UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Friday, 4 June, 2010 1:30 pm to 4:30 pm

PAPER 33

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

Define an autoregressive and moving average time series process of order p and q, i.e., ARMA(p, q). Now define the autoregressive and moving average operators, and use these to write your ARMA process in a concise form. What conditions must hold in order for this process to be *causal* and *invertible*?

Briefly describe how the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) can be used to obtain a good guess of appropriate p and q given a particular time series data set.

Consider the following process

$$x_t = 0.5\, x_{t-1} + w_t - 1.4\, w_{t-1} + 0.45\, w_{t-2} \,,$$

where w_t is IID white noise with variance σ_w^2. Argue that this is an ARMA(p, q) process, and specify the p and q. Is it causal and/or invertible? What is the ACF of this process?

2 Time Series

Consider the state-space model defined by

$$\text{the state equation} \qquad \mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{w}_t, \qquad \mathbf{w}_t \overset{\text{iid}}{\sim} \mathcal{N}_p(\mathbf{0}, Q),$$

$$\text{and the observation equation} \qquad \mathbf{y}_t = A\mathbf{x}_t + \mathbf{v}_t \qquad \mathbf{v}_t \overset{\text{iid}}{\sim} \mathcal{N}_q(\mathbf{0}, R).$$

The initial state is $\mathbf{x}_0 \sim \mathcal{N}_p(\boldsymbol{\mu}_0, \Sigma_0)$, and $\mathbf{w}_t$ and $\mathbf{v}_t$ are independent. Let

$$\begin{aligned} \mathbf{x}_t^s &= \mathbb{E}\{\mathbf{x}_t \,|\, \mathbf{y}_1, \ldots, \mathbf{y}_s\} \\ \text{and} \quad P_t^s &= \mathbb{E}\{(\mathbf{x}_t - \mathbf{x}_t^s)(\mathbf{x}_t - \mathbf{x}_t^s)^\top\} \end{aligned}$$

be the mean and covariance of the multivariate normal predictive distribution of the state vector at time t given data up to time s.

Suppose that $\mathbf{x}_{t-1}^{t-1}$ and P_{t-1}^{t-1} are known. Give and justify an expression for the mean and covariance of the forecasted states, $\mathbf{x}_t^{t-1}$ and P_t^{t-1}, in terms of the above quantities (and the definition of the state space model).

Now, assume that you also had expressions for mean and covariance of the filtered states, $\mathbf{x}_t^t$ and P_t^t, given $\mathbf{x}_t^{t-1}$ and P_t^{t-1}. Describe an algorithm for obtaining the entire set of forecasted means and covariances, $t = 1, \ldots, n$ starting from $t = 0$ with $\mathbf{x}_0^0 = \boldsymbol{\mu}_0$, and $P_0^0 = \Sigma_0$.

Obtain an expression for the innovations $\boldsymbol{\epsilon}_t = \mathbf{y}_t - \mathbb{E}\{\mathbf{y}_t \,|\, \mathbf{y}_1, \ldots, \mathbf{y}_{t-1}\}$, $\mathbb{E}\{\boldsymbol{\epsilon}_t\}$, and $\Sigma_t = \mathbb{V}\text{ar}(\boldsymbol{\epsilon}_t)$ in terms of the quantities calculated above, and use these to derive the likelihood for the parameters $\Theta = (\Phi, A, Q, R)$ implied by the state space model. Explain how the likelihood may be evaluated, and thereby how a maximum likelihood estimator $\hat{\Theta}$ for Θ may be obtained.

3 Monte Carlo Inference

Describe the rejection sampling algorithm for simulating a random variable with density f. Prove that the output of the algorithm does indeed have density f. Define the rejection rate M.

Give a rejection sampling algorithm for obtaining samples from a standard normal distribution using independent samples from a uniform distribution and independent samples generated from the density $g(y) = \lambda e^{-\lambda y} \mathbb{I}_{\{y \geqslant 0\}}$. Derive the setting of $\lambda > 0$ that minimises the rejection rate.

Briefly explain how you would modify the algorithm above to use only independent samples from a uniform distribution.

4 Monte Carlo Inference

Let $x_1, \ldots, x_n$ be distinct observed data values. Show that there are $\binom{2n-1}{n}$ distinct possible non-parametric bootstrap samples (i.e. up to rearrangement).

Describe the estimator $\hat{\theta}$ for a quantity

$$\theta = \int_2^\infty \frac{1}{\pi(1+x^2)}\,dx$$

that would be obtained by the following R code

```
R1a> n <- 100
R2a> x <- rcauchy(n)
R3a> theta.hat <- mean(x > 2)
```

Now consider the following R code (with the same `x` as above).

```
R1b> B <- 199
R2b> M <- matrix(NA, nrow=B, ncol=n)
R3b> for(b in 1:B) M[b,] <- sample(x, n, replace=TRUE)
R4b> v <- apply(M, 1, function(x){mean(x > 2)})
R5b> var(v)
```

Explain what is being calculated in the code, with particular attention paid to the expression in line `R5b`.

Describe the estimator $\tilde{\theta}$ obtained using the following R code. In particular, what role does `y` play?

```
R1c> u <- runif(n)
R2c> y <- 2/(1-u)
R3c> w <- y^2/(2*pi*(1+y^2))
R4c> theta.tilde <- mean(w)
```

Finally, consider the following R code (with the same `y` as above).

```
R1d> M2 <- matrix(NA, nrow=B, ncol=n)
R2d> for(b in 1:B) M2[b,] <- sample(y, n, replace=TRUE)
R3d> v2 <- apply(M2, 1, function(y){mean(y^2/(2*pi*(1+y^2)))})
R4d> var(v2)
```

Explain what is being calculated in the code. How do you think the value calculated in line `R4d` compares to the one in line `R5b` above?

5 Monte Carlo Inference

Briefly, and qualitatively, compare and contrast the Metropolis–Hastings algorithm and the Gibbs sampler. You should describe, in particular, when one might be preferred over the other, and remark on any possible variations.

Let $x_1, \ldots, x_n$ be independent normally distributed observations for which there is a suspicion of a *change–point* along the observation of the process for some random $m = 1, \ldots, n$. That is, given m we have that $x_i \mid \theta, \sigma^2 \sim N(\theta, \sigma^2)$, $i = 1, \ldots, m$ and $x_i \mid \phi, \sigma^2 \sim N(\phi, \sigma^2)$, $i = m+1, \ldots, n$. Our prior assumptions are that θ, ϕ, σ^2 have the scale-invariant (but improper) prior $p(\theta, \phi, \sigma^2) \propto 1/\sigma^2$ and that m has an independent discrete uniform distribution over $\{1, \ldots, n\}$.

Determine the joint posterior distribution of $(\theta, \phi, \sigma^2, m)$ given data $\mathbf{x} = (x_1, \ldots, x_n)$ up to a constant of proportionality.

Devise a Markov chain Monte Carlo scheme based on the Gibbs sampler to sample from this posterior distribution.

What is the meaning of $P(m = n \mid \mathbf{x})$? How would you estimate this quantity using the sample obtained under the scheme you devised above?

Note that a random variable with an inverse gamma distribution, $W \sim \mathrm{IG}(\alpha, \beta)$, has pdf

$$f(w; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{-(\alpha+1)} e^{-\beta/w} .$$

6 Monte Carlo Inference

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 $L(\mathbf{x} \mid \boldsymbol{\theta})$. In particular, state explicitly how the value of $\boldsymbol{\theta}^{(t+1)}$ is obtained in iteration $t+1$ conditional on $\boldsymbol{\theta}^{(t)}$ from the previous iteration, t.

Suppose you have four data points in two dimensions, one of which has a missing feature in the first component:

$$\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\} = \left\{ \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix}, \begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix}, \begin{pmatrix} x_{31} \\ x_{32} \end{pmatrix}, \begin{pmatrix} z \\ x_{42} \end{pmatrix} \right\},$$

where z represents the unknown value. Consider modelling this data with a bivariate normal distribution $N_2(\boldsymbol{\mu}, \Sigma)$ where

$$\boldsymbol{\mu} = (\mu_1, \mu_2), \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}.$$

Describe the EM steps that can be used to update the parameters at iteration t, and derive explicit expressions for the parameters $(\mu_1^{(t+1)}, \mu_2^{(t+1)}, \sigma_1^{2(t+1)}, \sigma_2^{2(t+1)})$ at iteration $t+1$ in terms of the observations and the parameters from the previous iteration t.

END OF PAPER