The Delta-Method applied to Sobol' indices

Michaël Baudin

February 12, 2019

Abstract

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

Contents

1 Convergence in distribution

Definition 1. *(*Convergence in distribution*) Assume that* $X_1, X_2, \ldots$ *is a sequence of real-valued random variables with cumulative distribution functions* $\{F_n\}_{n\geq 0}$. *Assume that* X *is a real-valued random variable with cumulative distribution function* F. *The sequence* X_n *converges in distribution to* X *if:*

$$\lim_{n\to\infty} F_n(X_n) = F(x).$$

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

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

The following theorem gives an example of such convergence.

Example 1. *(*Maximum of uniform random numbers*) Assume that* $X_1, X_2, \ldots$ *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:

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

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

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

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

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

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

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

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

2 Convergence in probability

Definition 2. *(*Convergence in probability*) Assume that $X_1, X_2, ...$ is a sequence of real-valued random variables. The sequence X_n converges in probability to X if, for any $\epsilon > 0$:*

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

In this case, we write:

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

Example 2. *(*Convergence of an exponential random variable*) Assume that $X_1, X_2, ...$ are independent random numbers such that $X_n \sim \mathcal{E}xp(n)$. 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:

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

which implies that

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

and concludes the proof. □

3 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*

$$P(Y \leq y) \leq P(X \leq y + \epsilon) + P(|Y - X| > \epsilon) \tag{1}$$

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 1 and concludes the proof. □

Theorem 2. *(*Convergence in probability implies convergence in distribution*) Assume that $X_1, X_2, ...$ 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 $X_1, X_2, ...$ 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 1 we must prove that

$$\lim_{n\to\infty} F_n(X_n) = F(x). \tag{2}$$

Let $\epsilon > 0$.

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

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

By hypothesis, the sequence $X_1, X_2, ...$ converges in probability to X, which implies:

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

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

$$\lim_{n\to\infty} P(X_n \leq x) \leq F(x + \epsilon). \tag{3}$$

Secondly, we consider the theorem 1 and derive a similar inequality. In the equation 1, 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 $X_1, X_2, ...$ converges in probability to X, which implies:

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

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

$$F(x - \epsilon) \leq \lim_{n\to\infty} P(X_n \leq x). \tag{4}$$

Combining the inequalities 3 and 4, we obtain:

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

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

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

which leads to the equation 2 and concludes the proof. $\square$

4 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 [**?**]. □

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 [**?**]. □

Theorem 5. *(*Slutky'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 continous fonction $g(x, y) = x + y$, 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 $y \neq 0$. □

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

$$\sqrt{n}(X_n - \theta) \xrightarrow{D} \mathcal{N}(0, \sigma^2). \tag{5}$$

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:

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

This implies:

$$\sqrt{n}(g(X_n) - g(\theta)) = g'\left(\tilde{\theta}\right)\sqrt{n}(X_n - \theta). \tag{6}$$

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

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

This implies:

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

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

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

The equation 5 and Slutky's theorem imply:

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

where the last equation is from the properties of the gaussian distribution. The equation 6 concludes the proof. □

Theorem 7. *(*Multivariate delta-method*) Assume that $X_1, X_2, ... \in \mathbb{R}^p$ is a sequence of real-valued random variables and $\theta \in \mathbb{R}^p$ are so that*

$$\sqrt{n}(X_n - \theta) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \Sigma\right), \tag{7}$$

where $\Sigma \in \mathbb{R}^{p\times p}$ is a positive semi-definite covariance matrix. Suppose that $g : \mathbb{R}^p \to \mathbb{R} \in C^1(\mathbb{R}^p)$. Let $\nabla g(\theta) \in \mathbb{R}^p$ be the gradient of g and assume that $\nabla g(\theta) \neq \mathbf{0}$. Therefore,

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \nabla g(\theta)^T \Sigma \nabla g(\theta)\right).$$

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

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

$$g(X_n) = g(\theta) + \nabla g\left(\tilde{\theta}\right)^T (X_n - \theta).$$

TODO : finish this. □

5 References

Most of the results in this document can be found in [**?**].