STAT2371 Assignment 1

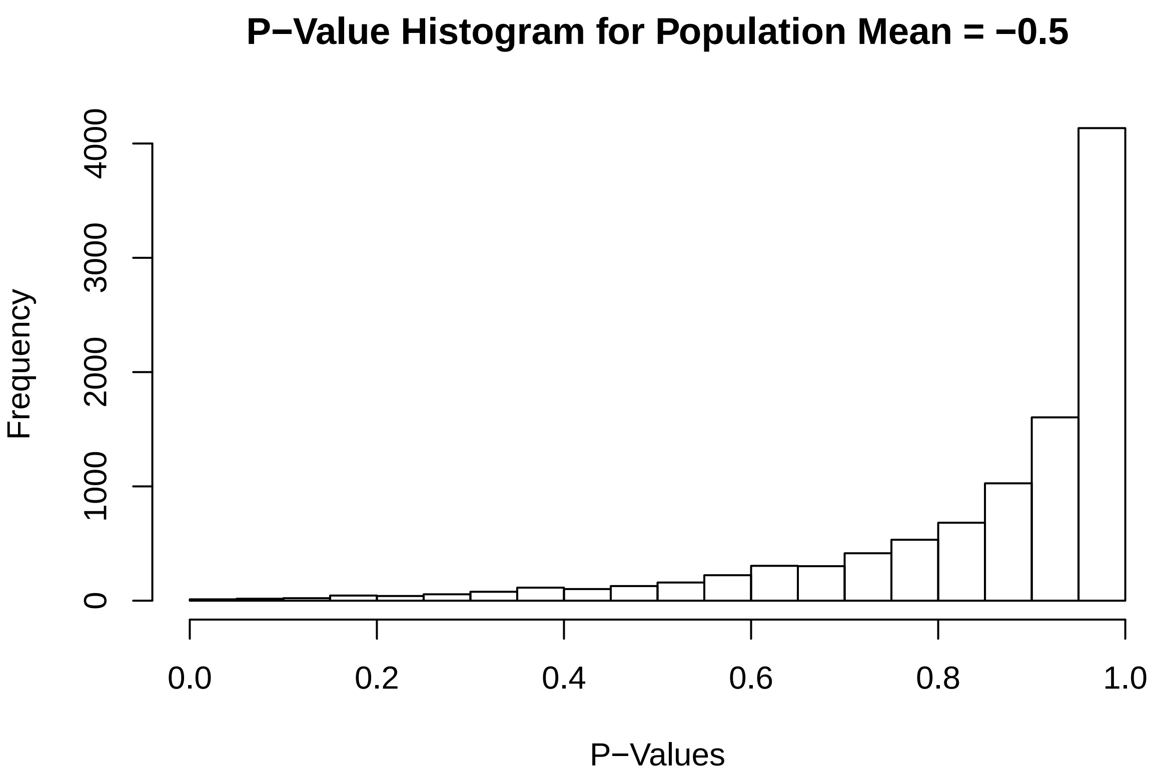
Guru Shyaam Shankar - Student ID: 46448608

13/08/2021

Question 1 Part A:

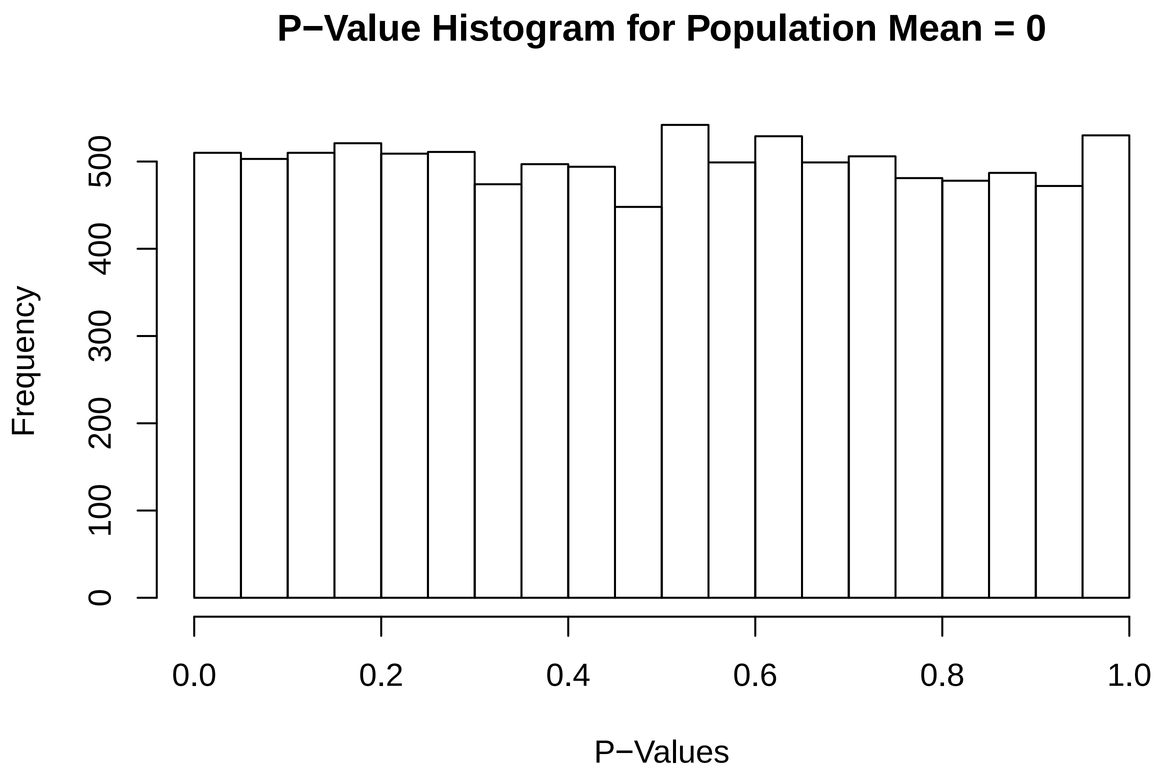
Scenario I: μ=-0.5

```
norm <- rnorm(10000, -0.5, 1)
my_pvalue_list<-0
my_t_list<-0
for (i in 1:10000){
n<- 10
my_sample<-sample(norm, n)
ttest<-t.test(my_sample, alternative =c('greater'))
my_t<- ttest$statistic
my_pvalue<-ttest$p.value
my_t_list[i]<-my_t
my_pvalue_list[i]<-my_pvalue
}
hist(my_pvalue_list, main = "P-Value Histogram for Population Mean = -0.5",
     xlab="P-Values",
     breaks= 20)
```



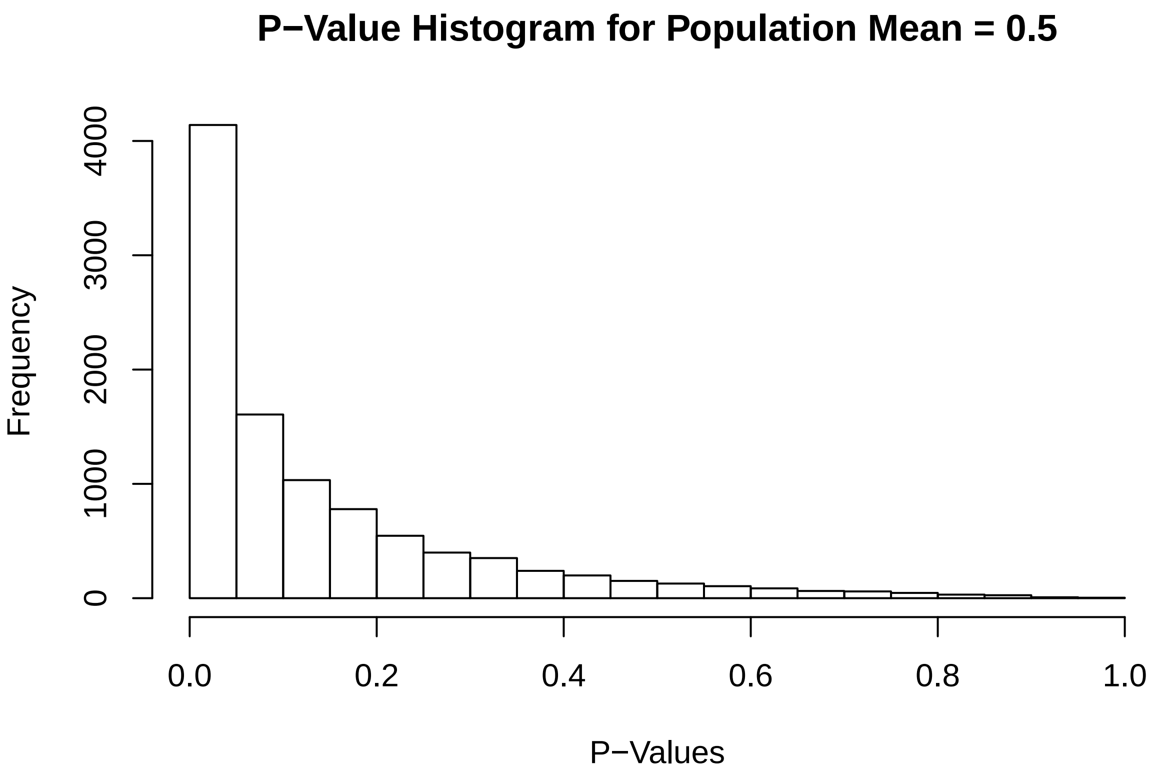
Scenario II: μ=0

```
norm <- rnorm(10000, 0, 1)
my_pvalue_list<-0
my_t_list<-0
for (i in 1:10000){
n<- 10
my_sample<-sample(norm, n)
ttest<-t.test(my_sample, alternative =c('greater'))
my_t<- ttest$statistic
my_pvalue<-ttest$p.value
my_t_list[i]<-my_t
my_pvalue_list[i]<-my_pvalue
}
hist(my_pvalue_list, main = "P-Value Histogram for Population Mean = 0",
     xlab="P-Values",
     breaks= 20)
```



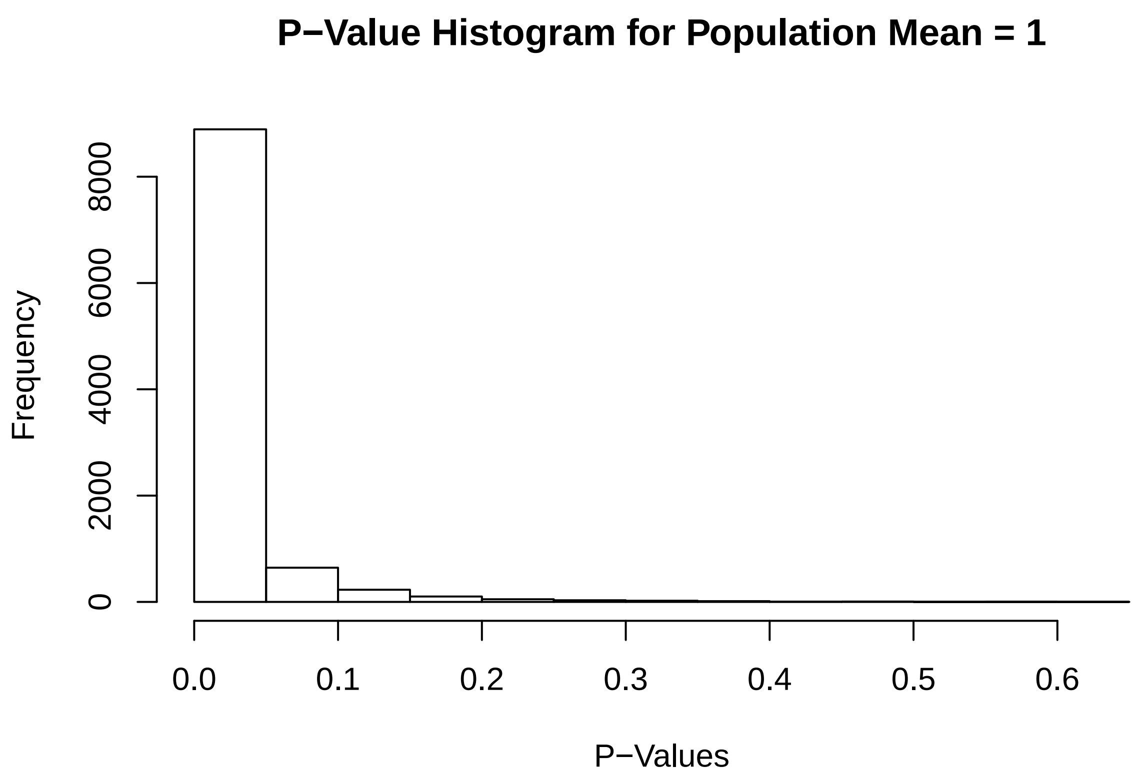
Scenario III: μ=0.5

```r
norm <- rnorm(10000, 0.5, 1)
my_pvalue_list<-0
my_t_list<-0
for (i in 1:10000){
n<- 10
my_sample<-sample(norm, n)
ttest<-t.test(my_sample, alternative =c('greater'))
my_t<- ttest$statistic
my_pvalue<-ttest$p.value
my_t_list[i]<-my_t
my_pvalue_list[i]<-my_pvalue
}
hist(my_pvalue_list, main = "P-Value Histogram for Population Mean = 0.5",
     xlab="P-Values",
     breaks= 20)
```



Scenario IV: μ=1

```
norm <- rnorm(10000, 1, 1)
my_pvalue_list<-0
my_t_list<-0
for (i in 1:10000){
n<- 10
my_sample<-sample(norm, n)
ttest<-t.test(my_sample, alternative =c('greater'))
my_t<- ttest$statistic
my_pvalue<-ttest$p.value
my_t_list[i]<-my_t
my_pvalue_list[i]<-my_pvalue
}
hist(my_pvalue_list, main = "P-Value Histogram for Population Mean = 1",
     xlab="P-Values",
     breaks= 20)
```



Question 1 Part 1 of Part B:

In the event that $\mu < 0$ we can observe a left skewed histogram, and the events that μ>0 we observe a right skewed histogram. However, when μ=0, the P-Value Histogram looks flat and uniformly distributed over [0,1]. Thus we can conclude, in Scenario I we observe a left skewed histogram, Scenario II an uniformly distributed histogram, and in Scenarios III and IV, a right skewed histogram.

Question 1 Part 2 of Part B:

The chances of observing a P-Value $< \alpha = 0.05$, are small in Scenario I, large in Scenario III and Scenario IV, and much larger in Scenario II. This is because of the null hypothesis, H_0: μ=0 versus the alternate hypothesis H_A: μ >0. Since, we can observe that when $\mu = 0$, it is uniformly distributed across [0,1], we can see it will have the highest chance of P-Value $< \alpha = 0.05$. In Scenario I, since the sample mean, μ=-0.5, is less than the null Hypothesis mean, H_0: μ=0, the chances are minuscule. In Scenarios III and IV, since the sample means μ=0.5 and μ=1 are equivalent to the alternate hypothesis, H_A: μ >0, we can observe a large chance. Thus, we can conclude that the chances of observing a P-Value $< \alpha = 0.05$ are definitely dependent on μ.

Guru Shyaam Shankar - 46448608

2)

$U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$

$$f_X(x) = \begin{cases} 1 & x \in [\theta - \frac{1}{2}, \theta + \frac{1}{2}] \\ 0 & \text{elsewhere} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < \theta - \frac{1}{2} \\ x - \theta + \frac{1}{2} & x \in [\theta - \frac{1}{2}, \theta + \frac{1}{2}] \\ 1 & x > \theta + \frac{1}{2} \end{cases}$$

$\therefore E(X) = \theta \qquad Var(X) = \frac{1}{12}$

Since X_1, X_2, X_3 are iid from the same distribution:
$E(X) = E(X_1) = E(X_2) = E(X_3)$

a)

$$\hat{\theta}_1 = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + X_3}{3}$$

$$\therefore E(\hat{\theta}_1) = E\left(\frac{X_1 + X_2 + X_3}{3}\right) = \frac{1}{3}(E(X_1 + X_2 + X_3)) = \frac{1}{3}(\theta + \theta + \theta) = \theta$$

$$E(\hat{\theta}_2) = E(X_2) = \theta$$

$$E(\hat{\theta}_3) = E\left(\frac{X_1 + X_3}{2}\right) = \frac{1}{2} E(X_1 + X_3) = \frac{1}{2}(\theta + \theta) = \theta$$

$\therefore \hat{\theta}_1, \hat{\theta}_2$ and $\hat{\theta}_3$ are all unbiased estimators of θ, because $E(\hat{\theta}_1) = E(\hat{\theta}_2) = E(\hat{\theta}_3) = \theta$.

2)

b)

$$\mathrm{var}(\hat{\theta}_1) = \mathrm{var}(\bar{X})$$
$$= \mathrm{var}\left(\tfrac{1}{3}(X_1 + X_2 + X_3)\right)$$
$$= \frac{1}{3^2}\left(3 \times \frac{1}{12}\right)$$
$$= \frac{1}{36}$$

$$\mathrm{var}(\hat{\theta}_2) = \mathrm{var}(X_2)$$
$$= \frac{1}{12}$$

$$\mathrm{var}(\hat{\theta}_3) = \mathrm{var}\left(\frac{X_1 + X_3}{2}\right)$$
$$= \frac{1}{2^2}\left(2 \times \frac{1}{12}\right)$$
$$= \frac{1}{24}$$

c)

We say that the most efficient unbiased estimator, is the one with the lowest variance.

Because, $\mathrm{var}(\hat{\theta}_1) < \mathrm{var}(\hat{\theta}_3) < \mathrm{var}(\hat{\theta}_2)$

We say that $\hat{\theta}_1$ is the most efficient estimator, and that is the estimator we prefer.

Estimators in order of preference: $\hat{\theta}_1, \hat{\theta}_3, \hat{\theta}_2$

Guru Shyaam Shankar - 46448608

3)

a)

$$f_X(x) = \binom{m}{x} p^x (1-p)^{m-x} \quad E(X) = mp \quad Var(X) = mp(1-p)$$

The joint pf of X_i:

$$f_X(x;p) = \prod_{i=1}^{n} \binom{m}{x_i} p^{x_i} (1-p)^{m-x_i} = \left(\prod_{i=1}^{n} \binom{m}{x_i}\right) p^{\sum_{i=1}^{n} x_i} (1-p)^{mn - \sum_{i=1}^{n} x_i}$$

The likelihood function:

$$L(p;X) = \left(\prod_{i=1}^{n} \binom{m}{X_i}\right) p^{\sum_{i=1}^{n} X_i} (1-p)^{mn - \sum_{i=1}^{n} X_i}$$

The log-likelihood function:

$$l(p;X) = \sum_{i=1}^{n} \log\binom{m}{X_i} + \sum_{i=1}^{n} X_i \log(p) + \left(mn - \sum_{i=1}^{n} X_i\right) \log(1-p)$$

b)

$$f_X(x;p) = \left(\prod_{i=1}^{n} \binom{m}{x_i}\right) p^{\sum_{i=1}^{n} x_i} (1-p)^{mn - \sum_{i=1}^{n} x_i}$$

$$= \left(\prod_{i=1}^{n} \binom{mn-1}{x_i - 1} \frac{m!(mn - x_i - 2)!}{x_i(mn-1)!(m-x_i)!}\right) p^{\sum_{i=1}^{n} x_i} (1-p)^{mn - \sum_{i=1}^{n} x_i}$$

$$= \left(\prod_{i=1}^{n} \frac{m!(mn - x_i - 2)!}{x_i(mn-1)!(m-x_i)!}\right)\left(\prod_{i=1}^{n} \binom{mn-1}{x_i - 1}\right) p^{\sum_{i=1}^{n} x_i} (1-p)^{mn - \sum_{i=1}^{n} x_i}$$

$$= h(X)\, g(s(X);p)$$

Where $s(X) = \sum_{i=1}^{n} X_i$,

Therefore, $\sum_{i=1}^{n} X_i$ is a sufficient statistic for p, by Factorisation Lemma as $g(s(X);p)$ is a negative binomial and $h(X)$ is independent of p.

Where: $h(X) = \left(\prod_{i=1}^{n} \frac{m!(mn - x_i - 2)!}{x_i(mn-1)!(m-x_i)!}\right)$

Guru Shyaam Shankar - 46448608

3)

c)

The MLE of p, is when $\frac{\partial l}{\partial p} = 0$

$$l(p; x) = \sum_{i=1}^{n} \log\binom{m}{x_i} + \sum_{i=1}^{n} X_i \log(p) + (mn - \sum_{i=1}^{n} X_i)\log(1-p)$$

$$\therefore \frac{\partial l}{\partial p} = 0 + \frac{1}{p}\sum_{i=1}^{n} X_i - \frac{1}{1-p}(mn - \sum_{i=1}^{n} X_i)$$

$$= \frac{1}{p}\sum_{i=1}^{n} X_i - \frac{1}{1-p}(mn - \sum_{i=1}^{n} X_i)$$

Let $\frac{\partial l}{\partial p} = 0$

$$\therefore \frac{1}{p}\sum_{i=1}^{n} X_i - \frac{1}{1-p}(mn - \sum_{i=1}^{n} X_i) = 0$$

$$\frac{1}{p}\sum_{i=1}^{n} X_i = \frac{mn - \sum_{i=1}^{n} X_i}{1-p}$$

$$mnp - p\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} X_i - p\sum_{i=1}^{n} X_i$$

$$\therefore mnp = \sum_{i=1}^{n} X_i$$

$$\therefore \text{MLE of } p : p = \hat{p} = \frac{1}{mn}\sum_{i=1}^{n} X_i$$

d)

Since, the MLE of p is $\frac{1}{mn}\sum_{i=1}^{n} X_i$,

$\therefore$, the MLE of $\log(p)$ is $\log\left(\frac{1}{mn}\sum_{i=1}^{n} X_i\right)$

e)

$$\frac{\partial l}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} X_i - \frac{mn - \sum_{i=1}^{n} X_i}{1-p}$$

$$\therefore \frac{\partial^2 l}{\partial p^2} = -\frac{1}{p^2}\sum_{i=1}^{n} X_i - \frac{1}{(1-p)^2}\left(mn - \sum_{i=1}^{n} X_i\right)$$

Since, $E(X) = mp$, $E\left(\sum_{i=1}^{n} X_i\right) = mnp$

$$E\left(\frac{\partial l}{\partial p}\right) = E\left(\frac{1}{p}\sum_{i=1}^{n} X_i - \frac{mn - \sum_{i=1}^{n} X_i}{1-p}\right)$$

$$= \frac{1}{p}(nmp) - \frac{mn - mnp}{1-p}$$

$$= mn - \frac{mn(1-p)}{1-p} = mn - mn = 0$$

$$\therefore I = -E\left(\frac{\partial^2 l}{\partial p^2}\right) = E\left(\frac{1}{p^2}\sum_{i=1}^{n} X_i + \frac{1}{(1-p)^2}\left(mn - \sum_{i=1}^{n} X_i\right)\right)$$

$$= \frac{1}{p^2}(mnp) + \frac{1}{(1-p)^2}(mn - mnp)$$

$$= \frac{mn}{p} + \frac{mn}{1-p} = \frac{mn}{p(1-p)}$$

$\therefore$ CRLB for the variance of unbiased estimators is:

$$\frac{p(1-p)}{mn}\{\tau'(p)\}^2$$

where $\tau(p)$ is an unbiased estimator

3)

f)

The CRLB holds if & only if $T = \tau(p)$ is a linear function of $U = \frac{\partial}{\partial p} l(p; x)$. Since $E(\sum_{i=1}^{\hat{n}} X_i) = mnp$, $\tau(p) = E(T)$, because it is unbiased, is also a linear function of p, only linear functions of p have unbiased estimators that attain the CRLB

In the case $\tau(p) = p(1-p)$, since it isn't linear of p, it will not attain the CRLB.

g)

The asymptotic properties of the MLE $\hat{p}$, we can see that:

$P(\sqrt{I}(\hat{p}-p) \leq x) \longrightarrow \Phi(x)$ where Φ is standard normal cdf.

$\therefore$ The Confidence Interval for MLE, $\hat{p} = \frac{1}{mn}\sum_{i=1}^{\hat{n}} X_i$, at 95% is:

$$= \overline{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n} X_i - 1.96\frac{\sigma}{\sqrt{n}}, \frac{1}{n}\sum_{i=1}^{n} X_i + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n} X_i - 1.96\frac{mp(1-p)}{\sqrt{n}}, \frac{1}{n}\sum_{i=1}^{n} X_i + 1.96\frac{mp(1-p)}{\sqrt{n}}\right)$$