Durham University

EXAMINATION PAPER

Examination Session:	Year:	Exam Code:
May/June	2022	MATH2711-WE01

Title:
Statistical Inference II

Time:	3 hours	
Additional Material provided:	Formula Sheet; Tables: Normal distribution, t-distribution, chi-squared distribution, signed-rank test statistic, rank-sum test statistic.	
Materials Permitted:		
Calculators Permitted:	Yes	Models Permitted: Casio FX83 series or FX85 series.

Instructions to Candidates:	Answer all questions. Section A is worth 40% and Section B is worth 60%. Within each section, all questions carry equal marks.

Revision:	

ED01/2022

SECTION A

Q1 Suppose that two random variables X_1 and X_2 are independent and identically distributed with pdf

$$f(x_i) = \begin{cases} 2x_i, & 0 < x_i < 1; \\ 0 & \text{otherwise.} \end{cases}$$

Consider the transformations $Y_1 = \frac{X_1}{X_2}$ and $Y_2 = X_1 X_2$.

1.1 Find the joint pdf of Y_1 and Y_2. Clearly define and sketch the region $\mathcal{Y}$ over which this pdf is non-zero.

1.2 Find the marginal pdf of Y_1.

Q2 Let $Y_1, \ldots, Y_n$ denote a random sample of size n from a population whose density is given by

$$f(y; \beta) = \begin{cases} 3\frac{\beta^3}{y^4}, & \beta \leq y \leq \infty; \\ 0, & \text{otherwise;} \end{cases}$$

where $\beta > 0$ is unknown. Consider the estimator $U = \min(Y_1, \ldots, Y_n)$ of β.

2.1 Show that the pdf of U can be expressed as

$$f(u) = 3n \frac{\beta^{3n}}{u^{3n+1}},$$

for $\beta \leq u \leq \infty$ and 0 otherwise.

2.2 Find expressions for the bias and mean-square error of the estimator U.

Q3 An independent, identically distributed sample, $\mathbf{x} = (x_1, \ldots, x_n)^T$, is drawn from a Borel distribution, which is a discrete probability distribution with probability mass function given by

$$f(x|\mu) = \mathbb{P}[X = x \mid \mu] = \frac{1}{x!} e^{-\mu x} (\mu x)^{x-1},$$

where $x \in \{1, 2, 3, \ldots\}$ and $\mu \in [0, 1]$. The mean and the variance of this distribution are

$$\mathbb{E}[X] = \frac{1}{1-\mu} \quad \text{and} \quad \mathbb{V}\text{ar}[X] = \frac{\mu}{(1-\mu)^3},$$

respectively. Suppose that we wish to test the null hypothesis $\mathcal{H}_0 : \mu = \mu_0$ against the alternative hypothesis $\mathcal{H}_1 : \mu = \mu_1$, where $\mu_0 < \mu_1$.

3.1 Derive the likelihood ratio for the data, and show that all likelihood ratio tests of $\mathcal{H}_0$ against $\mathcal{H}_1$ are of the form

$$\text{Reject } \mathcal{H}_0 \text{ if } T(\mathbf{x}) \geq c^*, \text{ for some constant } c^*,$$

where $T(\mathbf{x}) = \sum_{i=1}^n x_i$.

3.2 Suppose that $n = 100$, $\mu_0 = 0.5$, $\mu_1 = 0.75$. By using the central limit theorem, find, approximately,

(i) the sampling distribution for T as a function of μ.

(ii) the value of c^* for which the significance level of the test is 0.01.

Q4 A study compared the performances of engine bearings made of two different types of compounds. Five bearings of each type were tested. The following table gives the times until failure (in units of millions of cycles):

Type A	13.04	6.95	12.95	12.51	13.20
Type B	12.75	7.67	12.78	7.79	9.37

The above data yield the following statistics: $\bar{x} = 11.73$, $\sum_{i=1}^5 x_i^2 = 716.7867$, under Type A compound, and $\bar{y} = 10.072$, $\sum_{i=1}^5 y_i^2 = 533.2008$, under Type B compound.

4.1 Assuming that normal distributions with equal variances are appropriate models for the lifetimes of the two types of bearing, test the hypothesis that there is no difference in mean lifetime between the two types, at a significance level of 5%.

4.2 Without assuming normality, we can use the non-parametric Wilcoxon rank sum method to test the hypothesis that the two types of bearing have the same distribution. Perform this test at a significance level of 5%, using the exact tables of the rank-sum statistic.

4.3 Explain briefly how you would decide which of the two methods used in parts **4.1** and **4.2**, is most appropriate for these data.

SECTION B

Q5 A randomised placebo-controlled trial for the treatment of a severe form of liver disease is conducted. The patients enrolled on the trial are divided into two groups: the placebo group, and the treatment group.

The survival times of the $n = 90$ patients in the placebo group are denoted by x_i, while the survival times of the $m = 94$ patients in the treatment group are y_j. The survival times (in days) of the participants are assumed to follow exponential distributions with rates η and $\theta\eta$ in the placebo and treatment groups, respectively, where $\eta, \theta > 0$. The pdf for an exponentially distributed random variable X with rate parameter $\lambda > 0$ is given by

$$f(x; \lambda) = \lambda e^{-\lambda x},$$

for $x > 0$ and 0 otherwise.

5.1 Give a brief interpretation of the roles of the parameters η and θ in this problem, and find their joint log-likelihood.

5.2 Calculate the MLEs of the parameters η and θ, and also calculate the observed information matrix $\boldsymbol{I}(\hat{\eta}, \hat{\theta})$.

5.3 Show that the standard deviation of the MLE $\hat{\theta}$ can be approximated as

$$\mathrm{sd}(\hat{\theta}) \approx \hat{\theta}\sqrt{\frac{n+m}{nm}},$$

and hence derive a general formula for a $100(1-\alpha)\%$ Wald confidence interval for θ.

5.4 It is proposed to consider a transformation of the parameter θ of the form $\psi = \log(\theta)$. Briefly explain why this may be helpful when comparing the placebo and treatment groups. Derive a general expression for the $100(1-\alpha)\%$ Wald confidence interval for $\psi = \log(\theta)$.

5.5 In the study, it was found that $\sum_{i=1}^{90} x_i = 9344.667$ and $\sum_{j=1}^{94} y_j = 11577.36$. Use your results from **5.3** and **5.4** to calculate 95% confidence intervals for θ for these data. Comment on whether the results are consistent with the treatment having a positive effect on the patient condition.

Q6 The Maxwell-Boltzmann distribution is often used in physical applications to model the speed of a particle moving in three dimensional space. The pdf of the Maxwell-Boltzmann distribution is

$$f(y \mid \theta) = \sqrt{\frac{2}{\pi}}\theta^{-\frac{3}{2}} y^2 \exp\left\{-\frac{y^2}{2\theta}\right\},$$

for $y \geq 0$ and 0 elsewhere, and with parameter $\theta > 0$. A random variable Y following this distribution has $\mathbb{E}[Y] = \sqrt{\frac{8}{\pi}\theta}$ and $\mathbb{V}\mathrm{ar}[Y] = \frac{(3\pi-8)}{\pi}\theta$.

A number of observations $\boldsymbol{y} = [y_1, \ldots, y_n]^T$ related to the burning velocity of different chemicals were recorded in a laboratory experiment. Assume that the observations can be treated as a conditionally i.i.d sample from a Maxwell-Boltzmann distribution given an unknown parameter θ.

6.1 Find the Jeffreys prior for θ. Is this a proper prior? Justify your answer.

6.2 It is proposed that prior beliefs about the parameter θ can be represented by an inverse-Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ and pdf

$$f(\theta \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{\theta^{\alpha+1}} \exp\left\{-\frac{\beta}{\theta}\right\},$$

for $\theta > 0$ and zero otherwise.

Explain briefly what it means for a prior to be conjugate with a particular likelihood. Identify the posterior distribution of θ given $\boldsymbol{y}$ using the inverse-Gamma prior above, and hence determine whether or not this is a conjugate prior for the parameter θ.

6.3 Compare the two priors from **6.1** and **6.2**, and comment on how the Jeffreys prior can be viewed as noninformative for this problem.

6.4 Fifty-five observations of burning velocity were made during the experiment, giving data summarised as below:

$$n = 55, \qquad \sum_{i=1}^{55} y_i = 33.735, \qquad \sum_{i=1}^{55} y_i^2 = 29.549.$$

Using the prior from **6.2** with parameters $\alpha = \beta = 5$, find the posterior distribution for θ and thus find the optimal estimate d^* which minimises the posterior risk under the loss function

$$L(d, \theta) = \theta^2(\theta - d)^2.$$

Q7 Suppose that $Y_i, \ldots, Y_n$, are i.i.d. random variables whose probability mass function is given by

$$\mathbb{P}[Y = y] = \begin{cases} \theta_j, & \text{if } y = j, j \in \{1, 2, 3, 4\}; \\ 0, & \text{otherwise,} \end{cases}$$

where θ_j are unknown parameters such that $\sum_{j=1}^{4} \theta_j = 1$ and $\theta_j \geq 0$. Consider testing

$$\mathcal{H}_0 : \theta_1 = \theta_2 = \theta_3 = \theta_4$$
$$\text{vs.}$$
$$\mathcal{H}_1 : \mathcal{H}_0 \text{ is not true}$$

7.1 What are the full and restricted parameter spaces, Ω and Ω_0, respectively? State their dimensions.

7.2 Suppose that in a sample of n observations, we count n_j occurrences of $Y = j$, for $j = 1, 2, 3, 4$. Find the MLE of $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^T$ under the alternative and under the null hypotheses presented above.

7.3 Find the generalised likelihood ratio test statistic for a level α test of the two hypotheses.

7.4 Explain how we can generally use the generalised likelihood ratio test statistic to test hypotheses. Derive the critical region for testing the two hypotheses in this question at a significance level α. State the exact form of the test for $\alpha = 0.05$.

Q8 **8.1** Show that when we have a continuous parameter θ and a composite vs. composite hypothesis test of the form $\mathcal{H}_0 : \theta \in \Omega_0$ vs. $\mathcal{H}_1 : \theta \in \Omega_1$, the Bayes factor in favour of $\mathcal{H}_0$ against $\mathcal{H}_1$ is given by

$$B_{01} = \frac{\displaystyle\int_{\theta \in \Omega_0} f(\mathbf{x} \mid \theta) f_0(\theta) \mathrm{d}\theta}{\displaystyle\int_{\theta \in \Omega_1} f(\mathbf{x} \mid \theta) f_1(\theta) \mathrm{d}\theta},$$

where $f(\mathbf{x} \mid \theta)$ is the likelihood function of θ and $f_j(\theta)$ is the conditional prior of θ under hypothesis $\mathcal{H}_j$, for $j \in \{0, 1\}$.

8.2 Consider the setting where we have an exponential sampling distribution, with rate parameter $\lambda > 0$ and pdf given by

$$f(x \mid \lambda) = \lambda e^{-\lambda x},$$

when $x \in [0, \infty)$ and 0 elsewhere. We are given a sample $\mathbf{x} = (x_1, \ldots, x_n)^T$ of n observations which we assume to be conditionally i.i.d. given parameter λ. We wish to test the simple hypothesis $\mathcal{H}_0 : \lambda = 2$ vs. the composite alternative hypothesis $\mathcal{H}_1 : \lambda \neq 2$, considering equal prior probabilities for the two hypotheses. Under the alternative hypothesis, we entertain the general assumption of a gamma distribution for λ with prior mean equal to 2 and prior variance also equal to 2. Without providing general proofs:

(i) define the overall prior of λ as a mixture representation of the conditional prior distributions of λ under $\mathcal{H}_0$ and $\mathcal{H}_1$;

(ii) derive the Bayes factor in favour of $\mathcal{H}_0$ against $\mathcal{H}_1$.

8.3 Continuing within the context of **8.2**, suppose that we now observe the following sample of observations:

0.825 1.070 0.143

Draw conclusions based on the resulting posterior probability of $\mathcal{H}_0$ and the Bayes factor in favour of $\mathcal{H}_1$ against $\mathcal{H}_0$.