Faculty of Actuaries **Institute of Actuaries**

EXAMINATIONS

September 2000

Subject 103 — Stochastic Modelling

EXAMINERS' REPORT

1 (i) This is clearly a Markovian birth process. The state space is $\{0, 1, ..., N\}$.

Given that we have m infected and $N - m$ healthy individuals, the number of "dangerous" pair contacts is $m(N - m)$. Thus, the rate $\sigma_{m,m+1} = \lambda p \dfrac{m(N-m)}{P}$.

(ii) The expected total infection time is the sum of the reciprocal rates $\dfrac{P}{\lambda p}\sum_{m=1}^{N-1}\dfrac{1}{m(N-m)}$.

(Since $\dfrac{1}{m(N-m)} = \dfrac{1}{N}\left[\dfrac{1}{m}+\dfrac{1}{N-m}\right]$, this time may also be expressed as $\dfrac{N-1}{\lambda p}\sum_{m=1}^{N-1}\dfrac{1}{m}$.)

2 (i) Let N_{ij} denote the number of minutes when the car was in lane i at the start of the minute and in lane j at the end. The estimate of the transition probability p_{ij} is N_{ij} / N_{i+}, where $N_{i+} = \Sigma_j N_{ij}$.

(ii) The problem here is the alternative hypothesis. It would be possible to test whether the distribution of X_{n+1} conditional on $X_n = i$ was really independent of X_{n-1}. Or one might test, using a standard goodness-of-fit test, whether the distribution of the number of consecutive minutes spent in lane i really was geometrically distributed with parameter determined by i.

3 (i) Since $\{U_n\}$ is Gaussian and stationary, it is determined uniquely by its mean and autocovariance functions. For $k > 0$ we have $\gamma_k = \text{Cov}(U_n , U_{n-k})$ $= \dfrac{\tau^2}{2\theta} e^{-\theta k}$, so that the ACF is $\rho_k = e^{-\theta k}$ and the variance $\gamma_0 = \dfrac{\tau^2}{2\theta}$.

(ii) Compare this with the corresponding values for an AR(1): $\gamma_0 = \dfrac{\sigma^2}{1-\alpha^2}$ and $\rho_k = \alpha^k$ for $k > 0$.

The two are seen to match as long as $\alpha = e^{-\theta}$ and $\sigma^2 = (1 - e^{-2\theta})\,\dfrac{\tau^2}{2\theta}$.

4 (i) If X satisfies $dX_t = Y_t\, dB_t + Z_t\, dt$, then $f(X_t)$ satisfies

$$d[f(X_t)] = f'(X_t)\; Y_t\; dB_t + \{\, f'(X_t)\; Z_t + \tfrac{1}{2}\, f''(X_t)\; Y_t^2 \}\, dt.$$

(ii) $(x^4)' = 4x^3$, $(x^4)'' = 12x^2$.

$$d(B_t^4) \;=\; 4B_t^3\;\; dB_t + \tfrac{1}{2}\;.\; 12B_t^2 dt \;=\; 4B_t^3\;\; dB_t + \; 6B_t^2 dt.$$

(iii) $\int_0^t d(B_s^4) \;=\; 4\int_0^t\; B_s^3\;\, dB_s + 6\int_0^t\; B_s^2\;\, ds.$

Therefore $\int_0^t\; B_s^3\;\, dB_s \;=\; \tfrac{1}{4} B_t^4 - \tfrac{3}{2}\int_0^t\; B_s^2\;\, ds.$

5 (i) $P = \dfrac{1}{4}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 0 & 2 \\ 3 & 1 & 0 \end{pmatrix}, \quad P^2 = \dfrac{1}{16}\begin{pmatrix} 8 & 3 & 5 \\ 8 & 6 & 2 \\ 5 & 6 & 5 \end{pmatrix}, \quad P^3 = \dfrac{1}{64}\begin{pmatrix} 29 & 21 & 14 \\ 26 & 18 & 20 \\ 32 & 15 & 17 \end{pmatrix}.$

(ii) (a) $P_{13}^3 \;=\; \dfrac{14}{64} \;=\; \dfrac{7}{32} \;= 0.21875.$

(b) $P_{23}^2 \;=\; \dfrac{2}{16} \;=\; \dfrac{1}{8} \;= 0.125.$

(c) $\dfrac{14}{31} P_{13}^3 + \dfrac{9}{31} P_{23}^3 + \dfrac{8}{31} P_{33}^3 \;=\; \dfrac{14 \times 14 + 9 \times 20 + 8 \times 17}{31 \times 64} \;=\; \dfrac{512}{31 \times 64} \;=\; \dfrac{8}{31}.$

(iii) By time n = 300 the effects of the starting point have worn off. The answer is therefore indistinguishable from the stationary probability π_3 in all three cases.

It is easily observed that the distribution in (c) is stationary, so that $\pi_3 = \dfrac{8}{31}$.

6 (i) The daily change in value of a share is generally on a scale consistent with the value of the share: this tends to indicate that model II is preferable.

(ii) (a) Model II does appear to fit better than model I; the S dataset does indeed exhibit large variations when it is at a high level, and smaller ones when low.

However, the fit does not appear all that good, as Brownian increments are normally distributed, so are seldom as large as some of the jumps which appear in this dataset.

(b) If the Brownian model were accurate, the day-to-day increments $s_n - s_{n-1}$ should be independently normally distributed with constant mean and variance:

a test of normality (Anderson-Darling, Kolmogorov-Smirnov, χ^2 goodness-of-fit) would do fine; a test of independence (based on sample ACF, or the Durbin-Watson statistic) would also be a good suggestion.

(iii) (a) A Lévy process is the sum of three independent components: a deterministic part of the form $\mu + \alpha t$, a Brownian part of the form σB_t and a pure jump part which may be thought of as a compound Poisson process.

(b) One problem would be in estimating the distribution of the jump sizes, particularly with only 250 observations. Even if a family were assumed for the distribution (e.g. double exponential), there would be the additional difficulty that small jumps would not be detectable against the background of the Gaussian noise.

7 (i) (a) First the u_k need to be transformed so that their distribution is something suitable for the white noise sequence of a time series, since at the very least the mean of the sequence needs to be zero. $N(0, \sigma_e^2)$ is the standard choice: one method of achieving this is to define, for each integer t,

$$e_{2t} = \sigma_e \sqrt{-2 \log u_{2t}} \sin(2\pi u_{2t+1})$$

$$e_{2t+1} = \sigma_e \sqrt{-2 \log u_{2t}} \cos(2\pi u_{2t+1}),$$

but there are others, such as the polar method, inverse transform method or acceptance-rejection sampling.

The values of the e_t can now be fed into the formula to give the values of the X_t, whichever model is in use.

(b) The ability to re-use a pseudo-random number sequence is important when comparing the ability of different mechanisms to control a process which is affected by randomness: in order to ensure fair comparison of the mechanisms, the must be subjected to the same degree of "random" input.

(ii) The models do not possess the correct correlation structure.

(iii) (a) $\rho_1 = \text{Corr}(X_t, X_{t-1}) = \alpha_1\text{Corr}(X_{t-1}, X_{t-1}) + \alpha_2\text{Corr}(X_{t-2}, X_{t-1}) = \alpha_1 + \alpha_2\rho_1$.

Hence $\rho_1 = \alpha_1 / (1 - \alpha_2)$

$\rho_2 = \text{Corr}(X_t, X_{t-2}) = \alpha_1\text{Corr}(X_{t-1}, X_{t-2}) + \alpha_2\text{Corr}(X_{t-2}, X_{t-2}) = \alpha_1\rho_1 + \alpha_2$.

(b) We have $0.7 = \rho_1 = \alpha_1 / (1 - \alpha_2)$

and $0.5 = \rho_2 = \alpha_2 + \alpha_1^2 / (1 - \alpha_2) = \alpha_2 + 0.7\alpha_1$. Two equations in two unknowns. Solution: $\alpha_1 = \dfrac{35}{51}$, $\alpha_2 = \dfrac{1}{51}$.

(2 marks for the observation that $\alpha_1 = 0.7$ and $\alpha_2 = 0$ is very close to giving the right answer, as it gives $\rho_2 = 0.49$.)

8 (i) $P'_{0,0}(t) = \mu P_{0,1}(t) - \lambda P_{0,0}(t)$, or a more general form such as $P'_{0,0}(t) = \Sigma P_{0,k}(t)\sigma_{k,0}$

(ii) Since $P_{0,1}(t) = 1 - P_{0,0}(t)$,

we have $P'_{0,0}(t) = \mu(1 - P_{0,0}(t)) - \lambda P_{0,0}(t)$. Any solution method will do, e.g. $\dfrac{d}{dt}[e^{(\lambda+\mu)t} P_{0,0}(t)] = \mu e^{(\lambda+\mu)t}$, solved by $P_{0,0}(t) = \dfrac{\mu}{\lambda+\mu} + Ce^{-(\lambda+\mu)t}$, with C being determined by the fact that $P_{0,0}(0) = 1$.

(iii) $\mathbf{E}_0 O_t = \mathbf{E}_0 \int_0^t I_s\, ds = \int_0^t \mathbf{E}_0\, I_s\, ds = \int_0^t P_{0,0}(s) ds$

$$= \frac{\mu}{\lambda+\mu}t + \frac{\lambda}{(\lambda+\mu)^2}\,(1 - e^{-(\lambda+\mu)t})$$

(iv) Since the process must be in state 0 or state 1 at all times, the solution is just $t - \mathbf{E}_0\, 0_t = \dfrac{\lambda}{\lambda+\mu}t - \dfrac{\lambda}{(\lambda+\mu)^2}\,(1 - e^{-(\lambda+\mu)t})$.

(v) (a) Assuming a member who is initially healthy, expected outgoings (including expenses) by time t and expected income by time t, are respectively

$$\gamma t + \beta\left(\frac{\lambda}{\lambda+\mu}t - \frac{\lambda}{(\lambda+\mu)^2}(1 - e^{-(\lambda+\mu)t})\right)$$

and $$\alpha\left(\frac{\mu}{\lambda+\mu}t + \frac{\lambda}{(\lambda+\mu)^2}(1 - e^{-(\lambda+\mu)t})\right).$$

In the long run, then, as $t \to \infty$, we require $\alpha\mu = \beta\lambda + \gamma(\lambda + \mu)$ to break even.

(b) The assumptions required are that the rate of becoming ill and rate of recovery from illness are constant.

(c) This will certainly not be true of any individual member but, if membership is large and the age and health profiles of the members are constant by virtue of a constant influx of new members, it may be a reasonable approximation.

9 (i) If V_t is a martingale, then its expectation must be constant and equal to its initial value e^{uy}.

Therefore $\mathbf{E}e^{u(y+\mu t+\alpha B_t)-c(u)t} = e^{uy+(u\mu+u^2\alpha^2/2-c(u))t} = e^{uy}$.

Thus we must have $c(u) = u\mu + u^2\alpha^2/2$.

(ii) The optional stopping theorem states that if M_t is a martingale, and T is a random stopping time, then under some additional technical conditions (such as $M_{t\wedge T}$ being uniformly bounded) we have:

$$\mathbf{E}M_T = M_0.$$

It is frequently used to evaluate the expectation of a function of T, such as the moment generating function (as in this instance).

(iii) Applying the optional stopping theorem to the martingale V_t we find that

$$\mathbf{E}V_{T_a} = e^{uy} = \mathbf{E}e^{ua-c(u)T_a}.$$

The equation $c(u) = v$ has two roots u_+ , u_-, one being negative and the other positive (since v is positive).

Now $V_{t\wedge T_a} = e^{u(v)Y_t - vt}$ and $Y_t \geq a$ for $0 \leq t \leq T_a$. If $u(v) < 0$, then $0 < V_{t\wedge T_a} \leq e^{u(v)a}$ for all t, so that the technical condition is satisfied; the same cannot be said if $u(v) > 0$.

Therefore the positive root is unacceptable and $f(y, v) = \mathbf{E}_y e^{-vT_a} = e^{u_-(v)(y-a)}$.

Comment: For the record, there were two very slight errors in this question, both appearing as subscripts. In line 4, first formula: $T_{\{\alpha\}}$ should have read $T_{\{a\}}$, and in part (iii) line 1: $T_{\{u\}}$ should have read $T_{\{a\}}$. This was taken into account by the markers, and the examiners ensured that no marks were lost by students because of either small error.

10 (i) $P_{AA}(s, t) = e^{-\int_s^t \mu x dx} = e^{-\mu(t^2 - s^2)/2}$

(ii) $\mathbf{P}[R_s \geq w] = P_{AA}(s, s+w) = e^{-\mu((s+w)^2 - s^2)/2} = e^{-\mu s w - \mu w^2/2}$. Therefore

$$\mathbf{E}[R_s] = \int_0^\infty e^{-\mu s w - \mu w^2/2} dw.$$

Complete the square at the exponent to get

$$\mathbf{E}[R_s] = e^{\mu s^2/2} \int_0^\infty e^{-\mu(s+w)^2/2} dw = e^{\mu s^2/2} \int_{s\sqrt{\mu}}^\infty e^{-x^2/2} \frac{dx}{\sqrt{\mu}} = \frac{1}{\sqrt{\mu}} \frac{1 - G(s\sqrt{\mu})}{g(s\sqrt{\mu})}.$$

(iii) From (ii) and the given bound

$$\mathbf{E}[R_s] \leq \frac{1}{\sqrt{\mu}} \frac{1}{s\sqrt{\mu}} = \frac{1}{s\mu}$$

$$\mathbf{E}[R_s] \geq \frac{1}{\sqrt{\mu}} \left(\frac{1}{s\sqrt{\mu}} - \frac{1}{s^3 \mu^{3/2}} \right) = \frac{1}{s\mu} - \frac{1}{s^3 \mu^2}.$$

The first inequality yields

$$\mu \leq \frac{1}{s\mathbf{E}[R_s]} = \frac{1}{420} = 0.00238 \text{ . (year)}^{-2}.$$

The second inequality can be written as

$$\mu^2 \mathbf{E}[R_s] - \frac{\mu}{s} + \frac{1}{s^3} \geq 0,$$

so μ must lie **outside** the interval:

$$\frac{\frac{1}{s} \pm \sqrt{\frac{1}{s^2} - \frac{4\mathbf{E}[R_s]}{s^3}}}{2\mathbf{E}[R_s]} = \frac{1 \pm \sqrt{1 - \frac{4\mathbf{E}[R_s]}{s}}}{2s\mathbf{E}[R_s]} = \frac{1 \pm \sqrt{1 - \frac{24}{70}}}{2 \times 6 \times 70} = [0.00023,\ 0.00215]$$

In fact, since clearly $\mu \simeq \dfrac{1}{s\mathbf{E}[R_s]}$ we see that μ must lie in the interval $[0.00215,\ 0.00238]$.

(iv) Use the inverse transform method, $X = F^{-1}(U)$.

In this case $1 - F(x) = \exp(-\int_{70}^x [a + bt]\, dt) = \exp\{-a(x - 70) - ½b(x^2 - 70^2)\}$.

Therefore $\tfrac{1}{2}bx^2 + ax = -\log(1 - F(x)) + \tfrac{1}{2}70^2 b + 70a$. Replace $F(x)$ by u to get

$$x = F^{-1}(u) = \frac{-a + \sqrt{a^2 + 2b[-\log(1-u) + \tfrac{1}{2}70^2 b + 70a]}}{b}$$

$$= -r + \sqrt{70^2 + 140r + r^2 - 2b^{-1}\log(1-u)},$$

where $r = ab^{-1}$. If u is an observation of a uniform pseudo-random variate, then x is an observation from the required distribution.

11 (i) Consumer prices *do* tend to exhibit regular seasonal variation, though not a great deal these days. And, since prices tend to go up rather more than they come down, it is probably worth including a trend term in any model. It is certainly possible to test whether the trend term is equal to zero.

(ii) (a) $X_{n+1} - x_n = \alpha(x_n - x_{n-1}) + e_{n+1} + \beta e_n$.

(b) The parameters are α, β and σ_e^2. The trend removal process would have accounted for any μ parameter.

(iii) $\hat{x}_n(1) = \mathbf{E}(X_{n+1} \mid x_n, \ldots, x_1) = x_n + \alpha(x_n - x_{n-1}) + \mathbf{E}(e_{n+1} + \beta e_n \mid x_n, \ldots, x_1)$. Now e_{n+1} has mean 0 and is conventionally supposed independent of everything that happens before n.

On the other hand, e_n can be deduced from past data, e.g. $e_n = x_n - x_{n-1} - \alpha(x_{n-1} - x_{n-2}) - \beta e_{n-1}$, which may be iterated back to get e_n in terms of the known x and the known e_0.

Thus

$$\hat{x}_n(1) = x_n + \alpha(x_n - x_{n-1}) + \beta e_n .$$

Similarly,

$$\hat{x}_n(2) = \mathbf{E}(X_{n+2} \mid \mathbf{F}_n) = \mathbf{E}(X_{n+1} + \alpha(X_{n+1} - x_n) + e_{n+2} + \beta e_{n+1} \mid \mathbf{F}_n)$$

$$= (1 + \alpha)\,\hat{x}_n(1) - \alpha x_n .$$

We see that $X_{n+1} - \hat{x}_n(1) = e_{n+1}$, so that the prediction variance is just $\mathrm{Var}(e_{n+1}) = \sigma_e^2$.

(iv) Since $e_n = x_n - \hat{x}_{n-1}(1)$, we have

$$\hat{x}_n(1) = x_n + \alpha(x_n - x_{n-1}) + \beta(x_n - \hat{x}_{n-1}(1));$$

if we set $\alpha = 0$ and $\beta = -\xi$, the equation is identical to the updating equation for exponential smoothing.

(v) An ARIMA(p, d, q) model is $I(d)$; in this case, x is $I(1)$.

A stationary ($I(0)$) model has an equilibrium distribution: the distribution of the forecast of X_{n+k} would converge to equilibrium for large k. An $I(1)$ process is the partial sum of an $I(0)$ process, so would have increasing variance, even if the mean happened to be stable.

(vi) Two series $\{x\}$ and $\{y\}$ are cointegrated if both are $I(1)$ but there are some constants a and b such that $\{ax + by\}$ is stationary.

Two processes are likely to be cointegrated if one drives the other, or if both are driven by the same underlying process. In the given instance the suggestion is certainly worth investigating.