UNIVERSITY OF CAMBRIDGE

M. PHIL. IN STATISTICAL SCIENCE

Thursday 3 June 2004 1.30 to 4.30

STATISTICAL THEORY

Attempt **FOUR** *questions, not more than* **TWO** *of which should be from Section B.*

There are **ten** *questions in total.*

The questions carry equal weight.

You may not start to read the questions printed on the subsequent pages until instructed to do so by the Invigilator.

Section A

1 Let $X_1, \ldots, X_n$ be independent random variables with density $f(x)$. Define what is meant by a *kernel* $K(x)$, and by the *kernel density estimate* $\hat{f}_h(x)$ of $f(x)$, with kernel K and bandwidth $h > 0$.

Define the mean integrated squared error (MISE) of $\hat{f}_h$, and derive an exact expression for this quantity in terms of f and the scaled kernel K_h, where $K_h(x) = h^{-1}K(x/h)$.

For a symmetric, second-order kernel, under regularity conditions, the minimum value of the asymptotic MISE may be expressed as

$$\inf_{h>0} AMISE(\hat{f}_h) = \frac{5}{4}\{\mu_2(K)^2 R(K)^4 R(f'')\}^{1/5} n^{-4/5},$$

where $\mu_2(K) = \int_{-\infty}^{\infty} x^2 K(x)\,dx$, and $R(g) = \int_{-\infty}^{\infty} g(x)^2\,dx$ for a square integrable function $g : \mathbb{R} \to \mathbb{R}$. Show that $R(f'')$ may be made arbitrarily small by means of a scale transformation $af(ax)$ of $f(x)$, but that

$$D(f) = \sigma(f)^5 R(f'')$$

is scale invariant, where

$$\sigma(f)^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \left(\int_{-\infty}^{\infty} x f(x)\,dx\right)^2.$$

Let

$$f_0(x) = \frac{35}{32}(1-x^2)^3 \mathbf{1}_{\{|x|<1\}},$$

and let $h(x)$ be another twice continuously differentiable density satisfying $\int_{-\infty}^{\infty} xh(x)\,dx = 0$ and $\sigma(h) = \sigma(f_0)$. By considering $e(x) = h(x) - f_0(x)$ or otherwise, show that $R(h'') \geqslant R(f_0'')$.

2 Let $X_1, \ldots, X_n$ be independent random variables with continuous distribution function $F(x)$. Define the empirical distribution function, $\hat{F}_n(x)$, and show that the distribution of

$$D_n = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)|$$

does not depend on F. Explain how this result may be used to construct a confidence band for F of $(1-\alpha)$-level coverage.

State the Glivenko-Cantelli theorem.

Now suppose $X_1, \ldots, X_n$ are independent with distribution function F, and that $\theta = \theta(F)$ is a parametric function which may be expressed as $\theta(F) = \mathbb{E}_F\{h(X_1, \ldots, X_r)\}$. Explain why we may always choose h to be symmetric in its arguments. For $n \geqslant r$, define what is meant by a U-statistic for θ with kernel h.

Let $\theta(F)$ denote the variance of a random variable with distribution function F. Find a function $h : \mathbb{R}^2 \to \mathbb{R}$ which is symmetric in its arguments and which satisfies $\theta(F) = \mathbb{E}_F\{h(X_1, X_2)\}$. Evaluate the corresponding U-statistic and simplify your answer as much as possible.

3 Write brief accounts about Edgeworth expansions and saddlepoint approximations to the densities of sums of independent, identically distributed random variables. You should include a description of any notable ways in which the approximations differ.

Let $Y_1, \ldots, Y_n$ be independent random variables with the Laplace density

$$f_Y(y) = \frac{1}{2}\, e^{-|y|}, \quad y \in \mathbb{R},$$

for which the cumulant generating function is $K_Y(t) = -\log(1-t^2)$ for $|t| < 1$. Compute the Edgeworth expansion and saddlepoint approximation to the density of $S_n = \sum_{i=1}^n Y_i$, up to, but not including, terms of order n^{-1}.

4 Describe in detail *three* commonly-used techniques of bandwidth selection in kernel density estimation, mentioning briefly their asymptotic properties.

Hint: you may find the following formulae helpful:

$$h_{AMISE} = \left(\frac{R(K)}{R(f'')\mu_2(K)^2 n} \right)^{1/5}, \; AMISE(\hat{f}_h) = \frac{1}{nh} R(K) + \frac{1}{4} h^4 \mu_2(K)^2 R(f''),$$

and

$$\mathbb{E}\{R(\hat{f}_h'')\} = R(f'') + \frac{1}{nh^5} R(K'') + O(h^2)$$

as $n \to \infty$. *When estimating* $R(f'')$ *by* $\hat{R}_g^{(2)} = n^{-1} \sum_{i=1}^n \hat{f}_g^{(4)}(X_i)$, *the optimal AMSE bandwidth is*

$$g_{AMSE} \propto R(f''')^{-1/7} n^{-1/7}.$$

5 Give a brief description of marginal and profile likelihoods, contrasting the ways in which they are used to deal with nuisance parameters.

Let $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ be independent exponential random variables with $X_1, \ldots, X_m$ having mean $1/(\psi\lambda)$ and $Y_1, \ldots, Y_n$ having mean $1/\lambda$. Further, let $X = \sum_{i=1}^{m} X_i$ and $Y = \sum_{i=1}^{n} Y_i$. Write down the joint density of X and Y. Consider the transformation

$$T = \frac{X}{Y}, \quad U = Y.$$

By first computing the joint density of T and U, find the marginal density of T and show that the marginal log-likelihood for ψ based on T is

$$\ell(\psi; t) = m \log \psi - (m+n) \log(\psi t + 1).$$

Compute the maximum likelihood estimate of λ for fixed ψ, and hence show that the profile log-likelihood for ψ is identical to $\ell(\psi; t)$ above.

6 Describe the Wald, score and likelihood ratio tests for hypotheses concerning a multidimensional parameter θ. Explain briefly how they can be used to construct confidence regions for θ of approximate $(1-\alpha)$-level coverage.

Let $Y_0, Y_1, \ldots, Y_n$ be a sequence of random variables such that Y_0 has a Poisson distribution with mean θ and for $i \geqslant 1$, conditional on $Y_0, \ldots, Y_{i-1}$, the random variable Y_i has a Poisson distribution with mean θY_{i-1}. The parameter θ satisfies $0 < \theta \leqslant 1$. Find the log-likelihood for θ, and show that the maximum likelihood estimator, $\hat{\theta} = \hat{\theta}(Y_0, Y_1, \ldots, Y_n)$, may be expressed as $\hat{\theta} = \min(\tilde{\theta}, 1)$, where $\tilde{\theta} = \tilde{\theta}(Y_0, Y_1, \ldots, Y_n)$ is a function which should be specified.

For $\theta \in (0,1)$, compute the Fisher information $i(\theta)$, and show that

$$i(\theta) \leqslant \frac{1}{\theta(1-\theta)}$$

for all n.

Deduce that the Wald statistic for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$, where $0 < \theta_0 < 1$, does not have an asymptotic chi-squared distribution under the null hypothesis.

Section B

7 i) Suppose $(Y|U = u)$ has a Poisson distribution, with mean μu, and U has probability density function $f(u)$, where

$$f(u) = \theta^{\theta} u^{\theta-1} e^{-\theta u} / \Gamma(\theta), \quad \text{for } u \geq 0.$$

Show that

a) $E(Y) = \mu, \ \ var(Y) = \mu + \mu^2/\theta$,

b) Y has frequency function

$$g(y|\mu) = \frac{\Gamma(\theta + y)\mu^y \theta^{\theta}}{\Gamma(\theta) y! (\mu + \theta)^{\theta+y}} \text{ for } y = 0, 1, 2, \ldots.$$

ii) If $(Y_1, \ldots, Y_n)$ are independent observations, and Y_i has frequency function $g(y_i|\mu_i)$, where $\log \mu_i = \beta x_i$, and $x_1, \ldots, x_n$ are given, describe how to estimate β in the case where θ is a known parameter, and derive the asymptotic distribution of your estimator.

8 Let $Y_1, \ldots, Y_n$ be independent variables, such that

$$Y = \mu 1 + X\beta + \epsilon,$$

where X is a given $n \times p$ matrix of rank p, β is an unknown vector of dimension p, μ is an unknown constant, and 1 is the n-dimensional vector with every element 1. Assume that $X^T 1 = 0$, and that $\epsilon \sim N(0, \sigma^2 I)$, where σ^2 is unknown.

i) Derive an expression for $\hat{\beta}$, the least squares estimator of β, and derive its distribution.

ii) How would you test $H_0 : \beta = 0$?

iii) How would you check the assumption $\epsilon \sim N(0, \sigma^2 I)$?

(You may quote any standard theorems needed.)

9 What is meant by an improper prior in a Bayesian analysis?

Let $X_1, \ldots, X_n$ be independent identically distributed $N(\mu, \sigma^2)$, with both μ and σ^2 unknown. Suppose that μ and σ^2 are given independent prior densities. Show that in the case of the improper prior $\pi(\mu) \propto 1$ for μ, the marginal posterior density of σ^2 depends only on the sample variance $s^2 = (n-1)^{-1} \Sigma_{i=1}^n (x_i - \bar{x})^2$.

Show further that in the case of improper priors $\pi(\mu) \propto 1$, $\pi(\sigma^2) \propto \sigma^{-2}$, the posterior distribution of σ^2 is that of $(n-1)s^2/V$, where $V \sim \chi^2_{n-1}$.

10 Write an account of the main results in the frequentist (Neyman-Pearson) theory of optimal hypothesis testing. Your account should include discussion of all of the following: size of a test, Neyman-Pearson Lemma, uniformly most powerful tests, and unbiased tests.

(Proofs of results are not expected.)