UNIVERSITY OF
CAMBRIDGE

M. PHIL. IN STATISTICAL SCIENCE

Thursday, 4 June, 2009 9:00 am to 12:00 pm

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

Let

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + \epsilon_t + \sum_{k=1}^{q} \theta_k \epsilon_{t-k}, \tag{1}$$

where $\phi_1, \dots, \phi_p$ and $\theta_1, \dots, \theta_q$ are constants, and where $\{\epsilon_t\}$ is a white noise process with zero mean and variance σ^2. Write down conditions for $\{X_t\}$ to be a stationary, causal, invertible $ARMA(p,q)$ process.

For the rest of the question, assume that these conditions are satisfied and that $p = 1$. Find the Wold representation $X_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$, giving explicit expressions for each c_j, $j \geqslant 0$, in terms of ϕ_1 and $\theta_1, \dots, \theta_q$.

Let $\gamma_k = \text{cov}(X_t, X_{t-k})$. Show that

$$\gamma_k - \phi_1 \gamma_{k-1} = \begin{cases} 0 & \text{for } k > q \\ \sigma^2 \theta_q c_0 & \text{for } k = q \\ \sigma^2(\theta_{q-1} c_0 + \theta_q c_1) & \text{for } k = q-1 \\ \vdots & \vdots \\ \sigma^2(\theta_1 c_0 + \ldots + \theta_q c_{q-1}) & \text{for } k = 1 \\ \sigma^2(c_0 + \theta_1 c_1 + \ldots + \theta_q c_q) & \text{for } k = 0. \end{cases}$$

Write down the above equations when $p = q = 1$, and find γ_k, $k \in \mathbb{Z}$, in this case.

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

2 Time Series

Consider the state space model $X_t = FS_t + v_t$ and $S_t = GS_{t-1} + w_t$, where X_t and S_t are scalars that are observed and unobserved respectively, F and G are known constants, and $\{v_t\}$ and $\{w_t\}$ are uncorrelated white noise processes with variances V and W respectively. Let $\mathcal{F}_{t-1} = (X_1, \ldots, X_{t-1})$, and suppose that $S_{t-1} \mid \mathcal{F}_{t-1}$ has a normal distribution with mean $\hat{S}_{t-1}$ and variance P_{t-1}. Show that $S_t \mid \mathcal{F}_{t-1}$ has a normal distribution with mean $G\hat{S}_{t-1}$ and variance R_t , where $R_t = G^2 P_{t-1} + W$. Show that $X_t \mid S_t, \mathcal{F}_{t-1}$ is normally distributed with mean FS_t and variance V.

Using the hint below, show that the distribution of $(X_t, S_t)^T$ conditional on $\mathcal{F}_{t-1}$ is bivariate normal with mean vector $(FG\hat{S}_{t-1}, G\hat{S}_{t-1})^T$ and covariance matrix

$$\begin{pmatrix} Q_t & FR_t \\ FR_t & R_t \end{pmatrix},$$

where $Q_t = F^2 R_t + V$. Deduce that the conditional distribution of S_t given X_t and $\mathcal{F}_{t-1}$ is normal with mean $\hat{S}_t$ and variance P_t, where

$$\hat{S}_t = G\hat{S}_{t-1} + FR_t Q_t^{-1}(X_t - FG\hat{S}_{t-1}), \quad P_t = R_t - F^2 R_t^2 Q_t^{-1}.$$

Consider now the above model with $F = G = 1$. Show that the above recursions give

$$\hat{S}_t = (1-\alpha_t)\hat{S}_{t-1} + \alpha_t X_t, \quad P_t = \alpha_t V,$$

where you should give α_t in terms of P_{t-1}, V and W. Show that if P_t tends to a constant P as t tends to infinity, then α_t converges to $(-c + \sqrt{c^2+4c})/2$ as t tends to infinity, where c is the signal-to-noise ratio W/V.

[*Hint: Suppose that*

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right). \tag{1}$$

Then you are given that

$$Y_1 \mid Y_2 \sim N\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right). \tag{2}$$

Conversely, you are given that,
*if $Y_1 \mid Y_2$ satisfies (**??**) and $Y_2 \sim N(\mu_2, \Sigma_{22})$, then $(Y_1, Y_2)^T$ satisfies (**??**).*]

3 Monte Carlo Inference

(a) Describe the ratio of uniforms method for simulating random variables from a distribution with density $f_X(x)$, $-\infty < x < \infty$.

Prove that the resulting deviates do indeed have the density f_X.

(b) Explain why the ratio of uniforms method is usually combined with rejection sampling in most applications.

Give an algorithmic description of the combined method, requiring only uniform deviates $U \sim \mathrm{Unif}(0,1)$, that can be applied generally. *[You need not prove that the resulting method is correct.]*

What is the general form of the acceptance rate of the method?

(c) Show how you would use the ratio of uniforms method to obtain samples $X \sim N(0, \sigma^2)$, and calculate the acceptance rate in this case.

4 Monte Carlo Inference

(a) Describe the estimator $\hat{\theta}$ for a quantity θ (which you should also determine) that would be obtained by the following R code

```
> theta.hat <- mean(rcauchy(n) > 2)
```

and derive an expression for $\text{Var}(\hat{\theta})$ in terms of $n \equiv$ n.

(b) Now consider a function f1 for estimating the same quantity θ, given below.

```
f1 <- function(n)
{
   u <- runif(n)
   y <- 2/(1-u)
   w <- y^2/(2*pi*(1+y^2))
   return(mean(w))
}
```

Describe the estimator $\tilde{\theta}$ that would be obtained by the following call.

```
> theta.tilde <- f1(n)
```

What role does y play in the above function (f1)? Derive an expression for the variance of $\tilde{\theta}$ terms of n and θ.

(c) Finally, consider a function f2 for estimating the same quantity θ as above.

```
f2 <- function(n, beta)
{
   y <- runif(n, 0, 2)
   theta <-  sum(1/(pi*(1+y^2)) - beta*(y^2 - 4/3))
   return(1/2 - (2*mean(theta)))
}
```

Comment, qualitatively, on the principles being used by this function (f2) to construct the estimator. What is the form of the estimator $\check{\theta}$ obtained in the special case when $\beta = 0$ by the following call?

```
> theta.check <- f2(n, 0)
```

Can you suggest how β might be optimally chosen?

5 Monte Carlo Inference

(a) Describe the Metropolis Hastings (MH) algorithm and the Gibbs sampler, for obtaining a dependent sample from some distribution $\pi(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^k$, and prove that the Gibbs sampler is a special case of the MH algorithm.

(b) Now take the partially standardised bivariate normal distribution

$$\pi_{\sigma,\rho}(x,y) = \frac{1}{2\pi\sigma\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2+y^2-2\rho xy}{2\sigma^2(1-\rho^2)}\right\} \quad \text{for } \rho \in (-1,1), (x,y) \in \mathbb{R}^2, \sigma^2 > 0.$$

Find the full conditional distributions $\pi_{1,\rho}(x|y)$ and $\pi_{1,\rho}(y|x)$ and thereby illustrate how the Gibbs sampler can be used to obtain a dependent sample from $\pi_{1,\rho}(x,y)$, i.e., when $\sigma = 1$ in the expression above.

(c) Suppose that today is just not your day. Your analytical skills have failed you, and you cannot derive the full conditional distributions $\pi_{1,\rho}(x|y)$ and $\pi_{1,\rho}(y|x)$. Show how you can still obtain a dependent sample from $\pi_{1,\rho}(x,y)$ by taking proposals $(x',y') \sim \pi_{\sigma_q,0}(x,y)$, i.e., from a bivariate normal centred at (x,y) with covariance matrix $\sigma_q^2\mathbf{I}_2$, by applying the MH algorithm. Give an expression for the acceptance probability $\alpha((x,y),(x',y'))$ and thereby detail how the Markov chain transitions from one state to the next.

(d) Sketch a method for optimising the MH algorithm by tuning the proposal variance σ_q^2. Why is optimising $\alpha((x,y),(x',y'))$ not sensible?

6 Monte Carlo Inference

(a) Let $\mathbf{x}$ represent observed data, and $\mathbf{z}$ denote missing data, with joint distribution $f(\mathbf{x},\mathbf{z};\boldsymbol{\theta})$. Describe the iterative Expectation Maximisation (EM) algorithm for finding the $\hat{\boldsymbol{\theta}}$ that maximises the observed data likelihood $L(\mathbf{x}|\boldsymbol{\theta})$. In particular, state explicitly how the value of $\boldsymbol{\theta}^{(t+1)}$ obtained in iteration $t+1$ is derived from the value of $\boldsymbol{\theta}^{(t)}$ obtained in iteration t.

(b) Prove that every step of the EM algorithm increases the log likelihood. That is,

$$\log L(\mathbf{x}|\boldsymbol{\theta}^{(t+1)}) \geqslant \log L(\mathbf{x}|\boldsymbol{\theta}^{(t)}).$$

(c) Comment on the implications of the result in part **(b)** in the context of searching for a maximum likelihood estimator.

(d) Briefly compare EM with the method of Data Augmentation.

END OF PAPER