Lecture 21: Variance Reduction Methods and Sensitivity Analysis

Today’s Topics
1. Bootstrapping
2. Variance reduction methods
3. Importance sampling
4. Sensitivity analysis
1. Bootstrapping

• How do we get estimates of the standard errors in our estimators that don't have known distributions? e.g., in our estimate for the variance.
2. Variance Reduction Methods

• used to increase the accuracy of the Monte Carlo estimates that can be obtained for a given number of iterations

• "tricks" to make our MCS more "statistically efficient“ – more accuracy for a given number of samples, or fewer samples to achieve a given level of accuracy

• variance reduction methods: importance sampling, antithetic sampling, control variates, stratified sampling
3. Importance Sampling

• a general technique for estimating properties of one distribution while only having samples generated from another (different) distribution
3.1 Importance sampling general idea

A technique for estimating properties of one distribution while only having samples generated from another (different) distribution.

Consider random variable $X$, pdf $f_X(x)$.

\[ E_X[X] = \int x f_X(x) \, dx \]

\( E_X[X] \) denotes expectation under $f_X$.

Choose a random variable $Z \geq 0$ s.t.

\[ E_X[Z] = 1 = \int Z f_X(x) \, dx \]

\( E_X[Z] \) under $f_X$. 

\[ f_Z(x) \]
3.1 Importance sampling general idea

\[ \mu \triangleq \mathbb{E}_\chi [\chi] = \int \chi f_\chi (\chi) \, d\chi \]

\[ = \int \frac{\chi}{2} f_\chi (\chi) \, d\chi \]

\[ = \int \frac{\chi}{2} f_\chi (\chi) \, d\chi \quad \frac{\chi}{2} \]
3.1 Importance sampling general idea

We have

\[ E_X [X] = E_Z \left[ \frac{X}{Z} \right] \]

sample \( X \) under \( f_X(x) \)

estimator variance is

\[ \frac{\text{var}_X [X]}{N} \]

sample \( \frac{X}{Z} \) under \( f_Z(x) \)

estimator variance is

\[ \frac{\text{var}_Z \left[ \frac{X}{Z} \right]}{N} \]
3.1 Importance sampling general idea

We have

$$E_X[\frac{X}{Z}] = \frac{E_Z\left[\frac{X}{Z}\right]}{\int_{f_Z(x)}dx}$$

- Sample \( x \) under \( f_X(x) \)
- Sample \( \frac{X}{Z} \) under \( f_Z(x) \)

Estimator variance is

$$\text{var}_X[X] = \frac{\text{var}_Z\left[\frac{X}{Z}\right]}{N}$$

- Some values of \( X \) in our MCS have more impact on the parameter being estimated (here \( \mu_X \)) than others
- If "important" values are emphasized by sampling more frequently, then estimator variance can be reduced
- Key is to choose an appropriate "biasing distribution" → good choice means we can decrease \( N \) for same accuracy
3.2 Importance sampling for probability estimation

We saw $P(E_{A3})$ is estimated via MCS with $p(A_i)$. Consider $A$ is the event $y > y_{\text{limit}}$ (e.g., prob. of failure).

Define indicator function $I(A_i) = \begin{cases} 1, & y_i > y_{\text{limit}} \\ 0, & y_i \leq y_{\text{limit}} \end{cases}$

Then $p(A) = \frac{1}{N} \sum_{i=1}^{N} I(A_i)$

and $E[p(A)] = P(E_{A3})$ (unbiased)

$\text{var}[p(A)] = \frac{P(E_{A3})(1 - P(E_{A3}))}{N}$
3.2 Importance sampling for probability estimation

Introduce pdf \( w(x) \) (alternative pdf for \( x \))
→ called "biasing density"

→ choose \( w(x) \) so that event \( A \) occurs more frequently

Then

\[ P(A) = E_x[1(A)] = \int I(A) f(x) \, dx \]

\[ = \int I(A) \frac{f(x)}{w(x)} \, w(x) \, dx \]

\[ = E_w \left[ I(A) \frac{f(x)}{w(x)} \right] \]

↑ draw from pdf \( \frac{f(x)}{w(x)} \) to counter \( w(x) \)

↑ weight by \( \frac{f(x)}{w(x)} \)
3.2 Importance sampling for probability estimation

Then our MC IS estimator for $\text{PEA}_B$ is

$$\hat{\pi}_{IS}(A) = \frac{1}{N} \sum_{i=1}^{N} I(A_i) \frac{f(x_i)}{w(x_i)}$$

where the $x_i$ are drawn from $w(x)$

Summary:

- define $w(x)$
- draw samples of $x$ from $w(x)$ (idea: more samples in region of interest)
- estimate $\text{PEA}_B$, but we need to weight the samples by $\frac{f(x_i)}{w(x_i)}$ to account for the fact that we drew from $w(x)$, not $f(x)$

Then:

$$E \left[ \hat{\pi}_{IS}(A) \right] = \text{PEA}_B \quad \text{(unbiased)}$$

$$\text{var}_w \left[ \hat{\pi}_{IS}(A) \right] = \frac{1}{N} \left( E \left[ I(A) \frac{f(x)}{w(x)} \right] - (\text{PEA}_B)^2 \right)$$
3.3 How to pick the biasing distribution

Simple approach: scaling to shift probability mass into the event region

\[ w(x) = \frac{1}{a} f\left(\frac{x}{a}\right) \]

\[ a > 1 \]
3.3 How to pick the biasing distribution

\[ w(x) = f(x-c), \quad c > 0 \quad (\text{for } P(x > x_{\text{crit}})) \]
4. Sensitivity Analysis

• How do we use our MCS results to understand which uncertain inputs are contributing the most to output variability?
  
  – Important to understand where we should focus our uncertainty reduction efforts (improving manufacturing tolerances, improving models, installing sensors, etc.)
    → factor prioritization
  
  – Important to understand where there may be uncertainties but they are not important
    → factor fixing
4a. Vary-all-but-one (VABO) MCS

1. Run a MCS with all inputs varying

2. Fix input k to a deterministic value. Rerun the MCS with all inputs except k varying.

3. Compare statistics of the output (e.g., variance of Run #1 to variance of Run #2).

Questions:

• at what value should we fix factor k?

• would the results be different if we fixed factor k to a different deterministic value?

• what about possible interactions among the inputs?
4b. Global Sensitivity Analysis

\[
\begin{align*}
X_1 & \xrightarrow{} \text{model} \xrightarrow{} Y \\
X_2 & \xrightarrow{} \\
X_d & \xrightarrow{}
\end{align*}
\]

- d random inputs
- 1 random output

Main effect sensitivity index for input i:

\[
S_i = \frac{\text{var}(Y) - E[\text{var}(Y|X_i)]}{\text{var}(Y)}
\]

Relative expected reduction in output variance if the true value of \(X_i\) is learned (expectation over what that true value might be)

Measure of the effect of varying \(X_i\) alone, averaged over variations in other inputs
4b. Global Sensitivity Analysis

Also can compute higher-order interaction indices

\[ S_{ij}, S_{ijk}, \ldots \text{ etc.} \]

\[ \text{Var}(Y) \implies \begin{array}{c}
\text{Input } X_2 \\
\text{Input } X_1 \\
X_1X_2 \text{ interaction}
\end{array} \]

\[ S_1 + S_2 + S_3 = 1 \]

Generally

\[ \sum_{i=1}^{d} S_i + \sum_{i<j}^{d} S_{ij} + \ldots + S_{12\ldots d} = 1 \]

Total effect sensitivity index:

For input \( i \)

\[ S_{T_i} = \frac{\text{Var}(Y) - E[\text{Var}(Y/X_{ni})]}{\text{Var}(Y)} \]

\( \rightarrow \) measures contribution to \( \text{Var}(Y) \) of \( X_i \) including its main effect and all the interaction effects