NTNU | Norwegian University of Science and Technology

TMA4300 - COMPUTER INTENSIVE STATISTICAL METHODS

Exercise 1: Problem C

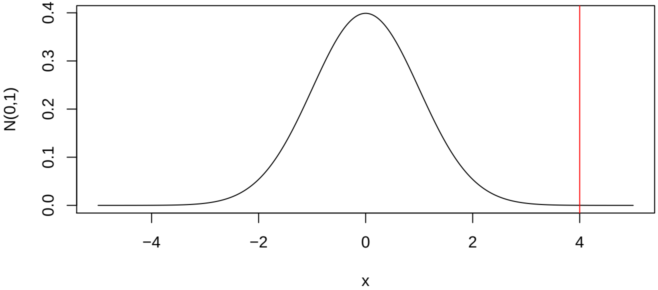
Jim Totland, Martin Tufte

March 1, 2022

Problem C: Monte Carlo integration and variance reduction

Aim: Learn about variance reduction techniques in Monte Carlo integration.

Problem: Find $\theta = \text{Prob}(X > 4)$ when $X \sim N(0,1)$.



C.1: Monte Carlo integration

Estimate θ by Monte Carlo integration using $n = 100000$ samples from the standard normal distribution. Compute also a 95% confidence interval for θ based on these n samples.

C.1: Monte Carlo integration

Estimate θ by Monte Carlo integration using $n = 100000$ samples from the standard normal distribution. Compute also a 95% confidence interval for θ based on these n samples.

Rewrite θ using expected value:

$$\theta = \Pr\{X > 4\} = \int_4^\infty \phi(x)\,\mathrm{d}x = \int_{-\infty}^\infty \mathbb{I}(x > 4)\,\phi(x)\,\mathrm{d}x = \mathrm{E}[\mathbb{I}(X > 4)]$$

C.1: Monte Carlo integration

Estimate θ by Monte Carlo integration using $n = 100000$ samples from the standard normal distribution. Compute also a 95% confidence interval for θ based on these n samples.

Rewrite θ using expected value:

$$\theta = \Pr\{X > 4\} = \int_4^\infty \phi(x)\,\mathrm{d}x = \int_{-\infty}^\infty \mathbb{I}(x > 4)\,\phi(x)\,\mathrm{d}x = \mathrm{E}[\mathbb{I}(X > 4)]$$

Use sample mean as estimator:

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^n \underbrace{\mathbb{I}(x_i > 4)}_{h(x_i)}$$

C.1: Confidence Intervals

Estimate variance:

$$\mathrm{Var}[\hat{\theta}] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}_f[h(X_i)] = \frac{1}{n}\mathrm{Var}_f[h(X)] \approx \frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}(h(x_i)-\hat{\theta})^2\right) =: \hat{\sigma}^2$$

C.1: Confidence Intervals

Estimate variance:

$$\mathrm{Var}[\hat{\theta}] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}_f[h(X_i)] = \frac{1}{n}\mathrm{Var}_f[h(X)] \approx \frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}(h(x_i)-\hat{\theta})^2\right) =: \hat{\sigma}^2$$

Using the central limit theorem: $\hat{\theta} \sim \mathcal{N}(\theta, \mathrm{Var}[\hat{\theta}]) \implies \frac{\hat{\theta}-\theta}{\hat{\sigma}} \sim t_{n-1}$.

A 95% confidence interval is therefore

$$[\hat{\theta} - t_{n-1,\,0.025}\cdot\hat{\sigma},\ \hat{\theta} + t_{n-1,\,0.025}\cdot\hat{\sigma}].$$

C.2: Importance sampling

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

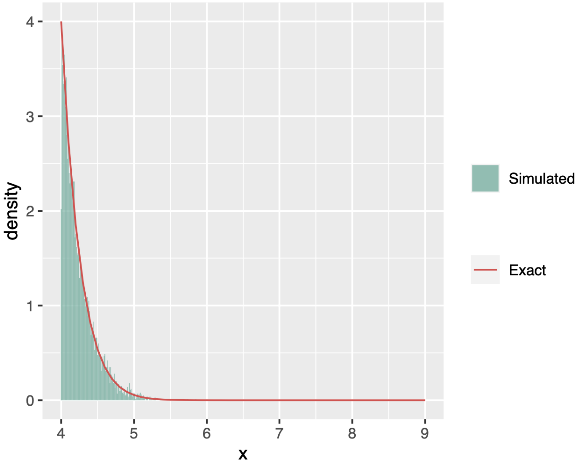
C.2: Importance sampling

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

Need to sample from proposal distribution $g(x)$.

C.2: Importance sampling

$$g(x) = \begin{cases} cx \exp -\frac{1}{2}x^2 & , x > 4 \\ 0 & , \text{else} \end{cases}$$



C.2: Importance sampling

$\hat{\theta}_{IS} = \frac{1}{n}\sum_{i=1}^{n} \frac{h(X_i)f(X_i)}{g(X_i)} = \frac{1}{n}\sum_{i=1}^{n} h(X_i)w(X_i)$. Find variance:

$$\begin{aligned}\mathrm{Var}[\hat{\theta}_{IS}] &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}_g[h(X_i)w(X_i)] = \frac{1}{n}\mathrm{Var}_g[h(X)w(X)] \\ &\approx \frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}(h(x_i)w(x_i) - \hat{\theta}_{IS})^2\right) =: \hat{\sigma}^2_{IS}.\end{aligned}$$

C.2: Importance sampling

$\hat{\theta}_{IS} = \frac{1}{n}\sum_{i=1}^{n} \frac{h(X_i)f(X_i)}{g(X_i)} = \frac{1}{n}\sum_{i=1}^{n} h(X_i)w(X_i)$. Find variance:

$$\begin{aligned}\mathrm{Var}[\hat{\theta}_{IS}] &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}_g[h(X_i)w(X_i)] = \frac{1}{n}\mathrm{Var}_g[h(X)w(X)] \\ &\approx \frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}(h(x_i)w(x_i) - \hat{\theta}_{IS})^2\right) =: \hat{\sigma}_{IS}^2.\end{aligned}$$

95% Confidence interval: $\left[\hat{\theta}_{IS} - t_{n-1,0.025}\cdot\hat{\sigma}_{IS},\ \hat{\theta}_{IS} + t_{n-1,0.025}\cdot\hat{\sigma}_{IS}\right]$

C.3: Antithetic sampling

$$\hat{\theta}_{AS} = \frac{1}{2}\left(\hat{\theta}_{IS}^{(1)} + \hat{\theta}_{IS}^{(2)}\right)$$

C.3: Antithetic sampling

$$\hat{\theta}_{AS} = \frac{1}{2}\left(\hat{\theta}_{IS}^{(1)} + \hat{\theta}_{IS}^{(2)}\right)$$

$$\hat{\theta}_{IS}^{(1)} = \frac{1}{n}\sum_{i=1}^{n} h\left(X_i^{(1)}\right) w\left(X_i^{(1)}\right) \text{ and } \hat{\theta}_{IS}^{(2)} = \frac{1}{n}\sum_{i=1}^{n} h\left(X_i^{(2)}\right) w\left(X_i^{(2)}\right).$$

The $X_j^{(1)}$'s are based on u while the $X_j^{(2)}$'s are based on $u-1$ in the inversion sampling of $g(x)$.

C.3: Antithetic sampling

$$\hat{\theta}_{AS} = \frac{1}{2}\left(\hat{\theta}_{IS}^{(1)} + \hat{\theta}_{IS}^{(2)}\right)$$

$$\hat{\theta}_{IS}^{(1)} = \frac{1}{n}\sum_{i=1}^{n} h\left(X_i^{(1)}\right) w\left(X_i^{(1)}\right) \text{ and } \hat{\theta}_{IS}^{(2)} = \frac{1}{n}\sum_{i=1}^{n} h\left(X_i^{(2)}\right) w\left(X_i^{(2)}\right).$$

The $X_j^{(1)}$'s are based on u while the $X_j^{(2)}$'s are based on $u-1$ in the inversion sampling of $g(x)$.

$\implies$ Only need $n = 50000$ to make $2n$ (antithetic) samples from $g(x)$.

C.3: Confidence intervals part 1

$$\hat{\theta}_{AS} = \frac{1}{2}\left(\hat{\theta}_{IS}^{(1)} + \hat{\theta}_{IS}^{(2)}\right).$$

C.3: Confidence intervals part 1

$\hat{\theta}_{AS} = \frac{1}{2}\left(\hat{\theta}_{IS}^{(1)} + \hat{\theta}_{IS}^{(2)}\right)$. Variance:

$$\mathrm{Var}\left[\hat{\theta}_{AS}\right] = \frac{1}{4}\left(\mathrm{Var}\left[\hat{\theta}_{IS}^{(1)}\right] + \mathrm{Var}\left[\hat{\theta}_{IS}^{(2)}\right] + 2\mathrm{Cov}\left[\hat{\theta}_{IS}^{(1)}, \hat{\theta}_{IS}^{(2)}\right]\right).$$

C.3: Confidence intervals part 1

$\hat{\theta}_{AS} = \frac{1}{2}\left(\hat{\theta}_{IS}^{(1)} + \hat{\theta}_{IS}^{(2)}\right)$. Variance:

$$\mathrm{Var}\left[\hat{\theta}_{AS}\right] = \frac{1}{4}\left(\mathrm{Var}\left[\hat{\theta}_{IS}^{(1)}\right] + \mathrm{Var}\left[\hat{\theta}_{IS}^{(2)}\right] + 2\mathrm{Cov}\left[\hat{\theta}_{IS}^{(1)}, \hat{\theta}_{IS}^{(2)}\right]\right).$$

How to make confidence interval? Rewrite $\hat{\theta}_{AS}$

C.3: Confidence intervals part 1

$\hat{\theta}_{AS} = \frac{1}{2}\left(\hat{\theta}_{IS}^{(1)} + \hat{\theta}_{IS}^{(2)}\right)$. Variance:

$$\text{Var}\left[\hat{\theta}_{AS}\right] = \frac{1}{4}\left(\text{Var}\left[\hat{\theta}_{IS}^{(1)}\right] + \text{Var}\left[\hat{\theta}_{IS}^{(2)}\right] + 2\text{Cov}\left[\hat{\theta}_{IS}^{(1)}, \hat{\theta}_{IS}^{(2)}\right]\right).$$

How to make confidence interval? Rewrite $\hat{\theta}_{AS}$

$$\hat{\theta}_{AS} = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\frac{1}{2}\left(h\left(X_i^{(1)}\right) w\left(X_i^{(1)}\right) + h\left(X_i^{(2)}\right) w\left(X_i^{(2)}\right)\right)}_{=Y_i} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

C.3: Confidence intervals part 1

$\hat{\theta}_{AS} = \frac{1}{2}\left(\hat{\theta}_{IS}^{(1)} + \hat{\theta}_{IS}^{(2)}\right)$. Variance:

$$\text{Var}\left[\hat{\theta}_{AS}\right] = \frac{1}{4}\left(\text{Var}\left[\hat{\theta}_{IS}^{(1)}\right] + \text{Var}\left[\hat{\theta}_{IS}^{(2)}\right] + 2\text{Cov}\left[\hat{\theta}_{IS}^{(1)}, \hat{\theta}_{IS}^{(2)}\right]\right).$$

How to make confidence interval? Rewrite $\hat{\theta}_{AS}$

$$\hat{\theta}_{AS} = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\frac{1}{2}\left(h\left(X_i^{(1)}\right)w\left(X_i^{(1)}\right) + h\left(X_i^{(2)}\right)w\left(X_i^{(2)}\right)\right)}_{=Y_i} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Y_i's independent $\implies$ can use t-distribution to create C.I.

C.3: Confidence intervals part 2

Use sample variance:

$$\hat{\sigma}^2_{AS} = \frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \hat{\theta}_{AS})^2\right)$$

C.3: Confidence intervals part 2

Use sample variance:

$$\hat{\sigma}_{AS}^2 = \frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \hat{\theta}_{AS})^2\right)$$

And we know:

$$\frac{\hat{\theta}_{AS} - \theta}{\hat{\sigma}_{AS}} \sim t_{n-1}$$

C.3: Confidence intervals part 2

Use sample variance:

$$\hat{\sigma}^2_{AS} = \frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \hat{\theta}_{AS})^2\right)$$

And we know:

$$\frac{\hat{\theta}_{AS} - \theta}{\hat{\sigma}_{AS}} \sim t_{n-1}$$

From which we construct a confidence interval.