MAST20005/MAST90058: Week 2 Solutions

1. Since this is a random sample, all of the variables are iid. Therefore they all have the same mean and variance.
   (a) $\mathrm{sd}(X_2) = \mathrm{sd}(X_4) = \sqrt{\mathrm{var}(X_4)} = \sqrt{4} = 2$.
   (b) By independence, $\mathrm{var}(X_7 + X_8) = \mathrm{var}(X_7) + \mathrm{var}(X_8) = \mathrm{var}(X_4) + \mathrm{var}(X_4) = 8$.
   (c) The covariance is zero since X_3 and X_4 are independent.
   (d) By the Central Limit Theorem, $\bar{X} \approx \mathrm{N}\left(\mathbb{E}(X_1), \frac{\mathrm{var}(X_1)}{9}\right) = \mathrm{N}\left(7, \frac{4}{9}\right)$.

2. (a) $\mathbb{E}(X_1) = \int_{-1}^{1} x(3/2)x^2\, dx = 0$.
   (b) First we need to know $\mathrm{var}(X_1)$, which we calculate as follows:

$$\mathbb{E}(X_1^2) = \int_{-1}^{1} x^2(3/2)x^2\, dx = \int_{-1}^{1} (3/2)x^4\, dx = \left[\frac{3}{10}x^5\right]_{-1}^{1} = \frac{3}{5},$$
$$\mathrm{var}(X_1) = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 = \frac{3}{5}.$$

   Then we do the calculations for Y,

$$\mathbb{E}(Y) = 15 \cdot \mathbb{E}(X_1) = 0,$$
$$\mathrm{var}(Y) = 15 \cdot \mathrm{var}(X_1) = 9.$$

   (c) The CLT implies $\bar{X}$ is approximately normally distributed. Since $Y = 15\bar{X}$ then Y will also be approximately normally distributed. Therefore, using the results from above, $Y \approx \mathrm{N}(0, 9)$. Thus,

$$\begin{aligned}\Pr(-0.3 < Y < 1.5) &= \Pr\left(\frac{-0.3-0}{3} < \frac{Y-0}{3} < \frac{1.5-0}{3}\right)\\ &\approx \Pr(-0.1 < Z < 0.5)\\ &= \Phi(0.5) - \Phi(-0.1) = 0.23\end{aligned}$$

3. (a) Yes. Tingjin has tried to run each experiment under the same conditions and each pot is separate to the others. Therefore it is reasonable to assume the measurements are iid. The 'population' here is the (hypothetically infinite) set of possible replicates of this particular experiment, i.e. imaginary extra pots prepared in the same way.
   (b) The family members are, presumably (since we aren't give any further information), a mixture of adults and children, which means their heights will not be identically distributed. Furthermore, the family members will be genetically related (again, a resonable assumption without being told otherwise), which means they will also not be independent. Therefore, this is not a random sample.
   The population of interest is also not clear from the given description. Was Damjan only interested in the heights of his family members? In that case, he has obtained **all** of the measurements (i.e. the whole population), so we wouldn't usually describe this as a sample. Was he interested in heights of people in general, e.g. the population of all humans? In that case, this **is** a sample from that population, but not a random one.

(c) The population of interest here is, presumably, the number of people sitting down on South Lawn at any point in time. However, this should really be specified more precisely, e.g. does Robert care about what happens at 2am?

For simplicity, suppose Robert does his sampling at the same time each day (e.g. noon), then it is better to think of the population as being just the number of people sitting down at that time of the day, for an arbitrary day. Then, for this to be a random sample, we need to assume they are iid. This is unlikely to be the case: weekends and weekdays are likely to differ, and also semester time versus non-semester time.

Suppose Robert did his sampling at different times of the day (and we therefore use the more expanded notion of the population). Then there we will be an even greater deviation from the iid assumption, since we expect this count to vary substantially across the day.

4. (a) i. For the sample mean:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0$$

For the sample variance:

$$\begin{aligned} s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \end{aligned}$$

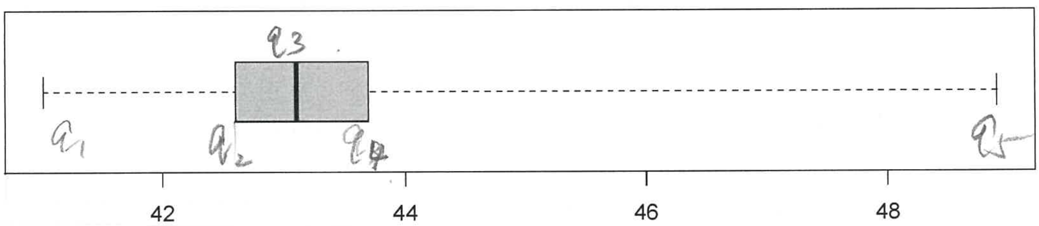
ii. $\bar{x} = 219.3/5 = 43.86$, $s = \sqrt{(1/4)(9654.27 - 5 \times 43.86^2)} = 2.99$

(b) i. The identity for the sample mean stays the same. The identity for the sample variance still holds, with the 'new' value of s^2 being 4 times greater than the 'old' value.

ii. Both the mean and the standard deviation will be double their previous values.

(c) i. The ordered data are: 41.0, 42.6, 43.1, 43.7, 48.9. The median is the 3rd ordered observation, 43.1. The first and third (Type 7) quartiles are the 2nd and 4th order statistics, 42.6 and 43.7 respectively. The five-number summary therefore is {41.0, 42.6, 43.1, 43.7, 48.9}.

ii. IQR = 43.7 − 42.6 = 1.1.



iii.

(d) i. The sample median is the same under this definition, but the 1st and 3rd quartiles differ:

$$(5+1)\cdot 0.25 = 1 + 0.5 \Rightarrow \tilde{\pi}_{0.25} = x_{(1)} + 0.5 \cdot (x_{(2)} - x_{(1)}) = 41.8$$
$$(5+1)\cdot 0.75 = 4 + 0.5 \Rightarrow \tilde{\pi}_{0.75} = x_{(4)} + 0.5 \cdot (x_{(5)} - x_{(4)}) = 46.3$$

两数的一半

Therefore, the five-number summary is $\{41.0, 41.8, 43.1, 46.3, 48.9\}$.

ii. Let $k = i + r$. Non-integer order statistics were defined by linear interpolation,

$$x_{(k)} = x_{(i+r)} = x_{(i)} + r \cdot (x_{(i+1)} - x_{(i)})$$

In other words, $x_{(k)} = \tilde{\pi}_p$ where $(n+1)p = k$. Rearranging the latter gives $p = k/(n+1)$, which is the definition of a 'Type 6' quantile.

(e) Using the Type 7 quantiles: $\hat{\pi}_{0.25} - 1.5 \times \text{IQR} = 40.95$ and $\hat{\pi}_{0.75} + 1.5 \times \text{IQR} = 45.35$. Since the observation 48.9 lies outside of these two extremes, it would be classified as an outlier. Repeating this calculation with Type 6 quantiles gives a larger IQR and no observations end up being classified as outliers.

5. The set $\{1, 1, 4, 4\}$ maximises the sample variance.

6. (a)
```
x <- c(10.39, 10.43, 9.99, 11.17,  8.91,
       11.20, 11.38, 7.74, 10.61, 11.11)
quantile(x, type = 6)
```

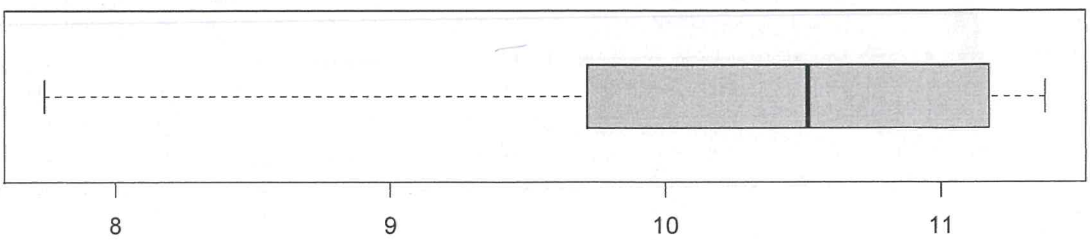
```
##       0%      25%      50%      75%     100%
##  7.7400   9.7200  10.5200  11.1775  11.3800
```

```
quantile(x, type = 7)
```

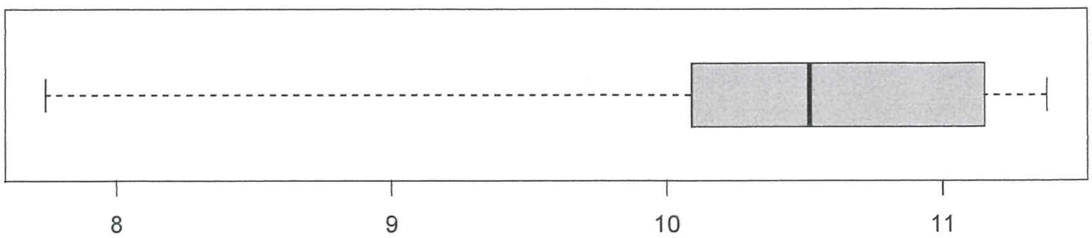
```
##     0%     25%    50%    75%   100%
##  7.740 10.090 10.520 11.155 11.380
```

(b) If using Type 6 quantiles, there are no outliers. If using Type 7 quantiles, then the observation 7.74 is an outlier.

(c) With Type 6 quantiles:

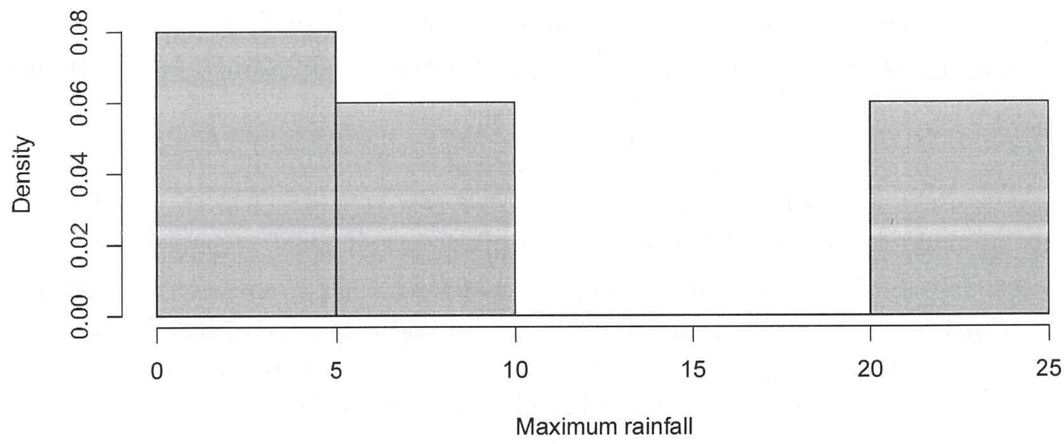


With Type 7 quantiles:



7. (a)

```
x <- c(9.9, 4.7, 20.5, 1.8, 4.7, 9.8, 20.5, 20.2, 6.5, 3.0)
hist(x, xlab = "Maximum rainfall", freq = FALSE, main = NULL, col = 8)
```



(b) Define $p_k = k/(n+1)$, for $k = 1, \ldots, n$. These have values:

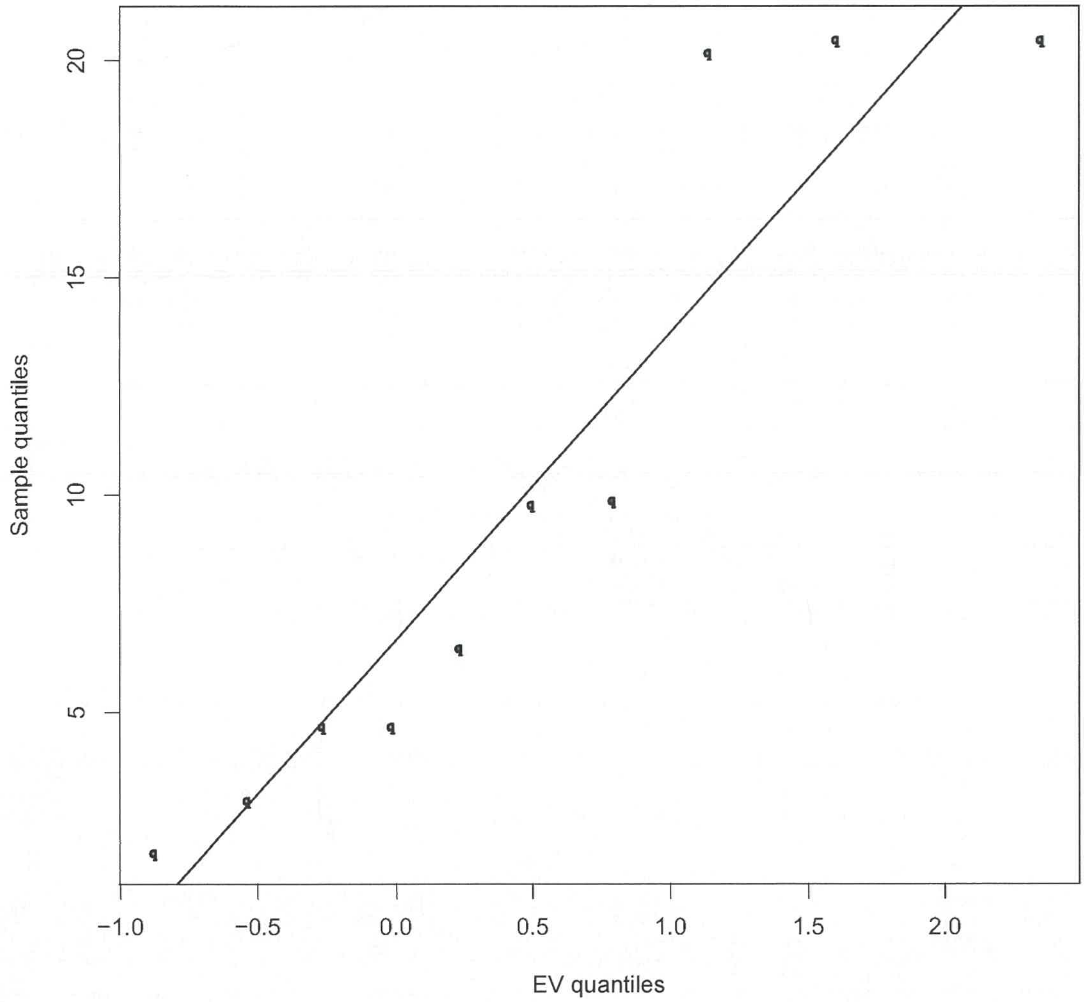
$$p_1 = 1/11 = 0.09, \ldots, p_{10} = 10/11 = 0.91$$

The theoretical quantiles are:

$$F^{-1}(p_1) = -0.9, \ldots, F^{-1}(p_{10}) = 2.4$$

We plot these against the corresponding sample quantiles, which are just the order statistics $(1.8, \ldots, 20.5)$. Some R code to make the plot:

```
p <- 1:10/11
Finv <- -log(-log(p))
x.sorted <- sort(x)
plot(Finv, x.sorted, pch = 19,
     xlab = "EV quantiles",
     ylab = "Sample quantiles")
# Adds best fitting line
fit <- lm(x.sorted ~ Finv)     ?
abline(fit)
```




```
coef(fit)
```

```
## (Intercept)        Finv
##    6.658273    7.071246
```

To estimate θ and ξ we can use the intercept and slope of the best fitting line computed according to some appropriate method. Based on the analysis above we obtain $\hat{\theta} = 6.66$ and $\hat{\xi} = 7.07$.

MAST20005/MAST90058: Week 3 Solutions

For all of the solutions below, we use the notation $\ell(\theta) = \ln L(\theta)$ for log-likelihood functions and $s(\theta) = \frac{\partial \ell}{\partial \theta}$ for their first derivatives.

1. (a)

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i - \mu)^2}$$

$$\ell(\mu) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$s(\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) \quad \boxed{?}$$

Setting $s(\mu) = 0$ and solving gives $\hat{\mu} = \bar{X}$.

(b) i.

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

$$\ell(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln\lambda - \ln\prod_{i=1}^{n} x_i!$$

$$s(\lambda) = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i$$

Setting $s(\lambda) = 0$ and solving gives $\hat{\lambda} = \bar{X}$.

ii. $\bar{x} = (5\cdot 0 + 7\cdot 1 + 12\cdot 2 + 9\cdot 3 + 5\cdot 4 + 1\cdot 5 + 1\cdot 6)/40 = 2.225$

(c) i.

$$L(\theta) = \left(\frac{1}{\theta^2}\right)^n \prod_{i=1}^{n} x_i \exp(-x_i/\theta)$$

$$\ell(\theta) = -2n\ln(\theta) + \sum_{i=1}^{n}\ln(x_i) - \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

$$s(\theta) = -\frac{2n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i$$

Setting $s(\theta) = 0$ and solving gives the estimator $\hat{\theta} = \sum_{i=1}^{n} X_i/(2n) = \bar{X}/2$.

ii.

$$L(\theta) = \left(\frac{1}{2\theta^3}\right)^n \prod_{i=1}^{n} x_i^2 \exp(-x_i/\theta)$$

$$\ell(\theta) = -n\ln 2 - 3n\ln(\theta) + \sum_{i=1}^{n} 2\ln(x_i) - \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

$$s(\theta) = -\frac{3n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i$$

Setting $s(\theta) = 0$ and solving gives the estimator $\hat{\theta} = \sum_{i=1}^{n} X_i/(3n) = \bar{X}/3$.

iii.

$$L(\theta) = \left(\frac{1}{2}\right)^n \prod_{i=1}^{n} \exp(-|x_i - \theta|)$$

$$\ell(\theta) = -n \ln 2 - \sum_{i=1}^{n} |x_i - \theta|$$

$$s(\theta) = \sum_{i=1}^{n} \operatorname{sgn}(x_i - \theta)$$

where $\operatorname{sgn}(\cdot)$ is the sign function: $\operatorname{sgn}(x) = 1$ if $x > 0$, $\operatorname{sgn}(x) = -1$ if $x < 0$, and $\operatorname{sgn}(0) = 0$. Note that $\ell(\theta)$ is piecewise linear and not differentiable when θ equals any x_i, so $s(\theta)$ is not defined at those points. If n is even, $s(\theta)$ is zero when there are an equal number of positive and negative signs, so $\hat{\theta}$ is between the middle two ordered values (i.e. any value between these will maximise the likelihood), and we would typically pick their average. If n is odd, for $s(\theta)$ to be zero $\hat{\theta}$ must equal the middle value. So, in general, $\hat{\theta}$ is the sample median.

2. The population mean and variance are $\mathbb{E}(X) = 5\theta/4$ and $\operatorname{var}(X) = (7\theta/4) - (5\theta/4)^2$.

   (a) We know that $\mathbb{E}(\bar{X}) = (5/4)\,\theta$ and therefore an unbiased estimator of θ based on $\bar{X}$ is $T_1 = (4/5)\bar{X}$.
   Note that $Z \sim \mathrm{Bi}(n, 1-\theta)$, which means that $\mathbb{E}(Z) = n(1-\theta)$ and $\mathbb{E}(Z/n) = 1-\theta$. Therefore, an unbiased estimator of θ based on Z is $T_2 = 1 - Z/n$.

   (b) Calculating the variance of the above estimators gives:

$$\operatorname{var}(T_1) = \frac{\theta\left(\frac{28}{25} - \theta\right)}{n}, \quad \operatorname{var}(T_2) = \frac{\theta(1-\theta)}{n}$$

?

   We can therefore see that $\operatorname{var}(T_1) > \operatorname{var}(T_2)$.

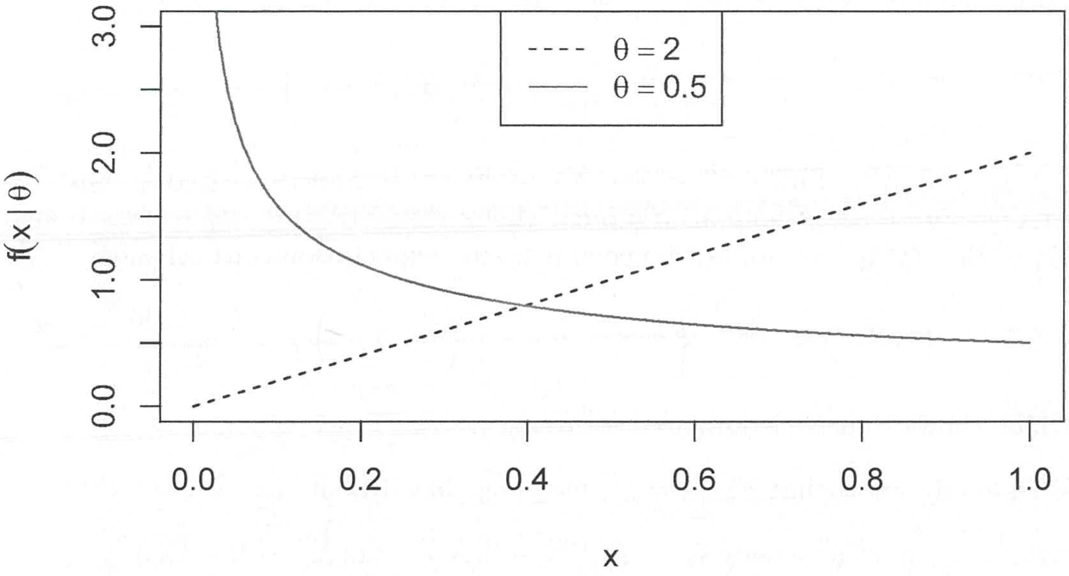
3. (a)
```
f1 <- function(x) {2   * x^(2   - 1)}
f2 <- function(x) {0.5 * x^(0.5 - 1)}
curve(f1, 0, 1, col = 1, lty = 2, ylim = c(0, 3),
      ylab = expression(f(x ~ "|" ~ theta)))
curve(f2, 0, 1, col = 2, lty = 1, add = TRUE)
legend("top", c(expression(theta == 2), expression(theta == 0.5)),
       col = c(1, 2), lty = c(2, 1))
```

$E(\frac{4}{5}X) = \theta$. $Var(\frac{4}{5}\bar{X}) = E(\frac{4}{5}X)^2 - E(\frac{4}{5}X)^2 = (\frac{28}{25}\theta - \theta^2) = \dfrac{\theta(\frac{28}{25} - \theta)}{n}$

$E((\frac{4}{5}X)^2) = E(\frac{16}{25}X^2) = \frac{16}{25}E(X^2) = \frac{16}{25} \times \frac{7}{4}\theta = \frac{28}{25}\theta$

~~$E(T_2) = E(1-\frac{Z}{n})$~~ $Var = np(1-p)$. $Z \sim Bi(n, 1-\theta)$ $\frac{1}{n}$

$Var(T_2) = \dfrac{n \cdot (1-\theta) \cdot \theta}{n^2} = \dfrac{(1-\theta)\theta}{n}$ $\quad \dfrac{Z}{n}$



(b)

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

$$\ell(\theta) = n \ln \theta + (\theta - 1) \sum_{i=1}^{n} \ln(x_i)$$

$$s(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \ln(x_i)$$

Setting $s(\theta) = 0$ and solving gives the estimator:

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \ln X_i}$$

(c) The MLEs are: $\hat{\theta}_X = 0.549$, $\hat{\theta}_Y = 2.210$, $\hat{\theta}_Z = 0.959$.
To find the method of moments estimator, we need to solve $\bar{X} = \theta/(\theta + 1)$ which gives $\tilde{\theta} = \bar{X}/(1 - \bar{X})$. Therefore, the MM estimates are: $\tilde{\theta}_X = 0.598$, $\tilde{\theta}_Y = 2.400$ and $\tilde{\theta}_Z = 0.865$.

4. Recall that for a random sample $X_1, \ldots, X_n$,

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = E(X_i)$$

and

$$\boxed{\mathrm{var}(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{var}(X_i) = \frac{\mathrm{var}(X_i)}{n}.}$$

??

(a) The X_i are iid exponential random variables with mean θ. Therefore, $\mathbb{E}(\bar{X}) = \mathbb{E}(X_i) = \theta$ and $\bar{X}$ is unbiased.

(b) We know that $\text{var}(X_i) = \theta^2$. Therefore, $\text{var}(\bar{X}) = \text{var}(X_i)/n = \theta^2/n$.

(c) Based on the above, an estimate of θ is $\hat{\theta} = \bar{x} = 3.48$.

5. From question 4(a)(i) from week 2, we know that:

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right).$$

Let $\mu = \mathbb{E}(X)$. Since $\sigma^2 = \text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mu^2$, we see that $\mathbb{E}(X^2) = \sigma^2 + \mu^2$. A similar argument shows that $\mathbb{E}(\bar{X}^2) = \sigma^2/n + \mu^2$. Using the above expression for the sample variance and taking expectations of both sides, ??

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

which shows that S^2 is unbiased for σ^2.

6. We already know that $\mathbb{E}(S^2) = \theta^2$, meaning that it is unbiased. Note that,

$$\theta^2 = \text{var}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = \mathbb{E}(X_i^2) - 0 = \mathbb{E}(X_i^2).$$

Therefore,

$$\mathbb{E}(\hat{\theta}^2) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i^2) = \frac{1}{n}\sum_{i=1}^{n}\theta^2 = \theta^2,$$

meaning that it is also unbiased.

To derive the variance of the estimator, first note that,

$$\text{var}(X_i^2) = \mathbb{E}(X_i^4) - \mathbb{E}(X_i^2)^2 = \mathbb{E}(X_i^4) - \theta^4.$$

The 4th moment for a normal distribution (easy to look up) is,

$$\mathbb{E}(X^4) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4.$$

Here we have $\mu = 0$ and $\sigma^2 = \theta^2$, so we have,

$$\mathbb{E}(X_i^4) = 3\theta^4.$$

Therefore we have,

$$\text{var}(X_i^2) = 2\theta^4.$$??

(After covering Module 3, you will learn that another way to do this is to note that $X_i^2/\theta^2 \sim \chi_1^2$ and therefore $\text{var}(X_i^2/\theta^2) = 2$.) Now we derive the variance of the estimator,

$$\text{var}(\hat{\theta}^2) = \frac{1}{n^2}\sum_i \text{var}(X_i^2) = \frac{2\theta^4}{n}.$$

Also, we know that,

$$\text{var}(S^2) = \frac{2\theta^4}{n-1}.$$

Therefore, $\text{var}(\hat{\theta}^2) < \text{var}(S^2)$ for any $n > 1$.

7. (a) We know this result from the lectures, but to show you some of the derivation: since $\text{var}(X_i) = \sigma^2$ for all $i = 1, \ldots, n$ and the observations are independent, we have $\text{var}(\bar{X}) = (\sigma^2 + \cdots + \sigma^2)/n^2 = \sigma^2/n$.

(b) According to the information given, $\mathrm{var}(\hat{\pi}_{0.5}) \approx \pi/2 \times \sigma^2/n$. Since $\sigma^2/n < \pi/2 \times \sigma^2/n$, we see that the sample mean has smaller variance than the sample median.

(c) We already know that the sample mean is unbiased. According to the information given, $\mathbb{E}(\hat{\pi}_{0.5}) \approx \pi_{0.5} = \mu$, so the sample median is at least approximately unbiased.

(d) Both estimators are unbiased (exactly or approximately) but the sample mean has smaller variance, so we would expect it to be more accurate.

MAST20005/MAST90058: Week 4 Solutions

1. (a) We use the sample mean and sample median as estimators for the population mean. From the R output, $\bar{x} = 10.29$ and $\hat{\pi}_{0.5} = 10.52$.

   (b) From problem 7 in week 3, we know that:

$$\mathrm{sd}(\bar{X}) = \sqrt{\mathrm{var}(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

$$\mathrm{sd}(\hat{\pi}_{0.5}) = \sqrt{\mathrm{var}(\hat{\pi}_{0.5})} = \sqrt{\frac{\pi}{2}}\frac{\sigma}{\sqrt{n}}$$

   We can approximate both of these by substituing $s = 1.159$ for σ (recall that S is an estimator for σ). This gives standard errors for each of the two estimates:

$$\mathrm{se}(\bar{x}) = \frac{s}{\sqrt{n}} = 0.367$$

$$\mathrm{se}(\hat{\pi}_{0.5}) = \sqrt{\frac{\pi}{2}}\frac{s}{\sqrt{n}} = 0.459$$

2. $73.8 \pm 1.96 \times 5/4 = [71.35, 76.25]$

3. $\bar{X}$ is approximately normally distributed. Using this gives:
   $2.09 \pm 1.96 \times 0.12/4 = [2.03, 2.15]$

4. $\bar{X}$ is approximately normally distributed. Using this gives:
   $11.95 \pm 1.96 \times 11.8/\sqrt{37} = [8.148, 15.75]$
   Some people prefer to use the t-distribution approximation, which is more conservative. Since the necessary quantiles were provided, let's try it:
   $11.95 \pm 2.028 \times 11.8/\sqrt{37} = [8.016, 15.88]$
   Note that the two are fairly similar.

5. $20.9 \pm 2.306 \times 1.858/3 = [19.47, 22.33]$.
   One sensible interpretation of the claim is that the *average* weight of a '22 kg' wheel is 22 kg (rather than claiming that *every* wheel is *exactly* 22 kg in weight). Since 22 is within our confidence interval for the mean, and the interval is relatively narrow, this claim seems to be reasonable given our data.

6. $937.4 - 988.9 \pm 1.96\sqrt{784/56 + 627/57} = [-61.3, -41.7]$.
   This confidence interval is very far from zero, meaning that our data show fairly strong evidence that the mean lifetimes are *not* the same.

MAST20005/MAST90058: Week 5 Solutions

1. $\left[\sqrt{\frac{12}{23.34}} \times 37.751, \sqrt{\frac{12}{4.404}} \times 37.751\right] = [4.41, 10.1]$

2. (a) The pooled estimate of the standard deviation is

$$s_p = \sqrt{\frac{9 \times 0.323^2 + 9 \times 0.210^2}{18}} = 0.2724.$$

Hence a 95% confidence interval is

$$2.548 - 1.564 \pm 2.101 \times 0.2724\sqrt{\frac{1}{10} + \frac{1}{10}} = [0.728, 1.240].$$

(b) Yes. The confidence interval is quite far from zero so we have good evidence that the mean force required to remove the seal when the wedge is in place is larger than when it is not.

(c) Inspect a box plot or compute a CI for $\mathrm{var}(X)/\mathrm{var}(Y)$.

3. A 90% confidence intervals for σ_x/σ_y is

$$\left[\sqrt{0.3821} \times \frac{0.197}{0.318}, \sqrt{2.475} \times \frac{0.197}{0.318}\right] = [0.383, 0.975].$$

This interval lies completely below 1, so we have good evidence that the standard deviations *differ* between the two shifts.

4. (a) $\hat{p} = 24/642 = 0.0374$

(b) $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n} = [0.0227, 0.0521]$

(c) Upper bound: $\hat{p} + 1.645\sqrt{\hat{p}(1-\hat{p})/n} = 0.0497$

5. Need $1.96 \times \sqrt{4.84}/\sqrt{n} = 0.4 \Rightarrow n = (1.96^2 \times 4.84/0.4^2) = 116.2$, so we take $n = 117$.

6. The error is $\epsilon = c\sqrt{p(1-p)/n}$, which gives the formula:

$$n = \frac{c^2 p(1-p)}{\epsilon^2}.$$

We don't know the value of p, but the maximum error occurs when $p = 0.5$ so we use that value for calculating the sample size. For 95% CIs we use $c = \Phi^{-1}(1-0.025) = 1.960$ and for 90% CIs we use $c = \Phi^{-1}(1-0.05) = 1.645$. In each case, we always round up to the nearest integer to get a valid sample size. This gives:

(a) 1068

(b) 2401

(c) 752

7. From the sample we have $n = 10$, $\bar{x} = 30.84$ and $s = 2.908$.

(a) $\hat{\mu} = 30.8$; 90% CI for μ: $30.84 \pm 1.833 \times 2.908 \times \sqrt{\frac{1}{10}} = [29.2, 32.5]$.

(b) $\hat{\sigma} = 2.91$; 95% CI for σ: $\left[\sqrt{\frac{9}{19.02}} \times 2.908, \sqrt{\frac{9}{2.700}} \times 2.908\right] = [2.00, 5.31]$.

(c) 90% PI for X: $30.84 \pm 1.833 \times 2.908 \times \sqrt{\frac{11}{10}} = [25.2, 36.4]$.

8. (a) Using the CLT approximation for $\hat{p}$,

$$\Pr\left(p - c\sqrt{\frac{p(1-p)}{n}} < \hat{p} < p + c\sqrt{\frac{p(1-p)}{n}}\right) \approx 0.95$$

where $c = \Phi^{-1}(1 - \alpha/2)$.

(b) Rearranging the above inequality gives,

$$-c\sqrt{\frac{p(1-p)}{n}} < \hat{p} - p < c\sqrt{\frac{p(1-p)}{n}}.$$

Since the middle term is smaller in absolute value than both endpoints, which are equal in absolute value, we ~~can~~ square all of the terms and write:

$$(\hat{p} - p)^2 < c^2 \frac{p(1-p)}{n}.$$

(c) Expanding both sides gives,

$$p^2 - 2p\hat{p} + \hat{p}^2 < -\frac{c^2}{n}p^2 + \frac{c^2}{n}p.$$

Moving all terms to one side and collecting them together based on p,

$$\left(1 + \frac{c^2}{n}\right)p^2 - 2\left(\hat{p} + \frac{c^2}{2n}\right)p + \hat{p}^2 < 0.$$

This is a quadratic in p. You can verify that it has two roots by calculating the discriminant and showing it is positive. Let the roots be a and b. Since the leading ~~coefficient is~~ positive (i.e. it is an 'upward' parabola), the inequality will be satisfied when $a < p < b$. We can calculate the roots either by completing the square and factorising, or by applying the quadratic formula. Either method will give you the required endpoints.

MAST20005/MAST90058: Week 6 Solutions

1. Note that $K = \sum_i (x_i - \bar{x})^2 = (n-1)s_x^2 = 594.9896$. A 95% confidence interval for β is $\hat{\beta} \pm c\hat{\sigma}/\sqrt{K}$, where c is the 0.975 quantile of t_{60}. In this case we get $0.75169 \pm 2.00 \times 0.6943/\sqrt{594.9896}$ or $(0.69, 0.81)$.

2. (a) Use the double expectation formula,

$$\mu_3 = \mathbb{E}(X_3) = \mathbb{E}(\mathbb{E}(X_3 \mid X_1, X_2)) = \alpha + \beta_1 \mathbb{E}(X_1 - \mu_1) + \beta_2 \mathbb{E}(X_2 - \mu_2) = \alpha.$$

(b) Note that $\sigma_{i3} = \text{cov}(X_i, X_3) = \mathbb{E}(X_i X_3) - \mathbb{E}(X_i)\mathbb{E}(X_3) = \mathbb{E}(X_i X_3) - \mu_i \mathbb{E}(X_3) = \mathbb{E}(X_i X_3) - \mathbb{E}(\mu_i X_3) = \mathbb{E}(X_i X_3 - \mu_i X_3) = \mathbb{E}(X_3(X_i - \mu_i))$. Using this together with the double expectation formula,

$$\begin{aligned}
\sigma_{13} &= \mathbb{E}(X_3(X_1 - \mu_1)) \\
&= \mathbb{E}\left[\mathbb{E}(X_3(X_1 - \mu_1) \mid X_1, X_2)\right] \\
&= \mathbb{E}\left[\mathbb{E}(X_3 \mid X_1, X_2)(X_1 - \mu_1)\right] \\
&= \mathbb{E}\left[\alpha(X_1 - \mu_1) + \beta_1(X_1 - \mu_1)^2 + \beta_2(X_2 - \mu_2)(X_1 - \mu_1)\right] \\
&= \beta_1\sigma_1^2 + \beta_2\sigma_{12}
\end{aligned}$$

and similarly,

$$\sigma_{23} = \beta_1\sigma_{12} + \beta_2\sigma_2^2.$$

Simultaneously solving these two equations gives the result.

3. (a) Firstly, adding and subtracting the parameter estimate terms:

$$\sum_{i=1}^{n}[Y_i - \alpha_0 - \beta(x_i - \bar{x})]^2 = \sum_{i=1}^{n}[(\hat{\alpha}_0 - \alpha_0) + (\hat{\beta} - \beta)(x_i - \bar{x}) + Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x})]^2.$$

In expanding this, consider the cross-terms and recall that $\hat{\alpha}_0 = \bar{Y}$ and $\hat{\beta} = \sum(x_i - \bar{x})(Y_i - \bar{Y})/\sum(x_i - \bar{x})^2$.

$$\sum_{i=1}^{n}(\hat{\alpha}_0 - \alpha_0)(\hat{\beta} - \beta)(x_i - \bar{x}) = (\hat{\alpha}_0 - \alpha_0)(\hat{\beta} - \beta)\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

$$\sum_{i=1}^{n}(\hat{\alpha}_0 - \alpha_0)\{Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x})\} = (\hat{\alpha}_0 - \alpha_0)\left(\sum_{i=1}^{n}(Y_i - \bar{Y}) - \hat{\beta}\sum_{i=1}^{n}(x_i - \bar{x})\right) = 0$$

$$\begin{aligned}
\sum_{i=1}^{n}(\hat{\beta} - \beta)(x_i - \bar{x})\{Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x})\} &= (\hat{\beta} - \beta)\left(\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y}) - \hat{\beta}\sum_{i=1}^{n}(x_i - \bar{x})^2\right) \\
&= 0
\end{aligned}$$

Note that we substituted for $\hat{\beta}$ in the last equation. Therefore the cross-terms disappear and we are left with,

$$\sum_{i=1}^{n}[Y_i - \alpha_0 - \beta(x_i - \bar{x})]^2 = n(\hat{\alpha}_0 - \alpha_0)^2 + (\hat{\beta} - \beta)^2\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}[Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x})]^2$$

as required.

(b) We know that $(\hat{\alpha}_0 - \alpha_0)/\sqrt{\hat{\sigma}^2/n} \sim t_{n-2}$. Hence,

$$\Pr\left(-c < \frac{\hat{\alpha}_0 - \alpha_0}{\sqrt{\hat{\sigma}^2/n}} < c\right) = 1 - \gamma$$

where c is the $1 - \gamma/2$ quantile of t_{n-2}. Rearrangement then gives the desired confidence interval.

(c) We know $(n-2)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-2}$. Therefore,

$$\Pr\left(F^{-1}(\gamma/2) < \frac{(n-2)\hat{\sigma}^2}{\sigma^2} < F^{-1}(1-\gamma/2)\right) = 1 - \gamma$$

which gives

$$\Pr\left(\frac{(n-2)\hat{\sigma}^2}{F^{-1}(1-\gamma/2)} < \sigma^2 < \frac{(n-2)\hat{\sigma}^2}{F^{-1}(\gamma/2)}\right) = 1 - \gamma$$

as required.

4. A linear model needs to be linear in the coefficients β_1 and β_2, so this is not a linear model.

5. Firstly,

$$\frac{\partial h}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i - \beta_3 x_i^2)$$

$$\frac{\partial h}{\partial \beta_2} = -2\sum_{i=1}^{n} x_i(y_i - \beta_1 - \beta_2 x_i - \beta_3 x_i^2)$$

$$\frac{\partial h}{\partial \beta_1} = -2\sum_{i=1}^{n} x_i^2(y_i - \beta_1 - \beta_2 x_i - \beta_3 x_i^2).$$

Setting these to zero and simplifying gives the normal equations.

6. (a) Firstly,

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{n}(x_i-\bar{x})y_i - \bar{y}\sum_{i=1}^{n}(x_i-\bar{x}) = \sum_{i=1}^{n}(x_i-\bar{x})y_i - \bar{y}\times 0 = \sum_{i=1}^{n}(x_i-\bar{x})y_i.$$

A similar argument shows that,

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{n} x_i(y_i-\bar{y}).$$

We also have,

$$\sum_{i=1}^{n} x_i(y_i-\bar{y}) = \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\bar{y} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}.$$

That completes the proof.

(b)

$$
\begin{aligned}
d^2 &= \sum_{i=1}^{n}(y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x}))^2 = \sum_{i=1}^{n}(y_i - \bar{y} - \hat{\beta}(x_i - \bar{x}))^2 \\
&= \sum_{i=1}^{n}(y_i - \bar{y})^2 - 2\hat{\beta}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + \hat{\beta}^2\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \sum_{i=1}^{n}(y_i - \bar{y})^2 - 2\hat{\beta}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + \hat{\beta}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\
&= \sum_{i=1}^{n}(y_i - \bar{y})^2 - \hat{\beta}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\
&= \sum_{i=1}^{n}(y_i - \bar{y})^2 - \frac{\left[\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right]^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.
\end{aligned}
$$

7. (a) $\hat{\beta} = \frac{741.1}{578.8} = 1.28$,
   $\hat{\alpha} = 313.2/13 - \hat{\beta} \times 230/13 = 1.44$,
   $\hat{\sigma} = \sqrt{(1000.2 - 741.1^2/578.8)/11} = 2.16$
   (b) $\text{se}(\hat{\alpha}) = \hat{\sigma}\sqrt{1/13 + (230/13)^2/578.8} = 1.70$,
   $\text{se}(\hat{\beta}) = \hat{\sigma}/\sqrt{578.8} = 0.898$
   (c) 95% CI for $\mu(18)$: $\hat{\mu}(18) \pm c \times \text{se}(\hat{\mu}(3))$
   $\hat{\mu}(18) = \hat{\alpha} + \hat{\beta} \times 18 = 24.48$
   $\text{se}(\hat{\mu}(18)) = 2.16 \times \sqrt{1/13 + (18 - 230/13)^2/578.8} = 0.278$
   $c = 2.201$ (0.975 quantile of t_{11})
   This gives the following interval: $23.2 < \mu(18) < 25.8$
   (d) 95% PI for $Y(18)$:

$$
24.48 \pm 2.201 \times 2.16 \times \sqrt{1 + \frac{1}{13} + \frac{(18 - 230/13)^2}{578.8}}
$$

   This gives the following interval: $19.6 < Y(18) < 29.4$

MAST20005/MAST90058: Week 7 Solutions

1. $\alpha = \Pr(X \in \{2,3\} \mid p = 1/3) = 0.22222 + 0.03703 = 0.25926$
   $\beta = \Pr(X \in \{0,1\} \mid p = 2/3) = 0.03703 + 0.22222 = 0.25926$

2. (a) Assuming H_0 gives $\mathbb{E}(Y) = 8$ and $\text{var}(Y) = 7.36 = 2.713^2$. Using a normal approximation,

$$\alpha = \Pr(Y \leqslant 6 \mid p = 0.08) \approx \Pr(Z < \frac{6-8}{2.713}) = \Phi(-0.737) = 0.23.$$

In this case we should ideally be using $\boxed{\text{continuity correction}}$ because it makes a noticeable difference, ??

$$\alpha = \Pr(Y \leqslant 6 \mid p = 0.08) \approx \Pr(Z < \frac{6.5-8}{2.713}) = \Phi(-0.553) = 0.29.$$

   (b) When $p = 0.04$, we have $\mathbb{E}(Y) = 4$ and $\text{var}(Y) = 3.84 = 1.96^2$. Using a normal approximation,

$$\alpha = \Pr(Y \geqslant 7 \mid p = 0.04) \approx \Pr(Z > \frac{7-4}{1.96}) = \Pr(Z > 1.531) = 1 - \Phi(1.531) = 0.063.$$

If we use continuity correction we get,

$$\alpha = \Pr(Y \geqslant 7 \mid p = 0.04) \approx \Pr(Z > \frac{6.5-4}{1.96}) = \Pr(Z > 1.276) = 1 - \Phi(1.276) = 0.10.$$

3. Under H_0, $\mathbb{E}(Y) = 14.63$ and $\text{var}(Y) = 13.606 = 3.689^2$. Hence,

$$z = \frac{23 - 14.63}{3.689} = 2.269$$

   (a) $z > 1.645$ so reject H_0 at 5% level of significance.
   (b) $z < 2.326$ so don't reject H_0 at the 1% level of significance.
   (c) The p-value is $\Pr(Z \geq 2.269) = 1 - \Phi(2.269) = 1 - 0.9883 = 0.0117$

4. $\hat{p}_m = 124/894 = 0.1387$ and $\hat{p}_f = 70/700 = 0.1$, which gives,

$$\hat{p}_m - \hat{p}_f \pm 1.96\sqrt{\frac{\hat{p}_m(1-\hat{p}_m)}{n_m} + \frac{\hat{p}_f(1-\hat{p}_f)}{n_f}} = (0.007, 0.07).$$

Under H_0, $\hat{p} = 194/1594 = 0.1217$. Reject H_0 if $|z| > 1.96$.

$$|z| = \frac{|\hat{p}_m - \hat{p}_f|}{\sqrt{\hat{p}(1-\hat{p})\,(1/n_m + 1/n_f)}} = 2.345 > 1.96$$

so we reject H_0.

5. (a) The critical region is:

$$T = \frac{\bar{X} - 47}{S/\sqrt{20}} < -1.729 \quad (0.05 \text{ quantile of } t_{19})$$

   (b) $t = (46.94 - 47)/(0.15/\sqrt{20}) = -1.789$. This is less than -1.729 so we reject H_0.

(c) Comparing the test statistic to the quantiles of t_{19} provided, we can deduce that the p-value is between 0.025 and 0.05.

6. (a) $H_0\colon \mu = 1.9$
   (b) $H_1\colon \mu \neq 1.9$
   (c) $|T| = |\bar{X} - 1.9|/(S/3) > 2.306$
   (d) $|t| = |2.05 - 1.9|/(0.17/3) = 2.647$
   (e) $|t| > 2.306$ so we reject H_0.
   (f) $2.306 < 2.647 < 2.896$ so the area of one tail (i.e. extreme values in one direction) is between 0.01 and 0.025. Since we have a two-sided alternative, the p-value will be double this, so we have: $0.02 <$ p-value < 0.05.

7. (a) The critical region is

$$\chi^2 = \frac{19S^2}{(0.095)^2} < 10.117$$

   and the observed value is

$$\chi^2 = \frac{19 \times (0.065)^2}{(0.095)^2} = 8.895$$

   so we reject H_0 and conclude there is evidence that the company was successful.
   (b) Since the 0.025 quantile of χ^2_{19} is $8.906 \approx 8.895$, the p-value is approximately 0.025.

8. (a) The critical region is given by:

$$|T| = \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{12S_X^2 + 15S_Y^2}{27}\left(\frac{1}{13} + \frac{1}{16}\right)}} > 2.052 \quad (0.975 \text{ quantile of } t_{27})$$

   (b) The observed value is:

$$|t| = \frac{|72.9 - 81.7|}{\sqrt{\frac{12 \times 25.6^2 + 15 \times 28.3^2}{27}\left(\frac{1}{13} + \frac{1}{16}\right)}} = 0.869 < 2.052$$

   so we cannot reject H_0.
   (c) $0.684 < 0.869 < 1.314$ so the area of one tail (i.e. extreme values in one direction) is between 0.1 and 0.25. Since we have a two-sided alternative, the p-value will be double this, so we have: $0.2 <$ p-value < 0.5.
   (d) The test of interest is $H_0\colon \sigma_X^2 = \sigma_Y^2$ against $H_1\colon \sigma_X^2 \neq \sigma_Y^2$.

$$\frac{s_X^2}{s_Y^2} = \frac{25.6^2}{28.3^2} = 0.818 \in (0.314, 2.96) \quad (0.025 \text{ and } 0.975 \text{ quantiles of } F_{12,15})$$

   so there is insufficient evidence that the variances differ (cannot reject H_0).

MAST20005/MAST90058: Week 9 Solutions

1. (a)

$$\mathbb{E}(\bar{X}_{\cdot\cdot}) = \mathbb{E}\left(\frac{1}{I}\sum_{i=1}^{I}\bar{X}_{i\cdot}\right) = \frac{1}{I}\sum_{i=1}^{I}\mu_i = \mu$$

(b)

$$\mathbb{E}(\bar{X}_{i\cdot}^2) = \text{var}(\bar{X}_{i\cdot}) + \mathbb{E}(\bar{X}_{i\cdot})^2 = \frac{\sigma^2}{J} + \mu_i^2$$

(c)

$$\mathbb{E}(\bar{X}_{\cdot\cdot}^2) = \text{var}(\bar{X}_{\cdot\cdot}) + \mathbb{E}(\bar{X}_{\cdot\cdot})^2 = \frac{\sigma^2}{IJ} + \mu^2$$

since we have

$$\text{var}(\bar{X}_{\cdot\cdot}) = \mathbb{E}\left((\bar{X}_{\cdot\cdot} - \mu)^2\right) = \mathbb{E}\left(\left(I^{-1}\sum_{i=1}^{I}\bar{X}_{i\cdot} - I^{-1}\sum_{i=1}^{I}\mu_i\right)^2\right)$$
$$= \mathbb{E}\left[\left(I^{-1}\sum_{i=1}^{I}(\bar{X}_{i\cdot} - \mu_i)\right)^2\right] = \frac{I\sigma^2}{I^2 J} = \frac{\sigma^2}{IJ}$$

(d)

$$\mathbb{E}(SS(T)) = \mathbb{E}\left(\sum_{i=1}^{I} J(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2\right) = \mathbb{E}\left(\sum_{i=1}^{I} J\bar{X}_{i\cdot}^2 + \sum_{i=1}^{I} J\bar{X}_{\cdot\cdot}^2 - 2\sum_{i=1}^{I} J\bar{X}_{i\cdot}\bar{X}_{\cdot\cdot}\right)$$
$$= \sum_{i=1}^{I} J\,\mathbb{E}(\bar{X}_{i\cdot}^2) + IJ\,\mathbb{E}(\bar{X}_{\cdot\cdot}^2) - 2IJ\,\mathbb{E}(\bar{X}_{\cdot\cdot}^2)$$
$$= \sum_{i=1}^{I} J\,\mathbb{E}(\bar{X}_{i\cdot}^2) - IJ\,\mathbb{E}(X_{\cdot\cdot}^2)$$
$$= \sum_{i=1}^{I} J\left(\frac{\sigma^2}{J} + \mu_i^2\right) - IJ\left(\frac{\sigma^2}{IJ} + \mu^2\right)$$
$$= (I-1)\sigma^2 + J\sum_{i=1}^{I}(\mu_i^2 - \mu^2)$$
$$= (I-1)\sigma^2 + J\sum_{i=1}^{I}(\mu_i - \mu)^2$$

as $\sum_{i=1}^{I}\mu_i\mu = I\mu^2$. Hence

$$\mathbb{E}(MS(T)) = \sigma^2 + \frac{J}{I-1}\sum_{i=1}^{I}(\mu_i - \mu)^2$$

as required. Note that we have used: $\sum_{i=1}^{I}(\mu_i - \mu)^2 = \sum_{i=1}^{I}(\mu_i^2 - 2\mu_i\mu + \mu^2) = \sum_{i=1}^{I}\mu_i^2 - 2I\mu^2 + I\mu^2 = \sum_{i=1}^{I}\mu_i^2 - I\mu^2 = \sum_{i=1}^{I}(\mu_i^2 - \mu^2)$.

(e) When H_0 is true, this is σ^2, otherwise it is larger than σ^2.

2.

$$F = \frac{MS(T)}{MS(E)} = \frac{2573.3}{1394.2} = 1.846$$

Under H_0, $F \sim \mathrm{F}_{4,15}$ so we reject H_0 if $F > 3.056$. Hence we cannot reject H_0.

3. (a) This is the same as a test that the slope parameter is equal to zero (versus not zero). The test statistic for this is equal to $t = 0.75169/0.02846 = 26.41$. This is much larger than the critical value 2.66 (note: this is a two-sided test, so we need the 0.995 quantile of t_{60}). Thus, we reject the null hypothesis and conclude that there is strong evidence of an association between body and brain weights.

   (b) Under $H_0\colon \rho = 0$, we have the following

$$Z = \frac{\frac{1}{2}\ln\frac{1+R}{1-R}}{\sqrt{\frac{1}{n-3}}} \approx \mathrm{N}(0,1).$$

   Therefore, we should reject H_0 if $|Z| > \Phi^{-1}(1-\alpha/2)$.

   (c) From the R output we see that $r^2 = 0.9208$ (under `Multiple R-squared`), so we can calculate that $r = 0.9595$.

   (d) The observed value for the above test statistic is $z = 14.91 > 2.58$. Thus we reject the null hypothesis that there is no correlation between brain and body weight.

   (e) Both tests indicate strong evidence against the null hypothesis of no association between the body and brain weights of mammals.

4. (a) Let $A_0 = \{(\mu,\sigma^2) : -\infty < \mu < \infty,\ \sigma^2 = \sigma_0^2\}$ and $A_1 = \{(\mu,\sigma^2) : -\infty < \mu < \infty,\ \sigma^2 > 0\}$ be the set of parameter values consistent with H_0 and H_1 respectively. To maximise L under H_0, we can show that we need $\hat{\mu} = \bar{x}$ and $\hat{\sigma}^2 = \sigma_0^2$, which gives:

$$L_0 = L(\hat{\mu},\hat{\sigma}^2) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{n/2} e^{-\frac{\sum_i (x_i-\bar{x})^2}{2\sigma_0^2}}.$$

   To maximise L under H_1, we can show that we need $\hat{\mu} = \bar{x}$ and $\hat{\sigma}^2 = n^{-1}\sum_i (x_i - \bar{x})^2$, which gives:

$$L_1 = L(\hat{\mu},\hat{\sigma}^2) = \left[\frac{ne^{-1}}{2\pi\sum_i (x_i-\bar{x})^2}\right]^{n/2}.$$

   (b) The LRT statistic is,

$$\lambda = \frac{L_0}{L_1} = \left(\frac{w}{n}\right)^{n/2} e^{-w/2+n/2}$$

   where $w = \sum_i (x_i-\bar{x})^2/\sigma_0^2$. We reject H_0 if $\lambda \leqslant k$. Solving this inequality in w gives a solution of the form $w \leqslant c_1$ and $w \geqslant c_2$, where c_1 and c_2 are constants depending on k and n. These need to be selected so to achieve the desired significance level. (Note: since $w \sim \chi^2_{n-1}$, for convenience we typically set c_1 and c_2 as equal-tailed quantiles from this distribution, even if these do not exactly correspond to the c_1 and c_2 from the LRT.)

MAST20005/MAST90058: Week 8 Solutions

1. In a truly random sequence of numbers, the probability of the next digit being the same as the preceding one is 1/10 and the probability of the next one differing by 1 from the preceding is 2/10. The null hypothesis is:

$$H_0\colon p_1 = 0.1,\ p_2 = 0.2,\ p_3 = 0.7$$

Suppose we obtain the following observed counts:

k=1

degree of freedom = 3 − 1 = 2

	Observed	Expected
Same	0	$50 \times 0.1 = 5$
Differ by 1	8	$50 \times 0.2 = 10$
Other	42	$50 \times 0.7 = 35$

The chi-squared statistic is:

reject H_0

$$\frac{(0-5)^2}{5} + \frac{(8-10)^2}{10} + \frac{(42-35)^2}{35} = 6.8 > 5.991 \quad (0.95 \text{ quantile of } \chi^2_2)$$

Thus we conclude that the string of 51 digits is unlikely to have been randomly generated.

2. (a) The table is:

i	x_i	$x_i - m$	Rank	Sign
1	41.195	1.195	5	1
2	39.485	−0.515	1	−1
3	41.229	1.229	6	1
4	36.840	−3.160	10	−1
5	38.050	−1.950	8.5	−1
6	40.890	0.890	4	1
7	38.345	−1.655	7	−1
8	34.930	−5.070	11	−1
9	39.245	−0.755	2	−1
10	31.031	−8.969	12	−1
11	40.780	0.780	3	1
12	38.050	−1.950	8.5	−1
13	30.906	−9.094	13	−1

and

$$W = 5 - 1 + 6 - 10 - 8.5 + 4 - 7 - 11 - 2 - 12 + 3 - 8.5 - 13 = -55.$$

Recall that

$$\mathbb{E}(W) = 0, \quad \operatorname{var}(W) = \frac{n(n+1)(2n+1)}{6} = 819$$

so

$$z = \frac{-55}{\sqrt{819}} = -1.922$$

which is less than −1.645 (0.05 quantile of a standard normal distribution), so we reject H_0 at a 5% level of significance.

(b) Bounding z against known quantiles from a standard normal gives: $-1.96 < z < -1.645$. Therefore, we deduce that, $0.025 <$ p-value < 0.05.

(c) There are 4 positive signs. Therefore, the p-value is

$$\Pr(Y \leqslant 4 \mid p = 0.5) = 0.1334.$$

This is greater than α so we cannot reject H_0.

(d) The null hypothesis is rejected using the signed-rank test but cannot be rejected using the sign test.

3. The observed and expected frequencies are:

	Red	Brown	Scarlet	White
O	254	69	87	22
E	243	81	81	27

and

$$\chi^2 = 3.646 < 7.815 \quad (0.95 \text{ quantile of } \chi^2_3)$$

so we cannot reject H_0 at the 5% level of significance.

4. This a problem where we want to do a goodness-of-fit test of a particular model but where we need to first estimate some of the parameters. We can set it up in one of two ways.

The first way is to think about the null distribution and work out which parameters need to be estimated. Under H_0 we have $p_{i1} = p_{i2}$, so let's call both of them p_i (since they are equal). These define the probabilities of each category (columns) that apply to each group of nurses (rows). Note that these are conditional probabilities, $p_i = \Pr(\text{category } i \mid \text{group I}) = \Pr(\text{category } i \mid \text{group II})$. To complete the model we also need to estimate the marginal probabilities of the two groups, let's call these $g_j = \Pr(\text{group } j)$, for $j = 1, 2$. The null model, therefore, is that the probability of an observation for category i in group j is $g_j p_i$. Note that there are 6 *independent* parameters to estimate (5 conditional column probabilities and one row probability), so ultimately we'll end up with a test with $12 - 6 - 1 = 5$ degrees of freedom.

The other way is to note that this model is equivalent to the usual test of independence of a contingency table, we end up estimating the same parameters and apply the same test as described above.

Under either setup, the observed and expected frequencies are:

$95+53=148$

$148 \times \frac{3}{5} = 88.8$

$148 \times \frac{2}{5} = 59.2$

		Category					
		1	2	3	4	5	6
Group I	O	95.0	36.0	71.0	21.0	45.0	32.0
	E	88.8	37.2	68.4	23.4	46.2	36.0
Group II	O	53.0	26.0	43.0	18.0	32.0	28.0
	E	59.2	24.8	45.6	15.6	30.8	24.0

and as there are 5 df

$$\chi^2 = \frac{(95 - 88.8)^2}{88.8} + \cdots + \frac{(28 - 24)^2}{24} = 3.23 < 11.07 \quad (0.95 \text{ quantile of } \chi^2_5)$$

so we cannot reject H_0.

MAST20005/MAST90058: Week 9 Solutions

1. (a)

$$\mathbb{E}(\bar{X}_{\cdot\cdot}) = \mathbb{E}\left(\frac{1}{I}\sum_{i=1}^{I}\bar{X}_{i\cdot}\right) = \frac{1}{I}\sum_{i=1}^{I}\mu_i = \mu$$

(b)

$$\mathbb{E}(\bar{X}_{i\cdot}^2) = \mathrm{var}(\bar{X}_{i\cdot}) + \mathbb{E}(\bar{X}_{i\cdot})^2 = \frac{\sigma^2}{J} + \mu_i^2$$

(c)

$$\mathbb{E}(\bar{X}_{\cdot\cdot}^2) = \mathrm{var}(\bar{X}_{\cdot\cdot}) + \mathbb{E}(\bar{X}_{\cdot\cdot})^2 = \frac{\sigma^2}{IJ} + \mu^2$$

since we have

$$\begin{aligned}\mathrm{var}(\bar{X}_{\cdot\cdot}) = \mathbb{E}\left((\bar{X}_{\cdot\cdot} - \mu)^2\right) &= \mathbb{E}\left(\left(I^{-1}\sum_{i=1}^{I}\bar{X}_{i\cdot} - I^{-1}\sum_{i=1}^{I}\mu_i\right)^2\right)\\ &= \mathbb{E}\left[\left(I^{-1}\sum_{i=1}^{I}(\bar{X}_{i\cdot} - \mu_i)\right)^2\right] = \frac{I\sigma^2}{I^2J} = \frac{\sigma^2}{IJ}\end{aligned}$$

(d)

$$\begin{aligned}\mathbb{E}(SS(T)) = \mathbb{E}\left(\sum_{i=1}^{I}J(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2\right) &= \mathbb{E}\left(\sum_{i=1}^{I}J\bar{X}_{i\cdot}^2 + \sum_{i=1}^{I}J\bar{X}_{\cdot\cdot}^2 - 2\sum_{i=1}^{I}J\bar{X}_{i\cdot}\bar{X}_{\cdot\cdot}\right)\\ &= \sum_{i=1}^{I}J\,\mathbb{E}(\bar{X}_{i\cdot}^2) + IJ\,\mathbb{E}(\bar{X}_{\cdot\cdot}^2) - 2IJ\,\mathbb{E}(\bar{X}_{\cdot\cdot}^2)\\ &= \sum_{i=1}^{I}J\,\mathbb{E}(\bar{X}_{i\cdot}^2) - IJ\,\mathbb{E}(X_{\cdot\cdot}^2)\\ &= \sum_{i=1}^{I}J\left(\frac{\sigma^2}{J} + \mu_i^2\right) - IJ\left(\frac{\sigma^2}{IJ} + \mu^2\right)\\ &= (I-1)\sigma^2 + J\sum_{i=1}^{I}(\mu_i^2 - \mu^2)\\ &= (I-1)\sigma^2 + J\sum_{i=1}^{I}(\mu_i - \mu)^2\end{aligned}$$

as $\sum_{i=1}^{I}\mu_i\mu = I\mu^2$. Hence

$$\mathbb{E}(MS(T)) = \sigma^2 + \frac{J}{I-1}\sum_{i=1}^{I}(\mu_i - \mu)^2$$

as required. Note that we have used: $\sum_{i=1}^{I}(\mu_i - \mu)^2 = \sum_{i=1}^{I}(\mu_i^2 - 2\mu_i\mu + \mu^2) = \sum_{i=1}^{I}\mu_i^2 - 2I\mu^2 + I\mu^2 = \sum_{i=1}^{I}\mu_i^2 - I\mu^2 = \sum_{i=1}^{I}(\mu_i^2 - \mu^2)$.

(e) When H_0 is true, this is σ^2, otherwise it is larger than σ^2.

2.

$$F = \frac{MS(T)}{MS(E)} = \frac{2573.3}{1394.2} = 1.846$$

Under H_0, $F \sim \mathrm{F}_{4,15}$ so we reject H_0 if $F > 3.056$. Hence we cannot reject H_0.

3. (a) This is the same as a test that the slope parameter is equal to zero (versus not zero). The test statistic for this is equal to $t = 0.75169/0.02846 = 26.41$. This is much larger than the critical value 2.66 (note: this is a two-sided test, so we need the 0.995 quantile of t_{60}). Thus, we reject the null hypothesis and conclude that there is strong evidence of an association between body and brain weights.

(b) Under $H_0\colon \rho = 0$, we have the following

$$Z = \frac{\frac{1}{2}\ln\frac{1+R}{1-R}}{\sqrt{\frac{1}{n-3}}} \approx \mathrm{N}(0,1).$$

Therefore, we should reject H_0 if $|Z| > \Phi^{-1}(1-\alpha/2)$.

(c) From the R output we see that $r^2 = 0.9208$ (under `Multiple R-squared`), so we can calculate that $r = 0.9595$.

(d) The observed value for the above test statistic is $z = 14.91 > 2.58$. Thus we reject the null hypothesis that there is no correlation between brain and body weight.

(e) Both tests indicate strong evidence against the null hypothesis of no association between the body and brain weights of mammals.

4. (a) Let $A_0 = \{(\mu, \sigma^2) : -\infty < \mu < \infty,\ \sigma^2 = \sigma_0^2\}$ and $A_1 = \{(\mu, \sigma^2) : -\infty < \mu < \infty,\ \sigma^2 > 0\}$ be the set of parameter values consistent with H_0 and H_1 respectively. To maximise L under H_0, we can show that we need $\hat{\mu} = \bar{x}$ and $\hat{\sigma}^2 = \sigma_0^2$, which gives:

$$L_0 = L(\hat{\mu}, \hat{\sigma}^2) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{n/2} e^{-\frac{\sum_i (x_i - \bar{x})^2}{2\sigma_0^2}}.$$

To maximise L under H_1, we can show that we need $\hat{\mu} = \bar{x}$ and $\hat{\sigma}^2 = n^{-1}\sum_i (x_i - \bar{x})^2$, which gives:

$$L_1 = L(\hat{\mu}, \hat{\sigma}^2) = \left[\frac{ne^{-1}}{2\pi\sum_i (x_i - \bar{x})^2}\right]^{n/2}.$$

(b) The LRT statistic is,

$$\lambda = \frac{L_0}{L_1} = \left(\frac{w}{n}\right)^{n/2} e^{-w/2+n/2}$$

where $w = \sum_i (x_i - \bar{x})^2/\sigma_0^2$. We reject H_0 if $\lambda \leqslant k$. Solving this inequality in w gives a solution of the form $w \leqslant c_1$ and $w \geqslant c_2$, where c_1 and c_2 are constants depending on k and n. These need to be selected so to achieve the desired significance level. (Note: since $w \sim \chi^2_{n-1}$, for convenience we typically set c_1 and c_2 as equal-tailed quantiles from this distribution, even if these do not exactly correspond to the c_1 and c_2 from the LRT.)

MAST20005/MAST90058: Week 10 Solutions

From the lectures, we know that the pdf of the kth order statistic $X_{(k)}$ is:

$$g_k(x) = k\binom{n}{k} F(x)^{k-1} (1 - F(x))^{n-k} f(x).$$

1. Here $f(x) = \frac{1}{3}e^{-x/3}$ and $F(x) = 1 - e^{-x/3}$.

   (a) Using the above result,

   $$\begin{aligned} g_3(x) &= 3\binom{5}{3} \left(1 - e^{-x/3}\right)^{3-1} \left(e^{-x/3}\right)^{5-3} \frac{1}{3}e^{-x/3} \\ &= 10\left(1 - e^{-x/3}\right)^2 e^{-x}, \quad x > 0. \end{aligned}$$

   (b) We need the probability that one or zero observations are larger than 5. The following derivation is very similar to the triangular distribution example from Module 9 (see the lecture notes).

   $$\begin{aligned} \Pr(X_{(4)} < 5) &= \binom{5}{4} F(x)^4 (1 - F(x)) + F(x)^5 \\ &= 5\left(1 - e^{-5/3}\right)^4 e^{-5/3} + \left(1 - e^{-5/3}\right)^5 \\ &= 0.7599 \end{aligned}$$

   (c) For $1 < X_{(1)}$ we need each observation to be larger than 1. In other words,

   $$\Pr(1 < X_{(1)}) = (1 - F(1))^5 = \left(e^{-1/3}\right)^5 = e^{-5/3} = 0.1889$$

2. (a) The likelihood is

   $$L(\theta) = \begin{cases} e^{-\sum_i (x_i - \theta)} & \theta \leqslant \min(x_i), \\ 0 & \text{otherwise.} \end{cases}$$

   This is maximised when each $(x_i - \theta)$ is minimised, and this happens when θ is as large as possible but still satisifies the constraint given by the inequality. Hence $\hat{\theta} = \min(X_i) = X_{(1)} = Y$.

   (b) Firstly,

   $$F(x) = \int_\theta^x e^{-(t-\theta)}\, dt = 1 - e^{-(x-\theta)}, \quad x \geqslant \theta.$$

   Then,

   $$\begin{aligned} g_1(y) &= n\,(1 - F(y))^{n-1} f(y) = 10\left(e^{-(y-\theta)}\right)^9 e^{-(y-\theta)} \\ &= 10e^{-10(y-\theta)}, \quad y \geqslant \theta. \end{aligned}$$

   (c) Firstly,

   $$\mathbb{E}(Y) = \int_\theta^\infty y 10e^{-10(y-\theta)}\, dy.$$

   Substitute $z = y - \theta$,

   $$\begin{aligned} \mathbb{E}(Y) &= \int_0^\infty (z + \theta) 10e^{-10z}\, dz \\ &= \theta \int_0^\infty 10e^{-10z}\, dz + \int_0^\infty z10e^{-10z}\, dz \\ &= \theta + \frac{1}{10} \end{aligned}$$

because the left integral evaluates to 1 since it integrates the pdf of an exponential distribution, and the right integral is the expected value of the same exponential distribution (so we know what its value is). Therefore, $\mathbb{E}(Y - \frac{1}{10}) = \theta$, which means $Y - \frac{1}{10}$ is an unbiased estimator of θ.

(d) Firstly (and substituting $z = y - \theta$ again),

$$\Pr(\theta < Y < \theta + c) = \int_{\theta}^{\theta+c} 10e^{-10(y-\theta)}\, dy = \int_0^c 10e^{-10z}\, dz = 1 - e^{-10c}.$$

Hence we need to solve $1 - e^{-10c} = 0.95$, which results in $c = 0.1\ln(20) = 0.300$. Now, simple rearranging gives

$$\Pr(\theta < Y < \theta + c) = \Pr(Y - c < \theta < Y)$$

so that a 95% confidence interval is $[y - 0.3, y]$.

(e) This was the 'boundary problem' example shown in the lectures as part of Module 2.

3. (a) $f(x) = 1$ and $F(x) = x$, so we have $g_1(x) = n(1-x)^{n-1}$, $0 < x < 1$.

(b) Using integration by parts,

$$\mathbb{E}(X_{(1)}) = \int_0^1 xn(1-x)^{n-1}\, dx = \left[-x(1-x)^n - \frac{1}{n+1}(1-x)^{n+1}\right]_0^1 = \frac{1}{n+1}.$$

4. (a) The pdf is symmetric about θ, with the function on either side being an exponential function that can be though of as two exponential distributions put 'back-to-back' (hence the nickname *double exponential distribution*). The expectation of X can be split into two integrals, one of each side of θ, and because of symmetry they will cancel out.

In more detail, let $Z = X - \theta$. This means Z has a symmetric pdf around 0, $f(z) = \frac{1}{2}e^{-|z|}$. Therefore,

$$\begin{aligned}\mathbb{E}(Z) &= \int_{-\infty}^{\infty} z\frac{1}{2}e^{-|z|}\, dz = \int_{-\infty}^{0} \frac{z}{2}e^{-|z|}\, dz + \int_0^{\infty} \frac{z}{2}e^{-|z|}\, dz = \int_{-\infty}^{0} \frac{z}{2}e^{z}\, dz + \int_0^{\infty} \frac{z}{2}e^{-z}\, dz \\ &= \int_0^{\infty} \frac{-z}{2}e^{-z}\, dz + \int_0^{\infty} \frac{z}{2}e^{-z}\, dz = -\int_0^{\infty} \frac{z}{2}e^{-z}\, dz + \int_0^{\infty} \frac{z}{2}e^{-z}\, dz = 0.\end{aligned}$$

This then implies $\mathbb{E}(X - \theta) = 0$, so we have $\mathbb{E}(X) = \theta$.

Using the hint, we can also exploit the symmetry to derive the variance,

$$\begin{aligned}\mathrm{var}(Z) = \mathbb{E}(Z^2) &= \int_{-\infty}^{\infty} z^2\frac{1}{2}e^{|z|}\, dz = \int_{-\infty}^{0} \frac{z^2}{2}e^{-|z|}\, dz + \int_0^{\infty} \frac{z^2}{2}e^{-|z|}\, dz \\ &= \int_{-\infty}^{0} \frac{z^2}{2}e^{z}\, dz + \int_0^{\infty} \frac{z^2}{2}e^{-z}\, dz = \int_0^{\infty} \frac{z^2}{2}e^{-z}\, dz + \int_0^{\infty} \frac{z^2}{2}e^{-z}\, dz \\ &= \int_0^{\infty} z^2e^{-z}\, dz = 2,\end{aligned}$$

and from this it follows that $\mathrm{var}(X) = 2$.

(b) For a sample mean we have $\mathbb{E}(\bar{X}) = \mathbb{E}(X) = \theta$ and $\mathrm{var}(\bar{X}) = \frac{1}{n}\mathrm{var}(X) = \frac{2}{n}$.

(c) Using the asymptotic distribution of the sample median, we have $\mathbb{E}(\hat{M}) \approx m$ and $\text{var}(\hat{M}) \approx (4nf(m)^2)^{-1}$. Due to symmetry, we know that $m = \theta$ (the population median is the same as the population mean), which means we have $f(m) = f(\theta) = \frac{1}{2}$ and thus: $\mathbb{E}(\hat{M}) \approx \theta$ and $\text{var}(\hat{M}) \approx \frac{1}{n}$.

(d) $\hat{M}$ is better. Both estimators are (approximately) unbiased but $\hat{M}$ has a smaller variance, so is more likely to be closer to the true value of θ. Note that this is the reverse of the situation of sampling from a normal distribution, where the sample mean is the better estimator.

(e) We did this already in the past! See the solution for question 1(c)iii from week 3. The MLE is the sample median, $\hat{M}$.

5. We use a confidence interval based on the order statistics. Since we are interested in the median, we would like a 'symmetric' interval formed by taking the ith lowest and ith largest order statistics, we just need to determine the most appropriate value of i. For $i = 1$ we have the interval $(x_{(1)}, x_{(14)})$, for $i = 2$ we have $(x_{(2)}, x_{(13)})$, and so on. Calculating the confidence levels for each of these leads us to using $(x_{(4)}, x_{(11)}) = (1.8, 6.26)$ as the best choice since it has a confidence level of 94.26%, which is very close to the desired 95%. This particular confidence level can be calculated in R using:

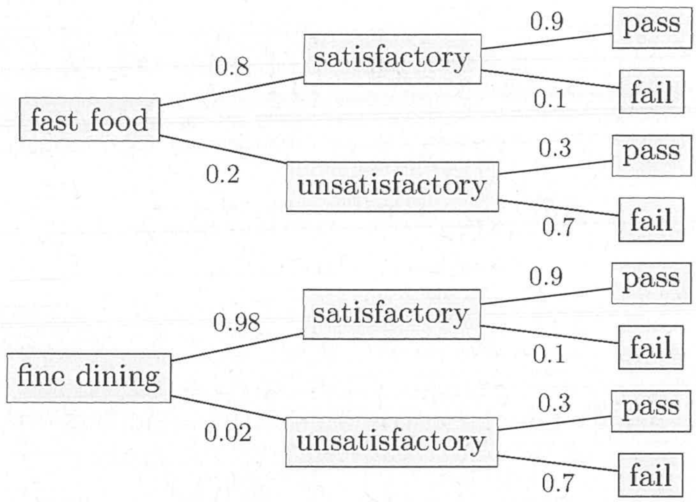
```
pbinom(10, size = 14, prob = 0.5) - pbinom(3, size = 14, prob = 0.5)
```

MAST20005/MAST90058: Week 11 Solutions

1. (a) $\Pr(\text{fail}) = 0.8 \times 0.1 + 0.2 \times 0.7 = 0.22$, $\Pr(\text{unsatisfactory} \mid \text{fail}) = \frac{0.2\times0.7}{0.22} = 0.636$

   (b) $\Pr(\text{fail}) = 0.98 \times 0.1 + 0.02 \times 0.7 = 0.112$, $\Pr(\text{unsatisfactory} \mid \text{fail}) = \frac{0.02\times0.7}{0.122} = 0.125$

   (c) Tree diagrams:



2. (a) The prior is Beta(1, 1). We observe 4 successes and 36 failures, so the posterior is Beta$(1+4, 1+36) = \text{Beta}(5, 37)$. This gives a posterior mean of $5/42 = 0.12$.

   (b) i. Using a Beta(α, β) prior, we need $\alpha + \beta = 60$ and $\alpha/(\alpha+\beta) = 0.04$. Solving these gives $\alpha = 2.4$ and $\beta = 57.6$.

   ii. The new posterior is Beta(6.4, 93.6).

   iii. The new posterior mean is $6.4/100 = 0.064$.

3. Let the prior be $\theta \sim \mathrm{N}(\mu_0, \sigma_0^2)$. The posterior pdf is,

$$
\begin{aligned}
f(\theta \mid x) &\propto \exp\left[-\frac{(x-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right] = \exp\left[-\frac{1}{2}\left(\frac{\theta^2 - 2x\theta + x^2}{\sigma^2} + \frac{\theta^2 - 2\mu_0\theta + \mu_0^2}{\sigma_0^2}\right)\right] \\
&\propto \exp\left[-\frac{1}{2}\left(\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}\right)\theta^2 - 2\left(\frac{x}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right)\theta\right)\right] = \exp\left[-\frac{1}{2\sigma_1^2}\left(\theta^2 - 2\mu_1\theta\right)\right] \\
&\propto \exp\left[-\frac{(\theta-\mu_1)^2}{2\sigma_1^2}\right]
\end{aligned}
$$

where,

$$
\mu_1 = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{x}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}} \quad \text{and} \quad \frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}.
$$

We recognise this as being a normal pdf, which means the posterior is $\theta \mid x \sim \mathrm{N}(\mu_1, \sigma_1^2)$.

4. Let the distribution of test scores be $X \sim \mathrm{N}(\theta, 25)$. We observe $\bar{x} = 70$ from a random sample of size $n = 16$. Using an improper uniform prior for θ, we get a posterior $\theta \mid \bar{x} \sim \mathrm{N}(70, 25/16)$. A central 95% credible interval for θ is $70 \pm 1.96 \times 5/4 = (67.55, 72.45)$.

5. (a) We have $Y \sim \mathrm{Pn}(n\theta)$ so the pdf is $f(y \mid \theta) = e^{-n\theta}(n\theta)^y/y!$. The posterior pdf is,

$$
f(\theta \mid y) \propto f(y \mid \theta) f(\theta) \propto \theta^{\alpha-1} e^{-\theta\beta} e^{-n\theta}\theta^y = \theta^{y+\alpha-1} e^{-\theta(n+\beta)}
$$

   This is a gamma distribution with parameters $y + \alpha$ and $n + \beta$.

(b) $E(\theta \mid y) = (y+\alpha)/(n+\beta)$

(c) The MLE is y/n and the prior mean is α/β. We can see that,

$$\frac{y+\alpha}{n+\beta} = \frac{n}{n+\beta} \cdot \frac{y}{n} + \frac{\beta}{n+\beta} \cdot \frac{\alpha}{\beta}.$$

6. The likelihood is,

$$f(x_1, \ldots, x_n \mid \theta) = (3\theta)^n \left(\prod_{i=1}^{n} x_i^2 \right) e^{-\theta \sum_{i=1}^{n} x_i^3}$$

and the prior pdf is,

$$f(\theta) = \frac{4^4}{\Gamma(4)} \theta^3 e^{-4\theta}, \quad 0 \leqslant \theta < \infty,$$

so the posterior pdf is,

$$f(\theta \mid x_1, \ldots, x_n) \propto \theta^n e^{-\theta \sum_{i=1}^{n} x_i^3} \theta^3 e^{-4\theta} = \theta^{n+3} e^{-\theta(4+\sum_{i=1}^{n} x_i^3)},$$

which we recognise as being a gamma distribution with parameters $n+4$ and $4+\sum_{i=1} x_i^3$. Therefore,

$$E(\theta \mid x_1, \ldots, x_n) = \frac{n+4}{4+\sum_{i=1}^{n} x_i^3}$$

7. Let $y = \sum_{i=1}^{n} x_i$ where the x_i are the individual times. Then the likelihood is,

$$f(x_1, \ldots, x_n \mid \theta) = \theta^n e^{-\theta y}$$

and the prior pdf is,

$$f(\theta) \propto \theta^{\alpha-1} e^{-\theta\beta}$$

which gives the posterior pdf,

$$f(\theta \mid x_1, \ldots, x_n) \propto \theta^{n+\alpha-1} e^{-\theta(y+\beta)}$$

which we recognise as being a gamma distribution with parameters $n+\alpha$ and $y+\beta$. Moreover, $\alpha/\beta = 0.2$ and $\sqrt{\alpha/\beta^2} = 0.1$ gives $\alpha = 4$ and $\beta = 20$, and from the data we have $n = 20$ and $y = n \times 3.8 = 76$ so that the posterior is a gamma distribution with parameters 24 and 96.

8. (a) The pdf for a single observation is $f(x \mid \theta) = \theta^{-1}$ with $0 < x < \theta$. Therefore, the likelihood is,

$$f(x_1, \ldots, x_n \mid \theta) = \theta^{-n}, \quad x_{(n)} < \theta.$$

Using an improper uniform prior for θ then gives the following for the posterior,

$$f(\theta \mid x_1, \ldots, x_n) \propto \theta^{-n}, \quad x_{(n)} < \theta.$$

We need to integrate this to calculate the normalising constant,

$$F(\theta) = \int_{x_{(n)}}^{\infty} \theta^{-n} \, d\theta = \frac{1}{(n-1)x_{(n)}^{n-1}}.$$

Therefore, the posterior pdf is,

$$f(\theta \mid x_1, \ldots, x_n) = (n-1)x_{(n)}^{n-1}\theta^{-n}, \quad x_{(n)} < \theta.$$

(b) The posterior is a decreasing function of θ, so its maximum occurs at the smallest possible value of θ, which is $x_{(n)}$.

(c) By integrating the posterior pdf we can show the posterior cdf is,

$$F(\theta \mid x_1, \ldots, x_n) = 1 - \left(\frac{x_{(n)}}{\theta}\right)^{n-1}, \quad x_{(n)} < \theta.$$

Solving for $F(\theta_u \mid x_1, \ldots, x_n) = 0.95$ will give us the upper bound of the desired credible interval.

$$0.95 = 1 - \left(\frac{x_{(n)}}{\theta_u}\right)^{n-1} \quad \Rightarrow \quad \theta_u = \frac{x_{(n)}}{\sqrt[n-1]{0.05}}$$

(d) Follow the steps outlined in the lectures (Module 9, slide 23 or page 6 of the notes), but do it in general for a sample size of n. For step 1 you need to use the cdf of $X_{(n)}$; the general formula for this is given earlier in the same module. The resulting confidence interval is,

$$\left(x_{(n)}, \frac{x_{(n)}}{\sqrt[n]{0.05}}\right).$$

Note that the denominator in the upper bound differs to that of the credible interval.

(e) If instead of a uniform prior we used one of the form $f(\theta) = \theta^{-1}$ we would get the two intervals to be the same. Note that this prior is also improper.

MAST20005/MAST90058: Week 12 Solutions

1. (a) Working with just a single observation to start with. The score function is,

$$\frac{\partial}{\partial p} \log p^x (1-p)^{1-x} = \frac{x}{p} - \frac{1-x}{1-p}.$$

Differentiating once more gives,

$$\frac{\partial}{\partial p}\left(\frac{x}{p} - \frac{1-x}{1-p}\right) = -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}.$$

Recall that $\mathbb{E}(X) = p$. Therefore the Fisher information is,

$$-\mathbb{E}\left[-\frac{X}{p^2} - \frac{1-X}{(1-p)^2}\right] = \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p(1-p)}.$$

Thus the Cramér–Rao lower bound, using the full sample, is $p(1-p)/n$.

(b) Since $\operatorname{var}(\bar{X}) = n^{-2}(\operatorname{var}(X_1) + \cdots + \operatorname{var}(X_n)) = p(1-p)/n$, the sample proportion $\bar{X}$ attains the Cramér–Rao lower bound.

2. (a) The score function is,

$$\frac{\partial}{\partial \theta} \log\left\{\left(\frac{1}{\sqrt{2\pi\theta}}\right)^n e^{-\sum_i (x_i-\mu)^2/(2\theta)}\right\} = -\frac{n}{2\theta} + \sum_i \frac{(x_i-\mu)^2}{2\theta^2}.$$

Setting this equal to 0 and solving for θ gives the result.

(b) Working with just a single observation to start with,

$$\frac{\partial^2}{\partial \theta^2} \log\left\{\left(\frac{1}{\sqrt{2\pi\theta}}\right) e^{-(x-\mu)^2/(2\theta)}\right\} = \frac{\partial}{\partial\theta}\left\{-\frac{1}{2\theta} + \frac{(x-\mu)^2}{2\theta^2}\right\} = \frac{1}{2\theta^2} - \frac{(x-\mu)^2}{\theta^3}.$$

Since $\mathbb{E}[(X-\mu)^2] = \theta$, the Fisher information is,

$$-\mathbb{E}\left[\frac{1}{2\theta^2} - \frac{(X-\mu)^2}{\theta^3}\right] = -\frac{1}{2\theta^2} + \frac{1}{\theta^2} = \frac{1}{2\theta^2},$$

and the Cramér–Rao lower bound, using the full sample, is $2\theta^2/n$.

(c) $\mathrm{N}(\theta, 2\theta^2/n)$

(d) Note that $(X_i-\mu)^2/\theta \sim \chi^2_1$. Thus, $n\hat{\theta}/\theta = \sum_i (X_i-\mu)^2/\theta \sim \chi^2_n$.

3. (a) Writing out the likelihood,

$$\begin{aligned} f(x_1,\dots,x_n \mid \theta) &\propto \theta^{-2n}\left(\prod_{i=1}^n x_i\right) e^{-\sum_{i=1}^n x_i/\theta} \\ &= \left(\prod_{i=1}^n x_i\right) e^{-y/\theta - 2n\log\theta} \end{aligned}$$

and the factorisation theorem yields that $Y = \sum_{i=1}^n X_i$ is sufficient for θ.

(b) The log-likelihood and score functions are, respectively:

$$\ell(\theta) = -2n\log(\theta) - \frac{y}{\theta} + \text{const.}$$

$$s(\theta) = \frac{\partial \ell}{\partial \theta} = -\frac{2n}{\theta} + \frac{y}{\theta^2}$$

(c) Solving $s(\theta) = 0$ gives the MLE, $\hat{\theta} = \bar{X}/2$.

(d) Differentiating the score function and treating the data as random gives,

$$-\frac{\partial^2 \ell}{\partial \theta^2} = -\frac{2n}{\theta^2} + \frac{2Y}{\theta^3},$$

which has expected value,

$$\mathbb{E}\left(-\frac{\partial^2 \ell}{\partial \theta^2}\right) = -\frac{2n}{\theta^2} + \frac{2n(2\theta)}{\theta^3} = \frac{2n}{\theta^2}.$$

Hence the Cramér–Rao lower bound is $\theta^2/(2n)$.

(e) The maximum likelihood estimate is $\hat{\theta} = 10.5/2 = 5.25$ and an approximate 95% confidence interval is $5.25 \pm 1.96 \times \sqrt{5.25^2/70} = (4.02, 6.48)$.

For questions 4–6, only the answers are given below. For full details of the derivation see the video consultation *Understanding Sufficiency* on the LMS.

4. Let the first five tosses of the coin be $X_1, \ldots, X_5$ and the second five be $Y_1, \ldots, Y_5$. Then $T = \sum_{i=1}^{5} X_i - \sum_{i=1}^{5} Y_i$ is sufficient for p.

5. Suppose we observe a random sample of size n.

   (a) $X_{(n)}$ is sufficient for θ.

   (b) $X_{(1)}$ and $X_{(n)}$ are jointly sufficient for θ.

6. Suppose we observe a random sample of size n.

   (a) $\sum_i X_i$ is sufficient for θ.

   (b) $X_{(1)}$ is sufficient for θ.

   (c) $\sum_i X_i$ and $X_{(1)}$ are jointly sufficient for θ.

7. (a) Let the sample frequencies (counts) of the three possible observations be f_0, f_1, f_2. Note that $f_0 + f_1 + f_2 = n$. The likelihood function is, $L(\theta) = (1-\theta)^{f_0}(\frac{3}{4}\theta)^{f_1}(\frac{1}{4}\theta)^{f_2} = \left[3^{f_1}\right]\left[(1-\theta)^{f_0}(\frac{1}{4}\theta)^{n-f_0}\right]$. Therefore, by the factorisation theorem we see that f_0 is sufficient for θ. (Note that we referred to this statistic as Z in the week 3 problems.)

   (b) We want to use an estimator based on this statistic in order to most efficiently capture the relevant information. This explains why the esimator based on $\bar{X}$ was not optimal: it is trying to use information that distinguishes between observations of 1 and 2, which is irrelevant information.

Week 2

1. (a) 2
(b) 8
(c) 0
(d) $\bar{X} \sim N(7, \frac{4}{9})$ $\frac{\sigma^2}{n}$ $\frac{4}{9}$

2. $Y = X_1 + \cdots + X_{15}$ sum of iid rvs.
pdf $= f(x) = \frac{3}{2}x^2 \quad (-1 < x < 1)$

(a). $E(X_1) = \int_{-1}^{1} x f(x)\,dx$
$= \int_{-1}^{1} x \cdot \frac{3}{2}x^2\,dx$
$= \int_{-1}^{1} \frac{3}{2}x^3\,dx$
$= [\frac{3}{8}x^4]_{-1}^{1}$
$= [\frac{3}{8}\times 1^4 - \frac{3}{8}\times(-1)^4] = 0.$

(b). $E(Y) = \cancel{E(15X_1)} = 15E(X_1) = 15\times 0 = 0.$
$Var(Y) = 15 \cdot Var(X_1) = 15 \times \frac{3}{5} = 9.$

$\boxed{Var(X) = E(X^2) - (E(X))^2}$ $= \frac{3}{5} - 0 = \frac{3}{5}$

$Var(X_1) = \underline{E(X_1^2)} - \underline{\underline{E(X_1)^2}}_{0} = \frac{3}{5} - 0 = \frac{3}{5}.$

$E(X_1^2) = \int_{-1}^{1} x^2 f(x)\,dx = \int_{-1}^{1} \frac{3}{2}x^4\,dx$
$= [\frac{3}{10}x^5]_{-1}^{1} = [\frac{3}{10}\times 1^5 - \frac{3}{10}\cdot(-1)^5]$
$= [\frac{3}{5}) =$

(c). CLT $\Rightarrow Pr(-0.3 < Y < 1.5)$

$\boxed{\cancel{Z = \frac{X-\mu}{\sigma/\sqrt{n}}}}$ $Y = 15\bar{X}$

$\boxed{Z = \frac{X-\mu}{\sigma}}$ $X \sim N(0, \frac{3}{5})$

$Y \sim N(0, 9)$ 3^2 Normal

$Pr(-0.3 < Y < 1.5) = Pr(\frac{-0.3-0}{3} < \frac{Y}{3} < \frac{1.5-0}{3})$
$= Pr(-0.1 < Z < 0.5) = \Phi(0.5) - \Phi(-0.1) = 0.6915 - 0.4602 = 0.23$

3.

4 (1) Show $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$
(2) $s^2 = (n-1)^{-1}(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)$

(1). $\sum_{i=1}^{n}(x_i - \bar{x}) = (x_1 - \bar{x}) + (x_2 - \bar{x}) + \cdots (x_n - \bar{x})$
$= (x_1 + x_2 + \cdots + x_n) - n\bar{x}$
$= \sum_{i=1}^{n} x_i - n\bar{x} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i$
$= 0.$

(2). Variance

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$-2x_1\bar{x} - 2x_2\bar{x} - 2x_3\bar{x}$
$-2\bar{x}(x_1 + x_2 + x_3 + \cdots)$
$n\bar{x}$ $\cdot 2$
$-2n\bar{x}$

$= \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$
$= \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2)$
$= \frac{1}{n-1}(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)$

(3) $\bar{x} = \frac{1}{n}\sum_{i=1}^{5} x_i = \frac{1}{5}\times 219.3 = 43.86$

Var: $s^2 = \frac{1}{4}(9654.27 - 5\times 43.86^2)$

$s = \sqrt{s^2} = 2.99$

$\boxed{(b)}$

(c) $\boxed{\tilde{\pi}_p = X_{(i)} + r\cdot(X_{(i+1)} - X_{(i)})}$

(d) $(n+1)p = i + r$

$\boxed{\tilde{\pi}_{0.25}}$ $(5+1)\times 0.25 = 1.5 = i + r$
$= 1 + 0.5$
$\Rightarrow \boxed{i = 1 \quad r = 0.5}$

$\therefore \tilde{\pi}_p = X_{(1)} + 0.5\times(X_2 - X_1)$
$\tilde{\pi}_{0.25} = 41.0 + 0.5\times(42.6 - 41.0)$
$= \boxed{41.8}$

$(5+1)\times 0.75 = 4.5 = 4 + 0.5$
$\boxed{i = 4 \quad r = 0.5}$

$\tilde{\pi}_{0.75} = X_{(4)} + 0.5\times(X_5 - X_4)$
$= \boxed{46.3}$ $\Rightarrow \{41.0, 41.8, \ldots\}$

$\boxed{iii}$ Let

(c) $Q_{0.25} - 1.5 \times IQR = 40.95$

$Q_{0.75} - 1.5 \times IQR = 45.35$

$\cancel{x < 41}$

$x < 40.95$ or $x > 45.35$.

$x 8.9$

$\boxed{6}$ (a)

Week 3

Week 3

1. (a) $L(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x_i-\mu)^2\right] = \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$

$\ln L(\mu) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$

$\frac{\partial \ln L(\mu)}{\partial \mu} = 0 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)\cdot(-2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$

$\Rightarrow \sum_{i=1}^{n}(x_i-\mu) = 0 \qquad \mu = \bar{x}$

(b) (i) poisson: pmf. $p(x) = e^{-\lambda}\frac{\lambda^x}{x!},\ x \in \{0, 1, \dots\}$

$\frac{\lambda^{x_1} \cdot \lambda^{x_2}}{x_1! \cdot x_2!}$

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

$\ln L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln\lambda - \ln\left(\prod_{i=1}^{n} x_i!\right)$

$\frac{\partial \ln L(\lambda)}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0$

$\Rightarrow \lambda = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \qquad \Rightarrow \lambda = \bar{x}$

(ii) $(5\times0 + 7\times1 + 12\times2 + 9\times3 + 5\times4 + 1\times5 + 1\times6)/40 = 2.25$

$\lambda = \bar{x} = 2.25$

1 (c) (i) $f(x/\theta) = \frac{1}{\theta^2} x e^{-\frac{x}{\theta}},\ 0<x<\infty,\ 0<\theta<\infty$

$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta^2} x_i e^{-\frac{x_i}{\theta}} = \left(\frac{1}{\theta^2}\right)^n \cdot \prod_{i=1}^{n} x_i \cdot e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}}$

$\ln x_i$

$\ln L(\theta) = -2n\ln\theta + \sum_{i=1}^{n}\ln x_i - \frac{1}{\theta}\sum_{i=1}^{n} x_i$

θ^{-1}

$-2 \quad \theta^{-2}$

$\frac{\partial \ln L(\theta)}{\partial \theta} = -2n\frac{1}{\theta} + 0 + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i = 0$

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

$\theta = \frac{\sum_{i=1}^{n} x_i}{2n} = \frac{\bar{x}}{2} = \frac{\bar{x}}{2}$

(iii) $f(x|\theta)=\frac{1}{2\theta^3}x^2 e^{-\frac{x}{\theta}},\ 0<x<\infty,\quad 0<\theta<\infty$

$L(\theta)=\prod_{i=1}^{n}\frac{1}{2\theta^3}x^2 e^{-\frac{x}{\theta}}=\left(\frac{1}{2\theta^3}\right)^n\cdot\prod_{i=1}^{n}x_i^2\cdot e^{-\frac{\sum_{i=1}^n x_i}{\theta}}$

$\left(\frac{1}{2}\right)^n \cancel{\left(\frac{1}{\theta^3}\right)^n}$

$\ln L(\theta)=\cancel{-3n\ln 2\theta}+2\sum_{i=1}^{n}\ln x_i-\frac{\sum_{i=1}^n x_i}{\theta}$

$=\cancel{2}\ -n\ln 2-3n\ln\theta$

$\frac{\partial\ln L(\theta)}{\partial\theta}=-3n\frac{1}{\theta}+\frac{\sum_{i=1}^n x_i}{\theta^2}=0$

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

$\theta=\frac{\sum_{i=1}^n x_i}{3n}=\frac{\bar{X}}{3}$ ✓

(iv) $f(x|\theta)=\frac{1}{2}e^{-|x-\theta|},\ -\infty<x<\infty,\ -\infty<\theta<\infty$

$L(\theta)=\prod_{i=1}^{n}\frac{1}{2}e^{-|x-\theta|}=\left(\frac{1}{2}\right)^n\cdot e^{-\sum_{i=1}^n|x_i-\theta|}$

$e^{-(|x_1-\theta|+|x_2-\theta|)}$

$\ln L(\theta)=-n\ln 2-\sum_{i=1}^{n}|x_i-\theta|$

$\frac{\partial\ln L(\theta)}{\partial\theta}=\sum_{i=1}^{n}\text{sgn}(x_i-\theta)$

2. $E(x)=0\times(1-\theta)+1\times(\frac{3}{4}\theta)+2\times\frac{\theta}{4}=\frac{5}{4}\theta$

$\because E(\bar{X})=\frac{5}{4}\theta\quad\therefore\ \theta=\frac{4}{5}\bar{X}\quad \boxed{T_1=\frac{4}{5}\bar{X}}\quad E(\frac{Z}{n})=1-\theta$

$Z\sim Bi(n,1-\theta)\quad E(Z)=n(1-\theta)\quad \cancel{\frac{\bar{X}}{n}=1-\theta}\quad E(-\frac{Z}{n})=\theta-1$

$\cancel{\theta=1-\frac{Z}{n}}\quad \boxed{\cancel{T_2=1-\frac{Z}{n}}}\quad E(1-\frac{Z}{n})=\theta$

$T_2=1-\frac{Z}{n}$

(b) ?

3. $f(x|\theta)=\theta x^{\theta-1}$

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

$\ln L(\theta)=n\ln\theta+(\theta-1)\sum_{i=1}^{n}\ln x_i$

$\frac{\partial\ln(L(\theta))}{\partial\theta}=\frac{n}{\theta}+\sum_{i=1}^{n}\ln x_i=0$ $\boxed{\text{Setting } s(\theta)=0}$

$\hat{\theta}=-\frac{n}{\sum_{i=1}^{n}\ln x_i}$

(c) $\hat{\theta}_x=\frac{10}{-18.2063}=0.549$

$\hat{\theta}_y=\frac{10}{-4.5246}=2.210$

$\hat{\theta}_z=\frac{10}{-10.0968}=0.959$

$\frac{\theta}{\theta+1}=\bar{X}$

$\theta=\theta\bar{X}+\bar{X}$

$\theta(1-\bar{X})=\bar{X}$

代进去算

$\boxed{\hat{\theta}=\frac{\bar{X}}{1-\bar{X}}}$

Week 3.

$$\cancel{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = \frac{1}{n-1}(}$$

$$\cancel{(x_1^2 - \bar{x}^2) + (x_2^2 - \bar{x}^2) + \cdots (x_n^2 - \bar{x})^2}$$

$\boxed{4.}$ $f(x|\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \ 0 < x < \infty, \ 0 < \theta < \infty)$

$\boxed{5.}$ We know that $s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{x}^2\right)$

Let $\mu = E(x)$ $\sigma^2 = Var(x) = E(X^2) - E(X)^2 = E(X^2) - \mu^2$

and $\boxed{E(X^2) = \sigma^2 + \mu^2}$

$$\left(\frac{1}{\theta}\right)' = (\theta^{-1})' = -\theta^{-2} = -\frac{1}{\theta^2}$$

$\boxed{6}$ $E(\bar{x}^2) = \frac{\sigma^2}{n} + \mu^2$

Week 4

1. a) $\bar{x} = 10.29$ b) $sd(\bar{x}) = \sqrt{Var(\bar{x})} = \frac{\sigma}{\sqrt{n}} = \frac{s}{\sqrt{n}} = 0.367$

$\hat{\pi}_{0.5} = 10.52$ $sd(\hat{\pi}_{0.5}) = \sqrt{Var(\hat{\pi}_{0.5})} = \sqrt{\frac{\pi}{2}}\frac{\sigma}{\sqrt{n}} = 0.459$

2. $N(\mu, 25)$ $\bar{x} = 73.8$ $n = 16$

$$\bar{x} \pm c \cdot \frac{\sigma}{\sqrt{n}} = 73.8 \pm 1.96 \times \frac{5}{\sqrt{16}} = (71.35, 76.25)$$

3. $N(2.09, 0.12^2)$ $n = 16$

$$2.09 \pm 1.96\frac{0.12}{\sqrt{16}} = (2.03, 2.15)$$

4. $11.95 \pm 1.96\frac{11.80}{\sqrt{37}}$

5.

$\boxed{\text{week 5}}$

7. (a) $n = 10$ $\bar{x} = 30.84$ $s = 2.908$

(a). 90% CI for μ.

$\hat{\mu} = \bar{x} = 30.84$ $30.84 \pm \cancel{1.64} \; 1.833 \cdot \frac{2.908}{\sqrt{10}}$

8.

Week 5

(1). Normal. single variance. σ^2

$\boxed{\text{Variance} \cdot \text{CI}}$ $\left[\frac{(n-1)s^2}{b}, \frac{(n-1)s^2}{a}\right]$. χ^2_{n-1} χ^2_{12} $a = F^{-1}_{12}(0.025) = 4.404$

$b = F^{-1}_{12}(0.975) = 23.34$.

* σ CI: $\left[\sqrt{\frac{12\times37.751}{23.34}}, \sqrt{\frac{12\times37.751}{4.404}}\right] = [4.41, 10.1]$.

2. (a). $sp = \sqrt{\frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}} = \sqrt{\frac{9\times0.323^2 + 9\times0.21^2}{18}} = 0.2724$

$\bar{x} - \bar{y} \pm c \cdot sp\sqrt{\frac{1}{n}+\frac{1}{m}} = 2.548 - 1.564 \pm 2.101 \times 0.2724 \times \sqrt{\frac{1}{10}+\frac{1}{10}}$
$= [0.728, 1.240]$

b). Yes, CI far away from 0.

~~c). 95% CI for σ_x/σ_y.~~

3. 90% CI for σ_x/σ_y $\left[\sqrt{c \cdot \frac{Sx^2}{Sy^2}}, \sqrt{d\frac{Sx^2}{Sy^2}}\right]$ $c = F^{-1}_{\cancel{15},12}(\cancel{0.95}\ 0.05) = \cancel{0.4640}$

$d = F^{-1}_{\cancel{18,12}}(0.95) = \cancel{2.47}$

$n = 16$. $\bar{x} = 21.95$ $Sx = 0.197$

$m = 13$, $\bar{y} = 21.88$, $Sy = 0.318$.

$c = F^{-1}_{12,15}(0.05) = 0.3821$

$d = F^{-1}_{12,15}(0.95) = 2.475$

$\left[\sqrt{c} \cdot \frac{Sx}{Sy}, \sqrt{d}\frac{Sx}{Sy}\right] = \left[\sqrt{0.3821} \cdot \frac{0.197}{0.318}, \sqrt{2.475}\frac{0.197}{0.318}\right]$

$= [0.383, 0.975]$

below 1, have good evidence differ

4 (a) $\hat{p} = \frac{24}{642} = 0.0374$

(b). Normal $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = [0.0227, 0.0521]$

(c). $\hat{p} + 1.645\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

5. $n = \left(\frac{c\sigma}{\epsilon}\right)^2 = \frac{1.96^2 \times \sigma^2}{\epsilon^2} = \frac{1.96^2 \times 4.84}{0.4^2} \approx 117$

6. (a) $n = \frac{1.96^2}{4\times0.03^2} = 1067.1 \approx 1068$

(b) $n = \frac{1.96^2}{4\times0.02^2} = 2401$

(c) $n = \frac{1.645^2}{4\times0.03^2} = 751.67 \approx 752$.

Week 6

$\boxed{11}$ $\hat\beta \pm c \dfrac{\hat\sigma}{\sqrt{K}}$ $\qquad \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar x)^2 = s^2$

$= 0.75169 \pm 2 \times \dfrac{0.6543}{\sqrt{594.9896}}$ $\qquad K = \sum_{i=1}^{n}(x_i-\bar x)^2 = (n-1)s^2 = 61 \times 3.123128^2 = 594.9896$

$= (0.69, 0.81)$

$\boxed{12}$ $\boxed{12\sim 6}$ 説明題.

$\boxed{7}$ (a) $\hat\alpha_0 = \bar Y = \dfrac{\sum y_i}{n} = \dfrac{313.2}{\cancel{12}\,13}$

$$\hat\beta = \frac{\sum_{i=1}^{n}(x_i-\bar x)(Y_i-\bar Y)}{\sum_{i=1}^{n}(x_i-\bar x)^2} = \frac{741.1}{578.8} = 1.28$$

$$\hat\alpha = \bar Y - \hat\beta \bar x = \frac{313.2}{13} - 1.28 \times \frac{230}{13} = 1.44$$

$$\hat\sigma^2 = \cancel{\sqrt{\tfrac{1}{n-2}D}}\ \frac{1}{n-2}D^2$$

$$= \frac{1}{11} \times \left(\sum_{i=1}^{n}(y_i-\bar y)^2 - \frac{\left[\sum_{i=1}^{n}(x_i-\bar x)(y_i-\bar y)\right]^2}{\sum_{i=1}^{n}(x_i-\bar x)^2} \right)$$

$$= \frac{1}{11} \times \left[1000.2 - \frac{741.1^2}{578.8} \right] =$$

$$\hat\sigma = \sqrt{\hat\sigma^2} = 2.16.$$

(b) $se(\hat\alpha) = \sqrt{Var(\hat\alpha)} = \sqrt{\left(\dfrac{1}{n} + \dfrac{\bar x^2}{K}\right)\hat\sigma^2} = \sqrt{\dfrac{1}{13} + \left(\dfrac{230}{13}\right)^2 / 578.8} \times 2.16 = 1.70$

$se(\hat\beta) = \dfrac{\hat\sigma}{\sqrt{K}} = \dfrac{2.16}{\sqrt{578.8}} = 0.898$

(c) $\hat\mu(18) \pm c \times se(\hat\mu(18))$

$\hat\mu(x) = \hat\alpha + \hat\beta x$
$\quad = \hat\alpha_0 + \hat\beta(x-\bar x)$
$\quad = \bar Y + \hat\beta(x - \bar x)$

$\hat\mu(18) = \hat\alpha + \hat\beta x = 1.44 + 1.28 \times 18 = 24.48$

$c = 2.201$ (0.975 quantile of t_{11})

$se(\hat\mu(18)) = \hat\sigma \sqrt{\dfrac{1}{n} + \dfrac{(x-\bar x)^2}{K}} = 2.16 \times \sqrt{\dfrac{1}{13} + \dfrac{\left(18 - \frac{230}{13}\right)^2}{578.8}} = 0.278$

$\boxed{CI}$ (23.2, 25.8)

(d) $\boxed{PI} = \hat\mu(x^*) \pm c \cdot \hat\sigma \cdot \sqrt{1 + \dfrac{1}{n} + \dfrac{(x^* - \bar x)^2}{K}} =$

~~24.48~~

$= 24.48 \pm 2.201 \times 2.16 \times \sqrt{1 + \dfrac{1}{13} + \dfrac{\left(18 - \frac{230}{13}\right)^2}{578.8}} = (19.6, 29.4)$

Week 7.

1. $\alpha = Pr(X \in \{2,3\} | p=\frac{1}{3}) = 0.222222 + 0.037037 04 = 0.25926$

$\beta = Pr(X \in \{0,1\} | p=\frac{2}{3}) = 0.037037 + 0.2222 = 0.25926.$

2. (a). $\alpha = Pr(Y \le 6 | p=0.08) =$

$E(Y) = 100 \times 0.08 = 8.$

$Var(Y) = np(1-p) = 100 \times 0.08 \times 0.92 = 7.36 = 2.713^2$

100次 ⇒ Normal

$\approx Pr(Z \le \frac{6-8}{2.713/\sqrt{100}}) = \cancel{\Phi^{-1}(-0.737)} = \cancel{0.23}\ 0.23$

$\alpha = Pr(Y \le 6 | p=0.08)$ 由 $\boxed{\frac{Y-\bar{Y}}{\sigma} = \frac{6-8}{2.713} = -0.737}$

$= Pr(Z \le -0.737) = \Phi^{-1}(-0.737) = 0.2305612$ ☆

$\boxed{\text{Continuity correction}}$

(b) $p=0.04$ $E(Y) = 0.04 \times 100 = 4$, $Var(Y) = np(1-p) = 3.84 = 1.96^2$

$\beta = Pr(Y \ge 7 | p=0.04) = Pr(Z \ge \frac{7-4}{1.96}) = Pr(Z > 1.531)$

$= 1 - \Phi^{-1}(1.531) = 0.063.$

3. (a). $\alpha = 0.05$ ~~$C = np_0 + \Phi^{-1}(1-\alpha)\sqrt{np_0(1-p_0)}$~~

~~$= 209 \times 0.07 + \Phi^{-1}$~~

reject H_0 if $Z = \frac{y/n - p_0}{\sqrt{p_0(1-p_0)/n}} > \Phi^{-1}(1-\alpha)$

$Z = \frac{\frac{23}{209} - 0.07}{\sqrt{0.07 \times 0.93/209}} = 2.269$

$\Phi^{-1}(0.95) = 1.64 < 2.269$

$Z > \Phi^{-1}(0.95)$

∴ reject H_0.

(b) $\alpha = 0.01$
$Z = 2.269$. $\Phi^{-1}(0.99) = 2.326 > Z$

do not reject H_0 at 1% level of significance.

(c). p-value $= P(Z \ge 2.269) = 1 - \Phi(2.269) = 1 - 0.9883618 = 0.0117$

Week 7

$\boxed{4}$. two proportions.

$$|Z| = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$$

$\hat{p}_1 = \frac{124}{894} = 0.1387$

$\hat{p}_2 = \frac{70}{700} = 0.1$

$\hat{p} = \frac{124+70}{894+700} = 0.1217.$

$\boxed{CI}$ for 95%

$$\hat{p}_1 - \hat{p}_2 \pm c\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$= 0.1387 - 0.1 \pm \Phi(0.975)\sqrt{\frac{0.1387\times(1-0.1387)}{894} + \frac{\hat{p}_2(1-\hat{p}_2)}{700}}$$

$(0.007, 0.07).$

$$|Z| = \frac{|0.1387 - 0.1|}{\sqrt{0.1217\times(1-0.1217)(\frac{1}{894}+\frac{1}{700})}} = 2.345.$$

$\Phi^{-1}(1-\frac{\alpha}{2}) = \Phi^{-1}(0.975) = 1.96 < |Z|$

$\therefore$ reject H_0.

$\boxed{5.}$ (a) $T = \frac{\bar{X} - 47}{S/\sqrt{20}} < \Phi^{-1}(0.05) \sim t_{19} = -1.729.$

(b). $T = \frac{46.94 - 47}{0.15/\sqrt{20}} = -1.7889 < -1.729.$

$\therefore$ reject H_0.

(c). $(0.025, 0.05) \sim t_{19}$

$-2.09 \quad -1.729.$

-1.7889

$\boxed{6}$

(a)(b) $H_0: \mu = 1.9 \quad H_1: \mu \neq 1.9$

(c). $|Z| = \frac{|\bar{X} - 1.91|}{S/\sqrt{n}} = \frac{|\bar{X} - 1.91|}{S/3} > \Phi^{-1}(0.975) \sim t_8 = \boxed{2.306}$

$\alpha = 0.05$

(d) $Z = \frac{2.05 - 1.90}{0.17/3} = 2.647 > 2.306$

$0.05 \quad 0.02$

$2.306 < 2.647 < 2.89$

(e) reject H_0

(f) $\boxed{0.02 < p\text{-value} < 0.05}$

Week 7.

$\boxed{7.}$ $\mu = 2.13$. $\sigma = 0.095$

$n = 20$ $\bar{x} = 2.10$ $s = 0.065$

(a) $\alpha = 0.05$ $\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{19 \times 0.065^2}{0.095^2} = 8.89 \; < \; \Phi^{-1}_{19}(0.05) = 10.117$

~~$F^{-1}(\frac{\alpha}{2}) = F^{-1}(0.025) = 8.9065$~~

~~$F^{-1}(1-\frac{\alpha}{2}) = F^{-1}(0.975) = 32.85$~~

~~$< F^{-1}(\frac{\alpha}{2}) < F^{-1}(0.975)$~~

$\therefore$ reject H_0
$\Rightarrow$ successful

(b). $\chi^2 = 8.89 < \Phi^{-1}(0.025) \approx 8.90$

$\therefore$ p-value $= Pr(\chi^2 < \Phi^{-1}(0.025)) \approx 0.025$

$\boxed{8}$. $n = 13$ $H_0 = \mu_x = \mu_y$ $H_1 = \mu_x \neq \mu_y$ $\alpha = 5\%$

$m = 16$.

(a). $s_p = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}} = \sqrt{\frac{12 s_x^2 + 15 s_y^2}{27}}$

$|T| = \frac{|\bar{X} - \bar{Y}|}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{|\bar{x} - \bar{Y}|}{s_p\sqrt{\frac{1}{13} + \frac{1}{16}}}$

t_{27}

$c = \Phi^{-1}(0.975) = 2.052$

critical region: reject H_0 if $t \geq c$ $|T| \geq 2.052$.

$t \geq 2.052$

(b). $t = \frac{|72.9 - 81.7|}{\sqrt{\frac{12 \times 25.6^2 + 15 \times 28.3^2}{27}} \times \sqrt{\frac{1}{13} + \frac{1}{16}}} = 0.8686 < 2.052$

$\therefore$ ~~reject~~ do not reject H_0.

(c). $0.6869 < 0.8686 < 1.31$ p-value $(0.2, 0.5)$

$\uparrow$ 0.75 $\uparrow$ 0.5 $\uparrow$ 0.9 $\uparrow$ 0.2

~~(d)~~ H_0

d) test variance $H_0: \sigma_x^2 = \sigma_y^2 \quad H_1: \sigma_x^2 \neq \sigma_y^2$

$\alpha = 0.05$

$$F = \cancel{\frac{(n-1)s_x^2}{\sigma^2} / n-1} \; \frac{s_x^2}{s_y^2} \sim F_{n-1, m-1} = F_{12,15}$$

$$= \frac{25.6^2}{28.3^2} = 0.818 \in (0.314, 2.96)$$

$\uparrow$ 0.025 $\qquad \uparrow$ 0.975

do not reject H_0.

Week 9.

$F \sim x, (+)$

① . 证明

② $MS(T) = 2573.3 \quad MS(E) = 1394.2$. $\quad \Phi^{-1}(1-\alpha) = \Phi^{-1}(0.95) = 3.05$.

F test. $F = \frac{MS(T)}{MS(E)} = \frac{2573.3}{1394.2} = 1.846 < c$. ~~reject~~ fail to ~~to~~ reject

③ (a) no association $\Rightarrow \beta = 0 \Leftarrow$ slope $= 0$. $\quad \alpha = 0.01 \quad$ 圭

$t = \frac{\hat{\beta} - 0}{se(\hat{\beta})} = \frac{0.75169}{0.02846} = 26.41 > c \Leftarrow c = \Phi^{-1}(0.995) \sim t_{60} \doteq 2.66$

2.66. reject Ho.

(b) $\frac{1}{2}\ln\left(\frac{1+R}{1-R}\right) \approx N\left(\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right), \frac{1}{n-3}\right)$

$\therefore t = \frac{\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) - 0}{\sqrt{\frac{1}{n-3}}} \approx N(0,1)$ $\quad$ reject Ho. if $t > c$. $c = \Phi^{-1}(1-\frac{\alpha}{2})$

(c). $R^2 = 0.9208 \Rightarrow R = 0.9596$

(d).

$t = \frac{\frac{1}{2}\ln\left(\frac{1+0.9596}{1-0.9596}\right)}{\sqrt{\frac{1}{59}}} = 14.908 > c = \Phi^{-1}(0.995) = ~~2.66~~ 2.58$. $\quad$ (for t_{60}) $\quad$ (norm)

reject Ho.

(e). have strong evidence against null hypothesis

④ . LRT

Week 10

① $f(x) = \lambda e^{-\lambda x} = \frac{1}{\theta} e^{-\frac{1}{\theta}x} = \frac{1}{3} e^{-\frac{1}{3}x}$

(a). $g_k(x) = k\binom{n}{k} F(x)^{k-1} (1-F(x))^{n-k} \cdot f(x)$

$F(x) = \int_0^x f(x)\,dx = \int_0^x \frac{1}{3} e^{-\frac{1}{3}x}\,dx = \cancel{\ }$

$\left[-e^{-\frac{1}{3}x}\right]_0^x = -e^{-\frac{1}{3}x} - (-1) = \underline{1-e^{-\frac{1}{3}x}}$

$F(5) = 1-e^{-\frac{5}{3}}$

$1-F(5) = e^{-\frac{5}{3}}$

$g_3(x) = 3\binom{5}{3}(1-e^{-\frac{1}{3}x})^2 \boxed{(e^{-\frac{1}{3}x})^2} (\frac{1}{3} \cdot e^{-\frac{1}{3}x})$

$= \frac{5\times4\times3\times2\times1}{3\times2\times1\times2\times1} = 10(1-e^{-\frac{1}{3}x})^2 \cdot e^{-x};\ x>0$

(b). $\underline{X_{(4)}<5}$ $Pr(X_{(4)}<5) = \sum_{i=4}^{5} \binom{5}{i} F(x)^i (1-F(x))^{5-i}$

$= \binom{5}{4} F(x)^4 (1-F(x))^1 + \binom{5}{5} \cdot F(x)^5 \cdot 1$

$= 5\times(1-e^{-\frac{5}{3}})^4 \times e^{-\frac{5}{3}} + (1-e^{-\frac{5}{3}})^5 = 0.75989$

(c). $Pr(1<X_{(1)}) = (1-F(x))^n = (1-F(1))^5 = (1-(1-e^{-\frac{1}{3}}))^5 = e^{-\frac{5}{3}} = 0.18887$

②. pdf $= f(x|\theta) = e^{-(x-\theta)},\ \theta \le x < \infty$

(a) $L(\theta) = \prod_{i=1}^{n} e^{-(x_i-\theta)} = \prod_{i=1}^{n} e^{\theta - x_i} = \cancel{e^{n\theta} \cdot e^{-\sum x_i}}\ \ e^{\sum_{i=1}^{n}(\theta - x_i)}$

when $x_i \le \theta$, makes maximum $\therefore\ \hat{\theta} = \min(X_i) = X_{(1)} = Y$

b) $Y = X_{(1)}$ $g_1(x) = n\cdot(1-F(x))^{n-1} \cdot f(x)$ $F(x) = \int_0^x e^{-(x-\theta)}\,dx = \left[-e^{-(x-\theta)}\right]_0^x$

$g_1(y)$ $= 10\cdot[1-(1-e^{-(x-\theta)})]^9 \cdot e^{-(x-\theta)}$ $= -e^{-(x-\theta)} - (-1) = 1-e^{-(x-\theta)}$

$\cancel{= 10 = 10(1-e^{9(x-\theta)})}$

$= 10\, e^{-9(x-\theta)} \cdot e^{-(x-\theta)} = 10 \cdot e^{-10(x-\theta)},\ x \ge \theta$

(c). $E(Y) =$ $\boxed{g_1(y) = 10 \cdot e^{-10(y-\theta)},\ y \ge \theta}$

$\int_\theta^\infty y \cdot g_1(y)\,dy = \int_\theta^\infty y \cdot 10 \cdot e^{-10(y-\theta)}\,dy$ $z = y-\theta$ $\therefore\ \underline{y = z+\theta}$

$= \int_0^\infty (z+\theta) \cdot 10 \cdot e^{-10z}\,dz$

$= 10\theta \int_0^\infty e^{-10z}\,dz + \int_0^\infty 10 \cdot z\, e^{-10z}\,dz$

??

$\int x \cdot e^{ax}\,dx = \frac{1}{a^2}(ax-1)e^{ax} + C$

$= \cancel{\theta + \frac{1}{10}}$

$10\theta\left[-10 e^{-10z}\right]_0^\infty$

$10\theta[0-(-10)]$

$= \theta + \frac{1}{10}$??

$\therefore\ E\ \bar{Y} = \theta + \frac{1}{10}$ $\therefore\ \theta = \bar{Y} - \frac{1}{10}$

$\therefore\ \underline{E(\bar{Y} - \frac{1}{10}) = \theta}$

Week 10.

(d). $Pr(\theta < Y < \theta + c)$. 95% CI for θ.

$= Pr(\theta < X_{(1)} < \theta + c) = Pr(X < \theta + c) - Pr(X < \theta)$.

$Pr(X_{(1)} \le c) = F_1(c) = \boxed{1 - e^{-(x-\theta)}}$ ✗. $10 \cdot e^{-10(y-\theta)}$

$Pr(X_{(1)} < \theta + c) - Pr(X_{(1)} < \theta)$

$= 1 - e^{-(\theta + c - \theta)} - (1 - e^{-(\theta - \theta)}) = 1 - e^{-c} - (1 - e^0) = 1 - e^{-10c}$

$Pr(\theta < Y < \theta + c) = 1 - e^{-10c}$. $Y - c < \theta < Y$

$Pr(Y - c < \theta < Y) = \boxed{1 - e^{-10c} = 0.95}$ $e^{-10c} = 0.05$

$\therefore$ 95% CI for θ. $(Y - 0.3, Y)$

$\cancel{\ln(e^{-10c})} = \ln(0.05) =$

$-10c = \ln(0.05) = \cancel{-2.9957}$ 0.299

$c = 0.300$

(e) boundry problem

3 (a) Unif $(0, 1)$. $X_1 < \dots X_{(n)}$. $f(x) = 1$

$\cancel{X_{(1)}} = g_1(x) = n \cdot (1 - F(x))^{n-1} \cdot f(x)$ $F(x) = X$.

$= n \cdot (1 - x)^{n-1}$.

$dt = -dx$.

$\cancel{n - 1 = t}$

$\cancel{n = t + 1}$

$1 - x = t$ $x = 1 - t$.

(b) $E(X_{(1)}) = \int_0^1 x \cdot g_1(x)\, dx = \boxed{\int_0^1 x \cdot n \cdot (1-x)^{n-1} \cdot dx}$

$= \cancel{\int_0^1 (t+1)\, x \cdot (1-x)^t}\, dx$

$= \cancel{\int_0^1 t \cdot x \cdot (1-x)^t dx + \int_0^1 x \cdot (1-x)^t dx}$

$= \int_0^1 (1-t) \cdot n \cdot t^{n-1} dt$

$= \int_0^1 n \cdot t^{n-1} dt - \int_0^1 t \cdot n\, t^{n-1} dt$

$= [t^n]_0^1 - [\frac{n}{n+1} t^{n+1}]_0^1$

$= 1 - \frac{n}{n+1} = \frac{1}{n+1}$

4 pdf: $f(x|\theta) = \frac{1}{2} e^{-|x-\theta|}$

(a). $x - \theta = z$. pdf $f(z) = \frac{1}{2} e^{-|z|}$ $x = \theta + z$.

symmetric

$E(z) = \int_{-\infty}^{\infty} z \cdot \frac{1}{2} e^{-|z|} dz = \int_{-\infty}^{0} z\, \frac{1}{2} e^{z} dz + \int_0^\infty z \cdot \frac{1}{2} e^{-z} dz$

$= \int_0^{\infty} -z\, \frac{1}{2} e^{z} dz + \int_0^\infty z \cdot \frac{1}{2} \cdot e^{-z} dz = 0$.

$\therefore E(X - \theta) = 0 \Rightarrow E(x = 0)$

$Var(Z) = E(Z^2) - E(Z)^2 = E(Z^2) = \int_{-\infty}^{\infty} z^2 \cdot \frac{1}{2} e^{z} dz = \int_{-\infty}^{0} z^2 \cdot \frac{1}{2} \cdot e^{z} dz$

$+ \int_{0}^{\infty} z^2 \cdot \frac{1}{2} e^{-z} dz$

$\therefore Var(x - \theta) = 2$ $= 2 \times \frac{1}{2} \times \int_0^\infty z^2 e^{-z} dz = 2$

$Var(x) = 2$

(b) $E(\hat{\theta}_1) = E(\bar{X}) = E(X) = 0$.

$Var(\bar{X}) = \frac{Var(X)}{n} = \frac{2}{n}$.

(c). $E(\hat{\theta}_2) = E(\hat{M}) \approx m$.

$Var(\hat{M}) \approx \frac{1}{4nf(m)^2}$

$\because$ Symmetric $\therefore$ mean = median.

$\therefore f(m) = f(\text{mean}) = \frac{1}{2}$ $\therefore Var(\hat{M}) = \frac{1}{n}$

(d). $Var(\hat{\theta}_2) < Var(\hat{\theta}_1)$ $\therefore$ median better.

(e). ---

$\boxed{5}$.

Week 11

1. $P(\text{pass} \mid \cancel{\text{sat}}\ \text{satisfactory}) = 0.9$ $P(\text{pass} \mid \text{unsat's}) = 0.3$

$P(\text{fail} \mid \text{sat's}) = 0.1$ $P(\cancel{\ }\ \text{fail} \mid \text{unsat}) = 0.7$

(a) $P(\text{unsa}) = 0.2$ $P(\text{sat's}) = 0.8$

$$Pr(\text{unsa} \mid \text{fail}) = \frac{\overset{0.7}{Pr(\text{fail} \mid \text{unsa})} \cdot \overset{0.2}{Pr(\text{unsat})}}{Pr(\text{fail})\ \ 0.22} = 0.63636$$

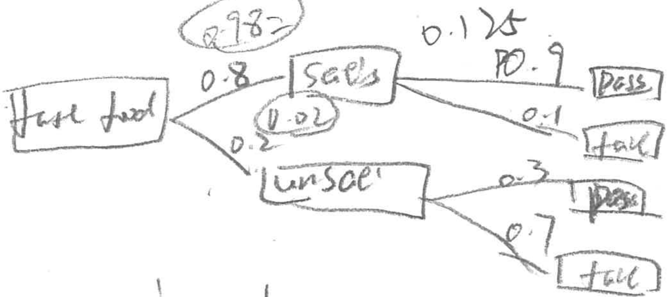
$Pr(\text{fail}) = Pr(\text{fail} \mid \text{sat})\, Pr(\text{sat}) + Pr(\text{fail} \mid \text{unsat})\, Pr(\text{unsat})$
$= 0.1 \times 0.8 + 0.7 \times 0.2 = 0.22$

(b) $Pr(\text{unsat}) = 0.02$ $Pr(\text{sats}) = 0.98$

$$Pr(\text{unsat} \mid \text{fail}) = \frac{\overset{0.7}{Pr(\text{fail} \mid \text{unsat})} \cdot \overset{0.02}{Pr(\text{unsta})}}{Pr(\text{fail})\ \ 0.112}$$

$Pr(\text{fail}) = 0.1 \times 0.98 + 0.7 \times 0.02 = 0.11$

(c). (a)



2. $Pr(\text{successful}) = \frac{1}{10} = 0.1$

$Pr(\text{fail}) = 0.9$

Beta $(\alpha, \beta) \Rightarrow$ Beta$(x+\alpha, \beta+(n-x))$

(a). ~~$\theta \sim B(n, 0.1)$~~ Prior $\sim$ Beta$(1,1)$

posterior is Beta$(1+4, 1+36)$ = Beta$(5, 37)$

$E = \frac{\alpha}{\alpha+\beta} = \frac{5}{5+37} = \frac{5}{42} = 0.12$

(b). ~~α~~ Beta(α, β). $\begin{cases} \alpha+\beta = 60 \\ \frac{\alpha}{\alpha+\beta} = 0.04 \end{cases}$

~~$\alpha = 0.04\alpha + 0.04\beta$~~

$\frac{0.96\alpha = \cancel{0.04\beta}}{0.04}$ $\beta = 24\alpha$

$24\alpha + \alpha = 60$
$25\alpha = 60$

$\begin{cases} \alpha = 2.4 \\ \beta = 57.6 \end{cases}$

i. prior Beta$(2.4, 57.6)$

ii. posterior $(2.4+4, 57.6+36)$ = Beta$(6.4, 93.6)$

iii. mean: $\frac{\alpha}{\alpha+\beta} = \frac{6.4}{6.4+93.6} = 0.064$

3. Normal

4. improper prior posterior is $N(70, \frac{25}{16})$

95% CI: $70 \pm 1.96 \frac{5}{4}$

~~$Y \sim Pn(n\lambda)$~~ ~~pdf: $e^{-n\lambda}\frac{(n\lambda)^y}{y!}$~~

~~pdf: $e^{-\lambda}\frac{\lambda^x}{x!}$~~

~~$f(y|\theta)$~~ =

$f(y|\theta) = \cancel{\prod_{i=1}^{n} e^{-\lambda}} \frac{\lambda^{y_i}}{y_i!}$

$= \cancel{e^{-n\lambda}}$

$Y \sim Pn(n\theta)$

pdf: $e^{-n\theta}\frac{(n\theta)^y}{y!}$

5. $f(\theta|y) \propto f(y|\theta) \cdot f(\theta)$

$\propto e^{-n\theta}\frac{(n\theta)^y}{y!}\, \theta^{\alpha-1} e^{-\theta\beta}$

$\propto e^{-\theta(n+\beta)}\ \theta^{y+\alpha-1}$

∴ Gamma distn with parameter $y+\alpha$, $n+\beta$.

(b) Gamma$(y+\alpha, n+\beta)$

$E(\theta|y)$ mean $= \dfrac{y+\alpha}{n+\beta}$

(c) MLE $\Rightarrow y/n$ prior mean $= \alpha/\beta$ $\quad \dfrac{y+\alpha}{n+\beta} = \dfrac{n}{n+\beta}\cdot\dfrac{y}{n} + \dfrac{\beta}{n+\beta}\cdot\dfrac{\alpha}{\beta}$

$\boxed{16.}$ prior: Gama$(4,4)$. $\quad f(x|\theta) = 3\theta x^2 e^{-\theta x^3}$

$f(\theta|x) \propto f(x|\theta)\, f(\theta)$ $\quad L\, f(x|\theta) = (3\theta)^n \prod_{i=1}^{n} x_i^2 \cdot e^{-\theta \cdot \sum_{i=1}^{n} x_i^3}$

☆ $\propto (3\theta)^n \cdot e^{-\theta \sum_{i=1}^{n} x_i^3} \cdot \theta^3 \cdot e^{-4\theta}$ $\quad f(\theta|\alpha=4,\beta=4) = \dfrac{4^4}{\Gamma(\alpha)}\theta^3 \cdot e^{-4\theta}$

$\propto \theta^{3+n} \cdot e^{-4\theta-\theta\sum_{i=1}^{n} x_i^3}$ $\quad \alpha - 1 = 3+n$

$e^{-\theta(4+\sum_{i=1}^{n} x_i^3)}$ $\quad \alpha = 4+n$

$\beta = 4 + \sum_{i=1}^{n} x_i^3$

Gamma$(4+n, 4+\sum_{i=1}^{n} x_i^3)$

$E(\theta) = \dfrac{4+n}{4+\sum_{i=1}^{n} x_i^3}$

$\boxed{17}$ $X \sim$ Exp(θ) $\quad f(y) = \theta \cdot e^{-\theta y}$ $x \in [0,\infty)$

prior Gamma

$\begin{cases} \text{mean} = \dfrac{\alpha}{\beta} = 0.2 \\ \sqrt{\dfrac{\alpha}{\beta^2}} = 0.1 \end{cases}$ ✗ $\quad \begin{cases} \alpha = 0.4 \\ \beta = 2 \end{cases}$ Gamma$(0.4, 2)$

$f(y|\theta) = \cancel{\prod}\,\theta^n \cdot e^{-\theta \sum_{i=1}^{n} y_i}$

$\bar{y} = 3.8 \quad n = 20$

posterior: $f(\theta|y) \propto \underbrace{f(y|\theta)}_{\text{likelihood}}\ \underbrace{f(\theta)}_{\text{prior}}$ $\quad f(\theta) = \dfrac{\cancel{20.4}}{\Gamma(\alpha)}$ Gama $\theta^{\cancel{-0.6}\ 3} \cdot e^{\cancel{-2\theta}\ -20\theta}$

$\alpha - 1 = \cancel{n+4}\ n+3$

$\propto n\theta^n \cdot e^{-\theta \sum_{i=1}^{n} y_i} \cdot \theta^{\cancel{-0.6}\ 3} \cdot e^{\cancel{-2\theta}\ -20\theta}$

$\theta^{\cancel{n-0.6}\ 3+n} \cdot e^{-\theta(20+\sum_{i=1}^{n} y_i)}$ $\quad$ posterior: Gamma$(\cancel{n+4}, \cancel{2+\sum y_i}\ 20+\sum_{i=1}^{n} y_i)$ 96

$\alpha = n+4$

$\begin{cases} \dfrac{\alpha}{\beta} = 0.2 \\ \sigma^2 = \dfrac{\alpha}{\beta^2} = 0.01 \end{cases}$ $\quad \begin{cases} \alpha = 4 \\ \beta = 20 \end{cases}$ Gamma$(4, 20)$ $\quad y_i = 3.8 \times 20$

~~Gamma$(4, 96)$~~ 7.6

$\boxed{8}$ (a) improper uniform prior: $f(\theta) = 1$. $\quad$ unif$(0,\theta)$ $\quad$ Gama$(n+4, 20+\sum_{i=1}^{n} y_i)$

$f(\theta|x) \propto \cancel{\phi}\, f(x|\theta) \cdot f(\theta)$ $\quad f(x|\theta) = \dfrac{1}{\theta}$ $\quad (24, 96)$

$\propto \theta^{-n}$, $\quad X_{(n)} < \theta$ $\quad$ likelihood: $\prod_{i=1}^{n} \dfrac{1}{\theta} = \theta^{-n}$ $\quad 0 < x < \theta$

$\int_{X_{(n)}}^{\infty} \theta^{-n}\, d\theta = \cancel{n\theta^{-n}} \left[\dfrac{1}{1-n}\theta^{-n+1}\right]_{X_n}^{\infty} = 0 - \dfrac{1}{1-n} X_n^{1-n} = \dfrac{1}{n-1} X_n^{1-n} = \dfrac{1}{n-1}\dfrac{1}{X_n^{n-1}}$

$\cancel{\int_{X_{(n)}}^{\infty} \left(\dfrac{1}{\theta}\right)^n d\theta = \dfrac{1}{n+1}}$ $\quad \therefore$ posterior $= \cancel{\dfrac{1}{n-1}}$??

$\boxed{(n-1) X_{(n)}^{n-1} \theta^{-n}}$ $\quad X_{(n)} < \theta$.

~~Week 11~~

(b) ~~decreasing $\therefore$ so maximum occurs at the smallest value of θ, $X_{(n)}$~~

(c)

$\boxed{\text{Week 12}}$

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

$L(p)=\prod_{i=1}^{n} p^x(1-p)^{1-x} = p^{\sum x_i}\cdot(1-p)^{n-\sum x_i}$

$\ln L(p) = \sum x_i \ln p + (n-\sum x_i)\ln(1-p)$

$\frac{\partial \ln L(p)}{\partial p} = \frac{\sum x_i}{p} - \frac{n-\sum x_i}{1-p}$

let $\frac{\sum x_i}{p} = \frac{n-\sum x_i}{1-p}$

$np - \cancel{p\sum x_i} = \sum x_i - \cancel{p\sum x_i}$

$\boxed{p=\bar{x}}$ $\quad \sum x_i = np$

$\frac{\partial^2 \ln(p)}{\partial p^2} = -\frac{\sum x_i}{p^2} - \frac{n-\sum x_i}{(1-p)^2}$

$I(\theta)=E(-V(p)) = -\left(-\frac{np}{p^2} - \frac{n-np}{(1-p)^2}\right) = \frac{n}{p}+\frac{n(1-p)}{(1-p)^2}$

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

$= \frac{n}{p(1-p)}$ ✓

$\therefore$ CR lower bound: $\boxed{\frac{1}{I(\theta)} = \frac{p(1-p)}{n}}$

②

~~$L(\theta)=\prod_{i=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right)=\frac{1}{n^2}$~~

~~$\ln L(\theta) = \ln n + \ln\sum_{i=1}^{n}$~~

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

~~like~~ $L(\theta) = \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{(x-\mu)^2}{2\theta}}\right)$

$= \left(\frac{1}{\sqrt{2\pi\theta}}\right)^n \cdot e^{-\frac{\sum(x-\mu)^2}{2\theta}} = \left(\frac{1}{\sqrt{2\pi}}\right)^n\cdot\left(\frac{1}{\sqrt{\theta}}\right)^n \cdot e^{-\frac{\sum_{i=1}^{n}(x-\mu)^2}{2\theta}}$

$\theta^{\frac{1}{2}}$ $\quad \theta^{-\frac{1}{2}n}$

$\ln L(\theta) = n\ln\frac{1}{\sqrt{2\pi}} - \frac{n}{2}\ln\theta - \frac{1}{2\theta}\sum(x_i-\mu)^2$

$\frac{\partial \ln L(\theta)}{\partial\theta} = -\frac{n}{2}\cdot\frac{1}{\theta} + 2\cdot\frac{1}{(2\theta)^2}\sum(x_i-\mu)^2 = -\frac{n}{2\theta} + \frac{\sum(x_i-\mu)^2}{2\theta^2}$ ✓

θ^{-2}. $\quad -2.\ \theta^{-3}$.

$V(\theta) = \frac{\partial^2 \ln L(\theta)}{\partial\theta^2} = +\frac{n\times 1}{2\cdot\theta^2} + \frac{\sum(x_i-\mu)^2}{2}(-2)\frac{1}{\theta^3}$

$= \frac{n}{2\theta^2} - \frac{\sum(x_i-\mu)^2 = n\theta}{\theta^3} = \frac{n}{2\theta^2} - \frac{n\theta}{\theta^3} = \frac{n}{2\theta^2} - \frac{2n}{2\theta^2} = -\frac{n}{2\theta^2}$

(b) $I(\theta) = E(V(\theta))$

CR lower bound $\frac{1}{I\theta} = \frac{2\theta^2}{n}$ $\quad \hat{\theta} \sim N\left(\hat{\theta}, \frac{2\theta^2}{n}\right)$

(c)

(d) $(x_i-\mu)^2/\theta \sim \chi_1$

$\frac{n\hat{\theta}}{\theta} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\theta} \sim \chi_n^2$

$\boxed{3}$ $f(x) = \frac{x}{\theta^2} e^{-\frac{x}{\theta}}$ $\theta^{-2n} e^{\ln \theta^{-2n}}$

$L(\theta) = \prod_{i=1}^{n} \left(\frac{x}{\theta^2} \cdot e^{-\frac{x}{\theta}}\right) = \left(\frac{1}{\theta^2}\right)^n \cdot \prod x_i \cdot e^{-\frac{\sum x_i}{\theta}}$

$= \prod x_i \cdot e^{-\frac{\sum x_i}{\theta} - 2n\ln\theta}$ $\therefore Y = \sum x_i$

Sufficient statistic

b) $\ln L(\theta) = -2n \ln\theta + \cancel{\ln \prod x_i} - \frac{1}{\theta}\sum x_i$

$\frac{\partial (\ln L(\theta))}{\partial \theta} = -2n\frac{1}{\theta} + \frac{1}{\theta^2}\sum x_i$

(c) $\frac{2n}{\theta} = \frac{\sum x_i}{\theta^2}$ $\hat{\theta} = \frac{\sum x_i}{2n} = \frac{\bar{x}}{2}$

d) $V(\theta) = \frac{\partial^2 (\ln L(\theta))}{\partial \theta^2} = 2n\frac{1}{\theta^2} - \frac{2}{\theta^3}\sum x_i$

$I(\theta) = E(V(\theta)) = \frac{2n}{\theta^2} - \frac{2\sum x_i}{\theta^3} = \frac{2n}{\theta^2} - \frac{2 \cdot 2n\hat{\theta}}{\theta^3}$

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

CR lower bound

$\therefore \boxed{\frac{1}{I(\theta)} = \frac{\theta^2}{2n}}$

(e) $\hat{\theta} \sim N\left(\hat{\theta}, \frac{\theta^2}{2n}\right)$ $\hat{\theta} = \frac{\bar{x}}{2} = \frac{10.5}{2} = 5.25$

$\hat{\theta} \pm 1.96 \times \sqrt{\frac{\theta^2}{2n}} = 5.25 \pm 1.96 \times \sqrt{\frac{5.25^2}{2 \times 35}} =$

1.229890 (4.02 , 6.47)