Sampling Techniques

MS&E348
Winter 2011/2012
Professor Gerd Infanger
Multidimensional integration

\[ V = (V_1, \ldots, V_h) \] random vector (independently distributed)

\[ V \] has realizations \( v \) or \( v^\omega \) with corresponding probability \( p(v) \) or \( p(v^\omega) \)

\( \Omega \) set of possible realizations

\[ \Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_h \]

\( z(V) \) cost function (random)
We want to compute the expected cost $E z(V)$

$$E z(V) = \int z(v) P(dv) = \int \int \cdots \int z(v)p(v)dv_{1} \cdots dv_{h}$$

where $p(v) = p(v_{1})p(v_{2}) \cdots p(v_{h})$

In the case of discrete distributions

$$E z(V) = \sum_{v_{1}} \sum_{v_{2}} \cdots \sum_{v_{h}} z(v)p(v)$$

$$E z(V) = \sum_{\omega_{1}} \sum_{\omega_{2}} \cdots \sum_{\omega_{h}} z(\omega)p(\omega)$$

where $\omega = (\omega_{1}, \omega_{2}, \ldots, \omega_{h})$, and $\omega_{1} \in \Omega_{1}, \omega_{2} \in \Omega_{2}, \ldots, \omega_{h} \in \Omega_{h}$
Monte Carlo methods

- Monte Carlo Methods are promising for multidimensional integration
- For example, determine the average height of a board:
Crude Monte-Carlo

\( v^\omega \) are observations sampled independently (with replacement) from their joint probability mass function \( p(v) = p(v_1)p(v_2)\ldots p(v_h) \)

Then \( z^\omega = z(v^\omega) \) are independent random variates with expectation \( z = E z(v) \)

\[ \hat{z} = \frac{1}{N} \sum_{\omega=1}^{N} z(v^\omega) \]

is an unbiased estimator of \( z \)
Its variance is

$$\sigma^2_{\tilde{z}} = \frac{1}{N} \sigma^2$$

where $\sigma^2 = \text{var} \ z(V)$ is the (population) variance of $z(V)$

$\tilde{z}$ is called the sample mean

$\sigma^2_{\tilde{z}}$ is called the variance of the sample mean

$\sigma^2$ is usually not known. It can be estimated

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{\omega=1}^{N} (z(v^\omega) - \tilde{z})^2$$

The true value of $\sigma^2$ is

$$\sigma^2 = \int \int \ldots \int (z(v) - z)^2 p(v) dv_1, \ldots dv_h$$
Motivational example: importance sampling

\[ F = F_1 + F_2 \]

\[ z_1, \text{ assume } z_2 = 0 \]
$$z = z_1 p(F_1) + z_2 p(F_2)$$

$$z = z_1 \frac{F_1}{F} + z_2 \frac{F_2}{F}$$

Assuming we know that $z_2 = 0$, we can estimate

$$\hat{z} = \hat{z}_1 \frac{F_1}{F}$$

We want to concentrate the sampling where it matters

$$\sigma_{\hat{z}}^2 = \left( \frac{F_1}{F} \right)^2 \sigma_{\hat{z}_1}^2$$
Variance reduction techniques

- Stratified sampling
- Importance sampling ★
- Control variates ★
- Antithetic variates
Importance sampling

\[ z = \sum_{\omega \in \Omega} z(\nu^\omega) p(\nu^\omega) \]

Suppose \( \Gamma(\nu^\omega) \approx z(\nu^\omega) \), and \( \Gamma = E \Gamma(\nu^\omega) \) known

\[ z = \sum_{\omega \in \Omega} z(\nu^\omega) \frac{\Gamma(\nu^\omega)}{\Gamma(\nu^\omega) \Gamma} p(\nu^\omega) \]

and we interpret it as

\[ z = \Gamma E \frac{z(\nu^\omega)}{\Gamma(\nu^\omega)} \]

distributed according to probability mass function

\[ q(\nu^\omega) = \frac{\Gamma(\nu^\omega)}{\Gamma} p(\nu^\omega) \]
Checking if \( q(v^\omega) = \frac{\Gamma(v^\omega)}{\Gamma} p(v^\omega) \) is a density:

Since \( \Gamma = \sum_{\omega \in \Omega} \Gamma(v^\omega) p(v^\omega) \)

\[
\sum q(v^\omega) = \frac{\sum_{\omega \in \Omega} \Gamma(v^\omega) p(v^\omega)}{\Gamma} = 1
\]

Since \( p(v^\omega) \geq 0 \), if all \( \Gamma(v^\omega) \geq 0 \), \( \Gamma \geq 0 \)

\[
q(v^\omega) = \frac{\Gamma(v^\omega)}{\Gamma} p(v^\omega) \geq 0
\]

If all \( \Gamma(v^\omega) \leq 0 \), \( \Gamma \leq 0 \)

\[
q(v^\omega) = \frac{-\Gamma(v^\omega)}{-\Gamma} p(v^\omega) \geq 0
\]
We obtain the estimate

\[ \hat{z} = \Gamma \frac{1}{N} \sum_{\omega=1}^{N} \frac{z(n^\omega)}{\Gamma(n^\omega)} \]

by sampling a sample \( S \) of size \( |S| = N \) from the distribution \( q(n^\omega) \)

And its variance

\[ \sigma_{\hat{z}}^2 = \frac{1}{N} \sum_{\omega \in \Omega} \left( \frac{z(n^\omega)}{\Gamma(n^\omega)} - z \right)^2 q(n^\omega) \]
The best importance distribution

One can see easily if $\Gamma(v^\omega) = z(v^\omega)$ the $\sigma_{\hat{z}}^2 = 0$ and

$$q^*(v^\omega) = \frac{z(v^\omega)}{z} p(v^\omega)$$

reflects the best importance distribution

If $\Gamma(v^\omega) \approx z(v^\omega)$ we obtain a “good” importance distribution, which is approximately proportional to the product $z(v^\omega)p(v^\omega)$
Additive and multiplicative approximation functions

Since $p(v^\omega) = \prod_{i=1}^h p_i(v_i^\omega)$,

if $\Gamma(v^\omega) = \sum_{i=1}^h \Gamma_i(v_i^\omega)$,

$$\Gamma = \sum_{i=1}^h E \Gamma_i(v_i^\omega) = \sum_{i=1}^h \sum_{\omega \in \Omega} \Gamma_i(v_i^\omega) p_i(v_i^\omega)$$

and if $\Gamma(v^\omega) = \prod_{i=1}^h \Gamma_i(v_i^\omega)$,

$$\Gamma = \prod_{i=1}^h E \Gamma_i(v_i^\omega) = \prod_{i=1}^h \sum_{\omega \in \Omega} \Gamma_i(v_i^\omega) p_i(v_i^\omega)$$
Multiplicative marginal cost model

\[ \Gamma(\omega) = z(\tau) \prod_{i=1}^{h} \Gamma_i(\omega_i), \]

\[ \Gamma_i(\omega_i) = \frac{z(\tau_1, \ldots, \tau_{i-1}, \omega_i, \tau_{i+1}, \ldots, \tau_h)}{z(\tau)} \]

yields

\[ z = \Gamma \sum_{\omega \in \Omega} \left[ \frac{z(\omega)}{\Gamma(\omega)} \right] \prod_{i=1}^{h} \frac{\Gamma_i(\omega_i)}{\Gamma_i} p_i(\omega_i) \]

The base case \( \tau \) can be chosen freely.
Variance Estimation

We estimate

\[ \sigma^2 = \text{var} \left[ \frac{z(\nu \omega)}{\Gamma(\nu \omega)} \right] \]

using sample $S$ (importance distribution) and obtain

\[ \hat{\sigma}^2_{\tilde{z}} = \frac{1}{N} \hat{\sigma}^2 \]
Additive marginal cost model

\[ \Gamma(v^\omega) = z(\tau) + \sum_{i=1}^{h} \Gamma_i(v_i^\omega), \]

\[ \Gamma_i(v_i^\omega) = z(\tau_1, \ldots, \tau_{i-1}, v_i^\omega, \tau_{i+1}, \ldots, \tau_h) - z(\tau) \]

yields

\[ z = z(\tau) + \sum_{i=1}^{h} \Gamma_i \sum_{\omega \in \Omega} \left[ \frac{z(v^\omega) - z(\tau)}{\sum_{i=1}^{h} \Gamma_i(v_i^\omega)} \right] \frac{\Gamma_i(v_i^\omega)}{\Gamma_i} \prod_{j=1}^{h} p_j(v_j^\omega) \]

where we assume that \( \sum_{i=1}^{h} \Gamma_i(v_i^\omega) > 0 \), so that at least one \( \Gamma_i(v_i^\omega) > 0 \).
Sample size and variance

Sample size $N = \sum_{i=1}^{h} N_i$

Assuming equal variance $\sigma_i^2$ of each term, choose $N_i$ approximately proportional to $\Gamma_i$

We estimate

$$\sigma_i^2 = \text{var} \left[ \frac{z(v^\omega) - z(\tau)}{\sum_{i=1}^{h} \Gamma_i(v_{i,i}^\omega)} \right]$$

using sample $S_i$ (importance distribution) and obtain

$$\hat{\sigma}^2 = \sum_{i=1}^{h} \frac{\Gamma_i^2 \hat{\sigma}_i^2}{\Gamma_i}$$
Control variates

Suppose $\Gamma(\nu^\omega) \approx z(\nu^\omega)$, i.e., positively correlated, and $\Gamma = E \Gamma(\nu^\omega)$ known

$$z = \sum_{\omega \in \Omega} z(\nu^\omega) p(\nu^\omega)$$

$$z = \sum_{\omega \in \Omega} [z(\nu^\omega) - \alpha \Gamma(\nu^\omega)] p(\nu^\omega) + \alpha \Gamma$$

estimator:

$$\hat{z} = \frac{1}{N} \sum_{\omega=1}^{N} [z(\nu^\omega) - \alpha \Gamma(\nu^\omega)] + \alpha \Gamma$$

by sampling a sample $S$ of size $N = |S|$ from the distribution $p(\nu^\omega)$
Variance estimation

\[ \sigma^2_{\tilde{z}} = \frac{1}{N} \text{var}[z(\nu^\omega) - \alpha \Gamma(\nu^\omega)] \]

\[ \sigma^2_{\tilde{z}} = \frac{1}{N} [\text{var}z(\nu^\omega) + \alpha^2 \text{var} \Gamma(\nu^\omega) - 2\alpha \text{cov}(z(\nu^\omega), \Gamma(\nu^\omega))] \]

where

\[ \alpha^* = \frac{\text{cov}(z(\nu^\omega), \Gamma(\nu^\omega))}{\text{var} \Gamma(\nu^\omega)} \]

represents the best value of \( \alpha \) minimizing the expression for the variance, usually estimated
Approximation functions

Additive Approximation Function

\[ \Gamma(\nu^\omega) = z(\tau) + \sum_{i=1}^{h} \Gamma_i(\nu_i^\omega), \]

\[ \Gamma_i(\nu_i^\omega) = z(\tau_1, \ldots, \tau_{i-1}, \nu_i^\omega, \tau_i+1, \ldots, \tau_h) - z(\tau) \]

Multiplicative Approximation Function

\[ \Gamma(\nu^\omega) = z(\tau) \prod_{i=1}^{h} \Gamma_i(\nu_i^\omega), \]

\[ \Gamma_i(\nu_i^\omega) = \frac{z(\tau_1, \ldots, \tau_{i-1}, \nu_i^\omega, \tau_i+1, \ldots, \tau_h)}{z(\tau)} \]