Stat 330 Assignment 5 Solutions

Mandy Yao

November 2020

5.1.2

a)

If $X_1, ..., X_n$ are iid with bernoulli(P) distribution, $\sum_{i=1}^{n} X_i \sim Bin(n,p)$. We have that $Y_n = \sum_{i=1}^{n} X_i$, and

$$E(\frac{Y_n}{n}) = E(\bar{X}_n) = \frac{1}{n}E(\sum_{i=1}^{n} X_i) = \frac{1}{n} \times n \times p = p.$$

Since the sequence $\{X_1, ..., X_n\}$ has finite variance, $Var(X_i) = p(1-p)$. By WLLN,

$$\bar{X}_n \xrightarrow{\text{p}} p \Longleftrightarrow \frac{Y_n}{n} \xrightarrow{\text{p}} p.$$

Alternatively, by Chebyshev's inequality,

$$lim_{n\to\infty} P(|\frac{Y_n}{n} - p| \geq \epsilon) = lim_{n\to\infty} \frac{np(1-p)}{n^2\epsilon^2} = 0.$$

b)

If $g(x) = 1 - x$, $g(x)$ is continuous. Since $\frac{Y_n}{n} \xrightarrow{\text{p}} p$, by theorem 5.1.4,

$$g(\frac{Y_n}{n}) = 1 - \frac{Y_n}{n} \xrightarrow{\text{p}} g(p) = 1 - p.$$

c)

First, we have that

$$(\frac{Y_n}{n})(1 - \frac{Y_n}{n}) = (\frac{Y_n}{n}) - (\frac{Y_n}{n})(\frac{Y_n}{n}).$$

Also, we already got that $\frac{Y_n}{n} \xrightarrow{\text{p}} p$, so by theorem 5.1.5,

$$(\frac{Y_n}{n})(\frac{Y_n}{n}) \xrightarrow{\text{p}} p^2.$$

Then by theorem 5.1.2,

$$(\frac{Y_n}{n}) - (\frac{Y_n}{n})(\frac{Y_n}{n}) \xrightarrow{\text{p}} p - p^2 = p(1-p).$$

5.2.3

We have that

$$F(Y_n) = P(max\{X_1, ..., X_n\} \leq y_n) = F_X(y_n)^n.$$

Then

$$F_{Z_n}(z) = P(Z_n \leq z) = P(n(1 - F(Y_n)) \leq z) = P(1 - F(Y_n) \leq \frac{z}{n}) = P(F(Y_n) \geq 1 - \frac{z}{n})$$

$$= P(Y_n \geq F^{-1}(1 - \frac{z}{n})) = 1 - P(Y_n < F^{-1}(1 - \frac{z}{n})) = 1 - F_{Y_n}(F^{-1}(1 - \frac{z}{n}))$$

$$= 1 - F(F^{-1}(1 - \frac{z}{n}))^n = 1 - (1 - \frac{z}{n})^n.$$

To compute $lim_{n\to\infty}\{1 - (1 - \frac{z}{n})^n\}$, consider the special limit

$$lim_{x\to\infty}(1 - \frac{k}{x})^{mx} = e^{-mk}.$$

With $k = z$, $x = n$, and $m = 1$, we get

$$lim_{n\to\infty}\{1 - (1 - \frac{z}{n})^n\} = 1 - e^{-z}, \;\; z > 0.$$

Therefore the limiting distribution is Exponential(1).

5.2.12

The mgf of Z_n is

$$M_{Z_n}(t) = e^{n(e^t - 1)},$$

so

$$M_{Y_n}(t) = E(e^{\frac{t}{\sqrt{n}}(Z_n - n)}) = e^{-t\sqrt{n}}e^{n(e^{\frac{t}{\sqrt{n}}} - 1)}.$$

Since the Maclaurin series of $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$,

$$lim_{n\to\infty}exp(-t\sqrt{n} + n(e^{t/\sqrt{n}} - 1)) = lim_{n\to\infty}exp(-t\sqrt{n} + n(1 + \frac{t}{\sqrt{n}} + \frac{t^2}{2n} + \frac{t^3}{6n^{1.5}} + ... - 1))$$

$$= lim_{n\to\infty}exp(\frac{t^2}{2} + \frac{t^3}{6n^{0.5}} + ...) = e^{\frac{t^2}{2}},$$

which is the mgf of $N(0, 1)$.

5.3.6

We have that

$$E(Y) = 400 \times \frac{1}{5} = 80,$$

$$Var(Y) = 400 \times \frac{1}{5} \times \frac{4}{5} = 64.$$

Then

$$P(0.25 < Y/400) = P(Y > 100) = 1 - P(Y < 100).$$

With continuity correction, this is approximately

$$1 - P(Y < 100.5) = 1 - \Phi(\frac{100.5 - 80}{\sqrt{64}}) = 1 - \Phi(2.5625) = 1 - 0.9948 = 0.0052.$$

6.1.2

a)

We get that

$$L(\theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1} = \theta^n (\prod_{i=1}^{n} x_i)^{\theta-1}$$

$$l(\theta) = nlog\theta + (\theta - 1)log(\prod_{i=1}^{n} x_i)$$

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} log(x_i)$$

Setting the last equation to 0, we obtain

$$\hat{\theta} = \frac{-n}{\sum_{i=1}^{n} log(x_i)}.$$

Also, we have that

$$\frac{\partial^2 l(\theta)}{\partial \theta^2} = \frac{-n}{\theta^2} < 0.$$

Therefore $\hat{\theta}$ is a local maximum, so $\hat{\theta}$ is the MLE.

b)

We get that

$$L(\theta) = \prod_{i=1}^{n} e^{-(x_i-\theta)} I(x_i \geq \theta) = e^{n\theta - \sum_{i=1}^{n} x_i} I((x_1, ..., x_n) \geq \theta) = e^{n\theta - \sum_{i=1}^{n} x_i} I(x_{(1)} \geq \theta).$$

Since $e^{n\theta - \sum_{i=1}^{n} x_i}$ is a monotone increasing function of θ, $L(\theta)$ attains its maximum at $\theta = x_{(1)}$, so $\hat{\theta}_{MLE} = x_{(1)}$.

6.1.9

We get that

$$L(\mu) = \prod_{i=1}^{n} \frac{e^{-\mu} \mu^{x_i}}{x_i!} = \frac{e^{-n\mu} \mu^{\sum x_i}}{\prod x_i!}$$

$$l(\mu) = -n\mu + \sum x_i log(\mu) - log(\prod x_i!)$$

$$\frac{\partial l(\mu)}{\partial \mu} = -n + \frac{\sum x_i}{\mu}$$

Setting the last equation to 0, we obtain

$$\hat{\mu}_{MLE} = \bar{x}.$$

Therefore

$$\widehat{P(X=2)}_{MLE} = \frac{e^{-\hat{\mu}_{MLE}} \hat{\mu}_{MLE}^2}{2!}.$$

Now we use R to find the estimator's realization for the data in the table:

```
mean=sum((0*7),(1*14),(2*12),(3*13),(4*6),(5*3))/55
dpois(2,mean)
```

From this we get that $\widehat{P(X=2)}_{MLE} = 0.27$.

6.2.3

We get that

$$log(f(x;\theta) = -log\pi - log(1+(x-\theta)^2)$$

so

$$\frac{\partial log(f(x;\theta))}{\partial\theta} = \frac{2(x-\theta)}{1+(x-\theta)^2}.$$

Then

$$I(\theta) = E((\frac{\partial log f(X;\theta)}{\partial\theta})^2) = \int_{\infty}^{\infty} \frac{4(x-\theta)^2}{\pi(1+(x-\theta)^2)^3}dx.$$

Let $x-\theta = tanz$. Then

$$x = \theta + tanz \Rightarrow dx = sec^2 z dz.$$

We get

$$I(\theta) = \int_{\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{4tan^2 z sec^2 z}{\pi(1+tan^2 z)^3}dz.$$

Since $1+tan^2 z = sec^2 z$,

$$I(\theta) = \int_{\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{4}{\pi}(\frac{sin^2 z}{cos^2 z})(cos^2 z)^2 dz = \int_{\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{4}{\pi} sin^2 z cos^2 z dz.$$

Using $cos^2 z = \frac{cos2z+1}{2}$ and $sin^2 z = \frac{1-cos2z}{2}$,

$$I(\theta) = \int_{\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{4}{\pi}(\frac{1-cos2z}{2})(\frac{cos2z+1}{2})dz = \frac{1}{2}.$$

Therefore the lower bound is

$$\frac{1}{nI(\theta)} = \frac{2}{n}.$$

Using theorem 6.2.2, we can also get that

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{\text{D}} N(0, \frac{1}{I(\theta)}) = N(0,2).$$

6.2.10

First, we have

$$E(|X_i|) = \int_{-\infty}^{\infty} \frac{|x_i|}{\sqrt{2\pi\theta}} e^{\frac{-1}{2\theta}x_i^2} dx_i = 2\int_0^{\infty} \frac{x_i}{\sqrt{2\pi\theta}} e^{\frac{-1}{2\theta}x_i^2} dx_i.$$

Let $u = x_i^2 \Rightarrow du = 2x_i dx_i$. Then

$$E(|X_i|) = \int_0^{\infty} \frac{1}{\sqrt{2\pi\theta}} e^{\frac{-u}{2\theta}} du = \sqrt{\frac{2\theta}{\pi}}.$$

We want

$$\sqrt{\theta}E(Y) = E(c\sum_{i=1}^{n}|X_i|) = c\sum_{i=1}^{n}E(|X_i|) = nc\sqrt{\frac{2\theta}{\pi}}.$$

Therefore for Y to be unbiased, $c = \frac{1}{n}\sqrt{\frac{\pi}{2}}$.

Now to calculate $Efficiency = \frac{1/(nI(\sqrt{\theta}))}{Var(Y)}$,

$$Var(Y) = c^2\sum_{i=1}^{n}Var(|X_i|) = c^2\sum_{i=1}^{n}(E(|X_i|^2) - E(|X_i|)^2) = c^2\sum_{i=1}^{n}(E(X_i^2) - \frac{2\theta}{\pi}).$$

Since $E(X_i^2) = Var(X_i) + E(X_i)^2 = \theta$,

$$Var(Y) = c^2 \sum_{i=1}^{n} (\theta - \frac{2\theta}{\pi}) = nc^2\theta(1 - \frac{2}{\pi}) = \frac{\theta}{n}(\frac{\pi}{2} - 1).$$

Also for $f(x_i, \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{\frac{x_i^2}{2\theta}}$,

$$log f(x_i; \theta) = \frac{-1}{2} log(2\pi) - \frac{1}{2} log\theta - \frac{x_i^2}{2\theta}$$

$$\frac{\partial log f(x_i;\theta)}{\partial\theta} = \frac{-1}{2\theta} + \frac{x_i^2}{2\theta^2}.$$

From this we get

$$I(\theta) = Var(\frac{\partial log f(X_i;\theta)}{\partial\theta}) = Var(\frac{-1}{2\theta} + \frac{X_i^2}{2\theta^2}) = Var(\frac{X_i^2}{2\theta^2}) = \frac{1}{4\theta^2} Var((\frac{X_i}{\sqrt{\theta}})^2).$$

Since $X_i \sim N(0, \theta)$, $\frac{X_i}{\sqrt{\theta}} \sim N(0, 1)$. Then $(\frac{X_i}{\sqrt{\theta}})^2 \sim \chi^2_{(1)}$, so

$$I(\theta) = \frac{1}{4\theta^2} \times 2 = \frac{1}{2\theta^2}.$$

Now

$$I(\sqrt{\theta}) = \frac{I(\theta)}{(\frac{d}{d\theta}\sqrt{\theta})^2} = \frac{\frac{1}{2\theta^2}}{(\frac{1}{2}\theta^{\frac{-1}{2}})^2} = \frac{2}{\theta}.$$

Therefore the efficiency is

$$\frac{1/(nI(\sqrt{\theta}))}{Var(Y)} = \frac{\frac{1}{n(2/\theta)}}{\frac{\theta}{n}(\frac{\pi}{2} - 1)} = \frac{1}{\pi - 2}.$$

6.3.6

We get that

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} exp(\frac{-1}{2\theta}(X_i - \mu_0)^2) = (\frac{1}{\sqrt{2\pi\theta}})^n exp(\frac{-1}{2\theta}\sum_{i=1}^{n}(X_i - \mu_0)^2)$$

$$l(\theta) = \frac{-n}{2} log 2\pi - \frac{n}{2} log\theta - \frac{1}{2\theta}\sum_{i=1}^{n}(X_i - \mu_0)^2$$

$$\frac{\partial l(\theta)}{\partial\theta} = \frac{-n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(X_i - \mu_0)^2.$$

Setting this to 0, we get

$$\hat{\theta} = \frac{\sum_{i=1}^{n}(X_i - \mu_0)^2}{n}.$$

Then we have

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} = \frac{(2\pi\theta_0)^{-n/2} exp(-\frac{1}{2\theta_0}\sum_{i=1}^{n}(X_i - \mu_0)^2)}{(2\pi\sum(X_i - \mu_0)^2/n)^{-n/2} exp(\frac{-n}{2})} = \left(\frac{W}{n}\right)^{\frac{n}{2}} exp(\frac{-1}{2}(W - n))).$$

We can see that this is based upon the statistic W. Now we find the null distribution of W:

$$X_i \sim N(\mu_0, \theta) \Rightarrow \frac{X_i - \mu_0}{\sqrt{\theta}} \sim N(0, 1) \Rightarrow \frac{(X_i - \mu_0)^2}{\theta} \sim \chi^2_{(1)} \Rightarrow \sum_{i=1}^{n} \frac{(X_i - \mu_0)^2}{\theta} \sim \chi^2_{(n)}$$

Therefore, we reject H_0 if $W > \chi^2_{\alpha/2}(n)$ or if $W < \chi^2_{1-\alpha/2}(n)$.

6.3.10

a)

We get that

$$L(\theta) = \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{\sum X_i}(1-\theta)^{n-\sum X_i}$$

$$l(\theta) = \sum X_i log\theta + (n - \sum X_i) log(1-\theta)$$

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{\sum X_i}{\theta} - \frac{n - \sum X_i}{1-\theta}.$$

Setting the last equation to 0,

$$\sum X_i - \hat{\theta}\sum X_i = n\hat{\theta} - \hat{\theta}\sum X_i \Rightarrow \hat{\theta} = \frac{\sum X_i}{n}$$

Then we get

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} = \frac{\theta_0^{\sum X_i}(1-\theta_0)^{n-\sum X_i}}{(\frac{Y}{n})^{\sum X_i}(1-\frac{Y}{n})^{n-\sum X_i}} = \left(\frac{n\theta_0}{Y}\right)^Y \left(\frac{n(1-\theta_0)}{n-Y}\right)^{n-Y}.$$

We can see that this is based upon the statistic Y. Now the null distribution of Y is $Y = \sum_{i=1}^{n} X_i \sim Bin(n, \theta)$ because $X_i \sim Bin(1, \theta)$.

b)

We want to find c_1 and c_2 so that

$$P(Y \geq c_2 \;\; or \;\; Y \leq c_1) = 0.05.$$

This is equal to

$$P(\frac{Y - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \geq \frac{100 - c_1 - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}) + P(\frac{Y - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \leq \frac{c_1 - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}})$$

$$= P(Z \geq \frac{50 - c_1}{5}) + P(Z \leq \frac{c_1 - 50}{5}) = 0.05$$

where $Z \sim N(0,1)$ by the Central Limit Theorem. Then

$$2P(Z \leq \frac{c_1 - 50}{5}) = 0.05 \Rightarrow \frac{c_1 - 50}{5} = z_{0.025} = -1.96$$

$$\Rightarrow c_1 = 40.2 \approx 40, \;\; and \;\; c_2 = 59.8 \approx 60.$$