Queen Mary
University of London

B. Sc. Examination by course unit 2009

MTH6136 Statistical Theory

Duration: 2 hours

Date and time: 29 May 2009, 1000h–1200h

Apart from this page, you are not permitted to read the contents of this question paper until instructed to do so by an invigilator.

You should attempt all questions. Marks awarded are shown next to the questions.

Calculators ARE permitted in this examination. The unauthorized use of material stored in pre-programmable memory constitutes an examination offence. Please state on your answer book the name and type of machine used.

Statistical functions provided by the calculator may be used provided that you state clearly where you have used them.

The New Cambridge Statistical Tables are provided.

Complete all rough workings in the answer book and cross through any work which is not to be assessed.

Candidates should note that the Examination and Assessment Regulations state that possession of unauthorized materials by any candidate who is under examination conditions is an assessment offence. Please check your pockets now for any notes that you may have forgotten that are in your possession. If you have any, then please raise your hand and give them to an invigilator now.

Exam papers must not be removed from the examination room.

Examiner(s): B. Bogacka

Question 1 Let $Y_i \underset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, $i = 1, \ldots, n$, be a random sample. Consider the following estimators of μ:

$$\widehat{\mu}_1 = \frac{1}{2}(Y_1 + Y_2)$$
$$\widehat{\mu}_2 = \frac{1}{4}Y_1 + \frac{Y_2 + \ldots + Y_{n-1}}{2(n-2)} + \frac{1}{4}Y_n$$
$$\widehat{\mu}_3 = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

(a) Show that each of the three estimators is unbiased. [4]

(b) Calculate the variance of each estimator and check the estimators' consistency. [7]

(c) Give the definition of the efficiency of an estimator and check which of the three estimators is efficient. [13]

Question 2 Suppose that $Y_1, Y_2, \ldots, Y_n$ are independent random variables with probability density function

$$f(y|\phi) = \begin{cases} \frac{2y}{\phi} \exp\left\{-\frac{y^2}{\phi}\right\}, & \text{for } y \geq 0; \\ 0, & \text{otherwise,} \end{cases}$$

where $\phi > 0$.

(a) Find a sufficient statistic for ϕ. [5]

(b) Obtain the maximum likelihood estimator of ϕ and show that it is a function of the sufficient statistic. [7]

Question 3 Suppose that $Y_1, Y_2, \ldots, Y_n$ are independent random variables with probability density function

$$f(y|\theta) = \begin{cases} \frac{1}{\theta} y^{-(\theta^{-1}+1)}, & \text{for } y > 1; \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta > 0$.

(a) Show that the distribution belongs to the exponential family of distributions. [9]

(b) Give a complete sufficient statistic for θ. [2]

(c) Use integration by parts to find the expectation of this complete sufficient statistic and hence obtain an unbiased estimator of θ. [11]

(d) What can you conclude about this estimator's properties? [2]

Question 4 Suppose that $Y_i \underset{iid}{\sim} \text{Bernoulli}(p_1)$, $i = 1, \ldots, n_1$ and $X_i \underset{iid}{\sim} \text{Bernoulli}(p_2)$, $i = 1, \ldots, n_2$, are two independent random samples.

(a) Knowing that, for $j = 1, 2$ independently, the MLE(p_j) is

$$\widehat{p}_j \sim \mathcal{AN}\left(p_j, \frac{p_j(1-p_j)}{n_j}\right)$$

obtain the approximate pivot for the difference of the two parameters, $p_1 - p_2$. Here $\mathcal{AN}$ means "asymptotically normal". [6]

(b) Derive a 95% confidence interval for $p_1 - p_2$. [6]

(c) In a political campaign one candidate has two polls taken at random: one among the voting population in Ohio State and one among the voting population in Texas. The question was "Will you vote for me?" Possible answers were "Yes" or "No" in each state. The results are: In Ohio State 350 out of $n_1 = 560$ answers are "Yes" and in Texas 290 out of $n_2 = 510$ answers are "Yes". Calculate 95% confidence interval for the difference in the proportions of positive answers in the two states. Is there evidence that the candidate is more popular in one of the two states? Explain your answer. [8]

Question 5 Suppose that $Y_1, \ldots, Y_n$ are independent $\text{Bin}(20, p)$ random variables and consider testing $H_0 : p = p_0$ against $H_1 : p = p_1$, where $p_1 > p_0$.

(a) Write down the likelihood, $L(p; \underline{y})$, and hence find the likelihood ratio given by [6]

$$\lambda(\underline{y}) = \frac{L(p_0; \underline{y})}{L(p_1; \underline{y})}.$$

(b) Show that [2]

$$\log\left\{\frac{p_0(1-p_1)}{p_1(1-p_0)}\right\} < 0.$$

(c) Use the Neyman-Pearson lemma to obtain the general form of the critical region of the most powerful test of H_0 against H_1. For large samples, find the form of the critical region for a test at the significance level $\alpha = 0.01$. [8]

(d) Show that a uniformly most powerful test of $H_0 : p = p_0$ against $H_1 : p > p_0$ exists and write down its critical region when the significance level is α. [4]

End of Paper