

Package 'sensitivity' : scientific appendix

Gilles Pujol

July 9, 2008

This document presents the formulas implemented in the 'sensitivity' package.

Notations

model	$y = f(x_1, \dots, x_p)$
n -sample	$X = (X_{ki})_{\substack{k=1\dots n \\ i=1\dots p}}$
vector extraction	$y_{\cdot} = (y_k)_{k=1\dots n}$ $X_{\cdot i} = (X_{ki})_{k=1\dots n}$
implicit loop	$f_1(x_{\cdot}) = f_2(y_{\cdot})$ means $\forall k = 1 \dots n, f_1(x_k) = f_2(y_k)$
estimators	$\widehat{\text{var}}$: variance $\widehat{\text{cor}}$: Pearson's correlation
rounding	$\lfloor x \rfloor$: largest integer not greater than x

1 srcpcc

Linear regressions :

$$\begin{aligned} y_{\cdot} &\simeq b_0 + \sum_{j=1}^p b_j X_{\cdot j} \\ y_{\cdot} &\simeq c_0 + \sum_{\substack{j=1 \\ j \neq i}}^p c_j X_{\cdot j} \\ X_{\cdot i} &\simeq d_{i0} + \sum_{\substack{j=1 \\ j \neq i}}^p d_{ij} X_{\cdot j} \quad (i = 1 \dots p) \end{aligned}$$

Sensitivity indices ($i = 1 \dots p$) :

$$\begin{aligned} \text{SRC}_i &= \frac{\widehat{\text{var}}(X_{\cdot i})}{\widehat{\text{var}}(y_{\cdot})} b_i^2 \\ \text{PCC}_i &= \widehat{\text{cor}}\left(y_{\cdot} - c_0 - \sum_{\substack{j=1 \\ j \neq i}}^p c_j X_{\cdot j}, X_{\cdot i} - d_{i0} - \sum_{\substack{j=1 \\ j \neq i}}^p d_{ij} X_{\cdot j}\right) \end{aligned}$$

2 morris

Notations about the domain :

- $\bigotimes_{i=1}^p [a_i, b_i]$: the domain
- n_1, \dots, n_p : number of levels
- k_1, \dots, k_p : “grid jump” coefficients

Delta ($i = 1 \dots p$) :

$$\Delta_i = k_i \frac{b_i - a_i}{n_i - 1}$$

Discretisation of the space :

$$\begin{aligned} G &= \bigotimes_{i=1}^p \left\{ a_i + k \frac{b_i - a_i}{n_i - 1} \right\}_{k=0 \dots n_i - 1} && \text{(grid on the whole domain)} \\ G' &= \bigotimes_{i=1}^p \left\{ a_i + k \frac{b_i - a_i}{n_i - 1} \right\}_{k=0 \dots n_i - 1 - k_i} && \text{(grid restricted to } \bigotimes_{i=1}^p [a_i, b_i - \Delta_i]) \end{aligned}$$

Random elements ($r = 1 \dots R$) :

- $d^{(r)}$: vector of length p composed of equiprobable + ones and - ones
- $x^{(r)}$: randomly chosen point on the grid G' (row vector of length p)

The $p + 1 \times p$ matrix of the design of experiments ($r = 1 \dots R$) :

$$X^{(r)} = J_{p+1,1} x^{(r)} + \frac{(2B - J_{p+1,p})D(d^{(r)}) + J_{p+1,p}D(\Delta_1, \dots, \Delta_p)}{2}$$

where

- $J_{i,j}$: $i \times j$ matrix filled with ones
- B : $(p + 1) \times p$ matrix with ones in the lower triangular part and zeros in the upper part, e.g.

$$B = \begin{pmatrix} 0 & \dots & 0 \\ 1 & \ddots & \vdots \\ \vdots & \ddots & 0 \\ 1 & \dots & 1 \end{pmatrix}$$

- $D(x)$: $p \times p$ diagonal matrix with the elements of the vector x on the diagonal

Predictions ($r = 1 \dots R, i = 1 \dots p + 1$) :

$$y_i^{(r)} = f(X_{i1}^{(r)}, \dots, X_{ip}^{(r)})$$

Elementary effects ($r = 1 \dots R, i = 1 \dots p$) :

$$EE_i^{(r)} = d_i^{(r)} \frac{y_{i+1}^{(r)} - y_i^{(r)}}{\Delta_i}$$

Sensitivity indices ($i = 1 \dots p$) :

$$\begin{aligned} \mu_i^* &= \frac{1}{R} \sum_{r=1}^R |EE_i^{(r)}| \\ \sigma_i &= \sqrt{\widehat{\text{var}}(EE_i^{(\cdot)})} \end{aligned}$$

3 sobol

Two initial n -samples, noted $X^{(1)}$ and $X^{(2)}$.

The corresponding response :

$$\begin{aligned} y_{\cdot}^{(1)} &= f(X_{\cdot,1}^{(1)}, \dots, X_{\cdot,p}^{(1)}) \\ y_{\cdot}^{(2)} &= f(X_{\cdot,1}^{(2)}, \dots, X_{\cdot,p}^{(2)}) \end{aligned}$$

3.1 method=sobol93

One more n -sample for each subset of indices $I = \{i_1, \dots, i_{n_I}\}$, noted $X^{(2,I,1)}$:

$$\begin{aligned} X_{\cdot,i}^{(2,I,1)} &= X_{\cdot,i}^{(2)}, \text{ if } i \notin I \\ X_{\cdot,i}^{(2,I,1)} &= X_{\cdot,i}^{(1)}, \text{ if } i \in I \end{aligned}$$

The response :

$$y_{\cdot}^{(2,I,1)} = f(X_{\cdot,1}^{(2,I,1)}, \dots, X_{\cdot,p}^{(2,I,1)})$$

Partial variances :

$$D_I = \frac{1}{n-1} \sum_{k=1}^n y_k^{(1)} y_k^{(2,I,1)} - \left(\frac{1}{n} \sum_{k=1}^n y_k^{(1)} \right)^2$$

Sobol index :

$$S_I = \frac{D_I - \sum_{\substack{J \subset \{1 \dots p\} \\ J \subsetneq I}} D_J}{\widehat{\text{var}}(y_{\cdot}^{(1)})}$$

3.2 method=saltelli02

The p more samples, noted $X^{(1,i,2)}$ ($i = 1 \dots p$) :

$$\begin{aligned} X_{\cdot,i}^{(1,i,2)} &= X_{\cdot,i}^{(2)} \\ X_{\cdot,j}^{(1,i,2)} &= X_{\cdot,j}^{(1)}, j \neq i \end{aligned}$$

The response :

$$y_{.i}^{(1,i,2)} = f(X_{.1}^{(1,i,2)}, \dots, X_{.p}^{(1,i,2)})$$

Partial variances :

$$D_i = \frac{1}{n-1} \sum_{k=1}^n y_k^{(2)} y_k^{(1,i,2)} - \frac{1}{n} \sum_{k=1}^n y_k^{(1)} y_k^{(2)}$$

$$D_i^{\text{tot}} = \frac{1}{n-1} \sum_{k=1}^n y_k^{(1)} y_k^{(1,i,2)} - \left(\frac{1}{n} \sum_{k=1}^n y_k^{(1)} \right)^2$$

First order and total indices :

$$S_i = \frac{D_i}{\widehat{\text{var}}(y_{.i}^{(1)})}$$

$$S_i^{\text{tot}} = 1 - \frac{D_i^{\text{tot}}}{\widehat{\text{var}}(y_{.i}^{(1)})}$$

4 fast

4.1 method=saltelli99

Maximum frequencies :

$$\omega_{\max} = \left\lfloor \frac{n-1}{2M} \right\rfloor$$

$$\omega'_{\max} = \left\lfloor \frac{\omega_{\max}}{2M} \right\rfloor$$

Frequencies (i = 1 ... p) :

$$\omega_i^{(i)} = \omega_{\max}$$

$$(\omega_j^{(i)})_{\substack{j=1 \dots p \\ j \neq i}} = (w_k)_{k=1 \dots p-1}$$

where

$$w_k = 1 + \left\lfloor (k-1) \frac{\omega'_{\max} - 1}{p-2} \right\rfloor \quad \text{if } \omega'_{\max} \geq p-1$$

$$w_k = 1 + ((k-1) \bmod \omega'_{\max}) \quad \text{if } \omega'_{\max} < p-1$$

Sampling (k = 1 ... n) :

$$s_k = \frac{2\pi(k-1)}{n}$$

$$X_{jk}^{(i)} = F_j^{-1} \left(\frac{1}{2} + \frac{1}{\pi} \arcsin(\sin(\omega_j^{(i)} s_k)) \right) \quad (j = 1 \dots p)$$

where F_j^{-1} is the inverse of the distribution function of the i th parameter.

Fourier coefficients ($j = 0 \dots n - 1$) :

$$c_j^{(i)} = \frac{1}{n} \sum_{k=1}^n f(X_{1k}^{(i)}, \dots, X_{pk}^{(i)}) e^{-is_k j}$$

Variance and partial variances ($i = 1 \dots p$) :

$$\begin{aligned} D^{(i)} &= \sum_{j=1}^{n-1} |c_j^{(i)}|^2 \\ D_i &= 2 \sum_{j=1}^M |c_{j\omega_i}^{(i)}|^2 \\ D_i^{\text{tot}} &= 2 \sum_{j=1}^{\omega_i/2} |c_j^{(i)}|^2 \end{aligned}$$

Sensitivity indices ($i = 1 \dots p$) :

$$\begin{aligned} S_i &= \frac{D_i}{D^{(i)}} \\ S_i^{\text{tot}} &= 1 - \frac{D_i^{\text{tot}}}{D^{(i)}} \end{aligned}$$

References

- [1] A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communication*, 145:280–297, 2002.
- [2] A. Saltelli, K. Chan, and E.M. Scott, editors. *Sensitivity analysis*. Wiley Series in Probability and Statistics. Wiley, 2000.
- [3] A. Saltelli, S. Tarantola, and K. Chan. A quantitative, model-independent method for global sensitivity analysis of model output. *Technometrics*, 41:39–56, 1999.