Computing first-order sensitivity indices with contribution to the sample mean plot

Saint-Geours Nathalie (1)  Tarantola, S. (2)
Kopustinskas, V. (2)  Bolado-Lavin, R. (3)

SAMO 2013 - Nice, July 1-4

(1) Irstea UMR TETIS, Montpellier, France
(2) Joint Research Center of the European Commission, Ispra, Italy
(3) Joint Research Center of the European Commission, Petten, Netherlands
Sensitivity analysis from given data
Sensitivity analysis from given data

1. sampling inputs
2. running the model

MODEL EXPERT
Sensitivity analysis from **given data**

1. **Sampling inputs**
2. **Running the model**
3. **Sensitivity analysis from given data?**
Sensitivity analysis from **given data**

1. **sampling inputs**
   - **MODEL EXPERT**

2. **running the model**
   - no specific Design of Experiments!
   - often **simple random sampling**

3. **sensitivity analysis from given data?**
   - **STATISTICIAN (YOU)**

---

**SAMO 2013 Nice July 1-4**

Saint-Geours et al.
Sensitivity analysis from **given data**

1. **sampling inputs**

2. **running the model**

3. **sensitivity analysis from given data?**

   - no specific Design of Experiments!
   - often **simple random sampling**

   one possible SA way:

   - Contribution to the Sample Mean Plot (CSM)

---

**MODEL EXPERT**

**STATISTICIAN (YOU)**
## Contribution to the Sample Mean (CSM)

Bolado-Lavin, R., Castaings, W., & Tarantola, S.

Contribution to the sample mean plot for graphical and numerical sensitivity analysis

Contribution to the Sample Mean (CSM)

Bolado-Lavin, R., Castaings, W., & Tarantola, S.

Contribution to the sample mean plot for graphical and numerical sensitivity analysis


- Model $Y = G(X_1, \ldots, X_m)$
- $X_1, \ldots, X_m$ independent random variables, pdf $p_j$, cdf $F_j$
- The **Contribution to the Sample Mean (CSM)** for $X_j$ is:

  $\forall q \in [0; 1],$

  $$C_j(q) = \int_{-\infty}^{F_j^{-1}(q)} \left( \int_{\mathbb{R}^{m-1}} G(x)p_{X_{\sim j}}(x_{\sim j})dx_{\sim j} \right) p_j(x_j)dx_j$$

  $\frac{\int_{\mathbb{R}^m} G(x)p_X(x)dx}{\int_{\mathbb{R}^m} G(x)p_X(x)dx}$

  (1)
Contribution to the Sample Mean (CSM)

Bolado-Lavin, R., Castaings, W., & Tarantola, S.
Contribution to the sample mean plot for graphical and numerical sensitivity analysis

- Model \( Y = G(X_1, \ldots, X_m) \)
- \( X_1, \ldots, X_m \) independent random variables, pdf \( p_j \), cdf \( F_j \)
- The **Contribution to the Sample Mean (CSM)** for \( X_j \) is:
  \[
  \forall q \in [0; 1],
  \begin{align*}
  C_j(q) &= \frac{F_j^{-1}(q)}{\int_{\mathbb{R}^m} G(x) p_{X_{\sim j}}(x_{\sim j}) dx_{\sim j}} \\
  &= \frac{\int_{-\infty}^{F_j^{-1}(q)} \left( \int_{\mathbb{R}^{m-1}} G(x) p_{X_{\sim j}}(x_{\sim j}) dx_{\sim j} \right) p_j(x_j) dx_j}{\int_{\mathbb{R}^m} G(x) p_X(x) dx}
  \end{align*}
  \tag{1}
  \]
Contribution to the Sample Mean (CSM)

\[ C_j(q) \] represents the fraction of the output mean due to the fraction \( q \) of smallest values of \( X_j \).
Contribution to the Sample Mean (CSM)

Procedure to approximate CSM plot from a set of \( n \) model runs. input sample \( (x_{ij})_{i=1...n,j=1...m} \) and output vector \( (y_i)_{i=1...n} \)
Contribution to the Sample Mean (CSM)

Procedure to approximate CSM plot from a set of $n$ model runs.
input sample $(x_{ij})_{i=1...n,j=1...m}$ and output vector $(y_i)_{i=1...n}$

1. compute the output mean $\hat{\mu}$
2. sort increasingly the $n$ random realisations of $X_j$:
   $$x_{\pi(1)j} \leq \cdots \leq x_{\pi(n)j}$$
3. compute $(c_1, \ldots, c_n)$:
   $$c_i = \frac{1}{n\hat{\mu}} \sum_{s=1}^{i} y_{\pi(s)}$$
Contribution to the Sample Mean (CSM)

Procedure to approximate CSM plot from a set of $n$ model runs. 
input sample $(x_{ij})_{i=1\ldots n, j=1\ldots m}$ and output vector $(y_i)_{i=1\ldots n}$

1. compute the output mean $\hat{\mu}$
2. sort increasingly the $n$ random realisations of $X_j$: 
   
   $$x_{\pi(1)} \leq \cdots \leq x_{\pi(n)}$$

3. compute $(c_1, \ldots, c_n)$:
   
   $$c_i = \frac{1}{n\hat{\mu}} \sum_{s=1}^{i} y_{\pi(s)}$$

4. plot $(c_1, \ldots, c_n)$ against $(q_1, \ldots, q_n)$ with $q_i = i/n$
Contribution to the Sample Mean (CSM)

CSM and first-order effects

$X_j$ with low first-order effect

CSM line close to the diagonal

(Bolado-Lavin et al., 2009)
Contribution to the Sample Mean (CSM)

CSM and first-order effects

\( X_j \) with low first-order effect

\( \sim \)

CSM line close to the diagonal

(Bolado-Lavin et al., 2009)
Contribution to the Sample Mean (CSM)

2 research questions

1. What relationship between CSM plot and $S_j$?
2. Is it possible to compute $S_j$ from a CSM plot?

CSM and first-order effects

$X_j$ with low first-order effect

$\sim$

CSM line close to the diagonal

(Bolado-Lavin et al., 2009)
1st question

What relation between CSM plot and first-order sensitivity indices $S_j$?
Relation between CSM and first-order indices $S_j$

**Property**

Let denote $c_v = \sigma(Y)/E(Y)$.

For any input $X_j$ we have:

$$S_j = \frac{1}{c_v^2} \cdot \int_0^1 \left[ \frac{d}{dq} (C_j(q) - q) \right]^2 dq \quad (2)$$
Property

Let denote \( c_v = \sigma(Y)/E(Y) \).

For any input \( X_j \) we have:

\[
S_j = \frac{1}{c_v^2} \cdot \int_0^1 \left[ \frac{d}{dq} \frac{C_j(q) - q}{\text{deviation to diagonal}} \right]^2 dq
\]  
\( (2) \)
Elements of proof
Elements of proof

**CSM expression using conditional expectation**

\[ \forall q \in [0; 1], \quad C_j(q) = \frac{1}{E(Y)} \int_{-\infty}^{F_j^{-1}(q)} E[Y | X_j = x_j] p_j(x_j) dx_j \]
Elements of proof

CSM expression using conditional expectation

\[ \forall q \in [0; 1], \quad C_j(q) = \frac{1}{E(Y)} \int_{-\infty}^{F_j^{-1}(q)} \mathbb{E}[Y \mid X_j = x_j] p_j(x_j) \, dx_j \]

CSM derivative

Using \[ \frac{d}{dq}(F_j^{-1}(q)) = \frac{1}{p_j(F_j^{-1}(q))} \]:

\[ \forall q \in [0; 1], \quad \frac{d}{dq} C_j(q) = \frac{\mathbb{E}[Y \mid X_j = F_j^{-1}(q)]}{E(Y)} \]
Elements of proof

CSM expression using conditional expectation

\[ \forall q \in [0; 1], \quad C_j(q) = \frac{1}{E(Y)} \int_{-\infty}^{F_j^{-1}(q)} E[Y \mid X_j = x_j] p_j(x_j) dx_j \]

CSM derivative

Using \[ \frac{d}{dq}(F_j^{-1}(q)) = 1/p_j(F_j^{-1}(q)) \]:

\[ \forall q \in [0; 1], \quad \frac{d}{dq} C_j(q) = \frac{S_j = \text{Var}[E(Y \mid X_j)] / V(Y)}{E(Y)} \]

\[ \begin{array}{c}
\text{SAMO 2013 Nice July 1-4} \\
Saint-Geours et al.
\end{array} \]
### 2nd question

**Computing first-order effects $S_j$ from a CSM plot?**
Computing $S_j$ from a CSM plot

Start from a sample of $n$ CSM points $(q_i, c_i)_{i=1,...,n}$.

A. Polynomial regression
- fit a polynomial model on CSM points $(q_i, c_i)$
- exact formula for $S_j$ from the regression coefficients

B. Spline smoothing
- fit a spline model on the CSM points
- approximate CSM derivative
- compute $S_j$ using Eqn.(2)
Computing $S_j$ from a CSM plot

Start from a sample of $n$ CSM points $(q_i, c_i)_{i=1,...,n}$.

A. Polynomial regression

- fit a polynomial model on CSM points $(q_i, c_i)$
- exact formula for $S_j$ from the regression coefficients

B. Spline smoothing

- fit a spline model on the CSM points
- approximate CSM derivative
- compute $S_j$ using Eqn.(2)
Computing $S_j$ from a CSM plot

Start from a sample of $n$ CSM points $(q_i, c_i)_{i=1,...,n}$.

A. Polynomial regression

- fit a polynomial model on CSM points $(q_i, c_i)$
- exact formula for $S_j$ from the regression coefficients

B. Spline smoothing

- fit a spline model on the CSM points
- approximate CSM derivative
- compute $S_j$ using Eqn.(2)
Computing $S_j$ from a CSM plot

Start from a sample of $n$ CSM points $(q_i, c_i)_{i=1,...,n}$.

A. Polynomial regression

- fit a polynomial model on CSM points $(q_i, c_i)$
- exact formula for $S_j$ from the regression coefficients

B. Spline smoothing

- fit a spline model on the CSM points
- approximate CSM derivative
- compute $S_j$ using Eqn.(2)
Polynomial regression
Polynomial regression

- expansion of CSM using **shifted Legendre polynomials** \( (P_k)_{k \in \mathbb{N}} \) which are orthogonal on \([0, 1]\)

\[
\forall i = 1 \ldots n, \quad c_i = \sum_{k=0}^{d} \alpha_k P_k(q_i) + \epsilon_i
\]
Polynomial regression

- expansion of CSM using **shifted Legendre polynomials** \((P_k)_{k \in \mathbb{N}}\)
  which are orthogonal on \([0, 1]\)

\[
\forall i = 1 \ldots n, \quad c_i = \sum_{k=0}^{d} \alpha_k P_k(q_i) + \epsilon_i
\]

- **selecting max order** \(d^*\) by minimizing AICc information criterion

\[
d^* = \arg \min_{d \in \mathbb{N}} \left[ \frac{n}{2} \cdot \log \left( \frac{2\pi}{n} \sum_{i=1}^{n} \epsilon_i(d)^2 \right) + \frac{n}{2} + \frac{n \cdot (d + 2)}{n - d - 3} \right]
\]
Polynomial regression

- **explicit formula for** $S_j$ **derived from Eqn.(2) using** $P_k$ **properties**

with:

\[
\tilde{\alpha}_k = \begin{cases} 
\alpha_k & \text{if } k > 1, \\
\alpha_k - \frac{1}{2} & \text{if } k = 1
\end{cases}
\]

we obtain:

\[
\hat{S}_j = \frac{2}{\hat{c}_v^2} \sum_{\substack{k,l=1 \atop k+l \in 2\mathbb{Z}}}^{d} \tilde{\alpha}_k \tilde{\alpha}_l \cdot \min(k, l) \left[1 + \min(k, l)\right]
\] (3)
3rd point

Numerical test cases
Test cases

1. Ishigami function
   - $X_1$ to $X_3$ i.i.d $\sim U[-\pi, \pi]$
   - $Y = \sin(X_1) + a \cdot \sin(X_2)^2 + b \cdot X_3^4 \cdot \sin(X_1)$

2. G-Sobol function
   - $X_1$ to $X_8$ i.i.d $\sim U[0, 1]$
   - fixed parameter vector $a = (0, 1, 4.5, 9, 99, 99, 99, 99)$:
   - $Y = \prod_{j=1}^{8} \frac{|4X_j - 2| + a_j}{1 + a_j}$
Scatterplots and CSM plots

sample size $n = 300$ (simple random sample)
Polynomial fit for input $X_1$

Sample size $n = 300$ (simple random sample)
Estimation of first-order effects

Convergence of $\hat{S}_1$ for increasing sample size $n$
Estimation of first-order effects

Convergence of $\hat{S}_2$ and $\hat{S}_3$ for increasing sample size $n$
Conclusion
Conclusion

Results

- explicit formula linking $S_j$ and CSM (derivative)

- $\hat{S}_j$ estimator based on polynomial expansion of the CSM plot (explicit formula from regression coefficients)

  $\rightarrow$ computation of $S_j$ from given data
Conclusion

+ Results

- explicit formula linking $S_j$ and CSM (derivative)

- $\hat{S}_j$ estimator based on polynomial expansion of the CSM plot
  (explicit formula from regression coefficients)
  $\rightarrow$ computation of $S_j$ from given data

- Limits

- minimum sample size $n \sim 1000$

- $\hat{S}_j$ does not compare well with other estimators based on given data such as EASI (Plischke, 2010)

- why? because it requires approximating derivatives
Conclusion

→ Further research

- Total-order effects?
  - Contribution to the Sample Variance (CSV plot)
  - first attempts were unsuccessful but...

Tarantola S., V. Kopustinskis, R. Bolado-Lavin, A. Kaliatka, E. Uspuras, M. Vaisnoras
Sensitivity analysis using contribution to sample variance plot: Application to a water hammer model
Thank you for your attention!

**Funding** (6 weeks stay in JRC, Ispra, Italy):
Appendix
References

Bolado-Lavin, R., Castaings, W., & Tarantola, S.
Contribution to the sample mean plot for graphical and numerical sensitivity analysis

Abramowitz, M. & Segun, I. (eds.)
Handbook of mathematical functions with Formulas, Graphs, and Mathematical Tables
1972, New York: Dover Publications
First-order variance-based sensitivity indices

\[ S_j = \frac{\text{Var}_{X_j} \left( E_{X \sim j} [Y | X_j] \right)}{\text{V}(Y)} \]  
(Saltelli et al., 2008)
First-order variance-based sensitivity indices

\[ S_j = \frac{\text{Var}_{X_j} (\mathbf{E}_{X \sim j} [Y | X_j])}{V(Y)} \]  

(Saltelli et al., 2008)
First-order variance-based sensitivity indices

\[ S_j = \frac{\text{Var}_{X_j} (E_{X \sim j} [Y | X_j])}{\text{V}(Y)} \]

\[ = \frac{1}{\text{V}(Y)} \int_{\mathbb{R}} \left( E [Y | X_j = x_j] - E(Y) \right)^2 p_j(x_j) dx_j \]

(Saltelli et al., 2008)
First-order variance-based sensitivity indices

\[ S_j = \frac{\text{Var}_{X_j} \left( E_{X_{\sim j}} [Y | X_j] \right)}{\text{V}(Y)} \]

\[ = \frac{1}{\text{V}(Y)} \int_{\mathbb{R}} \left( E[Y | X_j = x_j] - E(Y) \right)^2 p_j(x_j) dx_j \]

(Saltelli et al., 2008)
First-order variance-based sensitivity indices

\[ S_j = \frac{\text{Var}_{X_j} \left( \mathbb{E}_{X \sim j} [Y \mid X_j] \right)}{\mathbb{V}(Y)} \] (Saltelli et al., 2008)

\[ = \frac{1}{\mathbb{V}(Y)} \int_{[0,1]} \left( \mathbb{E} [Y \mid X_j = F_j^{-1}(q)] - \mathbb{E}(Y) \right)^2 dq \]
First-order variance-based sensitivity indices

\[ S_j = \frac{\text{Var}_{X_j} \left( \mathbb{E}_{X \sim j} [Y \mid X_j] \right)}{\text{V}(Y)} \]

(Saltelli et al., 2008)

\[ = \frac{1}{\text{V}(Y)} \int_{[0,1]} \left( \mathbb{E} [Y \mid X_j = F_j^{-1}(q)] - \mathbb{E}(Y) \right)^2 dq \]
First-order variance-based sensitivity indices

\[ S_j = \frac{\text{Var}_{X_j} \left( \mathbb{E}_{X \sim X_j} [Y \mid X_j] \right)}{\mathbb{V}(Y)} \]  
(Saltelli et al., 2008)

\[ = \frac{1}{\mathbb{V}(Y)} \int_{[0,1]} \left( \frac{d}{dq} C_j(q) \cdot \mathbb{E}(Y) - \mathbb{E}(Y) \right)^2 dq \]
Elements of proof (2)

First-order variance-based sensitivity indices

\[ S_j = \frac{\text{Var}_{X_j} (E_{X_{\sim j}} [Y | X_j])}{V(Y)} \]  
(Saltelli et al., 2008)

\[ = \frac{1}{V(Y)} \int_{[0,1]} \left( \frac{d}{dq} C_j(q) \cdot E(Y) - E(Y) \right)^2 dq \]

\[ = \frac{E(Y)^2}{V(Y)} \int_{0}^{1} \left[ \frac{d}{dq} C_j(q) - 1 \right]^2 dq \]
First-order variance-based sensitivity indices

\[
S_j = \frac{\text{Var}_{X_j} \left( E_{X \sim j} [Y \mid X_j] \right)}{V(Y)} \quad \text{(Saltelli et al., 2008)}
\]

\[
= \frac{1}{V(Y)} \int_{[0,1]} \left( \frac{d}{dq} C_j(q) \cdot E(Y) - E(Y) \right)^2 dq
\]

\[
= \frac{E(Y)^2}{V(Y)} \int_0^1 \left[ \frac{d}{dq} C_j(q) - 1 \right]^2 dq
\]

\[
= \frac{1}{C_v^2} \int_0^1 \left[ \frac{d}{dq} \left( C_j(q) - q \right) \right]^2 dq
\]
G-Sobol test case \((n = 300)\)

Scatterplots and CSM plots
G-Sobol test case

Convergence of $\hat{S}_1$ and $\hat{S}_4$ for increasing sample size $n$
Estimation of the coefficient of variation

Set of CSM points \( (q_i, c_i)_{i=1...n} \)

Coefficient of variation \( c_v = \sigma(Y)/E(Y) \)

Using \( c_i - c_{i-1} = y_{\pi(i)}/(n\hat{\mu}) \) we get:

\[
\hat{c}_v = n \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} (c_{i+1} - c_i - \frac{1}{n})^2}
\]

(4)
Shifted Legendre polynomials

Shifted Legendre polynomial $P_k$ are defined by

$$P_k(q) = P_k^{(s)}(2q - 1)$$

with $P_k^{(s)}$ the standardized Legendre polynomials, which are given by the Rodrigue’s formula [2, p.785, Eqn. 22.11.5]:

$$\forall k \in \mathbb{N}, \forall q \in [-1, 1], \quad P_k^{(s)}(q) = \frac{(-1)^k}{2^k \cdot k!} \frac{d^k}{dq^k} [(q^2 - 1)^k]$$
Detailed proof for Eqn.(3)

Using the approximation \( C(q) \approx \sum_k \alpha_k P_k(q) \), we get an approximation of the integral \( I = \int_0^1 \frac{d}{dq} (C(q) - q)^2 \) dq:

\[
\hat{I} = \int_0^1 \left[ \left( \sum_{k=1}^d \alpha_k P_k'(q) \right) - 1 \right]^2 dq
\]

We use the fact that \( P_1'(q) = 2 \) to define modified coefficients \( (\tilde{\alpha}_k)_{k=1,\ldots,d} \) as equal to coefficients \( (\alpha_k)_{k=1,\ldots,d} \) except for \( \tilde{\alpha}_1 = \alpha_1 - \frac{1}{2} \),:

\[
\hat{I} = \int_0^1 \left[ \sum_{k=1}^d \tilde{\alpha}_k P_k'(q) \right]^2 dq
\]

\[
= \sum_{k,l=1}^d \tilde{\alpha}_k \tilde{\alpha}_l \int_0^1 P_k'(q) P_l'(q) dq
\]

\[
= \sum_{k,l=1}^d \tilde{\alpha}_k \tilde{\alpha}_l I_{kl}
\]
Detailed proof for Eqn.(3)

Using the approximation \( C(q) \approx \sum_k \alpha_k P_k(q) \), we get an approximation of the integral \( I = \int_0^1 \frac{d}{dq} (C(q) - q)^2 \, dq \):

\[
\hat{I} = \int_0^1 \left[ \left( \sum_{k=1}^d \alpha_k P_k'(q) \right) - 1 \right]^2 \, dq
\]

We use the fact that \( P_1'(q) = 2 \) to define modified coefficients \((\tilde{\alpha}_k)_{k=1,\ldots,d}\) as equal to coefficients \((\alpha_k)_{k=1,\ldots,d}\) except for \( \tilde{\alpha}_1 = \alpha_1 - \frac{1}{2} \), :

\[
\hat{I} = \int_0^1 \left[ \sum_{k=1}^d \tilde{\alpha}_k P_k'(q) \right]^2 \, dq
\]

\[
= \sum_{k,l=1}^d \tilde{\alpha}_k \tilde{\alpha}_l \int_0^1 P_k'(q) P_l'(q) \, dq
\]

\[
= \sum_{k,l=1}^d \tilde{\alpha}_k \tilde{\alpha}_l I_{kl}
\]
Let assume that $k \leq n$.

Using an integration by parts we have:

$$I_{k,l} = \left[ P'_k(q) P_l(q) \right]_0^1 - \int_0^1 P''_k(q) P_l(q) dq$$  \hspace{1cm} (7)

$P''_k$ is a polynom of degree $k - 2$ : it can be decomposed on the finite orthogonal basis $(P_i)_{i=1,...,k-2}$. As $k - 2 < l$, using the orthogonality of shifted Legendre polynomials $(P_k)_{k \in \mathbb{N}}$ on $[0, 1]$, we find that the integral $\int_0^1 P''_k(q) P_l(q) dq$ is equal to 0. Hence:

$$I_{k,l} = P'_k(1) P_l(1) - P'_k(0) P_l(0)$$  \hspace{1cm} (8)
Detailed proof for Eqn.(3)

The values of $P_k(q)$ and its derivative $P'_k(q)$ at $q = 0$ and $q = 1$ can be found from the corresponding values of non-shifted Legendre polynomial $P^{(s)}_k(q)$ at $q = -1$ and $q = 1$, which are given in [2, p.777], Eqn.(22.4.6), (22.5.37) and (22.4.2). Using the relations $P_k(q) = P^{(s)}_k(2q - 1)$ and $P'_k(q) = 2(P^{(s)}_k)'(2q - 1)$ we have:

$$\forall k \in \mathbb{N} \quad \begin{cases} P_k(1) &= 1 \\
P'_k(1) &= k(k + 1) \\
P_k(0) &= (-1)^k \\
P'_k(0) &= (-1)^{k-1} k(k + 1) \end{cases} \tag{9}$$

We finally obtain:

$$\forall (k, l) \in \mathbb{N}^2, k \leq l, \quad I_{k,l} = k(k + 1)[1 + (-1)^{k+l}] \tag{10}$$

which we can also write this way:

$$\forall (k, l) \in \mathbb{N}^2, \quad I_{kl} = 2 \min(k, l) [1 + \min(k, l)] 1_{\{(k+l) \in 2\mathbb{N}\}} \tag{11}$$
Detailed proof for Eqn.(3)

The values of $P_k(q)$ and its derivative $P'_k(q)$ at $q = 0$ and $q = 1$ can be found from the corresponding values of non-shifted Legendre polynomial $P^{(s)}_k(q)$ at $q = -1$ and $q = 1$, which are given in [2, p.777], Eqn.(22.4.6), (22.5.37) and (22.4.2). Using the relations $P_k(q) = P^{(s)}_k(2q - 1)$ and $P'_k(q) = 2(P^{(s)}_k)'(2q - 1)$ we have:

$$
\forall k \in \mathbb{N} \quad \left\{ \begin{array}{l}
P_k(1) = 1 \\
P'_k(1) = k(k + 1) \\
P_k(0) = (-1)^k \\
P'_k(0) = (-1)^{k-1} k(k + 1)
\end{array} \right.
\quad (9)
$$

We finally obtain:

$$
\forall (k, l) \in \mathbb{N}^2, k \leq l, \\
l_{k,l} = k(k + 1)[1 + (-1)^{k+l}]
$$

which we can also write this way:

$$
\forall (k, l) \in \mathbb{N}^2, \\
l_{kl} = 2 \min(k, l) \left[ 1 + \min(k, l) \right] \mathbf{1}_{\{(k+l) \in 2\mathbb{N}\}}
$$

\quad (11)
Contribution to the Sample Variance

Contribution to the sample variance for input parameter $X_j$ at quantile $q$ is given by:

$$D_j(q) = \frac{1}{\sqrt{V(Y)}} \int_{-\infty}^{F_j^{-1}(q)} E \left[ (Y - \mathbf{E}(Y))^2 \mid X_j = x_j \right] p(x_j) dx_j$$

(12)
Contribution to the Sample Variance

Slope of the CSV plot

The slope of the CSV plot between the two points \((q_1, D(q_1))\) and \((q_2, D(q_2))\) is given by:

\[
\frac{D(q_2) - D(q_1)}{q_2 - q_1} = \frac{V(Y^{*[z_1,z_2]})}{V(Y)}
\]  

(13)

with variance \(V(Y^{*[z_1,z_2]})\), defined as the variance of the model output when the range of the parameter \(X_j\) is reduced to \([z_1, z_2]\), but with respect to constant mean \(E(Y)\) over the full range of all parameters:

\[
V(Y^{*[z]}) = E \left[ (Y - E(Y))^2 \mid X_j = z \right]
\]
Contribution to the Sample Variance

Relation with total order sensitivity indices?

Total order sensitivity indices:

\[ ST_j = 1 - \frac{\mathbb{E}_{X_j} \left[ \text{Var}_{X \sim j} (Y \mid X_j) \right]}{\text{V}(Y)} \]

\[ = 1 - \frac{\mathbb{E}_{X_j} \left( \mathbb{E}_{X \sim j} \left[ (Y - \mathbb{E}[Y \mid X_j])^2 \mid X_j = x_j \right] \right)}{\text{V}(Y)} \]

Let denote by \( \text{V}(Y^\circ\{x_j\}) \) the quantity \( \mathbb{E}_{X \sim j} \left[ (Y - \mathbb{E}[Y \mid X_j])^2 \mid X_j = x_j \right] \). It is the variance of model output when model input \( X_j \) is fixed to the value \( x_j \), but with respect to the conditional mean \( \mathbb{E}[Y \mid X_j = x_j] \). We then have:

\[ ST_j = \int_0^1 \left[ 1 - \frac{\text{V}(Y^\circ\{F_j^{-1}(q)\})}{\text{V}(Y)} \right] dq \]  

(14)
Problem is that the two variances $V(Y^o\{z\})$ and $V(Y^*\{z\})$ are not equal, as they are not computed with respect to the same mean value.

- **constant mean** $E(Y)$ for $V(Y^*\{z\})$
- **conditionnal mean** $E[Y | X_j = z]$ for $V(Y^o\{z\})$