UNIVERSITY OF CAMBRIDGE

M. PHIL. IN STATISTICAL SCIENCE

Friday 31 May 2002 9 to 12

TIME SERIES AND MONTE CARLO INFERENCE

*Attempt **FOUR** questions*

*There are **six** questions in total*

The questions carry equal weight

Note: *The following properties of the Inverse Gamma distribution may be used without proof. 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 $\mathbb{E}(x) = \dfrac{b}{a-1}$, *with* $Var(x) = \dfrac{b^2}{(a-1)^2(a-2)}$ *for* $a > 2$.

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

1 Monte Carlo Inference

(a) (i) Let $h : \mathbb{R} \to \mathbb{R}$ denote a non-negative function with finite integral over $\mathbb{R}$. If (U, V) is uniformly distributed in the region $C_h = \{(U,V) : 0 \leqslant U \leqslant \left(h\left(\frac{V}{U}\right)\right)^{1/2}\}$, calculate the joint density of $X_1 = V/U$ and $X_2 = U$.

(ii) Hence show that X_1 has density function

$$f(x) = \frac{h(x)}{\int_{-\infty}^{\infty} h(y)dy} \qquad -\infty < x < \infty.$$

(iii) Consider the Laplace distribution with probability density function

$$f(x) = \tfrac{1}{4}\exp\left(-\frac{|x-1|}{2}\right) \qquad -\infty < x < \infty.$$

For $h(x) = \exp\left(-\frac{|x-1|}{2}\right)$ and

$$a = \sqrt{\sup_x(h(x))}, \quad b_1 = -\sqrt{\sup_{x \leqslant 0}(x^2 h(x))}, \quad b_2 = \sqrt{\sup_{x \geqslant 0}(x^2 h(x))},$$

show that a, b_1, and b_2 are finite.

(iv) Hence describe how we may sample from the Laplace distribution using the ratio of uniforms methods given above.

(b) (i) For density function f, show that if $F(x) = \int_{-\infty}^{x} f(u)\,du$ and $U \sim U[0,1]$, then $X = F^{-1}(U) \sim f$.

(ii) Describe how we can sample from the Laplace distribution given in (a) using the method of inversion.

2 Monte Carlo Inference

(a) (i) 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.

(ii) How would your description in (i) change if the normalisation constant for f and/or g were unknown?

(iii) Show that the variance of the importance sampling estimator $\hat{\mu}_g$ is given by

$$\mathrm{Var}(\hat{\mu}_g) = \frac{1}{n} \int \frac{f^2(x)\theta^2(x)}{g(x)} dx - \frac{\mu^2}{n}.$$

(iv) Hence show that sampling from

$$g_0(x) \ \propto \ |\theta(x) f(x)|$$

minimises $\mathrm{Var}(\hat{\mu}_g)$ over all densities g.

[*Recall the Cauchy-Schwarz inequality*

$$\left(\int f(x) g(x) dx \right)^2 \ \leqslant \ \int f^2(x) dx \int g^2(x)\, dx.]$$

(b) (i) Suppose that $f(x) = \dfrac{1}{\pi(1+x^2)}$ and $\mu = \mathbb{P}_f(X \geqslant k)$. What function $\theta(x)$ could be used to estimate μ via importance sampling?

(ii) Take $g(x) \ \propto \ \frac{1}{x^2}$, $k \leqslant x < \infty$. What is the normalisation constant for g?

(iii) How would you obtain samples from g via inversion?

(iv) Hence show that if $u_i \sim U[0,1]$ $i = 1, \ldots, n$, then

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \frac{k}{\pi(u_i^2 + k^2)}$$

is an importance sampling estimator for μ.

3 Monte Carlo Inference

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

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

for design matrix

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

and where $\mathbf{a}_k = (a_0, a_1, \ldots, a_k)^T$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$, with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$, where I is the $n \times n$ identity matrix. For priors $\sigma^2 \sim \Gamma^{-1}(\alpha, \beta)$ and $\mathbf{a}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \Sigma_k)$ the posterior distribution is given by

$$\pi(\mathbf{a}_k, \sigma^2 | \mathbf{x}, \mathbf{y}) \propto \frac{1}{(\sigma^2)^{n/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - X_k\mathbf{a}_k)^T(\sigma^2 I)^{-1}(\mathbf{y} - X_k\mathbf{a}_k)\right)$$

$$\times (\sigma^2)^{-(\alpha+1)} \exp\left(-\frac{\beta}{\sigma^2}\right) \times \frac{1}{(2\pi)^{1/2}|\Sigma_k|^{k/2}} \exp\left(-\frac{1}{2}(\mathbf{a}_k - \boldsymbol{\mu}_k)^T \Sigma_k^{-1}(\mathbf{a}_k - \boldsymbol{\mu}_k)\right).$$

Calculate the conditional distributions $\pi(\mathbf{a}_k | \sigma^2, \mathbf{x}, \mathbf{y})$ and $\pi(\sigma^2 | \mathbf{a}_k, \mathbf{x}, \mathbf{y})$.

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

(b) Now suppose that the order of the polynomial model is unknown, and that we 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 $\mathbf{a}_k$, to the model of order $k+1$ with parameters $\mathbf{a}'_{k+1}$ (keeping σ^2 fixed) using the following procedure,

$$\begin{aligned} a'_i &= a_i \quad \text{for} \quad i = 1, \ldots, k \\ a'_{k+1} &= z \quad \text{for} \quad z \sim N(0, \sigma_a^2) \\ a'_0 &= a_0 - z\left(\frac{1}{n}\sum_{i=1}^{n} x_i^{k+1}\right). \end{aligned}$$

(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?

4 Monte Carlo Inference

(a) (i) Suppose $x_1, \ldots, x_n \overset{iid}{\sim} N(\mu, \sigma^2)$. State the likelihood function $L(\mathbf{x}|\mu, \sigma^2)$ for these data and derive the maximum likelihood estimates for μ and σ^2.

(ii) Describe the simulated annealing algorithm for maximising some function $f(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

(iii) Show how the simulated annealing algorithm can be used to maximise $L(\mathbf{x}|\mu, \sigma^2)$ above, with respect to $\boldsymbol{\theta} = (\mu, \sigma^2)$, using Gibbs Sampler updates.

(iv) Hence show that the simulated annealing algorithm converges to the maximum likelihood estimate as the system "freezes".

(b) (i) Describe the *EM* algorithm for maximising a likelihood in the presence of "missing data".

(ii) Suppose $x_1, \ldots, x_n$ are drawn independently from a mixture distribution comprising two normals so that

$$f(\mathbf{x}|\boldsymbol{\theta}) = \alpha f_1(\mathbf{x}) + (1 - \alpha) f_2(\mathbf{x})$$

where $f_j(\mathbf{x})$ denotes the density for a normal distribution with mean μ_j and common variance σ^2, and $\boldsymbol{\theta} = (\alpha, \mu_1, \mu_2, \sigma^2)$.

Suppose that we now introduce auxiliary variables Z_{ij} such that

$$Z_{ij} = \begin{cases} 1 & \text{if } x_i \sim N(\mu_j, \sigma^2) \\ 0 & \text{else.} \end{cases}$$

Show that the likelihood function can be written as

$$L(\mathbf{x}|\boldsymbol{\theta}, Z) = \prod_{i=1}^{n} \prod_{j=1}^{2} \left(\alpha_j f_j(x_i)\right)^{Z_{ij}} .$$

(iii) If we let $w_{ij}(\boldsymbol{\theta}) = \mathbb{E}(Z_{ij}|\boldsymbol{\theta}, \mathbf{x})$ and $n_j^{(m)} = \sum_{i=1}^{n} w_{ij}(\boldsymbol{\theta}^{(m)})$, show that an *EM* algorithm to find the maximum likelihood estimator for $\boldsymbol{\theta}$ uses the following transitions

$$\alpha^{(m+1)} = n_1^{(m)}/n$$

$$\mu_j^{(m+1)} = \frac{1}{n_j^{(m)}} \sum_{i=1}^{n} w_{ij}(\boldsymbol{\theta}^m) x_i.$$

What is the corresponding transition for $\sigma^{2(m+1)}$?

5 Time Series

Let X be a second order stationary process. Define its autocorrelations, correlogram and sample partial autocorrelations and explain how they can be used to diagnose AR and MA processes.

Suppose that $X_t = T_t + \varepsilon_t$, where ε_t is the Gaussian white noise of variance $\sigma^2 > 0$ and $T_t = a + bt + ct^2$ is a quadratic trend. Find the symmetric moving average on five points that provides an unbiased estimator of the trend with smallest quadratic error. Show that there is no such estimator on three points.

Explain how a centred moving average can be used to estimate seasonal components. Discuss a danger related to successive application of moving averages.

What is meant by exponential smoothing? Define the methods of simple and double exponential smoothing, and explain when they can be used to forecast time series.

6 Time Series

State carefully and prove the filter theorem. In your answer you should define the following terms: linear filter, filter generating function, filter transfer function.

Let $X = \{X_t : t \in \mathbb{Z}\}$ be an ARMA(p, q) process,

$$\varphi(B)X = \theta(B)\varepsilon,$$

where $\varphi(z) = \sum_{r=0}^{p} \varphi_r z^r$, $\theta(z) = \sum_{s=0}^{q} \theta_s z^s$, and $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$ is the Gaussian white noise, $\mathsf{E}\varepsilon_t^2 = \sigma^2 > 0$. Compute the spectral density of X. Explain what is meant by identifiability of a stationary process and specify when the ARMA process X above is stationary and identifiable. Show that in the latter case the process X is also invertible.

For a stationary ARMA(p, q) process X, show that the coefficients c_k, c_{k-1}, $\ldots$ of the Wold representation $X = C(B)\varepsilon$ satisfy a certain difference equation for all $k > \max(p - 1, q)$ and describe their limiting behaviour as $k \to \infty$. Relate this property to the limiting behaviour of the autocovariances of X.

Compute the autocovariance function for the stationary process X satisfying the equation

$$X_t = \frac{1}{3}X_{t-1} + \frac{2}{9}X_{t-2} + \varepsilon_t - \frac{2}{3}\varepsilon_{t-1},$$

ε being the Gaussian white noise of variance $\sigma^2 = \mathsf{E}\varepsilon_t^2 > 0$, and determine to which class of stationary processes X belongs.