A survey on sensitivity analysis - Part II

Agnès Lagnoux

Institut de Mathématiques de Toulouse
TOULOUSE - FRANCE

Workshop - Inverse calibration of DEB numerical models for Indian Ocean tunas

Victoria, 26-29 May 2015

Plan de l'exposé

Sobol indices drawbacks

Sobol indices are based on L^2 decomposition and thus on order two methods...

1. Imagine an output Y that is a symmetric function of the inputs X_1 and X_2 that do not share the same distribution but have the same first four moments and satisfy $\mathbb{E}\left[X_1^5\right] \neq \mathbb{E}\left[X_2^5\right]$.
2. They are well adapted to measure the contribution of an input to the deviation around the mean of Y. However, it seems very intuitive that the sensitivity of an extreme quantile of Y could depend on sets of variables different from that highlighted when studying the variance sensitivity.

Hence the need to introduce a new sensitivity index that takes into account such an importance.

There are several ways to generalize the Sobol indices.

- A first natural way consists in considering higher moment Sobol indices.
- One can also define new indices through contrast functions based on the quantity of interest. Unfortunately the Monte Carlo estimator of these new indices are computationally expensive.
- Another way is to proceed considering the whole distribution of the output instead only of its second moment.

Indices based on a constrast function

- Let $Y = X_1 + X_2$ with $X_1 \sim \mathcal{E}(1)$, $X_2 = -X_1$ and X_1 ind. of X_2.
- Here $S^1_{Sob} = S^2_{Sob} = 1/2$.
- Assume we are interested in the α-quantile $q_Y(\alpha)$ of Y and its sensitivity with respect to X_1 and X_2.
- We propose to use a "contrast" adapted to the α-quantile

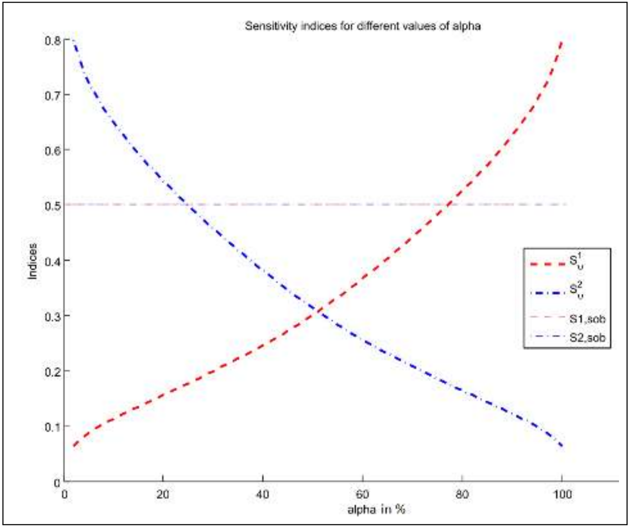
  $$\Psi(\theta) = \mathbb{E}(Y - \theta)(\alpha - \mathbf{1}_{Y \leq \theta})$$

  and define a new index

  $$S^k_\psi = \frac{\mathbb{E}\psi(Y; q_Y(\alpha)) - \mathbb{E}_{(X_k, Y)}\left(\psi(Y; q_{Y/X_k}(\alpha))\right)}{\mathbb{E}\psi(Y; q_Y(\alpha))},$$

  for $k = 1, 2$.
- Here, all the quantities can be explicitly computed.

Indices based on a constrast function



Indices based on a constrast function

As expected, we have

$$\begin{cases} S_\psi^1 < S_\psi^2 & \text{for} \quad \alpha < 1/2, \\ S_\psi^1 = S_\psi^2 & \text{for} \quad \alpha = 1/2, \\ S_\psi^1 > S_\psi^2 & \text{for} \quad \alpha > 1/2. \end{cases}$$

Moreover, naturally,

$$\begin{cases} \lim_{\alpha\to 1} S_\psi^1 = \lim_{\alpha\to 0} S_\psi^2 = 1, \\ \lim_{\alpha\to 0} S_\psi^1 = \lim_{\alpha\to 1} S_\psi^2 = 0. \end{cases}$$

On the contrary, the Sobol indices do not depend on α and do not include the fact that we are investigating quantiles.

Plan de l'exposé

We generalize the numerator of the classical Sobol index by considering higher order moments: for $p \geqslant 3$, we study

$$H_v^p := \mathbb{E}\left[(\mathbb{E}[Y|X_v] - \mathbb{E}[Y])^p\right]$$

instead of $H_v^2 := \mathrm{Var}(\mathbb{E}[Y|X_v]) = \mathbb{E}\left[(\mathbb{E}[Y|X_v] - \mathbb{E}[Y])^2\right]$.

We generalize the numerator of the classical Sobol index by considering higher order moments: for $p \geqslant 3$, we study

$$H_v^p := \mathbb{E}\left[(\mathbb{E}[Y|X_v] - \mathbb{E}[Y])^p\right]$$

instead of $H_v^2 := \mathrm{Var}(\mathbb{E}[Y|X_v]) = \mathbb{E}\left[(\mathbb{E}[Y|X_v] - \mathbb{E}[Y])^2\right]$.

The trick is to rewrite H_v^p as:

$$\mathbb{E}\left[(\mathbb{E}[Y|X_v] - \mathbb{E}[Y])^p\right] = \mathbb{E}\left[\prod_{i=1}^{p}\left(Y^{v,i} - \mathbb{E}[Y]\right)\right]$$

such as

$$H_v^2 := \mathrm{Var}(\mathbb{E}[Y|X_v]) = \mathrm{Cov}(Y, Y^v) = \mathbb{E}[(Y - \mathbb{E}[Y])\,(Y^v - \mathbb{E}[Y])].$$

Here, $Y^{v,1} = Y$ and for $i = 2, \ldots, p$, $Y^{v,i}$ is constructed independently as Y^v.

Design of experiments and estimation phase

In view of the estimation of H_v^p,

1. we first develop the product in H_v^p:

$$H_v^p = \sum_{l=0}^{p} \binom{p}{l} (-1)^{p-l} \mathbb{E}[Y]^{p-l} \, \mathbb{E}\left[\prod_{i=1}^{l} Y^{v,i}\right].$$

   with the usual convention $\prod_{i=1}^{0} Y^{v,i} = 1$.

2. we use the Pick and Freeze design of experiment constituted by the following $p \times N$-sample

$$\left(Y_j^{v,i}\right)_{(i,j)\in I_p \times I_N}$$

   to estimate all the expectations.

Properties of H_p^v

1. H_v^p is only non negative for even p.
2. For any p,
$$|H_v^p| \leqslant \mathbb{E}\left[|Y - \mathbb{E}[Y]|^p\right].$$
3. As the classical Sobol index, H_v^p is still invariant by translation of the output.

Properties of H_p^v

1. H_v^p is only non negative for even p.
2. For any p,
$$|H_v^p| \leqslant \mathbb{E}\left[|Y - \mathbb{E}[Y]|^p\right].$$
3. As the classical Sobol index, H_v^p is still invariant by translation of the output.

Asymptotic properties of $H_{p,N}^v$

$H_{p,N}^v$ is consistent and asymptotically normal:

$\sqrt{N}\left(H_{p,N}^v - H_v^p\right)$ converges in distribution to a Gaussian rv whose variance can be explicitly computed.

Drawbacks of $H_{p,N}^{v}$

The collection of all indices H_v^p is much more informative than the classical Sobol index.

Nevertheless it has several drawbacks: it may be negative when p is odd. To overcome this fact, we may have introduced $\mathbb{E}\left[|\mathbb{E}[Y|X_i, i \in v] - \mathbb{E}[Y]|^p\right]$ but proceeding in such a way, we would have loose the Pick and Freeze estimation procedure.

The Pick and Freeze estimation procedure is computationally expensive: it requires a $p \times N$ sample of the output Y. In a sense, if we want to have a good idea of the influence of an input on the law of the output, we need to estimate the first d indices H_v^p and hence we need to run the black-box code $K \times N$ times.

Plan de l'exposé

The code will be denoted by $Z = f(X_1, \dots, X_d) \in \mathbb{R}$.

Let F be the distribution function of Z

$$F(t) = \mathbb{P}(Z \leqslant t) = \mathbb{E}\left[\mathbb{1}_{\{Z \leqslant t\}}\right]$$

and $F^v(t)$ the conditional distribution function of Z conditionally on X_v:

$$F^v(t) = \mathbb{P}(Z \leqslant t | X_v,) = \mathbb{E}\left[\mathbb{1}_{\{Z \leqslant t\}} | X_v\right].$$

It is obvious that $\mathbb{E}[F^v(t)] = F(t)$.

We apply the framework presented previously with $Y(t) = \mathbb{1}_{\{Z \leqslant t\}}$ and $p = 2$.

We then have a consistent and asymptotically normal estimation procedure for the estimation of

$$\mathbb{E}\left[(F(t) - F^v(t))^2\right].$$

We define a Cramér Von Mises type distance of order 2 between $\mu := \mathcal{L}(Z)$ and $\mathcal{L}(Z|X_v)$ by

$$\begin{aligned} D^v_{2,CVM} &:= \int_{\mathbb{R}} \mathbb{E}\left[(F(t) - F^v(t))^2\right] d\mu(t) \\ &= \mathbb{E}\left[\mathbb{E}\left[(F(Z) - F^v(Z))^2\right]\right]. \end{aligned}$$

The aim of the rest of the section is dedicated to the estimation of $D^v_{2,CVM}$.

Design of experiments and estimation phase

We consider the following design of experiments consisting in:

1. two N-samples of Z: $(Z_j^{v,1}, Z_j^{v,2})$, $1 \leqslant j \leqslant N$;
2. a third N-sample of Z independent of $(Z_j^{v,1}, Z_j^{v,2})_{1 \leqslant j \leqslant N}$: W_k, $1 \leqslant k \leqslant N$.

Design of experiments and estimation phase

We consider the following design of experiments consisting in:

1. two N-samples of Z: $(Z_j^{v,1}, Z_j^{v,2})$, $1 \leqslant j \leqslant N$;
2. a third N-sample of Z independent of $(Z_j^{v,1}, Z_j^{v,2})_{1\leqslant j\leqslant N}$: W_k, $1 \leqslant k \leqslant N$.

The natural estimator of $D_{2,CVM}^v$ is then given by

$$\widehat{D}_{2,CVM}^v = \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}}\mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}} - \left[\frac{1}{2N}\sum_{j=1}^{N}\left(\mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}}+\mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}}\right)\right]^2\right\}.$$

Properties of $D^v_{2,CVM}$

1. $0 \leqslant D^v_{2,CVM} \leqslant \frac{1}{4}$.
   Moreover, if F is continuous, we have $0 \leqslant D^v_{2,CVM} \leqslant \frac{1}{6}$.
2. As the classical Sobol index, $D^v_{2,CVM}$ is still invariant by translation, by left-composition by any nonzero scaling of Y and by left-composition of Y by any isometry.

Properties of $D^v_{2,CVM}$

1. $0 \leqslant D^v_{2,CVM} \leqslant \frac{1}{4}$.
   Moreover, if F is continuous, we have $0 \leqslant D^v_{2,CVM} \leqslant \frac{1}{6}$.
2. As the classical Sobol index, $D^v_{2,CVM}$ is still invariant by translation, by left-composition by any nonzero scaling of Y and by left-composition of Y by any isometry.

Asymptotic properties of $\widehat{D}^v_{2,CVM}$

$\widehat{D}^v_{2,CVM}$ is consistent and asymptotically normal:

$\sqrt{N}\left(\widehat{D}^v_{2,CVM} - D^v_{2,CVM}\right)$ converges in distribution to a Gaussian rv whose variance can be explicitly computed.

Plan de l'exposé

Alternative definitions for measuring the strength of the statistical dependence of Y on X_k have been proposed, giving rise to the class of distribution-based sensitivity measures.

They define the importance of X_k as the distance between the unconditional distribution of Y and its conditional distribution.

These sensitivity measures are defined both in the presence and absence of correlations.

We present three examples of such sensitivity measures.

- the δ importance measure based on the L^1-norm between densities:

$$\delta_k = \frac{1}{2}\mathbb{E}\left[\int |p_Y(y) - p_{Y|X_k}(y)|\, dy\right]$$

where $p_Y(y)$ and $p_{Y|X_k}(y)$ stands respectively for the density function of Y and $Y|X_k$. Notice that $\delta_k = 0$ if and only if Y is independent of X_k.

- the β^{KS} sensitivity measure based on the Kolmogorov-Smirnov separation between cumulative distributions functions:

$$\beta_k = \mathbb{E}\left[\sup_y |F_Y(y) - F_{Y|X_k}(y)|\right].$$

Both sensitivity measures δ_k and β_k are monotonic transformation invariant.

- the θ probabilistic sensitivity measure based on the family of Shannon's cross-entropy:

$$\theta_k = \mathbb{E}\left[\int |p_{Y|X_k}(y)(\log p_{Y|X_k}(y) - \log p_{Y|X_k}(y))|\, dy\right].$$

θ_k can be interpreted as value of information sensitivity measures as the classical Sobol index for more details). We will see in the next subsection that this measure is part of a larger lass of sensitivity measures based on dissimilarity distances and the family of Csiszár's divergences.

A common rationale

Variance-based and the previous distribution-based sensitivity measures have a common conceptual aspect. In information theory, the distributions are statistical signals. In probabilistic sensitivity analysis, they are the conditional and unconditional model output distributions, $\mathbb{P}_Y$ and $\mathbb{P}_{Y|X_k=x_k}$. We call

$$\xi_k = \mathbb{E}[\gamma_k(X_k)] = \mathbb{E}[\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X_k})] \tag{1}$$

the global sensitivity measure of X_k based on operator $\zeta(\cdot,\cdot)$ and $\gamma_k(x_k)$ the inner statistic of ξ_k.

A common rationale

The above framework accommodates the definitions of the probabilistic sensitivity measures described previously. Selecting as inner operators

- $\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X_k}) = \mathbb{E}[(Y - \mathbb{E}[Y])^2 | X_k = x_k] = \mathbb{E}[(Y - \mathbb{E}[Y|X_k])^2 | X_k = x_k]$;
- $\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X_k}) = \frac{1}{2} \int |p_Y(y) - p_{Y|X_k}(y)| \, dy$;
- $\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X_k}) = \sup_y |F_Y(y) - F_{Y|X_k}(y)|$;
- $\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X_k}) = \int |p_{Y|X_k}(y)(\log p_{Y|X_k}(y) - \log p_{Y|X_k}(y))| \, dy$;

we obtain the inner statistics of the classical Sobol index, δ_k, β_k and θ_k respectively.

Plan de l'exposé

Motivation

A natural way of defining the impact of a given input X_k on Y is to consider a function which measures the similarity between the distribution of Y and that of $Y|X_k$. More precisely, the impact of X_k on Y is given by

$$S_{X_k} = \mathbb{E}_{X_k}[d(Y, Y|X_k)]$$

where $d(\cdot,\cdot)$ denotes a dissimilarity measure between two random variables.

The advantage of such a formulation is that many choices for d are available and we will see in what follows that some natural dissimilarity measures yield sensitivity indices related to well known quantities.

Motivation

A natural way of defining the impact of a given input X_k on Y is to consider a function which measures the similarity between the distribution of Y and that of $Y|X_k$. More precisely, the impact of X_k on Y is given by

$$S_{X_k} = \mathbb{E}_{X_k}[d(Y, Y|X_k)]$$

where $d(\cdot, \cdot)$ denotes a dissimilarity measure between two random variables.

The advantage of such a formulation is that many choices for d are available and we will see in what follows that some natural dissimilarity measures yield sensitivity indices related to well known quantities.

Let us note that the naïve dissimilarity measure

$$d(Y, Y|X_k) = (E[Y] - E[Y|X_k])^2$$

where rv are compared only through their mean values produces the unnormalized Sobol first-order sensitivity index $S^1_{X_k} = \mathrm{Var}(\mathbb{E}[Y|X_k])$.

Dissimilarity measures

Assuming all input random variables have an absolutely continuous distribution with respect to the Lebesgue measure on $\mathbb{R}$, the f-divergence between Y and $Y|X_k$ is given by

$$d_h(Y||Y|X_k) = \int_{\mathbb{R}} h\left(\frac{p_Y(y)}{p_{Y|X_k}(y)}\right) p_{Y|X_k}(y)\, dy$$

where h is a convex function such that $h(1) = 0$ and p_Y and $p_{Y|X_k}$ are the probability distribution functions of Y and $Y|X_k$, respectively.

Dissimilarity measures

Assuming all input random variables have an absolutely continuous distribution with respect to the Lebesgue measure on $\mathbb{R}$, the f-divergence between Y and $Y|X_k$ is given by

$$d_h(Y||Y|X_k) = \int_{\mathbb{R}} h\left(\frac{p_Y(y)}{p_{Y|X_k}(y)}\right) p_{Y|X_k}(y)\, dy$$

where h is a convex function such that $h(1) = 0$ and p_Y and $p_{Y|X_k}$ are the probability distribution functions of Y and $Y|X_k$, respectively.

Standard choices for the function h include for example

- Kullback-Leibler divergence: $h(t) = -\ln(t)$ or $h(t) = t\ln(t)$;
- Hellinger distance: $h(t) = (\sqrt{t} - 1)^2$;
- Total variation distance: $h(t) = |t - 1|$;
- Pearson χ^2 divergence: $h(t) = (t - 1)^2$ or $h(t) = t^2 - 1$;
- Neyman χ^2 divergence: $h(t) = (t - 1)^2/t$ or $h(t) = (1 - t^2)/t$.

From dissimilarity measures to sensitivity indices

Plugging this dissimilarity measure into the definition yields the following sensitivity index:

$$S^h_{X_k} = \int_{\mathbb{R}^2} h\left(\frac{p_Y(y)p_{X_k}(x)}{p_{X_k,Y}(x,y)}\right) p_{X_k,Y}(x,y)\,dxdy$$

where p_{X_k} and $p_{X_k,Y}$ are the probability distribution functions of X_k and (X_k, Y), respectively.

Properties

1. $S^h_{X_k} \geqslant 0$ and $S^h_{X_k} = 0$ iff Y and X_k are independent,
2. $S^h_{X_k}$ is invariant under any smooth and uniquely invertible transformation of the variables X_k and Y. This is a major advantage over variance-based Sobol sensitivity indices, which are only invariant under linear transformations.

Examples

- The total variation distance with $h(t) = |t-1|$ gives a sensitivity index equal to:
$$S^h_{X_k} = \int_{\mathbb{R}^2} |p_Y(y)p_{X_k}(x) - p_{X_k,Y}(x,y)|\, dxdy.$$
- The Kullback-Leibler divergence with $h(t) = -\ln(t)$ yields
$$S^h_{X_k} = \int_{\mathbb{R}^2} p_{X_k,Y}(x,y) \ln\left(\frac{p_{X_k,Y}(x,y)}{p_Y(y)p_{X_k}(x)}\right) dxdy,$$
that is the mutual information $I(X_k; Y)$ between X_k and Y.
- The Neyman χ^2 divergence with $h(t) = (1-t^2)/t$ leads to
$$S^h_{X_k} = \int_{\mathbb{R}^2} p_{X_k,Y}(x,y) \ln\left(\frac{p_{X_k,Y}(x,y)}{p_Y(y)p_{X_k}(x)}\right) dxdy,$$
which is the so-called squared-loss mutual information between X_k and Y (or mean square contingency).

The goal is to estimate

$$S_{X_k}^h = \int_{\mathbb{R}^2} h\left(\frac{1}{r(x,y)}\right) p_{X_k,Y}(x,y)\,dxdy = \mathbb{E}_{(X_k,Y)}\left[h\left(\frac{1}{r(X_k,Y)}\right)\right]$$

where $r(x,y) = p_{X_k,Y}(x,y)/(p_Y(y)p_{X_k}(x))$ is the ratio between the joint density of (X_k, Y) and the marginals.

The goal is to estimate

$$S^h_{X_k} = \int_{\mathbb{R}^2} h\left(\frac{1}{r(x,y)}\right) p_{X_k,Y}(x,y)\,dxdy = \mathbb{E}_{(X_k,Y)}\left[h\left(\frac{1}{r(X_k,Y)}\right)\right]$$

where $r(x,y) = p_{X_k,Y}(x,y)/(p_Y(y)p_{X_k}(x))$ is the ratio between the joint density of (X_k, Y) and the marginals.

Straightforward estimation is possible if one estimates the densities $p_{X_k,Y}(x,y)$, $p_{X_k}(x)$ and $p_Y(y)$ with e.g. kernel density estimators. However, it is well known that density estimation suffers from the curse of dimensionality.

Besides, since only the ratio function $r(x,y)$ is needed, we expect more robust estimates by focusing only on it.

Design of experiments and estimation phase

Let us assume now that we have a sample $(X_{k,i}, Y_i)$ for $i = 1, ..., N$ of (X_k, Y), the idea is to build first an estimate $\hat{r}(x, y)$ of the ratio.

Powerful estimating methods for ratios include

1. maximum-likelihood estimation,
2. unconstrained least-squares importance fitting,
3. k-nearest neighbors strategy dedicated to mutual information...

Design of experiments and estimation phase

Let us assume now that we have a sample $(X_{k,i}, Y_i)$ for $i = 1, \dots, N$ of (X_k, Y), the idea is to build first an estimate $\hat{r}(x, y)$ of the ratio.

Powerful estimating methods for ratios include

1. maximum-likelihood estimation,
2. unconstrained least-squares importance fitting,
3. k-nearest neighbors strategy dedicated to mutual information...

The final estimator $\hat{S}^h_{X_k}$ of $S^h_{X_k}$ will then be given by

$$\hat{S}^h_{X_k} = \frac{1}{N} \sum_{i=1}^{N} h\left(\frac{1}{\hat{r}(X_{k,i}, Y_i)}\right).$$

Bibliography

F. Gamboa, T. Klein and A. Lagnoux. "Sensitivity analysis based on Cramér von Mises distance", *Submitted*, 2015.

E. Borgonovo and B. Iooss. "Moment Independent Importance Measures and a Common Rationale", *Submitted*, 2014.

S. Da Veiga. "Global sensitivity analysis with dependence measures", *J. Stat. Comput. Simul.*, 2015.