Problem Set 2, Question 3: Solution

Statistics 506, Fall 2017

In this question you will design a Monte Carlo study to estimate a well known constant. Your code should use vectorization where possible.

a. Write a function to generate n iid samples from the square $\{(x_1, x_2) : |x_1| \leq 1, |x_2| \leq 1\}$.

The function `rsquare` below will generate data from any square centered at the origin.

```
# Generate n iid samples from the square centered
# at the origin with the given width.
rsquare = function(n, width=2){
  x = runif(2*n,-width/2, width/2)
  dim(x) = c(n,2)
  x
}
```

b. Generate data from this function and form a Monte Carlo estimate of the area of the unit circle $\{(x_1, x_2) : x_1^2 + x_2^2 \leq 1\}$. What well known constant are you estimating?

The unit circle has area π. The estimate $\hat{\pi} = 4p$ uses the proportion of points in the circle scaled by the area of the square.

```
## Estimate pi using the proportion of points in the unit circle
## multiplied by the area of the square.
n = 1e3
x = rsquare(n, 2)
p_hat = mean(x[,1]^2 + x[,2]^2 < 1)
est = 4*p_hat
```

An estimate based on 1000 Monte Carlo samples is $\hat{\pi} = 3.16$.

c. Report your results from part b with a 95% confidence interval. Does your interval cover the true value?

The the number of points in the circle is Binomial(n, p) and the standard error of the sample proportion $\hat{p}$ is $\sqrt{\hat{p}(1-\hat{p})/n}$. Our estimate is $4\hat{p}$ which has standard error $4\sqrt{\hat{p}(1-\hat{p})/n}$ since $\text{var}(4\hat{p}) = 16\text{var}(\hat{p})$.

```
## Our Monte Carlo estimate of the proportion is based on a Binomial(p) distribution.
## var(est) = var(4*p_hat) = 16*p*(1-p)/n
se = 4*sqrt(p_hat*{1-p_hat}/n)
z_a = qnorm(.975)
est_ci = est + c(-1, 1)*z_a*se
cover = ifelse(est_ci[1] < pi && est_ci[2] > pi, 'did', 'did not')
```

A 95% confidence interval for our previous estimate of $\hat{\pi} = 3.16$ is (3.05, 3.26) which did cover the true value of 3.14… on this trial.

d. Repeat part b with n large enough to estimate two significant digits accurately with 99% confidence. Briefly explain how you chose n and report your estimate with a 99% CI. How many digits is it accurate to?

An upper bound on the standard error is $4\sqrt{.5(.5)/n} = 2/\sqrt{n}$. For a 99% confidence interval, we need to use the multiplier $z_{.995} = 2.58$. To have accuracy for two significant digits (3.1), we need the width of the confidence interval to be less than .05. Solving the resulting inequality

$$2z_{.995}2/\sqrt{n} < .05 \implies n > 4z_{.995}/.05$$

so that $n > 4.2\text{e}{+}04$. I will round up and use `n = 5e5` below.

```
n = 5e5
x = rsquare(n, 2)
p_hat = mean(x[,1]^2 + x[,2]^2 < 1)
est = 4*p_hat
```

To report how many digits your estimate was accurate to you could set a random seed and compare by eye, making sure to account for rounding truncated digits. Below is a programmatic option using regular expressions.

```
## How many sig digits is est accurate to?
count_sig_digits = function(est, target){
  delta = abs(est - target)
  no_dec = regexpr('^[0]+', delta)
  dec = regexpr('^[0.]+', delta)
  if(dec!=1 & no_dec!=1){
    return(0)
  }
  max(attr(no_dec,"match.length"), attr(dec,"match.length")-1)
}
n_sig = count_sig_digits(est, pi)
```

In this case, our estimate of 3.141816 was accurate for 3.1415927 to 4 significant digits.

e. Repeat this exercise using the square $\{(x_1, x_2) : 0 \le x_1 \le 1, 0 \le x_2 \le 1\}$ and the portion of the unit circle in the positive quadrant. How do you need to adjust your Monte Carlo estimate to get an estimator of the same constant? How do you need to modify your confidence interval? (Hint: var$(aX) = a^2$var(X).)

In the modified question, the area of the square is now 1 but contains only $\frac{1}{4}$ of the area of the unit circle. These changes cancel out and our estimate remains $4\hat{p}$ with the standard error unchanged.

```
# Generate n iid samples from the square with lower
# left corner at the origin.
rsquare2 = function(n, width=1){
  x = runif(2*n,0, width)
  dim(x) = c(n,2)
  x
}
```

We can use the modified function above to verify that the interval is unchanged.

```
## Estimate pi using the proportion of points in the unit circle
## multiplied by four as the square contains only one fourth
## of the circle.
n = 5e5
x = rsquare2(n, 1)
p_hat = mean(x[,1]^2 + x[,2]^2 <= 1)
est = 4*p_hat
se = 4*sqrt(p_hat*{1-p_hat}/n)
z_a = qnorm(.975)
est_ci = est + c(-1, 1)*z_a*se
cover = ifelse(est_ci[1] < pi && est_ci[2] > pi, 'did', 'did not')
```

Our new estimate of 3.141 (3.136, 3.145) did cover the true value of the constant π.