Overall simulation pipeline

Bootstrap, χ^2, KS
MoM/MLE
'statistics'

Inversion
A/R
correlated Gaussian

Union bound ..
Confidence intervals

Input Modelling $\rightarrow$ Generate random vars $\rightarrow$ Simulation $\rightarrow$ Report Outcomes

ORIE 4580/5580: Simulation Modeling and Analysis

ORIE 5581: Monte Carlo Simulation

Unit 8: Variance Reduction

can 'improve' simulations to make the output less noisy

"variance reduction"

Sid Banerjee

School of ORIE, Cornell University

class poll: 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?

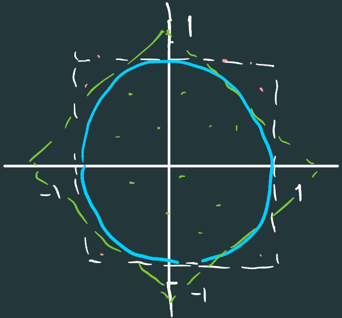
(a) $\frac{1}{n} \sum_{i=1}^{n} Z_i$

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

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

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

✓ (e) $\frac{4}{n} \sum_{i=1}^{n} Z_i$ (38%)



• $\mathbb{E}[Z_i] = \frac{\text{Area of } \bigcirc}{\text{Area of } \square} = \frac{\pi}{4}$

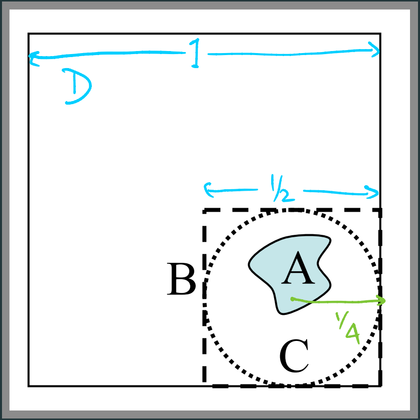
$Z_i = \mathbb{1}\{\text{pt } (x_i, y_i) \text{ is inside } \bigcirc\}$

$\Rightarrow \pi = 4\,\mathbb{E}[Z_i] \approx 4 \cdot \frac{1}{n} \sum_{i=1}^{n} Z_i$

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

- let $X_1, \dots, X_n$ be n points uniformly distributed in $\underbrace{[0,1]^2}_{D}$
- an estimator of v_A is $\left(Z_i = \mathbb{1}\{X_i \in A\} \right)$

$$\widetilde{V}_A = \left(\frac{\sum_{i=1}^n Z_i}{n} \right) \cdot \frac{1}{\mathrm{Vol}(D)} = \frac{\sum_{i=1}^n Z_i}{n}$$

$\mathrm{Vol}(D)$: 1 for $D = [0,1]^2$

- $Var(\underbrace{\mathbb{I}_{[X_i \in A]}}_{Z_i}) = v_A(1 - v_A)$
- $Var(\widetilde{V}_A) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n Z_i\right) = \frac{1}{n^2} \underbrace{\sum_{i=1}^n \mathrm{Var}(Z_i)}_{\text{since } Z_i \text{ are independent}} = \frac{n}{n^2} \mathrm{Var}(Z_i)$

$$= \frac{v_A(1-v_A)}{n}$$

If $W \sim \mathrm{Ber}(p)$

$\mathrm{Var}(W) = p(1-p)$

$\mathbb{E}[W] = \mathbb{E}[W^2] = p$

$\Rightarrow \mathrm{Var}(W) = \mathbb{E}[W^2] - \mathbb{E}[W]^2 =$

importance of smaller variance estimators

Assn - $A \subset B$
$\Rightarrow vol(A) = v_A \leq vol(B) = 1/4$

$vol(B) = 1/4$

- **method 2**: generate n points $Y_1, \ldots, Y_n$ uniformly in square $B = [0, \frac{1}{2}]^2$
- an estimator of v_A is $\left(W_i = \mathbb{1}\{Y_i \in A\} \Rightarrow \mathbb{E}[W_i] = v_A / \tfrac{1}{4} = 4v_A \right)$

$W_i \sim \mathrm{Ber}(4v_A)$

$$\widehat{V}_A = \left(\frac{1}{n} \sum_{i=1}^{n} W_i \right) \cdot \frac{1}{4}$$

- $Var(\mathbb{I}_{[Y_i \in A]}) = Var(W_i) = 4v_A(1 - 4v_A)$
- $Var(\widehat{V}_A) = Var\left(\frac{1}{4n} \sum W_i \right) = \frac{1}{16n^2} \cdot n \cdot 4v_A(1-4v_A) = \frac{v_A(1-4v_A)}{4n}$
- $Var(\widehat{V}_A) \leq Var(\widetilde{V}_A) \Rightarrow \widehat{V}_A$ gives more accurate estimates

$$\frac{v(1-4v)}{4n} \qquad \frac{v(1-v)}{n}$$

$$\frac{v}{n}\left(\frac{1}{4} - v\right) \leq \frac{v}{n}(1-v)$$

ie, smaller 'proposal set' for AR $\Rightarrow$ smaller variance

importance of smaller variance estimators

- **method 3**: generate n points $Z_1, \ldots, Z_n$ uniformly in circle C $\swarrow$ radius $1/4$ $\Rightarrow vol(C) = \pi/16$
- an estimator of v_A is $\left(L_i = \mathbb{1}\{Z_i \in C\} \sim \mathrm{Ber}(16 v_A/\pi) \right)$

$$\bar{V}_A = \frac{\pi}{16} \cdot \frac{1}{n} \sum_{i=1}^{n} L_i$$

- $Var(\mathbb{I}_{[Z_i \in A]}) = \frac{16 \cdot v}{\pi}\left(1 - \frac{16 v}{\pi}\right)$
- $Var(\bar{V}_A) = \frac{\pi^2}{256} \cdot \frac{1}{n} \cdot \frac{16 v}{\pi} \cdot \left(1 - \frac{16 v}{\pi}\right) = \frac{v}{n}\left(\frac{\pi}{16} - v\right)$
- $Var(\bar{V}_A) \leq Var(\widehat{V}_A) \leq Var(\widetilde{V}_A)$

$$\left[\left(\frac{\pi}{16} - v\right) \leq \left(\frac{1}{4} - v\right) \leq (1 - v)\right]\frac{v}{n}$$

Issue - sampling pts uniformly in a ⊙ is tricky

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

Upshot - Usual methods of getting better variance

Eg - better proposal sets for AR

may require more work

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 ← today
- importance sampling ← next class
- control variates
- stratified sampling ← 5582
- common random numbers ← 2nd half

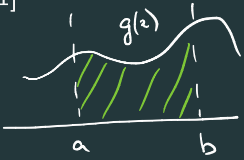
running example: Monte Carlo integration

Overall Q: compute $\int_a^b g(x)dx = I$

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

$$\int_a^b g(x)dx = (b-a)\underbrace{\int_a^b \frac{g(x)}{b-a}dx}$$

If $Z \sim \mathrm{Unif}[a,b] \Rightarrow I = (b-a)\mathbb{E}[g(Z)]$



$$= (b-a)\mathbb{E}\left[g\left(a+(b-a)U\right)\right]$$

$$= \mathbb{E}[f(U)], \text{ where } f(x) = (b-a)\,g(a+(b-a)x)$$

Note - properties of f matter for amnt of variance redn

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: X, X' = identically distributed random variables (not independent)

$$\mathbb{E}\left[\frac{X+X'}{2}\right] = \frac{1}{2}\left(\mathbb{E}[X] + \mathbb{E}[X']\right) = \mathbb{E}[X]$$

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

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

- Note - If $X \perp\!\!\!\perp X' \Rightarrow Var\left(\frac{X+X'}{2}\right) = \frac{Var(X)}{2}$

If $Cov(X,X') < 0 \Rightarrow Var\left(\frac{X+X'}{2}\right) < \frac{Var(X)}{2}$

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

$f(U)$ and $f(1-U)$ have the same distn

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

$$\alpha_{reg} = \frac{1}{2n}\sum_{i=1}^{2n} f(U_i) \quad \text{OR} \quad \frac{1}{2n}\sum_{i=1}^{2n} f(1-U)$$

$$\boxed{\mathrm{Var}(\alpha_{reg}) = \frac{1}{2n}\mathrm{Var}(f(U_i))}$$

- $\boxed{\text{when } U \text{ is large, } 1-U \text{ is small}} \Rightarrow \mathrm{Cov}(U, 1-U) < 0$
- $f(\cdot)$ monotone $\Rightarrow f(U)$ and $f(1-U)$ are negatively correlated

increasing or decreasing

- the antithetic variates estimator of $\mathbb{E}[f(U)]$

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

(Note - $2n$ function computations, same as α_{reg})

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

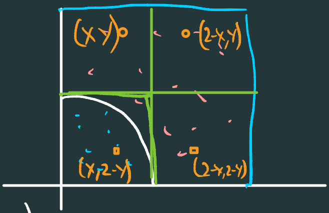
Eg - 'Estimating $\pi/4$'

$(X, Y) \sim U[0, 2]$

$\Rightarrow$ 4 'antithetic estimators'

$(X,Y), (2-X, 2-Y), (2-X, Y), (X, 2-Y)$

(All of these are Uniform in $[0,2]^2$)



$$\Rightarrow \quad Z_{\text{antithetic}} = \frac{1}{4}\left(Z_{xy} + Z_{x,2-y} + Z_{2-x,y} + Z_{2-x,2-y}\right)$$

example: Monte Carlo integration

to see why this works, compute the variance:

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

$$Var(\alpha^a) = \frac{1}{n}\left(\frac{Var(f(U_i)) + Cov(f(U_i), f(1-U_i))}{2}\right) = \frac{1}{2n}\left(Var(f(U_i)) + Cov(f(U_i), f(1-U_i))\right)$$

- 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)

Note -

$$\alpha^r = \frac{1}{2n}\sum_{i=1}^{2n} f(U_i)$$

$$\alpha^a = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{f(U_i)+f(1-U_i)}{2}\right)$$

class poll: variance of antithetic estimators

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

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

antithetic estimator

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

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

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

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

(d) all of the above

(e) none of the above

Variance Reduction

- Want to estimate $\int_a^b g(x)\,dx = \mathbb{E}\Big[\overbrace{(b-a)\,g(a+(b-a)U)}^{f(U)}\Big]$

'<u>Vanilla Monte Carlo</u>' - generate $U_1, U_2, \ldots, U_n$

compute $M_n = \frac{1}{n}\sum_{i=1}^{n} f(U_i)$

output (95% CI) $\equiv M_n \mp 2\underbrace{\sqrt{\left(\frac{1}{n}\cdot\frac{1}{n-1}\sum_{i=1}^{n}(f(U_i)-M_n)^2\right)}}_{\sqrt{\mathrm{Var}(M_n)}}$

- Variance reduction techniques reduce $\mathrm{Var}(M_n)$ with NO extra samples

- Antithetic Variates - $\hat{M}_n = \frac{1}{n}\sum_{i=1}^{n}\frac{f(U_i)+f(1-U_i)}{2}$

-ive correlated if f is monotone

Today - Eg - critical paths, importance sampling

class poll: variance of antithetic estimators

$$M_n = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{Y_i + Z_i}{2} \right), \quad Y_i = f(U_i),\ Z_i = f(1-U_i)$$

U_i are indep

$$\mathrm{Var}(M_n) = \frac{1}{n^2} \sum_{i=1}^{n} \underbrace{\mathrm{Var}\left(\frac{Y_i + Z_i}{2} \right)}$$

$$= \frac{1}{n} \text{ 'Sample Variance' of } \frac{Y_i + Z_i}{2}$$

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

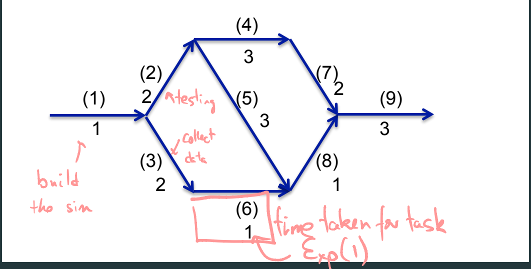
$$M_n = \frac{1}{2n} \left(\sum Y_i + \sum Z_i \right)$$

$$\mathrm{Var}(M_n) = \frac{1}{4n^2} \cdot n \left(\mathrm{Var}(Y_i) + \mathrm{Var}(Z_i) \right)$$

Wrong - $Y_i \not\perp\!\!\!\perp Z_i$

example: critical paths (example from scheduling)

'complex job' broken into multiple tasks



graph = precedence order between tasks

- arc length = duration of activity (assume *Exp*(label))
- activity durations are independent rvs $X_1, \ldots, X_9$ (not id, in particular $X_1 \sim Exp(1)$, $X_3 \sim Exp(2)$, ⋮
- project duration = length of longest source→sink path
- length of the critical path is

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

C is complicated, but its 'clearly' non-decreasing in each X_i

example: critical paths

- $C(\cdot, \dots, \cdot)$ is nondecreasing
- want identically distributed samples $\tilde{X}_1, \dots, \tilde{X}_9$ and $\hat{X}_1, \dots, \hat{X}_9$ such that when $C(\tilde{X}_1, \dots, \tilde{X}_9)$ is large, $C(\hat{X}_1, \dots, \hat{X}_9)$ is small
  $-\frac{1}{\lambda_i}\ln(1-U_i)$ $\quad -\frac{1}{\lambda_i}\ln(U_i)$
- suppose $X_i \sim F_i$ for each $i \in \{1, 2, \dots, 9\}$ $\longrightarrow \tilde{X}_i = F_i^{-1}(U_i),\ \hat{X}_i = F_i^{-1}(1-U_i)$

$\tilde{X}_{1r}, \tilde{X}_{2r}, \dots, \tilde{X}_{9r}$

$$\Rightarrow \quad C_n^{\text{antithetic}} = \frac{1}{n} \sum_{r=1}^{n} \left(C(\tilde{X}_r) + C(\hat{X}_r) \right) / 2$$

replication r

more generally, can take 2^9 different antithetic estimators

$$X_1 \dots X_9,\ \bar{X}_1 X_2 \dots X_9,\ X_1 \bar{X}_2 X_3 \dots X_9,\ \dots\dots,\ \bar{X}_1 \bar{X}_2 \dots \bar{X}_9$$

(this requires $2^9 \cdot n$ 'function calls')

example: critical paths

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

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

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

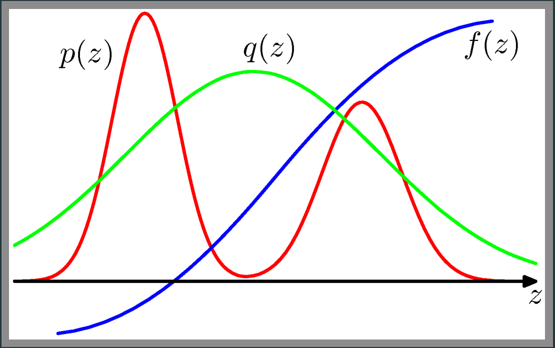
- $\mathrm{Var}(\hat{T}) = \frac{1}{N}\left(\frac{\mathrm{Var}(C(U)) + \overbrace{\mathrm{Cov}(C(U), C(1-U))}^{\text{-ive!}}}{2}\right)$

$\mathrm{Var}(T) = \frac{1}{2N}\mathrm{Var}((C(U))$

importance sampling

- given function $\phi(\cdot)$, want $\mathbb{E}[\phi(X)]$ where $X \sim P$ complicated
- can generate samples $Z \sim Q$ (but not from P) $\leftarrow X_1, X_2, \ldots, X_n$

simple - 'proposal'



· Idea - Acceptance-Rejection to get samples of P

$Y_1, Y_2 \ldots Y_M$, where $M \sim \text{Geom}(1/k)$, $k = \max\left(\frac{p(x)}{q(x)}\right)$

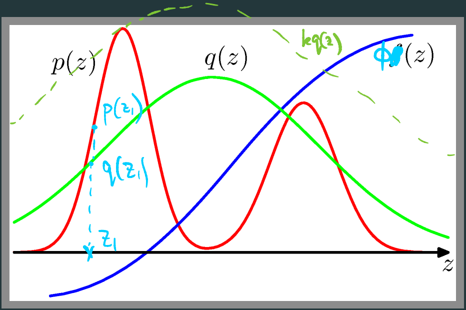
importance sampling (Generate (samples, 'importance wts'))

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

Recall - For AR, we accept Z_i w.p $\frac{kq(z_i)}{p(z_i)}$

importance sampling 'likelihood ratio'

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



Claim - $\mathbb{E}[W_i\phi(z_i)] = \mathbb{E}[\phi(X)]$

Pf - $\mathbb{E}[W_i\phi(z_i)]$

$$= \int_{-\infty}^{\infty} \phi(z) \cdot \underbrace{\frac{p(z)}{\cancel{q(z)}}}_{W(z)} \cdot \cancel{q(z)}\, dz$$

$$= \mathbb{E}[\phi(X)]$$

Note: MUST HAVE $q(z) > 0$ for all z s.t $p(z) > 0$ (Q is absolutely cont wrt P)

importance sampling: variance

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

$X \sim P$

$$\operatorname{Var}(M_L) = \frac{1}{L} \underline{\operatorname{Var}(W_i \phi(Z_i))}, \quad \mathbb{E}[M_L] = \mathbb{E}[\phi(X)]$$

$$\operatorname{Var}(W_i \phi(Z_i)) = \operatorname{Var}\left(\frac{p(Z_i)}{q(Z_i)} \phi(Z_i)\right)$$

$p(z)/q(z)$

$$\cdot \ \mathbb{E}\left[(W_i \phi(Z_i))^2\right] = \int_{-\infty}^{\infty} \frac{\phi(z)^2 \, p(z)^2}{q(z)^{\cancel{2}}} \cdot \cancel{q(z)} \, dz = \int_{-\infty}^{\infty} \phi(z)^2 w(z) p(z) \, dz$$

$$\Rightarrow \operatorname{Var}(M_L) = \frac{1}{L}\left(\mathbb{E}\left[\phi(x)^2 \cdot w(x)\right] - \left(\mathbb{E}[\phi(x)]\right)^2\right)$$

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 $\cancel{\mathbb{E}[\phi(X)]} = \frac{1}{L}\sum_{i=1}^{L} w_i \phi(Z_i)$, where $w_i = p(Z_i)/q(Z_i)$

- Suppose $q(x) = \dfrac{\phi(x)\, p(x)}{\mathbb{E}[\phi(X)]}$

Then $\mathrm{Var}[M_L] = 0$

True, but you need to know $\mathbb{E}[\phi(X)]$

(which was the question...)