Assignment 3 Q1 MAST20005 S2 2020

Question 1a

Hypotheses

where m is the median of daily new coronavirus cases

H0: m = 15

H1: m < 15

```
x = c(48,70,47,40,35,41,30,39,41,25,44,20,13,11,28,14,11,13,12,16,5)
#each number represents the number of coronavirus from Sep 2020 to Sep
20.
summary(x)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    5.00   13.00   28.00   28.71   41.00   70.00

#a) using the sign test
# m = 28 (our median value) this can help us indicate our
understanding whether or not we should reject the null given at 5%
significance level.

binom.test(sum(x<15),21)

##
##  Exact binomial test
##
## data:  sum(x < 15) and 21
## number of successes = 7, number of trials = 21, p-value = 0.1892
## alternative hypothesis: true probability of success is not equal to
0.5
## 95 percent confidence interval:
##  0.1458769 0.5696755
## sample estimates:
## probability of success
##              0.3333333
```

```r
#The p-value is 0.1892 higher than 0.05 from the significant level.
Therefore , we fail to reject the null hypothesis.
```

The p-value is 0.1892 higher than 0.05 from the significant level. Therefore , we fail to reject the H0.

Question 1b

Hypotheses

#where m(d) is the difference between the median row from row 2 and row 3 (2nd and 3rd weeks of median cases respectively)

H0: m(d) = 0
H1: m(d) > 0

```r
x2 = c(39,41,25,44,20,13,11)
sort(x2) #to verify

## [1] 11 13 20 25 39 41 44

x3 = c(28,14,11,13,12,16,5)
sort(x3) #to veRify

## [1]  5 11 12 13 14 16 28

wilcox.test(x2,x3,paired = TRUE)

##
##  Wilcoxon signed rank exact test
##
## data:  x2 and x3
## V = 27, p-value = 0.03125
## alternative hypothesis: true location shift is not equal to 0
```

The p value is 0.03125, so we reject the H0.

Q2a)

$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$

$(-e^{-\lambda x})\Big|_0^{x} = -e^{-\lambda x} + 1$

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0 & ; \text{otherwise} \end{cases}$$

$$\mathbb{P}(X \leq \pi_p) = 1 - F(x)$$
$$= 1 - e^{-\lambda x}$$

$\pi_p = F^{-1}(p) = ?$

let $y = p$, switch → simply $x = p$.

$$y = 1 - e^{-\lambda x}$$
$$y \longleftrightarrow x$$
$$x = 1 - e^{-\lambda y}$$
$$e^{-\lambda y} = 1 - x$$
$$-\lambda y = \ln(1-x)$$
$$y = \frac{\ln(1-x)}{-\lambda}$$

$$\Rightarrow \quad y = F^{-1}(p) = \frac{\ln(1-p)}{-\lambda} = \pi_p$$

Q2b) $\pi_{0.25} = \dfrac{\ln(0.75)}{-\lambda}$

$\lambda = ?$

first we find the MLE (maximum likelihood estimator for λ).

$$L(\lambda) = \lambda^n e^{-\lambda \sum x_i}$$
$$\ell(\lambda) = n\ln(\lambda) - \lambda \sum x_i$$
$$S(\lambda) = \frac{n}{\lambda} - \sum x_i$$

$$\Rightarrow S(\lambda) = 0 = \sum x_i = \frac{n}{\lambda}$$

$$\overline{X} = \frac{1}{\lambda}$$

$$\Rightarrow \hat{\lambda} = \frac{1}{\overline{X}}$$

use later for Q2c & d

$\overline{X} = (0.11 + 0.21 + 0.75 + 1.14 + 1.35 + 1.63 + 1.63 + 1.83 + 1.93 + 2.04 + 2.16$
$+ 2.25 + 2.41 + 2.52 + 2.65 + 2.83 + 2.92 + 2.92 + 4.83 + 7.23 + 8.80$
$+ 9.80 + 11.54 + 12.16 + 12.91 + 13.93 + 19.68 + 20.94 + 21.73 + 24.09) / 30$

$\bar{x} = 6.6973$

$\hat{\lambda} = 0.1493131595$ — use later for Q2c,d

$\pi_{0.25} = 0.8368$

$\hat{\pi}_p = X_{(k)}$ $\quad p = \frac{k-1}{n-1}$

$k = 8.25$

$X_{(8.25)} = X_{(8)} + 0.25(X_{(9)} - X_{(8)})$
$= 1.83 + 0.25(0.10)$
$= 1.855$ (final) answer for Q2b

Q2b)

$p = \frac{k-1}{n-1}$

$29(0.25)+1 = k$

$k = 8.25$

$$X_{(8.25)} = X_{(8)} + 0.25(X_{(9)} - X_{(8)})$$
$$= 1.83 + 0.10 \times 0.25$$
$$= 1.855 = \hat{\pi}_{0.25}$$

Q2c) $\hat{\pi}_{0.25} \approx N(\pi_{0.25}, \; p(1-p)/nf(\pi_p)^2)$

$$\pi_{0.25} = \frac{\ln(0.75)}{-0.1493} = 0.8368$$

$-0.1493 = -\lambda$

$$\frac{p(1-p)}{nf(\pi_p)^2} = \frac{0.25(0.75)}{30\,\boxed{f(\pi_{0.25})^2}}$$

$$\Rightarrow f(\pi_{0.25})^2 = f(0.8368)^2 = \left(\lambda e^{-\lambda(0.8368)}\right)^2$$
$$= \lambda^2 e^{-2\lambda(0.8368)}$$

$$\hat{\pi}_{0.25} \approx N\left(\pi_{0.25}, \frac{\frac{3}{16}}{30\lambda^2 e^{-2\lambda(0.8368)}}\right)$$

1.6736

$$\approx N\left(\pi_{0.25}, \frac{\cancel{3}\,1}{16 \times \cancel{30}_{10}\,\lambda^2 e^{-1.67367\lambda}}\right)$$

$$\approx N\left(\pi_{0.25}, \frac{1}{160\lambda^2 e^{-1.6736\lambda}}\right)$$

$\left[\text{i.e. } \approx N(0.8368, 0.3579)\right]$

$\pi_{0.25} = 0.8368$

$$\frac{p(1-p)}{nf(\pi_p)^2} \longrightarrow \frac{1}{160(0.149313\ldots)^2 e^{-1.6376(0.149313)}} = 0.3579936707$$

at $p = 0.25$

$n = 30$, given

$\hat{\lambda} = 0.149313$

Q2d)

from
Problem 1 Tute 1

$se(\hat{T}_{0.25}) = \sqrt{0.3579936707}$

$= 0.5983257229.$

Q3) a: prior,

$f(\beta) = e^{-\beta}$

$f(x_1, \ldots, x_n) = \beta^{2n} \prod_{i=1}^{n} x_i \, e^{-\beta \Sigma x_i}$
likelihood

$f(\beta | X = x) \propto Pr(\cancel{\beta}\,\theta | \beta) f(\beta)$

$= A \int_0^{\infty} f(x_1, \ldots, x_n | \beta) \underset{f(\beta)}{\underset{\parallel}{e^{-\beta}}} = 1$

$= A \int_0^{\infty} \beta^{2n} \prod_{i=1}^{n} x_i \, e^{-\beta \Sigma x_i} e^{-\beta} d\beta$

$= A \int_0^{\infty} \beta^{2n} \prod_{i=1}^{n} x_i \, e^{-\beta(\Sigma x_i + 1)} d\beta$

$= A \int_0^{\infty} \prod_{i=1}^{n} x_i \, \beta^{2n} e^{-\beta(\Sigma x_i + 1)} d\beta$

$= A \prod_{i=1}^{n} x_i \underbrace{\int_0^{\infty} \beta^{2n} e^{-\beta(\Sigma x_i + 1)} d\beta}_{I}$ $\boxed{\text{let } \infty = x}$

~~Integration by parts.~~

$u = \beta^{2n}$ $\quad v' = e^{-\beta(\Sigma x_i + 1)}$

$u' = 2n\beta^{2n-1}$ $\quad v = \frac{1}{-\beta(\Sigma x_i + 1)} e^{-\beta(\Sigma x_i + 1)}$

$I = \lim_{x \to \infty} \left[\cancel{\beta^{2n-1}} \frac{\beta^{2n}}{-\beta(\Sigma x_i + 1)} e^{-\beta(\Sigma x_i + 1)} \right]_0^x + \int \frac{2n\beta^{2n-1}}{\beta(\Sigma x_i + 1)} e^{-\beta(\Sigma x_i + 1)} d\beta$

$= \cancel{\frac{\beta^{2n}}{\beta(\Sigma x_i + 1)}}$ $\leftarrow$ Ignore

$=$

$= \frac{\cancel{\beta}^{2n}}{-\cancel{\beta}(\Sigma x_i + 1)} e^{-\cancel{\beta}(\Sigma x_i + 1)}$

$$\therefore 1 = \prod_{i=1}^{n} x_i \int_0^{x} A\beta^{2n} e^{-\beta(\sum x_i + 1)} \, d\beta = -\beta^{2n+1} A ((\sum x_i + 1)\beta)^{-2n-1} \Gamma(2n+1, (\sum x_i + 1)\cancel{\beta}) \prod_{i=1}^{n} x_i$$

$$\therefore f(\beta | X = x) = A \, \beta^{2n} \prod_{i=1}^{n} x_i \, e^{-\beta(\sum x_i + 1)}$$

$$\boxed{A = \frac{1}{-\beta^{2n+1}((\sum x_i + 1)\beta)^{-2n-1} \Gamma(2n+1, (\sum x_i + 1)\cancel{\beta}) \prod_{i=1}^{n} x_i}}$$

Wolfram α $= \dfrac{\beta^{2n} e^{-\beta(\sum x_i + 1)}}{-\beta^{2n+1}((\sum x_i + 1)\beta)^{-2n-1} \Gamma(2n+1, (\sum x_i + 1)\cancel{\beta})}$

$$= -\frac{e^{-\beta(\sum x_i + 1)}}{\beta((\sum x_i + 1)\beta)^{-2n-1} \Gamma(2n+1, (\sum x_i + 1)\beta)}$$

→ posterior ~~$\beta|x$~~ pdf

$$\beta | x \sim \Gamma(2n+1, (\sum x_i + 1)\beta)$$

Q3B) $E[\beta | x] = \dfrac{2n+1}{(\sum x_i + 1)\beta}$

$$sd[\beta | x] = \sqrt{\frac{2n+1}{(\sum x_i + 1)^2 \beta^2}}$$

$$= \frac{\sqrt{2n+1}}{(\sum x_i + 1)\beta}$$

NOTE!

$$\boxed{E[\beta | x] = \alpha / \beta \qquad Var[\beta | x] = \alpha / \beta^2}$$

Q4)

a. $X_1, \dots, X_n \sim N(\mu, \sigma^2)$.

$$f(x) = \frac{1}{\sigma(2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$ (Normal pdf).

$\Rightarrow$ μ is unknown, σ^2 known,

$\Rightarrow$ find a sufficient statistic for μ.

$$L(x_1, \dots, x_n \mid \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}\left\{\frac{x_i-\mu}{\sigma}\right\}^2\right\}$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left(\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2\right)$$

$\Rightarrow$ μ unknown,

$$L(x_1, \dots, x_n \mid \mu) = (2\pi\sigma^2)^{-n/2} \underbrace{\exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{n\mu^2}{2\sigma^2}\right)}_{\text{exponential family}}$$

By factorization theorem,

σ^2 is known, let

$$u(x_1, \dots, x_n) = (2\pi\sigma^2)^{n/2} \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right)$$

does not depend on μ

depends on $x_1, \dots, x_n$

$$v(r(x_1, \dots, x_n)\mid\mu) = \exp\left(-\frac{n\mu^2}{2\sigma^2} + \frac{\mu}{\sigma^2} r(x_1, \dots, x_n)\right)$$

where, $r(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i$

$\Rightarrow$ By factorization theorem, $\sum_{i=1}^{n} x_i$ is a sufficient statistic. Follows that the sample mean $\bar{X}_n$ is also a sufficient statistic for μ.

b). Next page.

family of exp family.

$$f(x_1, \ldots, x_n \mid \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum x_i^2 + \frac{\mu}{\sigma^2} \sum x_i - \frac{n\mu}{2\sigma^2}\right)$$

$$h(x_1, \ldots, x_n) = (2\pi)^{-n/2}$$

$$\phi(g_1(x_1, \ldots, x_n), g_2(x_1, \ldots, x_n)) = \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} \underbrace{\sum x_i^2}_{g_1(x_1,\ldots,x_n)} + \frac{\mu}{\sigma^2} \underbrace{\sum x_i}_{g_2(x_1,\ldots,x_n)} - \frac{n\mu^2}{2\sigma^2}\right\}$$

$g_3(x_1, \ldots, x_n)$

by factorization theorem, $\sum_{i=1}^n X_i$ is a sufficient statistic. $\boxed{\frac{\sum X_i}{n}} = g_3(x_1, \ldots, x_n)$

$$= \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} g_1(x_1, \ldots, x_n) + \frac{\mu}{\sigma^2} g_2(x_1, \ldots, x_n) - \frac{n\mu^2}{2\sigma^2}\right\}$$

* $g_1(x_1, \ldots, x_n) = \sum x_i^2$ is sufficient statistic of σ^2. And

joint

$g_2(x_1, \ldots, x_n) = \sum x_i$ is also a sufficient statistic of σ^2.

$\Rightarrow$ follows $\bar{X}_n$ is sufficient statistic.

$g_3(x_1, \ldots, x_n) = (\sum X_i)^2 \Rightarrow$ sufficient statistic.

*final answer $\sum_{i=1}^n (x_i - \bar{x})^2$ is sufficient statistic for σ^2.

(Q4c) $\phi(g(x_1, \ldots, x_n), \sigma)$

$$= \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2} \sum X_i^2 + \frac{\mu}{\sigma^2} \sum X_i - \frac{n\mu^2}{2\sigma^2}\right)$$

$$= \cancel{\sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}(\sigma^2 + n\mu^2)\right)}$$

$$\hat{\sigma} = \frac{1}{n} \sum_{i=1}^n X_i^2 - n\bar{X}^2$$

$$\hat{\mu} = \bar{X} = \sum X_i / n$$

— By def^n the MLE's for variance & the mean.

final answer

Bijection $\sigma^2 \mapsto \sigma$. $\sum_{i=1}^n (x_i - \bar{x})^2$ is sufficient statistic for σ^2.

Q5) a) The pdf of the exponential distribution with mean $1/\lambda$ is $f(x) = \lambda e^{-\lambda x}$. The likelihood is a product of this pdf across the n observations in our sample.

$$L(\lambda) = \prod_{i=1}^{n} f(x) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n \prod e^{-\lambda x_i}$$
$$= \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

let $y = \sum_{i=1}^{n} x_i$
$$= \lambda^n e^{-\lambda y},$$

$$L(\lambda_0) = \lambda_0^n e^{-\lambda_0 y}.$$
$$L(\lambda_1) = ?$$
$\Rightarrow$ we first need to compute the likelihood (MLE).

$$\ell(\lambda) = n \ln(\lambda) - \cancel{\lambda} \lambda \sum_i x_i$$

$$s(\lambda) = \frac{n}{\lambda} - \sum x_i$$

$$\Rightarrow s(\lambda) = 0 = \frac{n}{\lambda} - \sum x_i$$

$$\Rightarrow \frac{\sum x_i}{n} = \frac{1}{\lambda}$$

$$\Rightarrow \bar{x} = 1/\lambda$$

$$\Rightarrow \lambda = 1/\bar{x}$$
$$= \frac{1}{\sum x_i / n}$$ $\sum x_i = Y$ (taking the sufficient statistic)
$$= \frac{1}{Y/n} = \frac{n}{Y}$$

$$\frac{L_0}{L_1} = \frac{\lambda_0^n e^{-\lambda_0 y}}{\left(\frac{n}{Y}\right)^n e^{-\cancel{\lambda} y}}$$ (perform the likelihood test),

$$= \frac{\lambda_0^n e^{-\lambda_0 y}}{\left(\frac{n}{y}\right)^n e^{-n}} = \frac{\lambda_0^n}{\left(\frac{n}{y}\right)^n} e^{-\lambda_0 y + n} = \left(\frac{\overset{\text{constant}}{e} \lambda_0 y}{n}\right)^n e^{-\lambda_0 y}$$

where $A = \left(\frac{e\lambda}{n}\right)^n$ is arbitrary $= A\, y^n e^{-\lambda_0 y}$

from tutorial 9

~~$Y \geq$~~ $y < c_1$, $c_2 \leq y$,
for some c_1 & c_2 we reject.

Q5b) Since Y is the sum of iid exponential variables, we have $Y \sim Gamma(n, \lambda)$, the gamma distribution with mean

$\mu_1 = \mu_2 = \ldots = \frac{1}{\lambda}$

$\Rightarrow \mu_1 + \mu_2 + \ldots = n/\lambda$.

Under H_0, this becomes $Y \sim Gamma(n, \lambda_0)$
$(2\lambda_0 Y \sim \chi^2_{2n})$

Q5c) We reject H_0 if Y is too small or too large (consistent with the LRT, see part (a) above). For simplicity, we choose the two ~~others~~ thresholds such that they have equal tail probability under H_0.

R-code
sorry!!

$\Rightarrow$ we would reject H_0 if $Y < 37.111$ or $Y \geq 64.7806$