Stochastic Simulation
Variance reduction methods

Bo Friis Nielsen

Applied Mathematics and Computer Science

Technical University of Denmark

2800 Kgs. Lyngby – Denmark

Email: bfni@dtu.dk

Variance reduction methods

- To obtain better estimates with the same ressources
- Exploit analytical knowledge and/or correlation
- Methods:
  - ◇ Antithetic variables
  - ◇ Control variates
  - ◇ Stratified sampling
  - ◇ Importance sampling

Case: Monte Carlo evaluation of integral

Consider the integral

$$\int_0^1 e^x dx$$

We can interpret this interval as

$$\mathbb{E}\left(e^U\right) = \int_0^1 e^x dx = \theta \qquad U \in \mathsf{U}(0,1)$$

To estimate the integral: sample of the random variable e^U and take the average.

$$X_i = e^{U_i} \qquad \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

This is the crude Monte Carlo estimator , "crude" because we use no refinements whatsoever.

Analytical considerations

It is straightforward to calculate the integral in this case

$$\int_0^1 e^x dx = e - 1 \approx 1.72$$

The estimator X

$$\mathbb{E}(X) = e - 1 \qquad \mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

$$\mathbb{E}(X^2) = \int_0^1 (e^x)^2 \, dx = \frac{1}{2}\left(e^2 - 1\right)$$

Based on one observation

$$\mathbb{V}(X) = \frac{1}{2}\left(e^2 - 1\right) - (e - 1)^2 = 0.2420$$

Antithetic variables

General idea: to exploit correlation

- If the estimator is positively correlated with U_i (monotone function): Use $1-U$ also

$$Y_i = \frac{e^{U_i} + e^{1-U_i}}{2} = \frac{e^{U_i} + \frac{e}{e^{U_i}}}{2} \qquad \bar{Y} = \frac{\sum_{i=0}^{n} Y_i}{n}$$

- The computational effort of calculating $\bar{Y}$ should be similar to the effort needed to compute $\bar{X}$.
  - ◇ By the latter expression of Y_i we can generate the same number of Y's as X's

Antithetic variables - analytical

We can analyse the example analytically due to its simplicity

$$\mathbb{E}(\bar{Y}) = \mathbb{E}(\bar{X}) = \theta$$

To calculate $\mathbb{V}(\bar{Y})$ we start with $\mathbb{V}(Y_i)$.

$$\mathbb{V}(Y_i) = \frac{1}{4}\mathbb{V}\left(e^{U_i}\right) + \frac{1}{4}\mathbb{V}\left(e^{1-U_i}\right) + 2\cdot\frac{1}{4}\mathbb{C}ov\left(e^{U_i}, e^{1-U_i}\right)$$

$$= \frac{1}{2}\mathbb{V}\left(e^{U_i}\right) + \frac{1}{2}\mathbb{C}ov\left(e^{U_i}(e^{1-U_i}\right)$$

$$\mathbb{C}ov\left(e^{U_i}, e^{1-U_i}\right) = \mathbb{E}\left(e^{U_i}e^{1-U_i}\right) - \mathbb{E}\left(e^{U_i}\right)\mathbb{E}\left(e^{1-U_i}\right)$$

$$= e - (e-1)^2 = 3e - e^2 - 1 = -0.2342$$

$$\mathbb{V}(Y_i) = \frac{1}{2}(0.2420 - 0.2342) = 0.0039$$

Comparison: Crude method vs. antithetic

Crude method:

$$\mathbb{V}(X_i) = \frac{1}{2}\left(e^2 - 1\right) - (e-1)^2 = 0.2420$$

Antithetic method:

$$\mathbb{V}(Y_i) = \frac{1}{2}(0.2420 - 0.2342) = 0.0039$$

I.e, a reduction by 98 % , almost for free.

The variance on $\bar{X}$ - and $\bar{Y}$ - will scale with $1/n$, the number of samples.

Going from crude to antithetic method, reduces the variance as much as increasing number of samples with a factor 50.

Antethetic variables in more complex models

If

$$X = h(U_1, \ldots, U_n)$$

where h is monotone in each of its coordinates, then we can use antithetic variables

$$Y = h(1 - U_1, \ldots, 1 - U_n)$$

to reduce the variance, because

$$\mathbb{C}ov(X, Y) \leq 0$$

and therefore $\mathbb{V}(\frac{1}{2}(X + Y)) \leq \frac{1}{2}\mathbb{V}(X)$.

Antithetic variables in the queue simulation

Can you device the queueing model of yesterday, so that the number of rejections is a monotone function of the underlying U_i's?

Yes: Make sure that we always use either U_i or $1 - U_i$, so that a large U_i implies customers arriving quickly and remaining long.

Control variates

Use of covariates

$$Z = X + c(Y - \mu_y) \qquad \mathbb{E}(Y) = \mu_y(\text{ known})$$

$$\mathbb{V}(Z) = \mathbb{V}(X) + c^2\mathbb{V}(Y) + 2c\mathbb{C}ov(Y, X)$$

We can minimize $\mathbb{V}(Z)$ by choosing

$$c = \frac{-\mathbb{C}ov(X, Y)}{\mathbb{V}(Y)}$$

to get

$$\mathbb{V}(Z) = \mathbb{V}(X) - \frac{\mathbb{C}ov(X, Y)^2}{\mathbb{V}(Y)}$$

Example

Use U as control variate

$$Z_i = X_i + c\left(U_i - \frac{1}{2}\right) \qquad X_i = e^{U_i}$$

The optimal value can be found by

$$\mathbb{C}ov(X,Y) = \mathbb{C}ov\left(U, e^U\right) = \mathbb{E}\left(Ue^U\right) - \mathbb{E}(U)\mathbb{E}\left(e^U\right) \approx 0.14086$$

In practice we would not know this covariance, but estimate it empirically.

$$\mathbb{V}(Z_{c=\frac{-0.14086}{1/12}}) = \mathbb{V}\left(e^U\right) - \frac{\mathbb{C}ov\left(e^U, U\right)^2}{\mathbb{V}\left(U\right)} = 0.0039$$

Stratified sampling

This is a general survey technique: We sample in predetermined areas, using knowledge of structure of the sampling space

$$W_i = \frac{e^{\frac{U_{i,1}}{10}} + e^{\frac{1}{10}+\frac{U_{i,2}}{10}} + \cdots + e^{\frac{9}{10}+\frac{U_{i,10}}{10}}}{10}$$

What is an appropriate number of strata?

(In this case there is a simple answer; for complex problems not so)

Importance sampling

Suppose we want to evaluate

$$\theta = \mathbb{E}(h(X)) = \int h(x)f(x)dx$$

For $g(x) > 0$ whenever $f(x) > 0$ this is equivalent to

$$\theta = \int \frac{h(x)f(x)}{g(x)} g(x)dx = \mathbb{E}\left(\frac{h(Y)f(Y)}{g(Y)}\right)$$

where Y is distributed according to $g(y)$

This is an efficient estimator of θ, if we have chosen g such that the variance of $\left(\frac{h(Y)f(Y)}{g(Y)}\right)$ is small.

Such a g will lead to more Y's where $h(y)$ is large.

More important regions will be sampled more often.

Re-using the random numbers

We want to compare two different queueing systems.

We can estimate the rejection rate of system $i = 1, 2$ by

$$\theta_i = \mathbb{E}(g_i(U_1, \ldots, U_n))$$

and then rate the two systems according to

$$\hat{\theta}_2 - \hat{\theta}_1$$

But typically $g_1(\cdots)$ and $g_2(\cdots)$ are positively correlated: Long service times imply many rejections.

Then are more efficient estimator is based on

$$\theta_2 - \theta_1 = \mathbb{E}\left(g_2(U_1, \ldots, U_n) - g_1(U_1, \ldots, U_n)\right)$$

This amounts to letting the two systems run with the *same* input sequence of random numbers, i.e. same arrival and service time for each customer.

With some program flows, this is easily obtained by re-setting the seed of the RNG.

When this is not sufficient, you must store the sequence of arrival and service times, so they can be re-used.

Exercise 5: Variance reduction methods

- Estimate the integral $\int_0^1 e^x dx$ by simulation (the crude Monte Carlo estimator). Use eg. an estimator based on 100 samples and present the result as the point estimator and a confidence interval.
- Estimate the integral $\int_0^1 e^x dx$ using antithetic variables, with comparable computer ressources.
- Estimate the integral $\int_0^1 e^x dx$ using a control variable, with comparable computer ressources.
- Estimate the integral $\int_0^1 e^x dx$ using stratified sampling, with comparable computer ressources.
- Use control variates to reduce the variance of the estimator in exercise 4 (Poisson arrivals).