

Lecture 26: Importance Sampling

Ciaran Evans

Last time

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

Monte Carlo integration: write

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

where f is a density function.

- ▶ Sample $X_1, \dots, X_n \stackrel{iid}{\sim} f$
- ▶ **Monte Carlo estimate:** $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n g(X_i)$

Last time

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

A couple different options:

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

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

Question: How would we sample from the distribution with density f_2 ?

Inverse transform method

$$\text{pdf: } f_2(x) = \frac{4}{\pi(1+x^2)} \qquad \text{cdf: } F_2(x) = \frac{4}{\pi} \text{atan}(x)$$

To sample $X_1, \dots, X_n \stackrel{iid}{\sim} f_2$:

► Sample $U_1, \dots, U_n \stackrel{iid}{\sim} \text{Uniform}(0, 1)$

► $X_i = F_2^{-1}(U_i)$

Variance

```
g2 <- function(x){ pi/4 * exp(-x)}
```

```
n <- 10; nsim <- 1000
```

```
theta_hat2 <- rep(NA, nsim)
```

```
for(i in 1:nsim){  
  x <- tan(runif(n) * pi/4)  
  theta_hat2[i] <- mean(g2(x))  
}
```

```
var(theta_hat2)
```

```
## [1] 0.001882555
```

```
sd(theta_hat2)
```

```
## [1] 0.04338842
```

Comparing variance

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

► $f_1(x) = 1, g_1(x) = \frac{e^{-x}}{1+x^2}$: $Var(\hat{\theta}_1) \approx 0.006$

► $f_2(x) = \frac{4}{\pi(1+x^2)}, g_2(x) = \frac{\pi}{4}e^{-x}$: $Var(\hat{\theta}_2) \approx 0.002$

Relative efficiency: $\frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)} \approx \frac{1}{3}$

Reduction in variance: $100 \cdot \frac{Var(\hat{\theta}_1) - Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)} \approx 68\%$ reduction

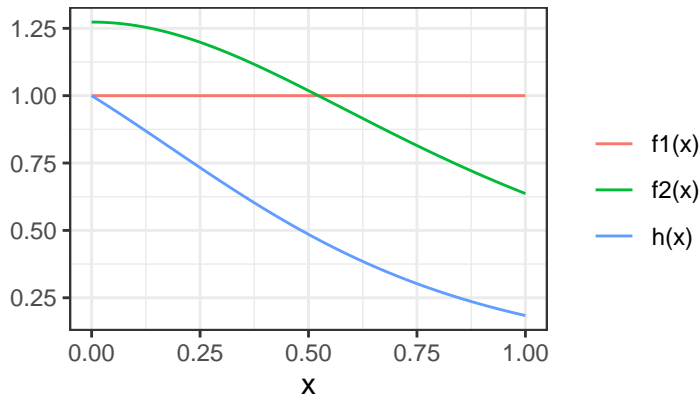
in variance

Comparing options

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

► $f_1(x) = 1, g_1(x) = \frac{e^{-x}}{1+x^2}: \text{Var}(\hat{\theta}_1) \approx 0.006$

► $f_2(x) = \frac{4}{\pi(1+x^2)}, g_2(x) = \frac{\pi}{4}e^{-x}: \text{Var}(\hat{\theta}_2) \approx 0.002$



Importance sampling

For any density function f supported on the domain of integration \mathcal{X} :

$$\theta = \int_{\mathcal{X}} h(x) dx = \int_{\mathcal{X}} \frac{h(x)}{f(x)} f(x) dx$$

► $X_1, \dots, X_n \stackrel{iid}{\sim} f$

► $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \frac{h(X_i)}{f(X_i)}$

$$\mathbb{E}[\hat{\theta}] = \theta \qquad \text{Var}(\hat{\theta}) = \frac{1}{n} \text{Var} \left(\frac{h(X)}{f(X)} \right)$$

Question: What choice of f would **minimize** this variance?

Importance sampling

$$\theta = \int_{\mathcal{X}} h(x) dx = \int_{\mathcal{X}} \frac{h(x)}{f(x)} f(x) dx$$

- ▶ $\text{Var}(\hat{\theta})$ is smaller if $f(x)$ is “similar to” $h(x)$
- ▶ If $f(x) = c \cdot h(x)$ for all x and some constant c , then
$$\text{Var}\left(\frac{h(X)}{f(X)}\right) = \text{Var}\left(\frac{1}{c}\right) = 0$$
- ▶ $\text{Var}(\hat{\theta})$ is minimized if $f(x) = \frac{|h(x)|}{\int |h(t)| dt}$
- ▶ But, if we can't integrate $h(x)$ anyway, then it isn't actually possible to use this $f(x)$

Your turn

Experiment with using different densities for Monte Carlo integration with another integral:

https://sta379-s25.github.io/practice_questions/pq_26.html

- ▶ Start in class
- ▶ You are welcome and encouraged to work with your neighbors
- ▶ Solutions posted on course website