

## STA380 Additional Practice Problems for Unit 2

These problems are not to be handed in, but they are for extra practice for students to be prepared for term tests.

1. Consider  $X \sim \text{Gamma}(\alpha = 2, \beta = 2)$  where  $\alpha$  is the shape parameter, and  $\beta$  is the scale parameter.

- (a) Use the hit or miss method to compute  $\mathbb{P}(X < 2)$ . You may use `rgamma()` and  $n = 10^5$ .
- (b) Recall in lecture, we said the Monte Carlo estimate for the variance of an estimator:

$$\widehat{\mathbb{V}\text{ar}}[\hat{\theta}] = \frac{\hat{\sigma}^2}{n} = \frac{\sum_{i=1}^n [g(x_i) - \bar{g}(x)]^2}{n^2}$$

Use this to construct a 90% confidence interval of this estimate.

2. Consider the integral:

$$\theta = \int_0^{10} x \ln(x) dx$$

- (a) Use the simple Monte Carlo estimator to **estimate** the expected value, where the importance function is the *Uniform(0, 1)* distribution. Call this  $\hat{\theta}$ .
  - (b) Write an exact expression for  $\mathbb{V}\text{ar}[\hat{\theta}]$ . **You do not need to evaluate this integral by hand.** Then, compute this integral using `integrate()` in R.
- Disclaimer:** you need to know how to integrate for the tests; we also just want you to learn the `integrate()` function as well. You will not be using the `integrate()` function in R for the term tests or exam.
- (c) Recall in lecture, we said the Monte Carlo estimate for the variance of the estimator for the integral is:

$$\widehat{\mathbb{V}\text{ar}}[\hat{\theta}] = \frac{\hat{\sigma}^2}{n} = \frac{\sum_{i=1}^n [g(x_i) - \bar{g}(x)]^2}{n^2}$$

Using this equation, compute  $\widehat{\mathbb{V}\text{ar}}[\hat{\theta}]$  in R. Use  $n = 10^4$ .

- (d) Compare the MC estimator against the theoretical variance of  $\hat{\theta}$ . Use `all.equal()`, which will report the **mean relative difference**.

3. Consider the integral:

$$\theta = \int_0^{10} e^x \ln(x) dx$$

Repeat the same steps from 2.(a) to (d). You should notice that the theoretical variance of  $\hat{\theta}$  and the MC estimate of the variance are not that close. Why do you think that is?

This is an example of when having a uniform importance function performs poorly. Unfortunately, there is no well-known probability density function that will fix the issue for this particular integral.

4. We want to estimate the following integral using importance sampling:

$$\theta = \int_1^4 \exp\left\{-\frac{(\ln x)^2}{2}\right\} dx$$

- (a) Use `integrate()` to compute the “true” value of  $\theta$ .
- Disclaimer:** you need to know how to integrate for the tests; we also just want you to learn the `integrate()` function as well. You will not be using the `integrate()` function in R for the term tests or exam.
- (b) Estimate  $\theta$  using the simple Monte Carlo estimator. That is, the importance function is the pdf of some uniform distribution.
  - (c) Consider the **log-normal** distribution:

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\}.$$

Estimate  $\theta$  using the log-normal pdf as the importance function. Use  $\sigma^2 = 1, \mu = 0$ .

- (d) Which estimator performs better, (b) or (c)? Why do you think one performed better than the other? This question is somewhat open-ended; my intended solution involves comparing against the variances.

5. Recall the beta distribution, which has pdf:

$$f_X(x) = \left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1.$$

Can we use antithetic variables to estimate the expected value of the  $X \sim Beta(\alpha = 2, \beta = 4)$  distribution? Why or why not? If the antithetic approach can be used, then use the antithetic approach to estimate the expected value of  $X$ .

6. Consider the exponential distribution with a scale parameter  $\theta$ .

(a) For what values of  $\theta$  can we compute  $\mathbb{E}[X \mathbf{1}_{\{0 < X < 1\}}]$  using antithetic variables? As a reminder,

$$\mathbb{E}[X \mathbf{1}_{\{0 < X < 1\}}] = \int_0^1 x f_X(x) dx.$$

- (b) Pick any value of  $\theta$  as long as the antithetic approach is valid. Then, use the antithetic approach to estimate  $\mathbb{E}[X \mathbf{1}_{\{0 < X < 1\}}]$ .  
(c) Compare the estimator of  $\mathbb{E}[X \mathbf{1}_{\{0 < X < 1\}}]$  to the true value. You will need to remember integration by parts to evaluate this integral.

Below are the solutions.

```
1. n <- 10^5
x <- rgamma(n, shape = 2, scale = 2)
cond <- (x<2)

est <- mean(cond)
var <- sum((cond - est)^2)/n^2
alpha <- 0.10
err_bds <- qnorm(1-alpha/2) * sqrt(var)
c(est - err_bds, est + err_bds)
```

2. (a)

$$\theta = \int_0^{10} x \ln(x) dx = \int_0^{10} \frac{1}{10} 10x \ln(x) dx = \mathbb{E}[10x \ln(x)]$$

Generate  $n = 10^4$  samples from  $U \stackrel{i.i.d.}{\sim} Uniform(0, 10)$ . Using SLLN, the simple MC estimator is

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n 10u_i \ln(u_i)$$

(b)

$$\begin{aligned} \text{Var}[\hat{\theta}] &= \frac{1}{n^2} \sum_{i=1}^n 100\text{Var}[u_i \ln(u_i)] \\ &= \frac{1}{n^2} \sum_{i=1}^n 100[\mathbb{E}[(u_i \ln(u_i))^2] - (\mathbb{E}[u_i \ln(u_i)])^2] \\ &= \frac{1}{n^2} \sum_{i=1}^n 100 \left[ \int_0^{10} (u_i \ln(u_i))^2 \frac{1}{10} du_i - \left( \int_0^{10} u_i \ln(u_i) \frac{1}{10} du_i \right)^2 \right] \\ &= \frac{100}{n} \left[ \int_0^{10} (u_i \ln(u_i))^2 \frac{1}{10} du_i - \left( \int_0^{10} u_i \ln(u_i) \frac{1}{10} du_i \right)^2 \right] \end{aligned}$$

```
fun1 <- function(x){log(x)^2*x^2/10}
fun2 <- function(x){log(x)*x/10}

ex2 <- integrate(fun1, 0, 10)
ex_sq <- integrate(fun2, 0, 10)
exact <- 100 * (ex2$value - (ex_sq$value)^2) / n
```

(c)

```
n <- 10^4
u <- runif(n, 0, 10)
mc_est <- sum((10*log(u)*u - mean(10*log(u)*u))^2) / n^2
```

(d) Answers will vary depending on the seed used.

```
> abs(mc_est - exact)
[1] 0.008770158
> all.equal(mc_est, exact)
[1] "Mean relative difference: 0.01724408"
```

3. We will skip part (a), as it is extremely similar to the previous question.

```
n <- 10^4
u <- runif(n, 0, 10)

# estimate
mc_est <- sum((10*log(u)*exp(u) - mean(10*log(u)*exp(u)))^2) / n^2

# exact
```

```

fun1 <- function(x){log(x)^2*exp(2*x)/10}
fun2 <- function(x){log(x)*exp(x)/10}

ex2 <- integrate(fun1, 0, 10)
ex_sq <- integrate(fun2, 0, 10)
exact <- 100 * (ex2$value - (ex_sq$value)^2) / n

# big difference, somewhat better mean relative difference
> abs(mc_est - exact)
[1] 43650.5
> all.equal(mc_est, exact)
[1] "Mean relative difference: 0.04583315"

```

4. (a) `fun <- function(x){exp(-(log(x))^2 / 2)}`  
`true <- integrate(fun, lower = 1, upper = 4)`

(b) At this point, you should be comfortable writing the algorithm, so we'll only show the code.

```

n <- 10^5 # arbitrary; students can use whatever they like.
x <- runif(n, 1, 4)
est1 <- 3 * mean(exp(-(log(x))^2 / 2))

```

(c) For the choice of  $\mu = 0, \sigma^2 = 1$  we have:

$$f_X(x) = \frac{1}{x\sqrt{2\pi}} \exp\left\{-\frac{(\ln x)^2}{2}\right\}.$$

And thus,

$$\begin{aligned} \int_1^4 \exp\left\{-\frac{(\ln x)^2}{2}\right\} dx &= \int_1^4 \exp\left\{-\frac{(\ln x)^2}{2}\right\} \frac{x\sqrt{2\pi} \exp\left\{-\frac{(\ln x)^2}{2}\right\}}{x\sqrt{2\pi} \exp\left\{-\frac{(\ln x)^2}{2}\right\}} dx \\ &= \int_1^4 \frac{x\sqrt{2\pi} \exp\left\{-\frac{(\ln x)^2}{2}\right\}}{x\sqrt{2\pi}} dx \\ &= \mathbb{E}[x\sqrt{2\pi} \mathbf{1}_{\{1 \leq X \leq 4\}}] \end{aligned}$$

The algorithm is as follows:

- i. Let  $X_1, \dots, X_n$  be a random variable generated from  $\text{LogNormal}(\mu = 0, \sigma^2 = 1)$ .
- ii. Compute:

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n x_i \sqrt{2\pi} \mathbf{1}_{\{1 \leq X_i \leq 4\}}.$$

The code is:

```

n <- 10^5
x <- rlnorm(n)
est2 <- mean(x * sqrt(2*pi) * (x >= 1 & x <= 4))

```

(d) I used `all.equal()` for comparison, although you're free to simply check the absolute distance or any other reasonable metric as well.

```

all.equal(true$value, est1)
## [1] "Mean relative difference: 0.0005045298"
all.equal(true$value, est2)
## [1] "Mean relative difference: 0.001480595"

```

In this case, (b) actually performs better than the one in (c). Recall in lecture, we derived the following variance of the estimator via the importance sampling method:

$$\text{Var}(\hat{\theta}) = \frac{1}{n} \left[ \left( \int \frac{g(x)^2}{f(x)} dx \right) - \theta^2 \right]$$

This means that the variance of the first estimator is then:

$$\text{Var}(\hat{\theta}_{(b)}) \left[ \left( \int_1^4 3 \exp\{-(\ln x)^2\} dx \right) - \theta^2 \right]$$

And then the variance of the second must be:

$$\text{Var}(\hat{\theta}) = \frac{1}{n} \left[ \left( \int_1^4 \exp\left\{-\frac{(\ln x)^2}{2}\right\} x \sqrt{2\pi} dx \right) - \theta^2 \right]$$

Let's make our lives easier and `integrate()` in R:

```
fun1 <- function(x){3 * exp(-(log(x))^2)}
fun2 <- function(x){exp(-(log(x))^2/2) * x * sqrt(2*pi)}

int1 <- integrate(fun1, 1, 4)
int2 <- integrate(fun2, 1, 4)

# variance with uniform importance function
(1/n) * (int1$value - true$value^2)
## [1] 3.442339e-06
# variance with lognormal importance function
(1/n) * (int2$value - true$value^2)
## [1] 7.335948e-05
```

we see here that the variance of the second estimator is indeed larger, and thus likely to give us a worse estimate.

5. Note that a  $Beta(\alpha = 2, \beta = 4)$  distribution has the pdf:

$$\begin{aligned} f_X(x) &= \left[ \frac{\Gamma(2+4)}{\Gamma(2)\Gamma(4)} \right] x^{2-1}(1-x)^{4-1} \\ &= \frac{(6-1)!}{(2-1)!(4-1)!} x(1-x)^3 \\ &= \frac{5!}{3!} x(1-x)^3 \\ &= \frac{5!}{3!} x(1-x)^3 \\ &= 20x(1-x)^3 \end{aligned}$$

Thus,  $\mathbb{E}[X] = \int_0^1 20x^2(1-x)^3 dx$ . Using the antithetic approach, we would then claim:

$$g(x) = 20x^2(1-x)^3$$

However, this function is not monotone. One way to verify this (this is just the first one that popped in my head without graphing) is to check when a local maximum or minimum occurs, and see that it occurs within the interval  $[0, 1]$ . Note,

$$\begin{aligned} g'(x) &= 20(2x)(1-x)^3 + 20x^2(3)(1-x)^2(-1) \\ &= 40x(1-x)^3 - 60x^2(1-x)^2 \\ 0 &\stackrel{\text{set}}{=} 40x(1-x)^3 - 60x^2(1-x)^2 \\ &= (1-x)^2 x(40(1-x) - 60x) \\ &= (1-x)^2 x(40 - 40x - 60x) \\ &= (1-x)^2 x(40 - 100x) \end{aligned}$$

The solutions are  $x = 0, 0.4, 1$ . Since clearly  $0.4 \in [0, 1]$ , we have a min/max occurs at  $x = 0.4$  and thus the function is likely not monotone. Hence, the antithetic approach would not work here as we cannot guarantee that  $g(u_i)$  and  $g(1-u_i)$  are negatively correlated.

6. (a) We need to observe when the value of  $h(x) := xf_X(x)$  is monotone between  $[0, 1]$ . Note that,

$$\begin{aligned} h(x) &= \frac{1}{\theta}xe^{-\frac{1}{\theta}x} \\ h'(x) &= \frac{1}{\theta}e^{-\frac{1}{\theta}x} + \frac{1}{\theta}xe^{-\frac{1}{\theta}x}\left(-\frac{1}{\theta}\right) \\ 0 &\stackrel{\text{set}}{=} e^{-x/\theta}\left(\frac{1}{\theta} - \frac{1}{\theta^2}x\right) \\ x &= \theta \end{aligned}$$

Hence, a critical value appears when  $x = \theta$ . Note that we only care about the function being monotone between  $[0, 1]$ . Hence, as long as this critical value occurs at or after 1, we're safe. We can use the antithetic approach if  $\theta \geq 1$ .

(b) For this solution we picked  $\theta = 3$ , but again, this is arbitrary.

```
n <- 10^4
u <- runif(n/2)
theta <- 3 # arbitrary; pick what you want
g1 <- (1/theta) * u * exp(-u/theta)
g2 <- (1/theta) * (1-u) * exp(-(1-u)/theta)
g_total <- mean((g1 + g2)/2)
```

(c) Note that:

$$\mathbb{E}[X\mathbf{1}_{\{0 < X < 1\}}] = \int_0^1 \frac{1}{\theta}xe^{-\frac{1}{\theta}x}dx$$

We'll use integration by parts. Recall:

$$u = \frac{x}{\theta}, \quad du = \frac{dx}{\theta}, \quad dv = e^{-x/\theta}dx, \quad v = -\theta e^{-x/\theta}$$

Thus,

$$uv - \int vdu := \frac{x}{\theta}(-\theta e^{-x/\theta}) + \int (-\theta e^{-x/\theta})\frac{dx}{\theta} = -xe^{-x/\theta} - \theta e^{-x/\theta}$$

Evaluate the previous integral with the bounds  $x = 0$  to  $x = 1$  grants us:

$$-e^{-1/\theta} - \theta e^{-1/\theta} - (0 - \theta) = -e^{-1/\theta}(1 + \theta) + \theta$$

For our comparison:

```
true_cond_exp <- -exp(-1/theta) * (1 + theta) + theta

library(testthat)
test_that("checking the MC cond exp versus true", {
  expect_equal(g_total, true_cond_exp, tol = 0.001)
})
```

Worked well on my end. ;)