EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2013

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* ${}^nC_r$.

This examination paper consists of 8 printed pages.
This front cover is page 1.
Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1. $W_1, W_2, \ldots, W_{n_1}$ are independent observations from a Poisson distribution with mean λ. $X_1, X_2, \ldots, X_{n_2}$ are independent observations from a Poisson distribution with mean $\alpha\lambda$. $Y_1, Y_2, \ldots, Y_{n_3}$ are independent observations from a Poisson distribution with mean $\alpha^2\lambda$. Here α and λ are positive parameters.

Firstly suppose that λ is known and α is unknown.

(i) Find the maximum likelihood estimator of α.

(8)

(ii) Assuming that the sample sizes are large, find an approximate 95% confidence interval for α. [You may assume that the appropriate regularity conditions hold.]

(5)

Suppose now that both λ and α are unknown.

(iii) Derive the likelihood function and show that $\left(\sum_{i=1}^{n_1} W_i + \sum_{i=1}^{n_2} X_i + \sum_{i=1}^{n_3} Y_i\right)$ and $\left(\sum_{i=1}^{n_2} X_i + 2\sum_{i=1}^{n_3} Y_i\right)$ are jointly sufficient for α and λ.

(3)

(iv) Find two equations satisfied by the maximum likelihood estimators of α and λ. [For this part, you may assume that the likelihood is maximised when both the first partial derivatives of the log likelihood are equal to zero.]

(4)

2. According to a genetic theory, the proportion of individuals in population 1 exhibiting a certain characteristic is p and the proportion in population 2 is $\frac{1}{2}p$. Independent random samples of n_1 and n_2 individuals are selected from populations 1 and 2 and X_1 and X_2 respectively are found to have the characteristic, so that X_1 and X_2 have binomial distributions. It is required to test the null hypothesis that $p = \frac{1}{2}$ against the alternative hypothesis that $p = \frac{2}{3}$.

(i) Show that the most powerful test has critical region of the form

$$X_1 \log(2) + X_2 \log(1.5) \geq k,$$

where k is a constant.

(11)

(ii) Use Normal approximations to find k so that the significance level of the test is approximately 5%.

(9)

3. U_1, U_2, …, U_n are independent observations from a uniform distribution between 0 and θ, where θ (> 0) is an unknown parameter. Denote the maximum of U_1, U_2, …, U_n (i.e. $\max(U_i)$) by X.

(i) Find and sketch the likelihood function and hence show that the maximum likelihood estimator of θ is X.

(6)

(ii) By noting that $X \le x$ if and only if all of U_1, U_2, …, U_n are less than or equal to x, show that the probability density function of X is $f_X(x) = \dfrac{nx^{n-1}}{\theta^n}$ for $0 < x < \theta$.

(3)

(iii) Show that $Y = \dfrac{X}{\theta}$ is a pivotal quantity for θ and hence find a 95% confidence interval for θ with lower limit X.

(7)

(iv) Explain why the standard (percentile) bootstrap method for finding a confidence interval for θ based on the maximum of each bootstrap sample is not appropriate in this case.

(4)

4. The time, in seconds, for a chemical reaction to take place at pressure i (i = 1, 2 and 3) is

$$f_i(t) = \begin{cases} 2\beta_i t e^{-\beta_i t^2} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

where β_1, β_2, β_3 are unknown positive constants. Independent observations $T_{i1}, T_{i2}, \ldots, T_{in}$ are made at each of pressures i = 1, 2 and 3. It is required to test the hypotheses

$$\mathrm{H}_0 : \frac{\beta_2}{\beta_1} = \frac{\beta_3}{\beta_2} \text{ versus } \mathrm{H}_1 : \frac{\beta_2}{\beta_1} \neq \frac{\beta_3}{\beta_2}.$$

(i) Find the maximum likelihood estimators of β_1, β_2, β_3.

(7)

(ii) By noting that, under H_0, $\beta_3 = \dfrac{\beta_2^2}{\beta_1}$, show that the restricted (or constrained) maximum likelihood estimators under H_0 satisfy

$$\hat{\hat{\beta}}_2 = \frac{3n}{\sum T_{2j}^2 + 2\sqrt{\sum T_{1j}^2 \sum T_{3j}^2}}, \quad \hat{\hat{\beta}}_1 = \beta_2 \sqrt{\frac{\sum T_{3j}^2}{\sum T_{1j}^2}} \quad \text{and} \quad \hat{\hat{\beta}}_3 = \frac{\hat{\hat{\beta}}_2^2}{\hat{\hat{\beta}}_1}.$$

[You may assume that the restricted maximum likelihood estimators can be found by setting the first partial derivatives to zero.]

(7)

(iii) Carry out the generalised likelihood ratio test at the 5% significance level for the case n = 20, $\sum_{j=1}^{n} T_{1j}^2 = 2.5$, $\sum_{j=1}^{n} T_{2j}^2 = 1.8$ and $\sum_{j=1}^{n} T_{3j}^2 = 1.0$, explaining clearly any asymptotic result that you use.

(6)

5. Independent observations X_1, X_2, …, X_n are available from the logarithmic distribution with unknown parameter α ($0 < \alpha < 1$) given by

$$P(X = x) = \frac{\alpha^x}{x(-\log(1-\alpha))} \text{ for } x = 1, 2, \ldots .$$

(i) It is required to test the null hypothesis $H_0 : \alpha = \frac{1}{2}$ against the alternative $H_1 : \alpha = \frac{3}{4}$, and the prior probability of H_0 is $\frac{1}{3}$. Find the prior odds of H_0 and the Bayes Factor and show that the posterior odds of H_0 is $\frac{2^{n-1+S}}{3^S}$, where $S = \sum X_i$.

(9)

Suppose instead that the prior distribution of α has probability density function

$$\pi(\alpha) = \theta\alpha^{\theta-1} \text{ for } 0 < \alpha < 1,$$

where θ is a known positive constant.

(ii) Find the posterior probability density function of α up to a constant of proportionality and find an expression for this constant in terms of an integral.

(5)

(iii) State, with reasons, whether the prior and likelihood functions are conjugate.

(3)

(iv) Suppose that A_1, A_2, …, A_{1000} constitute a random sample from the posterior distribution of α. Describe how this sample could be used to determine the predictive probability that X is less than or equal to 2.

(3)

6. What is meant by

(i) a *decision rule*,

(ii) a *minimax decision*,

(iii) a *Bayes decision*?

(5)

Show that the Bayes decision chooses an action which minimises the posterior expected loss.

(5)

The posterior distribution of the parameter p is

$$\pi(p \mid x) = \begin{cases} 12p^2(1-p) & 0 < p < 1 \\ 0 & \text{otherwise} \end{cases}$$

and it is required to estimate p. Evaluate the Bayes decision when the loss associated with estimating p by $\hat{p}$ is $(p-\hat{p})^2$.

(4)

Suppose instead that the loss is 1 if $|p-\hat{p}| > 0.1$ and 0 if $|p-\hat{p}| \le 0.1$. Use a sketch graph to explain the condition satisfied by the Bayes decision and describe how it could be evaluated iteratively.

(6)

7. $X_1, X_2, \ldots, X_{2n}$ are independent observations from the Bernoulli distribution with unknown parameter p ($0 < p < 1$) i.e.

$$P(X_i = x) = \begin{cases} 1-p & x = 0 \\ p & x = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{for } i = 1, 2, \ldots, 2n,$$

and let $Y = \sum_{i=1}^{2n} X_i$. It is required to estimate $p^2 = \theta$, say.

(i) Show that $\hat{\theta} = \left(\frac{Y}{2n}\right)^2$ is a biased estimator of θ. [You may use any results concerning standard distributions without proof but they must be stated clearly.]

(4)

(ii) Show that the jack-knife estimator of θ based on $\hat{\theta}$ is

$$\frac{Y(Y-1)}{2n(2n-1)},$$

and show that it is unbiased for p^2. [Hint: note that $Y = \sum_{i=1}^{2n} X_i^2$.]

(6)

(iii) Define $W_i = X_{2i-1}X_{2i}$ for $i = 1, 2, \ldots, n$. Show that $\tilde{\theta} = \frac{\sum_{i=1}^{n} W_i}{n}$ is an unbiased estimator of θ and find its variance.

(5)

(iv) Find the Cramér-Rao lower bound for the variance of unbiased estimators of θ and deduce the efficiency of $\tilde{\theta}$.

(5)

8. (a) Describe how *Spearman's rank correlation coefficient* is related to the product-moment correlation coefficient.

(2)

Explain carefully how permutations are used to evaluate the statistical significance of a particular value of the Spearman's rank correlation coefficient in testing the null hypothesis of no association between two variables against a two-sided alternative. Illustrate your answer by evaluating the significance level for the case when two judges independently rank five candidates identically.

(8)

(b) Give an account of the importance in estimation of *unbiasedness*, in relation to other desirable properties of estimators.

(10)