EXAMINATIONS

September 2002

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.

K G Forman
Chairman of the Board of Examiners
12 November 2002

© Faculty of Actuaries
© Institute of Actuaries

Questions involving straightforward applications of Markov Chains and Time Series were well answered, and the standard of answers to questions about martingales and Brownian motion is improving from year to year, but candidates appeared to experience unexpected difficulties in relation to the questions on Markov jump processes.

1 (i) Prove this using the Markov property. (Note that this mark can be earned for use of the property even if the word "Markov" is not mentioned.)

If $s_0 < s_1, \ldots < s_n < s < t$, then

$$P[X_t - X_s = j \mid X_{s_0} = i_0, \ldots, X_{s_n} = i_n, X_s = i]$$

$$= P[X_t = i + j \mid X_s = i] = \frac{(\lambda(t-s))^j}{j!} e^{-\lambda(t-s)},$$

independent of $i_0, i_1, \ldots, i_n$.

(ii) We need to prove that $E[X_t \mid \mathcal{F}_s] = X_s$.

With $s < t$ we have

$$E[X_t \mid \mathcal{F}_s] = E[X_t - X_s \mid \mathcal{F}_s] + E[X_s \mid \mathcal{F}_s]$$

$$= E[X_t - X_s] + X_s = \lambda(t-s) + X_s.$$

Thus $E[X_t - \lambda t \mid \mathcal{F}_s] = X_s - \lambda s.$

The key to part (i) was to use the Markov property; only a few candidates managed to do this part. Part (ii) was generally well answered.

2 (i) Since the equation can be written $(1 - \alpha B)\ Y = (1 + (1 - \alpha)B)Z$, the process is ARMA(1,1), or ARIMA(1,0,1).

(ii) There are a number of possible ways to calculate the γ_k, and one way which sidesteps the γ_k and calculates the ρ_k directly. The solution presented here is one of the possible answers; other methods were marked on their merits.

$\gamma_k = \text{Cov}\left[Y_t, Y_{t-k}\right].$

Rearrange the time series equation to give $Y_t = \alpha Y_{t-1} + Z_t + \left(1-\alpha\right) Z_{t-1}$

Now $\text{Cov}[Y_t, Z_t] = \sigma^2$

and $\quad \text{Cov}[Y_t, Z_{t-1}] = \alpha.\text{Cov}[Y_{t-1}, Z_{t-1}] + \text{Cov}[Z_t, Z_{t-1}] + (1-\alpha).\text{Cov}[Z_{t-1}, Z_{t-1}]$

$$= \alpha.\sigma^2 + 0 + (1-\alpha).\sigma^2 = \sigma^2$$

Therefore

$$\gamma_0 = \text{Cov}\left[Y_t, Y_t\right] = \alpha.\text{Cov}\left[Y_t, Y_{t-1}\right] + \text{Cov}\left[Y_t, Z_t\right] + (1-\alpha).\text{Cov}\left[Y_t, Z_{t-1}\right]$$

$$= \alpha.\gamma_1 + \sigma^2 + (1-\alpha).\sigma^2$$

$$\Rightarrow \gamma_0 = \alpha.\gamma_1 + (2-\alpha).\sigma^2 \qquad (1)$$

$$\gamma_1 = \text{Cov}\left[Y_t, Y_{t-1}\right]$$

$$= \alpha.\text{Cov}\left[Y_{t-1}, Y_{t-1}\right] + \text{Cov}\left[Z_t, Y_{t-1}\right] + (1-\alpha).\text{Cov}\left[Z_{t-1}, Y_{t-1}\right]$$

$$= \alpha.\gamma_0 + 0 + (1-\alpha).\sigma^2$$

$$\Rightarrow \gamma_1 = \alpha.\gamma_0 + (1-\alpha).\sigma^2 \qquad (2)$$

substitute for γ_0 from (1) into (2)

$$\Rightarrow \gamma_1 = \alpha.\left(\alpha.\gamma_1 + (2-\alpha)\sigma^2\right) + (1-\alpha).\sigma^2$$

$$\Rightarrow \gamma_1 = \frac{\left[(2-\alpha).\alpha + (1-\alpha)\right].\sigma^2}{1-\alpha^2} = \left(\frac{1+\alpha-\alpha^2}{1-\alpha^2}\right).\sigma^2$$

substitute for γ_1 back into (1)

$$\Rightarrow \gamma_0 = \alpha.\frac{\left(1+\alpha-\alpha^2\right).\sigma^2}{1-\alpha^2} + (2-\alpha).\sigma^2 = \left(\frac{2-\alpha^2}{1-\alpha^2}\right).\sigma^2$$

For $k \geq 2$,

$$\gamma_k = \text{Cov}\left[Y_t, Y_{t-k}\right]$$

$$= \alpha.\text{Cov}\left[Y_{t-1}, Y_{t-k}\right] + \text{Cov}\left[Z_t, Y_{t-k}\right] + (1-\alpha).\text{Cov}\left[Z_{t-1}, Y_{t-k}\right]$$

$$= \alpha.\gamma_{k-1} + 0 + 0$$

$$\Rightarrow \gamma_k = \alpha^{k-1}.\gamma_1$$

The autocorrelation function is: $\rho_k = \dfrac{\gamma_k}{\gamma_0}$. Therefore

$$\rho_0 = 1$$

$$\rho_1 = \frac{1+\alpha-\alpha^2}{2-\alpha^2}$$

$$\rho_k = \alpha^{k-1}\rho_1 \qquad k \geq 2$$

Part (i) was generally well answered, using a variety of different approaches. Candidates generally made good attempts at part (ii), the main problems occurring being a failure to correctly specify γ_0 and algebraic errors in solving the simultaneous equations.

3 (i) If the driver is to have had at least i accidents by time $t + dt$, either there must have been i accidents by time t or there must have been exactly $i - 1$ by time t and another between t and $t + dt$.

$P(\text{exactly } i-1) = P(\text{at least } i-1) - P(\text{at least } i)$.

Therefore

$$\frac{da_i}{dt} = i\beta(a_{i-1} - a_i).$$

(ii) Verification: $\dfrac{d}{dt}\{(1 - e^{-\beta t})^i\} = i\beta e^{-\beta t}(1 - e^{-\beta t})^{i-1}$.

$$a_{i-1} - a_i = (1 - e^{-\beta t})^{i-1}(1 - [1 - e^{-\beta t}]).$$

We should also verify that $a_i(0)$ is correct: the value should be 0 for $i > 0$, which it is.

Then, defining T_i as the time the process first hits i, we have

$$P_{0,i}(t) = P\{T_i \leq t\} - P\{T_{i+1} \leq t\} = (1 - e^{-\beta t})^i - (1 - e^{-\beta t})^{i+1} = e^{-\beta t}(1 - e^{-\beta t})^i$$

(iii) Probably a good suggestion. A driver who has had 2 accidents in 10 years is less likely to have another than a driver who has had 2 accidents in a month.

(iv) The proposed model does address the issue raised in (iii) but leads to a very high accident rate when t is close to 0, so is unsuitable.

Many candidates attempted part (i), but were unable to give a sufficiently clear description to convincingly display their understanding and so did not earn full marks. In some cases, *candidates did not correctly interpret the definition of $a_i(t)$. In part (ii) many candidates tried to solve the differential equation, where they could instead have simply shown the solution given to be valid. Candidates generally came up with sensible suggestions for part (iii), but only a few candidates were able to make suitable comments on part (iv).*

4 (i) When volatility is high, dX_t is strongly upwards; when volatility is low, dX_t is close to zero.

The first of these fits the assumptions, whereas for the second something more negative would be preferable. (It looks as though the process X has more opportunity to increase than to decrease.)

(ii) (a) Itô (time-independent case): if $dX_t = Y_t dt + Z_t dB_t$ then

$df(X_t) = \frac{df}{dx}(Y_t dt + Z_t dB_t) + ½\frac{d^2 f}{dx^2} Z_t^2 dt.$ or simply $f'(X_t)dX_t + ½ f''(X_t)(dX_t)^2$, with an explanation of what is meant by $(dX_t)^2$.

Equally acceptable is the time-dependent version:

$$df(X_t,t) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}(Y_t dt + Z_t dB_t) + ½\frac{\partial^2 f}{\partial x^2} Z_t^2 dt.$$

(b) Here

$$\begin{aligned} df(X_t) &= f'(X_t)dX_t + ½ f''(X_t)(dX_t)^2 \\ &= f'(X_t)[Y_t dB_t + Y_t^2 dt] + ½ f''(X_t) Y_t^2 dt \\ &= f'(X_t) Y_t dB_t + Y_t^2 [f'(X_t) + ½ f''(X_t)] dt. \end{aligned}$$

In this instance, $df(X_t) = -2e^{-2X_t} V_t dB_t$.

(c) This is a martingale by the disappearance of the dt term, as $E(dB_t \mid F_t) = 0$.

Alternatively,

$$e^{-2X_T} = e^{-2X_0} - 2\int_0^T e^{-2X_t} V_t dB_t,$$

which is a martingale.

(iii) $e^{-2X_0} = E(M_0) = E(M_t) = E(e^{-2X_t}) \geq e^{-2E(X_t)}$. It follows that $E(X_t) \geq X_0$, whatever the initial state.

This confirms the initial suggestion that downward movements are too unlikely in comparison to upward ones.

The question generally showed good attempts, with a variety of different but correct versions of Itô's Lemma being given. The only real problem evident here was in part (iv), where many candidates failed to use the given inequality to prove the point.

5 (i) Discrete state space, discrete time: Markov chain, simple random walk, anything like that;
Discrete state space, continuous time: Poisson process, Markov jump process;
Continuous state space, discrete time: time series, general random walk;
Continuous state space, continuous time: Itô process, Brownian motion, etc.

(ii) (a) by definition, is monthly. Therefore a continuous time variable is not appropriate. Something like a time series would do, containing an element of Autoregression.

(b) functions in continuous time; Brownian motion, Geometric BM or any kind of diffusion or Itô process would be a suitable candidate.

(c) has a discrete state space. It would be possible to review the member's status only once a year, say in which case a Markov chain would fit, but in the absence of a remark about frequency of membership status review a continuous-time model would seem more appropriate; a Markov jump process is what we might look for.

(iii) (a) Once a model has been decided upon, parameters may be estimated by standard methods. But it is necessary to check that the model, with parameters given by their estimated values, has sample paths which resemble the data actually observed, otherwise incorrect inferences can be drawn.

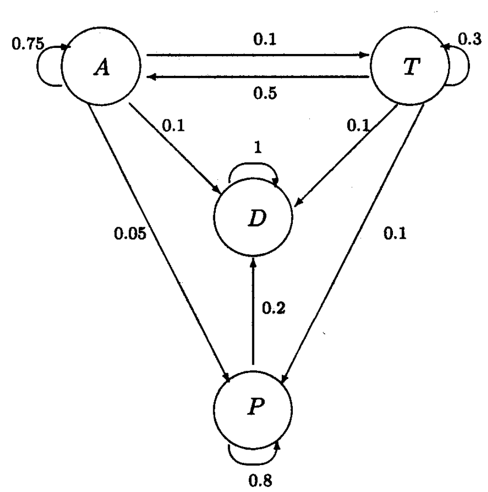
(b) Let $y_i = x_i - x_{i-1} = \nabla x_i$ and let n_u be the number of time $y_i = 1$. The parameter, θ (or p) is the probability of an up-jump, estimated as $\hat{\theta} = n_u / (n - 1)$.

We need to check that the y_i form a sequence of i.i.d. variables. A test based on chi-squared is *not* appropriate except as indicated below.

The test to apply is one which determines whether there is significant clustering of the +1's and the −1's, as opposed to their being randomly scattered through the sample: the runs test on the y_i, or equivalently the turning points test on the x_i, would be fine; a test based on the sample autocorrelation function or a contingency table ($y_i = -1$ or $+1$ as the row labels, $y_{i+1} = -1$ or $+1$ as the column labels) would also be acceptable.

In part (i) many candidates gave specific examples of observable processes, rather than describing the stochastic models themselves as requested in the question; some credit was however given for these cases. Part (ii) and (iii) (a) generally had good answers. In part (iii) (b) many candidates incorrectly suggested that a chi-squared test should be used,

6 (i)



(For full credit the probabilities should all be included on the diagram.)

Stationary distribution $\pi_A = \pi_T = \pi_P = 0$, $\pi_D = 1$. Derivation not required: the answer is obvious to anyone who understands.

(ii) The duration of a permanent disability benefit is a geometric r.v., T.

Since $P[X_{n+1} = P \mid X_n = P] = 0.8$, the parameter is 0.2:

$P[T = n] = (0.8)^{n-1}0.2$. Accordingly its mean is $\dfrac{1}{1-0.8} = 5$.

(iii) The required probability is obtained from $(1\ 0\ 0\ 0)P^3$.
$(1\ 0\ 0\ 0)P = (.75\ .1\ .05\ .1)$; $(1\ 0\ 0\ 0)P^2 = (.6125\ .105\ .0875\ .195)$;
$(1\ 0\ 0\ 0)P^3 = (.511875\ .09275\ .111125\ .28425)$.
The solution is $.09275 + .111125 = .203875$.
(The summation does not have to be performed to earn the mark: giving the two probabilities separately is a reasonable interpretation of the question.)

(iv) The probability of never visiting T or P starting from A is

$$0.1 + 0.75 \times 0.1 + (0.75)^2 \times 0.1 + \ldots$$

$$= 0.1\sum_{n=0}^{\infty}(0.75)^n = \frac{0.1}{1-0.75} = 0.4.$$

Most candidates correctly gave the transition graph in part (i), although there were a few cases where the transition probabilities were not included in the solution. Part (iii) was well answered. Solutions to (ii) and (iv) were mixed, with perhaps less than a third of students immediately recognising the required approach and thereby gaining full marks.

7 (i) The three elements are:

- the multiplier — usually denoted a
- the increment — usually denoted c; the increment is often set to zero (without any loss in the quality of the pseudo-random sequence) to speed up the generation process
- the modulus — usually denoted m — where $m > a$ and $m > c$; the generator will produce a series of pseudo-random numbers with period no more than m, so the modulus is usually set to as high a number as possible.

The recursive relationship is: $X_{n+1} = (aX_n + c) \pmod m$ for $n = 0, 1, 2, \ldots, N-1$. We then set $x_k = X_k/m$ for each $k = 1, 2, \ldots, N$.

(ii) Use the inverse transform method:

$$f(x) = \frac{\alpha\lambda^\alpha}{(\lambda+x)^{\alpha+1}} = \frac{2}{(1+x)^3}\text{, so that } F(x) = 1 - \left(\frac{1}{1+x}\right)^2$$

$$\Rightarrow x = \sqrt{\frac{1}{1-F(x)}} - 1.\text{, so}$$

- if $F(x) = 0.954$, then $x = 3.6625$
- if $F(x) = 0.462$, then $x = 0.3634$
- if $F(x) = 0.628$, then $x = 0.6396$

(iii) The structure of a Markov jump process implies that the time until the next jump has exponential distribution, with rate σ if the current state is H, ρ if S.

The even-numbered inter-jump times will have one distribution, the odd-numbered ones a different one.

Obtain numbers $y_0, y_2, y_4, \ldots, y_{2n}$ by the above procedure, being simulated outcomes of independent random variables $Y_0, Y_2, Y_4, \ldots, Y_{2n}$ exponentially distributed with parameter σ. Similarly, obtain $y_1, y_3, \ldots, y_{2n-1}$ using a parameter ρ instead of σ.

Put $t_j = y_0 + y_1 + \ldots + y_j$. Find j such that $t_{j-1} \le t < t_j$ and return

$$x_t = \begin{cases} H & \text{if } j \text{ is even} \\ S & \text{if } j \text{ is odd} \end{cases}$$

Candidates made reasonable attempts at part (i), although in many cases insufficient detail was provided to score full marks. Part (ii) was generally well answered, the only problem here being simple algebraic errors. Part (iii) was not answered well, with candidates on the whole failing to recognise that the transitions from healthy to sick and from sick to healthy should be modelled separately and then combined to provide the required process.

8 (i) $E(S_t - S_0) = \mu t$, $\text{Var}(S_t - S_0) = \sigma^2 t$.

$E(S_T - S_t) = \mu(T - t)$, $\text{Var}(S_T - S_t) = \sigma^2(T - t)$.

$\text{Cov}(S_t - S_0, S_T - S_t) = 0$, because of the independent increment property of Brownian motion.

(ii) The expectation is $\mu t - (t/T)\mu T = 0$.

$$\text{Var}(S_t - \hat{S}_t) = \text{Var}\left(\left(\frac{T-t}{T}\right)(S_t - S_0) - \frac{t}{T}(S_T - S_t)\right)$$

$$= \left(\frac{T-t}{T}\right)^2 \sigma^2 t + \left(\frac{t}{T}\right)^2 \sigma^2 (T-t) = \frac{t(T-t)\sigma^2}{T}.$$

(iii) The maximum of $t(T - t)$ is attained at $t = \frac{1}{2}T$.

This is not surprising. The graph of $S_t - \hat{S}_t$ against t is "tied down" at the ends, as the function is constrained to be equal to zero. The greatest scope for variation is bound to be in the middle.

(iv) Two possible reasons might be that a Brownian motion can become negative, which a stock price cannot, and that fluctuations in the value of a stock price are usually proportional to the price. Other reasons could also apply.

Under the revised model $\ln(S_t)$ has the same structure as S_t in the original model, so $\ln(S_t)$ will have its greatest variability at $\frac{1}{2}T$. The result will not differ greatly from the result above.

Candidates generally gave good answers for part (i). However, answers to part (ii) and (iii) were disappointing, particularly for part (iii) where the answer can be derived very simply from general reasoning. Part (iv) in general showed better answers, although many candidates failed to get full marks because they did not discuss how their response to (iii) would differ under the new model.

9 (i) $X_t = \mu + \alpha(X_{t-1} - \mu) + e_t + \beta e_{t-1}$ is the equation.

The parameters are α (the autoregressive parameter), β (the moving average parameter), μ (the mean level) and σ (the innovation standard deviation).

(ii) (a) Calculate theoretical ACF ρ_1, ρ_2 of ARMA(1, 1) in terms of α and β. Find sample ACF r_1, r_2 from the data. The required estimates are the values of α and β which ensure that $\rho_1 = r_1$ and $\rho_2 = r_2$.
The value of σ^2 is estimated using for example the calculated value of γ_0 and the sample variance.

(b) The assumptions are that the e_n are independent and Normally distributed.

(c) MLE in this case tries to minimise $\sum e_t^2$. Now e_t can be expressed as a function of x_t, x_{t-1} and e_{t-1}. e_{t-1} is unknown, but can be expressed as a function of x_{t-1}, x_{t-2} and e_{t-2}, etc. We need to estimate a suitable value of e_0.

This is done iteratively: assume that $e_0 = 0$ and estimate parameters on that basis; then use forecasting techniques on the time-reversed process $\{x_n, x_{n-1}, \ldots, x_1\}$ to gain a more accurate estimate of e_0; repeat this process until everything converges.

(iii) (a) $\hat{x}_{20}(1) = 5.67 + 0.61(8.2) + 0 - 0.23(-1.38) = 10.99.$
$\hat{x}_{20}(2) = 5.67 + 0.61(10.99) + 0 - 0 = 12.37.$

(b) For exponential smoothing the equation is
$\hat{x}_{20}(1) = \hat{x}_{19}(1) + \alpha(x_{20} - \hat{x}_{19}(1)) = 8.37 + 0.2(-0.17) = 8.34.$

(c) A variety of exponential smoothing might be better if the mean changes by some means other than a linear trend, or if there is multiplicative seasonal variation.

In part (i) a number of candidates omitted σ from the list of parameters. Part (ii) (a) showed some good answers, however in most cases insufficient details was provided to earn full marks. Part (ii) (c) was very poorly answered, with very few candidates showing that they understood the concept of backforecasting. In part (iii) most candidates understood what was generally required, though in some cases there were errors in applying the formulae.

10 (i) $P_{10}(t+h) = (1-\rho h).P_{10}(t) + \sigma.h.P_{11}(t)$

$$\Rightarrow \frac{d}{dt}P_{10}(t) = -\rho.P_{10}(t) + \sigma.P_{11}(t) \qquad (1)$$

and

$$P_{11}(t+h) = \rho.h.P_{10}(t) + (1-\sigma.h).P_{11}(t)$$

$$\Rightarrow \frac{d}{dt}P_{11}(t) = \rho.P_{10}(t) - \sigma.P_{11}(t) \qquad (2)$$

[(2) also follows from the fact that $P_{10}(t) + P_{11}(t) = 1$.]

(ii) $P_{10}(t) + P_{11}(t) = 1$

so from (1) $\frac{d}{dt}P_{10}(t) + (\sigma+\rho).P_{10}(t) = \sigma$

$$\Rightarrow \frac{d}{dt}e^{(\sigma+\rho)t}P_{10}(t) = \sigma.e^{(\sigma+\rho)t} + C$$

$$\Rightarrow P_{10}(t) = \frac{\sigma}{\sigma+\rho}.e^{(\sigma+\rho)t} + C.e^{-(\sigma+\rho)t}$$

$$\Rightarrow P_{10}(t) = \frac{\sigma}{\sigma+\rho}.\left(1-e^{-(\sigma+\rho)t}\right) \text{ since } P_{10}(0) = 0.$$

[Alternatively, instead of solving the DE, just verify that the function proposed as the solution does indeed satisfy the DE and also check that the value is correct at $t = 0$.]

Therefore

$$P_{11}(t) = 1 - \frac{\sigma}{\sigma+\rho}.\left(1-e^{-(\sigma+\rho)t}\right) = \frac{\rho+\sigma e^{-(\sigma+\rho)t}}{\sigma+\rho}.$$

(iii) (a) The generator matrix is now

$$\begin{pmatrix} -\rho & \rho & 0 \\ \sigma & -(\sigma+\rho) & \rho \\ 0 & 2\sigma & -2\sigma \end{pmatrix}$$

(b) Hence the Forward Equations are

$$\frac{d}{dt}p_0(t) = -\rho.p_0(t) + \sigma.p_1(t)$$

$$\frac{d}{dt}p_1(t) = \rho.p_0 - (\sigma+\rho)p_1(t) + 2.\sigma.p_2(t)$$

$$\frac{d}{dt}p_2(t) = \rho.p_1(t) - 2\sigma.p_2(t)$$

(c) Simply substituting in the suggested values gives the required result.

(d) The implication is that the given distribution is stationary. By the standard properties of Markov processes, it follows that it is the equilibrium distribution, so that the long-term probabilities of being in each of the three states are known.

Answers to this question were on the whole disappointing.

Problems in part (i) included transposing the parameters. Marks were available for deriving p_{11} from p_{10} (using the information given in the question) even where p_{10} could not be correctly identified.

Only a small proportion of candidates managed to provide a solution for the differential equation in part (ii).

Better answers were given for the generator matrix in part (iii).

Candidates that gave a generator matrix by and large were able to score marks by applying their matrix to produce the forward equations. Very few candidates provided any conclusions under part (iv).