Efficiency-Improvement Techniques

Reading: Ch. 11 in Law & Ch. 10 in Handbook of Simulation

Peter J. Haas

CS 590M: Simulation
Spring Semester 2020

Efficiency-Improvement Techniques

- Overview
- Common Random Numbers
- Antithetic Variates
- Conditional Monte Carlo
- Control Variates
- Importance Sampling
  - Likelihood ratios
  - Rare-event estimation

Many Different Techniques

- Common random numbers
- Antithetic variates
- Conditional Monte Carlo
- Control variates
- Importance sampling
- Stratified sampling
- Latin hypercube sampling (HW #1)
- Quasi-random numbers
- ...

Variance Reduction and Efficiency Improvement

Typical goal is variance reduction

- I.e., reduce variance of estimator α_n of α
- Narrower CIs $\Rightarrow$ less computational effort for given precision
- So methods often called "variance reduction" methods

Care is needed when evaluating techniques

- Reduction in effort must outweigh increased cost of executing V-R method
- Increase in programming complexity?
- In many cases, additional effort is obviously small
- What about more complicated cases?

Comparing Efficiency-Improvement Schemes

Trading off statistical and computational efficiency

- Suppose $\alpha = E[X] = E[Y]$
- Should we generate i.i.d. replicates of X or Y to estimate α?
- Assume large but fixed computer budget c
- Let $\tau_X(i)$ be (random) time to generate X_i
- Assume that $\big(X_1, \tau_X(1)\big), \big(X_2, \tau_X(2)\big), \ldots$ are i.i.d.
- Number of X-observations generated within budget c is $N_x(c) = \max\{n \geq 0 : \tau_X(1) + \cdots + \tau_X(n) \leq c\}$
- So estimator based on budget is $\alpha_X(c) = \frac{1}{N_X(c)} \sum_{i=1}^{N_X(c)} X_i$

Comparing Efficiency-Improvement Schemes, Continued

Hammersley-Handscomb Efficiency Measure

- Can show: $\lim_{c\to\infty} N(c)/c = \lambda_X$ a.s., where $\lambda_X = 1/E[\tau_X]$
- Hence

$$\alpha_X(c) - \alpha = \frac{1}{N_X(c)} \sum_{i=1}^{N_X(c)} X_i - \alpha \approx \frac{1}{\lfloor \lambda_X c \rfloor} \sum_{i=1}^{\lfloor \lambda_X c \rfloor} X_i - \alpha$$

$$\overset{\mathrm{D}}{\sim} \sqrt{\frac{\mathrm{Var}[X]}{\lambda_X c}} N(0,1) = \frac{1}{\sqrt{c}} \sqrt{E[\tau_X] \cdot \mathrm{Var}[X]}\, N(0,1)$$

- Similarly,

$$\alpha_Y(c) - \alpha \overset{\mathrm{D}}{\sim} \frac{1}{\sqrt{c}} \sqrt{E[\tau_Y] \cdot \mathrm{Var}[Y]}\, N(0,1)$$

- Efficiency measure: $\frac{1}{E[\tau_Y] \cdot \mathrm{Var}[Y]}$ (holds fairly generally)

Common Random Numbers (CRN)

Applies when comparing alternate systems

- Intuition: Sharper comparisons if systems experience same random fluctuations

More precisely:

- Goal: Compare two perf. measures distributed as X and Y
- Estimate $\alpha = E[X] - E[Y] = E[X - Y]$
- Generate i.i.d. pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$
- Point estimate: $\alpha_n = (1/n) \sum_{i=1}^{n} (X_i - Y_i)$

$$\mathrm{Var}[\alpha_n] = \frac{1}{n} Var[X - Y] = \frac{1}{n} \big(\mathrm{Var}[X] + \mathrm{Var}[Y] - 2\,\mathrm{Cov}[X, Y] \big)$$

- So want $\mathrm{Cov}[X, Y] > 0$
  - Note that $\mathrm{Cov}[X, Y] = 0$ if X and Y simulated independently

CRN, Continued

Simple case: One random number per sample of X and of Y

- Use same random number: $X_i = X_i(U_i)$ and $Y_i = Y_i(U_i)$
- If $X(u)$, $Y(u)$ both $\uparrow$ (or both $\downarrow$) in u, then $\text{Cov}[X, Y] > 0$
- True for inversion method: $X_i = F_X^{-1}(U_i)$ and $Y_i = F_Y^{-1}(U_i)$

In practice

- Sync random numbers between systems as much as possible
- Use multiple random number streams, one per event
- Jump-head facility of random number generator is crucial

CRN, Continued

Example: Long-run waiting times in two GI/G/1 queues

- Suppose that
  - Interarrival times are i.i.d according to cdf G for both systems
  - Service times are i.i.d. according to cdf H_i for queue i $(i = 1, 2)$
- Use one sequence $(U_j : j \geq 0)$ to generate a single stream of interarrival times for use in both systems
- Use one sequence $(V_j : j \geq 0)$ to generate service times in both systems: $S_{1,j} = H_1^{-1}(V_j)$ and $S_{2,j} = H_2^{-1}(V_j)$ for $j \geq 1$
- Note: Need two streams $\{U_i\}$ and $\{V_i\}$
  - Systems get out of sync with only one stream

Antithetic Variates

Applies when analyzing a single system

- Intuition: Combat "luck of the draw" by pairing each realized scenario with its opposite

More precisely:

- Goal: Estimate $E[X]$
- Generate $X_1, \ldots, X_{2n}$ and set $\alpha_n = \bar{X}_{2n}$
- Suppose pairs $(X_1, X_2), (X_3, X_4), \ldots, (X_{2n-1}, X_{2n})$ are i.i.d. (possible corr. within pairs)

$$\mathrm{Var}[\alpha_{2n}] = \frac{1}{4n^2}\big(\mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_{2n}] + 2\,\mathrm{Cov}[X_1, X_2] + \cdots + 2\,\mathrm{Cov}[X_{2n-1}, X_{2n}]\big)$$

- So want $\mathrm{Cov}[X_{2j-1}, X_{2j}] < 0$ for $j \geq 1$

Antithetic Variates, Continued

Simple case: One random number per sample of X and of Y

- Use same random number: $X_i = X_i(U_i)$ and $Y_i = Y_i(1 - U_i)$
- If $X(u)$, $Y(u)$ both $\uparrow$ (or both $\downarrow$) in u, then $\mathrm{Cov}[X, Y] < 0$
- E.g., inversion method: $X_i = F_X^{-1}(U_i)$ and $Y_i = F_Y^{-1}(1 - U_i)$

Ex: Avg. waiting time of first 100 cust. in GI/G/1 queue

- Interarrival times (service times) i.i.d according to cdf G (H)
- Replication $2k - 1$: $(I_j, S_j) = \big(G^{-1}(U_j), H^{-1}(V_j)\big)$
- Replication $2k$: $(I_j, S_j) = \big(G^{-1}(1 - U_j), H^{-1}(1 - V_j)\big)$

Ex: Alternative method for GI/G/1 queue (Explain?)

- Replication $2k - 1$: $(I_j, S_j) = \big(G^{-1}(U_j), H^{-1}(V_j)\big)$
- Replication $2k$: $(I_j, S_j) = \big(G^{-1}(V_j), H^{-1}(U_j)\big)$

Warning: CRN + AV together can backfire! [Law, p. 609]

Conditional Monte Carlo

Example: Markovian GSMP $\big(X(t) : t \geq 0\big)$

- All events are simple w. exponential clock-setting dist'ns
- Simulation algorithm (up to nth state transition time T_n)
  - Generate states $S_0, \ldots, S_{n-1} \overset{D}{\sim}$ DTMC w. transition matrix R
  - Generate holding time in each S_k: $H_k \overset{D}{\sim} exp\big(\lambda(S_k)\big)$
- Goal: Estimate $\alpha = E[Z]$ with $Z = \int_0^{T_n} f\big(X(u)\big)\,du = \sum_{k=0}^{n-1} f(S_k)H_k$
- Variance reduction trick:
  - Generate states $S_0, \ldots, S_{n-1}$ as above
  - Set holding time in S_k = mean holding time = $1/\lambda(S_k)$
- Q: Why does this work?

Conditional Monte Carlo, Continued

Law of total expectation

$E\big[E[U|V]\big] = E[U]$ ex: $E[U] = \sum_i E[U|V=v_i] \cdot P(V=v_i)$

Variance decomposition

$\text{Var}[U] = \text{Var}\big[E[U|V]\big] + E\big[\text{Var}[U|V]\big] \geq \text{Var}\big[E[U|V]\big]$

Key Idea

- Simulate V and compute $\tilde{U} = E[U|V]$
- Then $\tilde{U}$ has same mean as U but smaller variance
- So generate i.i.d replicates of $\tilde{U}$ to estimate $\alpha = E[U]$

Markovian GSMP example revisited

- $U = Z = \sum_{k=0}^{n-1} f(S_k)H_k$ and $V = (S_0, \ldots, S_{n-1})$
- So estimate $E[\tilde{Z}]$ from i.i.d replicates $\tilde{Z}_1, \ldots, \tilde{Z}_m$, where

$$\tilde{Z} = E[Z|S_0, \ldots, S_{n-1}] = \sum_{k=0}^{n-1} f(S_k)E[H_k|S_k] = \sum_{k=0}^{n-1} f(S_k)\frac{1}{\lambda(S_k)}$$

Control Variates

Intuition: Exploit extra system knowledge

- Goal: Estimate $\alpha = E[X]$
- Suppose that there exists a random variable Y such that
  - Y is strongly correlated with X
  - $E[Y]$ can be computed analytically

$\mu_Y = E[Y]$

- Control variable: $C = Y - E[Y]$ $\quad E[Y - \mu_Y] = E[Y] - \mu_Y = 0$
- Controlled estimator: $X(\lambda) = X - \lambda C$
- $E[X(\lambda)] =$ $E[X - \lambda C] = E[X] - \lambda E[C] = E[X]$
- $v(\lambda) = \mathrm{Var}[X(\lambda)] = \mathrm{Var}[X] - 2\lambda \,\mathrm{Cov}[X, C] + \lambda^2 \,\mathrm{Var}[C]$
- $v(\lambda)$ is minimized at $\lambda^* =$ $\mathrm{Cov}[X, C] / \mathrm{Var}[C]$
- Minimizing variance is $v(\lambda^*) = (1 - \rho^2)\,\mathrm{Var}[X]$, where

$$\rho = \frac{\mathrm{Cov}[X,C]}{\sqrt{\mathrm{Var}[X]\cdot\mathrm{Var}[C]}} \quad = \text{correlation coefficient of } X \text{ and } C$$

Control Variates, Continued

The method

1. Simulate i.i.d. pairs $(X_1, C_1), \ldots, (X_n, C_n)$
2. Estimate λ^* by

$$\lambda_n^* = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n) C_i \Bigg/ \frac{1}{n} \sum_{i=1}^{n} C_i^2$$

3. Apply usual estimation techniques to $Z_1, \ldots, Z_n$, where $Z_i = X_i - \lambda^* C_i$ for $1 \le i \le n$

Ex: E[avg delay] for first n customers in GI/G/1 queue

- X_i = average delay in ith replication
- $V_{i,k}$ = kth service time in i replication, with $E[V_{i,k}] = 5$
- Take $C_i = (1/n) \sum_{k=1}^{n} V_{i,k} - 5$
- Q: Why is this a good choice?

Control Variates, Continued

Internal and External Controls

- C_i in queueing example is an internal control, generated internally to the simulation
- Example of an external control:
  - Simplify original simulation model M to a version M' where performance measure α' can be computed analytically
  - Generate replications of M and M' using common random numbers to obtain $(X_1, X_1'), \ldots, (X_n, X_n')$
  - Take $C_i = X_i' - \alpha'$

Multiple controls

- $X(\lambda_1, \ldots, \lambda_m) = X - \lambda_1 C^{(1)} - \cdots - \lambda_m C^{(m)}$
- Can compute $(\lambda_1^*, \ldots, \lambda_m^*)$ by solving linear syst. of equations
- Essentially, we fit a linear regression model and simulate the leftover uncertainty (i.e., the residuals)

Importance Sampling

Likelihood ratios for i.i.d. random variables

- Goal: Estimate $\alpha = E[g_n(X_0, X_1, \ldots, X_n)]$
- $X_0, \ldots, X_n$ are i.i.d. replicates of X with pmf $p(s) = P(X = s)$
- Let Y be another RV with pmf $q(s) = P(Y = s)$
- Suppose that Y is "easier" to simulate than X
- We will estimate α by simulating Y and then "correcting"

Likelihood ratio for i.i.d. random variables

$$L_n = \frac{\prod_{i=0}^{n} p(Y_i)}{\prod_{i=0}^{n} q(Y_i)}$$ (rel. likelihood of seeing $\mathbf{Y}$ under p vs under q)

- To avoid blowups, define $0/0 = 0$ and assume that $q(x) = 0 \Rightarrow p(x) = 0$ ("absolute continuity")

Importance Sampling, Continued

Likelihood-ratio identity for i.i.d. random variables

$$E[g_n(Y_0, Y_1, \ldots, Y_n)L_n] = E[g_n(X_0, X_1, \ldots, X_n)]$$

Proof

$$\begin{aligned}
&E[g_n(Y_0, \ldots, Y_n)L_n] \\
&= \sum_{s_0 \in S} \cdots \sum_{s_n \in S} g_n(s_0, \ldots, s_n) \left(\frac{\prod_{i=0}^n p(s_i)}{\prod_{i=0}^n q(s_i)} \right) P(Y_0 = s_0, \ldots, Y_n = s_n) \\
&= \sum_{s_0 \in S} \cdots \sum_{s_n \in S} g_n(s_0, \ldots, s_n) \left(\frac{\prod_{i=0}^n p(s_i)}{\prod_{i=0}^n q(s_i)} \right) \prod_{i=0}^n q(s_i) \\
&= \sum_{s_0 \in S} \cdots \sum_{s_n \in S} g_n(s_0, \ldots, s_n) \prod_{i=0}^n p(s_i) \\
&= E[g_n(X_0, \ldots, X_n)]
\end{aligned}$$

Importance Sampling, Continued

$$g_n(Y_0, \ldots, Y_n)L_n = \prod g(s_i) \frac{\prod p(s_i)}{\prod q(s_i)} = \prod g(s_i) \cdot \frac{\prod p(s_i)}{\prod g(s_i) \prod p(s_i) / \alpha} = \alpha$$

General guidance for choosing q

- Somewhat of an art (depends on details of model)
- But if $g_n(s_0, \ldots, s_n) = \prod_{i=0}^{n} g(s_i)$ for some $g \geq 0$ and we take $q(s) = g(s)p(s)/\alpha$, then $g_n(Y_0, \ldots, Y_n)L_n \equiv \alpha$ and var $= 0$
- Can't actually choose q as above (since α is unknown) but can guide choice
  - $q(s)$ is large if s is "important", i.e., $g(s)$ and/or $p(s)$ is large

Implementation

- Set $L = 1$ initially & update whenever new Y_i is generated:

$$L \leftarrow L \times \frac{p(Y_i)}{q(Y_i)} \quad \text{for } i \geq 1$$

Importance Sampling, Continued

Importance sampling for DTMCs

- Goal: Estimate $E[g_n(X_0, \ldots, X_n)]$ where $M = (X_i : i \geq 0)$ is a DTMC with initial dist'n μ and transition matrix P
- Simulate DTMC $\tilde{M} = (Y_i : i \geq 0)$ w. building blocks $\tilde{\mu}$ and $\tilde{P}$

$$L_n = \frac{\mu(Y_0) \prod_{i=1}^{n} P(Y_{i-1}, Y_i)}{\tilde{\mu}(Y_0) \prod_{i=1}^{n} \tilde{P}(Y_{i-1}, Y_i)}$$

- Assume absolute continuity: if initial state or a jump has zero probability in $\tilde{M}$, it has zero probability in M
- Can be computed incrementally: set $L = 1$ and then

$$L \leftarrow L \times \frac{\mu(Y_0)}{\tilde{\mu}(Y_0)} \quad \text{and} \quad L \leftarrow L \times \frac{P(Y_{i-1}, Y_i)}{\tilde{P}(Y_{i-1}, Y_i)} \quad \text{for } i \geq 1$$

- Can generalize to $E[g_N(X_0, \ldots, X_N)]$ where N is random

Importance Sampling, Continued

Importance sampling for GSMPs

- Goal: Estimate $E[g_t(X(u) : 0 \leq u \leq t)]$ where $G = \big(X(t) : t \geq 0\big)$ is a GSMP with bldg blocks ν, F_0, p, F
- Simulate GSMP $\tilde{G} = \big(\tilde{X}(t) : t \geq 0\big)$ with building blocks $\tilde{\nu}$, $\tilde{F}_0$, $\tilde{p}$, $\tilde{F}$ (all other building blocks, e.g., S and $E(s)$, the same)
- Assume that cdfs F_0, F, $\tilde{F}_0$, $\tilde{F}$ have pdf's f_0, f, $\tilde{f}_0$, $\tilde{f}$
- Assume absolute continuity: if jump or clock reading has zero prob. in $\tilde{G}$, it has zero prob. in G

Importance Sampling, Continued

Simulation algorithm for GSMPs: as usual except

- Set $L = 1$ initially
- After generating initial state $\tilde{S}_0$, set $L \leftarrow L \times \frac{\nu(\tilde{S}_0)}{\tilde{\nu}(\tilde{S}_0)}$
- After generating $\tilde{C}_{0,i}$ for e_i, set $L \leftarrow L \times \frac{f_0(\tilde{C}_{0,i};e_i,\tilde{S}_0)}{\tilde{f}_0(\tilde{C}_{0,i};e_i,\tilde{S}_0)}$
- After generating $\tilde{C}_{n,i}$ for e_i, set $L \leftarrow L \times \frac{f(\tilde{C}_{n,i};\tilde{S}_n,e_i,\tilde{S}_{n-1},e_n^*)}{\tilde{f}(\tilde{C}_{n,i};\tilde{S}_n,e_i,\tilde{S}_{n-1},e_n^*)}$
- After generating a jump $\tilde{S}_{n-1} \to \tilde{S}_n$, set $L \leftarrow L \times \frac{p(\tilde{S}_n;\tilde{S}_{n-1},e_n^*)}{\tilde{p}(\tilde{S}_n;\tilde{S}_{n-1},e_n^*)}$

Efficiency-Improvement Techniques

Overview

Common Random Numbers

Antithetic Variates

Conditional Monte Carlo

Control Variates

Importance Sampling

Likelihood ratios

Rare-event estimation

Application to Rare-Event Estimation

Example: DTMC model of machine reliability

- State space of $(X_n : n \geq 0)$: $S = \{0, 1, 2, 3\}$
  - $X_n = 0$: machine fully operational at nth inspection
  - $X_n = 1$ or 2: machine operational but degraded
  - $X_n = 3$: machine has failed
- $\nu(0) \stackrel{\Delta}{=} P(X_0 = 0) = 1$

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & 0 \\ 0 & \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

- $\mu \gg \lambda$, so failures take a long time to occur

Rare-Event Estimation, Continued

- Set $N = \min\{n > 0 : X_n = 3\}$ (time to failure)
- Goal: Estimate $\alpha = P(N \le j) = E[I(N \le j)]$ with j small
- Challenge: Event $A = \{N \le j\}$ is very rare
- Can write $\alpha = E[g_j(X_0, \ldots, X_j)]$, where

$$g_j(x_0, \ldots, x_j) = \begin{cases} 1 & \text{if } x_i = 3 \text{ for some } 0 \le i \le j; \\ 0 & \text{otherwise} \end{cases}$$

- Use importance sampling with $\lambda = \mu$
- I.e., simulate DTMC $(\tilde{X}_n : n \ge 0)$ with

$$\tilde{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array} \quad \text{and} \quad \tilde{\nu} = \nu$$

Rare-Event Estimation, Continued

Rare-Event Estimation Algorithm for Machine Reliability

1. Choose sample size n
2. Simulate $(\tilde{X}_n : n \geq 0)$ up to time $T = \min(j, N)$
3. Compute $W = I(N \leq j) \dfrac{\prod_{i=1}^{T} P(\tilde{X}_{i-1}, \tilde{X}_i)}{\prod_{i=1}^{T} \tilde{P}(\tilde{X}_{i-1}, \tilde{X}_i)}$
4. Repeat Steps 2–3 n times, independently, to produce i.i.d. replicates $W_1, \ldots, W_n$
5. Compute point estimates and confidence intervals as usual

Extensions of basic method include dynamic importance sampling