STAT 305 – Solution: Introduction to Statistical Inference   Winter 2007

Lab Assignment # 2: The Central Limit Theorem and Simulations in R

Question 1.

a)
X is a discreet random variable and hence the mean for X

$$\mu = \sum_{i=1}^{6} x_i p_i = \sum_{i=1}^{6} (i)(\frac{1}{6}) = (\frac{1}{6}) \sum_{i=1}^{6} (i) = 3.5$$

and the variance

$$\sigma^2 = \sum_{i=1}^{6} (x_i - \mu)^2 p_i = \frac{1}{6} \sum_{i=1}^{6} (i - 3.5)^2 = 3.5$$

By the central limit theorem, for large n $\bar{X} \sim N(\mu, \sigma^2/n)$. Hence,

$$\bar{X} \ \sim \ N(3.5, 0.117)$$

The standard deviation is $sd = \sqrt{0.117} = 0.34$

We can plot this in R

```
> x=seq(2.5,4.5,0.001)
>plot(x,dnorm(x,mean=3.5,sd=0.34),type='l',main='The distribution of the sample mean us
```

The plot is given in the figure 1.

b) and c)
We need to get samples for $\bar{X}$. One sample from $\bar{X}$ means sampling from X 30 times and computing the mean. We choose to get 1000 samples for $\bar{X}$. The code in R is given in the following:

```
>barX=rep(0,1000)
>for (i in 1:1000){barX[i]=mean(sample(1:6,30,replace=T))}
>mean(barX)
>var(barX)
>hist(barX,freq=NULL)
```

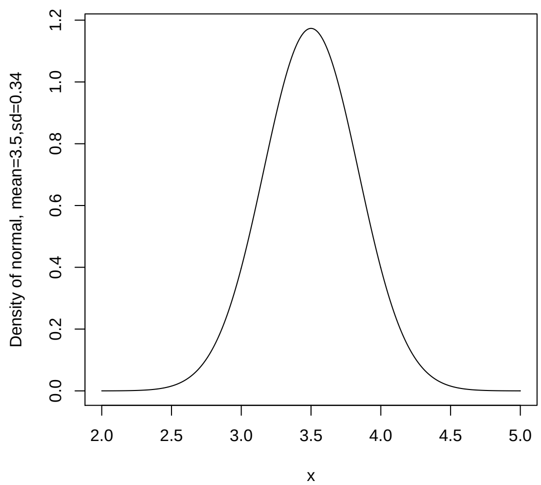

Figure 1: The distribution of the sample mean using CLT.

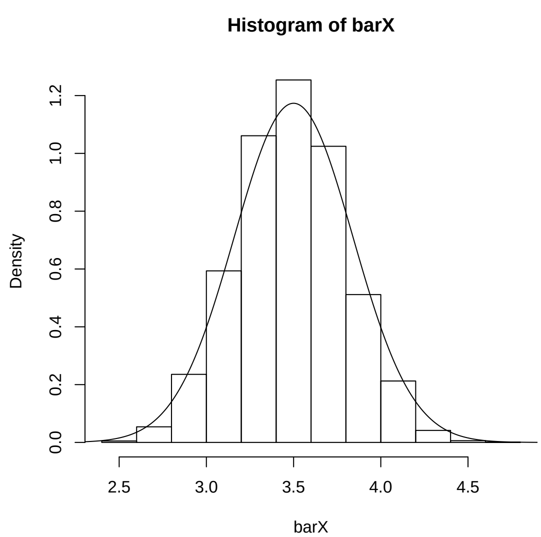

Figure 2: The distribution of $\bar{X}$ using CLT and the histogram using simulations

We get mean=3.501 and var=0.103 using R. The plot obtained in (a) and the histogram obtained in (b) using simulations is given in the figure 2.
Question 2.

a)
We want to find the probability using different number of iterations. Lets change the number from 1 to 10000. We can change this if necessary after observing the results. We create a vector p to record the computed probability for each iteration. Hence, p should have a length of 10000. If $1 \leq i \leq 10000$ is the number of iterations, we compute the probability by sampling from the uniform distribution for both X and Y, i times. Then, we count how many times $X + Y$ exceeds 2. The probability will be the number of counts divided by i, the number of iterations.

```
>p=seq(1,10000)
>for (i in 1:10000){X=runif(i,0,1)
Y=runif(i,0,1)
p[i]=sum(X+Y>2)/i
 }
>plot(1:10000,p)
```

b)

```
>p=seq(1,10000)
>for (i in 1:10000){X=runif(i,0,1)
Y=runif(i,0,1)
p[i]=sum(X+Y-5*X*sqrt(Y)>0)/i
 }
>plot(1:10000,p)
> mean(p[9900:10000])
```

```
[1] 0.25
```

Based on the figure, we observe after 9000 iterations, the estimated probability is almost fixed and hence we take the mean of p between 9900 and 10000 iterations as our final answer.

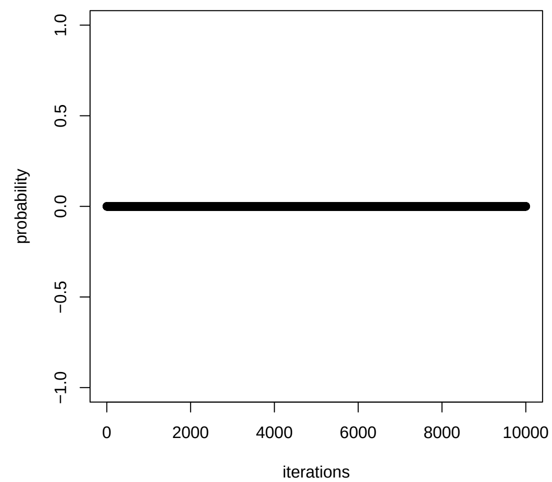

Figure 3: The probability computed using different number of iterations for part a)

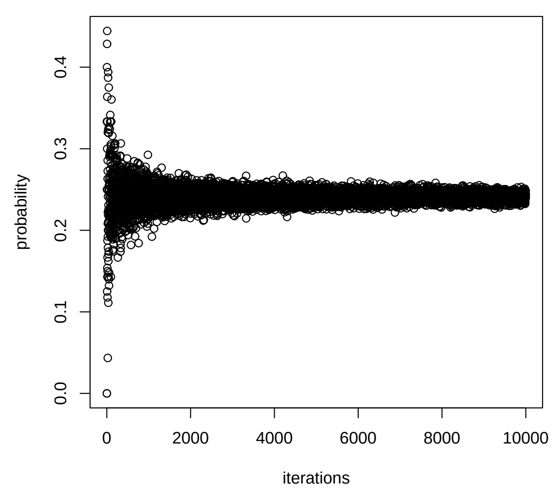

Figure 4: The probability computed using different number of iterations for part b)