2/26/2025

Tuesday, February 25, 2025 7:55 PM

$\boxed{\text{Sobol Indices}}$

$y = F(\vec{\theta}),\ \vec{\theta} \in \Gamma,\ \Gamma = [0,1]^p$

$$\rightarrow F(\vec{\theta}) = f_0 + \sum_{i=1}^{p} f_i(\theta_i) + \sum_{1 \le i < j \le p} f_{ij}(\theta_i, \theta_j) + \dots$$

$$f_0 = \int_\Gamma F(\vec{\theta})\, d\vec{\theta} = E[F(\vec{\theta})]$$

$$f_i(\theta_i) = \int_{\Gamma^{p-1}} F(\vec{\theta})\, d\vec{\theta}_{\sim i} - f_0 = E[F(\vec{\theta}) \mid \theta_i] - f_0$$

$$f_{ij}(\theta_i, \theta_j) = \int_{\Gamma^{p-2}} F(\vec{\theta})\, d\vec{\theta}_{\sim i,j} - f_i - f_j - f_0 = E[F(\vec{\theta}) \mid \theta_i, \theta_j] - f_i - f_j - f_0$$

- Let the Variance of $F(\vec{\theta}) = y$

$$D = Var[y] = \int_\Gamma (F(\theta))^2\, d\vec{\theta} - f_0^2$$

- Since each f term is orthogonal, then

$$D = \sum_{i=1}^{p} D_i + \sum_{1 \le i < j \le p} D_{ij} + \dots$$

where $D_i = \int_0^1 f_i^2(\theta_i)\, d\theta_i = Var[f_i(\theta_i)]$

Because $Var[f_i] = E[f_i^2] - (E[f_i])^2$ + $E[f_i] = \int_0^1 f_i(\theta_i)\, d\theta_i = 0$

$$D_{ij} = \int_0^1 \int_0^1 f_{ij}^2(\theta_i, \theta_j)\, d\theta_i\, d\theta_j = Var[f_{ij}(\theta_i, \theta_j)]$$

Sobol' Indices

- the sobol' Indices are given by

$$S_i = \frac{D_i}{D} = \frac{Var[f_i(\theta_i)]}{Var[F(\vec{\theta})]}$$ ⇒ The "first-order" Sobol' index describing the proportion of variance attributed to θ_i.

$$S_{ij} = \frac{D_{ij}}{D} = \frac{Var[f_{ij}(\theta_i, \theta_j)]}{Var[F(\vec{\theta})]}$$ ⇒ Second order Sobol' index " . . . . " attributed to θ_i-θ_j interaction.

Note: $\sum_{i=1}^{p} S_i + \sum_{1 \le i < j \le p} S_{ij} + \dots = 1$

- We use S_i (+ S_{ij} occasionally) to define an importance index for how θ_i impacts our system.

- We also have the total sensitivity index

$$S_{T_i} = S_i + \sum_{\substack{j=1 \\ j \neq i}}^{p} S_{ij} + \sum_{\substack{j,k=1 \\ j,k \neq i}}^{p} S_{ijk} + \dots$$

- S_{T_i} assesses the global effects of θ_i on $F(\vec{\theta})$ through all possible interactions.
- we often don't go above S_{ij}

- Note that if $\sum_{i=1}^{p} S_i \approx 1$, then first-order effects dominate

- also $$S_{T_i} \approx S_i + \sum_{\substack{j=1 \\ j \neq i}}^{p} S_{ij} \Rightarrow S_{T_i} - S_i = \sum_{\substack{j=1 \\ j \neq i}}^{p} S_{ij}$$

⇒ $S_{T_i} - S_i$ reflects higher order interactions

$\Rightarrow \quad S_{T_i} - S_i$ reflects higher order interactions

- We can interpret S_i via $\left(F(\vec{\theta}) = Y\right)$

$$S_i = \frac{Var\left[E\left[Y \mid \theta_i\right]\right]}{Var(Y)}$$

- We can also write

$$S_{T_i} = 1 - \frac{Var\left[E\left[Y \mid \vec{\theta}_{\sim i}\right]\right]}{Var[Y]}$$ "1 − % Variance attributed to all but θ_i"

$$= \frac{E\left[Var\left[Y \mid \vec{\theta}_{\sim i}\right]\right]}{Var[Y]}$$ "Expected change in Y for $\vec{\theta}_{\sim i}$ fixed"

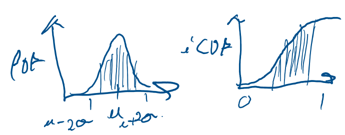
- So S_{T_i} reflects the contributions of θ_i on $F(\vec{\theta}) = Y$.

- We can always express each θ_i with their own PDF, so long as θ_i are independent

- If $\underline{S_{T_i}} \approx 0$ (or $< \eta$, $\eta \in [0,1]$), then θ_i is functionally non-influential

Computation

- We need to approximate $Var\left[E\left[Y \mid \theta_i\right]\right]$ + $Var[Y]$.
  - We often employ Monte Carlo (MC) sampling to get these estimates.
    - ▫ MC Sampling draws random values from a known pdf



★ Why not use quadrature?

- For high dimensions ($P = 5$), MC outperform

- If we approximate $E[Y|\theta_i]$, we need M samples. ($M \approx 500 - 5000$)
  - We need to replicate this M times for variance
  - With P parameters, each need $E[Y|\theta_i]$

  $\Rightarrow O(P \cdot M^2)$ evaluations

Algorithms

- First, recall that $Var[Y] = E[Y^2] - E[Y]^2$

Because

$$f_i = E[Y|\theta_i] - f_0$$

$$E[f_i] = 0$$

- Note

$$Var\left[E[Y|\theta_i]\right] = Var\left[f_i(\theta_i) + f_0\right] = E\left[(f_i(\theta_i) + f_0)^2\right] - \underbrace{E\left[f_i(\theta_i) + f_0\right]^2}_{E[f_0]^2}$$

$$= E\left[\underbrace{E[Y|\theta_i]^2}_{(*)}\right] - f_0^2$$

where $(*)$ is the mean when all but θ_i are varied, ie

$$E\left[E[Y|\theta_i]^2\right] = \int_{\Gamma^{P-1}} E[Y|\theta_i]^2 \, d\vec{\theta}_{\sim i}$$

- This gives rise to the Saltelli Algorithm