Lecture 25: Monte Carlo Integration

Ciaran Evans

Motivating example

Suppose we wish to calculate the integral

$$\int_{-\infty}^{\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

Question: How could we calculate or approximate this integral?

Estimation via simulation

We want to approximate

$$\theta = \int_{-\infty}^{\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \mathbb{E}[X^4] \qquad X \sim N(0,1)$$

- Sample $X_1, ..., X_n \overset{iid}{\sim} N(0,1)$
- Estimate:

$$\widehat{\theta} =$$

Estimation via simulation

```r
n <- 100
x <- rnorm(n)
mean(x^4)
```

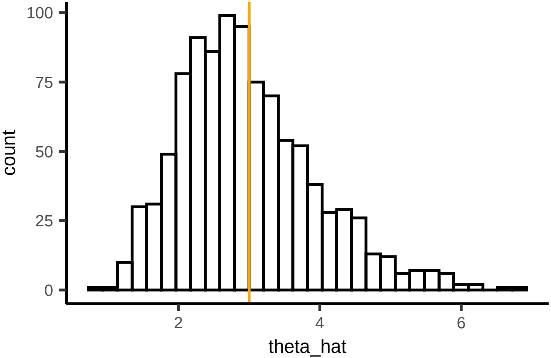
```
## [1] 1.410003
```

True value: $\theta = 3$

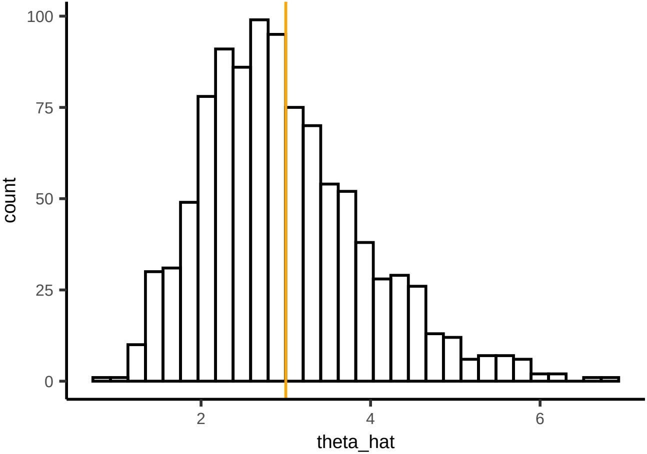
Question: Why are these numbers different?

Repeating many times

```r
n <- 100; nsim <- 1000
theta_hat <- rep(NA, nsim)
for(i in 1:nsim){
  x <- rnorm(n)
  theta_hat[i] <- mean(x^4)
}
```



Repeating many times



Question: Since $\widehat{\theta}$ is different for each sample, how do I measure the overall performance of $\widehat{\theta}$ as an estimate of θ?

MSE

$$MSE(\widehat{\theta}) = \text{Bias}(\widehat{\theta})^2 + Var(\widehat{\theta}) = (\mathbb{E}(\widehat{\theta}) - \theta)^2 + Var(\widehat{\theta})$$

- $\mathbb{E}(\widehat{\theta}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^4\right) =$

- $Var(\widehat{\theta}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i^4\right) =$

Monte Carlo Integration

Goal: Want to estimate

$$\theta = \int\limits_{\mathcal{X}} g(x) f(x) dx = \mathbb{E}[g(X)]$$

where f is a density and $X \sim f$ is a random variable with density f.

- Sample $X_1, ..., X_n \overset{iid}{\sim} f$
- **Monte Carlo estimate:** $\widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} g(X_i)$
- Numerical integration:
  - error comes from approximation of integrand (e.g. polynomial interpolation)
  - error decreases as number of nodes increases
- Monte Carlo integration:
  - error comes from variability of random samples
  - error decreases as sample size n increases

Activity, Part I

Work with your neighbor on the questions on the handout / course website:

Then we will discuss as a class

- Use Monte Carlo integration to approximate another integral
- Explore variability of estimate

Activity, Part I

$$\theta = \int_0^1 \frac{e^{-x}}{1+x^2} dx$$

Monte Carlo integration: write

$$\theta = \int_0^1 g(x) f(x) dx = \mathbb{E}[g(X)] \qquad X \sim f$$

Question: What could we use for g and f here?

Activity, Part I

```r
g <- function(x){
  exp(-x)/(1 + x^2)
}

n <- 10
x <- runif(n)
mean(g(x))
```

```
## [1] 0.5288093
```

Question: How variable is this estimate?

Activity, Part I

```r
n <- 10
nsim <- 1000
theta_hat <- rep(NA, nsim)

for(i in 1:nsim){
  x <- runif(n)
  theta_hat[i] <- mean(g(x))
}

var(theta_hat)
```

```
## [1] 0.006171067
```

```r
sd(theta_hat)
```

```
## [1] 0.07855614
```

Another perspective

$$\theta = \int_0^1 \frac{e^{-x}}{1+x^2} dx = \int_0^1 g(x) f(x) dx$$

So far:

- $g(x) = \dfrac{e^{-x}}{1+x^2}$, $f(x) = 1$

Question: Is this the only way we can choose g and f for this problem?

Activity, Part II

$$\theta = \int_0^1 \frac{e^{-x}}{1+x^2} dx = \int_0^1 g(x) f(x) dx$$

A couple different options:

- $f_1(x) = 1$, $g_1(x) = \dfrac{e^{-x}}{1+x^2}$

- $f_2(x) = \dfrac{4}{\pi(1+x^2)}$, $g_2(x) = \frac{\pi}{4} e^{-x}$

Activity: Work with your neighbor to compare the MSE of these two options. Which one better estimates θ?