Generalized Sobol' index crib sheet

Saint-Venant flood model

Height of flood, $H_{\text{flood}} \equiv$ river base level + river height $= Z_v + H$
Height of barrier, $H_{\text{barrier}} \equiv$ bank height + dyke height $= C_b + H_d$
Overflow $S = H_{\text{flood}} - H_{\text{barrier}}$, prefer $S < 0$.

Index subsets

$\mathcal{D} = \{1, 2, \ldots, d\}$. For $u \subseteq \mathcal{D}$, $|u|$ is cardinality $-u$ or u^c is complement.
For $u \neq \varnothing$, $\lceil u \rceil = \max\{1 \leq j \leq d \mid j \in u\}$.

ANOVA of $L^2[0,1]^d$, Efron-Stein (1981), Hoeffding (1948), Sobol' (1969)

$f(\boldsymbol{x}) = \sum_{u \subseteq \mathcal{D}} f_u(\boldsymbol{x})$. For $u \neq \varnothing$, $\sigma_u^2 = \int f_u(\boldsymbol{x})^2 \,\mathrm{d}\boldsymbol{x}$, and $\sigma_\varnothing^2 = 0$.

$u \neq v \implies \int f_u(\boldsymbol{x}) f_v(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} = 0$, Variance $\sigma^2 = \sum_u \sigma_u^2$. $\mu = \int f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$

Subset importance measures

$\underline{\tau}_u^2 = \sum_{v \subseteq u} \sigma_v^2$ $\overline{\tau}_u^2 = \sum_{v \cap u \neq \varnothing} \sigma_v^2$ $\Upsilon_u^2 = \sum_{v \supseteq u} \sigma_v^2$ Mean dimension $\sum_{u \subseteq \mathcal{D}} |u| \sigma_u^2 / \sigma^2$.

Frankenpoints

$\boldsymbol{y} = \boldsymbol{x}_u{:}\boldsymbol{z}_{-u}$ means $y_j = \begin{cases} x_j, & j \in u \\ z_j, & j \notin u. \end{cases}$

Basic identities

$\int f(\boldsymbol{x}) f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}) \,\mathrm{d}\boldsymbol{x} = \mu^2 + \underline{\tau}_u^2$ $\frac{1}{2} \int \big(f(\boldsymbol{x}) - f(\boldsymbol{x}_{-u}{:}\boldsymbol{z}_u)\big)^2 \,\mathrm{d}\boldsymbol{x} = \overline{\tau}_u^2$

GSIs

$\mathrm{NXOR}(u,v) = \mathrm{XOR}(u,v)^c = \big((u \cap v^c) \cup (u^c \cap v)\big)^c = \cdots = (u \cap v) \bigcup (u^c \cap v^c)$

$\Theta_{uv} \equiv \iint f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}) f(\boldsymbol{x}_v{:}\boldsymbol{z}_{-v}) \,\mathrm{d}\boldsymbol{x} \,\mathrm{d}\boldsymbol{z} = \mu^2 + \underline{\tau}^2_{\mathrm{NXOR}(u,v)}$

$\mathrm{GSI} = \sum_{u \subseteq \mathcal{D}} \sum_{v \subseteq \mathcal{D}} \Omega_{uv} \Theta_{uv} = \mathrm{tr}\big(\Omega^{\mathsf{T}} \Theta\big)$

$\widehat{\mathrm{GSI}} = \frac{1}{n} \sum_{i=1}^{n} f(\boldsymbol{x}_{i,u}{:}\boldsymbol{z}_{i,-u}) f(\boldsymbol{x}_{i,v}{:}\boldsymbol{z}_{i,-v})$ where $\boldsymbol{x}_i, \boldsymbol{z}_i \overset{\text{iid}}{\sim} \mathbf{U}[0,1]^d$