MAST20005 - Assignment 3

Lucas Fern

18/10/2020

Question 1

1a)

$$H_0 : m = 15$$
$$H_1 : m < 15$$

We will reject H_0 if the proportion of days with a count below 15 is large enough.

```
cases <- c(48, 70, 47, 40, 35, 41, 30,
           39, 41, 25, 44, 20, 13, 11,
           28, 14, 11, 13, 12, 16, 5)
n.under.15 <- sum(cases < 15)
binom.test(n.under.15, length(cases),
           alternative = "greater", conf.level = 0.95)
```

```
##
##  Exact binomial test
##
## data:  n.under.15 and length(cases)
## number of successes = 7, number of trials = 21, p-value = 0.9608
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.1681758 1.0000000
## sample estimates:
## probability of success
##              0.3333333
```

Since the p-value of the test is 0.9608 there is a low chance that this data came from a distribution with a median below 15, so we have insufficient evidence to reject the null.

1b)

Let m_2 be the median of the 2nd week and m_3 be the median of the third week.

$$H_0 : m_2 - m_3 = 0$$
$$H_1 : m_2 - m_3 > 0$$

We will reject the null hypothesis if scores in the 2nd week are significantly higher than in the third week.

```
week.2 <- cases[8:14]
week.3 <- cases[15:21]

wilcox.test(week.2, week.3,
```

```
            alternative = "greater", conf.level = 0.95,
            exact = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  week.2 and week.3
## W = 38, p-value = 0.04798
## alternative hypothesis: true location shift is greater than 0
```

Since the p-value of the test is 0.04798 there is just sufficient evidence to reject the null hypothesis at a 0.05 level of significance.

Question 2

2a)

$$\pi_p = a \text{ , where } \int_0^a \lambda e^{-\lambda x} dx = p$$

$$\int_0^a e^{-\lambda x} dx = -\left[e^{-\lambda x}\right]_0^a$$

$$= -\left[e^{-\lambda a} - 1\right]$$

$$e^{-\lambda a} - 1 = -p$$

$$-\lambda a = \ln(1-p)$$

$$\pi_p = a = \frac{-1}{\lambda} \ln(1-p)$$

2b)

$$\hat{\pi}_{0.25} = x_{(k)} \quad \text{where } k = p(n-1)+1 = 0.25(29)+1 = 8.25$$

$$\hat{\pi}_{0.25} = x_{(8.25)}$$

$$= x_{(8)} + \frac{1}{4}(x_{(9)} - x_{(8)})$$

$$= 1.83 + \frac{1}{4}(1.93 - 1.83)$$

$$= 1.855$$

2c)

$$\hat{\pi}_p \approx N\left(\pi_p, \frac{p(1-p)}{n \cdot f(\pi_p)^2}\right)$$

$$= N\left(\frac{-1}{\lambda}\ln(1-p), \frac{p(1-p)}{n \cdot f(\pi_p)^2}\right)$$

$$\hat{\pi}_{0.25} \approx N\left(\frac{-1}{\lambda}\ln(0.75), \frac{0.25 \times 0.75}{30 \times f\left(\frac{-1}{\lambda}\ln(0.75)\right)^2}\right)$$

$$f\left(\frac{-1}{\lambda} \cdot \ln(0.75)\right) = \lambda \times \exp\left(-\lambda\left(\frac{-1}{\lambda}\ln(0.75)\right)\right)$$
$$= 0.75\lambda$$

$$\Rightarrow \hat{\pi}_{0.25} \approx N\left(\frac{-1}{\lambda}\ln(0.75), \frac{0.1875}{30 \times 0.75^2 \times \lambda^2}\right)$$

$$\hat{\pi}_{0.25} \approx N\left(\frac{-1}{\lambda}\ln(0.75), \frac{1}{90\lambda^2}\right)$$

2d)

The MLE for λ is $\hat{\lambda} = 1/\bar{X}$

$$\bar{X} \approx 6.697 \Rightarrow \hat{\lambda} \approx \frac{1}{6.697} = 0.1493$$

$$se(\hat{\pi}_{0.25}) = \sqrt{\frac{1}{90(0.1493)^2}}$$

$$\approx 0.71$$

Question 3

3a)

$$L(\beta) = \prod_{i=1}^{n} f(x_i|\beta)$$
$$= \prod_{i=1}^{n} \beta^2 x_i e^{-\beta x_i}$$
$$= \beta^{2n} \left(\prod_{i=1}^{n} x_i\right)\left(e^{-\beta \Sigma x_i}\right)$$
$$= \beta^{2n} e^{-\beta n \bar{x}} \left(\prod_{i=1}^{n} x_i\right)$$

$$\therefore f(\beta|x) \propto \beta^{2n} e^{-\beta n\bar{x}} \left(\prod_{i=1}^{n} x_i\right) e^{-\beta}$$
$$\propto \beta^{2n} e^{-\beta(n\bar{x}+1)}$$

This is recognisable as a gamma distribution so the posterior is:

$$(\beta|x) \sim \gamma(2n+1, 1+n\bar{x})$$

3b)

$\gamma(\alpha, \beta)$ has mean α/β and variance α/β^2 so

$$E(\beta|x) = \frac{2n+1}{1+n\bar{x}}$$

$$Var(\beta|x) = \frac{2n+1}{(1+n\bar{x})^2} \implies sd(\beta|x) = \frac{\sqrt{2n+1}}{1+n\bar{x}}$$

Question 4

4a)

The joint PDF is:

$$f(x_1,\dots,x_n|\mu) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x_i-\mu)^2}{2\sigma^2}\right)$$
$$= \frac{1}{(\sigma^2 2\pi)^{\frac{n}{2}}} \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$
$$= \frac{1}{(\sigma^2 2\pi)^{\frac{n}{2}}} \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2-2\mu x_i+\mu^2)\right)$$
$$= \frac{1}{(\sigma^2 2\pi)^{\frac{n}{2}}} \left[e^{\sum x_i^2} e^{-2\mu\sum x_i} e^{n\mu^2}\right]^{\frac{-1}{2\sigma^2}}$$

Can be factorised into $\phi(g(x_1,\dots,x_n),\mu)\,h(x_1,\dots,x_n)$

as:

$$e^{\frac{-\mu}{\sigma^2}\sum x_i} e^{\frac{-n\mu^2}{2\sigma^2}} \; e^{\frac{-1}{2\sigma^2}\sum x_i^2}\left(\frac{1}{(\sigma^2 2\pi)^{\frac{n}{2}}}\right)$$

Since the purple portion doesn't depend on μ

$\sum_{i=1}^{n} x_i$

is a sufficient statistic for μ.

(This also means that $\bar{x}$ is sufficient since it's only a constant transformation from $\sum_{i=1}^{n} x_i$)

4b)

The joint PDF is:

$$f(x_1,\ldots,x_n|\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x_i-\mu)^2}{2\sigma^2}\right)$$

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

$$\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(x_i-\mu)^2 = s^2$$

$$\Rightarrow f(x_1,\ldots,x_n|\sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot \frac{1}{\sigma^n}\exp\left(\frac{1-n}{2\sigma^2}s^2\right)$$

Can be factorised into $\phi(g(x_1,\ldots,x_n),\sigma^2)h(x_1,\ldots,x_n)$:

$$\frac{1}{\sigma^n} e^{\left(\frac{1-n}{2\sigma^2}s^2\right)} \times \frac{1}{(2\pi)^{\frac{n}{2}}}$$

So the sample variance s^2 is a sufficient statistic for σ^2.

4c)

$$f(x_1,\ldots,x_n|\sigma) = f(x_1,\ldots,x_n|\sigma^2)$$

Since our $h(x_1,\ldots,x_n)$ from b) doesn't contain any σ, and s^2 is the only part of $\phi(g(x_1,\ldots,x_n),\sigma)$ which depends on the x_i, s^2 is a sufficient statistic for σ.

Question 5

5a)

$X_1, \ldots, X_n \sim \exp(\lambda) \implies \hat{\lambda}_{MLE} = \frac{1}{\bar{X}}$

$\therefore$ the maximum likelihood under H_1 is $L(\frac{1}{\bar{x}})$.

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$
$$= \lambda^n \exp\{-\lambda \textstyle\sum_{i=1}^{n} x_i\}$$
$$= \lambda^n \exp\{-\lambda n \bar{x}\}$$

So the likelihood ratio is

$$\frac{L_0}{L_1} = \frac{L(\lambda_0)}{L(\frac{1}{\bar{x}})}$$
$$= \frac{\lambda_0^n \exp\{-\lambda_0 n \bar{x}\}}{(\frac{1}{\bar{x}})^n \exp\{\frac{-1}{\bar{x}} n \bar{x}\}}$$

$$= (\bar{x}\lambda_0)^n \exp\{n - \lambda_0 n \bar{x}\}$$

$$Y = \sum_{i=1}^{n} x_i = \bar{x} n \implies \bar{x} = Y/n$$

so

$$\frac{L_0}{L_1} = (\tfrac{Y}{n}\lambda_0)^n \exp\{n - \lambda_0 Y\}$$

and the LRT is based on Y.

5b)

Under H_0, $\lambda = \lambda_0$ so $X_1, \ldots, X_n \sim Exp(\lambda_0)$

It is known that the sum of n exponential random variables with parameter λ is distributed as $\gamma(n, \lambda)$

$$\Rightarrow \sum_{i=1}^{n} X_i = Y \sim \gamma(n, \lambda_0)$$

under H_0.

5c)

It was shown in a) that the LRT is based on Y.
We will reject H_0 if L_0 is significantly larger than L_1, so the critical region will be

$$\frac{L_0}{L_1} \leq C$$

With $n=50$, $\lambda_0=1$, the LRT is

$$\frac{L_0}{L_1} = \left(\frac{Y}{50}\right)^{50} \exp\{50-Y\} \leq C$$

$$Y^{50}e^{-Y} \leq 50^{50}c/e^{50} = c_1$$

$$50\cdot\ln(Y)(-Y) \leq \ln(c_1) = c_2$$
$$Y\cdot\ln(Y) \geq \frac{-1}{50}c_2 = c_3$$

The 0.95 quantile of $Y \sim \gamma(50,1)$ is 62.17
So we reject H_0 if

$$Y\times\ln(Y) \geq 62.17\times\ln(62.17)$$
$$Y\times\ln(Y) \geq 256.76$$

This test has significance level 0.05.