REPORT OF THE BOARD OF EXAMINERS

April 2003

Subject 103 — Stochastic Modelling

EXAMINERS' REPORT

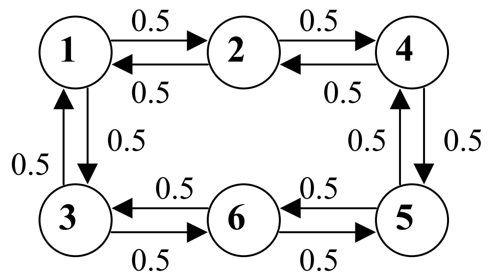
Introduction

The attached subject report has been written by the Principal Examiner with the aim of helping candidates. The questions and comments are based around Core Reading as the interpretation of the syllabus to which the examiners are working. They have however given credit for any alternative approach or interpretation which they consider to be reasonable.

J Curtis
Chairman of the Board of Examiners

3 June 2003

1 (i) Transition diagram:



(ii) (a) and (c) are both zero, as it is not possible to return to 1 in an odd number of steps.

For (b): $p_{1,1}^{(2)} = p_{12}p_{21} + p_{13}p_{31} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.

(iii) (a) Yes every chain with finite state space has a stationary distribution.

(b) No. The chain is periodic, so the probabilities do not converge.

It was possible to read the question as implying the possibility that the mouse could stay where it was: candidates who did this were still able to obtain full marks for the question.

Every candidate was able to draw the diagram correctly and most of them correctly evaluated the required probabilities.

Some candidates were not clear about the distinction between a stationary distribution and a limiting distribution.

2 (i) Prices and salaries are notoriously non-stationary processes, having a tendency to increase rather than a tendency to stay in the vicinity of some central value. What is more, the increase is more likely to be geometric than linear. There is some hope that $\{\nabla \ln P_t\}$, which is equal to $\{\ln (P_t/P_{t-1}\}$, may be a stationary process, and similarly for S.

(ii) If prices have recently increased, it is reasonable that workers will demand salary increases; if salaries have recently increased, there is more money in the economy, generating a tendency for prices to rise. The dependences are reasonable.

(iii) $$\begin{pmatrix} \nabla \ln S_t \\ \nabla \ln P_t \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{pmatrix} \begin{pmatrix} \nabla \ln S_{t-1} \\ \nabla \ln P_{t-1} \end{pmatrix} + \begin{pmatrix} \theta_S \\ \theta_P \end{pmatrix} + \begin{pmatrix} e_{S,t} \\ e_{P,t} \end{pmatrix}.$$

This is a first-order VAR.

(iv) Sensitivity analysis comes after model verification. In model verification you check that simulations of the process (with parameter values equal to the values estimated from data) look similar to what has actually been observed. For sensitivity analysis you check that this still works when the parameter values used for the simulation are a little bit different from the estimates. The purpose is to guard against the possibility that you have by chance used parameter values with untypical properties.

Parts (i) and (ii) were done quite well. In part (i) a lot of people concentrated just on the logs or just on the differences rather than both. In part (ii) quite a few people gave answers such as "both are linked to inflation" rather than explaining why, or only considered one way (eg why should salaries be related to prices, but not the other way around).

Most people got part (iii). In part (iv) most people mentioned varying the parameters slightly, but not many conveyed the idea of then studying the simulated output of the model and assessing whether it still looked similar to real life.

3 (i) $E(Y_t) = \mu t$, $\mathrm{Var}(Y_t) = \sigma^2 t$.

(ii) $E(X_{nh}) = n(2p - 1)\, D$, $\mathrm{Var}(X_{nh}) = nD^2 \times 4p(1 - p)$.

Therefore $E(X_t) = (2p - 1)\, D\left[\dfrac{t}{h}\right]$, $\mathrm{Var}(X_t) = 4p(1 - p)D^2\left[\dfrac{t}{h}\right]$.

(iii) (a) We require $\mu = (2p - 1)D/h$ and $\sigma^2 = 4p(1 - p)D^2/h$.

(b) This implies that $p = \dfrac{1}{2}\left(1+\dfrac{\mu h}{D}\right)$, $D^2 = \sigma^2 h + \mu^2 h^2$.

This applies only when $0 \le |h| < \mu^{-1}$. Small values of h should be used because the random walk model converges to Brownian motion / diffusion as $h \to 0$.

Parts (i) and (ii) were well answered — maybe half the people successfully equated the random walk moments with the Brownian motion ones — although many candidates failed to see the derivation of the conditions on p and D through to completion.

The biggest problem was the last part: "h small" is on the whole a bit too vague to get the full credit.

4 (i) Let $X = -(\log U)/\theta$. Then, for $x > 0$,

$$P(X > x) = P(\log U < -\theta x) = \exp(-\theta x).$$

Differentiating, $f(x) = \theta e^{-\theta x}$, as required.

Alternatively: use the inverse distribution function method.
$F(x) = \int_0^x \theta e^{-\theta y} dy = 1 - e^{-\theta x}$. We need to invert this: set $U = F(X)$ and express X as a function of U. This gives $X = -\theta^{-1} \log(1-U)$.

(ii) (a) Use $f(x) = \theta \exp(-\theta x)$ as the base density. We need to find a constant C such that $k(\theta)\dfrac{e^{-\theta x}}{1+x} \le C\theta e^{-\theta x}$ for all $x > 0$.

$C = k(\theta)/\theta$ is the best that we can do.

The procedure is:

1. Generate a value y from the density $f(x) = \theta \exp(-\theta x)$.

2. Take another Uniform pseudo-random variable U_2; if this is less than $g(y)/(C f(y))$ [which is equal to $1/(1 + y)$] then we accept the value y, otherwise reject it and return to 1.

(b) On average it takes C repetitions of steps 1 and 2 to generate a value. Each such repetition requires two uniform pseudo-random variables. So the answer is $2\, k(\theta)/\theta$.

Most candidates did well here, although a few people inverted the density function instead of the distribution function.

Some experienced real difficulty in providing a clear statement of the algorithm for Acceptance-Rejection method. A lot of people answered this part in abstract without realising that they should use the density from part (i) as the base density, and hence didn't get the marks for calculating C.

5 (i) (a) $EW(t) = 0$, $\text{Cov}(W(s), W(t)) = 16\text{Cov}(B(ks), B(kt)) = 16k \min(s, t)$.

Thus $k = 1/16$ is the required value.

To prove that a process is a BM, it is necessary to check the covariance, not just the variance.

(b) In addition to the expectation and covariance function, we need to show

* that W is Normally distributed. Any linear transformation of a Normal random variable is itself Normal.
* that W has continuous sample paths. For small h, we have $W(t + h) - W(t) = -4(B(kt + kh) - B(kt))$, which clearly tends to 0 as h approaches 0.

* An alternative to proving continuity is to state that the increments of W are independent of past values and are also Normally distributed.

(ii) Consider $\exp[2B(t)–2t]$ with respect to the filtration F_t.

$B(t) – B(s)$ is independent of F_t and $B(s)$ is F_s-measurable. Then

$$\begin{aligned} E\left(e^{2B(t)} \mid F_s\right) &= E\left(e^{2[B(t)-B(s)]}e^{2B(s)} \mid F_s\right) \\ &= e^{2B(s)}E\left(e^{2[B(t)-B(s)]} \mid F_s\right) \\ &= e^{2B(s)}E\left(e^{2[B(t)-B(s)]}\right) \end{aligned}$$

The increment $B(t) – B(s)$ has the normal distribution with mean 0 and variance $t – s$, so the expectation of $e^{2[B(t) – B(s)]}$ is equal to $M(2) = \exp(2(t – s))$, where $M(\cdot)$ is the moment generating function of the $N(0, t – s)$ distribution.

It follows that

$$E\left(e^{2B(t)-2t} \mid F_s\right) = e^{2B(s)-2s}$$

and therefore $e^{2[B(t)- t]}$ is a martingale.

For part (i), in (a) many candidates just checked variance rather than the covariance, but got 1/16 correct. Not so many people got part (b) — generally people assumed that (a) implied (b), rather than considering other properties of Brownian motion

Most candidates fared well on part (ii), identifying where necessary the mgf of a normal random variable to successfully demonstrate the given process is a martingale.

6 (i) $\{X(t)\}$ is **not** Markov because, for example, $P[X_{t+1} = 2 \mid X_t = 3, X_{t-1} = 2, \ldots]$ cannot be reduced to $P[X_{t+1} = 2 \mid X_t = 3]$.

(ii) (a) Define new states

$3a$ = Level 3 this year following Level 2 last year
$3b$ = Level 3 this year following Level 4 last year

(b) The transition matrix is then

	1	2	3a	4	3b
1	0.2	0.8	0	0	0
2	0.2	0	0.8	0	0
3a	0	0.2	0	0.8	0
4	0	0	0	0.8	0.2
3b	0.2	0	0	0.8	0

(c) We have

$$\begin{aligned}
&\pi_1 = 0.2\pi_1 + 0.2\pi_2 + 0.2\pi_{3b} && \rightarrow \pi_1 = 0.25\pi_2 + 0.25\pi_{3b} \\
&\pi_2 = 0.8\pi_1 + 0.2\pi_{3a} && \rightarrow \pi_2 = 0.25(\pi_{3a} + \pi_{3b}) \\
&\pi_{3a} = 0.2\pi_{3b} + 0.2\pi_{3a} && \rightarrow \pi_{3a} = 0.25\pi_{3b} \\
&\pi_{3b} = 0.2\pi_4 &&
\end{aligned}$$

Applying the condition $\sum \pi_i = 1$, we obtain the solution as

$$(\pi_1, \pi_2, \pi_{3a}, \pi_4, \pi_{3b}) = \frac{1}{441}(21, 20, 16, 320, 64).$$

It follows that the long-run proportion of time spent in Level 3 is $(16 + 64)/441 = 80/441$.

A large number of candidates fared well on this question. Some committed the mostly harmless error of including too many (redundant) states. Most people had a reasonable attempt at (ii)(c) too, although there were some errors of arithmetic or matrix multiplication.

7 (i) The generator matrix is

$$\begin{pmatrix} -\sigma-\mu & \sigma & \mu \\ \rho & -\rho-\nu & \nu \\ 0 & 0 & 0 \end{pmatrix}.$$

(ii) Let B and C be the levels of benefits and contributions, and define h (respectively s) as the expected time spent by a policyholder in state H (respectively S) before finally entering state D. For solvency the company requires $Ch - Bs \geq 0$.

The model allows h and s to be calculated. [The equations are

$$h = \frac{1}{\sigma+\mu} + \frac{\sigma}{\sigma+\mu} s, \quad s = \frac{1}{\rho+\nu} + \frac{\rho}{\rho+\nu} h.]$$

(iii) (a) This might improve predictive power as far as the individual policyholder is concerned, but in a large population of policyholders with constant age profile it is not likely to make much difference.

(b) Certain kinds of sickness are more likely to strike at particular times of year, but they may more or less balance out. If there is a significant increase in the incidence of sickness in one season, it should be helpful to include this in the model.

(c) Similarly to (a), the inclusion of duration dependence will significantly improve the goodness of fit when only one policyholder is being regarded, but in a large population of policyholders it is unlikely to make a difference overall.

(iv) If the population is split up into categories depending on age, duration of illness and time of year, it is highly unlikely that there will be sufficient data to estimate all the transition rates reliably.

(v) This can be regarded a Time Series problem: we seek a test for the existence of seasonal variation. One suggestion could be to find the best-fitting model without seasonal variation and the best-fitting with seasonal variation, then to compare the values of the Akaike Information Criterion.

Alternatively, if the number of policyholders is roughly constant from year to year, one could use one-way Analysis of Variance to determine whether Season has a significant effect on claims. If the number of policyholders varies greatly from year to year, a two-way ANOVA with Year and Season as explanatory variables could be fitted.

A test based on χ^2 would have to be carefully constructed. For example, drawing up a contingency table with the quarters as the rows and with column titles like "Stayed healthy" and "Fell sick" would be a reasonable approach.

Again done quite well generally.

Many candidates weren't clear that they needed to write a formula that characterises the company's policy in part (ii).

Parts (iii) and (iv) done well —most people seemed to get the general ideas and write sensible comments, even if they didn't quite get full marks. Only a few candidates *noticed that in a large stable population of policyholders the trainee's suggestions would not greatly improve the model's predictive power.*

In part (v) many candidates merely mention using the chi-squared test without explicitly describing it. The question was carefully written to elicit responses which showed whether the candidate understood what was to be tested.

8 (i) State space for X_t is $\{0, 1, 2, 3\}$.

(ii) (a) Let T be the time taken for a cash dispenser to break down. From the question:

$$P\big(T \in (t, t+dt) \mid T > t\big) = \alpha dt + o(dt)$$

or in other words $\dfrac{P\big(T \in (t, t+dt)\big)}{P(T > t)} = \alpha dt + o(dt)$

If we set $F(t) = P(T \leq t)$, then

$$\frac{F'(t)\,dt}{1 - F(t)} = \alpha dt + o(dt)$$

and letting $dt \to 0$,

$$\frac{F'(t)}{1 - F(t)} = \alpha$$

Integrate this to get $-\ln\big(1 - F(t)\big) = \alpha \quad t + const$

Since $F(0) = 0$, we can set $const = 0$, and hence $F(t) = 1 - e^{-\alpha t}$

That is, T has an exponential distribution with mean $1/\alpha$.

(b) $P(\text{no breakdowns by time } t) = (e^{-\alpha t})^3 = e^{-3\alpha t}$. Thus the time until the first breakdown is exponentially distributed with parameter 3α.

(iii) $P(X(t+h) = m) = P(X(t) = m, X(t+h) = m) + P(X(t) = m-1, X(t+h) = m)$
$+ P(X(t) = m+1, X(t+h) = m) + o(h)$.

Thus in general

$$P_m(t+h) = P_m(t)[1 - ((3-m)\alpha + m\beta)h] + P_{m-1}(t)(4-m)\alpha h 1_{m>0}$$
$$+ P_{m+1}(t)(m+1)\beta h 1_{m<3} + o(h),$$

where 1_A is the indicator function of event A.

Rearranging and letting $h \to 0$ gives the required equations.

(iv) First look at the LHS: if P_m has the given form, then

$$P_m'(t) = \binom{3}{m}\theta(t)^{m-1}(1-\theta(t))^{2-m}\theta'(t)[m(1-\theta(t))-(3-m)\theta(t)]$$

On the other hand, the RHS is equal to

$$-\binom{3}{m}\theta(t)^m(1-\theta(t))^{3-m}\left[m\beta+(3-m)\alpha\right]+$$

$$(4-m)\alpha\binom{3}{m-1}\theta(t)^{m-1}(1-\theta(t))^{4-m}+(m+1)\beta\binom{3}{m+1}\theta(t)^{m+1}(1-\theta(t))^{2-m}$$

$$=\binom{3}{m}\theta(t)^{m-1}(1-\theta(t))^{2-m}\left\{-[m\beta+(3-m)\alpha]\theta(1-\theta)+m\alpha(1-\theta)^2+(3-m)\beta\theta^2\right\}$$

$$=\binom{3}{m}\theta(t)^{m-1}(1-\theta(t))^{2-m}(m-3\theta)\left\{\alpha(1-\theta)-\beta\theta\right\}$$

Thus the LHS and RHS are equal as long as $\theta'(t)=\alpha-(\alpha+\beta)\theta(t)$.

Almost everyone got part (i). In part (ii), stating the time until first breakdown is exponential in (a) is not sufficient: it was necessary to derive this from first principles. Similarly, a reasonable number got the answer to (b) without deriving it properly.

The most successful strategy for part (iii) was writing the generator matrix and getting the equations from there.

Only the exceptional candidate attempted (iv); the algebra was tough, but did not involve any tricks.

9 (i) We need to prove that $E(Y_{n+1} \mid X_1, X_2, \ldots X_n) = Y_n$.

We have

$$E(Y_{n+1} \mid X_1, X_2, \ldots X_n) = E((q/p)^{S_n+X_{n+1}} \mid X_1, X_2, \ldots X_n)$$

$$= (q/p)^{S_n} E((q/p)^{X_{n+1}} \mid X_1, X_2, \ldots X_n)$$

$$= (q/p)^{S_n}(p(q/p) + q(q/p)^{-1} + (1-p-q)) = Y_n,$$

which shows that Y_n is a martingale.

By taking expectations, we obtain that $E(Y_{n+1}) = E(Y_n)$, and this in turn implies that $E(Y_n) = E(Y_0) = (q/p)^m$.

(ii) (a) It is clear that T is a stopping time with respect to this martingale, since the event that the share has reached 0 or N by time n depends only on $X_1, X_2, \ldots, X_n$ and m.]

For $n \le T$, we have $0 \le S_n \le N$,

$$Y_n = (q/p)^{S_n} \le (q/p)^N \text{ if } q > p, \text{ or } Y_n \le 1 \text{ if } q \le p.$$

(b) The conditions of the optional stopping theorem are satisfied. It follows that $E(Y_T) = E(Y_0) = (q/p)^m$.

(iii) $$\left(\frac{q}{p}\right)^m = E(Y_T) = \left(\frac{q}{p}\right)^N P(S_T = N) + \left(\frac{q}{p}\right)^0 P(S_T = 0) = 1 - \left(1 - \left(\frac{q}{p}\right)^N\right) P(S_T = N).$$

Thus $P(S_T = N) = \dfrac{1-(q/p)^m}{1-(q/p)^N}$ so that $P(S_T = 0) = \dfrac{(q/p)^m - (q/p)^N}{1-(q/p)^N}$.

(iv) (a) We have

$$\begin{aligned} E(Z_{n+1} \mid X_1, X_2, \ldots, X_n) &= E(S_n^2 + 2S_n X_{n+1} + X_{n+1}^2 - 2(n+1)p \mid X_1, X_2, \ldots, X_n) \\ &= S_n^2 - 2(n+1)p + 2S_n E(X_{n+1}) + E(X_{n+1}^2). \end{aligned}$$

But $E(X_{n+1}) = 0$, $E(X_{n+1}^2) = p + q = 2p$, and the last equation gives $E(Z_{n+1} \mid X_1, X_2, \ldots, X_n) = S_n^2 - 2np = Z_n$, so that $\{Z_n\}$ is a martingale.

(b) The conditions of the optional stopping theorem are not satisfied, since $S_n^2 - 2np$ is not bounded below for $0 \le n \le T$. But we can work with a truncated stopping time $T_K = \min(T, K)$, for which the conditions of the OST *are* satisfied, then let $K \to \infty$.

Applying the optional stopping theorem we get

$$E(Z_T) = E(Z_0) = m^2.$$

But $E(Z_T) = E(S_T^2) - 2pE(T) = N^2P(S_T = N) - 2p\,E(T)$ and we are given that $P(S_T = N) = m/N$. Thus $m^2 = E(Z_T) = Nm - 2pE(T)$, which implies that $E(T) = m(N - m)/(2p)$, as required.

Parts (i) and (iv)(a) were generally answered successfully.

Part (ii): in (a), the bounds on $(q/p)^{S_n}$ must be shown to apply for any p and q, not just some; the main problem with (b) was that many did not mention the Optional Stopping Theorem as justification for claiming that $E(Y_T) = E(Y_0)$.

Of the candidates who attempted (iv)(b), very few noticed that the stopping time did not meet the requirements of the OST, so needs to be truncated.

10 (i) The model can be written as

$$(1 - 1.7B + 0.4B^2 + 0.3B^3)\,X_t = (1 - 0.7B + 0.12B^2)\,e_t.$$

The term in the brackets on the LHS above is divisible by $1 - B$. We have

$$(1 - 1.7B + 0.4B^2 + 0.3B^3) = (1 - B)(1 - 0.7B - 0.3B^2).$$

The last term is also divisible by $1 - B$, giving

$$(1 - 1.7B + 0.4B^2 + 0.3B^3) = (1 - B)(1 - 0.7B - 0.3B^2) = (1 - B)^2(1 + 0.3B).$$

(ii) The model can be identified as an ARIMA(1, 2, 2) process.

(iii) $\{X_t : t \geq 0\}$ is clearly a non-stationary process, as the AR operator $(1 - B)^2(1 + 0.3B)$ has two roots with modulus one; any ARIMA(p, d, q) process with $d > 0$ is non-stationary.

(iv) $(1 + 0.3B)W_t = (1 - 0.7B + 0.12B^2)e_t$, which can be expressed as $W_t = (1 + 0.3B)^{-1}(1 - 0.7B + 0.12B^2)e_t$.
Expanding the first term, $W_t = (1 - 0.3B + 0.09B^2 - \ldots)(1 - 0.7B + 0.12B^2)e_t$ $= (1 - B + 0.42B^2 - \ldots)e_t$.
Therefore $\psi_0 = 1$, $\psi_1 = -1$, $\psi_2 = 0.42$.

(v) Just invert the previous representation: $e_t = (1 - B + 0.42B^2 - \ldots)^{-1}\,W_t = (1 + B - 0.42B^2 + B^2 - \ldots)\,W_t$, so that $\pi_1 = -1$, $\pi_2 = -0.58$.

(vi) None of the three processes is Markov; we know that if $\{X_t : t \geq 0\}$ is an ARIMA(p, d, q) process with $q > 0$, then any finite collection $(X_n, X_{n-1}, \ldots, X_{n-m+1})^T$ is non-Markov.

(vii) First transform the data, so that the differenced observations $y_t = (1 - B)x_t$ may be thought of as realisations of a stationary ARIMA(1, 1) model.

Estimate the parameters of the model (by Maximum Likelihood or Method of Moments) and obtain forecasts for future values of y.

Transform these back to obtain forecasts for the x values.

Part (i) was done very well; of people with the wrong answer, most stopped after taking out one (1–B) factor rather than trying for the 2nd. (ii) and (iii) were also well answered, following on from (i).

In (iv) there were difficulties with the expansion of the denominator, meaning that few candidates did well. Similar problems were encountered with part (v).

For the verification or otherwise of the Markov property in (vi) it would have helped to write out $\{Y_t\}$ *and* $\{Z_t\}$ *explicitly in terms of lagged terms.*

Most candidates who attempted (vii) got at least part of the method, but few wrote down all the main steps in the correct sequence.