Assignment 11 Solutions
Asymptotics of the LRT and Posterior Distributions, and Wald and Score Tests
(April 24, 2005)

1. (*More Estimation in a Zero-Inflated Poisson Model*)
Consider the situation of Problem 1 of Assignment 10, and in particular, inference based on the $n = 20$ observations Vardeman simulated from the distribution.

   (a) Find a large sample 90% joint confidence region for (p, λ) based on the loglikelihood function (based on inverting *LRT's*). Plot this in the (p, λ)-plane and to the extent possible, compare it to the elliptical region you found in Assignment 10.

   (b) Note that for a fixed value of λ,
$$p = \frac{\frac{n_0}{n} - 1}{\exp(-\lambda) - 1}$$
maximizes the likelihood. Use this fact to find and plot the profile loglikelihood for λ. Use this plot and make an approximate *90% confidence interval* for λ. How does this interval compare to the one you found in part e) of Problem 1 from Assignment 10?

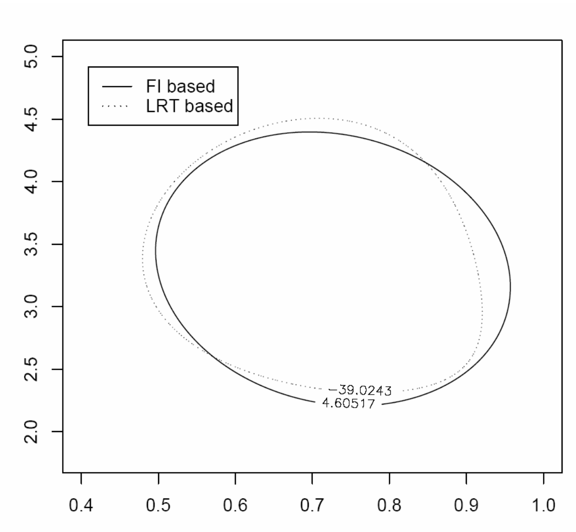
Solution:

(a)

Continue with problem 1 in homework 10, we have
$$l_n(x) = 6\ln\left(pe^{-\lambda} + 1 - p\right) + 14\ln p - 14\lambda + 48\ln\lambda$$
Then, the large sample confidence set of (p, λ) is
$$\left\{(p, \lambda) \,|\, \ln(x) \geq \underset{(p,\lambda)}{Sup} \ln(p, \lambda) - \tfrac{1}{2}\chi_2^2\right\}$$
$$= \left\{(p, \lambda) \,|\, \ln(x) \geq \ln\left(\hat{p}_{MLE}, \hat{\lambda}_{MLE}\right) - \tfrac{1}{2}\chi_2^2\right\}$$
Accomplish this in R:

```
p=.72675
lam=3.30242
data=c(0,3,2,5,3,4,
0,0,4,0,0,5,
3,5,5,4,2,0,
1,2)
m=sum(data==0)
n=length(data)
lhood=function(x){
p=x[1]
lam=x[2]
6*log(p*exp(-lam)+(1-p))+14*log(p)-14*lam+48*log(lam)-sum(log(factorial(data)))
}
```

```
th1=seq(0.4,1,length=100)
th2=seq(1.8,5,length=100)
th=expand.grid(th1,th2)
sum=apply(th,1,lhood)
sum=matrix(sum,100,100)
p=.72675
lam=3.30242
contour(th1,th2,sum,levels=(lhood(c(p,lam))-0.5*qchisq(0.9,2)),lty=3,add=T)
legend(locator(1),legend=c("FI based","LRT based"),lty=c(1,3))
```



(b)

The profile loglikelihood for λ is

$$\left\{\lambda | \ln^*(\lambda) \geq \underset{\lambda}{Sup} \ln^*(\lambda) - \frac{1}{2}\chi_1^2\right\}$$

$$= \left\{\lambda | \ln\left(p = \frac{\frac{n_0}{n}-1}{e^{-\lambda}-1}, \lambda\right) \geq \ln\left(\hat{p}_{MLE}, \hat{\lambda}_{MLE}\right) - \frac{1}{2}\chi_1^2\right\},$$

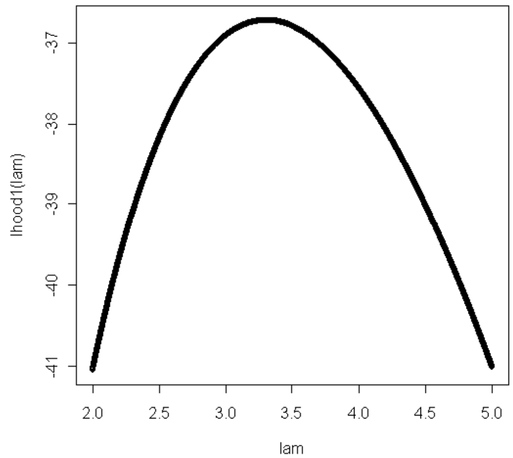
where $n_0 = m$.

Accomplish this in R:

```
> lhood1=function(lam){
+ p=(m/n-1)/(exp(-lam)-1)
+ 6*log(p*exp(-lam)+(1-p))+14*log(p)-14*lam+48*log(lam)-sum(log(factorial(data)))
+ }
> lam=seq(2,5,length=10000)
>plot(lam,lhood1(lam),mail="Plot of Profile log-likelihood")
> lb=min(lam[lhood1(lam)>=(lhood(c(.72675,3.30242))-0.5*qchisq(.9,1))])
> ub=max(lam[lhood1(lam)>=(lhood(c(.72675,3.30242))-0.5*qchisq(.9,1))])
> c(lb,ub)
[1] 2.526253 4.204920
```

The 90% confidence interval for λ is (**2.526253, 4.204920**)

And the plot of the profile log-likelihood is



We see that the 90% confidence interval obtained in homework 10 1.(e) is symmetric about $\hat{\lambda}$, while here the 90% confidence interval got by profile log-likelihood is not necessarily symmetric about $\hat{\lambda}$.

2. Consider again the model of Problem 2 of Assignment 10. Below are $n = 20$ observations that Vardeman simulated from this model using $\alpha = .3$.

$$1.36, 1.35, 0.78, 1.85, 2.32, 0.55, 1.07, -0.57, -0.38, 0.25,$$
$$-0.36, 1.71, 1.40, 0.46, 3.16, -0.78, 0.69, -0.03, 1.26, 0.44$$

   (a) Plot the *loglikelihood* for this sample. What, approximately, is the maximum likelihood estimate for α?

   (b) If you wished to test the hypothesis $H_0 : \alpha = .4$ with Type I error probability .1, what would be your decision here? Carefully explain. (Use a *likelihood ratio test*).

   (c) Use the plot from *a)* and make an approximate *90% confidence interval* for α based on the likelihood function (based on inverting *LRT's*). How does this interval compare to the one you made in part *f)* of Problem 2 on Assignment 10?

Solution:

(a)

As $f(x|\alpha) = \alpha f_1(x) + (1-\alpha) f_0(x)$ and $f_0(x) \sim N(0,1)$, $f_1(x) \sim N(1,1)$, then the likelihood is $L\left(\alpha|\underset{\sim}{x}\right) = \prod_{i=1}^{20} f(x_i|\alpha)$, and the loglikelihood is

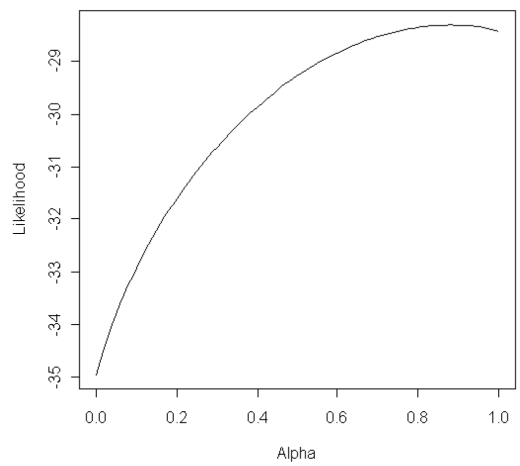
$\log L\left(\alpha|\underset{\sim}{x}\right) = \sum_{i=1}^{20} \log f(x_i|\alpha) = \sum_{i=1}^{20} \log\{\alpha f_1(x_i) + (1-\alpha) f_0(x_i)\};$

Program in R:

```
> data=c(1.36,1.35,0.78,1.85,2.32,0.55,1.07,-0.57,-0.38,0.25,-0.36,1.71,1.40,0.46,3.16,-0.78,0.69,-0.03,1.26,0.44)
> lhood=function(a){
+ sum(log((1-a)*dnorm(data,mean=0,sd=1)+a*dnorm(data,mean=1,sd=1)))
+ }
> a=seq(0,1,.01)
> a=as.matrix(a)
> b=apply(a,1,lhood)
> a[b==max(b)] ####mle
[1] 0.88
> plot(a,b,type="l",xlab="Alpha",ylab="Likelihood")
```

The maximum likelihood estimate is $\hat{\alpha}_{MLE} = 0.88$.

The plot of loglikelihood is



(b)

```
> chi=2*(lhood(0.88)-lhood(0.4))
> 1-pchisq(chi,1)#####p-value
[1] 0.07823232
```

With Type I error probability .1, we need to reject the null hypothesis.

(c)

```
> alpha=seq(0,1,.001)
> alpha=as.matrix(alpha)
> lll=apply(alpha,1,lhood)
> lb=min(alpha[lll>=(lhood(0.88)-0.5*qchisq(.9,1))])
> ub=max(alpha[lll>=(lhood(0.88)-0.5*qchisq(.9,1))])
> c(lb,ub) #####confidence intervals
[1] 0.431 1.000
```

The approximate 90% confidence interval for α based on the likelihood funciton is

$(\mathbf{0.431},\ \mathbf{1.000})$.
By following the similar way as 1.(f) in homework 10, we can get another 90% confidence interval as
$\hat{\alpha}_n \pm Z(0.95)\frac{1}{\sqrt{nI(\hat{\alpha}_n)}}$
where n=20, $\hat{\alpha}_n = 0.88$, and $I(\hat{\alpha}_n) = \frac{1}{1.1}$ which can be read from the Figure 1.
Then the 90% confidence interval is:
$\left(0.88 \pm 1.645\sqrt{\frac{1.1}{20}}\right) = (0.4942133, 1.265787)$.

3. Problems 1.2.4 and 1.2.5 of *B&D.*

(Problem 1.2.4)

Let $X_1, ..., X_n$ be distributed as

$$p(x_1, ..., x_n|\theta) = \frac{1}{\theta^n}$$

where $x_1, ..., x_n$ are natural numbers between 1 and θ and $\Theta = \{1, 2, 3, ...\}$.

(a) Suppose θ has prior frequency,

$$\pi(j) = \frac{c(a)}{j^a}, j = 1, 2, ...,$$

where $a > 1$ and $c(a) = \left|\sum_{j=1}^{\infty} j^{-a}\right|^{-1}$.Show that

$$\pi(j|x_1, ..., x_n) = \frac{c(n+a, m)}{j^{n+a}}, j = m, m+1, ...$$

where $m = \max(x_1, ..., x_n), c(b, t) = \left|\sum_{j=t}^{\infty} j^{-b}\right|^{-1}, b > 1$.

(b) Suppose that $\max(x_1, ..., x_n) = x_1 = m$ for all n. Show that $\pi(m|x_1, ..., x_n) \to 1$ as $n \to \infty$ whatever be a. Interpret this result.

Solution:

(a) As $P(x_1, ..., x_n|\theta = j) = \frac{1}{j^n} I(x(n) \le j)$ and $\pi(j) = \frac{c(a)}{j^a}, c(a) = \left(\sum_{j=1}^{\infty} j^{-a}\right)^{-1}$ then

$$\begin{aligned}
\pi(m|x_1, ..., x_n) &= \frac{P(x_1, ..., x_n|\theta = j)\pi(j)}{\sum_{j=1}^{\infty} P(x_1, ..., x_n|\theta = j)\pi(j)} = \frac{\frac{1}{j^n} I(x(n) \le j)\frac{c(a)}{j^a}}{\sum_{j=1}^{\infty}\frac{1}{j^n} I(x(n) \le j)\frac{c(a)}{j^a}} \\
&= \frac{\frac{c(a)}{j^{n+a}} I(x(n) \le j)}{\sum_{j=1}^{\infty}\frac{c(a)}{j^{n+a}} I(x(n) \le j)} = \frac{\frac{c(a)}{j^{n+a}} I(x(n) \le j)}{\sum_{j=x(n)}^{\infty}\frac{c(a)}{j^{n+a}}} \\
&= \frac{\frac{1}{j^{n+a}} I(x(n) \le j)}{\sum_{j=x(n)}^{\infty}\frac{1}{j^{n+a}}} = \frac{c(n+a, m)}{j^{n+a}}, \quad j = m, m+1, ...
\end{aligned}$$

where $m = \max(x_1, ..., x_n)$.

(b)

$$\pi\left(m|x_1,...,x_n\right) = \frac{c\left(n+a,m\right)}{m^{n+a}} = \frac{1/m^{n+a}}{\sum_{j=m}^{\infty}\frac{1}{j^{n+a}}}$$
$$= \frac{1}{\sum_{j=0}^{\infty}\left(\frac{m}{j+m}\right)^{n+a}} = \frac{1}{1+\sum_{j=1}^{\infty}\left(\frac{m}{j+m}\right)^{n+a}}$$

As

$$\left(\frac{m}{j+m}\right)^{n+a} = \frac{1}{\left(1+\frac{j}{m}\right)^{n+a}}$$

and

$$\sum_{j=1}^{\infty}\left(\frac{m}{m+j}\right)^{n+a} \sim m\int_0^{\infty}\frac{1}{\left(1+x\right)^{n+a}}dx = \frac{m}{n+a-1} \to 0 \text{ as } n\to\infty$$

so

$$\sum_{j=0}^{\infty}\left(\frac{m}{m+j}\right)^{n+a} \to 0 \text{ as } n\to\infty$$

$$\Rightarrow \pi\left(m|x_1,...,x_n\right) \to 1 \text{ as } n\to\infty$$

(Problem 1.2.5)

In example *1.2.1* suppose n is large and $\frac{1}{n}\sum_{i=1}^{n}x_i = \bar{x}$ is not close to 0 or 1 and the prior distribution is beta, $\beta\left(r,s\right)$. Justify the following approximation to the posterior distribution

$$P\left[\theta\le t|X_1=x_1,...,X_n=x_n\right] \approx \Phi\left(\frac{t-\tilde{\mu}}{\tilde{\sigma}}\right)$$

where Φ is the standard normal distribution function and $\mu = \frac{n}{n+r+s}\bar{x}+\frac{r}{n+r+s}$, $\tilde{\sigma}^2 = \frac{\tilde{\mu}(1-\tilde{\mu})}{n+r+s}$

Solution:

Since

$$f(\theta|X) \propto f(X|\theta)\pi(\theta)$$
$$\propto (1-\theta)^{n-n\overline{x}}\theta^{n\overline{x}}(1-\theta)^{s-1}\theta^{r-1}$$
$$= \theta^{n\overline{x}+r-1}(1-\theta)^{n-n\overline{x}+s-1}$$

i.e. $\theta|X \sim Beta\left(n\bar{x}+r, n-n\bar{x}+s\right)$.
Consider the distribution of $\theta|\frac{X-\tilde{\mu}}{\tilde{\sigma}}$,where $\widetilde{\mu} = \frac{n\overline{x}+r}{n+r+s}$ and $\widetilde{\sigma}^2 = \frac{\widetilde{\mu}(1-\widetilde{\mu})}{n+r+s+1}$
Denote $\theta|X \sim Beta(a,b)$, where $a = n\overline{x}+r$ and $b = n-n\overline{x}+s$

Known that for $V_1, ..., V_a$ i.d.d exp(1) and $W_1, ...W_b$ i.d.d exp(1), $\frac{\sum_1^a V_i}{\sum_1^a V_i + \sum_1^b W_i} \sim Beta(a, b)$. Then

$$\begin{aligned}
\theta|\frac{X-\tilde{\mu}}{\tilde{\sigma}} &= \frac{\frac{\sum_1^a V_i}{\sum_1^a V_i+\sum_1^b W_i} - \frac{a}{a+b}}{\sqrt{\frac{ab}{(a+b)^2(a+b+1)}}} \\
&= \frac{b\sum_1^a V_i - a\sum_1^b W_i}{(a+b)\left(\sum_1^a V_i + \sum_1^b W_i\right)\sqrt{\frac{ab}{(a+b)^2(a+b+1)}}} \\
&= \frac{\frac{b\sum_1^a V_i - a\sum_1^b W_i}{(a+b)}}{\frac{\sum_1^a V_i+\sum_1^b W_i}{(a+b)}\sqrt{\frac{ab}{(a+b+1)}}} \qquad (*)
\end{aligned}$$

Let $Y = \frac{b\sum_1^a V_i - a\sum_1^b W_i}{(a+b)}$ and $E(Y) = 0, Var(Y) = \frac{ab}{a+b}$
Then (*) equals

$$(*) = \frac{1}{\frac{(\sum_1^a V_i+\sum_1^b W_i)}{(a+b)}} \frac{Y}{\sqrt{\frac{ab}{a+b}}} \frac{\sqrt{\frac{ab}{a+b}}}{\sqrt{\frac{ab}{(a+b+1)}}}$$

By Central Limit Theorem, $\frac{Y}{\sqrt{\frac{ab}{a+b}}} \to N(0, 1)$; and by *Weak Law* of Large Number, $\frac{\sum_1^a V_i+\sum_1^b W_i}{(a+b)} \to 1$ and also $\frac{\sqrt{\frac{ab}{(a+b)}}}{\sqrt{\frac{ab}{(a+b+1)}}} \to 1$, So, by *Slutskey's Theorem*, $\theta|\frac{X-\tilde{\mu}}{\tilde{\sigma}} \to N(0, 1)$, i.e. $\theta|X \to N(\tilde{\mu}, \tilde{\sigma}^2)$. Thus,

$$P(\theta \leq 0|X_1 = x_1, ..., X_n = x_n) \approx \Phi(\frac{t - \widetilde{\mu}}{\widetilde{\sigma}})$$

where Φ is the standard normal distribution function.

4. Consider Bayesian inference for the binomial parameter p. In particular, for sake of convenience, consider the Uniform $(0, 1)$ ($Beta(\alpha, \beta)$ for $\alpha = \beta = 1$) prior distribution.

   (a) It is possible to argue from reasonably elementary principles that in this binomial context, where $\Theta = (0, 1)$, the Beta posteriors have a consistency property. That is, simple arguments can be used to show that for any fixed p_0 and any $\varepsilon > 0$, for $X_n\tilde{}Binomial(n, p_0)$, the random variable

$$Y_n = \int_{p_0-\varepsilon}^{p_0+\varepsilon} \frac{1}{B(\alpha + X_n, \beta + (n - X_n))} p^{\alpha+X_n-1}(1-p)^{\beta+(n-X_n)-1}\, dp$$

   (which is the posterior probability assigned to the interval $(p_0 - \varepsilon, p_0 + \varepsilon)$.) converges in p_0 probability to 1 as $n \to \infty$. This part of the problem is meant to lead you through this argument. Let $\varepsilon > 0$ and $\delta > 0$.

(i) Argue that there exists m such that if $n \geq m$, $\left|\frac{x_n}{n} - \frac{\alpha+x_n}{\alpha+\beta+n}\right| < \frac{\varepsilon}{3}$, $\forall x_n = 0, 1, ..., n$.

(ii) Note that the posterior variance is $\frac{(\alpha+x_n)(\beta+n-x_n)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}$. Argue there is an m' such that if $n \geq m'$ the probability that the posterior assigns to $\left(\frac{\alpha+x_n}{\alpha+\beta+n} - \frac{\varepsilon}{3}, \frac{\alpha+x_n}{\alpha+\beta+n} + \frac{\varepsilon}{3}\right)$ is at least $1-\delta$ $\forall x_n = 0, 1, ..., n$.

(iii) Argue that there is an m'' such that if $n \geq m''$ the p_0 probability that $\left|\frac{X_n}{n} - p_0\right| < \frac{\varepsilon}{3}$ is at least $1-\delta$.
Then note that if $n \geq \max(m, m', m'')$ i) and ii) together imply that the posterior probability assigned to $\left(\frac{x_n}{n} - \frac{2\varepsilon}{3}, \frac{x_n}{n} + \frac{2\varepsilon}{3}\right)$is at least $1-\delta$ for any realization x_n. Then provided $\left|\frac{x_n}{n} - p_0\right| < \frac{\varepsilon}{3}$ the posterior probability assigned to $(p_0 - \varepsilon, p_0 + \varepsilon)$ is also at least $1-\delta$. But iii) says this happens with p_0 probability at least $1-\delta$. That is, for large n, with p_0 probability at least $1-\delta, Y_n \geq 1-\delta$. Since δ is arbitrary, (and $Y_n \leq 1$) we have the convergence of Y_n to 1 in p_0 probability.

(b) *Vardeman* intends to argue in class that posterior densities for large n tend to look normal (with means and variances related to the likelihood material). The posteriors in this binomial problem are Beta$(\alpha + x_n, \beta + (n - x_n))$ (and we can think of $X_n \sim \text{Bi}(n, p_0)$ as derived as the sum of n *iid* Bernoulli(p_0) variables). So we ought to expect *Beta* distributions for large parameter values to look roughly normal. To illustrate this do the following. For $\rho = .3$ (for example ... any other vlaue would do as well), consider the Beta$(\alpha + n\rho, \beta + n(1-\rho))$ (posterior) distributions for $n = 10, 20, 40$ and 100. For $p_n \sim \text{Beta}(\alpha + n\rho, \beta + n(1-\rho))$ plot the probability densities for the variables

$$\sqrt{\frac{n}{\rho(1-\rho)}}(p_n - \rho)$$

on a single set of axes along with the standard normal density. Note that if W has pdf $f(\cdot)$, then $aW+b$ has pdf $g(\cdot) = \frac{1}{a} f\left(\frac{\cdot - b}{a}\right)$. (Your plots are translated and rescaled posterior densities of p based on possible observed values $x_n = .3n$.) If this is any help in doing this plotting, *Vardeman* tried to calculate values of the *Beta* function using *MathCad* and got the following: $(B(4,8))^{-1} = 1.32 \times 10^3$, $(B(7,15))^{-1} = 8.14 \times 10^5$, $(B(13,29))^{-1} = 2.291 \times 10^{11}$ and $(B(31,71))^{-1} = 2.967 \times 10^{27}$.

Solution: (a)

i)

$$\begin{aligned}
& \left|\frac{x_n}{n} - \frac{\alpha + x_n}{\alpha + \beta + n}\right| \\
= & \left|x_n\left(\frac{1}{n} - \frac{1}{\alpha+\beta+n}\right) - \frac{\alpha}{\alpha+\beta+n}\right| \\
\leq & |x_n| \frac{\alpha+\beta}{n(\alpha+\beta+n)} + \frac{\alpha}{\alpha+\beta+n} \qquad (x_n \in \{0, 1, ..., n\}) \\
\leq & \frac{\alpha+\beta}{\alpha+\beta+n} + \frac{\alpha}{\alpha+\beta+n}
\end{aligned}$$

$\forall \varepsilon, \delta > 0, \exists m$, s.t. $\forall n \geq m$.

$$\left| \frac{x_n}{n} - \frac{\alpha + x_n}{\alpha + \beta + n} \right| \leq \frac{2\alpha + \beta}{\alpha + \beta + n} < \frac{\varepsilon}{3}$$

ii) Let $\hat{P}_B = E\left(p|x\right)$, by *Chebechev*, for $\varepsilon \& \delta$ fixed above, $\exists m'$, s.t. $\forall n \geq m'$,

$$\begin{aligned} P\left(|\hat{p}_n - \hat{p}_B| \geq \frac{\varepsilon}{3}\right) &\leq \frac{9Var\left(\hat{p}_n\right)}{\varepsilon^2} = \frac{9\left(\alpha + x_n\right)\left(\beta + n - x_n\right)}{\left(\alpha + \beta + n\right)^2 \left(\alpha + \beta + n + 1\right)\varepsilon^2} \\ &\leq \frac{9}{4\left(\alpha + \beta + n + 1\right)\varepsilon^2} < \frac{\delta}{3} \end{aligned}$$

$\Rightarrow P\left(\hat{p}_n \in \left(\hat{p}_B - \frac{\varepsilon}{3}, \hat{p}_B + \frac{\varepsilon}{3}\right)\right) \geq 1 - \frac{\delta}{3}$

iii) By *WLLN*, $\frac{x_n}{n} \overset{P_{p_0}}{\to} p_0$, for $\varepsilon \& \delta$ fixed above, $\exists m$",s.t $\forall n \geq m$", $\Rightarrow P_0\left(\left|\frac{x_n}{n} - p_0\right| \geq \frac{\varepsilon}{3}\right) < \frac{\delta}{3}$.
Then, by i), ii), & iii), if $n \geq \max\left(m, m', m''\right)$

$$\begin{aligned} & P_0\left(|\hat{p}_n - p_0| \geq \varepsilon\right) \\ \leq\ & P\left(|\hat{p}_n - \hat{p}_B| \geq \frac{\varepsilon}{3}\right) + P\left(\left|\hat{p}_B - \frac{x_n}{n}\right| \geq \frac{\varepsilon}{3}\right) + P\left(\left|p_0 - \frac{x_n}{n}\right| \geq \frac{\varepsilon}{3}\right) \\ <\ & \frac{\delta}{3} + \frac{\delta}{3} + \frac{\delta}{3} = \delta \end{aligned}$$

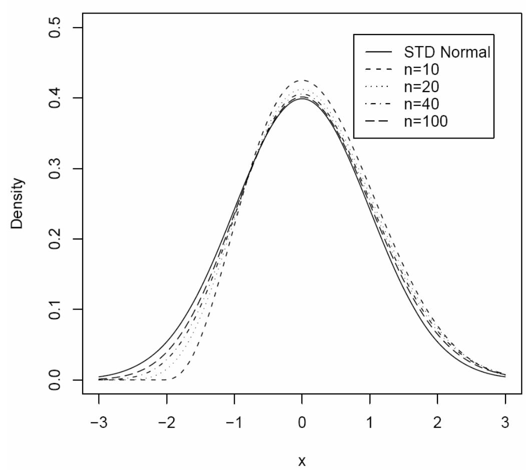
$$\Rightarrow P\left(\hat{p}_n \in \left(p_0 - \varepsilon, p_0 + \varepsilon\right)\right) \geq 1 - \delta$$

b)

Coding in R as:

```
alpha=1
beta=1
rho=0.3
n=100
a=sqrt(n/(rho*(1-rho)))
b=a*rho
x=seq(-3,3,length=100)
xx=((x+b)/a)
dens=dbeta(xx,(alpha+0.3*n),(beta+0.7*n))/a
plot(x,dens,,ylim=c(0,0.5),type="l",xlab="x",ylab="Density",lty=5)
points(x,dnorm(x),type="l",lty=1)
n=10
a=sqrt(n/(rho*(1-rho)))
b=a*rho
x=seq(-3,3,length=100)
xx=((x+b)/a)
```

```
dens=dbeta(xx,(alpha+0.3*n),(beta+0.7*n))/a
points(x,dens,type="l",lty=2)
n=20
a=sqrt(n/(rho*(1-rho)))
b=a*rho
x=seq(-3,3,length=100)
xx=((x+b)/a)
dens=dbeta(xx,(alpha+0.3*n),(beta+0.7*n))/a
points(x,dens,type="l",lty=3)
n=40
a=sqrt(n/(rho*(1-rho)))
b=a*rho
x=seq(-3,3,length=100)
xx=((x+b)/a)
dens=dbeta(xx,(alpha+0.3*n),(beta+0.7*n))/a
points(x,dens,type="l",lty=4)
legend(locator(1),legend=c("STD Normal","n=10","n=20","n=40","n=100"),lty=(c(1:5)))
```



5. Suppose that X, Y and Z are independent *binomial* variables, $X \sim bin(n, p_1), Y \sim bin(n, p_2)$ and $Z \sim bin(n, p_3)$. For the parameter space (for (p_1, p_2, p_3)) $\Theta = [0, 1]^3$, we will consider testing H_0:$p_1 = p_2 = p_3$ based on (X, Y, Z).

   (a) Find the general forms of the likelihood ratio tests, the Wald tests and the score tests of this hypothesis.

   (b) Use the fact that the parameter space here is basically 3-dimensional while Θ_0 is basically 1-dimensional so that there are 2 independent constraints involved 2 and the limiting χ^2 distributions of the test statistics thus have $\nu = 2$ associated degrees of freedom to actually carry out these tests with $\alpha \approx .05$ if $X = 33$, $Y = 53$ and $Z = 59$, all based on $n = 100$.

Solution:
$X \sim Bin(n, p_1)\,,\; Y \sim Bin(n, p_2)\,,\; Z \sim Bin(n, p_3)$.
Let $\theta = (p_1, p_2, p_3)'$, then $\hat{p}_{mle} = \left(\frac{x}{n}, \frac{y}{n}, \frac{z}{n}\right)$ over all $\theta \in \Theta$.
And also the restricted $\hat{p}_{mler} = \frac{x+y+z}{3n}$ under $H_0 : p_1 = p_2 = p_3$.

(a)

- *Likelihood Ratio Test (LRT):*

$$\begin{aligned}
\lambda_n(x) &= \frac{\underset{(p_1,p_2,p_3)}{Sup} \begin{pmatrix} n \\ x \end{pmatrix} p_1^x (1-p_1)^{n-x} \cdot \begin{pmatrix} n \\ y \end{pmatrix} p_2^y (1-p_2)^{n-y} \cdot \begin{pmatrix} n \\ z \end{pmatrix} p_3^z (1-p_3)^{n-z}}{\underset{p_1=p_2=p_3=p}{Sup} \begin{pmatrix} n \\ x \end{pmatrix}\begin{pmatrix} n \\ y \end{pmatrix}\begin{pmatrix} n \\ z \end{pmatrix} p^{x+y+z} (1-p)^{3n-(x+y+z)}} \\
&= \frac{\left(\frac{x}{n}\right)^x \left(1-\frac{x}{n}\right)^{n-x} \cdot \left(\frac{y}{n}\right)^y \left(1-\frac{y}{n}\right)^{n-y} \cdot \left(\frac{z}{n}\right)^z \left(1-\frac{z}{n}\right)^{n-z}}{\left(\frac{x+y+z}{3n}\right)^{x+y+z} \cdot \left(1-\frac{x+y+z}{3n}\right)^{3n-(x+y+z)}} \\
&= \frac{(3n)^{3n} \cdot x^x (n-x)^{n-x} \cdot y^y (n-y)^{n-y} \cdot (z)^z (n-z)^{n-z}}{n^{3n} \cdot (x+y+z)^{x+y+z} \cdot (3n-(x+y+z))^{3n-(x+y+z)}}
\end{aligned}$$

By $2\ln\lambda_n \xrightarrow{L_\theta} X_2^2$, we can do *LRT*, and reject $H_0 : \theta = \theta_0$ when $2\log\lambda_n \geq X_2^2(1-\alpha)$

- *Wald Test*

$$g(\theta) = \begin{pmatrix} p_1 - p_2 \\ p_1 - p_3 \end{pmatrix}, \quad G(\theta) = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$$

$$\begin{aligned}
\log L(\theta) &= x\ln p_1 + (n-x)\log(1-p_1) \\
&\quad + y\ln p_2 + (n-y)\log(1-p_2) + z\ln p_3 + (n-z)\log(1-z) + const
\end{aligned}$$

$$\begin{aligned}
I_1(\theta) &= -E\left(\frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'}\right) \\
&= \begin{pmatrix} \frac{1}{p_1(1-p_1)} & 0 & 0 \\ 0 & \frac{1}{p_2(1-p_2)} & 0 \\ 0 & 0 & \frac{1}{p_3(1-p_3)} \end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
& B\left(\hat{\theta}_n\right)^{-1} \\
= \; & G(\theta) I_1^{-1}(\theta) G(\theta)^{-1} \\
= \; & \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} p_1(1-p_1) & 0 & 0 \\ 0 & p_2(1-p_2) & 0 \\ 0 & 0 & p_3(1-p_3) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} \\
= \; & \begin{pmatrix} p_1(1-p_1) & -p_2(1-p_2) & 0 \\ p_1(1-p_1) & 0 & -p_3(1-p_3) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} \\
= \; & \begin{pmatrix} p_1(1-p_1) + p_2(1-p_2) & p_1(1-p_1) \\ p_1(1-p_1) & p_1(1-p_1) + p_3(1-p_3) \end{pmatrix}
\end{aligned}
$$

Then, the wald statistic is

$W_n \hat{=} n g\left(\hat{\theta}_n\right)' B\left(\hat{\theta}_n\right)^{-1} g\left(\hat{\theta}_n\right)$

$= n \begin{pmatrix} \frac{x-y}{n} & \frac{x-z}{n} \end{pmatrix} \begin{pmatrix} p_1(1-p_1) + p_2(1-p_2) & p_1(1-p_1) \\ p_1(1-p_1) & p_1(1-p_1) + p_3(1-p_3) \end{pmatrix}^{-1} \begin{pmatrix} \frac{x-y}{n} \\ \frac{x-z}{n} \end{pmatrix}$

By $W_n \underrightarrow{\mathcal{L}_\theta} \chi_2^2$ under H_0, we can do *Wald Test*, and reject $H_0 : \theta = \theta_0$ when $W_n \geq X_2^2(1-\alpha)$.

- *Score Test*

  $R_n^* = \begin{pmatrix} \frac{x}{p_1} - \frac{n-x}{1-p_1} & \frac{y}{p_2} - \frac{n-y}{1-p_2} & \frac{z}{p_3} - \frac{n-z}{1-p_3} \end{pmatrix} |_{p_1=p_2=p_3=\hat{p}_{mler}}$

  $\cdot I_1\left(\hat{p}_{mler}\right)^{-1} \cdot \begin{pmatrix} \frac{x}{p_1} - \frac{n-x}{1-p_1} \\ \frac{y}{p_2} - \frac{n-y}{1-p_2} \\ \frac{z}{p_3} - \frac{n-z}{1-p_3} \end{pmatrix} |_{p_1=p_2=p_3=\hat{p}_{mler}}$

  Then, by $R_n^* \underrightarrow{\mathcal{L}_\theta} \chi_2^2$ under H_0, we can do *Score Test*, and reject $H_0 : \theta = \theta_0$ when $R_n^* \geq X_2^2(1-\alpha)$.

(b)

```
> x=33
> y=53
> z=59
> n=100
> pn1=x/n
> pn2=y/n
> pn3=z/n
> in1=matrix(c(1/(pn1*(1-pn1)),0,0,0,1/(pn2*(1-pn2)),0,0,0,1/(pn3*(1-pn3))),nrow=3,byrow=T)
> g=c(pn1-pn2,pn1-pn3)
> G=matrix(c(1,-1,0,1,0,-1),nrow=2,byrow=T)
> ##Likehood Ratio Test
> loglamn=x*log(x)+(n-x)*log(n-x)+y*log(y)+(n-y)*log(n-y)+z*log(z)+(n-z)*log(n-z)
+3*n*log(3)-(x+y+z)*log(x+y+z)-(3*n-x-y-z)*log(3*n-x-y-z)
> lrt=2*loglamn
> lrt
```

```
[1] 15.07826
> qchisq(0.95,2)
[1] 5.991465
> 1-pchisq(rn,2) ##p-value
[1,] 0
> ##Wald Test
> bn=G%*%solve(in1)%*%t(G)
> wn=n*t(g)%*%solve(bn)%*%g
> wn
[1,] 16.17791
> 1-pchisq(wn,2)
[1,] 0
> ##Score test
> pnl=(x+y+z)/(3*n)
> in1l=matrix(c(1/(pnl*(1-pnl)),0,0,0,1/(pnl*(1-pnl)),0,0,0,1/(pnl*(1-pnl))),nrow=3,byrow=T)
> dll=c(x/pnl-(n-x)/(1-pnl),y/pnl-(n-y)/(1-pnl),z/pnl-(n-z)/(1-pnl))
> rn=t(dll)%*%solve(in1l)%*%dll/n
> rn
[,1]
[1,] 14.84316
> 1-pchisq(rn,2)
[,1]
[1,] 0.0005982035
```

From above results, we will **reject** the null hypothesis either using Likelihood Ratio test or Wald test or Score test.