Homework 7

Louis Bensard

April 14, 2018

Problem 1:

Let $g(x) = \frac{e^{-x}}{1+x^2}$ and $X_1,\ldots,X_n \sim unif(0,1)$, then

$$\theta = (1-0)\int_0^1 \frac{g(x)}{1-0}dx \simeq (1-0)\cdot\frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

Thus,

$$\theta \simeq \frac{1}{n}\sum_{i=1}^{n} g(X_i) = \hat{\theta}_{MC}$$

```
g <- function(x){

    return(exp(-x)/(1+x^2))
}

n = 20000
X = runif(n,0,1)
```

(a)

```
theta_hat_MC = (1/n)*sum(g(X))
print(theta_hat_MC)
```

```
## [1] 0.52038
```

```
#standard error
g_bar_MC = (1/n)*sum(g(X))
ste_MC = sqrt((1/(n*(n-1)))*sum((g(X)-g_bar_MC)^2))
print(ste_MC)
```

```
## [1] 0.0017275
```

Thus, the value of the integral is approximated by $\hat{\theta}_{MC} = 0.52038$ with standard error $ste_{MC} = 1.7\cdot 10^{-3}$.

(b)

Since $X_1,\ldots,X_n \sim unif(0,1)$, then $F^{-1}(X_i) = X_i$, then $\hat{\theta}_{AS} = \frac{1}{2n}\sum_{i=1}^{n} g(X_i) + g(1-X_i)$.

```
# n=10000; X = runif(n,0,1) for test

theta_hat_AS = (1/(2*n))*sum(g(X) + g(1-X))
print(theta_hat_AS)
```

```
## [1] 0.52486
```

```r
#standard error
g_bar_AS = (1/n)*sum(g(X) + g(1-X))
ste_AS = sqrt((1/((4*n)*(n-1)))*sum((g(X) + g(1-X) - g_bar_AS)^2))
print(ste_AS)
```

```
## [1] 0.00023545
```

Thus, the value of the integral is approximated by $\hat{\theta}_{AS} = 0.52486$ with standard error $ste_{AS} = 2.3 \cdot 10^{-4}$.

conclusion:

The standard error obtained using Antithetic Sampling is way smaller than the one obtained from Monte-Carlo for almost the same approximation. Moreover, even if we use only n=10000 for AS (to even the total number of functions evaluations with MC), we still get a smaller standard error with AS than with MC (0.0017 vs 0.0003). Therefore, the Antithetic Method does a very good job reducing the standard error of a parameter estimate without being less accurate.

Problem 2:

(a)

If $X \sim f(x) = \frac{e^{-x}}{1-e^{-1}}$ $(0 < x1)$, then X has cdf $F(x) = \frac{1-e^{-x}}{1-e^{-1}}$ and thus:

$$F^{-1}(x) = -ln(1-(1-e^{-1})x)$$

Moreover, if $U \sim unif(0,1)$, then $F^{-1}(U) \sim X$, thus we have:

```r
F_inv <- function(x){

    return(-log(1-(1-exp(-1))*x))
}

#genrate n random values from X
n=20000
U = runif(n,0,1)
X = F_inv(U)
```

(b)

Let X_1,...,$X_n \sim f(x)$, then by Law of Large Numbers:

$$\theta = E[\frac{g(X)}{f(X)}] \simeq \frac{1}{n}\sum_{i=1}^{n}\frac{g(X_i)}{f(X_i)} = \frac{1}{n}\sum_{i=1}^{n}h(X_i)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\frac{1-e^{-1}}{1+x^2} = \hat{\theta}_{MC_2}$$

```r
h <- function(x){

    return((1-exp(-1))/(1+x^2))
```

```r
}

theta_hat_MC2 = (1/n)*sum(h(X))
print(theta_hat_MC2)
```

```
## [1] 0.52499
```

```r
#standard error
h_bar_MC = (1/n)*sum(h(X))
ste_MC2 = sqrt((1/(n*(n-1)))*sum((h(X)-h_bar_MC)^2))
print(ste_MC2)
```

```
## [1] 0.00068412
```

Thus, the value of the integral is approximated by $\hat{\theta}_{MC_2} = 0.52499$ with standard error $ste_{MC_2} = 6.8 \cdot 10^{-4}$. The standard error is significantly less than the standard error from MC in problem 1, so we can see that generating random values from $X \sim f(x)$ is a better choice than from $X \sim unif(0,1)$.

(c)

Let $Y_1^{(j)} = h(F^{-1}(U_j))$ and $Y_2^{(j)} = h(F^{-1}(1-U_j))$, then:

$$\hat{\theta}_{AS_2} = \frac{1}{2n}\sum_{j=1}^{n} Y_1^{(j)} + Y_2^{(j)}$$

```r
# n=10000; U = runif(n,0,1); X = F_inv(U) for test

Y_1 = h(F_inv(U))
Y_2 = h(F_inv(1-U))

theta_hat_AS2 = (1/(2*n))*sum(Y_1+Y_2)
print(theta_hat_AS2)
```

```
## [1] 0.52483
```

```r
#standard error
y_bar_AS = (1/n)*sum(Y_1 + Y_2)
ste_AS2 = sqrt((1/((4*n)*(n-1)))*sum((Y_1 + Y_2 - y_bar_AS)^2))
print(ste_AS2)
```

```
## [1] 0.00016814
```

Thus, the value of the integral is approximated by $\hat{\theta}_{AC_2} = 0.52483$ with standard error $ste_{AS_2} = 1.7 \cdot 10^{-4}$. The standard error is less than the standard error from AS in problem 1, so here as well, we can see that generating random values from $X \sim f(x)$ is a better choice than from $X \sim unif(0,1)$.

We can still note that the standard error from AS is still significantly less than the standard error from MC, as expected.

Problem 3:

(a)

Let $\boldsymbol{Y_1}, ..., \boldsymbol{Y_n} \sim f(\boldsymbol{x}|\Sigma)$, then:

$$\hat{\theta}_{HM} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{-\infty < Y_i^{(1)} < 1, -\infty < Y_i^{(2)} < 4, -\infty < Y_i^{(3)} < 2\}}$$

```
library(MASS)
n = 20000
mu = c(0,0,0)
Sigma = matrix(c(1,3/5,1/3,3/5,1,11/15,1/3,11/15,1),3,3)

Y = mvrnorm(n, mu, Sigma)

theta_hat_HM = sum((Y[,1]<1) & (Y[,2]<4) & (Y[,3]<2))/n
print(theta_hat_HM)
```

```
## [1] 0.82335
```

Thus, $\hat{\theta}_{HM} = 0.82335$.

(b)

```
#standard error
ste_HM = sqrt((theta_hat_HM - theta_hat_HM^2)/n)
print(ste_HM)
```

```
## [1] 0.0026967
```

So, $ste_{HM} = 2.7 \cdot 10^{-3}$.

(c)

```
L = theta_hat_HM-1.96*ste_HM
U = theta_hat_HM+1.96*ste_HM
cat("CI = [",L,", ", U,"]")
```

```
## CI = [ 0.81806 ,  0.82864 ]
```

$$\begin{aligned} CI_{95\%} &= \hat{\theta}_{HM} \pm t^* \cdot ste_{HM} \\ &= [0.81806, 0.82864] \end{aligned}$$

My confidence interval does contain the true value $\theta^* = 0.8279849897$.