UNIVERSITY OF CAMBRIDGE

M. PHIL. IN STATISTICAL SCIENCE

Wednesday 7 June 2006 9 to 12

TIME SERIES AND MONTE CARLO INFERENCE

*Attempt **FOUR** questions. There are six questions in total.*
Marks for each question are indicated on the paper in square brackets.
Each question is worth a total of 20 marks.

Note: *The following properties of the Inverse Gamma and Beta distributions may be used without proof:*

If $X \sim \Gamma^{-1}(a,b)$ *then*

$$f_X(x) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} \exp(-b/x), \quad x > 0$$

and $\mathbb{E}(X) = \frac{b}{a+1}$, *with* $Var(X) = b^2/(a-1)^2(a-2)$.

If $X \sim \beta(a,b)$ *then*

$$f_X(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1 \ ,$$

and $\mathbb{E}(X) = \frac{a}{a+b}$, $Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$, *and mode* $(X) = \frac{a-1}{a+b-2}$.

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 Time Series

Suppose $X_1, \ldots, X_T$ are observations of a weakly stationary autoregressive process of order p. Write down a recursive method for obtaining forecasts $\hat{X}_{T,k}$, $k \geqslant 1$, of X_{T+k} for $T \geqslant p$, and write down $\hat{X}_{T,1}$ and $\hat{X}_{T,2}$. [5]

Let $Z_t = \mu + X_t$ where

$$X_t = \phi X_{t-1} + \epsilon_t, \tag{1.1}$$

with $|\phi| < 1$, where $\{\epsilon_t\}$ is a white noise process. Show that $\{Z_t\}$ is weakly stationary and derive its autocorrelation function. [2]

Suppose $Z_1, \ldots, Z_T$ are observed. By considering the recursive forecasts for the X_t's, find forecasts $\hat{Z}_{T,k}$, $k \geqslant 1$, for Z_{T+k}, $T \geqslant 1$, in terms of ϕ, μ and $Z_1, \ldots, Z_T$. What happens to $\hat{Z}_{T,k}$ as $k \to \infty$? [7]

Now let $W_t = \mu t + Y_t$ where $\mathbb{E}Y_t = 0$ and $(I - B)Y_t = X_t$ where B is the backwards shift operator and X_t is as in equation (1.1). Show that $(I - B)W_t = Z_t$. By considering the forecasts $\hat{Z}_{T,k}$, derive forecasts $\hat{W}_{T,k}$ of W_{T+k} in terms of μ, ϕ and observations $W_1, \ldots, W_T$. What happens to $(\hat{W}_{T,k})/k$ as $k \to \infty$? [6]

2 Time Series

For a weakly stationary process $\{X_t\}$, define the autocovariance function and the spectral distribution function. If a spectral density function $f(\lambda)$ exists, express $f(\lambda)$ in terms of the autocovariances γ_k, $k \in \mathbb{Z}$ and express γ_k in terms of $f(\lambda)$. [4]

(a) Find γ_k and $f(\lambda)$ if $X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$, where $\{\epsilon_t\}$ is a white noise process with mean 0 and variance σ^2, and θ_1, θ_2 are constants. Find the spectral distribution function for this process. [6]

(b) Given that a weakly stationary process has spectral density

$$f(\lambda) = \pi - |\lambda|,$$

find the autocovariance function. [5]

(c) Let $X_t = A\cos(\omega_0 t) + B\sin(\omega_0 t)$, where A and B are uncorrelated random variables A and B, each with mean 0 and variance 1, and ω_0 is in $(0, \pi)$. Show that $\{X_t\}$ is weakly stationary and find its spectral distribution function. [5]

3 Monte Carlo Inference

(a) Let $f(x)$ and $g(x)$ be probability density functions with $f(x) \leqslant Mg(x)$ for all x and some $0 < M < \infty$. Write down the algorithm for the rejection method for simulating random variables with probability density f, using observations from g. Prove that this method works. [6]

Show that the probability of acceptance is greatest when $M = \sup_x \left[\frac{f(x)}{g(x)}\right]$. [3]

(b) Consider generating random numbers from a distribution with density

$$f_X(x) = \begin{cases} \frac{1}{6}(x-1) & 1 \leqslant x \leqslant 3 \\ \frac{1}{12}(7-x) & 3 \leqslant x \leqslant 7 \\ 0 & \text{otherwise} \end{cases}$$

given a set of pseudo-random numbers $\{U_i\} \in [0,1]$.

Describe how this can be done via

(i) the rejection method [5]

(ii) the method of inversion. [4]

Which method do you prefer and why? [2]

4 Monte Carlo Inference

(a) Describe the bootstrap method for estimating the standard error of an estimator $\hat{\theta}$ for some parameter $\theta(F)$, on the basis of a random sample $x_1, \dots x_n$ from F. Your description should include the form of the empirical estimator $\hat{F}$ of F used in the algorithm. [5]

Show that, given a sample of n distinct observations, the probability that any subsequent bootstrap sample has at least one repeated value is given by

$$1 - \frac{n!}{n^n}$$ [4]

(b) Suppose we observe paired data, $(X_1, Y_1), \dots, (X_{100}, Y_{100})$. Construct for $r = \mathrm{cor}(X, Y)$

(i) a 95% percentile interval [3]

(ii) a 95% bootstrap$-t$ interval. [3]

Would you use similar methods to construct a confidence interval for the mean of X? Explain your answer. [1]

(c) Describe what is meant by a control variate for Y, an unbiased estimate of θ. Show how the variance of Y can be minimised using control variates. [4]

5 Monte Carlo Inference

(a) Describe the Metropolis-Hastings algorithm for obtaining a dependent sample from some distribution $\pi(\theta)$, $\theta \in \mathbb{R}^p$. [2]

In the context of the Metropolis Hastings algorithm, describe what is meant by the terms

(i) random-walk update [2]

(ii) independence sampler [2]

Discuss the relative advantages and disadvantages of these. [2]

(b) Suppose we observe data $\mathbf{y} = (y_1, \ldots, y_n)^T$ with corresponding known covariates $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ each of length k. We express the model in the form

$$\mathbf{y} = X_k \mathbf{a}_k + \boldsymbol{\epsilon}$$

for design matrix

$$X_k = \begin{pmatrix} 1 & \mathbf{x}_1^T \\ \vdots & \vdots \\ 1 & \mathbf{x}_n^T \end{pmatrix}$$

$$\begin{aligned} \text{where} \quad & \mathbf{a}_k = (a_0, a_1, \ldots, a_k)^T \\ \text{and} \quad & \boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T \\ \text{with} \quad & \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n) \end{aligned}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix.

(i) Given priors $\sigma^2 \sim \Gamma^{-1}(\alpha, \beta)$ and $\mathbf{a}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ give the posterior distribution $\pi(\mathbf{a}_k, \sigma^2 | \mathbf{x}, \mathbf{y})$ up to a constant of proportionality. [3]

(ii) Describe reversible jump updates for a change in model order from k to $k+1$ and from $k+1$ to k (i.e., introducing/deleting a covariate x_{k+1}). [9]

6 Monte Carlo Inference

Consider the following likelihood used in the fitting of capture-recapture models in population ecology.

$$L(\mathbf{n},\mathbf{m}|\boldsymbol{\phi},\mathbf{P},\mathbf{R}) \propto \prod_{j=1}^{J-1}(1-\phi_j)^{\sum_{i=1}^{j} n_{ij}}\phi_j{}^{A_{1j}} \times \prod_{j=2}^{J} P_j^{\sum_{i=1}^{j-1} m_{ij}}(1-P_j)^{A_2 j}$$

$$\text{where} \quad A_{1j} = \sum_{i=1}^{j}\left(\sum_{s=j+1}^{J} m_{is} + \sum_{s=j+1}^{J} n_{is}\right)$$

$$\text{and} \quad A_{2j} = \sum_{i=1}^{j-1}\left(\sum_{s=j+1}^{J} m_{is} + \sum_{s=j}^{J} n_{is}\right)$$

$$\text{and where} \quad \boldsymbol{\phi} = (\phi_1,\ldots,\phi_{J-1}), \quad \mathbf{P} = (P_1,\ldots,P_J)$$

are parameters to be estimated

$$\mathbf{m} = \{m_{ij}\} \quad i=1,\ldots,I \quad , \quad j=2,\ldots,J$$
$$\text{and} \quad \mathbf{R} = \{R_i\} \quad i=1,\ldots,I$$

are data

$$\text{and} \quad \mathbf{n} = \{n_{ij} : i=1,\ldots,I \ , \ j=1,\ldots,J-1\} \cup \{n_i = i=1,\ldots,I\}$$

are introduced 'missing data' such that

$$n_i = R_i - \sum_{j=1}^{J-1} n_{ij} = \sum_{j=1}^{J} m_{ij}.$$

(a) By differentiating the log-likelihood, show that the maximum likelihood estimates of ϕ_i and P_i are given by

$$\hat{\phi}_j = \frac{\sum_{i=1}^{j}\sum_{s=j+1}^{J}(m_{is}+n_{is})}{\sum_{i=1}^{j}(\sum_{s=j+1}^{J} m_{is} + \sum_{s=j}^{J} n_{is})} \ , \ j=1,\ldots,J-1 \qquad [5]$$

and

$$\hat{P}_j = \frac{\sum_{i=1}^{j-1} m_{ij}}{\sum_{i=1}^{j-1}\sum_{s=j}^{J}(m_{is}+n_{is})} \ , \ j=2,\ldots,J \qquad [4]$$

(b) Writing $n_{i,j} = n_{ij}$ for clarity, use the fact that

$$(n_{i,i} \ , \ n_{i,i+1},\ldots,n_{i,J-1} \ , \ n_i|\mathbf{m},\boldsymbol{\phi},f)$$

is multinomial with number of trials $R_i - \sum_{j=i+1}^{J} m_{ij}$ and cell probabilities

$$\left(\frac{1-\phi_i}{C_i}\ ,\ \frac{1-\phi_{i+1}}{C_i}\phi_i(1-P_{i+1})\ ,\ \ldots\ ,\ \frac{1-\phi_{J-1}}{C_i}\prod_{j=i}^{J-2}\phi_j(1-P_{j+1})\ ,\ \frac{1}{C_i}\prod_{j=i}^{J-1}\phi_j(1-P_{j+1}) \right)$$

where C_i is a normalising constant, to derive an EM algorithm to estimate ϕ_i and P_i. Take care to explain in detail both the E and M steps. [7]

(c) Show that the maximum likelihood estimates of P_i can be derived without using differentiation, via standard distribution results. [4]

END OF PAPER