UiO : Matematisk institutt

Det matematisk-naturvitenskapelige fakultet

STK-4051/9051 Computational Statistics Spring 2022
Variance reduction

Instructor: Odd Kolbjørnsen, oddkol@math.uio.no

Recap

- Exact methods
  - Inversion/transformation methods
  - Rejection sampling
- Approximate methods
  - Sampling importance resampling
  - Sequential Monte Carlo
  - Markov chain Monte Carlo (Chapter 7 and 8)
- Variance reduction methods
  - Importance sampling
  - Antithetic sampling
  - Control variates
  - Rao-blackwellization
  - Common random numbers

Today

Exercise

Monte Carlo methods

- Aim (following notation from book):

$$\mu = E^{f(\mathbf{X})}[h(\mathbf{X})] = \begin{cases} \int_{\mathbf{x}} h(\mathbf{x})f(\mathbf{x}))d\mathbf{x} & \mathbf{x} \text{ continuous} \\ \sum_{\mathbf{x}} h(\mathbf{x})f(\mathbf{x}) & \mathbf{x} \text{ discrete} \end{cases}$$

- Monte Carlo:
  1. Simulate $\mathbf{X}_i \sim f(\mathbf{x}), i = 1, ..., n$
  2. Approximate μ by

$$\hat{\mu}_{MC} = \frac{1}{n}\sum_{i=1}^{n} h(\mathbf{x}_i)$$

- Properties:
  - Unbiased $E[\hat{\mu}_{MC}] = \mu$
  - If $X_1, ..., X_n$ are independent
    - Variance: $\text{var}[\hat{\mu}_{MC}] = \frac{1}{n}\text{var}[h(\mathbf{X})]$
    - Consistent: $\hat{\mu}_{MC} \to \mu$ as $n \to \infty$ if $\text{var}[h(\mathbf{X})] < \infty$
  - Estimate of variance:

$$\widehat{\text{var}}[\hat{\mu}_{MC}] = \frac{1}{n-1}\sum_{i=1}^{n}(h(\mathbf{x}_i) - \hat{\mu}_{MC})^2$$

- Can we do better than this?

Last time

- Sequential Monte Carlo
- Importance sampling (normalized or not)
- Control variates
  - We know something about the distribution
  - Formalizes the correlation argument from importance sampling

$$\hat{\mu}_{CV} = \hat{\mu}_{MC} + \lambda(\hat{\theta}_{MC} - \theta)$$

$$\lambda = -\frac{\text{cov}[\hat{\mu}_{MC}, \hat{\theta}_{MC}]}{\text{var}[\hat{\theta}_{MC}]}$$

$$\text{var}[\hat{\mu}_{CV}] = \text{var}[\hat{\mu}_{MC}] + \lambda^2\text{var}[\hat{\theta}_{MC}] + 2\lambda\text{cov}[\hat{\mu}_{MC}, \hat{\theta}_{MC}]$$

- Rao-Blacwellization
  - We know something about a conditional distribution
  - We can make a part of the computation analytically
  - Particular useful with hyper parameters

$$\text{var}[h(\mathbf{X}_i)] = E[\text{var}[h(\mathbf{X}_i)|\mathbf{X}_2]] + \text{var}[E[h(\mathbf{X})|\mathbf{X}_2]] \geq \text{var}[E[h(\mathbf{X})|\mathbf{X}_2]$$

Antithetic sampling

Things that are **antithetic** to one another contradict or oppose each other.

- Assume available $\hat{\mu}_1$ and $\hat{\mu}_2$, identically distributed with $\text{var}[\hat{\mu}_j] = \sigma^2/n$
- Assume $\text{cov}[\hat{\mu}_1, \hat{\mu}_2] < 0$.
- Define $\hat{\mu}_{AS} = \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2)$

$$\begin{aligned}\text{var}[\hat{\mu}_{AS}] =& \tfrac{1}{4}(\text{var}[\hat{\mu}_1] + \text{var}[\hat{\mu}_2]) + \tfrac{1}{2}\text{cov}[\hat{\mu}_1, \hat{\mu}_2] \\ =& \tfrac{(1+\rho)\sigma^2}{2n}\end{aligned}$$

  where $\rho = \text{cor}[\hat{\mu}_1, \hat{\mu}_2]$.
- Gain by including $\hat{\mu}_2$ a factor of $\frac{1+\rho}{2}$!

Possible to construct such $\hat{\mu}_1, \hat{\mu}_2$?

Antithetic sampling

- Main idea: Most simulation procedures for generating $\mathbf{X} \sim f(\mathbf{x})$ is based on some transformation $X = h(\mathbf{U})$ where $\mathbf{U} = (U_1, ..., U_m)$ are iid uniform variables
- If U_j is uniform[0,1], then also $1 - U_j$ is uniform[0,1]
- $h(\mathbf{U})$ and $h(\mathbf{1} - \mathbf{U})$ will typically have negative correlation.
- Choose $\mathbf{X}_i = h(\mathbf{U}_i)$, $\mathbf{Y}_i = h(\mathbf{1} - \mathbf{U}_i)$

$$
\begin{aligned}
\hat{\mu}_1 =& n^{-1} \sum_{i=1}^{n} h(\mathbf{U}_i) \\
\hat{\mu}_2 =& n^{-1} \sum_{i=1}^{n} h(\mathbf{1} - \mathbf{U}_i)
\end{aligned}
$$

- Can be generalized to other settings as well.
- The following slides:
  - Proof of $\text{cor}[h(\mathbf{U}_i), h(\mathbf{1} - \mathbf{U}_i)] \leq 0$ for h monotone function in each U_j.

Antithetic sampling-theoretical derivations

- Assume $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ iid sample
- Assume $\hat{\mu}_j = n^{-1} \sum_{i=1}^n h_j(\mathbf{x}_i)$ with $E[h_j(\mathbf{x}_i)] = \mu$.
- Assume $h_j(\mathbf{X}_i)$ is increasing in each argument
- Result: $\text{cor}[h_1(\mathbf{X}_i), h_2(\mathbf{X}_i)] \geq 0$.

If $h_1(\boldsymbol{x})$, $h_2(\boldsymbol{x})$ is non decreasing in each argument $\boldsymbol{x} = (x_1, \dots, x_m)$
$h_j(\boldsymbol{x}) > h_j(\boldsymbol{x} - \boldsymbol{h})$, for all $\boldsymbol{h}$ such that $h_i > 0, i = 1, \dots, m$
then $\text{cor}(h_1(\boldsymbol{X}_i), h_2(\boldsymbol{X}_i)) \geq 0$

Proof by induction on dimension:

1) Prove that it is true in dimension 1
2) Prove that if it is true for dimension $m - 1$ then it is true for dimension m

Note slightly confusing the way we use the index on $\boldsymbol{X}_i$

Could have had $\mathrm{E}\left(h_j(\boldsymbol{X})\right) = \mu_j$ but this is not the case in question In antithetic sampling

Antithetic sampling-theoretical derivations

First: dimension 1

$$[h_1(X) - h_1(Y)][h_2(X) - h_2(Y)] \geq 0 \quad \text{Same sign}$$

$$\Downarrow$$

$$E[[h_1(X) - h_1(Y)][h_2(X) - h_2(Y)]] \geq 0 \quad \text{For any X and Y}$$

$$\Downarrow$$

$$E[h_1(X) - \mu - (h_1(Y) - \mu)][h_2(X) - \mu - (h_2(Y) - \mu)] \geq 0$$

$$\Downarrow \quad \text{Assuming } X, Y \text{ ind}$$

$$\mathrm{cov}[h_1(X), h_2(X)] + \mathrm{cov}[h_1(Y), h_2(Y)] \geq 0$$

$$\Downarrow \quad \text{Assuming } X, Y \text{ iid}$$

$$\mathrm{cov}[h_1(X), h_2(X)] \geq 0$$

Select joint distribution of X and Y to suit us

X and Y is selected to have the same distribution as X_i and X and Y are selected to be independent

Antithetic sampling-theoretical derivations

- Practical application in dimension 1
  - If h_1 increasing, h_2 decreasing:

$$\mathrm{cor}[h_1(X), h_2(X)] = -\mathrm{cor}[h_1(X), -h_2(X)] \leq 0$$

  - If X uniform: Then choose $h_1(X) = h(X)$, $h_2(X) = h(1 - X)$

$$\begin{aligned}\mathrm{Var}[\tfrac{1}{2}(h_1(X) + h_2(X))] =& \tfrac{1}{4}\mathrm{Var}[h_1(X)] + \tfrac{1}{4}\mathrm{Var}[h_2(X)] + \tfrac{1}{2}\mathrm{Cov}[h_1(X), h_2(X)] \\ \leq& \tfrac{1}{2}\mathrm{Var}[h_1(X)]\end{aligned}$$

  - If X Gaussian: Then choose $h_1(X) = h(X)$, $h_2(X) = h(-X)$

Works the same way, since: $\Phi(\cdot)$ is monotone and

$$\Phi(x) = u \Leftrightarrow \Phi(-x) = 1 - u$$

Warm up computations

$h_j(\mathbf{X}_i)$ is increasing in each argument

$E[h_j(\mathbf{X})|X_m] = \tilde{h}_j(X_m)$ is an increasing function in X_m.

For any $(x_1, x_2, \dots, x_{m-1})$, we have that

$$h_j(x_1, \dots, x_{m-1}, x_m) \geq h_j\,(x_1, \dots, x_{m-1}, x_m - h), \text{ for } h > 0$$

$$E\{h_j(X_1, \dots, X_{m-1}, x_m)\} \geq E\{h_j\,(X_1, \dots, X_{m-1}, x_m - h)\} \text{ for } h > 0$$

The relation is valid for any distribution,
we select: $f(\boldsymbol{x}|x_m)$ which gives the result

- $(E[\tilde{h}_j(X_m)] = E[E[h_j(\mathbf{X})|X_m]] = E[h_j(\mathbf{X})] = \mu)$

law of:
Total Expectation

- Assume $\mathrm{cor}[h_1(\mathbf{X}), h_2(\mathbf{X})] \geq 0$ for $\mathbf{X}$ $(m-1)$ dimensional. Then

$$\mathrm{cov}[h_1(\mathbf{X}), h_2(\mathbf{X})|X_m] \geq 0$$

Use result for $m-1$ for $f(\boldsymbol{x}|x_m)$

Taking expectations gives

$$\begin{aligned} 0 \leq & E[\mathrm{cov}[h_1(\mathbf{X}), h_2(\mathbf{X})|X_m] \\ = & E[E[h_1(\mathbf{X})h_2(\mathbf{X})|X_m]] - \underbrace{E[E[h_1(\mathbf{X})|X_m] \cdot E[h_2(\mathbf{X})|X_m]]} \end{aligned}$$

$$\begin{aligned} & E[E[h_1(\mathbf{X})|X_m] \cdot E[h_2(\mathbf{X})|X_m]] \\ & = E[\tilde{h}_1(X_m)\tilde{h}_2(X_m)] \\ & = \mathrm{cov}[\tilde{h}_1(X_m)\tilde{h}_2(X_m)] + E[\tilde{h}_1(X_m)]E[\tilde{h}_2(X_m)] \\ & \geq E[\tilde{h}_1(X_m)]E[\tilde{h}_2(X_m)] = \mu^2 \end{aligned}$$

Use result for 1 dimension

which gives

$$0 \leq E[h_1(\mathbf{X})h_2(\mathbf{X})] - \mu^2 = \mathrm{cov}[h_1(\mathbf{X}), h_2(\mathbf{X})]$$

Example

- $\mu = E[x/(2^x - 1)]$ for $x \sim N(0, 1)$
- Note: $x \sim N(0, 1)$ imply $-x \sim N(0, 1)$
- `Example_6_10.R`

```
set.seed(231171)
N = 2e5
x = rnorm(N)
N2 = N/2
x1 = x[1:N2]
x12=matrix(t(cbind(x1,-x1)), 1,N)

plot(1:N, cumsum(h(x))/(1:N),type='l',col='red'
lines(1:N, cumsum(h(x12))/(1:N)  , col='blue')
```