UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Thursday 29 May, 2003 1:30 to 4:30

PAPER 43

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 and Beta distributions may be used without proof. If* $X \sim \Gamma^{-1}(a, b)$, *then*

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

and $\mathbb{E}(x) = \dfrac{b}{a-1}$, *with* $\mathrm{Var}(x) = \dfrac{b^2}{(a-1)^2(a-2)}$ *for* $a > 2$.

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

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

and

$$\mathbb{E}(x) = \frac{a}{a+b}, \text{ with } \mathrm{Var}(x) = \frac{ab}{(a+b)^2(a+b+1)}$$

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) Let f and g be probability density functions and set

$$M = \sup_{x \in \mathbb{R}} \left(\frac{f(x)}{g(x)} \right) \tag{1}$$

(i) Describe the rejection sampling algorithm for generating observations from f using a set of observations from g.

(ii) Prove that this rejection sampling algorithm obtains observations from f.

(iii) Calculate the probability that an observation drawn from g is rejected. Hence deduce that this rejection sampling algorithm is optimised for M given above in equation (1).

(b) Suppose we wish to sample from the distribution with density function,

$$f(x) \propto x^{\alpha-1} \exp(-x\beta)$$

defined over the region (a, ∞) where $0 < a < \alpha/\beta$ and $\alpha > 1$. Then, let the sampling distribution have density

$$g(x) \propto x^{-(b+1)},$$

for $x \in (a, \infty)$.

(i) Show that the optimal value of M is

$$\left(\frac{\alpha + b}{\beta} \right)^{\alpha+b} \exp(-(\alpha + b)) \left(ba^b \int_a^\infty x^{\alpha-1} \exp(-x\beta)\, dx \right)^{-1}.$$

(ii) Using the method of inversion, describe how we could sample from g.

(iii) Suppose that we generate 100 observations from g. Derive an expression for the probability that at least half of these observations are accepted to be from f.

2 Monte Carlo Inference

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

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

$$1 - \prod_{j=0}^{n-1} \left(1 - \frac{j}{n}\right) .$$

(ii) Let X_i, $i = 1, \ldots, n$ denote a set of n independent random variables from a uniform $U[0, \theta]$ distribution, and consider the functional

$$G(\mathbf{X}, \theta) \;=\; \frac{n\Big(\theta - X_{(n)}\Big)}{\theta} ,$$

where $X_{(n)} = \max(X_1, \ldots, X_n)$.

By using the fact that $\left(1 - \dfrac{x}{n}\right)^n \to e^{-x}$ as $n \to \infty$ or otherwise, show that for large n, $G(\mathbf{X}, \theta)$ has an exponential distribution.

Describe the plug-in principle and use it to argue that

$$G(\mathbf{Y}, \hat{\theta}) = \frac{n(X_{(n)} - Y_{(n)})}{X_{(n)}}$$

is a bootstrap realisation of $G(\mathbf{X}, \theta)$, given a bootstrap sample $\mathbf{Y}(Y_1, \ldots, Y_n)$, selected randomly with replacement from $\mathbf{X}$.

Finally, show that if $\hat{\theta}$ denotes the MLE for θ, then, as $n \to \infty$,

$$\mathbb{P}(G(\mathbf{Y}, \hat{\theta}) = 0) \to 1 - \exp(-1).$$

Explain what this last result tells you about the suitability of the bootstrap method for this particular functional.

3 Monte Carlo Inference

Consider the mixture distribution density,

$$f(x) = \sum_{i=1}^{k} \omega_i f_i(x),$$

where $f_i \sim N(\mu_i, \sigma_i^2)$, ω_i is the corresponding weight for density f_i, $i = 1, \ldots, k$ $(k \geqslant 2)$ and $\sum_{i=1}^{k} \omega_i = 1$. For each $i = 1, \ldots, k$ the priors on the parameters are independent of each other, such that

$$\mu_i \sim N(0, \sigma^2) \ ; \ {\sigma_i}^2 \sim \Gamma^{-1}(\alpha, \beta)$$

and for the mixture weights, $\boldsymbol{\omega} = \{\omega_1, \ldots, \omega_k\}$ we have $\boldsymbol{\omega} \sim$ Dirichlet $(\boldsymbol{\epsilon})$, with corresponding density function,

$$f(\boldsymbol{\omega}) \propto \prod_{i=1}^{k} \omega_i^{\epsilon_i - 1}$$

for $\epsilon_i > 0$, $i = 1, \ldots, k$ such that $\sum_{i=1}^{k} \omega_i = 1$.

(a) (i) Calculate the posterior density of the parameters $\boldsymbol{\theta} = \{\mu_1, \ldots, \mu_k, \sigma_1^2, \ldots, \sigma_k^2, \boldsymbol{\omega}\}$ given the data, $\mathbf{x} = \{x_1, \ldots, x_n\}$, up to proportionality.

(ii) Describe how we may update each μ_i using the Metropolis Hastings algorithm.

(iii) Suggest a Metropolis Hastings updating procedure for each σ_i^2, using a symmetric random walk proposal distribution, and give an explicit expression for the corresponding acceptance probability.

(iv) Describe an independent Metropolis Hastings algorithm for updating the mixture weights $\boldsymbol{\omega}$, once again giving the corresponding acceptance probability.

(b) An alternative approach for sampling from the posterior distribution is to use a data augmentation approach, and introduce the auxiliary variables $\mathbf{z} = \{z_1, \ldots, z_n\}$ where $z_j \in \{1, 2, \ldots, k\}$ indicates which density the data x_j is drawn from. We treat $\mathbf{z}$ as missing data and construct the joint posterior distribution over the parameters $\boldsymbol{\theta}$ and $\mathbf{z}$. The corresponding posterior distribution is given by

$$\pi(\boldsymbol{\theta}, \mathbf{z}|\mathbf{x}) \propto \prod_{j=1}^{n} \omega_{z_j} \frac{1}{\sqrt{2\pi\sigma_{z_j}^2}} \exp\left(-\frac{(x_j - \mu_{z_j})^2}{2\sigma_{z_j}^2}\right)$$
$$\times \prod_{i=1}^{k} \left[\exp\left(-\frac{(\mu_i - \mu)^2}{2\sigma^2}\right) (\sigma_i^2)^{-(\alpha+1)} \exp(-\beta\sigma_i^2) \omega_i^{\epsilon_i - 1}\right] .$$

(i) Describe how we can use the Gibbs sampler to update each μ_i, σ_i^2 and $\boldsymbol{\omega}$.

(ii) By calculating the conditional posterior probability that observation x_j comes from density i or otherwise, describe how we can sample z_j from its posterior conditional distribution.

(c) Discuss the practical implementation issues involved with the approaches outlined in (a) and (b).

4 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 Binomial, $\mathrm{Bin}(n, p)$ (for fixed n), or a Normal, $N(\mu, \sigma^2)$, provides the best model for these data. Calculate the MLE's for p, 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 MLE in this case.

Now calculate the Boltzmann distribution with $f(p)$ equal to the log-likelihood under the Binomial model and show that this converges to a point mass at the MLE 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.

5 Time Series

Consider the time series data $y_1, y_2, \ldots, y_T$, where $T = 2m + 1$. Let $\omega_j = 2\pi j/T$, $j = 1, \ldots, m$. Motivate and interpret the periodogram,

$$I(\omega_j) = \frac{1}{\pi T}\left[\Big(\sum_{t=1}^{T} y_t \cos(\omega_j t)\Big)^2 + \Big(\sum_{t=1}^{T} y_t \sin(\omega_j t)\Big)^2\right], \qquad j = 1, \ldots, m.$$

Show that if $\{y_t\}$ is Gaussian white noise with variance σ^2 then $I(\omega_1), \ldots, I(\omega_m)$ are independent $\big(\sigma^2/\pi\big)\chi_2^2/2$. Deduce that $I(\omega)$ is an unbiased but not consistent estimator of the spectrum. Assuming that the spectal density $f(\omega)$ is smooth, explain how to construct a consistent estimator of $f(\omega)$.

Consider the hypothesis that the time series $y_1, y_2, \ldots, y_T$ is purely random. Describe the turning point test of this hypothesis. Show that for all T large enough the number of turning points P_T has mean $2(T-2)/3$ and variance $(16T-29)/90$.

Hint: you may use without proof the following facts: for $j, k \in \{1, \ldots, m\}$

$$\sum_{t=1}^{T} \cos(\omega_j t)\cos(\omega_k t) = \sum_{t=1}^{T} \sin(\omega_j t)\sin(\omega_k t) = (T/2)\delta_{jk},$$

$$\sum_{t=1}^{T} \cos(\omega_j t) = \sum_{t=1}^{T} \sin(\omega_j t) = \sum_{t=1}^{T} \cos(\omega_j t)\sin(\omega_k t) = 0.$$

6 Time Series

Let X_t be defined via

$$X_t = \phi X_{t-1} + \epsilon_t + \theta \epsilon_{t-1},$$

where ϕ and θ are some constants and ϵ is a white noise process of variance σ^2. Specify conditions on ϕ and θ under which X is a stationary invertible ARMA(1,1) process.

Compute the Wold representation, the spectral density and autocovariance function of X.

Determine the linear least-square predictor of X_t in terms of X_{t-1}, X_{t-2}, $\ldots$.