SUMMER SCHOOL IN STATISTICS 2016
ROYAL UNIVERSITY OF PHNOM PENH

Applications of the central limit theorem and the Delta method in statistics.

The particular case of the Statistical Study of Sobol Indices

Thierry Klein
thierry.klein@math.univ-toulouse.fr
http://perso.math.univ-toulouse.fr/klein/

INSTITUT de MATHEMATIQUES de TOULOUSE

Contents

1 Gaussian Vectors

1.1 Gaussian Random variables

Definition 1.1. *Let $m \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_+$, we say that a random variable X is a gaussian random variable with parameter m and σ^2 is for any borelian set $A \in \mathcal{B}(\mathbb{R})$ we have*

$$\mathbb{P}\left(X \in A\right) = \int_A e^{-\frac{(x-m)^2}{2\sigma^2}} \frac{dx}{\sqrt{2\pi\sigma^2}}.$$

The function $f(x) = \frac{e^{-\frac{(x-m)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$ is the density of the law of X with respect to the Lebesgue measure on $\mathbb{R}$. We will the note $X \sim \mathcal{N}\left(m, \sigma^2\right)$.
If $m = 0$ and $\sigma^2 = 1$ X is called a standard Gaussian random variable. We will use the following notation

$$\phi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}},$$
$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt = \mathbb{P}\left(X \le t\right).$$

Φ is the cumulative distribution function of the standard gaussian random variable $\mathcal{N}\left(0, 1\right)$

Exercise 1. *Let $X \sim \mathcal{N}\left(m, \sigma^2\right)$.*

1. *Show that $\mathbb{E}\left(X\right) = m$ and $\mathrm{Var}\left(X\right) = \sigma^2$.*
2. *Let a and b be two real numbers show that $aX + b \sim \mathcal{N}\left(am + b, a^2\sigma^2\right)$.*
3. *Compute for any $k \in \mathbb{N}$ when $m = 0$ and $\sigma^2 = 1$, $\mathbb{E}\left(X^k\right)$.*

Lemma 1.1. *If $X \sim \mathcal{N}\left(m, \sigma^2\right)$ then*

1. *his cumulative function at point $t \in \mathbb{R}$ is $\mathbb{P}\left(X \le t\right) = \Phi(\frac{t-m}{\sigma})$.*
2. *his characteristic function is $\varphi(t) = \mathbb{E}\left[e^{itX}\right] = e^{itm - \sigma^2 t^2/2}$.*
3. *his Laplace transform is $L(t) = \mathbb{E}\left[e^{tX}\right] = e^{tm + \sigma^2 t^2/2}$*

Proof. 1. $\mathbb{P}\left(X \le t\right) = \mathbb{P}\left(\frac{X-m}{\sigma} \le \frac{t-m}{\sigma}\right) = \Phi(\frac{t-m}{\sigma})$.

2. Let us assume first that $m = 0$ and σ^2 and let Y be a standard Gaussian random variable then

$$h(t) := \mathbb{E}\left[e^{itY}\right] = \int_{\mathbb{R}} e^{itx} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}.$$

Now, one can apply the theorem of derive under the integral and get

$$h'(t) = \int_{\mathbb{R}} ixe^{itx} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}.$$

We now integrate by part and get that ($u'(x) = xe^{-x^2/2}$ and $v(x) = ie^{its}$) and get that $h'(t) = -th(t)$ and hence since h(0)=1 we have that $h(t) = e^{-t^2/2}$. Now,

$$\varphi(t) = \mathbb{E}\left[e^{itX}\right] = \mathbb{E}\left[e^{it(\sigma Y + m)}\right] = e^{itm} h(\sigma t),$$

the result follows.

3. Let us assume first that $m = 0$ and σ^2 and let Y be a standard Gaussian random variable then

$$g(t) := \mathbb{E}\left[e^{tY}\right] = \int_{\mathbb{R}} e^{tx} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} = e^{t^2/2} \int_{\mathbb{R}} e^{-(x-t)^2/2} \frac{dx}{\sqrt{2\pi}} = e^{t^2/2}.$$

Now

$$L(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{t(\sigma Y + m)}\right] = e^{tm} g(\sigma t),$$

the result follows.

□

Exercise 2. *1. Show that if X and Y are independent and if $X \sim \mathcal{N}\left(m_1, \sigma_1^2\right)$ and $X \sim \mathcal{N}\left(m_2, \sigma_2^2\right)$ then $X+Y \sim \mathcal{N}\left(m_1+m_2, \sigma_1^2+\sigma_2^2\right)$*

2. Let $(X_1, \ldots, X_n)$ be i.i.d. standard gaussian random variables what is the law of $\frac{X_1+\ldots+X_n}{\sqrt{n}}$?

1.2 Gaussian Vectors

Definition 1.2. *A random vector $X := (X_1, \ldots, X_d)$ is a Gaussian vector of dimension d if for any $\alpha \in \mathbb{R}^d$ the real random variable*

$$< X, \alpha >= \sum_{i=1}^{d} \alpha_i X_i$$

is a real Gaussian variable. The expectation of X is the vector $\mathbb{E}[X] := (\mathbb{E}[]X_1, \ldots, \mathbb{E}[X_d])$ and his Covariance matrix is a $d \times d$ matrix Γ

$$\Gamma = \mathrm{Cov}\,(X) := \left(Cov\,(X_i, X_j)\right)_{1 \le i,j \le d}.$$

Lemma 1.2. *If $X_1, \ldots, X_d$ are i.i.d standard Gaussian random variables then the vector $X = (X_1, \ldots, X_d)$ is a Gaussian random vector with $\mathbb{E}[X] = 0$ and $\Gamma = I_d$ where I_d is the identity matrix of size d.*

Proof. Let $\alpha \in \mathbb{R}^d$ and let us consider the characteristic function of $Y =< x, \alpha >$

$$\phi_Y(t) = \mathbb{E}\left[e^{itY}\right] = \mathbb{E}\left[e^{it\sum_{j=1}^d \alpha_j X_j}\right] = \prod_{j=1}^{d} \mathbb{E}\left[e^{it\alpha_j X_j}\right] = \prod_{j=1}^{d} e^{-\alpha_j^2 t^2/2} = e^{-\sum_{j=1}^d \alpha_j^2 t^2/2}.$$

Hence $Y \sim \mathcal{N}\left(0, \sum_{j=1}^d \alpha_j^2\right)$. □

Exercise 3. *1. Show that $X_1, \ldots, X_d$ are independant standard gaussian random variable then $X = (X_1, \ldots, X_d) \sim \mathcal{N}_d\left(\mathbb{E}[X], \Gamma\right)$ where Γ is a diagonal matrix.*

2. Let $X \sim \mathcal{N}(0,1)$ and ϵ independent of X such that $\mathbb{P}(\epsilon = 1) = \mathbb{P}(\epsilon = 1) = 1/2$. Show that both $\epsilon X \sim \mathcal{N}(0,1)$ but that the vector $X, \epsilon X$ is not Gaussian. In particular a vector X whose coordinate are Gaussian is not necessary a Gaussian vector.

Proposition 1.1. *The law of a gaussian vector $X \sim \mathcal{N}_d(m, \Gamma)$ in $\mathbb{R}^d$ is characterized by the mean vector $m = \mathbb{E}[X]$ and is covariance matrix Γ. More precisely, for any $\alpha \in \mathbb{R}^d$*

$$\Phi_X(\alpha) = \mathbb{E}\left[e^{i<\alpha, X>}\right] = \exp\left(i < \alpha, m > - \frac{< \alpha, \Gamma\alpha >}{2}\right)$$

Proof. Just recall that by definition $< \alpha, X >$ is a gaussian random variable with mean $< \alpha, m >$ and with variance

$$\sigma^2 := \mathrm{Var}\,(< \alpha, X >)$$

$$\sigma^2 = \mathrm{Var}\left(\sum_{j=1}^{d} \alpha_j X_j\right) = \sum_{j,k} \alpha_j \alpha_k \Gamma_{j,k} =< \alpha, \Gamma\alpha > .$$

□

Proposition 1.2. *1. Let $X \sim \mathcal{N}_d(m, \Gamma)$, A be a linear mapping fron $\mathbb{R}^d$ to $\mathbb{R}^m$ and b be a fixed vector in $\mathbb{R}^m$. Then $Y = AX + b$ is the gaussian vector $Y \sim \mathcal{N}_m\left(Am + b, A\Gamma A^t\right)$, where A^t is the transpose of the matrix A.*

2. If $X \sim \mathcal{N}_m(0, I_d)$, $m \in \mathbb{R}^d$ and Γ a symmetric matrix with non negative eigenvalues then $m + \sqrt{\Gamma}X \sim \mathcal{N}_m(m, \Gamma)$.

3. Let $m \in \mathbb{R}^d$ and Γ a symmetric matrix with positive eigenvalues then the Gaussian vector $X \sim \mathcal{N}_m(m, \Gamma)$ admits a density with respect to the Lebesgue's measure on $\mathbb{R}^d$ defined by

$$\exp\left(-\frac{1}{2} < x - m, \Gamma^{-1}(x - m) >\right) \frac{dx}{(2\pi)^{d/2}\sqrt{Det(\Gamma)}}.$$

1.3 Practical and Exercices

Exercise 4. *1. Let $Z_1, \dots, Z_n$ be i.i.d $\mathcal{N}(0,1)$ random variables. we set $\overline{Z_n} = \frac{1}{n}\sum_{k=1}^{n} Z_k$. Show that $G = \left(Z_1 - \overline{Z_n}, Z_2 - \overline{Z_n}, \dots, Z_n - \overline{Z_n}, \overline{Z_n}\right)$ is a Gaussian vector, compute its covariance matrix Γ_n.*

2. *Show that $U_n = \left(Z_1 - \overline{Z_n}, Z_2 - \overline{Z_n}, \dots, Z_n - \overline{Z_n}\right)$ and $\overline{Z_n}$ are independant.*

3. *Let Σ_n be the covariance matrix of the Gaussian vector U_n. Determine the eigenvalues of Σ_n. What is the law of $\|U_n\|$.*

Exercise 5. *Let (X, Y) be a random vector uniformly distributed on unit ball*

$$D = \left\{(x, y); x^2 + y^2 < 1\right\}.$$

Using polar coordinate we can write

$$X = R\cos(\Theta),\ Y = R\sin(\Theta).$$

Set

$$\begin{aligned} R' &= \sqrt{-4\log(R)} \\ U &= R'\cos(\Theta) \\ V &= R'\sin(\Theta). \end{aligned}$$

1. *Show that U and V are two independent Standard gaussian random variables ($U \sim \mathcal{N}(0,1)$ and $U \sim \mathcal{N}(0,1)$).*

2. *Write a program (in matlab or R or...) that simulate random variables uniformly distributed on D. We assume here that our computer can simulate independent and uniformly distributed random variable on the interval $[0,1]$.*

3. *Write a program that simulate N independent gaussian vector $\mathcal{N}_d(0, I_d)$.*

Exercise 6. *Let U_1 and U_2 be a two independent uniformly distributed on $[0,1]$. Set*

$$\begin{aligned} X_1 &= \sqrt{-2\log(U_1)}\cos(2\pi U_2), \\ X_1 &= \sqrt{-2\log(U_1)}\sin(2\pi U_2). \end{aligned}$$

1. *Show that X_1 and X_2 are two independent Standard gaussian random variables ($U \sim \mathcal{N}(0,1)$ and $U \sim \mathcal{N}(0,1)$).*

2. *Write a program that simulate N independent gaussian vector $\mathcal{N}_d(0, I_d)$.*

Exercise 7. *Let Σ be a symmetric positive matrix of size d and μ be a vector in $\mathbb{R}^d$. The aim of this exercice is to perform the simulation of a gaussian $\mathcal{N}_d(\mu, \Sigma)$.*

1. *Case $d = 1$, show that if $X \sim \mathcal{N}(0,1)$ then $\Sigma X + \mu \sim \mathcal{N}_d(\mu, \Sigma)$. Write the corresponding code to simulate a $\mathcal{N}_d(\mu, \Sigma)$ random variable.*

2. *Case $d = 2$, let $(Y_1, Y_2) \sim \mathcal{N}_2\left((\mu_1, \mu_2), \Sigma\right)$ with*

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$$

   where $\sigma_i^2 = \mathrm{Var}(Y_i)$ and $\rho = \frac{\mathrm{Cov}(Y_1, Y_2)}{\sigma_1\sigma_2}$.

   (a) *Show that if*

   $$A = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_2\rho & \sigma_2\sqrt{1-\rho^2} \end{pmatrix}$$

   then $AA^t = \Sigma$

   (b) *Let $X \sim \mathcal{N}_2(0, I_2)$ what is the law of $AX + (\mu_1, \mu_2)$?*

   (c) *Write the corresponding code to simulate a $\mathcal{N}_2(\mu, \Sigma)$ random variable.*

3. *General case. Use the Cholesky to decompose Σ as AA^t.*