Queen Mary
University of London

Main Examination period 2018

MTH6136 / MTH6136P: Statistical Theory

Duration: 2 hours

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 available are shown next to the questions.

Only non-programmable calculators that have been approved from the college list of non-programmable calculators are permitted in this examination. Please state on your answer book the name and type of machine used.

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

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any notes, mobile phones, smartwatches or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately.

It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms, it shall be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

Exam papers must not be removed from the examination room.

Examiners: Dr D. S. Stark, Dr D. S. Coad

Question 1. [20 marks] Suppose that $Y_1, \ldots, Y_n$ are independent beta random variables with probability density function

$$f_Y(y) = \theta(1-y)^{\theta-1}, \quad 0 < y < 1,$$

where $\theta > 0$.

(a) Use Neyman's factorisation theorem to show that $\prod_{i=1}^n (1 - Y_i)$ is a sufficient statistic for θ. **[7]**

(b) Show that the Cramér-Rao lower bound for unbiased estimators of θ^{-1} is $(n\theta^2)^{-1}$. **[6]**

(c) (i) Show that Y is a member of the exponential family of distributions and use the Lehmann-Scheffé Theorem to show that the statistic $\sum_{i=1}^n \log(1 - Y_i)$ is complete and therefore that the statistic $\prod_{i=1}^n (1 - Y_i)$ is also complete. **[4]**

(ii) Given that $\mathbb{E}(\log(1 - Y)) = -1/\theta$, explain why $-\sum_{i=1}^n \log(1 - Y_i)/n$ is the unique minimum variance unbiased estimator of $1/\theta$. **[3]**

Question 2. [20 marks] Let $Y_1, \ldots, Y_n$ be independent $\text{Bin}(m, \pi)$ random variables, where m is known.

(a) Consider the estimator for π

$$T_n = \frac{1}{(n+1)m} \sum_{i=1}^n Y_i.$$

Show that

$$\text{bias}(T_n) = -\frac{\pi}{n+1}$$

and

$$\text{Var}(T_n) = \frac{n\pi(1-\pi)}{(n+1)^2 m}.$$

Show that the sequence of estimators T_n is consistent. **[8]**

(b) Show that the least squares estimator of π is

$$\frac{1}{m}\overline{Y} = \frac{1}{mn} \sum_{i=1}^n Y_i.$$

[8]

(c) Now, suppose that $Y_i \sim \text{Bin}(m, \pi_i)$ independently for $i = 1, 2, \ldots, n$, where the π_i are not all equal. Why is it no longer appropriate to use least squares as a method of estimation? **[4]**

Question 3. [20 marks] Suppose that $Y_1, \ldots, Y_n$ are independent Pareto distributed random variables with mean $2\theta/(\theta-1)$ and probability density function

$$f_Y(y) = \frac{\theta 2^\theta}{y^{\theta+1}}, \quad y \geq 2,$$

where $\theta > 1$.

(a) Show that the maximum likelihood estimator of θ is

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n \log(Y_i/2)}.$$ **[7]**

(b) Show that the Cramér-Rao lower bound for estimating θ is θ^2/n and obtain the asymptotic distribution of $\hat{\theta}$. Hence, write down an approximate $100(1-\alpha)\%$ confidence interval for θ. **[7]**

(c) Show that the method of moments estimator of θ is

$$\frac{\overline{Y}}{\overline{Y}-2},$$

where $\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$. **[6]**

Question 4. [20 marks] Suppose that $Y_1, \ldots, Y_{n_1}$ are $\mathrm{N}(\mu_1, \sigma^2)$ random variables and $Y_{n_1+1}, \ldots, Y_{n_1+n_2}$ are $\mathrm{N}(\mu_2, \sigma^2)$ random variables, all independent, where σ^2 is known.

(a) Show that the maximum likelihood estimators of μ_1 and μ_2 are

$$\hat{\mu}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_i$$

and

$$\hat{\mu}_2 = \frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2} Y_i.$$ **[7]**

(b) State a pivot for $\mu_1 - \mu_2$ and give an exact $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$. **[7]**

(c) Use the confidence interval found in part (b) to obtain a test of $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$ at the 5% level of significance. **[6]**

Question 5. [20 marks] Let $Y_1, \ldots, Y_n$ be independent mean zero normal random variables with probability density function

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{y^2}{2\sigma^2}\right\}, \qquad -\infty < y < \infty,$$

where $\sigma^2 > 0$, and consider testing $H_0 : \sigma = \sigma_0$ against $H_1 : \sigma = \sigma_1$ where $\sigma_1 > \sigma_0$ for fixed σ_0 and σ_1.

(a) Write down the likelihood, $L(\sigma^2; \underline{y})$, and hence find the likelihood ratio given by $\Lambda(\underline{y}) = L(\sigma_0^2; \underline{y})/L(\sigma_1^2; \underline{y})$. [**8**]

(b) Show that the general form of the most powerful test of H_0 against H_1 is to reject H_0 if $\sum_{i=1}^n y_i^2 > c$ for a constant c. [**6**]

(c) Given that under H_0, $\sum_{i=1}^n Y_i^2/\sigma_0^2 \sim \chi_n^2$, derive the form of the critical region of the test with significance level α. [**6**]

End of Paper.