THE UNIVERSITY OF MELBOURNE

Semester 2 Final Exam, 2016

School of Mathematics and Statistics

MAST90058 Elements of Statistics

Writing time: 3 hours

Reading time: 15 minutes

This is NOT an open book exam

This paper consists of 5 pages (including this page)

Authorised materials:

- Hand-held scientific calculators (not CAS or graphics) may be used.
- Students may bring three double-sided A4 sheet of handwritten notes.

Instructions to Students

- You may NOT remove this question paper at the conclusion of the examination
- This examination contains 9 questions.
- All questions may be attempted. The total number of marks available is 70.

Instructions to Invigilators

- Students may NOT remove this question paper at the conclusion of the examination
- All graphics or CAS calculators should be confiscated.
- Students may use three double-sided A4 sheet of handwritten notes.

Blank page (ignored in page numbering)

Question 1 (12 marks) Let X be the time headway in traffic flow. Let $X_1, \ldots, X_n$ be a random sample from the the shifted exponential pdf

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

and 0 otherwise. Then time headway observations were made

```
3.11 0.64 2.55 2.20 5.44 3.42 10.39 8.93 17.82 1.30
```

(a) Write the log-likelihood function.

(b) Determine the maximum likelihood estimators of θ and λ.

(c) Determine a two-dimensional sufficient statistic of θ and λ.

(d) Give the Crámer-Rao lower bound of unbiased estimators of θ.

(e) Determine the maximum likelihood estimate of θ and give an approximate 99% confidence interval for θ. Some R output that may help.

```
>  z <- c(0.95,0.975,0.99,0.995)
> qnorm(z)
[1] 1.644854 1.959964 2.326348 2.575829
```

Question 2 (10 marks) Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution on $[0, \theta]$ with pdf

$$f(x;\theta) = \frac{1}{\theta}, \ 0 \leq x \leq \theta,$$

and 0 otherwise. Recall that the maximum likelihood estimator for θ is $Y = \max_{1 \leq i \leq n} X_i$ and it can be shown that Y has pdf $g(y) = ny^{n-1}/\theta^n$ if $0 \leq y \leq \theta$ and 0 otherwise.

(a) Derive an unbiased estiamator of θ using the maximum likelihood estimator Y.

(b) Verify that $P(\alpha^{1/n} \leq Y/\theta \leq 1) = 1-\alpha$ and use this probability statement to find a $100(1-\alpha)\%$ confidence interval for θ.

(c) Suppose your lecturers waiting time for the morning tram is uniformly distributed on $[0, \theta]$ and observed weighting times (in minutes) are

```
3.1  8.0  8.9  9.4  3.7
```

Find a 95% confidence interval for θ.

Question 3 (12 marks) The following data are the lead concentration (μg/l) in eight samples:

```
17.0   21.4   30.6 5.0   12.2   11.8   17.3   18.8
```

corresponding to sample mean $\overline{x} = 16.76$ and sample standard deviation $s = 7.57$.

(a) Give a point estimate of the median, $m = \pi_{0.5}$.

(b) Test the null hypothesis that $H_0 : m = 15$ against the alternative $H_1 : m > 15$ at the 0.05 significance level using the sign test.

(c) Find a confidence interval for the 40th percentile $\pi_{0.40}$ with confidence level close to 90%. Give the exact confidence level of your interval.

(d) Assume the data is normally distributed and find a 95% confidence interval for the mean. How does this compare with your confidence interval for the median?

The following R output may be useful:

```
> z <- c(0.95,0.975,0.99,0.995)
> qnorm(z)
[1] 1.644854 1.959964 2.326348 2.575829
> qt(z, df=7)
[1] 1.894579 2.364624 2.997952 3.499483
> pbinom(0:7, size = 8, prob = 0.4)
[1] 0.02 0.11 0.32 0.59 0.83 0.95 0.99 1.00
> pbinom(0:7, size=8, prob=0.5)
[1] 0.00 0.04 0.14 0.36 0.64 0.86 0.96 1.00
```

Question 4 (10 marks) Suppose that the length of a certain species of fish has a uniform distribution over the interval $[0, \theta]$ where the prior distribution of θ has the Pareto ppdf.

$$h(\theta) = \begin{cases} \frac{\alpha x_0^\alpha}{\theta^{\alpha+1}}, & \theta \geq x_0 \\ 0 & \theta < x_0. \end{cases}$$

The mean for a Pareto random variable with the above pdf is $E[\theta] = \alpha x_0/(\alpha - 1)$ if $\alpha > 1$ and $E[\theta] = \infty$ if $\alpha \leq 1$.

Let $X_1, \ldots, X_n$ be the lengths of a random sample of these fish and let $M = \max\{X_1, \ldots, X_n\}$.

(a) Show the posterior distribution of θ is proportional to

$$\frac{\alpha x_0^\alpha}{\theta^{\alpha+1+n}}, \ \theta > M^* = \max(M, x_0).$$

(b) Hence deduce the posterior density of θ has the Pareto density

$$k(\theta \mid x_1, \ldots, x_n) = \frac{(\alpha + n)M^{*(\alpha+n)}}{\theta^{\alpha+1+n}}, \ \theta > M.$$

(c) Find the posterior mean of the distribution of θ.

(d) If $x_0 = 20$cm, $\alpha = 5$, and the maximum length of a random sample of 15 fish was 30 cm use the posterior mean to estimate θ.

Question 5 (5 marks) A lecturer wishes to dermine what proportion of students in a certain unversity support gay marriage. Let θ denote the proportion of students supporting gay marriage.

(a) What sample size is necessary to obtain a 95% confidence interval for θ of with no smaller than 0.1 irrespective of the true value of θ?

(b) If the lecturer has reason to believe that at least 2/3 of the students support gay marriage, how large a sample size would you recommend.

Question 6 (7 marks) Suppose that among males aged 50-59 the Prostate Specific Antigen (PSA) follows a $N(\mu, 1)$ distribution (i.e. $\sigma = 1$ is assumed known). Consider a a test for the null hypothesis $H_0 : \mu = 4$ rejecting H_0 when $|Z| = \sqrt{n}|\overline{X} - 4| > 1.645$.

(a) Determine the significance level of the above test.

(b) A sample of 9 males is randomly selected. Compute the type II error probability and the power when the true mean for the PSA level is $\mu = 5$.

The following R output may be useful:

```
> pnorm(-1-1.645/3)
[1] 0.06077103
> pnorm(-1+1.645/3)
[1] 0.3257546
> qnorm(c(0.995, 0.975, 0.95))
[1] 2.575829 1.959964 1.644854
>
```

Question 7 (6 marks) A 60 years old female patient with low bone density was treated with drug D for 100 months. Bone density (g/mm^2) was measured irregularly over a period of 100 months (at 0,8,18,48,64,66,79,92 and 100 months from baseline). The following partial R output was obtained from the sample.

```
Coefficients:
             Estimate Std. Error
(Intercept)  0.79498   0.008249
time         0.00088   0.000145

Residual standard error: 0.0132 on 7 degrees of freedom
Multiple R-squared:  0.836
```

(a) Write down the regression model equation and state all the assumptions about the errors. How could you check graphically such assumptions?

(b) Interpret the value of slope and intercept in the context of this study.

(c) Assess whether there is a significant change of bone density over time using an appropriate test at the 0.05 significance level. Provide a test statistic and state your conclusion in the context of these data.

You may use the followint R output:

```
> p <- c(0.995, 0.975, 0.95, 0.90)
> qt(p, df=9)
[1] 3.249836 2.262157 1.833113 1.383029
> qt(p, df=7)
[1] 3.499483 2.364624 1.894579 1.414924
> qt(p, df=8)
[1] 3.355387 2.306004 1.859548 1.396815
```

Question 8 (6 marks) A study investigated the relationship between smoking and kidney cancer. The resulting data are reported in the following data.

		Cases	Controls
No. cigarettes	0	11	36
smoked per day:	1-10	11	16
	>10	16	14

(a) Is there evidence to suggest that the number of cigarettes smoked per day is associated to cancer? Answer by carrying out a chi-square test: state null and alternative hypotheses, give the distribution of the test statistic under the null hypothesis and state your conclusion about the null hypothesis.

(b) Is smoking associated to kidney cancer? Justify your answer using your the results from (a).

You may use the following R output:

```
> p <- c(0.995, 0.975, 0.95, 0.90)
> qchisq(p, df=1)
[1] 7.879439 5.023886 3.841459 2.705543
> qchisq(p, df=2)
[1] 10.596635  7.377759  5.991465  4.605170
> qchisq(p, df=14)
[1] 31.31935 26.11895 23.68479 21.06414
> qchisq(p, df=12)
[1] 28.29952 23.33666 21.02607 18.54935
```