Probabilistic sensitivity analysis: contribution to the sample mean plot and moment-independent importance measures

William Castaings‡

‡Institut de Mécanique des Fluides de Toulouse
travaux menés au Centre Commun de Recherche de la Commission Européenne
Contribution to the sample mean plot
- Contribution to the sample mean plot
- Statistical test for inputs prioritisation

Moment independent sensitivity analysis
- Moment-independent importance measures
- Numerical and computational aspects
- Application examples
Outline

Contribution to the sample mean plot

- Contribution to the sample mean plot
- Statistical test for inputs prioritisation

Moment independent sensitivity analysis

- Moment-independent importance measures
- Numerical and computational aspects
- Application examples
Contribution to the sample mean plot for graphical and numerical sensitivity analysis

R. Bolado

European Commission, Joint Research Centre (IE, Petten)

W. Castaings, S. Tarantola

European Commission, Joint Research Centre (IPSC, Ispra)
\[ Y = f(X) \]

- \( f \) a deterministic scalar function
- \( X = (X_1, \cdots, X_k) \) and \( Y \) random variables
- \( x = (x_1, \cdots, x_k) \) realization of the model inputs \( X \)
- \( y \) realization of the model output \( Y \)
\[ Y = f(X) \]

- **Objective:** understand the behaviour of the system with very few model runs
Overview

Context

- *Sinclair, (1993)* investigated how finite changes in inputs pdfs affect the mean and variance of the output
- Contribution to the sample mean (CSM) plot recognized as a general tool for sensitivity analysis

Objectives

- Revive and further develop CSM plot
- Exploit the full potential of this graphical tool
- CSM plot, primary building block of a statistical test for inputs prioritisation
Different steps for the construction of a CSM plot

1. realizations of $X_i$ are sorted, generating \{\(x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(N)}\}\);

2. realizations of $Y$ are sorted accordingly, generating \{\(y^{(i,1)}, y^{(i,2)}, \ldots, y^{(i,N)}\}\)

3. new variable $M_i$ defined by

\[
m_i^{(q)} = \frac{1}{N} \sum_{j=1}^{q} y^{(i,j)} \quad q = 1, \ldots, N
\]

4. normalization of the $M_i$ using the sample mean of $Y$;

5. plot $M_i$ against $F_{X_i}(x_i)$

**Underlying features**

- For both axes, values lie in the interval $[0, 1]$
- $(F_{X_i}(x_i^{(q)}), m_i^{q})$: fraction of the output mean due to any given fraction of values of the input $X_i$. 

William Castaings

IMPEC, 13 Oct 2008
Different steps for the construction of a CSM plot

1. realizations of $X_i$ are sorted, generating \( \{x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(N)}\} \); 
2. realizations of $Y$ are sorted accordingly, generating \( \{y^{(i,1)}, y^{(i,2)}, \ldots, y^{(i,N)}\} \); 
3. new variable $M_i$ defined by 
   \[
   m_i^{(q)} = \frac{1}{N} \sum_{j=1}^{q} y^{(i,j)} \quad q = 1, \ldots, N 
   \]
4. normalization of the $M_i$ using the sample mean of $Y$; 
5. plot $M_i$ against $F_{X_i}(x_i)$

Underlying features

- For both axes, values lie in the interval $[0, 1]$
- \( (F_{X_i}(x_i^{(q)}), m_i^{(q)}) \): fraction of the output mean due to any given fraction of values of the input $X_i$. 

William Castaings
IMPEC, 13 Oct 2008
Didactic example
All parameters, single output

Analytic function

- $Y = 2\exp(X_1) - \exp(X_2) + \sin(X_3)$
- $X_i, i = 1, 2, 4 \sim U(0, 1), X_3 \sim U(0, \pi)$
High-level waste repository model (LevelE)
Single parameter, several outputs

Information provided by the plot

- Effects on the mean of the output of changes in the inputs pdfs
- Underlines the limitations of the sample size/design
- Global importance measures
CSM plot with increasing sample size

![Graph showing CSM plot with increasing sample size. The graph plots the relative contribution to the mean against CDF $V_1$. The lines represent different sample sizes: N=50 (black), N=500 (grey), and N=5000 (light grey). The x-axis represents the cumulative distribution function of $V_1$, ranging from 0 to 1, while the y-axis represents the relative contribution to the mean, ranging from 0 to 1.]
Prioritisation of model inputs

CSM plot

- If $F_{X_i}(x_i^{(q)}) \simeq m_i^q \forall q$, any quantile range of $X_i$ has a similar influence on the output mean, i.e. non-influent model input.

Relation with VB

- Variance-based first-order effect

\[ S_i = \frac{Var(E[Y|X_i])}{Var(Y)} \]

- CSM plot, variability of $E[Y|X_i < x_i^*]$ (rather than $E[Y|X_i = x_i^*]$) across the range.
Statistical test keynotes

Features

- **Hypotheses** (null hypothesis $H_0$ and alternative hypothesis $H_1$):
  - $H_0: f_{Y|X_i}(y|x_i = x^*_i) = f_Y(y) \forall x^*_i \in R_i$ ($R_i$ is the support of $X_i$);
  - $H_1: \exists x^*_i, x'_i \in R_i \ (f_{Y|X_i}(y|x_i = x^*_i) \neq f_{Y|X_i}(y|x_i = x'_i))$.
- **Test statistic**: $D_m$, the maximum distance to the diagonal

Different steps

1. **Empirical distribution of $D_m$**
   - Random permutations of the inputs realizations
   - For each permutation, compute $D_m$ from CSM plot
2. **Compute $D_{m\alpha}$**, value of the test statistic for the critical level $\alpha$
3. **Estimation of $D_{mX_i}$**, $i = 1, \cdots, k$ from the original CSM plot
4. **null hypothesis $H_0$ rejected if $D_{mX_i} > D_{m\alpha}$** (i.e. $X_i$ is an important input)
**Statistical test keynotes**

### Features

- **Hypotheses** (null hypothesis $H_0$ and alternative hypothesis $H_1$):
  - $H_0$: $f_{Y|X_i}(y|x_i = x_i^*) = f_Y(y)$ $\forall x_i^* \in R_i$ ($R_i$ is the support of $X_i$);
  - $H_1$: $\exists x_i^*, x_i' \in R_i$ $/$ $f_{Y|X_i}(y|x_i = x_i^*) \neq f_{Y|X_i}(y|x_i = x_i')$.

- **Test statistic**: $D_m$, the maximum distance to the diagonal

### Different steps

1. **Empirical distribution of $D_m$**
   - Random permutations of the inputs realizations
   - For each permutation, compute $D_m$ from CSM plot
2. **Compute $D_{m\alpha}$, value of the test statistic for the critical level $\alpha$**
3. **Estimation of $D_{mx_i}$, $i = 1, \cdots, k$ from the original CSM plot**
4. **null hypothesis $H_0$ rejected if $D_{mx_i} > D_{m\alpha}$ (i.e. $X_i$ is an important input)**
Convergence of importance measures

Test statistic and SDP (*Ratto et al, 2007*) first order indices (LHS samples 50-3000)
Robustness of importance measures

Test statistic across 20 LHS replicates of size 500
Conclusions

Potential

- CSM plot: simple, versatile and very informative graphical tool
- Statistical test: identifies important model inputs for very low sample size, no additional model run for robustness analysis

Limitations

- Inputs prioritisation assessment restricted to first order effects
- Statistical test prone to type I error (treating non-influential inputs as important)

To be done ...

- Systematic approach for non-monotonic mappings
- Second order interactions with surfaces
- Investigate the potential of the contribution to the sample variance (CSV) plot
Outline

Contribution to the sample mean plot
- Contribution to the sample mean plot
- Statistical test for inputs prioritisation

Moment independent sensitivity analysis
- Moment-independent importance measures
- Numerical and computational aspects
- Application examples
Relative importance of model inputs on the output probability distribution function

Emanuele Borgonovo
Bocconi University, Department of Decision Sciences (ELEUSI, Milan)

W. Castaings, S. Tarantola
European Commission, Joint Research Centre (IPSC, Ispra)
$Y = f(X)$

- $f$ a deterministic scalar function
- $X = (X_1, \cdots, X_k)$ and $Y$ random variables
- $x = (x_1, \cdots, x_k)$ realization of the model inputs $X$
- $y$ realization of the model output $Y$
\[ Y = f(X) \]

- Variance not necessarily adapted to describe the output variability
- Analysis of the entire output distribution \( f_Y(y) \) rather than \( V(Y) \)
Conditional variance $V(Y|X_i)$ to be compared with $V(Y)$

$$S_i = \frac{V_i}{V(Y)}$$

$$V_i = V(E(Y|X_i)) = V(Y) - E(V(Y|X_i))$$

$$V(Y) = E(V(Y|X_i)) + V(E(Y|X_i))$$
- Conditional PDF $f_{Y|X_i}(y)$ to be compared with $f_Y(y)$

$$s(X_i) = \int |f_Y(y) - f_{Y|X_i}(y)| \, dy$$

$$\delta_i = \frac{1}{2} E_{X_i}[s(X_i)]$$

- Conditional variance $V(Y|X_i)$ to be compared with $V(Y)$

$$S_i = \frac{V_i}{V(Y)}$$

$$V_i = V(E(Y|X_i)) = V(Y) - E(V(Y|X_i))$$

$$V(Y) = E(V(Y|X_i)) + V(E(Y|X_i))$$
Conditional PDF $f_{Y|X_i}(y)$ to be compared with $f_Y(y)$

$$s(X_i) = \int |f_Y(y) - f_{Y|X_i}(y)| \, dy$$

$$\delta_i = \frac{1}{2} E_{X_i}[s(X_i)]$$
• Conditional PDF \( f_{Y|X_i}(y) \) to be compared with \( f_Y(y) \)

\[
s(X_i) = \int |f_Y(y) - f_{Y|X_i}(y)| \, dy
\]

\[
\delta_i = \frac{1}{2} E_{X_i}[s(X_i)]
\]

• Other moment-independent important measures based on CDF (Park et al, 1994; Chun et al, 2000),

• The measures proposed by Borgonovo, (2006) have interesting normalization properties
Essential properties

✓ Individual importance

\[ 0 \leq \delta_i \leq 1 \]
Essential properties

✓ Individual importance

$$0 \leq \delta_i \leq 1$$

Joint importance of $$X_i$$ and $$X_j$$

$$\delta_{ij} = \frac{1}{2} \int f_{X_i,X_j}(x_i,x_j) \left[ \int |f_Y(y) - f_{Y|X_i,X_j}(y)|dy \right] dx_i dx_j$$

$$\delta_{ij} = \delta_i$$ if $$Y$$ is dependent on $$X_i$$ but independent on $$X_j$$

🧰 $$\delta$$ can be extended to any set of inputs (i.e. analysis by groups)
Essential properties

- Individual importance: $0 \leq \delta_i \leq 1$

- Normalization of joint importance: $\delta_{1,2,\ldots,k} = 1$

Joint importance of $X_i$ and $X_j$

$$\delta_{ij} = \frac{1}{2} \int f_{X_i,X_j}(x_i,x_j) \left[ \int |f_Y(y) - f_{Y|X_i,X_j}(y)| dy \right] dx_i dx_j$$

$$\delta_{ij} = \delta_i \text{ if } Y \text{ is dependent on } X_i \text{ but independent on } X_j$$

- $\delta$ can be extended to any set of inputs (i.e. analysis by groups)
Essential properties

- Individual importance
  \[ 0 \leq \delta_i \leq 1 \]

- Normalization of joint importance
  \[ \delta_1,2,\ldots,k = 1 \]

- Subadditivity
  \[ \delta_i \leq \delta_{ij} \leq \delta_i + \delta_{j|i} \]

\[
\delta_{j|i} = \frac{1}{2} \int f_{X_i,X_j}(x_i, x_j) \times \left[ \int |f_{Y|X_i}(y) - f_{Y|X_i,X_j}(y)| dy \right] dx_i dx_j
\]
Essential properties

✓ Individual importance

\[ 0 \leq \delta_i \leq 1 \]

✓ Normalization of joint importance

\[ \delta_{1,2,\ldots,k} = 1 \]

✓ Subadditivity

\[ \delta_i \leq \delta_{ij} \leq \delta_i + \delta_{j|i} \]

- All properties hold for dependent inputs
- Proofs are provided in *Borgonovo, (2006; 2007)*
Essential aspects of the computational approach

- Focus on $\delta_i \quad i = 1, \cdots, k$

\[
\delta_i = \frac{1}{2} \int f_{X_i}(x_i) \left[ \int |f_Y(y) - f_{Y|X_i}(y)| \, dy \right] \, dx_i
\]

for $i=1$ to $k$ Loop on model inputs

for $j=1$ to $N$ Loop on different values of $X_i$

\[
s(x_i) = \int |f_Y(y) - f_{Y|X_i}(y|X_i = x_i^{(j)})| \, dy
\]
endfor

endfor

Key features

1. Sample generation
2. Evaluation of the area $s(X_i)$
Evaluation of $s(X_i)$

1. Discrete model outputs
   - Histograms are perfectly suited
   - Zero width and number of bins calculated from the sample

2. Continuous model outputs
   - Non-parametric estimation of PDFs (e.g. kernel density estimation)
   - Monotonic transformations can be applied without altering $\delta$ properties
   - Area calculated from CDFs (Liu and Homma, 2008)
Critical aspects of sample generation

Comparison with variance-based

- Approximation errors are potentially larger when dealing with the entire PDF
- Shortcuts are more difficult to elaborate

1. **Shift between** $f_Y(y)$ **and** $f_{Y|X_i}(y)$ **should be due to the fact that**

   $X_i = x_i^*$

   i.e. $\int |f_{Y|X_i} - f_Y(y)| dy < \varepsilon$ **if** $X_i$ **is a dummy input**

   **otherwise type I error (treating non influential inputs as important)**

2. **A sufficient number of** $x_i^*$ **should be explored for estimating**

   $E_{X_i}[s(X_i)]$. 

William Castaings
IMPEC, 13 Oct 2008
Sample generation for independent inputs

- Unconditional sample for $f_X(x)$ and $f_Y(y)$

$$
\begin{pmatrix}
    x_1^{(1)} & x_2^{(1)} & \cdots & x_i^{(1)} & \cdots & x_k^{(1)} \\
    x_1^{(2)} & x_2^{(2)} & \cdots & x_i^{(2)} & \cdots & x_k^{(2)} \\
     \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
    x_1^{(N-1)} & x_2^{(N-1)} & \cdots & x_i^{(N-1)} & \cdots & x_k^{(N-1)} \\
    x_1^{(N)} & x_2^{(N)} & \cdots & x_i^{(N)} & \cdots & x_k^{(N)} \\
\end{pmatrix}
= 
\begin{pmatrix}
    y^{(1)} \\
    y^{(2)} \\
    \vdots \\
    y^{(N-1)} \\
    y^{(N)}
\end{pmatrix}
$$
Sample generation for independent inputs

- Unconditional sample for $f_X(x)$ and $f_Y(y)$

\[
\begin{pmatrix}
  x_1^{(1)} & x_2^{(1)} & \ldots & x_i^{(1)} & \ldots & x_k^{(1)} \\
  x_1^{(2)} & x_2^{(2)} & \ldots & x_i^{(2)} & \ldots & x_k^{(2)} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_1^{(N-1)} & x_2^{(N-1)} & \ldots & x_i^{(N-1)} & \ldots & x_k^{(N-1)} \\
  x_1^{(N)} & x_2^{(N)} & \ldots & x_i^{(N)} & \ldots & x_k^{(N)} 
\end{pmatrix}
= \begin{pmatrix}
  y^{(1)} \\
  y^{(2)} \\
  \vdots \\
  y^{(N-1)} \\
  y^{(N)}
\end{pmatrix}
\]

- $X_{\sim i}$ not influenced by the fact that $X_i = x_i^{(j)}$
- Substituted column sampling can be applied
Sample generation for independent inputs

• Unconditional sample for \( f_X(x) \) and \( f_Y(y) \)

\[
\begin{pmatrix}
  x_1^{(1)} & x_2^{(1)} & \cdots & x_i^{(1)} & \cdots & x_k^{(1)} \\
  x_1^{(2)} & x_2^{(2)} & \cdots & x_i^{(2)} & \cdots & x_k^{(2)} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_1^{(N-1)} & x_2^{(N-1)} & \cdots & x_i^{(N-1)} & \cdots & x_k^{(N-1)} \\
  x_1^{(N)} & x_2^{(N)} & \cdots & x_i^{(N)} & \cdots & x_k^{(N)}
\end{pmatrix}
\begin{pmatrix}
  y^{(1)} \\
  y^{(2)} \\
  \vdots \\
  y^{(N-1)} \\
  y^{(N)}
\end{pmatrix}

= \begin{pmatrix}
  y^{(1)} \\
  y^{(2)} \\
  \vdots \\
  y^{(N-1)} \\
  y^{(N)}
\end{pmatrix}

• Ex. Conditional sample for \( f_{X|X_i}(x|X_i = x_i^{(1)}) \) and \( f_{Y|X_i}(y|X_i = x_i^{(1)}) \)

\[
\begin{pmatrix}
  x_1^{(1)} & x_2^{(1)} & \cdots & x_i^{(1)} & \cdots & x_k^{(1)} \\
  x_1^{(2)} & x_2^{(2)} & \cdots & x_i^{(2)} & \cdots & x_k^{(2)} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_1^{(N-1)} & x_2^{(N-1)} & \cdots & x_i^{(N-1)} & \cdots & x_k^{(N-1)} \\
  x_1^{(N)} & x_2^{(N)} & \cdots & x_i^{(N)} & \cdots & x_k^{(N)}
\end{pmatrix}
\begin{pmatrix}
  y^{(1)}’ \\
  y^{(2)}’ \\
  \vdots \\
  y^{(N-1)}’ \\
  y^{(N)}’
\end{pmatrix}

= \begin{pmatrix}
  y^{(1)}’ \\
  y^{(2)}’ \\
  \vdots \\
  y^{(N-1)}’ \\
  y^{(N)}’
\end{pmatrix}

Sample generation for independent inputs

- $N$ conditional samples of size $N$ required for the calculation of $\delta_i$
- Total number of model evaluations

$$COST = N(1 + kN)$$
Sample generation for independent inputs

- $N$ conditional samples of size $N$ required for the calculation of $\delta_i$
- Total number of model evaluations
  
  \[
  \text{COST} = N(1 + kN)
  \]

Slightly more efficient calculation strategy

1. Less than $N$ sample points for approaching $f_X(x|X_i = x_i^*)$, i.e.
   $N_{\text{int}} < N$
2. Less than $N$ different values $x_i^*$ of $X_i$ for approaching $E_{X_i}[s(X_i)]$, i.e. $N_{\text{ext}} < N$
Sample generation for independent inputs

- $N$ conditional samples of size $N$
  \[ \text{COST} = N(1 + kN) \]

- $N_{\text{ext}}$ conditional samples of size $N_{\text{int}}$
  \[ \text{COST} = N + kN_{\text{int}}N_{\text{ext}} \]
Sample generation for independent inputs

- $N$ conditional samples of size $N$
  
  \[
  \text{COST} = N(1 + kN)
  \]

- $N_{\text{ext}}$ conditional samples of size $N_{\text{int}}$
  
  \[
  \text{COST} = N + kN_{\text{int}}N_{\text{ext}}
  \]

- Reducing $N_{\text{int}}$ can lead to type I error
- $N_{\text{ext}}$ more likely to be reduced given the shape of $s(X_i)$,
Sample generation for independent inputs

- $N$ conditional samples of size $N$
  \[ COST = N(1 + kN) \]

- $N_{\text{ext}}$ conditional samples of size $N_{\text{int}}$
  \[ COST = N + kN_{\text{int}}N_{\text{ext}} \]

- No constraints for the design of the unconditional sample

- Efficient sampling strategies like Latin Hypercube Sampling (McKay, 1979) or Quasi-Random sampling (ex. Sobol, 1976) can be used
Sample generation for dependent inputs

- Unconditional correlated sample for \( f_X(x) \) and \( f_Y(y) \)

\[
\begin{pmatrix}
x_1^{(1)} & x_2^{(1)} & \ldots & x_i^{(1)} & \ldots & x_k^{(1)} \\
x_1^{(2)} & x_2^{(2)} & \ldots & x_i^{(2)} & \ldots & x_k^{(2)} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
x_1^{(N-1)} & x_2^{(N-1)} & \ldots & x_i^{(N-1)} & \ldots & x_k^{(N-1)} \\
x_1^{(N)} & x_2^{(N)} & \ldots & x_i^{(N)} & \ldots & x_k^{(N)}
\end{pmatrix}
= 
\begin{pmatrix}
y^{(1)} \\
y^{(2)} \\
\vdots \\
y^{(N-1)} \\
y^{(N)}
\end{pmatrix}
\]
Sample generation for dependent inputs

- Unconditional correlated sample for $f_X(x)$ and $f_Y(y)$

\[
\begin{pmatrix}
  x_1^{(1)} & x_2^{(1)} & \cdots & x_i^{(1)} & \cdots & x_k^{(1)} \\
  x_1^{(2)} & x_2^{(2)} & \cdots & x_i^{(2)} & \cdots & x_k^{(2)} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_1^{(N-1)} & x_2^{(N-1)} & \cdots & x_i^{(N-1)} & \cdots & x_k^{(N-1)} \\
  x_1^{(N)} & x_2^{(N)} & \cdots & x_i^{(N)} & \cdots & x_k^{(N)}
\end{pmatrix}
= 
\begin{pmatrix}
  y^{(1)} \\
  y^{(2)} \\
  \vdots \\
  y^{(N-1)} \\
  y^{(N)}
\end{pmatrix}
\]

- $X_i$ influenced by the fact that $X_i = x_i^{(j)}$
- Generation of conditional correlated samples for $f_X|X_i(x)$
- Permutated columns sampling plans can be used
Sample generation for dependent inputs

- Unconditional correlated sample for $f_X(x)$ and $f_Y(y)$

$$
\begin{pmatrix}
    x_1^{(1)} & x_2^{(1)} & \cdots & x_i^{(1)} & \cdots & x_k^{(1)} \\
    x_1^{(2)} & x_2^{(2)} & \cdots & x_i^{(2)} & \cdots & x_k^{(2)} \\
    \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
    x_1^{(N-1)} & x_2^{(N-1)} & \cdots & x_i^{(N-1)} & \cdots & x_k^{(N-1)} \\
    x_1^{(N)} & x_2^{(N)} & \cdots & x_i^{(N)} & \cdots & x_k^{(N)}
\end{pmatrix}
= 
\begin{pmatrix}
    y^{(1)} \\
    y^{(2)} \\
    \vdots \\
    y^{(N-1)} \\
    y^{(N)}
\end{pmatrix}
$$

- Replicated Latin Hypercube Sampling (McKay, 1995)
  - $r$ matrices generated through column permutation
  - Correlations induced through permutations (Iman et al., 1987; Stein et al., 1987)
Approach based on rLHS

Tutorial example for sample generation

- $X_i \ (i = 1, 2, 3) \sim U[-\pi - \pi]$
- rLHS sample, number of variables $k = 3$, base sample size $N = 4$, number of replicates $r = 2$

<table>
<thead>
<tr>
<th>Base sample</th>
<th>1st Replicate</th>
<th>2nd Replicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.11</td>
<td>-2.11</td>
<td>0.79</td>
</tr>
<tr>
<td>-2.38</td>
<td>3.09</td>
<td>0.28</td>
</tr>
<tr>
<td>-2.18</td>
<td>-2.18</td>
<td>3.03</td>
</tr>
<tr>
<td>0.79</td>
<td>-0.14</td>
<td>-0.39</td>
</tr>
<tr>
<td>-0.14</td>
<td>0.28</td>
<td>3.03</td>
</tr>
<tr>
<td>0.53</td>
<td>3.03</td>
<td>-0.99</td>
</tr>
<tr>
<td>1.71</td>
<td>-2.38</td>
<td>0.79</td>
</tr>
<tr>
<td>0.28</td>
<td>3.09</td>
<td>-0.14</td>
</tr>
<tr>
<td>3.03</td>
<td>-0.99</td>
<td>1.71</td>
</tr>
<tr>
<td>-0.39</td>
<td>3.03</td>
<td>1.71</td>
</tr>
<tr>
<td>-0.99</td>
<td>-0.14</td>
<td>-2.11</td>
</tr>
<tr>
<td>0.28</td>
<td>0.53</td>
<td>0.79</td>
</tr>
<tr>
<td>0.53</td>
<td>0.79</td>
<td>-2.38</td>
</tr>
<tr>
<td>-0.14</td>
<td>3.09</td>
<td>-2.11</td>
</tr>
</tbody>
</table>
Approach based on rLHS

Tutorial example for sample generation

- \( X_i \ (i = 1, 2, 3) \sim U[-\pi, \pi] \)
- rLHS sample, number of variables \( k = 3 \), base sample size \( N = 4 \), number of replicates \( r = 2 \)

<table>
<thead>
<tr>
<th>Base sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.11</td>
</tr>
<tr>
<td>0.79</td>
</tr>
<tr>
<td>1.71</td>
</tr>
<tr>
<td>-0.39</td>
</tr>
<tr>
<td>-2.38</td>
</tr>
<tr>
<td>0.14</td>
</tr>
<tr>
<td>0.28</td>
</tr>
<tr>
<td>3.09</td>
</tr>
<tr>
<td>-2.18</td>
</tr>
<tr>
<td>0.53</td>
</tr>
<tr>
<td>3.03</td>
</tr>
<tr>
<td>-0.99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1st Replicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.11</td>
</tr>
<tr>
<td>-0.39</td>
</tr>
<tr>
<td>0.79</td>
</tr>
<tr>
<td>1.71</td>
</tr>
<tr>
<td>3.09</td>
</tr>
<tr>
<td>-0.14</td>
</tr>
<tr>
<td>0.28</td>
</tr>
<tr>
<td>-2.38</td>
</tr>
<tr>
<td>-2.18</td>
</tr>
<tr>
<td>-0.99</td>
</tr>
<tr>
<td>3.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd Replicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.79</td>
</tr>
<tr>
<td>-0.39</td>
</tr>
<tr>
<td>1.71</td>
</tr>
<tr>
<td>-2.11</td>
</tr>
<tr>
<td>-2.38</td>
</tr>
<tr>
<td>0.14</td>
</tr>
<tr>
<td>0.28</td>
</tr>
<tr>
<td>3.09</td>
</tr>
<tr>
<td>-2.18</td>
</tr>
<tr>
<td>-0.99</td>
</tr>
<tr>
<td>3.03</td>
</tr>
</tbody>
</table>
### Approach based on rLHS

**Tutorial example for sample generation**

- $X_i$ ($i = 1, 2, 3$) ~ $U[-\pi, \pi]$
- rLHS sample, number of variables $k = 3$, base sample size $N = 4$, number of replicates $r = 2$

| $f_X(x)$ (Base sample) |  
|------------------------|--
| -2.11 -2.38 -2.18     |
| 0.79 -0.14 0.53       |
| 1.71 0.28 3.03       |
| -0.39 3.09 -0.99      |
| 1st Replicate         |
| -2.11 3.09 -2.18     |
| -0.39 -0.14 -0.99    |
| 0.79 0.28 3.03      |
| 1.71 -2.38 0.53     |
| 2nd Replicate         |
| 0.79 -2.38 -2.18    |
| -0.39 3.09 -0.99   |
| 1.71 -0.14 3.03     |
| -2.11 0.28 0.53    |

Sorting replicates according to values of $X_1$

$$
\begin{bmatrix}
-2.11 & 3.09 & -2.18 \\
-2.11 & 0.28 & 0.53 \\
-0.39 & -0.1409 & -0.99 \\
-0.39 & 3.0980 & -0.99 \\
0.79 & 0.2820 & 3.03 \\
0.79 & -2.3887 & -2.18 \\
1.71 & -2.38 & 0.53 \\
1.71 & -0.14 & 3.03 \\
\end{bmatrix}
$$
Approach based on rLHS

Tutorial example for sample generation

- $X_i \ (i = 1, 2, 3) \sim U[-\pi - \pi]$
- rLHS sample, number of variables $k = 3$, base sample size $N = 4$, number of replicates $r = 2$

<table>
<thead>
<tr>
<th>$f_X(x)$</th>
<th>(Base sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.11</td>
<td>-2.38</td>
</tr>
<tr>
<td>0.79</td>
<td>-0.14</td>
</tr>
<tr>
<td>1.71</td>
<td>0.28</td>
</tr>
<tr>
<td>-0.39</td>
<td>3.09</td>
</tr>
</tbody>
</table>

1\textsuperscript{st} Replicate

| -2.11     | 3.09          | -2.18     |
| -0.39     | -0.14         | -0.99     |
| 0.79      | 0.28          | 3.03      |
| 1.71      | -2.38         | 0.53      |

2\textsuperscript{nd} Replicate

| 0.79      | -2.38         | -2.18     |
| -0.39     | 3.09          | -0.99     |
| 1.71      | -0.14         | 3.03      |
| -2.11     | 0.28          | 0.53      |

Conditional samples for $f_{X|X_1}(x)$

| $f_{X|X_1}(x|X_1 = x_1^{(1)})$ |
|-----------------------------|
| -2.11                       | 3.09 | -2.18     |
| -2.11                       | 0.28 | 0.53      |

| $f_{X|X_1}(x|X_1 = x_1^{(2)})$ |
|-----------------------------|
| -0.39                       | -0.1409 | -0.99     |
| -0.39                       | 3.0980  | -0.99     |

| $f_{X|X_1}(x|X_1 = x_1^{(3)})$ |
|-----------------------------|
| 0.79                        | 0.2820 | 3.03      |
| 0.79                        | -2.3887 | -2.18    |

| $f_{X|X_1}(x|X_1 = x_1^{(4)})$ |
|-----------------------------|
| 1.71                        | -2.38  | 0.53      |
| 1.71                        | -0.14  | 3.03      |
Approach based on rLHS

Tutorial example for sample generation

- \( X_i \ (i = 1, 2, 3) \sim U[-\pi, \pi] \)
- rLHS sample, number of variables \( k = 3 \), base sample size \( N = 4 \), number of replicates \( r = 2 \)

| \( f_X(x) \) (Base sample) | 1
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.11</td>
<td>-2.38</td>
</tr>
<tr>
<td>0.79</td>
<td>-0.14</td>
</tr>
<tr>
<td>1.71</td>
<td>0.28</td>
</tr>
<tr>
<td>-0.39</td>
<td>3.09</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1st Replicate</th>
<th>1.71</th>
<th>-2.38</th>
<th>0.53</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.11</td>
<td>3.09</td>
<td>-2.18</td>
<td></td>
</tr>
<tr>
<td>-0.39</td>
<td>-0.14</td>
<td>-0.99</td>
<td></td>
</tr>
<tr>
<td>0.79</td>
<td>0.28</td>
<td>3.03</td>
<td></td>
</tr>
<tr>
<td>1.71</td>
<td>-2.38</td>
<td>0.53</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd Replicate</th>
<th>1.71</th>
<th>-2.38</th>
<th>0.53</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.79</td>
<td>-2.38</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>-0.39</td>
<td>-0.14</td>
<td>-0.99</td>
<td></td>
</tr>
<tr>
<td>1.71</td>
<td>0.28</td>
<td>3.03</td>
<td></td>
</tr>
<tr>
<td>-2.11</td>
<td>3.09</td>
<td>-2.18</td>
<td></td>
</tr>
<tr>
<td>-0.39</td>
<td>3.09</td>
<td>-0.99</td>
<td></td>
</tr>
</tbody>
</table>

Sorting replicates according to values of \( X_2 \)
**Approach based on rLHS**

**Tutorial example for sample generation**

- $X_i \ (i = 1, 2, 3) \sim U[-\pi, \pi]$  
- rLHS sample, number of variables $k = 3$, base sample size $N = 4$, number of replicates $r = 2$

<table>
<thead>
<tr>
<th>$f_X(x)$ (Base sample)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$-2.11$</td>
<td>$-2.38$</td>
</tr>
<tr>
<td>$0.79$</td>
<td>$-0.14$</td>
</tr>
<tr>
<td>$1.71$</td>
<td>$0.28$</td>
</tr>
<tr>
<td>$-0.39$</td>
<td>$3.09$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$1^{st}$ Replicate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$-2.11$</td>
<td>$3.09$</td>
</tr>
<tr>
<td>$-0.39$</td>
<td>$-0.14$</td>
</tr>
<tr>
<td>$0.79$</td>
<td>$0.28$</td>
</tr>
<tr>
<td>$1.71$</td>
<td>$-2.38$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$2^{nd}$ Replicate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.79$</td>
<td>$-2.38$</td>
</tr>
<tr>
<td>$-0.39$</td>
<td>$3.09$</td>
</tr>
<tr>
<td>$1.71$</td>
<td>$-0.14$</td>
</tr>
<tr>
<td>$-2.11$</td>
<td>$0.28$</td>
</tr>
</tbody>
</table>

**Conditional samples for $f_{X|X_2}(x)$**

| $f_{X|X_2}(x|X_2 = x_2^{(1)})$ |  |
|-----------------------------|--|
| $1.71$ | $-2.38$ | $0.53$ |
| $0.79$ | $-2.38$ | $-2.18$ |

| $f_{X|X_2}(x|X_2 = x_2^{(2)})$ |  |
|-----------------------------|--|
| $-0.39$ | $-0.14$ | $-0.99$ |
| $1.71$ | $-0.14$ | $3.03$ |

| $f_{X|X_2}(x|X_2 = x_2^{(3)})$ |  |
|-----------------------------|--|
| $0.79$ | $0.28$ | $3.03$ |
| $-2.11$ | $0.28$ | $0.53$ |

| $f_{X|X_2}(x|X_2 = x_2^{(4)})$ |  |
|-----------------------------|--|
| $-2.11$ | $3.09$ | $-2.18$ |
| $-0.39$ | $3.09$ | $-0.99$ |
Approach based on rLHS

Tutorial example for sample generation

- $X_i \quad (i = 1, 2, 3) \sim U[-\pi, \pi]$
- rLHS sample, number of variables $k = 3$, base sample size $N = 4$, number of replicates $r = 2$

\[
\begin{bmatrix}
-2.11 & -2.38 & -2.18 \\
0.79 & -0.14 & 0.53 \\
1.71 & 0.28 & 3.03 \\
-0.39 & 3.09 & -0.99 \\
\end{bmatrix}
\]

<table>
<thead>
<tr>
<th>$f_X(x)$ (Base sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-2.11$</td>
</tr>
<tr>
<td>$0.79$</td>
</tr>
<tr>
<td>$1.71$</td>
</tr>
<tr>
<td>$-0.39$</td>
</tr>
</tbody>
</table>

$1^{st}$ Replicate

\[
\begin{bmatrix}
-2.11 & 3.0980 & -2.18 \\
0.79 & -2.3887 & -2.18 \\
-0.39 & -0.1409 & -0.99 \\
-0.39 & 3.0980 & -0.99 \\
1.71 & -2.3887 & 0.53 \\
\end{bmatrix}
\]

$2^{nd}$ Replicate

\[
\begin{bmatrix}
0.79 & -2.38 & -2.18 \\
-0.39 & 3.09 & -0.99 \\
1.71 & -0.14 & 3.03 \\
-2.11 & 0.28 & 3.03 \\
\end{bmatrix}
\]

Sorting replicates according to values of $X_3$.

William Castaings
IMPEC, 13 Oct 2008
Approach based on rLHS

Tutorial example for sample generation

- \( X_i \ (i = 1, 2, 3) \sim U[-\pi, \pi] \)
- rLHS sample, number of variables \( k = 3 \), base sample size \( N = 4 \), number of replicates \( r = 2 \)

| \( f_X(x) \) (Base sample) |  
|---|---|---|---|
| -2.11 | -2.38 | -2.18 |
| 0.79 | -0.14 | 0.53 |
| 1.71 | 0.28 | 3.03 |
| -0.39 | 3.09 | -0.99 |

| 1\(^{st}\) Replicate |  
|---|---|---|---|
| -2.11 | 3.09 | -2.18 |
| -0.39 | -0.14 | -0.99 |
| 0.79 | 0.28 | 3.03 |
| 1.71 | -2.38 | 0.53 |

| 2\(^{nd}\) Replicate |  
|---|---|---|---|
| 0.79 | -2.38 | -2.18 |
| -0.39 | 3.09 | -0.99 |
| 1.71 | -0.14 | 3.03 |
| -2.11 | 0.28 | 0.53 |

Conditional samples for \( f_{X|X_3}(x) \)

| \( f_{X|X_3}(x|X_3 = x_3^{(1)}) \) |  
|---|---|---|---|
| -2.11 | 3.0980 | -2.18 |
| 0.79 | -2.3887 | -2.18 |

| \( f_{X|X_3}(x|X_3 = x_3^{(2)}) \) |  
|---|---|---|---|
| -0.39 | -0.1409 | -0.99 |
| -0.39 | 3.0980 | -0.99 |

| \( f_{X|X_3}(x|X_3 = x_3^{(3)}) \) |  
|---|---|---|---|
| 1.71 | -2.3887 | 0.53 |
| -2.11 | 0.2820 | 0.53 |

| \( f_{X|X_3}(x|X_3 = x_3^{(4)}) \) |  
|---|---|---|---|
| 0.79 | 0.28 | 3.03 |
| 1.71 | -0.14 | 3.03 |
Sample generation for dependent inputs

- Replicated Latin Hypercube Sampling (*McKay, 1995*)
  - \( r \) matrices of size \( N \) generated through column permutation
  - Correlations induced through permutations (*Iman et al, 1987; Stein et al, 1987*)
Sample generation for dependent inputs

- Replicated Latin Hypercube Sampling (McKay, 1995)
  - $r$ matrices of size $N$ generated through column permutation
  - Correlations induced through permutations (Iman et al, 1987; Stein et al, 1987)

- Sample size used for approaching $f_{X|X_i}(x)$ (i.e. $N_{int}$) is given by the number of replicates $r$
- Number of values of $x_i$ explored (i.e. $N_{ext}$) for estimation $E_{X_i}[s(X_i)]$ given by the base sample size $N$
Sample generation for dependent inputs

- Replicated Latin Hypercube Sampling (*McKay, 1995*)
  - \( r \) matrices of size \( N \) generated through column permutation
  - Correlations induced through permutations (*Iman et al, 1987; Stein et al, 1987*)

\[
\text{COST} = r \times N
\]

- Sample size used for approaching \( f_{X|X_i}(x) \) (i.e. \( N_{int} \)) is given by the number of replicates \( r \)
- Number of values of \( x_i \) explored (i.e. \( N_{ext} \)) for estimation \( E_{X_i}[s(X_i)] \) given by the base sample size \( N \)
Sample generation for dependent inputs

- Replicated Latin Hypercube Sampling (*McKay, 1995*)
  - $r$ matrices of size $N$ generated through column permutation
  - Correlations induced through permutations (*Iman et al, 1987; Stein et al, 1987*)

\[ \text{COST} = r \times N \]

- $r$ should be close to $N$ in order to ensure that
\[ \int |f_{Y|X_i} - f_Y(y)| dy < \varepsilon \text{ if } X_i \text{ is a dummy input} \]

\[ \text{COST} \sim N^2 \]

Number of model evaluations independent from $k$
Brute force approach

- $X_i \ (i = 1, \cdots, 4) \sim U[-\pi, \pi]$
- Ishigami function $f(Y) = \sin X_1 + 7 \sin^2 X_2 + 0.1 X_3^4 \sin X_1$
- $X_4$ is a dummy input

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_i$</td>
<td>0.3139</td>
<td>0.4424</td>
<td>0.</td>
<td>0.</td>
</tr>
<tr>
<td>$\tilde{\delta}_i$</td>
<td>0.2110</td>
<td>0.4073</td>
<td>0.1568</td>
<td>0.</td>
</tr>
</tbody>
</table>
Brute force approach

\[ V(Y|X_i) \]

\[ E(Y|X_i) \]
Brute force approach
rLHS approach
Validation for independent inputs

- Number of replicates essential in order to ensure that
  \[ \int |f_{Y|X_i} - f_Y(y)|dy < \varepsilon \] for dummy input factors
rLHS approach

Validation for independent inputs

- Lack of correspondence for $X_i$ lead to approximation error for $s(X_i)$
- Reasonable accuracy for $\delta_i$ estimates

$rLHS$ approach

$\delta_1 = 0.21, \delta_2 = 0.40$
$\delta_2 = 0.16, \delta_4 = 0.06$

Brute force approach

$\delta_1 = 0.20, \delta_2 = 0.40$
$\delta_2 = 0.13, \delta_4 = 0.$
rLHS approach
Effect of dependence among inputs

- \( X_i \ (i = 1, \ldots, 4) \sim U[01] \)
- \( f(Y) = X_1 + X_2 + X_3 \)
- \( X_4 \) is a dummy input
rLHS approach

Effect of dependence among inputs

- $X_i \ (i = 1, \ldots, 4) \sim U[01]
- f(Y) = X_1 \cdot X_2 \cdot X_3
- X_4$ is a dummy input
• The presence of correlations or/and interactions increases the approximation error

• The additional terms generated by dependence create a non-null effect for a dummy factor

• Correlations increase the importance of the correlated parameters for both VB and MI

• In the presence of interactions, influence on other factors can be different
Conclusions

- Moment-independent importance measures with interesting properties
- Any shortcut is prone to substantial approximation errors when dealing with the entire PDF
- Computationally intensive assessment
- Calculation methods for independent and dependent model inputs, other sampling plans to be investigated

Rather than the entire PDF, a specific portion might be of interest
Ex. Focus on the variability of extremes

- Relative importance of model inputs in determining the variability of extremes
  \[ s(x_i) = \int |f_Y(y|Y > y^{90%}) - f_{Y|X_i}(y|X_i, Y > y^{90%})|dy \]

Monte Carlo Filtering
- Select the sample points verifying \( Y > y^{90%} \)
- Induced correlation structure for \( f_X(x|Y > y^{90%}) \)
- Conditional samples generation (i.e. \( f_{Y|X_i}(y|X_i, Y > y^{90%}) \)) might be difficult

Adaptation of importance measure
- Restrict the area calculation to the targeted region of the model output
  \[ s(x_i) = \int_{\Omega} |f_Y(y) - f_{Y|X_i}(y|X_i)|dy \]
Merci de votre attention ...

Q?