Semester 2 Final Exam, 2016

School of Mathematics and Statistics

MAST20005 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 8 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 median $\pi_{0.50}$ 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 (8 marks) Consider drawing a random sample of size n from $X \sim N(\mu_X, \sigma^2)$ and another sample of size m from $Y \sim N(\mu_Y, \sigma^2)$, where X and Y are independent. Let S_X^2 and S_Y^2 denote the sample variances based on these two samples.

(a) Show that the pooled estimator

$$S_{\text{pool}}^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

is an unbiased estimator of σ^2.

(b) Show that S_{pool}^2 is optimal in the sense that has minimum variance among all combined estimators of form $\alpha S_X^2 + (1-\alpha)S_Y^2$, where α is a constant between 0 and 1.

[Hint: If the Z follows the chi-square distribution $\chi^2(p)$ with p degrees of freedom, then $E(Z) = p$ and $Var(Z) = 2p$.]

Question 8 (6 marks) Let $X_1, \dots, X_n$ be a random sample from a negative binomial distribution with pmf

$$f(x; r, p) = \binom{k+r-1}{k} p(1-p)^r, \quad x = 0, 1, 2, \dots,$$

where $E(X) = r(1-p)/p$ and $Var(X) = r(1-p)/p^2$. Find the method of moment estimator for the parameters p and r.