UNIVERSITY OF CAMBRIDGE

M. PHIL. IN STATISTICAL SCIENCE

Thursday 9 June, 2005 9:00 to 12:00

TIME SERIES AND MONTE CARLO INFERENCE

*Attempt **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.

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

If $X \sim \Gamma(a, b)$ then

$$f_X(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \qquad x > 0$$

and $\mathbb{E}(x) = \dfrac{a}{b}$, with $\mathrm{var(x)} = \dfrac{\mathrm{a}}{\mathrm{b}^2}$, for a, $b > 0$.

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

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

and

$$\mathbb{E}(x) = \frac{b}{a-1}, \quad \text{with} \quad \mathrm{var\ (x)} = \frac{\mathrm{b}^2}{(\mathrm{a}-1)^2(\mathrm{a}-2)} \quad \text{for} \quad \mathrm{a} > 2 \quad \text{and} \quad \mathrm{b} > 0.$$

1 Time Series

Explain what is meant by a weakly stationary process $\{X_t\}$ with autocovariance function γ_k and spectral density $f_X(\lambda)$. Write down an expression for $f_X(\lambda)$ in terms of the γ_k's

Define a white noise process, $WN(0, \sigma^2)$, and find its spectral density function.

Let $Y_t = \sum_{r \in \mathbb{Z}} c_r X_{t-r}$ with $\sum_{r \in \mathbb{Z}} |c_r| < \infty$. Show that $\{Y_t\}$ is weakly stationary, and find its spectral density function in terms of $f_X(\lambda)$ and $C(z) = \sum_{r \in \mathbb{Z}} c_r\, z^r$.

Hence, or otherwise, find the spectral density function of $\{X_t\}$ where

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}\,,$$

and $\{\epsilon_t\}$ is a white noise process with mean 0 and variance σ^2.

For the process defined by

$$X_t = \epsilon_t + \theta \epsilon_{t-1}.$$

with $|\theta| < 1$, find $c_0, c_1, c_2, \ldots$ such that $\sum_{r=0}^{\infty} c_r X_{t-r}$ is a white noise process.

2 Time Series

Let

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t \qquad (*)$$

where $\phi_1, \ldots, \phi_p$ are constants and $\{\epsilon_t\}$ is a white noise process with mean 0 and variance σ^2. State a condition that ensures that there is a weakly stationary solution to $(*)$.

Suppose that $p = 2$, $\phi_1 = \alpha$ and $\phi_2 = \alpha^2$. Show that the stationarity condition is satisfied if and only if $|\alpha| < \alpha_0$ for some α_0, and find α_0.

For $|\alpha| < \alpha_0$, find the Wold representation of $\{X_t\}$.

Assume $|\alpha| < \alpha_0$ and α is known. Write down the linear minimum mean square error forecasts for X_{T+1} and X_{T+2} based on $X_T, X_{T-1}, \ldots$

3 Monte Carlo Inference

(a) Define the congruential generator and explain how it can be used to generate a sequence of pseudo-random numbers $\{U_i\} \in [0, 1]$.

State the conditions necessary for a non-multiplicative congruential generator to have maximal possible cycle length when the shift $c > 0$.

In a congruential generator with modulus, $M = 12^k$, for some positive integer $k > 1$, describe a condition on the multiplier, a, such that the cycle length is equal to M.

(b) (i) Show how you could sample from a density $f(x)$, corresponding to a $N(0, \sigma^2)$ distribution, via the ratio of uniforms method using $h(x) = \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right)$ and

$$f(x) = \frac{h(x)}{\int\limits_{-\infty}^{\infty} h(x)dx} \qquad -\infty \leqslant x < \infty$$

(ii) Show how you could sample from a Weibull distribution with density,

$$g(x) = \frac{\alpha}{\beta} x^{\alpha-1}\, e^{-x^{\alpha}/\beta} \qquad 0 \leqslant x < \infty$$

using the method of inversion.

(iii) Therefore show how you would sample from a Weibull and half normal distribution mixture

$$k(x) = 2pf(x) + (1-p)g(x) \qquad 0 \leqslant x < \infty$$

4 Monte Carlo Inference

(a) Explain how the method of importance sampling may be used to estimate $\mu = \mathbb{E}_f[\theta(x)]$ from a sample $x_1, \ldots, x_n \sim g(x)$, where $f(x)$ and $g(x)$ are normalised densities with common support, and $\theta(x)$ denotes any general scalar function of x.

Now suppose

$$f(x) = \frac{4}{\pi}\sqrt{1-x^2}, \qquad x \in (0,1)\,.$$

Show how we can use a function

$$g(x) \propto x^2 \qquad x \in (0,1)$$

to estimate $\mu_g = \mathbb{E}_f(x)$ via importance sampling.

Show that the variance of an importance sampling estimate is minimised by sampling from

$$g = g_0 = \frac{|\theta(x)f(x)|}{\int |\theta(x)f(x)|\,dx}.$$

Hence, show that the variance of μ_g above is minimised by using

$$g = g_0 = 3x\sqrt{1-x^2}.$$

(b) Consider the model

$$y_i = \mathbf{x}_i^T\boldsymbol{\beta} + e_i$$
$$\mathbb{E}[e_i] = 0$$

where $\mathbf{x}_i$ is a (2×1) vector of observations, β_1 and β_2 are the first and second elements of $\boldsymbol{\beta}$ respectively, and $\boldsymbol{\beta}$ is estimated by $\hat{\boldsymbol{\beta}} = (X^TX)^{-1}(X^TY)$, where X is the matrix of rows x_i^T, and Y is the vector with components y_i.

Suggest an estimate for the statistic $\theta = \beta_1\beta_2$.

Show how you could construct a 95% confidence interval for θ using the following two methods:

(i) Percentile bootstrap

(ii) Equal-tailed bootstrap-t.

5 Monte Carlo Inference

(a) Describe under what circumstances, and for what purpose, you might consider using the Metropolis Hastings algorithm.

Describe the random-walk Metropolis Hastings algorithm.

Briefly explain the following MCMC implementation issues and how they may be dealt with in practice.

(i) convergence;

(ii) single component versus block updates;

(iii) the tuning of proposal variances to achieve good mixing qualities.

Figure 1 (see over the page) gives some initial trace plots for each parameter from a 3-parameter posterior, $\pi(\theta_1, \theta_2, \theta_3|X)$, obtained using a single-component random-walk Metropolis Hastings MCMC algorithm. Assuming this to be one step in an algorithm tuning exercise, what changes would you make to the algorithm before rerunning?

(b) The following model can be used to study the spread of infectious diseases in trees:

$$P_i = 1 - \exp\left\{-\alpha \sum_{j \in I_t} K(d_{ij})\right\}$$

where P_i	is the probability of a susceptible tree being infected
α	is a scaling parameter
d_{ij}	is the distance between tree i and tree j
$K(d_{ij})$	is a distance kernel which characterises the chance of infection occurring over a given distance d_{ij}
I_t	is the set of all trees infected at time t

The likelihood is given by

$$f(\mathbf{X}|\boldsymbol{\theta}) = \prod_{i \in I_{t+1}} P_i \prod_{i \notin I_{t+1}} 1 - P_i$$

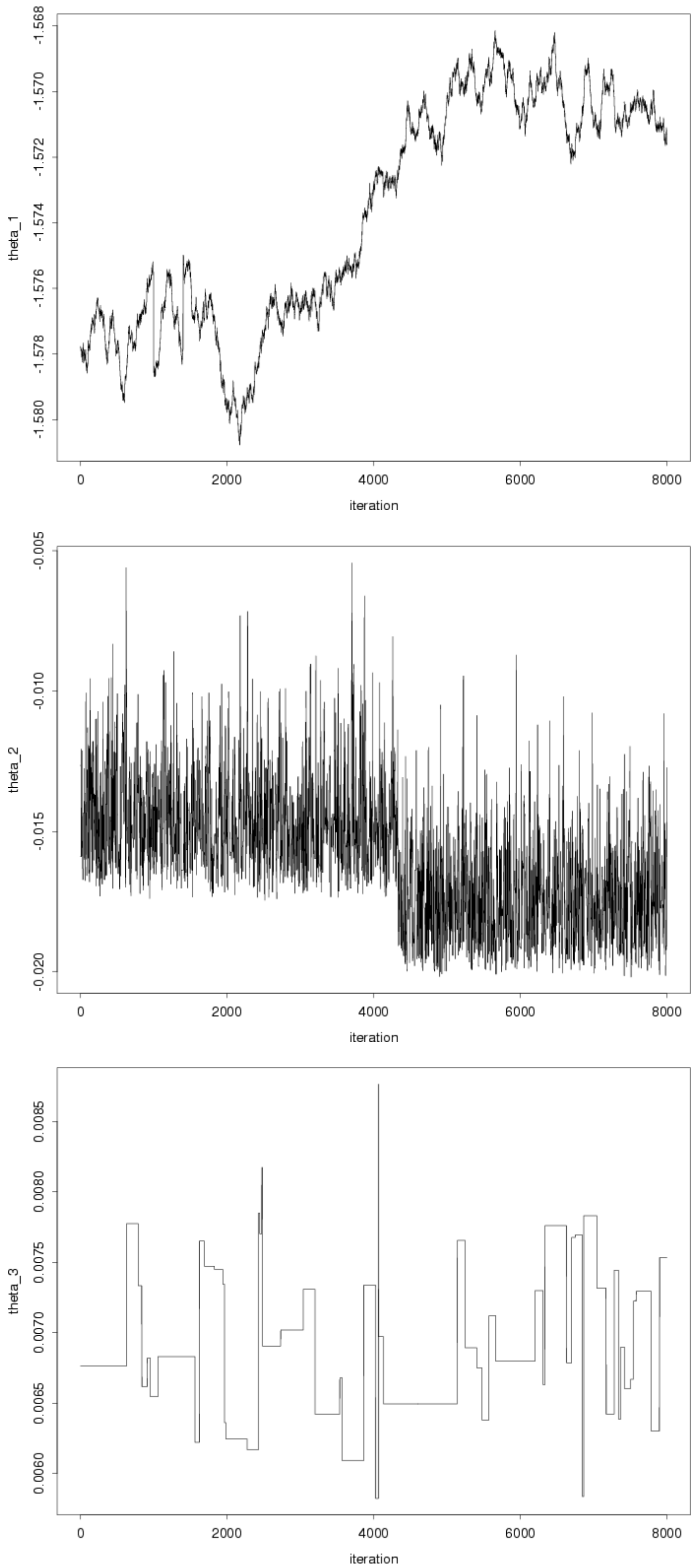
where $\mathbf{X}$ represents the data and $\boldsymbol{\theta}$ is a vector of parameters to be estimated.

Let $K(d_{ij}) = K_1(d_{ij}) = d_{ij}^{-\beta}$, and place priors on α and β such that, $p(\alpha) \sim N(\mu_\alpha, \sigma_\alpha^2)$ and $p(\beta) \sim N(\mu_\beta, \sigma_\beta^2)$, respectively.

Explain why a Gibbs sampler would not be appropriate for estimating the marginal posterior distributions of α and β.

Explain how we could use MCMC methods to compare the model with $K_1(d_{ij})$ above with that with a second kernel, $K_2(d_{ij}) = \exp\{-\gamma d_{ij}\}$, where the prior for γ is such that $p(\gamma) \sim N(\mu_\gamma, \sigma_\gamma^2)$.

Figure 1. Sample Path Trace Plots for parameters θ_1, θ_2 and θ_3



6 Monte Carlo Inference

Describe the annealing algorithm for minimising some function $f(\theta)$ with respect to θ.

Suppose that we observe data $x_1, \ldots, x_m$ and that we wish to decide whether a Poisson, $P_0(\lambda)$, or a normal, $N(\mu, \sigma^2)$, provides the best model for these data. Calculate the maximum likelihood estimates for λ, μ and σ^2.

Derive an annealing algorithm to fit the normal model using Gibbs updates. Hence, show that the annealing algorithm converges to the maximum likelihood estimate in this case.

Now calculate the Boltzmann distribution with $f(\lambda)$ equal to the log-likelihood under the Poisson model and show that this converges to a point mass at the maximum likelihood estimate as the temperature decreases.

Finally, calculate the form of the AIC statistic for each model. Hence, describe how your annealing algorithm can be extended to distinguish between the two models. Make clear (and fully describe) any proposal distributions, Jacobian terms and acceptance ratios that you need for your trans-dimensional simulated annealing algorithm.

END OF PAPER