

PAPER 42

Statistical Theory

*Attempt **FOUR** questions.*

*There are **six** 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.**

**1** Let  $X_1, \dots, X_n$  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  $g : \mathbb{R} \rightarrow \mathbb{R}$ . 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  $R(h'') \geq R(f_0'')$ .

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

Let  $\theta(F)$  denote the variance of a random variable with distribution function  $F$ . Find a function  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$  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  $Y_1, \dots, Y_n$  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:*

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

and

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

as  $n \rightarrow \infty$ . 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  $X_1, \dots, X_m, Y_1, \dots, Y_n$  be independent exponential random variables with  $X_1, \dots, X_m$  having mean  $1/(\psi\lambda)$  and  $Y_1, \dots, Y_n$  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  $\psi$  based on  $T$  is

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

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

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

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

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

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

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.