EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2017

MODULE 2 : Statistical Inference

Time allowed: Three hours

Candidates should answer **FIVE** *questions.*

All questions carry equal marks.
The number of marks allotted for each part-question is shown in brackets.

Graph paper and Official tables are provided.

Candidates may use calculators in accordance with the regulations published in the Society's "Guide to Examinations" (document Ex1).

The notation log *denotes logarithm to base* ***e***.
Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Note also that $\binom{n}{r}$ *is the same as* ${}^{n}C_{r}$.

This examination paper consists of 12 printed pages.
This front cover is page 1.
Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1. (a) $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with probability density function $f(x;\theta)$, where θ is an unknown parameter. $T(X_1, X_2, \ldots, X_n)$ is a function of $X_1, X_2, \ldots, X_n$ which is to be used to estimate θ. Define what is meant by

(i) $T(X_1, X_2, \ldots, X_n)$ is an *efficient unbiased estimator* of θ;

(ii) the *efficiency of an unbiased estimator* of θ;

(iii) $T(X_1, X_2, \ldots, X_n)$ is a *consistent estimator* of θ;

(iv) $T(X_1, X_2, \ldots, X_n)$ is a *sufficient statistic* for θ. [You should explain what is meant by sufficiency and not simply quote the factorisation criterion.]

(6)

(b) Now suppose that $f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}$, $x > 0$, and $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

(i) Show that $\bar{X}$ is an unbiased estimator of θ.

(2)

(ii) State the factorisation criterion for sufficient statistics and use it to show that $\bar{X}$ is a sufficient statistic for θ.

(5)

(iii) State the Cramér-Rao inequality for unbiased estimators of θ, and the Cramér-Rao lower bound associated with this inequality. Show that $\bar{X}$ attains this bound for the distribution above. Explain what this means regarding the efficiency of $\bar{X}$ as an estimator of θ.

(7)

2. Let $X_1, X_2, \ldots, X_n$ be a random sample from the uniform distribution on the interval $(0, \theta)$, where $\theta > 0$.

(i) Find the moment estimator $\tilde{\theta}$ for θ based on the first moment.

(3)

(ii) Show that the maximum likelihood estimator for θ is $\hat{\theta} = \max_i X_i$.

(3)

(iii) Find the mean and variance of $\tilde{\theta}$ and $\hat{\theta}$.

[You are given that the probability density function for $\hat{\theta}$ is

$$g(y) = \begin{cases} \dfrac{ny^{n-1}}{\theta^n} & \text{for } 0 < y < \theta \\ 0 & \text{elsewhere.} \end{cases}$$]

(7)

(iv) Show that $\tilde{\theta}$ and $\hat{\theta}$ are both consistent estimators for θ.

(3)

(v) Find an unbiased estimator for θ of the form $c\hat{\theta}$ for some constant c, and calculate the relative efficiency, and asymptotic relative efficiency, of $\tilde{\theta}$ compared to this unbiased estimator.

(4)

3. (i) State the *Neyman-Pearson lemma*.

(4)

(ii) $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with probability density function $f(x)$. It is required to test the null hypothesis

$$\mathrm{H}_0 : f(x) = 1, \quad 0 < x < 1,$$

against the alternative

$$\mathrm{H}_1 : f(x) = \frac{\beta e^{\beta x}}{e^{\beta} - 1}, \quad 0 < x < 1, \text{ where } \beta > 0 \text{ is known.}$$

Use the Neyman-Pearson lemma to show that the most powerful test of H_0 against H_1 has critical region of the form $\sum_{i=1}^{n} x_i > B$, for some constant B.

(5)

(iii) Define a *uniformly most powerful (UMP) test*.

(4)

(iv) Now suppose that β in part (ii) is unknown, so that the alternative hypothesis H_1 is now composite. Use part (ii) to deduce a UMP test for H_0 against the composite H_1.

(3)

(v) You are given that for n = 2 the probability density function for $Y = X_1 + X_2$ under H_0 is

$$g(y) = \begin{cases} y & \text{for } 0 < y \le 1 \\ 2 - y & \text{for } 1 < y < 2 \\ 0 & \text{elsewhere.} \end{cases}$$

Show that, for a test of size α with $\alpha < \frac{1}{2}$, the UMP test of H_0 against the composite H_1 in part (iv) with n = 2 rejects H_0 when $x_1 + x_2 > 2 - \sqrt{2\alpha}$.

(4)

4. (a) Define what is meant by a *generalised likelihood ratio test*. If λ is the test statistic in such a test, what is the widely used asymptotic distribution of $-2\log\lambda$ for such tests?

(4)

(b) Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the inverse Gaussian distribution with probability density function

$$f(x;\nu,\mu)=\left(\frac{\nu}{2\pi x^3}\right)^{\frac{1}{2}}\exp\left\{\frac{-\nu(x-\mu)^2}{2\mu^2 x}\right\},\quad x>0,\ \mu>0,\ \nu>0.$$

It is required to test the null hypothesis $H_0: \nu=\nu_0$ against the alternative hypothesis $H_1: \nu\neq\nu_0$, where the parameter μ is unknown under both hypotheses.

(i) Given that the maximum likelihood estimate of μ is the mean of the sample, $\bar{x}$, under both H_0 and $H_0 \cup H_1$, show that

$$\hat{\nu}=\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)-\frac{1}{\bar{x}}\right]^{-1}$$

is the maximum likelihood estimate for ν. Hence show that the generalised likelihood ratio test statistic λ is $\left(\frac{\nu_0}{\hat{\nu}}\right)^{\frac{n}{2}}\exp\left\{-\frac{n}{2}\left(\frac{\nu_0}{\hat{\nu}}-1\right)\right\}$ and that the corresponding test has a critical region of the form $\{\hat{\nu}\geq B \text{ or } \hat{\nu}\leq A\}$, where A, B are constants such that $A < B$.

(10)

(ii) A random sample of size 50 is taken from an inverse Gaussian distribution and it is required to test the null hypothesis $H_0: \nu = 2.0$ against a two-sided alternative H_1. The maximum likelihood estimate of ν is 3.2. Use these data and the asymptotic distribution of $-2\log\lambda$ to perform a generalised likelihood ratio test of H_0 versus H_1. Without doing any calculations, explain how you would construct an approximate 95% confidence interval for ν.

(6)

5. (a) (i) The independent random variables $X_1, X_2, \ldots, X_n$ each have the same distribution with a single unknown parameter $\theta > 0$. Let $\hat{\theta}_n$ denote an estimator for θ based on $X_1, X_2, \ldots, X_n$, such that $E(\hat{\theta}_n) = \theta + \frac{k}{n}$, where k is a constant that does not depend on n. Define what is meant by the *jack-knife* estimator of θ and show that it is an unbiased estimator for θ.

(6)

(ii) Now suppose that $\theta = \sigma^2$, the variance of the distribution of the X_i, and $\hat{\theta}_n = \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Deduce that the corresponding jack-knife estimator of σ^2 is unbiased.

(2)

(iii) For large n derive an expression for an approximate 95% confidence interval for σ^2 based on the jack-knife estimator in part (ii).

(5)

(iv) Describe how to find an approximate 95% confidence interval for σ^2 based on $\hat{\sigma}^2$, using a *bootstrap* method.

(4)

(b) Define what is meant by a *pivotal quantity*, and outline how such quantities can be used to construct confidence intervals.

(3)

6. (i) Independent random samples $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ have been drawn from two distributions and it is required to decide whether or not there is significant evidence that values from the first distribution tend to be smaller than those from the second distribution. Formulate this requirement as a test of hypotheses using the Wilcoxon rank sum test and explain how such a test is implemented, when n_1, n_2 are small and when they are large, noting any assumptions that are made.

(6)

(ii) In the Royal Statistical Society tables Table 10, 0.025 level, the critical value corresponding to $n_1 = 5,\ n_2 = 5$ is given as 17. Explain how this value is obtained.

(7)

(iii) Suppose that $n_1 = 15,\ n_2 = 15$. For a one-sided Wilcoxon rank sum test at level 0.05, compare the approximate critical value given by the expression below Table 10 with that in the Table itself.

(3)

(iv) You are now told that for each value of i = 1, 2, …, 15, X_i and Y_i are the scores for the ith student on two different tests. Define the test statistic for the sign test of the hypotheses in part (i) and explain why the sign test might be preferred to the Wilcoxon rank sum test in these circumstances.

(4)

7. (a) Suppose that $x_1, x_2, \ldots, x_n$ are a random sample of observations of a random variable X that has a probability density function depending on a single unknown parameter θ. Define what is meant by a *predictive distribution* for a further observation x_{n+1} in the context of Bayesian inference. (2)

(b) (i) Suppose that the distribution of θ is unimodal. Define the $(1-\alpha)$ *highest posterior density (HPD) credible interval (Bayesian confidence interval)* for a scalar parameter θ, with posterior density $q(\theta)$. (2)

(ii) Suppose again that the distribution of θ is unimodal. Show that the shortest possible $(1-\alpha)$ credible interval for θ is a $(1-\alpha)$ HPD credible interval. (4)

(iii) Suppose that the posterior density of θ is

$$q(\theta) = \begin{cases} \frac{1}{2} & \text{for } 0 \le \theta \le 1 \\ \frac{3}{4} - \frac{1}{4}\theta & \text{for } 1 \le \theta \le 3 \\ 0 & \text{otherwise.} \end{cases}$$

Using a sketch of the posterior density, or otherwise, find a 95% HPD for θ. (4)

(c) In three separate trials a drug is given to n_1, n_2, n_3 patients and the numbers of patients for which the drug is effective are X_1, X_2, X_3 respectively, where X_1, X_2, X_3 are independent binomial random variables with probabilities of success θ_1, θ_2, θ_3. A hierarchical model is to be fitted in which the prior distribution for θ_i, $i = 1, 2, 3$ is beta(α, β) with density function

$$\frac{1}{B(\alpha,\beta)}\theta_i^{\alpha-1}(1-\theta_i)^{\beta-1},$$

and mean $\dfrac{\alpha}{\alpha+\beta}$, and the prior distribution of α and β is

$$g(\alpha,\beta) \propto \text{constant}, \ \alpha > 0, \ \beta > 0.$$

(i) Write down the kernel (the part that depends on the parameters, but not any constants) of the joint posterior distribution of $\theta_1, \theta_2, \theta_3, \alpha, \beta$. (3)

(ii) Are the prior distributions for θ_i, $i = 1, 2, 3$, conjugate priors? Explain your answer. (2)

(iii) Suppose that you have a procedure for simulating values from the joint posterior distribution of $(\alpha, \beta, \theta_1, \theta_2, \theta_3)$. Explain how you could calculate an estimate of the mean of the beta prior distribution. (3)

8. (a) In the context of a decision problem with decision procedures $\{\delta_i, i = 1, 2, \ldots, m\}$, states of nature $\{\theta_j, j = 1, 2, \ldots, n\}$, risk function $\{R(\delta_i, \theta_j), i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n\}$ and probabilities $\{p_j, j = 1, 2, \ldots, n\}$ of the states of nature, define the terms *Bayes decision procedure* and *admissible decision procedure*. Explain whether a Bayes decision procedure is necessarily admissible when all $p_j > 0$ in the context above.

(5)

(b) A manufacturing company is about to launch a new product. In a simplified model of the company's marketing strategy there are two states of nature that describe the market potential of its product as being 'high' or 'low' and there are two possible actions open to the company, d_1 and d_2. The loss function (in coded units) is as follows.

Action	Market Potential	
	High (θ_1)	Low (θ_2)
d_1	1	4
d_2	4	2

To determine which strategy to adopt, the company will conduct a survey. The response is regarded as favourable if at least 30% of sampled customers express a liking for the product, and unfavourable otherwise. It is known that if the market potential is high then there is a probability of 0.9 of a favourable response, whereas if potential is low the corresponding probability is 0.2. The company requires a decision procedure specifying which action to take according to whether the survey response is favourable or unfavourable.

(i) List all possible decision procedures.

(2)

(ii) Determine the risk function for each procedure.

(4)

(iii) Identify, with a reason, any inadmissible decision procedures.

(2)

(iv) Find the minimax decision procedure.

(2)

(v) Let *p* be the prior probability that the market potential is high. Find the range of values of *p* for which the minimax decision procedure is also the Bayes decision procedure. Which other decision procedures can be Bayes decision procedures, and for what values of *p*?

(5)

BLANK PAGE

BLANK PAGE

BLANK PAGE