Homework 3 - PCS

Elena Volpi

10/21/2021

#Problem 1: Excercise 9.1 Suppose we want to estimate θ, where:

$$\int_0^1 e^{x^2} dx$$

Show that generating a random number U and then using the estimator $\frac{e^{U^2}(1+e^{1-2U})}{2}$ is better than generating two random number U_1 and U_2 and using

Let our estimators be $\theta_1 = \frac{e^{U^2}(1+e^{1-2U})}{2}$ and $\theta_2 = \frac{e^{U_1^2}+e^{U_2}{}^2}{2}$. We calculate and compare variances.

$$\begin{aligned}
\text{var}(\theta_1) &= \text{var}(\frac{e^{U^2}(1+e^{1-2U})}{2}) \\
&= \frac{1}{4}\text{var}(e^{U^2} + e^{1-2U+U^2}) \\
&= \frac{1}{4}\text{var}(e^{U^2} + e^{(1-U)^2}) \\
&= \frac{1}{4}(\text{var}(e^{U^2}) + \text{var}(e^{(1-U)^2}) + 2\text{cov}(e^{u^2}, e^{(1-U)^2}))
\end{aligned}$$

Since $U\ (0,1)$, utilizing WolframAlpha we find

$$\begin{aligned}
E(e^{2U^2}) &\approx 2.36445 \\
(E(e^{U^2}))^2 &\approx (1.46265)^2 \approx 2.13935 \\
\text{var}(e^{U^2}) &= E((e^{U^2})^2) - (E(e^{U^2}))^2 \approx 0.2251 \\
E(e^{2(1-U)^2}) &\approx 2.36445 \\
E(e^{(1-U)^2}) &\approx 1.46265 \\
\text{var}(e^{(1-U)^2}) &= E(e^{2(1-U)^2}) - (E(e^{(1-U)^2}))^2 \approx .2251 \\
E(e^{U^2} * e^{(1-U)^2}) &= E(e^{1-2U+2U^2}) \approx 1.97015 \\
\text{cov}(e^{U^2}, e^{(1-U)^2}) &= E(e^{U^2}e^{(1-U)^2}) - E(e^{U^2})E(e^{(1-U)^2}) \\
&\approx 1.97015 - (1.46265)(1.46265) \\
&\approx -0.169195
\end{aligned}$$

Therefore,

$$\begin{aligned}
\text{var}(\theta_1) &= \frac{1}{4}(\text{var}(e^{U^2}) + \text{var}(e^{(1-U)^2}) + 2\text{cov}(e^{u^2}, e^{(1-U)^2})) \\
&\approx \frac{1}{4}(0.2251 + 0.2251 + 2(-0.169195)) \\
&\approx 0.0279525
\end{aligned}$$

Similarly, the variance of θ_2 can be found:

$$\begin{aligned}\text{Var}(\theta_2) &= var(\frac{e^{U_1^2} + e^{{U_2}^2}}{2}) \\ &= \frac{1}{4}(var(e^{U_1^2}) + var(e^{U_2^2} + 2cov(e^{U_1^2}, e^{U_2^2}))\end{aligned}$$

Since U_1 and U_2 are identically and independently distributed, $\text{Cov}(U_1, U_2) = 0$ and $\text{Var}(U_1) = \text{Var}(U_2)$. We found previously, $\text{var}(e^{U_1^2}) \approx 0.2251$. Therefore, $\text{var}(e^{{U_2}^2}) \approx 0.2251$. Thus

$$\begin{aligned}var(\theta_2) &\approx \frac{1}{4}(2 * 0.2251) \\ &\approx 0.11255\end{aligned}$$

. We see that generating a random number U and then using the estimator θ_1 is better than using θ_2 since $\text{var}(\theta_1) < \text{var}(\theta_2)$.

#Problem 2: Excercise 9.3 Let X_i, i=1,..,5, be independent exponential random variables each with mean 1, and consider the quantity θ defined by

$$\theta = P(\sum_{i=1}^{5} iX_i \geq 21.6)$$

a) Explain how we can use simulation to estimate θ. There are two ways I see that we can approach this problem. We can do a 1000 samples of size 5 using the exponential distribution with mean 1. We can check if each sample multiplied by its index is greater than 21.6 and find the estimate for theta by computing the mean of the proportion to fit the criteria. Here it is shown below:

```
Sum_by_index <- function(n){
  i <- seq(1,n,by=1)  #get a sequence of 1 through n
  xi <- rexp(1,n=n)   #generate n exponential variables
  sum(i*xi)  #return each sum
}

theta_est <- function(N,n){
  sums <- rerun(N, Sum_by_index(n))
  flat_sums <- sums %>% flatten_dbl()  #Convert to dbl vector
  proportion <- (flat_sums < 21.6)
  theta <- 1-mean(proportion)
  return(theta)
}

theta_est(10000,5)
```

```
## [1] 0.163
```

b) Give the antithetic variable estimator.

```
sum_Anti <- function(n){
  i <- seq(1,n,by=1)
  U <- runif(ceiling(n/2))
  if(n%%2 == 0){
    xi <- c(-log(U), -log(1-U[1:(n/2)]))
  }else{
```

```
    xi <- c(-log(U), -log(1-U[1:floor(n/2)]))
  }
  sum(i*xi)
}

theta_est_anti <- function(N,n){
  sums <- rerun(N, sum_Anti(n))
  flat_sums <- sums %>% flatten_dbl()
  proportion <- (flat_sums >= 21.6)
  theta <- mean(proportion)
  return(theta)
}
theta_est_anti(1000,5)
```

```
## [1] 0.147
```

c) Is the use of antithetic variables efficient in this case?

```
reg_approach <- rerun(100,theta_est(100,5))%>%flatten_dbl()
anti_approach <- rerun(100, theta_est_anti(100,5))%>%flatten_dbl()
(var_reg <- var(reg_approach))
```

```
## [1] 0.001306778
```

```
(var_anti <- var(anti_approach))
```

```
## [1] 0.001346222
```

```
(diff <- var_reg - var_anti)
```

```
## [1] -3.944444e-05
```

Our antithetic variable approach has a variance 0.0008141414 smaller than my first approach, with a variance of 0.001117687. This suffices to show that the antithetic variable approach is the more efficient of the two.

#Problem 3: Excercise 9.10 In certain situations a random variable X, whose mean is known, is simulated so as to obtain an estimate of $P(X \leq a)$ for a given constant a. The raw simulation estimator from a single run is I, where

$$I = \begin{cases} 1, & \text{if } X \leq a \\ 0, & \text{if } X > a \end{cases}$$

Because I and X are clearly negative correlated, a natural attempt to reduce the variance is to use X as a control - and so use an estimator of the form I+c(X-E[X]).

a) Determine the percentage of variance reduction over the raw estimator I that is possible (by using the best c) if X were uniform(0,1)

Our control variate formula is $X + c(Y - \mu_Y)$ with mean following $E(X + c(Y - \mu_Y)) = E(X) + cE(Y - \mu_Y) = E(X)$ and variance following $\text{var}(X + c(Y - \mu_Y)) = \text{var}(X) + c^2\text{var}(Y) + 2c\,\text{cov}(X, Y)$. Our variance reduction is maximized when $c^* = -\frac{\text{cov}(X,Y)}{\text{var}(Y)}$, as was shown in recitation.

We then have,

$$\text{var}(X + c^*(Y - \mu_Y)) = \text{var}(X) - \frac{\text{cov}(X,Y)^2}{\text{var}(Y)}$$

$$= \frac{\text{var}(X + c^*(Y - \mu_Y))}{\text{var}(X)} = 1 - \text{cor}(X,Y)^2$$

Let $X\ Unif(0,1)$, then the c for maximization is $c^* = -\frac{\text{cov}(I,X)}{\text{var}(X)}$.

If $a \geq 1$ and $a \leq 0$, then $P(X \leq a) = 0$ or 1, respectively. We will consider $a\ (0,1)$

We then have, $\text{var}(I + c^*(X - \mu_X)) = \text{var}(I) - \frac{\text{cov}(I,X)^2}{\text{var}(X)}$.

Then, $cov(I,X) = E(IX) - E(I)E(X)$, where $E(X) = \frac{1}{2}$ and $E(I) = P(X \leq a) = a$ since X is uniformly distributed. Also,

$$IX = \begin{cases} X, & \text{if } X \leq a \\ 0, & \text{if } X > a \end{cases}$$

Then

$$\begin{aligned} E(IX) &= \int_0^a x\mathbf{1}_{x \leq a} f(x)dx \\ &= \int_0^a xdx \\ &= \frac{1}{2}a^2 \end{aligned}$$

Moreover, $cov(I,X) = \frac{1}{2}a^2 - \frac{1}{2}a = \frac{1}{2}(a^2 - a)$. It follows

$$\begin{aligned} c^* &= -\frac{\frac{1}{2}(a^2 - a)}{\frac{1}{12}} \\ &= 6(a^2 - a) \\ &= 6a(a-1) \end{aligned}$$

Futhermore,

$$\begin{aligned} var(I) &= E(I^2) - E(I)^2 \\ I^2 &= \begin{cases} 1^2, & \text{if } X \leq a \\ 0^2, & \text{if } X > a \end{cases} \\ I^2 &= \begin{cases} 1, & \text{if } X \leq a \\ 0, & \text{if } X > a \end{cases} \quad = I \\ var(I) &= a - a^2 = a(1-a) \\ cor(I,X)^2 &= \frac{cov(I,X)^2}{\text{var}(X)\text{var}(I)} \\ &= \frac{(\frac{1}{2}(a^2 - a))^2}{a(1-a)\frac{1}{12}} \\ &= 3a(1-a) \end{aligned}$$

b) Repeat a) if X were exponential with mean 1.

Let X ~Exp(1).

$$\begin{aligned} E(IX) &= \int_0^a xe^{-x}dx \\ &= 1 - (a+1)e^{-a} \\ cov(I,X) &= E(IX) - E(I)E(X) \\ &= 1 - (a+1)e^{-a} - a(1) \end{aligned}$$

Furthermore,

$$\begin{aligned}
c^* &= = -\frac{1-(a+1)e^{-a}-a}{1} \\
&= 1-a-(a+1)e^{-a}
\end{aligned}$$

Therefore,

$$\begin{aligned}
cor(I,X)^2 &= \frac{cov(I,X)^2}{var(X)var(I)} \\
cor(I,X)^2 &= \frac{(1-a-(a+1)e^{-a})^2}{a(1-a)}
\end{aligned}$$

c) Explain why we knew that I and X were negatively correlated. Its easy to show using the definition of covariance:

$$\begin{aligned}
cov(I,X) &= E(IX) - E(I)E(X) \\
&= \int_{-\infty}^{\infty} x\mathbf{1}_{x\le a} f(x)dx - P(x \le a)\mu_x \\
&= \int_{-\infty}^{a} xf(x)dx - \mu_x \int_{-\infty}^{a} f(x)dx \\
&= \int_{-\infty}^{a} (x-\mu_x)f(x)dx \\
&\le \int_{-\infty}^{\infty} (x-\mu_x)f(x)dx \\
&= E(X-\mu_x) \\
&= \mu_x - \mu_x \\
cov(I,X) &\le 0
\end{aligned}$$

##Problem 4: Excercise 9.12 a) Explain how control variables may be used to estimate θ in exercise 1. We may use control variables to estimate θ if we can introduce a random variable Y correlated with random variable X and the mean of Y, μ_Y is known. We would then estimate θ using $X + c(Y-\mu_Y)$, where we let $X = e^{u^2}$ and Y= U where U~Unif(0,1).We know the Var(Y) = $\frac{1}{12}$.

$$\begin{aligned}
Var(X + c^*(Y-\mu_y)) &= var(X) + (c^*)^2 var(Y) + 2c^* cov(X,Y) \\
&= var(X) + \left(-\frac{cov(X,Y)}{var(Y)}\right)^2 var(Y) + 2\left(-\frac{cov(X,Y)}{var(y)}\right) cov(X,Y) \\
&= var(X) - \frac{cov(X,Y)^2}{var(Y)} \\
&= var(X) - 12cov(X,Y)^2
\end{aligned}$$

b) Do 100 simulation runs, using the control given in a), to estimate first c* and then the variance of the estimator

```
est_cstar <- function(n, anti = FALSE){
  Y <- runif(n)
  X <- exp(Y^2)
  c_star <- -cov(X,Y)/(1/12)

  if(anti){
    if(n%%2 == 0){
      X_anti <- c(exp(Y[1:(n/2)]^2),exp((1-Y[1:(n/2)])^2))
    }else{
```

```
      X_anti <-c(exp(Y[1:ceiling(n/2)]^2),exp((1-Y[1:floor(n/2)])^       2))
    }
    (anti_var_est <- var(X_anti) - 12*cov(X,Y)^2)
    cat("The estimated of c* is ",c_star,"\n with an estimated          variance of ", anti_var_est,"\n
  }else {
    var_est <- var(X) - 12*cov(X,Y)^2
    cat("The estimate of c* is ",c_star,"\n with an estimated         variance of ", var_est,"\n")
  }
}

est_cstar(1000)
```

```
## The estimate of c* is  -1.517981
##  with an estimated         variance of  0.03274371
```

Our estimate for c* was -1.584297 with an estimated variance of 0.02846646.

c) Using the same data as in b), determine the variance of the antithetic variable estimator.

```
c <-est_cstar(1000, anti = TRUE)
```

```
## The estimated of c* is  -1.48982
##  with an estimated          variance of  0.03516818
```

Our estimated variance of the antithetic variable estimator was 0.0362307.

d) Which of the two types of variance reduction techniques worked better in this example?

In this example the control variable approach had the lesser variance and therefore was the better approach.