STAT400 - Homework 6

Henrique Rio

Due 10/15/2020 by 4pm

Be sure to `set.seed(400)` at the beginning of your homework.

```
#reproducibility
set.seed(400)
```

```
# useful libraries
library(tidyverse)
```

1. Develop two Monte Carlo integration approaches to estimate $\int_0^5 x^2 \exp(-x)dx$. (You must use different distributions in the two approaches). Check your answer using the `integrate()` function.

```
# method 1
m <- 1000
x <- runif(m,min = 0,max=5)
theta.hat <-mean(x^2*exp(-x))*5
print(theta.hat)
```

```
## [1] 1.777716
```

```
# method 2
x2 <-rexp(m,10)

theta.hat2 <- mean(x2^2*exp(-x2))


# check answers with integrate()'
integrand <- function(x) {x^2*exp(-x)}
integrate(integrand,0,5)
```

```
## 1.750696 with absolute error < 1.9e-14
```

2. Estimating the cdf of a normal distribution. Use $m = 1000$.

   a. Implement all 3 methods that we discussed in class (Example 1.7, Page 9-10 of Ch. 6 Notes) to estimate the cdf of a normal distribution $\Phi(x)$. Note that you will need to show some derivations for method 2.
   b. Compare your estimates with the output from the R function `pnorm()` for $x = 0.5, 1, 2, 3$. Summarise your findings comparing the performance of the methods.
   c. For each method, compute an estimate of the variance of your Monte Carlo estimate of $\Phi(2)$. Summarise your findings.
   d. For each method, compute a 95% confidence interval for $\Phi(2)$. Summarise your findings. Which CI is narrower and why does that matter?

```
# method 1\
 m <- 1000
 y1 <- runif(m)
```

```r
x <- 2

f1 <- function(x) {(x/(sqrt(2*pi))*exp(-1/2*(x*y1)^2))*(x>0)}


fihat1 <- 1/2 + mean(f1(x))




# method 2
m <- 1000

y <- runif(m,0,2)

f2 <- function(x) {(x/(sqrt(2*pi))*exp(-1/2*(x*y)^2))*(x>0)}


fihat2 <- 1/2 + mean(f2(x))*2




# method 3
z <- rnorm(m,0,1)



fihat3 <- 1-mean(z[z<=x])


# compare to pnorm for x = 0.5, 1, 2, 3
fi <- pnorm(x)

# compute estimates of variance for Phi(2)
var1 <- sum((f1(x)-fihat1)^2) / m
var2 <- sum((f2(x)-fihat2)^2) / m
var3 <- sum((z[z<=x]-fihat2)^3) / m





# compute CIs for Phi(2)


CI1 <- c(fihat1-1.96*sqrt(var1),fihat1+1.96*sqrt(var1))
CI2 <- c(fihat2-1.96*sqrt(var2),fihat2+1.96*sqrt(var2))
CI3 <- c(fihat3-1.96*sqrt(var3),fihat3+1.96*sqrt(var3))
```

```
## Warning in sqrt(var3): NaNs produced

## Warning in sqrt(var3): NaNs produced
```

```r
print(c(CI1,CI2,CI3))
```

```
## [1] -0.1144926  2.0490877 -0.5677177  2.5700155        NaN        NaN
```

part c) they seem to perform more or less similar when changing the X value

part d) the variance for method 1 is the smalles out of all of them.

3. Compute a Monte Carlo estimate $\hat{\theta}_1$ of

$$\theta = \int_0^{0.5} e^{-x} dx$$

by sampling from the Uniform$(0, 0.5)$ and estimate the variance of $\hat{\theta}_1$. Find another Monte Carlo estimator $\hat{\theta}_2$ by sampling from the Exponential(1) distribution and estimating its variance.

Which of the variances (of $\hat{\theta}_1$ or $\hat{\theta}_2$) is smaller?

```r
# number of samples
m <- 1000

## first MC estimator and variance
## use the Unif(0, .5) dsn
m <- 1000
x <- runif(m,min = 0,max=0.5)
theta.hat <- mean(exp(-x))*0.5
print(theta.hat)
```

```
## [1] 0.3909057
```

```r
var1 <- sum((exp(-x)*0.5-theta.hat)^2) / m
print(var1)
```

```
## [1] 0.003205271
```

```r
## second MC estimator and variance
## use the Exp(1) dsn
m <- 1000
x <- rexp(m)
theta.hat <- mean(x/2)
print(theta.hat)
```

```
## [1] 0.4732539
```

```r
var1 <- sum((x/2-theta.hat)^2) / m
print(var1)
```

```
## [1] 0.209666
```

```r
## Compare estimated variances
```

$\hat{\theta}_1$ appears to have a much smaller variance when compared to $\hat{\theta}_1$