M.PHIL. IN STATISTICAL SCIENCE

Wednesday 4 June 2008 9.00 to 12.00

TIME SERIES AND MONTE CARLO INFERENCE

Attempt no more than **FOUR** *questions.*

There are **SIX** *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 Time Series

Explain what is meant by a *weakly stationary process* $\{X_t\}$. Define the *autocovariance function* and the *autocorrelation function* of $\{X_t\}$.

Let

$$X_t \;=\; \alpha\,(X_{t-1} - X_{t-2}) \;+\; \epsilon_t\,, \tag{1}$$

where α is a real constant and $\{\epsilon_t\}$ is a white noise process with mean zero and variance σ^2. Determine the range of possible values of α for which (1) has a unique weakly stationary solution.

For $\alpha = -1/12$, find the Wold representation of $\{X_t\}$ and determine the autocovariance function of $\{X_t\}$.

[*Results from lectures may be quoted and used without proof.*]

2 Time Series

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a weakly stationary process with autocovariance function γ_k and spectral density function $f_X(\lambda)$. Write down an expression for γ_k in terms of $f_X(\lambda)$.

The process $\{Y_t\}$ is obtained from $\{X_t\}$ by applying the filter $\{a_r\}_{r\in\mathbb{Z}}$, with $a_r \in \mathbb{R}$ for all $r \in \mathbb{Z}$ and $\sum_{r\in\mathbb{Z}} |a_r| < \infty$, so that $Y_t = \sum_{r\in\mathbb{Z}} a_r\, X_{t-r}$. Show that $\{Y_t\}$ is weakly stationary and find its spectral density function $f_Y(\lambda)$ in terms of $f_X(\lambda)$ and $a(\lambda) = \sum_{r\in\mathbb{Z}} a_r\, e^{ir\lambda}$.

Let $\{Z_t\}$ be obtained from $\{Y_t\}$ by applying the filter $\{b_r\}$, with $b_r \in \mathbb{R}$ for all $r \in \mathbb{Z}$ and $\sum_{r\in\mathbb{Z}} |b_r| < \infty$. Write down the spectral density function $f_Z(\lambda)$ of $\{Z_t\}$. Show that $\{Z_t\}$ can be obtained from $\{X_t\}$ by applying a linear filter $\{c_r\}$, and find c_r in terms of the a_k's and the b_k's.

Let the *gain* of a filter $\{a_r\}$ be $G_a(\lambda) = |\,a(\lambda)|\,,\ \lambda \in [0, \pi]$.

(a) Suppose that $Y_t = X_t - X_{t-1}$. Find $f_Y(\lambda)$. Sketch the gain of the filter taking $\{X_t\}$ to $\{Y_t\}$ and comment.

(b) Suppose that $Z_t = Y_t - Y_{t-12}$. Find $f_Z(\lambda)$. Sketch the gain of the filter taking $\{Y_t\}$ to $\{Z_t\}$ and comment.

(c) Find the filter that takes $\{X_t\}$ onto $\{Z_t\}$ and find its gain.

3 Monte Carlo Inference

(a) Monte Carlo methods depend crucially on the ability to generate pseudo random numbers in the interval $(0, 1)$. Make a short list of what you consider to be the most important properties of a good pseudo random number generator.

(b) Suppose you had an ideal pseudo random number generator giving you the ability to generate arbitrarily many independent uniform variates $U_1, U_2, \ldots$, i.e., $U_i \sim U(0, 1)$.

(i) Describe how the method of inversion can be used to obtain draws from the $\mathrm{Exp}(\lambda)$ distribution. How can this method be extended to obtain draws from a double–exponential (Laplace) distribution? What property of the pseudo random number generator is crucial to ensure that your algorithm gives samples from the correct distribution?

(ii) Give two distinct algorithms for obtaining draws from a χ^2_ν distribution where $\nu \in \{2, 3, \ldots\}$. Say which algorithm you prefer, and why. Your answer may depend on ν.

(iii) Consider obtaining draws from a $\mathrm{Beta}(\alpha, \beta)$ distribution, when $\alpha, \beta \in \{1, 2, \ldots\}$. Give one algorithm which uses one of the methods from (ii), and one new method.

4 Monte Carlo Inference

(a) Describe the jackknife *and* nonparametric bootstrap methods for estimating the variance of an estimator $\hat{\theta}$ of some parameter $\theta(F)$, on the basis of a random sample $x_1, \ldots, x_n$ of distinct observations from F. Your description of the nonparametric bootstrap should include the form of the empirical distribution function $\hat{F}_n$ used in the algorithm.

(b) Find the probability that a bootstrap sample contains at least one repeated value.

(c) Consider the following R code where `x` is a vector of length `n` containing the random sample $x_1, \ldots, x_n$, where `x` and `n` have been set earlier in the code.

```
R1a> mat <- matrix(NA, nrow=n, ncol=n-1)
R2a> for(i in 1:n) mat[i,] <- x[-i]
R3a> vect <- apply(mat, 1, mean)
R4a> (n-1)*mean((vect - mean(vect))^2)
R5a> (n-1)*(mean(vect) - mean(x))
```

Explain what is being calculated in lines `R4a` and `R5a`. Give the numerical value of the expression in line `R5a`, and justify your answer.

Now consider another piece of R code below (with the same `x` as above).

```
R1b> alpha <- 0.05
R2b> B <- 199
R3b> mat <- matrix(NA, nrow=B, ncol=n)
R4b> for(b in 1:B) mat <- sample(x, n, replace=TRUE)
R5b> vect <- apply(mat, 1, mean)
R6b> s <- sort(vect)
R7b> c(s[(B+1)*alpha/2], s[(B+1)*(1-alpha/2)])
```

Explain what is being calculated in the code, with particular attention paid to the value of the expression in line `R7b`.

(d) Suppose that we had another random sample $y_1, \ldots, y_m$ from a distribution $G \neq F$, where $\theta = \mathbb{E}_F\{X\} = \mathbb{E}_G\{Y\}$ and $\mathrm{Cov}(X, Y) < 0$. Give an algorithm for constructing an efficient, unbiased estimator $\tilde{\theta}$ of θ that uses the combined sample $x_1, \ldots, x_n, y_1, \ldots, y_m$. How could you estimate $\mathrm{Var}(\tilde{\theta})$?

5 Monte Carlo Inference

(a) (i) Describe the Gibbs Sampler for obtaining a dependent sample from some distribution $\pi(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \mathbb{R}^p$.

(ii) Suppose that we observe data $\mathbf{y} = (y_1, \ldots, y_n)^\top$, with corresponding known (scalar) covariates $\mathbf{x} = (x_1, \ldots, x_n)^\top$ and that we want to fit a polynomial regression model of order k to the data. Then we can express the model in the form

$$\mathbf{y} = \mathbf{X}_k \boldsymbol{\beta}_k + \boldsymbol{\varepsilon}$$

for design matrix

$$\mathbf{X}_k = \begin{pmatrix} 1 & x_1 & \cdots & x_1^k \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \cdots & x_n^k \end{pmatrix}$$

where $\boldsymbol{\beta}_k = (\beta_0, \beta_1, \ldots, \beta_k)^\top$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$, with $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$, where $\mathbf{I}$ is the $n \times n$ identity matrix. For independent priors $\sigma^2 \sim \Gamma^{-1}(a, b)$ and $\boldsymbol{\beta}_k \sim N_{k+1}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ the posterior distribution is given by

$$\pi(\boldsymbol{\beta}_k, \sigma^2 | \mathbf{x}, \mathbf{y}) \propto (\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}_k \boldsymbol{\beta}_k)^\top (\mathbf{y} - \mathbf{X}_k \boldsymbol{\beta}_k) \right\} \\ \times (\sigma^2)^{-(a+1)} \exp\left\{ -\frac{b}{\sigma^2} \right\} \times \exp\left\{ -\frac{1}{2} (\boldsymbol{\beta}_k - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{\beta}_k - \boldsymbol{\mu}_k) \right\} .$$

Show that the conditional distributions $\pi(\boldsymbol{\beta}_k | \sigma^2, \mathbf{x}, \mathbf{y})$ and $\pi(\sigma^2 | \boldsymbol{\beta}_k, \mathbf{x}, \mathbf{y})$ are multivariate normal and inverse gamma, respectively, and calculate the parameters of each distribution.

(iii) Hence describe how we can use the Gibbs Sampler to obtain a dependent sample from the joint posterior distribution of $\pi(\boldsymbol{\beta}, \sigma^2 | \mathbf{x}, \mathbf{y})$.

(b) Now suppose that the order of the polynomial is unknown, and that we wish to use a reversible jump procedure to update the order of the polynomial model. We propose to move from the model of order k, with parameters $\boldsymbol{\beta}_k$, to the model of order $k+1$ with parameters $\boldsymbol{\beta}'_{k+1}$ (keeping σ^2 fixed) using the following procedure,

$$\begin{aligned} \beta'_i &= \beta_i && \text{for} \quad i = 1, \ldots, k \\ \beta'_{k+1} &= z && \text{for} \quad z \sim N(0, \sigma^2_\beta) \quad \text{and } \sigma^2_\beta \text{ known} \\ \beta'_0 &= \beta_0 - \frac{z}{n} \sum_{i=1}^n x_i^{k+1}. \end{aligned}$$

(i) Calculate an explicit expression for the corresponding acceptance probability for this move.

(ii) Define the reverse move, for moving from the model of order $k+1$ to the model of order k .

(iii) What is the corresponding acceptance probability for this reverse move, from the model of order $k+1$, to the model of order k?

6 Monte Carlo Inference

(a) Let $\mathbf{x}$ represent observed data, and $\mathbf{z}$ denote missing data, with joint distribution $f(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta})$. Briefly describe the iterative Expectation Maximisation (EM) algorithm for finding the $\hat{\boldsymbol{\theta}}$ that maximises the observed data likelihood $L(\mathbf{x}|\boldsymbol{\theta})$.

(b) Suppose that $\mathbf{y} = (y_1, y_2, y_3, y_4)$ is a data vector of observed counts from a multinomial distribution with parameters n and $\mathbf{p}$, where the cell probabilities

$$\mathbf{p} = (p_1,\, p_2,\, p_3,\, p_4) \;=\; \left(\frac{1}{2} - \frac{\theta}{2}, \frac{\theta}{4}, \frac{\theta}{4}, \frac{1}{2}\right),$$

are parameterised by $\theta \in [0, 1]$.

(i) Find the maximum likelihood estimator $\hat{\theta}$ based on the complete data likelihood $L(\theta|\mathbf{y})$.

(ii) Now suppose instead that only three counts

$$\mathbf{x} = (x_1, x_2, x_3)$$

were observed, where

$$\mathbf{x} = (y_1, y_2, y_3 + y_4).$$

That is, $y_3 = x_3 - z$, $y_4 = z$ and z is missing. Consider using the EM algorithm for estimating $\hat{\theta}$ based on the observed data log likelihood $L(\mathbf{x}|\theta)$. Derive the "E–step" of the EM algorithm and write the resulting expression(s) in terms of $\log L(\mathbf{y}^{(t)}|\theta)$, for

$$\mathbf{y}^{(t)} = (y_1, y_2, y_3^{(t)}, y_4^{(t)})$$

where $y_3^{(t)} = x_3 - z^{(t)}$, $y_4^{(t)} = z^{(t)}$, and $z^{(t)} = \mathbb{E}\{z|\mathbf{x}, \theta^{(t)}\}$ which you should calculate. In other words, show that the "E–step" is the same as "filling in the missing values" in this case.

(iii) Combine the "E–step" in part (ii) with an "M–step" derived from the appropriate application of your result from part (i). That is, give a complete description of your EM algorithm in this case for iteratively finding $\tilde{\theta}$, the maximum likelihood estimator of the observed data likelihood.

(iv) Suppose that $\mathbf{x} = (38, 34, 125)$ and $\theta^{(t)} = 0.5$. What are the values of $y_3^{(t)}$ and $\theta^{(t+1)}$?

END OF PAPER