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 $X_1, ..., X_n \overset{iid}{\sim} f$
- **Monte Carlo estimate:** $\widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} g(X_i)$

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:

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

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

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

Inverse transform method

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

To sample $X_1, ..., X_n \overset{iid}{\sim} f_2$:

- Sample $U_1, ..., U_n \overset{iid}{\sim} Uniform(0, 1)$
- $X_i = F_2^{-1}(U_i)$

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] 0.001982649
```

```
sd(theta_hat2)
```

```
## [1] 0.04452695
```

Comparing variance

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

- $f_1(x) = 1,\ g_1(x) = \dfrac{e^{-x}}{1+x^2}$: $Var(\widehat{\theta}_1) \approx 0.006$
- $f_2(x) = \dfrac{4}{\pi(1+x^2)},\ g_2(x) = \frac{\pi}{4} e^{-x}$: $Var(\widehat{\theta}_2) \approx 0.002$

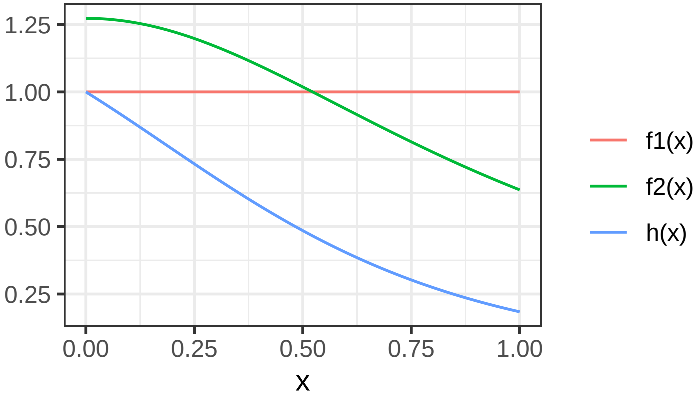
Relative efficiency: $\dfrac{Var(\widehat{\theta}_2)}{Var(\widehat{\theta}_1)} \approx \dfrac{1}{3}$

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

Comparing options

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

- $f_1(x) = 1,\ g_1(x) = \dfrac{e^{-x}}{1+x^2}$: $Var(\widehat{\theta}_1) \approx 0.006$
- $f_2(x) = \dfrac{4}{\pi(1+x^2)},\ g_2(x) = \frac{\pi}{4}e^{-x}$: $Var(\widehat{\theta}_2) \approx 0.002$



Importance sampling

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

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

- $X_1, ..., X_n \overset{iid}{\sim} f$
- $\widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \frac{h(X_i)}{f(X_i)}$

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

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

Importance sampling

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

- $Var(\widehat{\theta})$ 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 $Var\left(\frac{h(X)}{f(X)}\right) = Var\left(\frac{1}{c}\right) = 0$
- $Var(\widehat{\theta})$ is minimized if $f(x) = \dfrac{|h(x)|}{\int |h(x)|dx}$
- 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