UNIVERSITY OF CAMBRIDGE

M. Phil. in STATISTICAL SCIENCE

Monday, 1 June, 2009   1:30 pm to 3:30 pm

STATISTICAL THEORY

Attempt no more than ***THREE*** *questions.*

There are ***FOUR*** *questions in total.*

The questions carry equal weight.

STATIONERY REQUIREMENTS
Cover sheet
Treasury Tag
Script paper

SPECIAL REQUIREMENTS
None

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

1 (a) Let A be a symmetric $n \times n$ matrix of rank $n-p$, and let B be a $p \times n$ matrix of rank p. Suppose that $BA = 0$. You are given that we can write $A = LL^T$, where L is an $n \times (n-p)$ matrix of rank $n-p$. Show that $L^T L$ is positive definite, and by considering $BLL^T L(L^T L)^{-1}$, show that $BL = 0$.

Let $Y \sim N_n(\mu, \sigma^2 I)$. Find the distribution of the random vector $Z = \begin{pmatrix} BY \\ L^T Y \end{pmatrix}$, and deduce that BY and $Y^T AY$ are independent.

(b) Consider the linear model $Y = X\beta + \epsilon$, where X is an $n \times p$ design matrix of full rank p $(< n)$, $\beta \in \mathbb{R}^p$ is an unknown vector of regression coefficients and $\epsilon \sim N_n(0, \sigma^2 I)$. Write down expressions for the maximum likelihood estimators $\hat{\beta}$ and $\hat{\sigma}^2$, and also write down their marginal distributions.

Using the result of part (a), or otherwise, show carefully that $\hat{\beta}$ and $\hat{\sigma}^2$ are independent.

2 For $n = 1, 2, \ldots$, let $Y = (Y_1, \ldots, Y_n)^T$ have independent and identically distributed components with density $f(\cdot;\theta)$ for some $\theta \in \Theta \subseteq \mathbb{R}^d$ on a sample space $\mathcal{Y}$. Let θ_0 denote the true value of θ. Assume Θ is closed and bounded and that for each $y \in \mathcal{Y}$, the likelihood $L(\theta; y)$ is a continuous function of θ. Suppose that, for each n, the maximum likelihood estimator $\hat{\theta}_n$ based on $Y_1, \ldots, Y_n$ is unique, the model is identifiable and $\mathbb{E}_{\theta_0}\bigl\{\sup_{\theta \in \Theta} |\log f(Y_1;\theta)|\bigr\} < \infty$.

Prove that $\hat{\theta}_n$ is consistent; i.e. $\hat{\theta}_n \xrightarrow{p} \theta_0$ as $n \to \infty$.

[You may use the fact that $\mathbb{E}_{\theta_0}\{\log f(Y_1;\theta)\}$ is a continuous function of θ and

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \log f(Y_i;\theta) - \mathbb{E}_{\theta_0}\{\log f(Y_1;\theta)\} \right| \xrightarrow{p} 0$$

as $n \to \infty$].

Under regularity conditions that you need *not* specify, state a result about the asymptotic normality of $\hat{\theta}_n$.

Now let $Y_1, \ldots, Y_n$ be independent $U[0, \theta]$ random variables. Find $\hat{\theta}_n$ and prove from first principles that $\hat{\theta}_n$ is consistent. By considering the distribution function of $n(\theta - \hat{\theta}_n)/\theta$, show that $\hat{\theta}_n = \theta + o_p(n^{-1/2})$ as $n \to \infty$. Give one regularity condition for your asymptotic normality result that is violated in this case.

3 Let $X = (X_1, \ldots, X_p)^T$ denote a random vector distributed as $N_p(\theta, I)$, where $p \geqslant 4$. Consider estimating θ with

$$\hat{\theta} = \bar{X}1_p + \left(1 - \frac{p-3}{\|X - \bar{X}1_p\|^2}\right)(X - \bar{X}1_p),$$

where $\bar{X} = p^{-1}\sum_{j=1}^{p} X_j$, where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^p$ and where 1_p is a p-vector of ones. Describe very briefly the action of $\hat{\theta}$ on each component X_j.

Write down the distribution of $\bar{X}$ in terms of $\bar{\theta} = p^{-1}\sum_{j=1}^{p}\theta_j$, and show that if $R(\hat{\theta}, \theta) = \mathbb{E}_\theta(\|\hat{\theta} - \theta\|^2)$ denotes the risk function, then

$$R(\hat{\theta}, \theta) = 1 + \mathbb{E}_\theta\left\{\left\|\left(1 - \frac{p-3}{\|X - \bar{X}1_p\|^2}\right)(X - \bar{X}1_p) - (\theta - \bar{\theta}1_p)\right\|^2\right\}.$$

Show further that

$$R(\hat{\theta}, \theta) = p - (p-3)^2\mathbb{E}_\theta\left(\frac{1}{\|X - \bar{X}1_p\|^2}\right).$$

Using the fact that $\|X - \bar{X}1_p\|^2$ has a non-central chi-squared distribution with $p-1$ degrees of freedom and non-centrality parameter $\|\theta - \bar{\theta}1_p\|^2$, describe the set on which the risk function attains its minimum, and find the value of the risk function on this set.

4 Describe in detail the Least Angle Regression (LARS) algorithm for a linear model with n observations and p linearly independent covariates, where $n > p$.

[You may assume, without derivation, that at the kth iteration, the LARS algorithm moves to $\hat{\boldsymbol{\mu}}^k = \hat{\boldsymbol{\mu}}^{k-1} + \hat{\gamma}^k \mathbf{u}^k$, where

$$\hat{\gamma}^k = \min_{j \in (\mathcal{A}^k)^c}{}^{+}\left(\frac{C^k - c_j^k}{\alpha^k - a_j^k}, \frac{C^k + c_j^k}{\alpha^k + a_j^k}\right),$$

but your answer should define all the terms in these formulae.]

Define the LASSO estimator $\hat{\boldsymbol{\beta}}_\lambda^{LASSO}$ with penalty parameter λ. Describe the modification to the LARS algorithm that yields all LASSO solutions $\{\hat{\boldsymbol{\beta}}_\lambda^{LASSO} : \lambda > 0\}$.

[*Hint: Write $\tilde{\gamma}^k$ for the smallest step in the positive γ-direction along the LARS line $\boldsymbol{\mu}(\gamma) = \hat{\boldsymbol{\mu}}^{k-1} + \gamma\mathbf{u}^k$ for which some active index j_k satisfies $\beta_{j_k}(\tilde{\gamma}^k) = 0$, where $\boldsymbol{\beta}(\gamma) = (\beta_1(\gamma), \ldots, \beta_p(\gamma))^T$ satisfies $\boldsymbol{\mu}(\gamma) = X\boldsymbol{\beta}(\gamma)$.*]

END OF PAPER