

Lecture 9 - Monte Carlo Integration

Björn Andersson (w/ Xijia Liu)

Department of Statistics, Uppsala University

February 27, 2015

Table of Contents

- 1 Numerical integration
 - Basic methods for numerical integration
- 2 Monte Carlo integration
- 3 Variance reduction in Monte Carlo integration

Quadrature rules

Midpoint rule:

$$\int_a^b f(x) dx \approx (b - a) f\left(\frac{a + b}{2}\right)$$

Trapezoidal rule:

$$\int_a^b f(x) dx \approx (b - a) \frac{f(a) + f(b)}{2}$$

Composite trapezoidal rule:

$$\int_a^b f(x) dx \approx \frac{b - a}{2n} \left[f(a) + 2 \sum_{i=1}^{n-1} f\left(a + i \frac{b - a}{n}\right) + f(b) \right]$$

Quadrature rules

Consider estimating $\int_1^3 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$.

```
R> pnorm(3) - pnorm(1)
```

```
[1] 0.1573054
```

```
R> (3 - 1) * dnorm((4/2))
```

```
[1] 0.1079819
```

```
R> (3 - 1) * (dnorm(1) + dnorm(3)) / 2
```

```
[1] 0.2464026
```

```
R> ((3 - 1) / (2*20)) * (dnorm(1) + 2 * (sum(dnorm(1 +  
+ (1:19) * (3 - 1) / 20))) + dnorm(3))
```

```
[1] 0.157496
```

Simpson's rule

We want to calculate $\int_a^b f(x)dx$. Let $t_i = a + ih$, $h = \frac{b-a}{2n}$ and divide $[a, b]$ into $2n$ intervals. Simpson's rule:

$$\int_a^b f(x)dx \approx \frac{h}{3} \left(f(a) + 4 \sum_{i=1}^n f(t_{2i-1}) + 2 \sum_{i=1}^{n-1} f(t_{2i}) + f(b) \right).$$

Exercise: implement Simpson's rule in R.

Table of Contents

1 Numerical integration

2 Monte Carlo integration

- Simple Monte Carlo estimation
- General Monte Carlo estimation
- Hit or Miss estimation
- Variance and efficiency

3 Variance reduction in Monte Carlo integration

Monte Carlo methods

The textbook contains only a brief introduction to the Monte Carlo integration methods. This lecture will introduce the topic in a somewhat more formal way.

For a thorough treatment of the topic:

- *Simulation and the Monte Carlo Method*, Rubinstein and Kroese (2008)
- *Monte Carlo Statistical Methods*, Robert and Casella (2004)

Simple Monte Carlo

The object is to estimate the integral $\theta = \int_0^1 g(x)dx$.

- We can re-write θ as

$$\theta = \int_0^1 g(x)dx = E(g(X)),$$

where $X \sim U(0, 1)$.

- If X_1, \dots, X_m is a random sample from $U(0, 1)$, then we retrieve the simple Monte Carlo (SMC) estimator

$$\hat{\theta}_{SMC} = \overline{g_m(X)} = \frac{1}{m} \sum_{i=1}^m g(X_i).$$

- $\hat{\theta}_{SMC}$ converges in probability to $E(g(X)) = \theta$ by the strong law of large numbers.

Simple Monte Carlo example

We want to compute a Monte Carlo estimate of $\theta = \int_0^1 e^{-x} dx$ and compare it to the exact value.

```
R> x <- runif(10000)
```

```
R> mean(exp(-x))
```

```
[1] 0.6342469
```

```
R> 1 - exp(-1)
```

```
[1] 0.6321206
```

Hence, the estimate $\hat{\theta}$ is very close to the exact value.

General Monte Carlo estimation

Here, the object is to estimate $\theta = \int_a^b g(x)dx$.

- We re-write θ as

$$\theta = (b - a) \int_a^b g(x) \frac{1}{b - a} dx = (b - a) E[g(X)],$$

where $X \sim U(a, b)$.

- The crude Monte Carlo (CMC) estimate is then:

$$\hat{\theta}_{CMC} = \frac{b - a}{m} \sum_{i=1}^m g(X_i),$$

where X_1, \dots, X_m are i.i.d. from $U(a, b)$.

General Monte Carlo estimation: R

We want to estimate $\int_1^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$.

```
R> x <- runif(10000, 1, 3)
```

```
R> ((3 - 1) / 10000) * sum(dnorm(x))
```

```
[1] 0.1573236
```

```
R> pnorm(3) - pnorm(1)
```

```
[1] 0.1573054
```

General Monte Carlo estimation

- More generally, suppose we want to estimate $\theta = \int_D g(x)dx$, where $g(x)$ can be decomposed as $h(x)f(x)$, where $f(x)$ is a probability density on D .
- Then we have

$$\theta = \int_D h(x)f(x)dx = E_f[h(X)],$$

where X is a R.V. with density f on the set D .

- The general crude Monte Carlo estimate is then

$$\hat{\theta}_{GCMC} = \frac{1}{m} \sum_{i=1}^m h(X_i),$$

where X_1, \dots, X_m are i.i.d. R.V.s from density f on D .

General Monte Carlo estimation example

Consider estimating $\theta = \int_{-\infty}^{\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-(x-2)^2/2}$. Our estimator is then

$$\hat{\theta} = \frac{1}{m} \sum_{j=1}^m x_j^4,$$

where the x_j are generated from f .

```
R> x4 <- rnorm(100000, mean = 2)
```

```
R> sum(x4^4) / 100000
```

```
[1] 42.80055
```

For $X \sim N(\mu, \sigma^2)$, $E(X^4) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$.

So for $\mu = 2, \sigma^2 = 1$, $E(X^4) = 43$.

Hit or Miss estimation

- We consider estimating $\theta = \int_a^b g(x)dx$. We assume that $0 \leq g(x) \leq c$, and re-write θ as

$$\theta = \int_a^b \int_0^{g(x)} 1 dy dx = c(b-a) \int_a^b \int_0^c \frac{\mathbf{1}\{y \leq g(x)\}}{c(b-a)} dy dx.$$

- Hence, $\theta = c(b-a)E(\mathbf{1}\{Y \leq g(X)\})$, where $Y \sim U(0, c)$ and $X \sim U(a, b)$.

Hit or Miss estimation

Hence, by independently generating X_1, \dots, X_m from $U(a, b)$ and Y_1, \dots, Y_m from $U(0, c)$ we get the following hit or miss (HM) estimator of θ

$$\hat{\theta}_{HM} = \frac{c(b-a)}{m} \sum_{i=1}^m \mathbf{1}\{Y_i \leq g(X_i)\}.$$

Hit or Miss estimation

Consider again to estimate $\int_1^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$, this time with Hit or Miss. We have that $c = g(0) = 1/\sqrt{2\pi}$. Hence we generate $x_i \sim U(1, 3)$ and $y_i \sim U(0, 1/\sqrt{2\pi})$ and compare y_i with $g(x_i)$:

```
R> x <- runif(10000, 1, 3)
R> y <- runif(10000, 0, 1 / sqrt(2 * pi))
R> ((2 / sqrt(2 * pi)) / 10000) * sum(y < dnorm(x))

[1] 0.1588588

R> pnorm(3) - pnorm(1)

[1] 0.1573054
```


Variances of the Monte Carlo estimators

- For $\hat{\theta}_{CMC} = \frac{b-a}{m} \sum_{i=1}^m g(X_i)$, we have

$$\text{Var}(\hat{\theta}_{CMC}) = \text{Var} \left[\frac{b-a}{m} \sum_{i=1}^m g(X_i) \right] = \frac{(b-a)^2}{m} \text{Var}[g(X)],$$

where $X \sim U(a, b)$.

- For $\hat{\theta}_{GCMC} = \frac{1}{m} \sum_{i=1}^m h(X_i)$, we have

$$\text{Var}(\hat{\theta}_{GCMC}) = \text{Var} \left[\frac{1}{m} \sum_{i=1}^m h(X_i) \right] = \frac{1}{m} \text{Var}[h(X)],$$

where X is generated from f on set D .

Estimating the variance of the Monte Carlo estimators

- We may estimate the variance of $g(X)$ from the sample variance, i.e. by

$$\hat{\text{Var}}[g(X)] = \sum_{i=1}^m \frac{[g(x_i) - \overline{g(x)}]^2}{m-1}.$$

- The standard error estimate of $\hat{\theta}_{CMC}$ is then

$$\hat{\text{se}}(\hat{\theta}_{CMC}) = \frac{b-a}{\sqrt{m}} \left\{ \sum_{i=1}^m \frac{[g(x_i) - \overline{g(x)}]^2}{m-1} \right\}^{1/2}.$$

- An alternative is to repeat the estimation of θ m times to retrieve the sampling distribution of the estimator. Then the sample variance of the repetitions can be used to estimate the variance of the Monte Carlo estimator.

Estimating the variance of the Monte Carlo estimators

Consider $\int_1^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$.

```
R> xMC <- numeric(100)
R> for(i in 1:100){
+   x <- runif(1000, 1, 3)
+   xMC[i] <- ((3 - 1) / 1000) * sum(dnorm(x))
+ }
R> var(xMC)

[1] 1.492597e-05

R> 4 * var(dnorm(x)) / 1000

[1] 2.036098e-05
```

Efficiency

- For two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ of θ , $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if

$$\frac{\text{Var}(\hat{\theta}_1)}{\text{Var}(\hat{\theta}_2)} < 1.$$

- If the variances of the estimators are unknown, they can be estimated from the Monte Carlo sample.
- The variance of the Monte Carlo estimators can be reduced by increasing the number of replications, hence the speed of the estimators must also be considered.

Crude Monte Carlo vs. Hit or Miss Monte Carlo

For a given m , the variance of $\hat{\theta}_{HM}$ is always larger than $\hat{\theta}_{GCMC}$. The Hit or Miss method is worse because an extra R.V. Y is introduced, which leads to extra uncertainty.

```
R> xHM <- numeric(100)
R> for(i in 1:100){
+   x <- runif(1000, 1, 3)
+   y <- runif(1000, 0, 1 / sqrt(2 * pi))
+   xHM[i] <- ((2 / sqrt(2 * pi)) / 1000) * sum(y < dnorm(x))
+ }
R> var(xHM)

[1] 0.0001227415

R> var(xMC)

[1] 1.492597e-05
```

Table of Contents

- 1 Numerical integration
- 2 Monte Carlo integration
- 3 Variance reduction in Monte Carlo integration
 - Common and antithetic variables
 - Importance sampling

Introduction to variance reduction

- Increasing the number of replications m reduces the variance of the Monte Carlo estimator
- A large increase in m is required to get a small reduction to the standard error
- Generally, if the standard error is to be at most e then $m \geq \lfloor \sigma^2 / e^2 \rfloor$, where $\sigma^2 = \text{Var}[g(X)]$
- Reducing the standard error thus has a high computational cost
- Considering this, it is useful to attempt other variance reduction methods

Common and antithetic random variables

- Let X and Y be R.V.s with known cdfs F and G . We want to estimate $\theta = E(X - Y)$ by Monte Carlo methods.
- We draw X and Y using the inverse transform method:

$$X = F^{-1}(U_1), Y = G^{-1}(U_2),$$

where U_1 and U_2 are generated from $U(0, 1)$

- Note that:

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y).$$

Hence, the variance of $X - Y$ will be reduced if $\text{Cov}(X, Y)$ is positive.

Common and antithetic random variables

- We say that common R.V.s are used if $U_1 = U_2$ and that antithetic R.V.s are used if $U_2 = 1 - U_1$.
- Since F^{-1} and G^{-1} are non-decreasing functions, using common R.V.s implies

$$\text{Cov}[F^{-1}(V), G^{-1}(V)] \geq 0,$$

for $V \sim U(0, 1)$. Thus we achieve a variance reduction.

- Using the same reasoning, $\text{Var}(X + Y)$ is reduced when using antithetic R.V.s.

Variance reduction strategy

We want to estimate $\theta = E[H(X)]$ for a monotone function H , where X can be generated using the inverse transform method, i.e. $X = F_X^{-1}(V)$, where $V \sim U(0, 1)$.

- Using antithetic R.V.s, $m/2$ replicates

$$Y_j = H[F_X^{-1}(V_j)]$$

are generated along with $m/2$ replicates

$$Y'_j = H[F_X^{-1}(1 - V_j)],$$

where $V_j \sim U(0, 1)$, $j \in \{1, \dots, m/2\}$.

- The antithetic estimator is then

$$\hat{\theta}_{AT} = \frac{1}{m} \sum_{j=1}^{m/2} (Y_j + Y'_j).$$

Variance reduction strategy example

Consider a R.V. X . Want to estimate $E[h(X)]$ with antithetic variables.

```
R> ATest <- function(Finv, h, n){  
+   if(n %% 2 != 0) return("n must be a positive even  
+   integer.")  
+   m <- n / 2  
+   u <- runif(m)  
+   y1 <- h(Finv(1 - u))  
+   y2 <- h(Finv(u))  
+   sum(y1 + y2) / n  
+ }  
R> regest <- function(Finv, h, n){  
+   u <- runif(n)  
+   sum(h(Finv(u))) / n  
+ }
```

Variance reduction strategy example

Example with $h(x) = x^2$ (second moment) and $X \sim U(0,1)$:

```
R> uniAT <- numeric(100)
R> for(i in 1:100){
+   uniAT[i] <- ATest(function(x) x, function(x) x^2, 10000)
+ }
R> unireg <- numeric(100)
R> for(i in 1:100){
+   unireg[i] <- regest(function(x) x, function(x) x^2, 10000)
+ }
R> var(uniAT)

[1] 7.796065e-07

R> var(unireg)

[1] 7.996479e-06
```

Importance sampling: idea

- We want to estimate $\theta = \int h(x)f(x)dx$
- For $\hat{\theta}_{GCMC} = \frac{1}{m} \sum_{j=1}^m h(X_j)$, we see that its variance depends on the variance of $h(X)$.
- If we instead of simulating from f choose another distribution g to simulate from, we retrieve

$$\theta = \int h(x) \frac{f(x)}{g(x)} g(x) dx$$

- g is called the importance sampling function

Importance sampling

Let X be a R.V. with pdf g , such that $g(x) > 0$ on the set $A = \{x : f(x) > 0\}$. Let Y be the R.V. such that $Y = \frac{f(X)}{g(X)} h(X)$. Then

$$\int_{x \in A} h(x) f(x) dx = \int_{x \in A} h(x) \frac{f(x)}{g(x)} g(x) dx = E_g \left[\frac{f(X)}{g(X)} h(X) \right] = E_g(Y).$$

The importance sampling estimator is

$$\hat{\theta}_{IS} = \frac{1}{m} \sum_{i=1}^m \frac{f(X_i)}{g(X_i)} h(X_i),$$

where X_1, \dots, X_m are generated from g . The variance of the importance sampling estimator depends on the variance of Y . Hence, for the variance to be reduced by this method, the density g should be close to f .

Importance sampling

Note that while $\hat{\theta}_{IS}$ converges almost surely to θ , the variance is only finite when

$$\begin{aligned} E_g(Y^2) &= E_g \left[\left(\frac{f(X)}{g(X)} h(X) \right)^2 \right] = E_f \left[[h(X)]^2 \frac{f(X)}{g(X)} \right] \\ &= \int_{x \in A} [h(x)]^2 \frac{[f(x)]^2}{g(x)} dx < \infty. \end{aligned}$$

So the choice of g should be such that g does not have lighter tails than f . The ratio f/g should be bounded.

Importance sampling

An alternative to $\hat{\theta}_{IS} = \frac{1}{m} \sum_{i=1}^m \frac{f(X_i)}{g(X_i)} h(X_i)$ which avoids the finite variance requirement is to instead use

$$\hat{\theta}_{wIS} = \frac{\sum_{j=1}^m h(X_j) f(X_j) / g(X_j)}{\sum_{j=1}^m f(X_j) / g(X_j)}.$$

This estimator has a small bias but has lower variance under some circumstances, such as when h is nearly constant.

Importance sampling: example

Consider $\theta = \int_4^\infty f(x)dx$, where f is the standard normal density function. We could estimate θ with the empirical average

$$\hat{\theta}_{\text{MC1}} = \frac{1}{m} \sum_{j=1}^m \mathbf{1}(x_j > 4),$$

where the x_j are generated from $\sim N(0, 1)$. However, very few of the generated numbers will be larger than 4. Instead we could use importance sampling with importance sampling function $g(x) = e^{-(x-4)}$, $x \in [4, \infty)$, the density for a shifted $\sim \text{Exp}(1)$ random variable. (Note that $h(x) = 1$.) Our estimator is then

$$\hat{\theta}_{\text{IS}} = \frac{1}{m} \sum_{j=1}^m \frac{f(x_j)}{g(x_j)} = \frac{1}{m} \sum_{j=1}^m \frac{\frac{e^{-x_j^2/2}}{\sqrt{2\pi}}}{e^{-(x_j-4)}} = \frac{1}{m} \sum_{j=1}^m \frac{e^{-x_j^2/2+x_j-4}}{\sqrt{2\pi}}$$

where the x_j are generated from g . Exercise: compare $\hat{\theta}_{\text{IS}}$ and $\hat{\theta}_{\text{MC1}}$.