Introduction to Sensitivity Analysis

Cédric J. Sallaberry & Laura P. Swiler
What is sensitivity Analysis? (1/2)

- Involve generation and exploration of mapping from analysis inputs to analysis results

- Analysis input: \( x = [x_1, x_2, \ldots, x_n] \)

- Analysis results: \( y(x) = [y_1(x), y_2(x), \ldots, y_n(x)] \)

How important are the individual elements of \( x \) with respect to the uncertainty in \( y(x) \)?
What is sensitivity Analysis ? (2/2)

Sensitivity Analysis can be used for

- Ranking parameters in term of their importance relative to the uncertainty in the output
- Verification and validation of the model. It is a powerful tool to check whether the system performs as expected
- Leading further uncertainty quantification towards the parameters that really matters in an iterative process

Sensitivity is usually **not** used for

- Prediction: The purpose is not to construct a meta-model

- Determining the importance of one parameter or one feature of the system to the response. It only looks at the influence of the **uncertainty** in the input on the **uncertainty** of the output
Methods for sensitivity analysis

Several techniques have been developed to perform sensitivity analysis, which can be classified into three groups:

- Deterministic
- Local
- Global
Deterministic Sensitivity Analysis (1/2)

- The basis of such methods, such as **adjoint method** or **gradient-based method** is to invert the problem.

- They can be very powerful, since the influence of input parameters is instantaneous and mathematically proven.

- However, they can be pretty hard to set up mathematically and numerically and any change in the system may lead to a complete new analysis.
At high level, the inverse problem can be seen as a minimization of the functional form derived from the original problem.

For linear systems, *linear algebra* and *functional analysis* give theoretical stability leading to general existence results and efficient algorithms in order to solve the problem.

For non-linear systems most of the results are specific to a given problem and it is often necessary to use optimization algorithm from *Numerical Analysis* course to converge to the solution.

In most of big analysis such methods are impractical due to the complexity of the (often non-linear) equations and potential coupling between the systems.
Local Sensitivity Analysis (1/2)

- Local methods can be applied on deterministic or probabilistic problems.
- As their name suggests, they calculate sensitivity of the output at one location (or around a neighborhood) in the input hypercube.
- Most of these methods calculate derivative of the function relative to one or multiple parameter.
- In order to calculate derivatives, Taylor series expansion is often used.

\[
f(x + h) = f(x) + h \frac{\partial f}{\partial x} + \frac{h^2}{2!} \frac{\partial^2 f}{\partial x^2}(x) + \cdots + \frac{h^n}{n!} \frac{\partial^n f}{\partial x^n}(x + \theta h)
\]
We are usually interested more on the global effect of an input variables than on its effect around a specific point in the hypercube.

• Technically, there is no assumptions on the relation between any input $x_i$ and the output $y$. The accuracy of this approach will depend on the method and the number of points used to estimate the derivative.

• It is also possible to reduce the accuracy of the method but looking at a larger region around the point of interest by selecting further point in order to calculate the derivative.

• It is theoretically possible, although costly, to calculate higher order derivatives including partial derivatives over two or more input parameters.

• however, new runs (to estimate the derivatives) have to be done for each instance of interest of each input variable. This method can be then very costly if applied to a system with many inputs and outputs.
One particular case of local regression that may be useful is when dealing with reliability methods such as FORM/SORM:

- The gradients are already estimated so no more calculations are necessary
- The point of the hypercube is of particular interest as its stability may affect the probability of interest.
- Sensitivity indices derived not only tell which input is the most important in this area but also which parameters of the associated distributions would more likely to affect the output result.
Local Sensitivity Analysis with reliability methods (2/3)

Because of the way the derivatives are calculated (not taking into account the correlation between inputs variables), it is more appropriate to group the correlated variables and study the effect of such groups.
Local Sensitivity Analysis with reliability methods (3/3)

For each input, the sensitivity indices indicate how the change in the parameters distribution used would affect the result in the neighborhood of the point (and potentially the may
Global methods

• These methods try to capture the influence of the inputs on the whole hypercube.

• Working on the overall for both the input and output of interest, they had to consider the all range of distribution of the variables.

• Several methods cover the whole range of uncertainty in inputs and output of interest. They include response surface, variance based decomposition and sampling-based methods. The first two will be the subject of futures lectures. The rest of this presentation will be on tools used on sampling-based methods.

• For sampling-based methods, the same sample set is used to perform uncertainty analysis and sensitivity analysis. While, depending on the problem, the sample size may be large, there is no need to rerun the code a second time.
Complete Variance Decomposition (CVD)

• The driving idea of complete variance decomposition stands on rewriting the function as a sum of functions depending on one or more parameters.

\[ f(x) = f_0 + \sum_{i=1}^{k} f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) + \ldots + f_{1,2,\ldots,k}(x_1, x_2, \ldots, x_k) \]

• CVD will estimate the influence of each parameter and interactions of parameters without any assumption, but at the cost of more runs.

This topic and two specific methods (Sobol and FAST) will be covered in more detailed in a future lecture.
Response Surface

• Response surface or Meta-models are used to represent the response domain with an analytical function (or at least less expensive numerical function than the original code).

• They can be used either directly to estimate the importance of parameters or as a surrogate of the model to be able to perform greater number of runs at a lower cost (then making a CVD possible for instance)

This topic and some selected methods will be covered in more detailed in future lectures.
Scatterplots

• Scatterplots consists in plotting the relation between inputs (on the x-axis) and output (on the y-axis)

• They are the simplest way to observe any dependency between input and output without making any assumption. And are usually enough to understand relationship between input and output

• They are usually bidimensional (one output vs. one input) but can be sometime tri-dimensional to show three way interactions.

• But, they may be impractical if hundreds of inputs and outputs are in consideration
2D - Scatterplots examples

Color coding or use of different symbols may help represent conjoint influence from discrete and continuous parameter.

Box-plots may be a good alternative to scatterplots when dealing with discrete variables.
• When the number of input and output is big in the analysis, it is necessary to screen out most of the relations before plotting them.

• Regression analysis and Correlation are therefore used to determine which variables influence the uncertainty in the model result.

• Several other methods can be used. Some of them will be presented in future lectures. This lecture will focus on linear regression, coefficient of correlation and on the rank transform.
Correlation and regression: Example set up

inputs

- $X_1, X_2, X_3$ and $X_4$ uniformly distributed between 0 and 1
- $\text{Corr}(X_1, X_4) = 0.8$. All other correlations are set to 0
- $\varepsilon$ is an input representing random noise which varies between -0.2 and 0.2

output

$$Y_1 = 10X_1 + X_2 + 0.1X_3 + 0.01X_4$$

$$Y_2 = 10X_1 + X_2 + 0.1X_3 + 0.01X_4 + \varepsilon$$

sampling

- Latin Hypercube Sampling (LHS)
- Iman-Conover correlation technique
- Sample size = 1,000
Example : Results for Y1

Slope=10

Slope=1
Random noise does not seem to influence the scatterplots

Example: Results for Y2
Coefficient of Correlation (Pearson)

Measures the strength and direction of a **linear** relationship between two quantities $X_i$ and $Y_j$.

\[
\rho(X_i, Y) = \frac{\text{cov}(X_i, Y)}{\sqrt{\text{var}(X_i) \cdot \text{var}(Y)}}
\]

\[
\rho(X_i, Y) = \frac{n \sum_{i=1}^{n} x_{i,1} y_{i} - \sum_{i=1}^{n} x_{i,1} \sum_{i=1}^{n} y_{i}}{\sqrt{n \sum_{i=1}^{n} x_{i,1}^2 - \left( \sum_{i=1}^{n} x_{i,1} \right)^2} \sqrt{n \sum_{i=1}^{n} y_{i}^2 - \left( \sum_{i=1}^{n} y_{i} \right)^2}}
\]

**Note 1:** as the covariance is normalized by the variance of the two terms in consideration, the slope of the relation does not change the correlation value.

**Note 2:** there may be dependency amongst variables that will not be captured by correlation.

### Coefficient of correlation (CC) in example

#### Results for $Y_1$

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$Y_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>1.000</td>
<td>0.040</td>
<td>-0.001</td>
<td>0.800</td>
<td>0.995</td>
</tr>
<tr>
<td>$X_2$</td>
<td>0.040</td>
<td>1.000</td>
<td>0.021</td>
<td>0.112</td>
<td>0.139</td>
</tr>
<tr>
<td>$X_3$</td>
<td>-0.001</td>
<td>0.021</td>
<td>1.000</td>
<td>0.045</td>
<td>0.011</td>
</tr>
<tr>
<td>$X_4$</td>
<td>0.800</td>
<td>0.112</td>
<td>0.045</td>
<td>1.000</td>
<td>0.804</td>
</tr>
<tr>
<td>$Y_1$</td>
<td>0.995</td>
<td>0.139</td>
<td>0.011</td>
<td>0.804</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Even if $X_4$ has negligible effect in the equation of $Y$, its correlation coefficient is strong due to its correlation with $X_1$.

#### Results for $Y_2$

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$Y_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>1.000</td>
<td>0.040</td>
<td>-0.001</td>
<td>0.800</td>
<td>0.995</td>
</tr>
<tr>
<td>$X_2$</td>
<td>0.040</td>
<td>1.000</td>
<td>0.021</td>
<td>0.112</td>
<td>0.139</td>
</tr>
<tr>
<td>$X_3$</td>
<td>-0.001</td>
<td>0.021</td>
<td>1.000</td>
<td>0.045</td>
<td>0.011</td>
</tr>
<tr>
<td>$X_4$</td>
<td>0.800</td>
<td>0.112</td>
<td>0.045</td>
<td>1.000</td>
<td>0.804</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>0.995</td>
<td>0.139</td>
<td>0.011</td>
<td>0.804</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Random noise does not seem to influence the correlation.
Partial Correlation Coefficient (PCC)

Measures the strength and direction of a linear relationship an input $X_i$ and an output $Y_j$ after the linear effect of the remaining input parameters has been taken out from both $X_i$ and $Y_j$

Step 1: linear regression models of $Y_j$ and $X_i$

$$\tilde{Y}_{i,j} = a_0 + a_1 X_1 + a_2 X_2 + \cdots a_{i-1} X_{i-1} + a_{i+1} X_{i+1} + \cdots a_n X_n$$

$$\tilde{X}_i = b_0 + b_1 X_1 + b_2 X_2 + \cdots b_{i-1} X_{i-1} + b_{i+1} X_{i+1} + \cdots b_n X_n$$

Step 2: Calculation of residual

$$r_{y_{i,j}} = Y_j - \tilde{Y}_{i,j}$$

$$r_{x_i} = X_i - \tilde{X}_i$$

Step 3: Calculation of correlation between $r_{y_{i,j}}$ and $r_{x_i}$

$$PCC(X_i, Y) = \frac{\text{cov}(r_{x_i}, r_{y_{i,j}})}{\sqrt{\text{var}(r_{x_i}) \cdot \text{var}(r_{y_{i,j}})}}$$
In absence of noise, as the model is perfectly linear, each PCC is equal to 1.00 this reflects that each parameter has a perfect linear influence on the output $Y_1$.

When random noise is added, the linear influence of the parameter is partially or totally hidden. Depending of it’s importance in the equation, a parameter may have a PCC varying from 1.00 to 0.02 in the analysis of $Y_2$.

Note that $\varepsilon$ would be in the group of input parameters, all PCC would be equal to 1.00 for $Y_2$. 

<table>
<thead>
<tr>
<th>PCC values</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>1.00</td>
<td>0.93</td>
<td>0.25</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Partial Correlation Coefficient in example (2/2)

\[ r_{y3,1} = Y_1 - \tilde{Y}_{3,1} = 0.1X_3 \]
\[ r_{x3} = X_3 - 0.5 \]
\[ Y_1 \in [0,0.1] \]
\[ X_3 \in [0,1] \]

\[ \tilde{Y}_{3,1} = \tilde{Y}_{3,2} = 10X_1 + X_2 + 0.01X_4 \]
\[ \tilde{X}_3 = 0.5 \]

\[ r_{y3,2} = Y_2 - \tilde{Y}_{3,2} = 0.1X_3 + \varepsilon \]
\[ r_{x3} = X_3 - 0.5 \]
\[ X_3 \in [0,1] ; \varepsilon \in [-0.2, 0.2] \]
\[ Y_2 \in [-0.2, 0.3] \]
Coefficients of the Linear Regression model, estimated using Least Square

\[ X = (X_1, \ldots, X_n) \rightarrow Y_j \]

\[ \tilde{Y}_j = \sum_{i=1}^{n} \theta_i X_i \]

**Linear regression**

- We want to select the \( \theta_i \) such that they minimize the square difference between the output \( Y_j \) and its linear regression. (That’s why it’s called **Least Square**)
- The minimum of the function is obtained when its derivative is zero.

\[
\min_{\theta} f(\theta) = \min_{\theta} \| Y_j - \tilde{Y}_j \|^2 = \min_{\theta} \| Y_j - \theta X \|^2 \\
\rightarrow f'(\theta) = 0 \\
\rightarrow 2X^T(Y_j - \theta X) = 0 \\
\rightarrow \left( X^T X \right) \theta = X^T Y \\
\rightarrow \theta = \left( X^T X \right)^{-1} X^T Y
\]
Regression Coefficient in example (1/2)

\[ \theta = \left( X^T X \right)^{-1} X^T Y_1 \rightarrow \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix} = \begin{pmatrix} -2.6 \times 10^{-8} \\ 10 \\ 1 \\ 0.1 \\ 0.01 \end{pmatrix} \]

\( Y_1 = 10X_1 + X_2 + 0.1X_3 + 0.01X_4 \)

\( \tilde{Y}_1 = 10X_1 + X_2 + 0.1X_3 + 0.01X_4 - 2.6 \times 10^{-8} \)

\[ \theta = \left( X^T X \right)^{-1} X^T \tilde{Y}_1 \rightarrow \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix} = \begin{pmatrix} -0.03 \\ 10.03 \\ 1.02 \\ 0.101 \\ 0.014 \end{pmatrix} \]

\( Y_2 = 10X_1 + X_2 + 0.1X_3 + 0.01X_4 + \epsilon \)

\( \tilde{Y}_2 = 10.03X_1 + 1.02X_2 + 0.101X_3 + 0.014X_4 - 0.03 \)
Regression Coefficient in example (2/2)

**Note:** values are not normalized. Changing the unit will change the result

Let change $X_2$ as uniformly distributed between 0 and 0.1.

$Y_1$ will then be defined with:

$$Y_1 = 10X_1 + 10X_2 + 0.1X_3 + 0.01X_4$$

$$\theta = \left( X^T X \right)^{-1} X^T Y_1 \rightarrow \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix} = \begin{pmatrix} -2.6 \times 10^{-8} \\ 10 \\ 10 \\ 0.1 \\ 0.01 \end{pmatrix}$$

The change in unit changes the value of the regression coefficient associated with $X_3$ while its influence does not change.
Standardized Coefficients of the Linear Regression model corresponds to linear coefficients of the Standardized model. The standardization of a variable is performed by subtracting the mean and dividing the result by the standard deviation.

\[
X = (X_1, \ldots, X_n) \rightarrow Y_j
\]

**Linear regression**

\[
\widetilde{Y}_j^* = \sum_{i=1}^{n} \theta_i^* X_i^*
\]

**Standardization**

\[
Y_j^* = \frac{Y_j - \mu_{Y_j}}{\sigma_{Y_j}}; \quad X_i^* = \frac{X_i - \mu_{X_i}}{\sigma_{X_i}}
\]

The standardized values will **not** be affected by unit change.
Standard Regression Coefficient in example

<table>
<thead>
<tr>
<th>SRC values</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1$</td>
<td>0.99</td>
<td>0.1</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>0.99</td>
<td>0.1</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Results are the same whether there is random noise or not.
All the previous techniques suppose that the relation between input and output is linear. It is generally NOT the case. A simple way to relax part of this constraint is to work with RANK.
Advantages

- It extends the relation from linear to monotonic patterns
- It does not involve any additional calculation than ranking
- It does not create over fitting

Drawbacks

- As it is a non-parametric method, it cannot be used for prediction
- It still does not capture non-monotonic pattern or conjoint influence of multiple input parameters
The coefficient of Determination, noted R² measures the part of the variance of the output that is explained by the regression model.

**R² ~ 1**: Most of the variance of Y is explained

No other analysis is required to determine the most influent parameters

**R² ~ 0**: Most of the variance of Y is NOT explained

Some influential parameters may be missing OR the influence of the parameters selected is misrepresented. It may be thus necessary to look at scatterplots or/and apply different regression techniques.
Conclusive remarks on regression, correlation, rank transform and $R^2$

- Partial correlation coefficients (PCCs) and standardized regression coefficients (SRCs) are usually not calculated as described, as they can all be calculated by inverting the correlation matrix.
- The rank regression is a powerful tool for performing sensitivity analysis.
  - It captures monotonic relations which are the most common in the type of problem we deal with
  - It is fast and can thus be applied on large set of data
- This technique is pretty robust and does not usually over-fit (i.e. over-estimate the value of $R^2$)
- The $R^2$ is an indicator of the quality of the regression, informing when more analyses are necessary to understand the system.
- Experience shows that most of the time, even when the $R^2$ is low, the most influential parameters are selected (their influence is however underestimated)
General Conclusion

• While several techniques exist and bring, each, different aspect and information in sensitivity analysis, the most commonly used in the type of system we study is Sampling-based methods
  ➢ They are easy to setup and do not need more run than for the uncertainty analysis
  ➢ They are not so expensive considering that the same sample set can work on a large input set and for many output parameters
• Furthermore, simple regression, associated with rank transform, will give pretty good information on the behavior of a particular output about 80% of the time.

• However, that does not mean that other methods should not be used. Sampling-based methods allow to simplify the analysis (results can be processed automatically) and reduce the number of cases on which more sophisticated methods need to be used. These methods will be presented in the subsequent lectures.
