Ishaan Taneja

Grading:
30 marks- Complete submission of Problem 1,5
70 marks- Problem 1

From Probability and Statistics with Examples Using R

1. The following result is a Berry-Eseen Type bound.

   Theorem: *Let $X_n \sim Binomial(n, p)$, then there exists $C > 0$ such that*

$$\sup_{x\in R}\left|P\left(\frac{X_n - np}{\sqrt{np(1-p)}} \leq x\right) - \int_{\infty}^{x} \frac{\exp(-\frac{y^2}{2})}{\sqrt{2\pi}}dt\right| \leq \frac{(p^2 + (1-p)^2)}{2\sqrt{np(1-p)}}$$

   We shall prove it by simulation by the below algorithm.

```
For x = -2,-1.9,-1,8,...0,...,1.9,2
    using inbuilt pnorm find z[x]:- pnorm(x)
Set p
For m = 1,50,100,150,...,1000
    For x = -2,-1.9,-1,8,...0,...,1.9,2
      1)Generate B: 1000 Samples of Binomial (m,p) using inbuilt rbinom function
        Compute SB:  (B-m*p)/((m*p(1-p))^(0.5))
      2)Compute y[x] : the proportion of samples  in SB less than equal to x
      3)Repeat steps 1) and 2) 100 times and compute average --  my[x] over each trial.
      4)Calculate   diff[m]=  max(abs(my[x]- z[x]))

For m = 1,50,100,150,...,1000
    Calculate  error(m)= [p^2+(1-p)^2]/[2*(m*p*(1-p))^0.5]

Plot diff and error.
```

   See if result is verified by picture. Can you do anything additional to verify the Theorem?

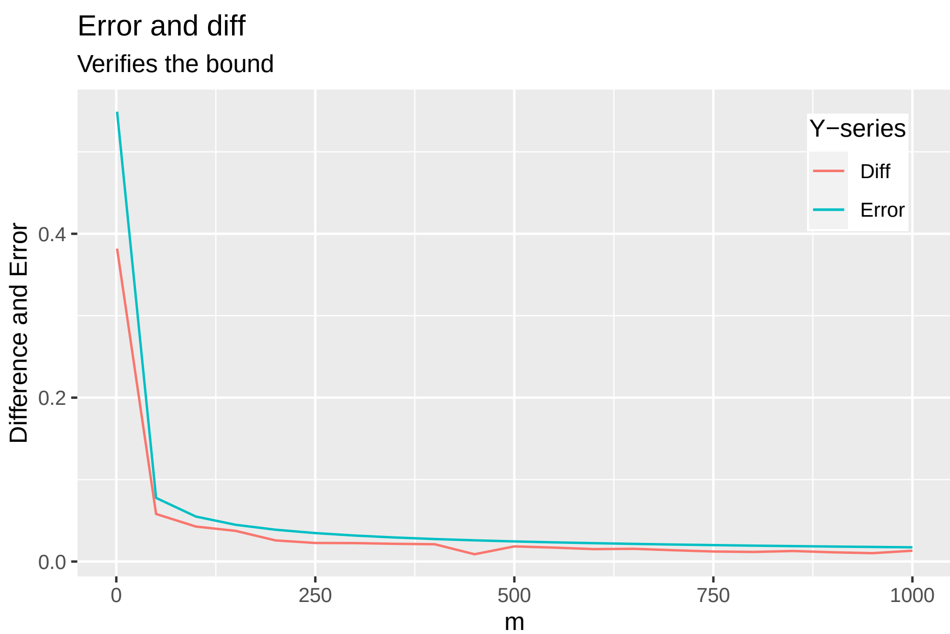
   Solution: 1

```
> x=seq(-2,2,0.1)
> z.x=pnorm(x)
> p=runif(1)
> p
```

```
[1] 0.6253851

> my.x=c()
> diff.m=c()
> l=c()
> m=c(1,seq.int(50,1000,50))
> itr=1:100
> for (i in m){
+   for (j in x){
+     prop=c()
+     for (k in itr){
+       c=0
+       B=rbinom(1000,i,p)
+       SB=(B-i*p)/((i*p*(1-p))^0.5)
+       for (s in SB){
+         if(s<= j){
+          c=c+1
+         }
+       }
+       prop=append(prop,c/1000)
+     }
+     my.x=append(my.x,mean(prop))
+
+   }
+ }
> d=abs(my.x-z.x)
> max=41
> r= seq_along(d)
> diff=split(d,ceiling(r/max))
> diff.m=c()
> for (i in 1:21){
+   diff.m=append(diff.m,max(diff[[i]]))
+ }
> error=c()
> for (i in m) {
+   error= append(error,(p^2+(1-p)^2)/(2*(i*p*(1-p))^0.5))
+ }
> library(ggplot2)
> data=data.frame(m,diff.m,error)
> C=ggplot(data, aes(x=m)) +
+   geom_line(aes(y = diff.m, color = "darkred")) +
+   geom_line(aes(y = error, color="steelblue"))+
+   labs(y='Difference and Error',title = 'Error and diff',
+   subtitle = 'Verifies the bound')+
+   labs(color = 'Y series')+ theme(
+     legend.position = c(.95, .95),
+     legend.justification = c("right", "top"),
```

```
+     legend.box.just = "right",
+     legend.margin = margin(1, 1, 1, 1)
+   )+ scale_color_discrete(name="Y-series",labels=c("Diff", "Error"))
> C
```



Additionally, we can increase the number of samples which will increase the accuracy of our result.

2. Ex 5.2.10 (c)

Solution: 2

Z $\sim$ Normal(μ, σ^2)

Let the median of Z be 'm'.

i.e.

$$P(Z < m) = \frac{1}{2}$$
$$\implies P(\frac{Z-\mu}{\sigma} < \frac{m-\mu}{\sigma}) = \frac{1}{2}$$
$$\implies P(X < \frac{m-\mu}{\sigma}) = \frac{1}{2}$$

where, $X \sim$ Normal(0,1)

Therefore,

$$\int_{-\infty}^{\frac{m-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \, dx = \frac{1}{2} \qquad ------(1)$$

Using R, we get:

```
> qnorm(1/2)

[1] 0
```

$$P(X < 0) = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \, \mathrm{d}x = \frac{1}{2} \qquad ------(2)$$

From (1) and (2):

$$\frac{m-\mu}{\sigma} = 0 \implies m = \mu$$

Hence, median of Z is μ.

3. Ex 5.2.11
   Solution: 3
   X $\sim$ Normal(μ, σ^2)

$$\begin{aligned} P(|X-\mu| < k\sigma) &= P(-k\sigma < X - \mu < k\sigma) \\ &= P(-k < \frac{X-\mu}{\sigma} < k) \\ &= P(-k < Z < k) \end{aligned}$$

   where, Z $\sim$ Normal(0,1)
   Therefore,

$$\begin{aligned} P(|X-\mu| < k\sigma) &= P(-k < Z < k) \\ P(|X-\mu| < k\sigma) &= \int_{-k}^{k} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz \end{aligned}$$

   which depends only on the value of k and is independent of μ and σ.
   Hence, $P(|X-\mu| < k\sigma)$ does not depend on the values of μ or σ.

4. Example 5.2.11
   Solution: 4
   X $\sim$ Normal(0,1)
   $E(X) = \mu = 0$ and $\sigma = 1$
   Therefore,

$$\begin{aligned} P(E(X)-\sigma < X < E(X)+\sigma) &= P(-1 < X < 1) \\ &= \int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \\ &= \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx - \int_{-\infty}^{-1} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \\ &= P(X \le 1) - P(X \le -1) \end{aligned}$$

   Using R, we get:

```
> pnorm(1)-pnorm(-1)
```

```
[1] 0.6826895
```

Therefore, $P(-1 \leq X \leq 1) \approx 0.68$
i.e. There is roughly 68% chance that a standardized normal random variable will produce a value within one standard deviation of expected value.

5. Example 5.2.12
Solution: 5
Let, Y be the random variable representing the weight (in grams) of cashews in the bag filled by machine.
Y $\sim$ Normal$(200, 4^2)$
The probability that a bag filled by machine will have fewer than 195 grams of cashew is given by:

$$P(Y < 195) = P\left(\frac{Y - 200}{4} < \frac{195 - 200}{4}\right)$$
$$= P(X < -1.25)$$

where, X $\sim$ Normal(0,1)
Using R, we get:

```
> pnorm(-1.25)

[1] 0.1056498
```

Therefore,

$$P(X < -1.25) = \int_{-\infty}^{-1.25} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \approx 0.106$$

Hence, there is slightly more than a 10% chance of a bag this light being produced by the machine.