HOMEWORK 5

problem 1

(a) Given iid sample $\{X_i\}_{i=1}^{n}$ from the poisson, find MLE $\hat{\lambda}$.

$$L(X;\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{(x_i)!} e^{-\lambda}$$

$$\ell = \log(L(X;\lambda)) = \log\left(\prod_{i=1}^{n} \frac{\lambda^{x_i}}{(x_i)!} e^{-\lambda}\right) = -n\lambda - \sum_{i=1}^{n} \ln(x_i!) + \ln\lambda \sum_{i=1}^{n} x_i$$

$$\frac{d\ell}{d\lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0$$

$$\therefore \quad n = \frac{1}{\lambda}\sum_{i=1}^{n} x_i$$

$$\boxed{\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

$$\frac{d^2\ell}{d\lambda^2} = \sum_{i=1}^{n} x_i \cdot -\frac{1}{\lambda^2}.$$

(b) Fisher information $var(X_i) = \lambda$

$$E\left(-\frac{\partial^2}{\partial\lambda^2} \log f(X|\lambda)\right) = \left.\frac{n\lambda}{\lambda^2}\right|_{\lambda=\lambda_0} = \frac{n}{\lambda_0}$$

(c) asymptotic variance of MLE $\hat{\lambda}$.

$$\frac{1}{nI(\theta_0)} = \frac{1}{n\left(\frac{n}{\lambda_0}\right)} = \frac{\lambda_0}{n^2}$$

(d) show that $\tilde{\lambda}$ is a consistent estimator of λ_0

if $\frac{Y}{n} \to \theta$ then, continuous mapping theorem would state $f(Y/n) \to f(\theta)$.

$f = -\log(\lambda)$

$$Y = \sum_{i=1}^{n} \underbrace{1\{x_i = 0\}}_{\text{Bernoulli}(p_0)}$$

$\frac{Y}{n} = \bar{x}_n \longrightarrow p_0$ by the law of large number (LLN)

$$-\log\left(\frac{Y}{n}\right) \to -\log(p_0) = -\log(e^{-\lambda_0}) = \lambda_0$$

since $\tilde{\lambda} \to \lambda_0$, we have shown that $\tilde{\lambda}$ is a consistent estimator of λ_0.

e) delta states $\sqrt{n}(g(\bar{x}) - g(\theta)) \to N(0, (g'\theta)^2 \sigma^2)$

CLT: $\sqrt{n}\left(\frac{Y}{n} - p_0\right) \to N(0, p_0(n-p_0))$

deltamethod: $\sqrt{n}\left(-\log\left(\frac{Y}{n}\right) - \lambda_0\right) \to N\left(0, \frac{1}{\lambda_0^2} p_0(1-p_0)\right)$

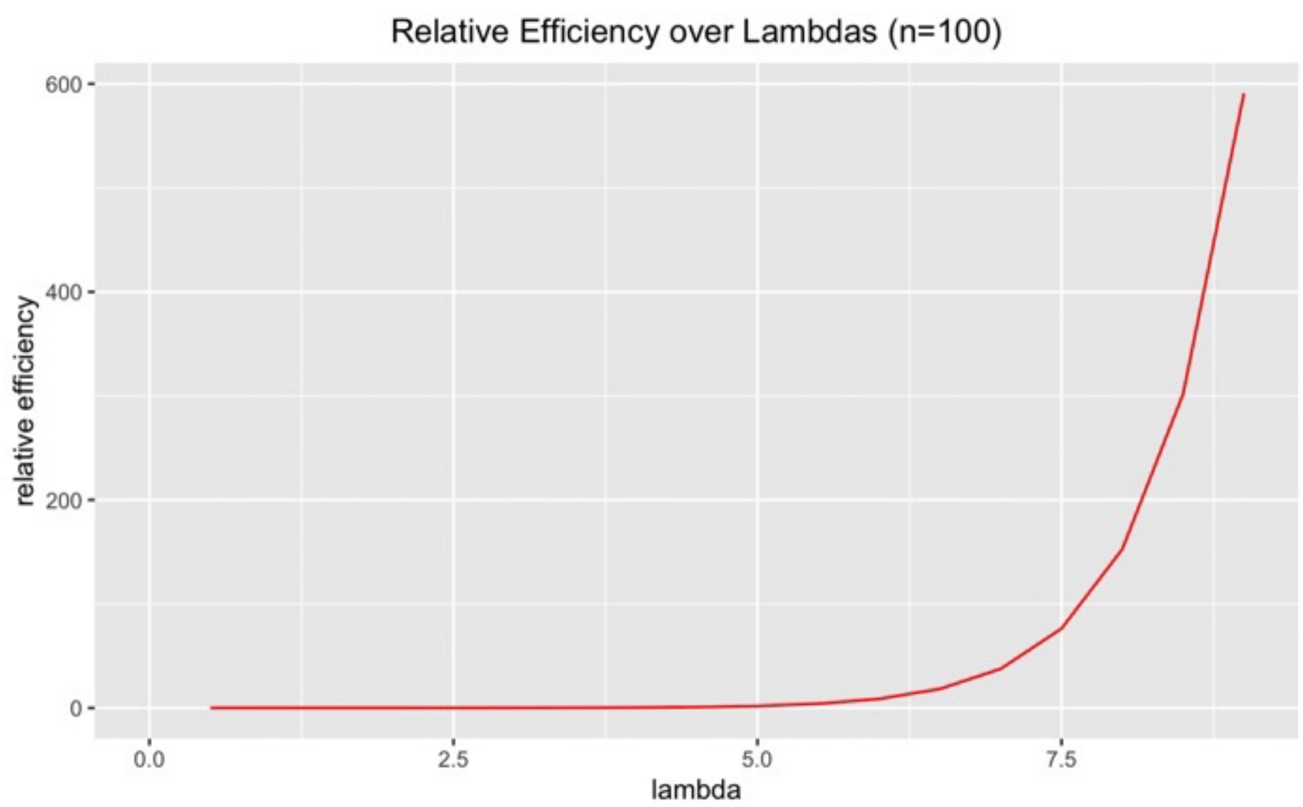
the variance of $\tilde{\lambda}$ is $\frac{1}{\lambda_0^2} p_0(1-p_0)$

f) $ARE(\tilde{\lambda}, \hat{\lambda}) = \dfrac{\text{asymptotic variance of MLE } \hat{\lambda}}{\text{asymptotic variance of } \tilde{\lambda}} = \dfrac{\frac{\lambda_0}{n^2}}{\frac{1}{\lambda_0^2} p_0(1-p_0)} = \dfrac{\lambda_0^3}{n^2 p_0(1-p_0)}$

when $n \to \infty$, the ARE $= 0$, suggesting that the variance of MLE $\hat{\lambda}$ gets relatively smaller than the variance of $\tilde{\lambda}$.

As $\lambda \to \infty$, ARE $\to \infty$ as well, suggesting that $\tilde{\lambda}$ is asymptotically more efficient than $\hat{\lambda}$.

(g) Using ggplot, plot the relative efficiency of $\hat{\lambda}$ versus $\tilde{\lambda}$ from the last part (on Y-axis) against different values of true λ_0 (on X-axis). What do you conclude? Does the true λ_0 affect your choice on which estimator to choose?




```{r}
#1f
rel_eff <- function(lambda) {
  p = exp(-lambda)
  n = 100
  return (lambda^3/(n^2*p*(1-p)))
}
lambdas <- seq(0,9,0.5)
y <- rel_eff(lambdas)
df <- data.frame(x=lambdas, y=y)
plot<- ggplot(df, aes(x=lambdas, y=y)) + geom_line(color='red') + labs(x="lambda", y="relative efficiency", title= 'Relative Efficiency over Lambdas (n=100)') + theme(plot.title = element_text(hjust = 0.5))
plot
```

problem 1. $f(x;\theta) = \frac{x}{\theta^2} e^{-x^2/2\theta^2}$

(a) method of moment

$$\mu_1 = E(X|\theta) = \int_0^\infty x f(x|\theta)\,dx = \int_0^\infty x \frac{x}{\theta^2} e^{-x^2/2\theta^2} dx = \frac{1}{\theta^2}\int_0^\infty x^2 e^{-x^2/2\theta^2} dx$$

$$= \frac{1}{\theta^2}\int_0^\infty v\, e^{-v/2\theta^2} \cdot \frac{dv}{2\sqrt{v}} \qquad v = x^2$$

$\Gamma(n+1) = n\Gamma(n)$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$

$$= \frac{1}{2\theta^2}\int_0^\infty v^{3/2-1} \cdot \exp\left(\frac{-v}{2\theta^2}\right) dv$$

$\Gamma(3/2)$

$$= \frac{1}{2\theta^2}\Gamma\left(\frac{3}{2}\right)(2\theta)^{3/2} = \sqrt{2}\theta\,\Gamma(3/2) = \sqrt{2}\theta\,\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \sqrt{2}\theta\cdot\frac{1}{2}\sqrt{\pi}$$

$$= \boxed{\frac{\sqrt{\pi}\,\theta}{\sqrt{2}}}$$

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}$$

$E(X) = \lambda\sqrt{\pi/2}$

$var(X) = \frac{4-\pi}{2}\lambda^2$

$$\therefore \theta\frac{\sqrt{\pi}}{\sqrt{2}} = \bar{X}$$

$$\therefore \boxed{\tilde{\theta} = \frac{\bar{X}\sqrt{2}}{\sqrt{\pi}}}$$

(b) $var(\tilde{\theta}) = var\left(\frac{\bar{X}\sqrt{2}}{\sqrt{\pi}}\right) = \frac{2}{\pi} var(\bar{x}) = \frac{2}{\pi}\frac{var(X)}{n} = \boxed{\frac{2\theta^2(2-\pi/2)}{\pi n}}$

$var(X) = E(X^2) - E(X)^2 = 2\theta^2 - \left(\sqrt{\frac{\pi}{2}}\,\theta\right)^2 = \theta^2(2-\pi/2)$

(c)

$$lik(\theta) = \prod_{i=1}^{n} \frac{x_i}{\theta^2} e^{-x_i^2/2\theta^2}$$

$$\ell(\theta) = \log[lik\,\theta] = \sum_{i=1}^{n}\log(x_i) - 2n\log(\theta) - \frac{1}{\theta^2}\sum_{i=1}^{n}\frac{x_i^2}{2}$$

$$\frac{d\,\ell(\theta)}{d\theta} \Rightarrow -2n\left(\frac{1}{\theta}\right) + 2\left(\frac{1}{\theta^3}\right)\sum_{i=1}^{n}\left(\frac{x_i^2}{2}\right) = 0$$

$$\frac{-2n\theta^2 + 2\sum\frac{x_i^2}{2}}{\theta^3} = 0$$

$$\therefore 2n\theta^2 = 2\sum\frac{x_i^2}{2}$$

$$\theta^2 = \frac{\sum\frac{x_i^2}{2}}{n}$$

$$\hat{\theta} = \sqrt{\sum_{i=1}^{n}\left(\frac{x_i^2}{2}\right)/n}$$

(d) $E(\hat{\theta}^2) = \theta_0^2$ if $\hat{\theta}^2$ is unbiased.

$$\frac{1}{n}E\left(\sum_{i=1}^{n} x^2/2\right) = \frac{1}{2n}E\left(\sum_{i=1}^{n} x_i^2\right) = \frac{1}{2n}\cdot n\,E(x_i^2) = \frac{1}{2n}\cdot n\left(var(x_i) + E(x)^2\right)$$

$$= \frac{1}{2}\left(\frac{4-\pi}{2}\theta^2 + \theta^2\cdot\pi/2\right) = \frac{\theta^2}{2}\cdot 2 = \theta_0^2$$

$\hat{\theta} = \sqrt{\frac{\sum_{i=1}^{n}(x_i^2/2)}{n}}$. let $g(x) = \sqrt{x}$ .. since g is strictly concave $E(g(x)) < g(E(x))$

(e) Therefore, $E(\hat{\theta}) = E\left(\sqrt{\frac{\sum(x_i^2/2)}{n}}\right) < \sqrt{E\left[\frac{\sum(x_i^2/2)}{n}\right]}$

since $E(\hat{\theta}) \neq \theta_0$, $\hat{\theta}$ is a biased estimator of θ_0.

f) $\frac{\partial^2 \log L}{\partial \theta^2} = \frac{d}{d\theta}\left(-2n\left(\frac{1}{\theta}\right) + 2\left(\frac{1}{\theta^3}\right)\sum_{i=1}^{n}\left(\frac{x_i^2}{2}\right)\right) = \frac{2n}{\theta^2} - 6\theta^{-4}\sum_{i=1}^{n}\frac{x_i^2}{2}$

$$= \frac{2n\theta^2 - 6\sum_{i=1}^{n}\frac{x_i^2}{2}}{\theta^4}$$

$$= \frac{2n\theta^2 - 3\sum_{i=1}^{n}x_i^2}{\theta^4}$$

$$-E\left(\frac{2n\theta^2 - 3\sum_{i=1}^{n}x_i^2}{\theta^4}\right) = -E\left(\frac{2n\theta^2 + 3\sum_{i=1}^{n}-x_i^2}{\theta^4}\right)$$

$$E(T(x)) = E(-x_i^2) = -2\theta^2$$

$$\therefore E\left(\frac{2n\theta^2 + 3\sum_{i=1}^{n}-x_i^2}{\theta^4}\right) = \frac{2n}{\theta^2} + \frac{3nE(T(x))}{\theta^4}$$

$$= \frac{2n}{\theta^2} + \frac{3n\cdot(-2)}{\theta^2}$$

$$= \frac{-4n}{\theta^2}$$

since E is linear and x_i are iid,

$$-E\left(\frac{2n\theta^2 + 3\sum_i x_i^2}{\theta^4}\right) = \boxed{\frac{4n}{\theta^2}}$$

g) the approximate variance of the MLE is $var(\hat{\theta}_{MLE}) \approx \frac{1}{nI(\theta)}$

$$\therefore var(\hat{\theta}_{MLE}) \approx \frac{\theta^2}{4n \cdot n}$$

The large sample distribution of a maximum likelihood estimate is approximately normal with mean θ_0 and variance $1/[nI(\theta_0)]$. Since this is merely a limiting result, which holds as the sample size tends to infinity, we say that the mle is **asymptotically unbiased** and refer to the variance of the limiting normal distribution as the **asymptotic variance of the mle.**

(h) asymptotic variance of MOME : $\boxed{\frac{2\theta^2(2-\pi/2)}{\pi n}}$

asymptotic variance of $\hat{\theta}_{MLE}$: $\frac{\theta^2}{4n^2}$

$$\frac{var(\hat{\theta}_{MLE})}{var(\tilde{\theta}_{MOM})} = \frac{\frac{\cancel{\theta^2}}{4\cancel{n^2}n}}{\frac{2\cancel{\theta^2}(2-\pi/2)}{\pi\cancel{n}}} = \frac{\pi}{8n(2-\pi/2)}$$

As n increases, $\frac{var(\hat{\theta}_{MLE})}{var(\tilde{\theta}_{MOM})} \to 0$, suggesting that the $var(\tilde{\theta}_{mom})$ grows big. $var(\hat{\theta}_{MLE})$ is more efficient for large n.

Twe θ_0 doesn't affect the choice on which estimator to choose but the sample size does affect.

(i) No, MLE is not necessarily efficient in finite samples. So, efficiency in infinite samples doesn't gaurantee efficiency in the finite sample.

problem 3 $f(x;\eta) = \frac{1}{4\eta}\exp\left(-\frac{|x|}{2\eta}\right)$

A random variable has a Laplace (μ, b) dist if the pdf is $f(x|\mu,b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$

therefore, x has Laplace $(0, 2\eta)$.

(a) MLE: $L = \prod_{i=1}^{n} \frac{1}{4\eta}\exp\left(-\frac{|x_i|}{2\eta}\right)$

$$\ell = -n\log(4\eta) + \sum_{i=1}^{n} -\frac{|x_i|}{2\eta} = -n\log(4\eta) - \frac{1}{2\eta}\sum_{i=1}^{n}|x_i|$$

$$\frac{d\ell}{d\eta} = -\frac{n}{4\eta} + \sum_{i=1}^{n}|x_i|\cdot(2\eta)^{-2}\cdot(2) = \frac{-n}{4\eta} + \frac{\sum_{i=1}^{n}|x_i|\cdot 2}{(2\eta)^2} = 0$$

$$\frac{-n\eta + \sum_{i=1}^{n}|x_i|\cdot 2}{4\eta^2} = 0 \Leftrightarrow -n\eta + \sum_{i=1}^{n}|x_i|\cdot 2 = 0 \Leftrightarrow \eta = \boxed{\frac{\sum_{i=1}^{n}|x_i|\cdot 2}{n}}$$

(b) If unbiased, $E(\hat{\eta}_{MLE}) = \eta$.

Each $X_i \sim$ Laplace$(0, 2\eta)$

$|X_i| \sim$ exponential$\left(\frac{1}{2\eta}\right)$

$E(|X|) = 2\eta$

$$E(\eta) = E\left(\frac{\sum_{i=1}^{n}|x_i|\cdot 2}{n}\right) = \frac{2}{n}\cdot\sum_{i=1}^{n}E|x_i| = \frac{2}{\cancel{n}}\cdot\cancel{n}\cdot 2\eta$$
$$= \boxed{4\eta}$$

since $E(\eta) \neq \eta$, it is biased.

X_n Laplace$(0, 2\eta)$

$|X| \sim \exp\left(\frac{1}{2\eta}\right)$

$E(|X|) = 2\eta$

(c) $$\text{var}(\eta) = \text{var}\left(\frac{\sum_{i=1}^{n}|x_i|\cdot 2}{n}\right) = \frac{4}{n^2}\sum_{i=1}^{n}\text{var}|x_i| = \frac{4}{n^2}(4\eta^2)n$$
$$= \boxed{\frac{16\eta^2}{n}}$$

var of $\exp(1/2\eta) = \frac{1}{\left(\frac{1}{2\eta}\right)^2} = 4\eta^2$

(d) Asymptotic variance of MLE.

$$\frac{d^2\ell}{d\eta^2} = \frac{d}{d\eta}\left(\frac{-n}{4\eta} + \frac{\sum_{i=1}^{n}|x_i|\cdot 2}{4\eta^2}\right) = \frac{n}{4\eta^2} - \frac{1}{2\eta^3}\sum_{i=1}^{n}2|x_i|$$

$$I(\theta_0) = -E\left(\frac{d^2\ell}{d\eta^2}\right) = -E\left(\frac{n}{4\eta^2} - \frac{1}{2\eta^3}\sum_{i=1}^{n}2|x_i|\right) = \frac{-n}{4\eta^2} + E\left[\frac{1}{2\eta^3}\sum_{i=1}^{n}2|x_i|\right]$$
$$= \frac{-n}{4\eta^2} + \frac{1}{2\eta^3}\sum_{i=1}^{n}2E|x_i|$$
$$= \frac{-n}{4\eta^2} + \frac{1}{2\eta^3}\cdot 2n\cdot 2\eta = \frac{-n}{4\eta^2} + \frac{4n}{2\eta^2} = \frac{-n+8\eta}{4\eta^2} = \frac{-n+8\eta}{4\eta^2}$$

asymptotic Var $= \dfrac{1}{n I(\theta_0)} = \dfrac{1}{n\left(\dfrac{-n+8\eta}{4\eta^2}\right)} = \dfrac{4\eta^2}{-n^2+8n\eta}$

e) $\dfrac{\text{finite sample variance}}{\text{asymptotic variance}} = \dfrac{16\eta^{2\,4}}{\cancel{n}} \cdot \dfrac{\overset{n}{-\cancel{n^2}+8n\eta}}{4\cancel{\eta^2}} = 4(8\eta - n)$

The ratio depends on n. ✗

f) is MLE an efficient estimator?

$\Rightarrow$ An unbiased estimator that achieves the cramer Rao lower bound on variance is an efficient estimator

$var(\hat{\theta}) = \dfrac{16\eta^2}{n}$ since $n \geq 0$, $var(\hat{\theta}) = \dfrac{16\eta^2}{n} \geq \dfrac{1}{n I(\theta_0)} = \dfrac{-4n^2}{-\eta^2+8n\eta}$

$\dfrac{1}{n I(\theta_0)} = \dfrac{-4n^2}{-n^2+8n\eta}$

Since the estimator achieves the cramer Rao lower bound, it is an efficient estimator.

Problem 4

$X \sim \text{binom}(n, p)$

a) MLE of p.

$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$l(p) = \ln\left(\binom{n}{x} p^x (1-p)^{n-x}\right)$$

$$= \ln\binom{n}{x} + x\ln(p) + (n-x)\ln(1-p)$$

$$\frac{\partial l(p)}{dp} = \frac{\partial}{\partial p}\left[\ln\binom{n}{x} + x\ln(p) + (n-x)\ln(1-p)\right]$$

$$= 0 + \frac{x}{p} - \frac{n-x}{1-p} = 0$$

when $\frac{x}{p} - \frac{n-x}{1-p} = 0$

$$\Leftrightarrow \frac{x(1-p) - p(n-x)}{1-p} = 0$$

$$\Leftrightarrow x(1-p) - p(n-x) = 0$$

$$x - \cancel{xp} - pn + \cancel{xp} = 0$$

$$x = pn$$

$$\boxed{\hat{p} = \frac{x}{n}}$$

(b) $E(\hat{p}) = E\left(\frac{x}{n}\right) = \frac{1}{n}E(x) = \frac{1}{n}\cdot np = p$

(c) $I(p) = -E\left[\frac{\partial^2}{\partial p^2}\log f(x|p)\right]$

$$\frac{\partial^2}{\partial p^2} = \frac{\partial}{\partial p}(\log f(x|p)) = \frac{-x}{p^2} - \frac{n-x}{(1-p)^2}$$

$$= -E\left[\frac{-x}{p^2} - \frac{n-x}{(1-p)^2}\right]$$

$$= E\left[\frac{x}{p^2}\right] + E\left[\frac{n-x}{(1-p)^2}\right]$$

$$= \frac{np}{p^2} + \frac{1}{(1-p)^2}\left[E[n] - E(x)\right]$$

$$= \frac{np}{p^2} + \frac{n-np}{(1-p)^2} = \frac{n}{p} + \frac{n}{1-p} = \frac{(1-p)n + np}{(1-p)p} = \boxed{\frac{n}{(1-p)p}}$$

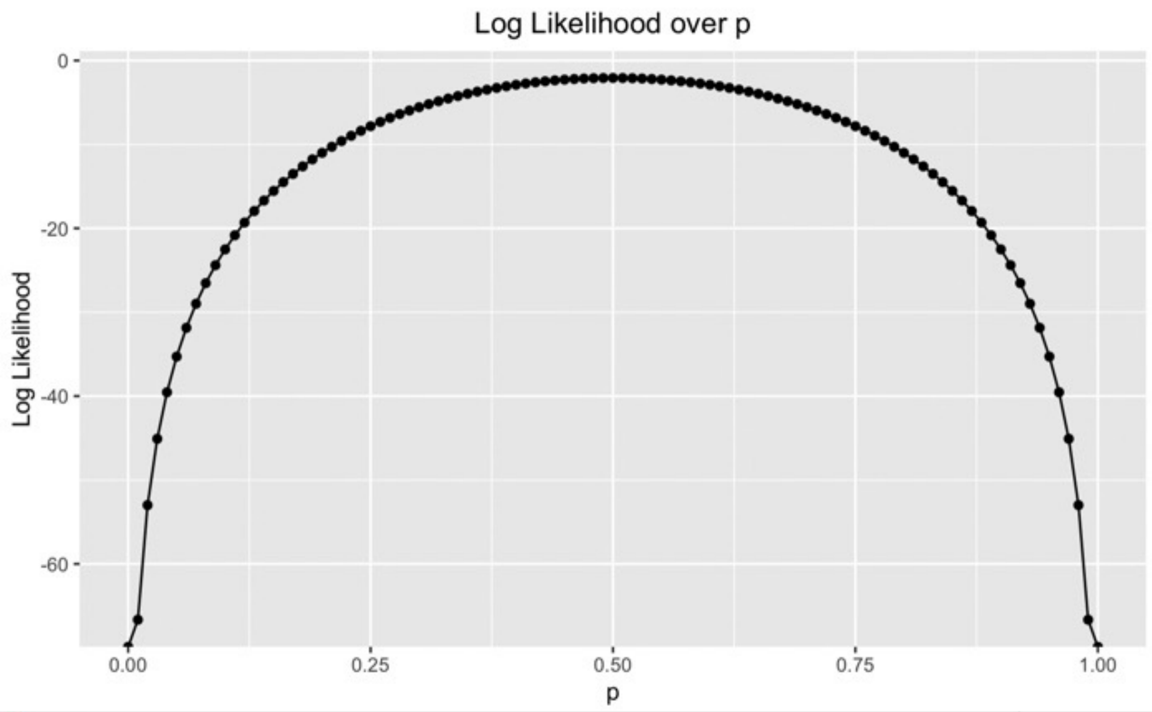
Cramer's Lower Bound:

$$\text{var}(p) = \frac{1}{I(p)} = \boxed{\frac{(1-p)p}{n}}$$

MLE of p: $var\left(\frac{X}{n}\right) = \frac{1}{n^2} var(X) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$.

since $var(\hat{p}) = \frac{1}{I(p)}$, the MLE achieves Cramer Rao bound.

(d)




```{r}
library(ggplot2)
log_likelihood <- function(p){
  return(log(choose(40,20)) + 20*log(p)+ (40-20)*log(1-p))
}
x<-seq(from = 0, to = 1, by= 0.01)
y<-log_likelihood(x)
df <-data.frame(x, y)
ggplot(df, aes(x,y)) + geom_point()+ geom_line()+ labs(x="p", y="Log Likelihood", title= 'Log Likelihood over p') + theme(plot.title =
element_text(hjust = 0.5))
```

5. a) $X_i \sim N(0,1)$ Derive the log-likelihood function given $\sigma^2 = 1$.

$$f(x|\mu,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

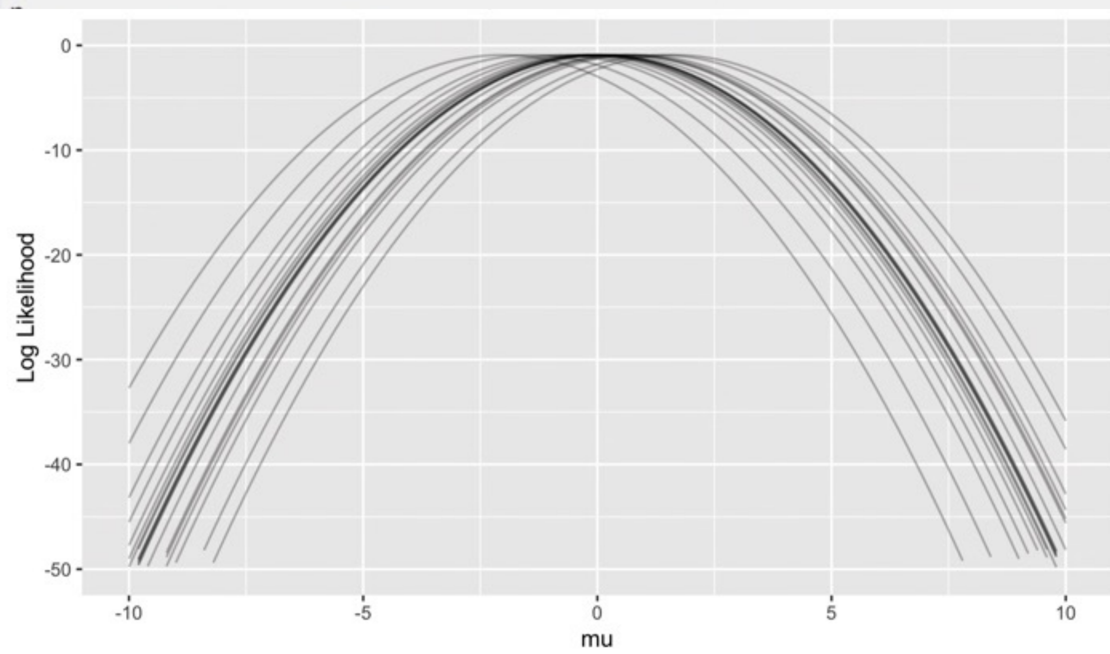
$$lik = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} = \prod_{i=1}^{n} (2\pi)^{-1/2} e^{-(x-\mu)^2/2}$$

$$\boxed{\log lik = -\frac{n}{2}\log(2\pi) - \sum_{i=1}^{n} \frac{(x-\mu)^2}{2}}$$

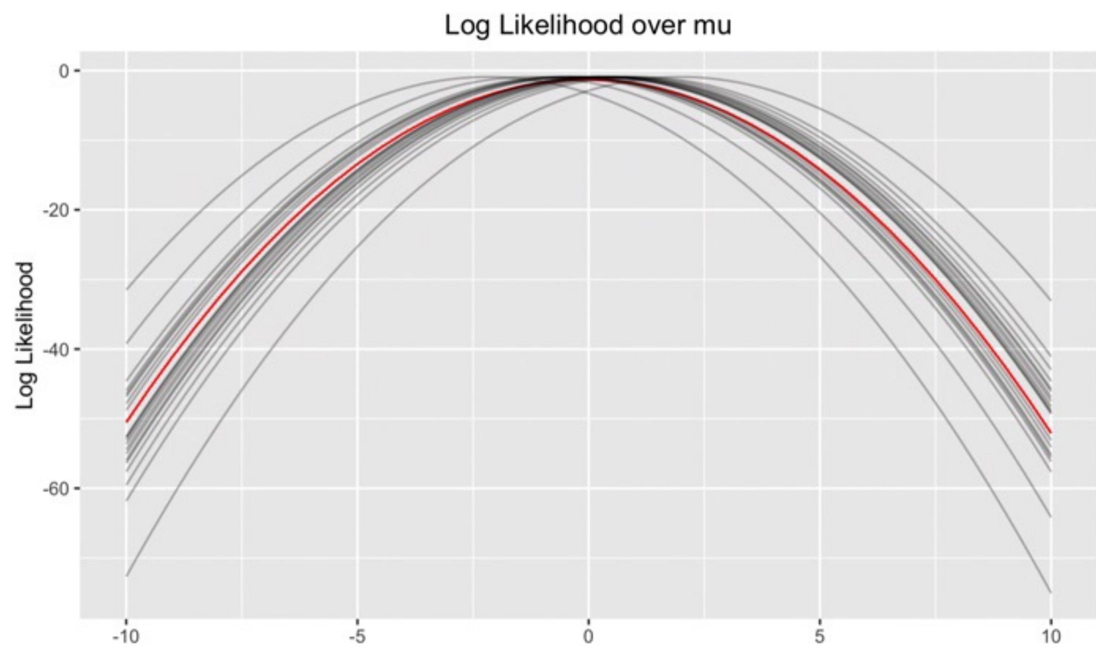
b)

```r
# 5b
sample_size <- 20
y <- rep(0,41)
samples <- rnorm(sample_size)
log_likelihood_func <- function(mu,X) {
  return (-1/2*log(2*pi) - (X-mu)^2/2)
}
p <- p_average
for (i in 1: sample_size){
  p<- p +stat_function(fun=log_likelihood_func, args =list(X= samples[i]),alpha = 0.3) +
labs(x="mu", y="Log Likelihood", title= 'Log Likelihood over mu') + theme(plot.title =
element_text(hjust = 0.5))

}
```



c)




```r
```{r}
new_sample_size <- 5000
new_samples <- rnorm(new_sample_size)
mu <- seq(-10,10,0.5)
y <- numeric(40)
counter <- 1
for (k in mu){
  for (i in 1:sample_size){
    # summing all the log_likelihood function at each corresponding mu values.
    y[counter] = y[counter] + log_likelihood_func(k, new_samples[i])
  }
  y[counter] = y[counter] / sample_size
  counter = counter+1
}
df <- data.frame(x=mu, y=y)
p_average <- ggplot(df, aes(x=x, y=y)) + geom_line(color='red')

p_average
```
```

```{r}
new_sample_size <- 5000
new_samples <- rnorm(new_sample_size)
mu <- seq(-10,10,0.5)
y <- numeric(40)
counter <- 1
for (k in mu){
  for (i in 1:sample_size){
    # summing all the log_likelihood function at each corresponding mu values.
    y[counter] = y[counter] + log_likelihood_func(k, new_samples[i])
  }
  y[counter] = y[counter] / sample_size
  counter = counter+1
}
df <- data.frame(x=mu, y=y)
p_average <- ggplot(df, aes(x=x, y=y)) + geom_line(color='red')

p_average
```

d) Fisher information is $\frac{1}{\sigma^2}$. since $\sigma^2 = 1$, FI = 1.

The average log likelihood function is centered around 0, suggesting that its mean is 0 and the variance is 1, as given.

normal pdf: $f(x|0,1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$\ell(x) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}x^2$$

$$\ell'(x) = -x \qquad \therefore I(\theta_0) = -E(\ell''(x)) = -(-1) = \boxed{1}$$

$$\ell''(x) = -1$$

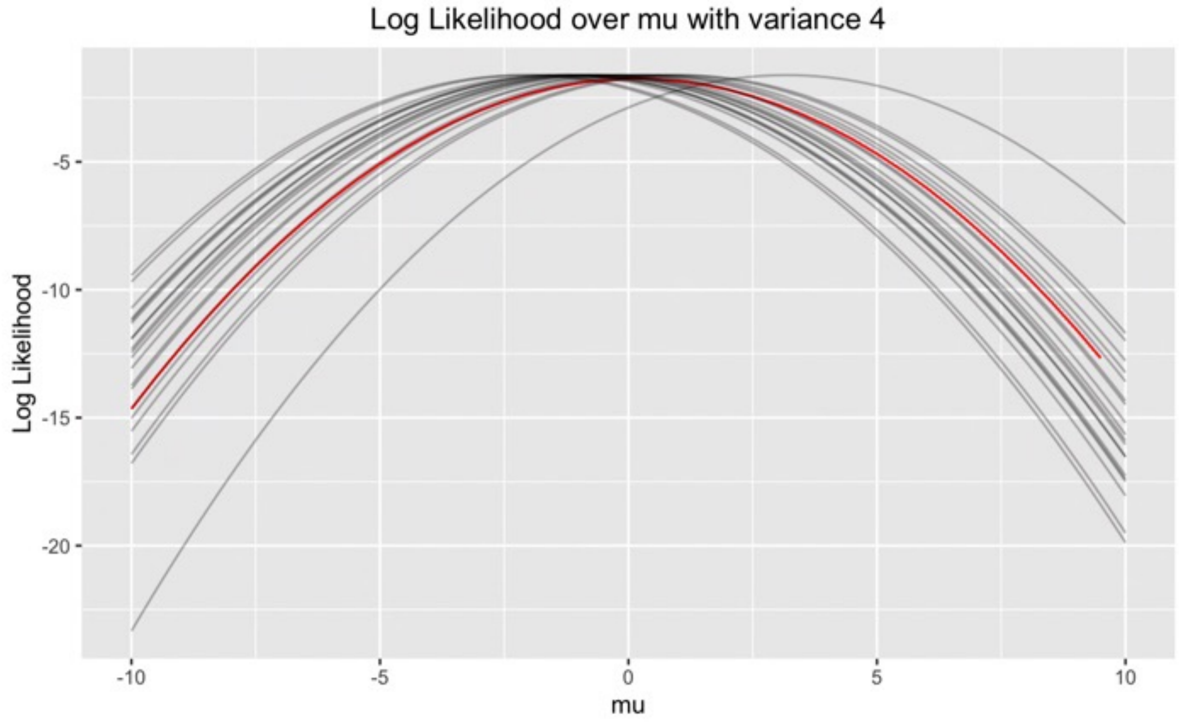
e) when $\sigma^2 = 4$,

$$f(x|\mu, \sigma^2=4) = \frac{1}{\sqrt{2\cdot 4\pi}} e^{-\frac{(x-\mu)^2}{2(4)}}$$

$$= \frac{1}{\sqrt{8\pi}} e^{-\frac{(x-\mu)^2}{8}}$$

$$\text{lik} = \prod_{i=1}^{n} \frac{1}{\sqrt{8\pi}} e^{-\frac{(x-\mu)^2}{8}} = \prod_{i=1}^{n} (8\pi)^{-1/2} e^{-(x-\mu)^2/8}$$

$$\boxed{\log \text{lik} = \frac{-n}{2}\log(8\pi) - \sum_{i=1}^{n} \frac{(x-\mu)^2}{8}}$$




```r
# 5d-(c)
log_likelihood_func <- function(mu,X) {
  return (-1/2*log(8*pi) - (X-mu)^2/8)
}
new_sample_size <- 5000
new_samples <- rnorm(new_sample_size)
mu <- seq(-10,10,0.5)
y <- numeric(40)
counter <- 1
for (k in mu){
  for (i in 1:sample_size){
    # summing all the log_likelihood function at each corresponding mu values.
    y[counter] = y[counter] + log_likelihood_func(k, new_samples[i])
  }
  y[counter] = y[counter] / sample_size
  counter = counter+1
}
df <- data.frame(x=mu, y=y)
p_average <- ggplot(df, aes(x=x, y=y)) + geom_line(color='red')

p_average

# 5d-(b)
sample_size <- 20
y <- rep(0,41)
samples <- rnorm(sample_size)
p <- p_average
for (i in 1: sample_size){
  p<- p +stat_function(fun=log_likelihood_func, args =list(X= samples[i]),alpha = 0.3) +
labs(x="mu", y="Log Likelihood", title= 'Log Likelihood over mu with variance 4') +
theme(plot.title = element_text(hjust = 0.5))

}
p
```

f) $N(0,4)$ has FI of $\frac{1}{4}$. $N(0,1)$ has FI of 1.

since FI of $N(0,1)$ is greater than FI of $N(0,4)$, and when $I(\theta_0)$ is large, it means that it is easier to detect true parameter θ_0 over different samples, the estimation of μ is easier for $N(0,1)$ than $N(0,4)$.

normal pdf: $f(x \mid 0,4) = \frac{1}{\sqrt{8\pi}} e^{-x^2/8}$

$$\ell(x) = -\frac{1}{2}\log(8\pi) - \frac{1}{8}x^2$$

$$\ell'(x) = -\frac{1}{4}x \qquad \therefore I(\theta_0) = -E(\ell''(x)) = -(-1/4) = 1/4$$

$$\ell''(x) = -\frac{1}{4}$$