

Practical: Monte Carlo and Markov chain theory

Instructors: Kari Auranen, Elizabeth Halloran and Vladimir Minin

July 20 – July 22, 2020

Estimating the tail of the standard normal distribution

Let $Z \sim \mathcal{N}(0, 1)$. We would like to estimate the tail probability $\Pr(Z > c)$, where c is large (e.g., $c = 4.5$).

Naïve Monte Carlo: simulate $Z_1, \dots, Z_n \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$. Then

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n 1_{\{Z_i > c\}} \approx \mathbb{E}(1_{\{Z > c\}}) = \Pr(Z > c).$$

This estimator will most likely give you 0 even for $n = 10,000$. The problem is the large variance of the integrand:

$$\text{Var}(\hat{\mu}) = \frac{1}{n} \text{Var}(1_{\{Z_1 > c\}}) = \frac{1}{n} \Pr(Z_1 > c)[1 - \Pr(Z_1 > c)] = \mathbf{3.4} \times \mathbf{10^{-10}} \text{ for } n = 10,000 \text{ and } c = 4.5.$$

This variance is huge, because the quantity of interest is $\Pr(Z_1 > c) = 3.39 \times 10^{-6}$ and the standard deviation of our estimator is 1.84×10^{-5} . So the standard deviation of our estimator exceeds to quantity we are trying to estimate.

Importance sampling: Simulate $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} \text{Exp}(c, 1)$ from a shifted exponential with density

$$g(y) = e^{-(y-c)} 1_{\{y > c\}}.$$

Generating such random variables is very easy: just simulate a regular exponential $\text{Exp}(1)$ and add c to the simulated value. Then the importance sampling estimator becomes

$$\tilde{\mu} = \frac{1}{n} \sum_{i=1}^n \frac{\phi(Y_i)}{g(Y_i)} 1_{\{Y_i > c\}} = \tilde{\mu} = \frac{1}{n} \sum_{i=1}^n \frac{\phi(Y_i)}{g(Y_i)},$$

where $\phi(x)$ is the standard normal density. Notice that we dropped the indicator function because the way we simulated Y_i s guarantees that its value is always 1.0. The variance of this estimator amounts to

$$\begin{aligned} \text{Var}(\tilde{\mu}) &= \frac{1}{n} \text{Var} \left[\frac{\phi(Y)}{g(Y)} \right] = \frac{1}{n} \left\{ \mathbb{E}_g \left[\frac{\phi^2(Y)}{g^2(Y)} \right] - \left[\mathbb{E}_g \left(\frac{\phi(Y)}{g(Y)} \right) \right]^2 \right\} \\ &= \frac{1}{n} \left[\int_c^\infty \frac{\phi^2(y)}{g(y)} dy - \Pr(Z > c)^2 \right] = \mathbf{1.95} \times \mathbf{10^{-15}} \text{ for } n = 10,000 \text{ and } c = 4.5. \end{aligned}$$

The standard deviation of the importance sampling estimator is 4.4×10^{-8} . This means that importance sampling reduced Monte Carlo error roughly by a factor of 400 from using 1.84×10^{-5} to 4.4×10^{-8} .

Your task

Implement naive and importance sampling Monte Carlo estimates of $\Pr(Z > 4.5)$, where $Z \sim \mathcal{N}(0, 1)$. Download ‘import_sampl_reduced.R’ from the course web page. The code has a couple of things to get you started.