Z$
3. Les N observations du vecteur X s'obtiennent par la résolution

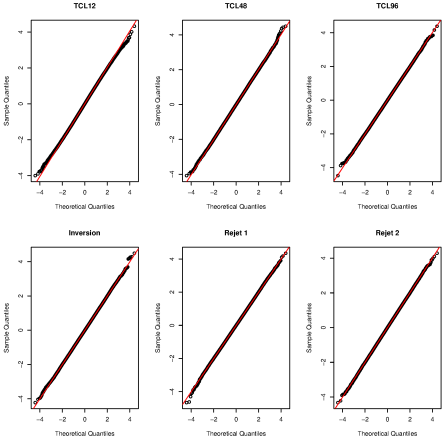
$$Y = U^{-1}\,(X-\mu) \quad\Rightarrow\quad X \;=\; U\,Y + \mu \quad\Rightarrow\quad X \;=\; U\,D_\lambda\,Z \;+\; \mu.$$

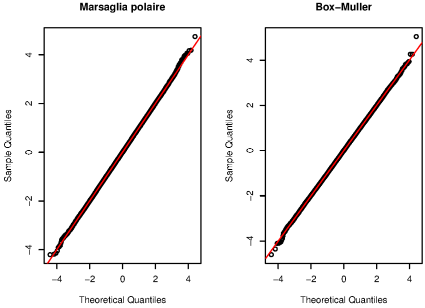
Loi Gaussienne - Tests statistiques



- Test de Student
- Test de Shapiro-Wilk
- Test de Lillierfors
- Test d'Anderson-Darling
- Test de D'Agostino
- Test de Jarque-Bera
- Test de Wilcoxon

QQplot Interpétation





Sommaire.

Loi discrète à support fini

Soit X une V. A. discrète dont l'ensemble des valeurs possibles est fini :

$$X(\Omega) = \{x_1, \ldots, x_d\}.$$

Notons $p_1, p_2, \ldots, p_d$ les probabilités associées telles que :

$$p_k := P(X = x_k), \quad s_0 := 0,\ s_k := \sum_{i \leq k} p_i, \quad 1 \leq k \leq d.$$

Les points $s_0, s_1, \ldots, s_d$ induisent une partition de $[0, 1]$ et si U est une variable de loi uniforme sur $[0, 1[$, $P(U \in [s_{k-1},\ s_k[) = s_k - s_{k-1} = p_k$.

$$\text{On en déduit que} \quad Y := \sum_{k=1}^{d} x_k\, \mathbf{1}_{[s_{k-1}, s_k[}(U) \quad \text{a même loi que } X.$$

Le principe à utiliser pour générer une instance de X est le suivant :

1. générer U de distribution $\mathcal{U}\Big([0, 1]\Big)$
2. donner à X la valeur x_j si

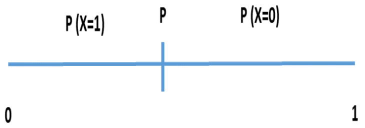
$$\sum_{i=1}^{j-1} p_i \ \leq\ U\ <\ \sum_{i=1}^{j} p_i.$$

Pratiquement on divise l'intervalle $[0, 1]$ en d intervalles de longueur $p_1, p_2, \ldots, p_d$. On regarde dans quel intervalle U tombe et on donne à X la valeur x_i associée au sous-intervalle.

Loi de Bernoulli

On veut simuler une variable aléatoire de loi de Bernoulli de paramètre p ($X \sim B(p)$).
Si

$$U \sim \mathcal{U}([0,\ 1]) \quad \mathrm{X} = \mathbf{1}\mathrm{I}_{\{U \leq p\}} \sim \mathcal{B}(p)$$



En effet

$$P(X = 1) = P(U \leq p) = \int_0^1 \mathbf{1}\mathrm{I}_{\{U \leq p\}}\, du = p$$

Application ”Pile ou face” :
$X \sim \mathcal{B}(\frac{1}{2})$.
On tire $U \rightarrow$ ”Pile” si $U \leq \frac{1}{2}$; ”Face” sinon. Formellement, cette procédure s’écrit

$$\mathbf{1}_{\{U \leq 1/2\}}.$$

Loi binomiale

Une v.a. X à valeurs entières comprises entre 1 et n suit une loi binômiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0, 1]$ si :

$$\forall k = 1, \ldots, n, \qquad P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Proposition

Soient $U_i, i = 1, \ldots, n$ des v.a. uniformes sur $[0, 1]$ indépendantes.
Soit l'ensemble des V.A. de Bernoulli $\mathbf{1}_{\{U_i \le p\}};\ 1 \le i \le n$, de paramètre p indépendantes :

$$X = \sum_{i=1}^{n} \mathbf{1}_{\{U_i \le p\}} \sim \mathcal{B}(n, p)$$

alors X suit une loi binomiale de paramètres n et p.

Méthode

Pour simuler la réalisation d'une v.a.r. de loi binomiale de paramètres (n, p), on peut partir de l'observation que si U suit une loi uniforme sur $[0, 1]$, alors la v.a.r. $Y = \mathbf{1}_{\{U \le p\}}$ suit une loi de Bernoulli de paramètre p. Par conséquent, si $(U_1, \ldots, U_n)$ est un échantillon de la loi uniforme sur $[0, 1]$, alors la v.a.r. $X = \sum_{i=1}^{n} \mathbf{1}_{U_i \le p}$ est une somme de n variables de Bernouilli de même paramètre p ; X suit la loi binomiale de paramètres (n, p).

Loi géométrique

Une v.a. X à valeurs entières strictement positives suit une loi géométrique de paramètre $p \in]0,\ 1[$ si :

$$\forall k \geq 1, \quad P(X = k) = p(1-p)^{k-1}$$

avec

1. $E(X) = \frac{1}{p}, \qquad Var(X) = \frac{1-p}{p^2}$
2. $P(X \leq k) = (1-p)^{k-1}$.

C'est la loi du temps d'attente d'un premier succès dans une suite d'épreuves répétées indépendantes ayant chacune même probabilité de succès p. Par conséquent si $(U_i)_{i\geq 1}$ est une suite i.i.d. de variables suivant la loi uniforme sur $[0,\ 1]$, la V. A.

$$N := min\{k \in \mathbb{N}^*;\ U_k \leq p\}$$

suit la loi géométrique de paramètre p.
(k : 1er essai partir duquel on a $U_k \leq p$)

L'algorithme correspondant consiste donc à générer une v.a. U_k de loi uniforme tant que la valeur obtenue est supérieure ou égale à p et à s'arrêter à la première valeur de U_k strictement inférieure à p en retournant son indice.
Défaut : le nombre moyen de variables uniformes utilisées égal à $E[N] = 1/p$ est grand pour des valeurs de p petites.

Loi de Poisson

Une v.a. X à valeurs entières positives suit une loi de Poisson de paramètre $\lambda > 0$ si :

$$\forall k \geq 0; \quad P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Proposition :
Soit $(E_i)_{i \geq 1}$ une suite de v.a. i.i.d. exponentielles de paramètre λ.
Notons $S_1 := E_1$ et pour $n \geq 2$, $S_n := E_1 + ... + E_n$. On a alors

$$\forall n \geq 1, \qquad P(S_n \leq 1 < S_{n+1}) = e^{-\lambda} \frac{\lambda^n}{n!}.$$

Méthode :
La variable aléatoire

$$Y := \sum_{k=1}^{+\infty} k \, \mathbf{1}_{\{S_k \leq 1 < S_{k+1}\}} = \mathbf{1}_{S_1 \leq 1 < S_2} + 2 \times \mathbf{1}_{S_2 \leq 1 < S_3} + ... + n \times \mathbf{1}_{S_n \leq 1 < S_{n+1}} + ...$$

suit la loi de Poisson de paramètre λ. Pour déduire de cette formule un algorithme effectif, on simule les E_i en posant,

$$E_i = -\lambda^{-1} \, ln U_i,$$

Loi de Poisson

Les U_i étant i.i.d. de loi uniforme sur $\mathcal{U}_{[0,1]}$, on dispose alors des équivalences

$$S_k \leq 1 < S_{k+1} \iff \frac{-1}{\lambda} \sum_{i=1}^{k} lnU_i \leq 1 < \frac{-1}{\lambda} \sum_{i=1}^{k+1} lnU_i$$

$$\iff \prod_{i=1}^{k+1} U_i < e^{-\lambda} \leq \prod_{i=1}^{k} U_i$$

Algorithme :

1. On génére l'une après l'autre les variables U_i et on compare leur produit à $e^{-\lambda}$, en s'arrétant lorsque la relation précédente est vérifiée.
2. On attribue alors à Y la valeur correspondante de k.