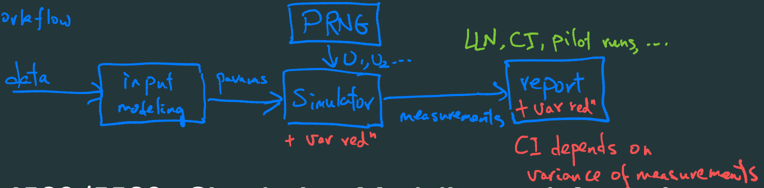
Sim Workflow



ORIE 4580/5580: Simulation Modeling and Analysis

ORIE 5581: Monte Carlo Simulation

Unit 8: Variance Reduction

Sid Banerjee

School of ORIE, Cornell University

clicker question: back to the circle

generate n rv pairs $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, where all $X_i, Y_i \sim U[-1, 1]$, and compute n observations $Z_i = \mathbb{1}_{\{X_i^2+Y_i^2 \leq 1\}}$
which of the following is an estimator for π?

20
(a) $\frac{1}{n}\sum_{i=1}^{n} Z_i$

20
(b) $\frac{1}{n}\sum_{i=1}^{n} Z_i^2$

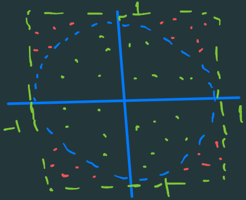
21
(c) $\frac{1}{2n}\sum_{i=1}^{n} Z_i$

9
(d) $\frac{2}{n}\sum_{i=1}^{n} Z_i$

30
(e) $\frac{4}{n}\sum_{i=1}^{n} Z_i$ ✓

$$\mathbb{E}[Z_i] = \frac{\text{Area of } \odot}{\text{Area of } [-1,1]^2} = \frac{\pi \cdot 1^2}{2^2} = \frac{\pi}{4}$$

$$\Rightarrow \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Z_i\right] = \frac{\pi}{4}$$



$Z_i = \mathbb{1}\{(x_i, y_i) \text{ in } \odot\}$

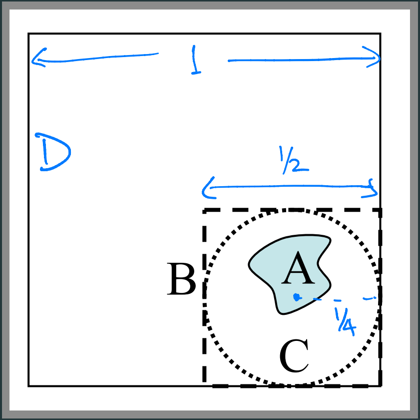
$\mathrm{Var}(Z_i)$

$$\pi \in \left[\frac{4}{n}\sum_{i=1}^{n} Z_i - \frac{2}{\sqrt{n}}\hat{\sigma},\ \frac{4}{n}\sum Z_i + \frac{2}{\sqrt{n}}\hat{\sigma}\right]$$ with 95% conf

variance reduction

- construct estimators with lower variance
- fewer replications to build CI of given width
- may need to exploit problem-specific information
  depending on application, this effort can be worthwhile

importance of smaller variance estimators



- aim: compute volume v_A of region A in the unit square
- **method 1**: generate points uniformly over the unit square (outermost box) and compute the fraction of points falling in region A

importance of smaller variance estimators

$X_i \sim U[0,1],\ Y_i \sim U[0,1]$

- let $X_1, \ldots, X_n$ be n points uniformly distributed in $[0,1]^2$

  $Y_1 \qquad Y_n$
- an estimator of v_A is

Let $Z_i = \mathbb{1}\{i\text{th sample falls in } A\}$

$= \mathbb{1}\{(X_i, Y_i) \in A\}$

$$\widetilde{V}_A = \frac{1}{n} \sum_{i=1}^{n} Z_i \leftarrow$$

$$\mathbb{E}[\widetilde{V}_A] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Z_i\right] = \mathbb{E}[Z_1] = \frac{v_A}{\underbrace{vol([0,1]^2)}_{=1}}$$

- $Var(\overset{Z_i}{\mathbb{I}_{[X_i \in A]}}) = v_A(1 - v_A) \leftarrow Var(Z_i) = v_A(1-v_A)$

  $\nwarrow Ber(v_A)$
- $Var(\widetilde{V}_A) = Var\left(\frac{1}{n}\sum_{i=1}^{n} Z_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} Z_i\right) = \frac{\sum_{i=1}^{n} Var(Z_i)}{n^2} = \frac{v_A(1-v_A)}{n}$

  $\uparrow$ iid

$\mathbb{E}[Z_i] = 0$

$$Var\left(\sum Z_i\right) = \mathbb{E}\left[\Big(\underbrace{\sum Z_i}_{\text{not } nZ_1}\Big)^2\right] = \mathbb{E}\left[\underbrace{\sum_{i=1}^{n} Z_i^2}_{} + 2\overbrace{\sum_{i \neq j} Z_i Z_j}^{\mathbb{E}[Z_i]\mathbb{E}[Z_j]}\right] = n\mathbb{E}[Z_1^2] + 0$$

importance of smaller variance estimators

- **method 2**: generate n points $Y_1, \ldots, Y_n$ uniformly in square $B = [0, \tfrac{1}{2}]^2$

$v_B = 1/4$

- an estimator of v_A is

$$\widehat{V}_A = v_B \cdot \frac{1}{n} \sum_{i=1}^{n} Z_i = \frac{1}{4} \cdot \frac{1}{n} \sum_{i=1}^{n} Z_i$$

$$Z_i = \mathbb{1}\{Y_i \in A\} = \mathrm{Ber}\left(\frac{v_A}{v_B}\right)$$

$$\Rightarrow \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} Z_i\right] = \mathbb{E}[Z_1] = \frac{v_A}{v_B}$$

- $Var(\mathbb{I}_{[Y_i \in A]}) = \frac{v_A}{v_B}\left(1 - \frac{v_A}{v_B}\right)$

- $Var(\widehat{V}_A) = Var\left(\frac{v_B}{n} \sum Z_i\right) = \frac{v_B^2}{n^2} \cdot n\,Var(Z_1) = \frac{v_B^{\cancel{2}}}{n} \cdot \frac{v_A}{\cancel{v_B}}\left(1 - \frac{v_A}{v_B}\right)$

- $Var(\widehat{V}_A) \leq Var(\widetilde{V}_A) \Rightarrow \widehat{V}_A$ gives more accurate estimates

$$= \frac{v_A(v_B - v_A)}{n}$$

dec as v_B decreases $\longrightarrow$

$\left(\text{here } \frac{v_A(1/4 - v_A)}{n}\right)$

importance of smaller variance estimators

- **method 3**: generate n points $Z_1, \dots, Z_n$ uniformly in circle C
- an estimator of v_A is

$$\bar{V}_A = \quad v_C \cdot \frac{1}{n} \sum W_i \,, \quad \text{where } W_i = \mathbb{1}_{\{Z_i \in A\}}$$

- $Var(\mathbb{I}_{[Z_i \in A]}) =$
- $Var(\bar{V}_A) = \quad \dfrac{v_A (v_C - v_A)}{n} = \dfrac{v_A \left(\frac{\pi}{16} - v_A\right)}{n}$
- $Var(\bar{V}_A) \leq Var(\widehat{V}_A) \leq Var(\widetilde{V}_A)$

complexity vs. variance reduction

- $\bar{V}_A$ requires points that are uniformly distributed over a circle
- to generate points uniformly in circle centered at $(0,0)$ with radius a:
  1. generate $U_1 \sim U[0,1]$, $U_2 \sim U[0,1]$ and $U_3 \sim U[0,1]$.
  2. set $R = a \max[U_1, U_2]$, $\theta = 2\pi U_3$.
  3. return $(R\cos\theta, R\sin\theta)$.
- requires cosine and sine computations
- faster to generate points uniformly in rectangle
  $\Rightarrow$ more points in same computation time

complexity vs. variance reduction

- although $\bar{V}_A$ has smaller variance than $\widehat{V}_A$, may be better to use $\widehat{V}_A$
- trade-off between reduction in variance and extra computation needed for variance reduction

variance reduction: techniques that help reduce estimator variance

– antithetic variates
– importance sampling
– control variates
– stratified sampling
– common random numbers

general but limited

general, widely applicable, but needs tuning

running example: Monte Carlo integration

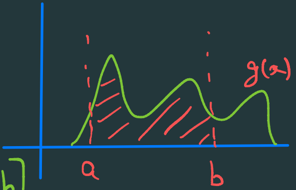
compute $\int_a^b g(x)dx$

- we know how to compute $\mathbb{E}[f(U)]$, where $U \sim U[0,1]$

$$\int_a^b g(x)\,dx = (b-a)\underbrace{\int_a^b \frac{g(x)}{(b-a)}\,dx}_{\mathbb{E}[g(Z)],\ Z \sim \mathrm{Unif}[a,b]}$$

$$\Rightarrow Z = a + (b-a)\,U[0,1]$$

$$\Rightarrow \int_a^b g(x)\,dx = \mathbb{E}\left[(b-a)\cdot g\big(a+(b-a)U\big)\right] = \mathbb{E}[f(U)]$$



running example: Monte Carlo integration

compute $\int_a^b g(x)dx$

$$\begin{aligned}(b-a)\int_a^b g(x)\frac{1}{b-a}dx &= (b-a)\ \mathbb{E}\left[g(Z)\right] \\ &= (b-a)\ \mathbb{E}\left[g(a+(b-a)U)\right] \\ &= \mathbb{E}\left[f(U)\right]\end{aligned}$$

$$f(x) = (b-a)g(a+(b-a)x)$$

antithetic variates

- observation: $\underline{X, X'}$ = identically distributed random variables $\left(\begin{smallmatrix}\text{not}\\ \text{indep}\end{smallmatrix}\right)$

$$\mathbb{E}\left[\frac{X+X'}{2}\right] = \frac{\mathbb{E}[X]+\mathbb{E}[X']}{2} = \mathbb{E}[X]$$

$$Var\left(\frac{X+X'}{2}\right) = \frac{1}{4}\underbrace{Var(X+X')}_{\underbrace{Var(X)+Var(X')}_{2Var(X)}+2Cov(X,X')}$$

$$= \frac{Var(X)+Cov(X,X')}{2}$$

$Cov(X,X') = \mathbb{E}[(X-\mu)(X'-\mu)]$

- If $X = X'$ $\Rightarrow Cov(X,X') = Var(X)$
- If $X \perp\!\!\!\perp X'$, $Cov(X,X')=0$
- Note - $Cov(X,X')$ can be ≤ 0

antithetic variates

- if X and X' are independent,

$$Var\left(\frac{X+X'}{2}\right) = \frac{1}{2}Var(X).$$

- if X and X' are negatively correlated,

$$Var\left(\frac{X+X'}{2}\right) < \frac{1}{2}Var(X).$$

- want simulation model to give two estimates of the performance measure X and X' such that $Cov(X, X') < 0$.

antithetic variates in Monte Carlo integration

- compute $\mathbb{E}[f(U)]$, where $U \sim U[0,1]$ Note - $1-U \sim U[0,1]$
- if $U_1, \ldots, U_{2n} \sim U[0,1]$, the regular MC estimator of $\mathbb{E}[f(U)]$ is

$$\alpha_{reg} = \frac{1}{2n}\sum_{i=1}^{2n} f(U_i), \quad \mathrm{Var}(\alpha_{reg}) = \frac{\mathrm{Var}(f(U))}{2n} \quad (\because U_i \text{ are } \mathrm{II})$$

- when U is large, $1-U$ is small
- $f(\cdot)$ monotone $\Rightarrow f(U)$ and $f(1-U)$ are negatively correlated
- the antithetic variates estimator of $\mathbb{E}[f(U)]$

$$\alpha_a = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{f(U_i)+f(1-U_i)}{2}\right), \quad \mathrm{Var}(\alpha_a) = \frac{1}{n}\mathrm{Var}\left(\frac{f(U)+f(1-U)}{2}\right)$$

$$= \frac{1}{2n}\left(\mathrm{Var}(f(U)) + \mathrm{Cov}(f(U), f(1-U))\right)$$

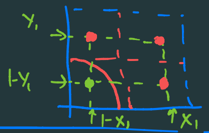
example: Monte Carlo integration

to see why this works, compute the variance:

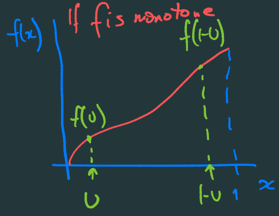
$$Var(\alpha^r) = \frac{1}{2n} Var(f(U))$$

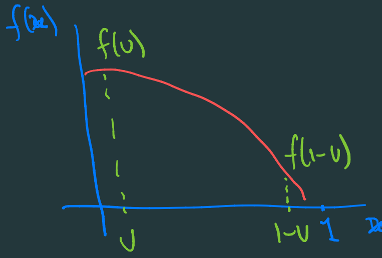
$$Var(\alpha^a) = \frac{1}{2n}\Big(Var(f(U)) + Cov\big(f(U), f(1-U)\big)\Big)$$

Eg - ⊙ in $[0,2]^2$



- since $Cov(f(U_i), f(1-U_i)) \leq 0$, we have $Var(\alpha^a) \leq Var(\alpha^r)$.
- a sufficient condition for antithetic variates to work is that the performance measure is monotone (increasing or decreasing)





For $f(U_1,U_2)$

$$-\frac{1}{4}\Big(f(U_1,U_2) + f(U_1,1-U_2) + f(1-U_1,U_2) + f(1-U_1,1-U_2)\Big)$$

clicker question: variance of antithetic estimators

to estimate $\mathbb{E}[f(U)]$, use n uniform rv $U_1, U_2, \ldots, U_n$ to get $2n$ observations:
$Y_1 = f(U_1), Y_2 = f(U_2), \ldots, Y_n = f(U_n)$
$Z_1 = f(1-U_1), Z_2 = f(1-U_2), \ldots, Z_n = f(1-U_n)$
and use these to get the antithetic estimator:

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i + Z_i}{2}$$

$\mathbb{E}[Y_i] = \mathbb{E}[Z_i] = \mu$
$= \mathbb{E}[\mu_n]$
(ie, $\approx \mu_n$)

this estimator has variance σ^2/n, where σ^2 is approximated by:

16 (a) sample variance of all $2n$ observations: $\frac{1}{2n-1}\sum_{i=1}^{n}(Y_i - \mu_n)^2 + (Z_i - \mu_n)^2$

47 (b) ✓ sample variance of combined observations: $\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{Y_i+Z_i}{2} - \mu_n\right)^2$

11 (c) average of $\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \mu_n)^2$ and $\frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \mu_n)^2$

25 (d) all of the above

(e) none of the above

clicker question: variance of antithetic estimators

$$\mathbb{E}\left[\frac{1}{2n-1}\sum_{i=1}^{n}(Y_i-\mu_n)^2+(Z_i-\mu_n)^2\right]$$

$$=\frac{1}{2n-1}\left(\sum_{i=1}^{n}\mathbb{E}\left[(Y_i-\mu_n)^2\right]+\sum_{i=1}^{n}\mathbb{E}\left[(Z_i-\mu_n)^2\right]\right)$$

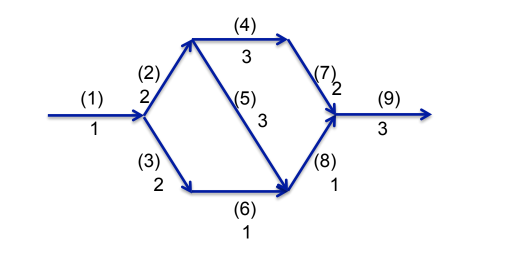
$$=\frac{2n}{2n-1}\mathbb{E}\left[(Y_i-\mu_n)^2\right]\approx \mathrm{Var}(Y_i)$$

$$\left(\text{not } \frac{\mathrm{Var}(Y)+\mathrm{Cov}(Y,Z)}{2}\right)$$

On the other hand

$$\mathbb{E}\left[\frac{1}{n-1}\sum\left(\frac{Y_i+Z_i}{2}-\mu_n\right)^2\right]=\mathrm{Var}\left(\frac{Y+Z}{2}\right)=\frac{\mathrm{Var}(Y)+\mathrm{Cov}(Y,Z)}{2}$$

example: critical paths



- arc length = duration of activity (assume *Exp*(label))
- activity durations are independent rvs $X_1, \ldots, X_9$
- project duration = length of longest source→sink path
- length of the critical path is

$$C(X_1, \ldots, X_9) = X_9 + \max\left[X_1 + X_2 + X_4 + X_7,\right.$$
$$\left.\max\left[X_1 + X_2 + X_5, X_1 + X_3 + X_6\right] + X_8\right].$$

example: critical paths

- $C(\cdot, \ldots, \cdot)$ is nondecreasing
- want identically distributed samples $\tilde{X}_1, \ldots, \tilde{X}_9$ and $\hat{X}_1, \ldots, \hat{X}_9$ such that when $C(\tilde{X}_1, \ldots, \tilde{X}_9)$ is large, $C(\hat{X}_1, \ldots, \hat{X}_9)$ is small
- suppose $X_i \sim F_i$ for each $i \in \{1, 2, \ldots, 9\}$

example: critical paths

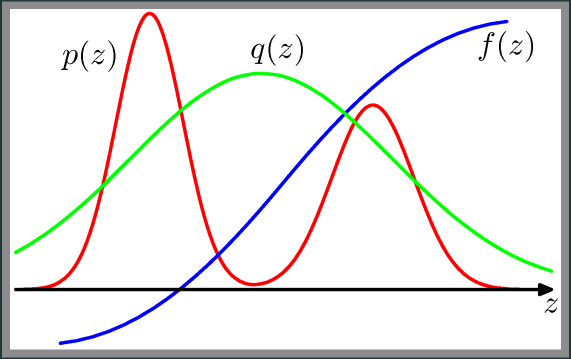
- $(U_1^n, \ldots, U_9^n) =$ 9-dim vector of iid $U[0,1]$ rvs
- $C(F_1^{-1}(U_1^n) \ldots, F_9^{-1}(U_9^n))$ and $C(F_1^{-1}(1-U_1^n), \ldots, F_9^{-1}(1-U_9^n))$ are negatively correlated
- the antithetic estimator

$$\widehat{T} = \frac{1}{N}\sum_{n=1}^{N} \frac{\left[C(F_1^{-1}(U_1^n), \ldots, F_9^{-1}(U_9^n)) + C(F_1^{-1}(1-U_1^n), \ldots, F_9^{-1}(1-U_9^n))\right]}{2}$$

should have smaller variance than the estimator $\frac{1}{2N}\sum_{n=1}^{2N} C(F_1^{-1}(U_1^n), \ldots, F_9^{-1}(U_9^n))$.

importance sampling

- given function $\phi(\cdot)$, want $\mathbb{E}[\phi(X)]$ where $X \sim P$
- can generate samples $Z \sim Q$ (but not from P)

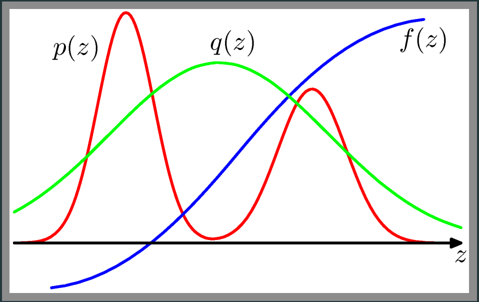


importance sampling

- given function $\phi(\cdot)$, want $\mathbb{E}[\phi(X)]$ where $X \sim P$
- can generate samples $Z \sim Q$

importance sampling

1. generate $Z_1, Z_2, \ldots, Z_L \sim Q$
2. compute $\mathbb{E}[\phi(X)] = \frac{1}{L} \sum_{i=1}^{L} w_i \phi(Z_i)$, where $w_i = p(Z_i)/q(Z_i)$



importance sampling: why does it work?

given function $\phi(\cdot)$, want $\mathbb{E}[\phi(X)]$ where $X \sim P$

importance sampling

1. generate $Z_1, Z_2, \ldots, Z_L \sim Q$
2. compute $\mathbb{E}[\phi(X)] = \frac{1}{L}\sum_{i=1}^{L} w_i \phi(Z_i)$, where $w_i = p(Z_i)/q(Z_i)$

importance sampling: variance

1. generate $Z_1, Z_2, \ldots, Z_L \sim Q$
2. compute $\mathbb{E}[\phi(X)] = \frac{1}{L}\sum_{i=1}^{L} w_i \phi(Z_i)$, where $w_i = p(Z_i)/q(Z_i)$

importance sampling: comments

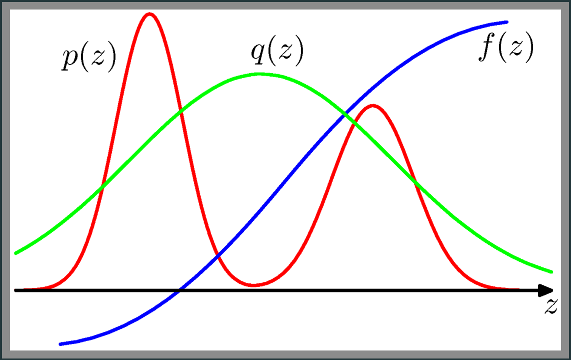
given function $\phi(\cdot)$, want $\mathbb{E}[\phi(X)]$ where $X \sim P$

importance sampling

1. generate $Z_1, Z_2, \ldots, Z_L \sim Q$
2. compute $\mathbb{E}[\phi(X)] = \frac{1}{L}\sum_{i=1}^{L} w_i \phi(Z_i)$, where $w_i = p(Z_i)/q(Z_i)$

(advanced) sampling-importance-resampling (SIR)

what if we want samples from P?



(advanced) sampling-importance-resampling (SIR)

sampling-importance-resampling (SIR)

1. generate $Z_1, Z_2, \ldots, Z_L \sim Q$
2. compute $w_i = p(Z_i)/q(Z_i)$ for each i 3. resample $X_1, X_2, \ldots, X_L$, where $X_i = Z_k$ with probability $\frac{w_k}{\sum_{j=1}^{L} w_j}$