

The Delta-Method applied to Sobol' indices

Michaël Baudin

February 18, 2019

Abstract

We explore the use of the Delta-method in order to estimate the Sobol' indices.

Contents

1	Random vector	1
2	Convergence in distribution	2
3	Convergence in probability	3
4	Convergence in probability implies convergence in distribution	4
5	Delta method	5
6	References	7

1 Random vector

Definition 1. (Random vector) *A random vector in \mathbb{R}^k is a vector $\mathbf{X} = (X_1, \dots, X_k) \in \mathbb{R}^k$ of random variables. The cumulative distribution function of \mathbf{X} is F is the function defined by:*

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

for any $\mathbf{x} \in \mathbb{R}^k$.

In the remaining of this section, most results can be derived equally for the real univariate case (i.e. for $k = 1$) or for the general multivariate case. In order to introduce multivariate definitions, a distance on \mathbb{R}^k must be defined. For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$, let $d(\mathbf{x}, \mathbf{y})$ be a distance function on \mathbb{R}^k that generates the topology. In the particular univariate case, the most common distance is the absolute value:

$$d(x, y) = |x - y|,$$

for any $x, y \in \mathbb{R}$. In the general multivariate case, the most common distance is the euclidian distance:

$$d(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^k (x_i - y_i)^2 \right)^{1/2},$$

for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$.

2 Convergence in distribution

Definition 2. (Convergence in distribution) Assume that $\mathbf{X}_1, \mathbf{X}_2, \dots$ is a sequence of real-valued random variables with cumulative distribution functions $\{F_n\}_{n \geq 0}$. Assume that \mathbf{X} is a real-valued random variable with cumulative distribution function F . The sequence \mathbf{X}_n converges in distribution to \mathbf{X} if:

$$\lim_{n \rightarrow \infty} F_n(\mathbf{x}) = F(\mathbf{x}). \quad (1)$$

for any $\mathbf{x} \in \mathbb{R}^k$ at which F is continuous. In this case, we write:

$$\mathbf{X}_n \xrightarrow{D} \mathbf{X}. \quad (2)$$

By definition of the cumulative distribution function, the equation 1 can be written:

$$\lim_{n \rightarrow \infty} P(\mathbf{X}_n \leq \mathbf{x}) = P(\mathbf{X} \leq \mathbf{x}). \quad (3)$$

for any $\mathbf{x} \in \mathbb{R}^k$ at which F is continuous.

The following theorem gives an example of such convergence.

Example 1. (Maximum of uniform random numbers) Assume that $X_1, X_2, \dots \in \mathbb{R}$ are independent uniform random numbers such that $X_n \sim \mathcal{U}(0, 1)$. Let Y_n be the maximum:

$$Y_n = \max_{1 \leq i \leq n} X_i.$$

Therefore the sequence $n(1 - Y_n)$ converges in distribution to an exponential random variable, i.e.:

$$n(1 - Y_n) \xrightarrow{D} \mathcal{E}(1).$$

Proof. By definition, the exponential distribution with rate λ has the cumulative distribution function:

$$F(y) = 1 - \exp(-\lambda y),$$

for any real $y \geq 0$. We apply the previous equality to $\lambda = 1$ which shows that we must prove that:

$$F(y) = 1 - \exp(-y),$$

for any real $y \geq 0$. Let F_n be the cumulative distribution function of the random variable $n(1 - Y_n)$. By definition of the cumulative distribution function,

$$\begin{aligned} F_n(z) &= P(n(1 - Y_n) \leq z) \\ &= P(1 - Y_n \leq z/n) \\ &= P(1 - z/n \leq Y_n) \\ &= 1 - P(Y_n \leq 1 - z/n), \end{aligned}$$

for any $z \in \mathbb{R}$. However, the cumulative distribution function of the maximum is:

$$\begin{aligned} P(Y_n \leq y) &= P(X_1 \leq y, X_2 \leq y, \dots, X_n \leq y) \\ &= P(X_1 \leq y)P(X_2 \leq y) \cdots P(X_n \leq y) \text{ (by independence)} \\ &= yy \cdots y \text{ (by definition of the C.D.F. of a uniform random variable)} \\ &= y^n, \end{aligned}$$

for any $y \in [0, 1]$. Therefore,

$$F_n(z) = 1 - (1 - z/n)^n$$

for any $z \in \mathbb{R}$. But

$$\lim_{n \rightarrow \infty} (1 + z/n)^n = \exp(z)$$

for any $z \in \mathbb{R}$. Therefore,

$$\lim_{n \rightarrow \infty} F_n(z) = 1 - \exp(-z)$$

for any $z \in \mathbb{R}$ which concludes the proof. \square

3 Convergence in probability

Definition 3. (Convergence in probability) Assume that $\mathbf{X}_1, \mathbf{X}_2, \dots \in \mathbb{R}^k$ is a sequence of random vector variables. The sequence \mathbf{X}_n converges in probability to $\mathbf{X} \in \mathbb{R}^k$ if, for any $\epsilon > 0$:

$$\lim_{n \rightarrow \infty} P(d(\mathbf{X}_n, \mathbf{X}) > \epsilon) = 0. \quad (4)$$

In this case, we write:

$$\mathbf{X}_n \xrightarrow{P} \mathbf{X}. \quad (5)$$

In the particular univariate case, the equation 6 is generally written:

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0. \quad (6)$$

Example 2. (Convergence of an exponential random variable) Assume that $X_1, X_2, \dots \in \mathbb{R}$ are independent exponential random numbers, i.e. such that $X_n \sim \text{Exp}(n)$. Therefore the sequence X_n converges in probability to the zero random variable:

$$X_n \xrightarrow{P} 0.$$

Proof. By definition of the exponential cumulative distribution function, we have

$$P(X_n \leq x) = 1 - \exp(-nx),$$

for any real number $x \geq 0$. Let $X = 0$. For any $\epsilon > 0$, we have:

$$\begin{aligned} P(|X_n - X| > \epsilon) &= P(|X_n| > \epsilon) \text{ (since } X = 0) \\ &= P(X_n > \epsilon) \text{ (since } X_n \geq 0) \\ &= 1 - P(X_n \leq \epsilon) \\ &= \exp(-n\epsilon) \end{aligned}$$

by definition of the exponential cumulative distribution function. Since $\epsilon > 0$, this implies:

$$\lim_{n \rightarrow \infty} \exp(-n\epsilon) = 0$$

which implies that

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$$

and concludes the proof. \square

4 Convergence in probability implies convergence in distribution

Theorem 1. Assume that X and Y are random variables. Let $y \in \mathbb{R}$ and let $\epsilon > 0$. Therefore

$$P(Y \leq y) \leq P(X \leq y + \epsilon) + P(|Y - X| > \epsilon) \quad (7)$$

Proof.

$$P(Y \leq y) = P(Y \leq y, X \leq y + \epsilon) + P(Y \leq y, X > y + \epsilon).$$

But $P(Y \leq y, X \leq y + \epsilon) \leq P(X \leq y + \epsilon)$, which implies :

$$\begin{aligned} P(Y \leq y) &\leq P(X \leq y + \epsilon) + P(Y \leq y, X > y + \epsilon) \\ &\leq P(X \leq y + \epsilon) + P(Y - X \leq y - X, y + \epsilon < X) \\ &\leq P(X \leq y + \epsilon) + P(Y - X \leq y - X, y - X < -\epsilon). \end{aligned}$$

Moreover, we have $P(Y - X \leq y - X, y - X < -\epsilon) \leq P(Y - X < -\epsilon)$ which implies :

$$P(Y \leq y) \leq P(X \leq y + \epsilon) + P(Y - X < -\epsilon).$$

By definition of a probability, we have $P(Y - X > \epsilon) \geq 0$. Therefore,

$$P(Y \leq y) \leq P(X \leq y + \epsilon) + P(Y - X < -\epsilon) + P(Y - X > \epsilon)$$

which leads to the equation 7 and concludes the proof. \square

Theorem 2. (Convergence in probability implies convergence in distribution) Assume that X_1, X_2, \dots is a sequence of real-valued random variables. If

$$X_n \xrightarrow{P} X$$

then

$$X_n \xrightarrow{D} X.$$

Proof. Assume that $\{F_n\}_{n \geq 0}$ are the cumulative distribution functions of X_1, X_2, \dots . Assume that X is a real-valued random variable with cumulative distribution function F and that $X_n \xrightarrow{P} X$. Let $x \in \mathbb{R}$ be a point at which F is continuous. By the definition 2 we must prove that

$$\lim_{n \rightarrow \infty} F_n(X_n) = F(x). \quad (8)$$

Let $\epsilon > 0$.

First, we first apply the theorem 1 to the random variables X_n and X . The equation 7 implies:

$$P(X_n \leq x) \leq P(X \leq x + \epsilon) + P(|X_n - X| > \epsilon).$$

By hypothesis, the sequence X_1, X_2, \dots converges in probability to X , which implies:

$$\lim_{n \rightarrow \infty} P(X_n \leq x) \leq P(X \leq x + \epsilon).$$

since $\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$. By definition of the cumulative distribution function F , this implies:

$$\lim_{n \rightarrow \infty} P(X_n \leq x) \leq F(x + \epsilon). \quad (9)$$

Secondly, we consider the theorem 1 and derive a similar inequality. In the equation 7, we exchange the random variables X and Y :

$$P(X \leq y) \leq P(Y \leq y + \epsilon) + P(|X - Y| > \epsilon)$$

and we apply the inequality to the real $y - \epsilon$; this leads to:

$$P(X \leq y - \epsilon) \leq P(Y \leq y) + P(|X - Y| > \epsilon).$$

Back to the theorem 2, we apply the previous inequality to the random variables X and X_n , at the point x . We obtain:

$$P(X \leq x - \epsilon) \leq P(X_n \leq x) + P(|X - X_n| > \epsilon).$$

By hypothesis, the sequence X_1, X_2, \dots converges in probability to X , which implies:

$$P(X \leq x - \epsilon) \leq \lim_{n \rightarrow \infty} P(X_n \leq x).$$

By definition of the cumulatif distribution function F , this implies:

$$F(x - \epsilon) \leq \lim_{n \rightarrow \infty} P(X_n \leq x). \quad (10)$$

Combining the inequalities 9 and 10, we obtain:

$$F(x - \epsilon) \leq \lim_{n \rightarrow \infty} P(X_n \leq x) \leq F(x + \epsilon).$$

The function F is, by hypothesis, continuous at the point x which implies than we can take the limit of the previous inequality when $\epsilon \rightarrow 0$. This implies:

$$F(x) \leq \lim_{n \rightarrow \infty} P(X_n \leq x) \leq F(x)$$

which leads to the equation 8 and concludes the proof. \square

5 Delta method

Theorem 3. (Joint random vector convergence in distribution) *Let X_n and Y_n be two sequences of real random variables such that:*

$$X_n \xrightarrow{D} X \quad \text{and} \quad Y_n \xrightarrow{P} c,$$

where c is a non-random constant. Therefore, the joint random vector (X_n, Y_n) converges in distribution to (X, c) :

$$(X_n, Y_n) \xrightarrow{D} (X, c).$$

Proof. See [?], chapter 2, Stochastic convergence, theorem 2.7, page 10. \square

Theorem 4. (Continuous mapping) *Let X_n be a sequence of real random variables such that:*

$$X_n \xrightarrow{D} X.$$

Assume that g is a continous function on \mathbb{R} . Therefore:

$$g(X_n) \xrightarrow{D} g(X).$$

Proof. See [?]. \square

Theorem 5. (Slutsky's theorem) Let X_n and Y_n be two sequences of real random variables such that:

$$X_n \xrightarrow{D} X \quad \text{and} \quad Y_n \xrightarrow{P} c,$$

where c is a non-random constant. Therefore,

1. $X_n Y_n \xrightarrow{D} Xc$,
2. $X_n + Y_n \xrightarrow{D} X + c$,
3. $X_n/Y_n \xrightarrow{D} X/c$.

Proof. Let us prove that $X_n Y_n \xrightarrow{D} Xc$. Let g be the continuous function defined by $g(x, y) = xy$, for any $x, y \in \mathbb{R}$. By the theorem 3, we have $(X_n, Y_n) \xrightarrow{D} (X, c)$. By the continuous mapping theorem, this implies:

$$X_n Y_n = g((X_n, Y_n)) \xrightarrow{D} g((X, c)) = Xc.$$

In order to prove the equation $X_n + Y_n \xrightarrow{D} X + c$, we use the continuous function $g(x, y) = x + y$, which immediately leads to the required result.

Finally, let g be the continuous function defined by $g(x, y) = x/y$, for any $x, y \in \mathbb{R}$ such that $y \neq 0$. \square

Theorem 6. (Univariate delta-method) Assume that $X_1, X_2, \dots \in \mathbb{R}$ is a sequence of real-valued random variables and $\theta \in \mathbb{R}$ are so that

$$\sqrt{n}(X_n - \theta) \xrightarrow{D} \mathcal{N}(0, \sigma^2). \quad (11)$$

Assume that $g \in C^1(\mathbb{R})$ and that $g'(\theta) \neq 0$. Therefore,

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{D} \mathcal{N}(0, \sigma^2 g'(\theta)^2).$$

Proof. By hypothesis, we have $g \in C^1(\mathbb{R})$ so that we can apply Taylor's theorem. Taylor's expansion of g at the point θ implies that there exists a $\tilde{\theta}$ between X_n and θ such that:

$$g(X_n) = g(\theta) + g'(\tilde{\theta})(X_n - \theta).$$

This implies:

$$\sqrt{n}(g(X_n) - g(\theta)) = g'(\tilde{\theta}) \sqrt{n}(X_n - \theta). \quad (12)$$

Since $\tilde{\theta}$ is between X_n and θ , for any $\epsilon > 0$, we have:

$$\lim_{n \rightarrow \infty} P(|\tilde{\theta} - \theta| > \epsilon) = 0.$$

This implies:

$$\tilde{\theta} \xrightarrow{P} \theta.$$

By hypothesis, the function g is continuously differentiable, which implies that g' is a continuous function. Therefore, the continuous mapping theorem implies:

$$g'(\tilde{\theta}) \xrightarrow{P} g'(\theta).$$

The equation 11 and Slutsky's theorem imply:

$$g'(\tilde{\theta}) \sqrt{n}(X_n - \theta) \xrightarrow{D} g'(\theta) \mathcal{N}(0, \sigma^2) = \mathcal{N}(0, \sigma^2 g'(\theta)^2),$$

where the last equation is from the properties of the gaussian distribution. The equation 12 concludes the proof. \square

Theorem 7. (Multivariate delta-method) Assume that $\mathbf{X}_1, \mathbf{X}_2, \dots \in \mathbb{R}^p$ is a sequence of random variables and $\boldsymbol{\theta} \in \mathbb{R}^p$ are so that

$$\sqrt{n}(\mathbf{X}_n - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Sigma), \quad (13)$$

where $\Sigma \in \mathbb{R}^{p \times p}$ is a positive semi-definite covariance matrix. Suppose that $\mathbf{g} : \mathbb{R}^p \rightarrow \mathbb{R}^m \in C^1(\mathbb{R}^p)$. For any $\boldsymbol{\theta} \in \mathbb{R}^p$, let $J(\boldsymbol{\theta}) \in \mathbb{R}^{m \times p}$ be the Jacobian matrix of \mathbf{g} :

$$J(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} & \dots & \frac{\partial g_1}{\partial \theta_p} \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial \theta_1} & \dots & \frac{\partial g_m}{\partial \theta_p} \end{pmatrix}$$

Assume that $\nabla \mathbf{g}(\boldsymbol{\theta}) \neq \mathbf{0}$. Therefore,

$$\sqrt{n}(\mathbf{g}(\mathbf{X}_n) - \mathbf{g}(\boldsymbol{\theta})) \xrightarrow{D} \mathcal{N}(\mathbf{0}, J(\boldsymbol{\theta})^T \Sigma J(\boldsymbol{\theta})).$$

In the previous theorem, notice that $\mathbf{0} \in \mathbb{R}^p$.

Proof. By hypothesis, we have $\mathbf{g} \in C^1(\mathbb{R}^p)$ so that we can apply Taylor's theorem. Taylor's expansion of the function \mathbf{g} at the point $\boldsymbol{\theta}$ implies that there exists a $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p$ between \mathbf{X}_n and $\boldsymbol{\theta}$ such that:

$$\mathbf{g}(\mathbf{X}_n) = \mathbf{g}(\boldsymbol{\theta}) + J(\tilde{\boldsymbol{\theta}})(\mathbf{X}_n - \boldsymbol{\theta}).$$

This implies:

$$\sqrt{n}(\mathbf{g}(\mathbf{X}_n) - \mathbf{g}(\boldsymbol{\theta})) = J(\tilde{\boldsymbol{\theta}}) \sqrt{n}(\mathbf{X}_n - \boldsymbol{\theta}). \quad (14)$$

By definition of the variance,

$$V(\sqrt{n}(g(X_n) - g(\theta))) = V(J(\tilde{\theta}) \sqrt{n}(X_n - \theta))$$

TODO : finish this. □

6 References

Most of the results in this document can be found in [?], chapter 2, Stochastic convergence.