THE ROYAL STATISTICAL SOCIETY

2009 EXAMINATIONS – SOLUTIONS

GRADUATE DIPLOMA

MODULAR FORMAT

MODULE 3

STOCHASTIC PROCESSES AND TIME SERIES

The Society provides these solutions to assist candidates preparing for the examinations in future years and for the information of any other persons using the examinations.

The solutions should NOT be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids.

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

Note. In accordance with the convention used in the Society's examination papers, the notation log denotes logarithm to base *e*. Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Graduate Diploma, Module 3, 2009. Question 1

(i) $G(z) = \sum_{i=0}^{\infty} p_i z^i$.

(ii) $G_n(z) = E\left(z^{X_n}\right) = \sum_{i=0}^{\infty} p_i E\left(z^{X_n} \mid X_1 = i\right) = \sum_{i=0}^{\infty} p_i \left[G_{n-1}(z)\right]^i = G\left(G_{n-1}(z)\right)$.

(iii) $\theta_n = P(X_n = 0) = G_n(0)$. Setting $z = 0$ in the relationship of part (ii), we obtain $\theta_n = G(\theta_{n-1})$ $(n \geq 2)$.

(iv) Letting $n \to \infty$ in the result of part (iii), and noting that G is a continuous function of z so that $G(\theta_{n-1}) \to G(\theta)$ as $n \to \infty$, we obtain the equation $\theta = G(\theta)$.

We now have the special case as identified in the question.

(v) In this special case, quoting the standard result for a binomial distribution, $G(z) = (1 - p + pz)^2$.

(vi) θ_1 is simply the zero term of the binomial distribution, so $\theta_1 = (1 - p)^2$. Equivalently, $\theta_1 = G_1(0) = G(0) = (1 - p)^2$.

(vii) $\theta_2 = G\left(\theta_1\right) = \left[1 - p + p(1-p)^2\right]^2 = \left(1-p\right)^2\left[1 + p(1-p)\right]^2$

$= \left(1-p\right)^2\left(1 + p - p^2\right)^2$, as required.

(viii) θ is the smallest positive root of the equation $\theta = G(\theta)$, i.e. of $\theta = (1 - p + p\theta)^2$, so we must solve

$$p^2\theta^2 + (2p - 2p^2 - 1)\theta + (1 - p)^2 = 0.$$

Because $\theta = 1$ is necessarily a root of the equation $\theta = G(\theta)$, it is easy to factorize the quadratic to give

$$(\theta - 1)[p^2\theta - (1 - p)^2] = 0.$$

It follows that the extinction probability is given by $\min\{1, [(1 - p)/p]^2\}$.

Hence the extinction probability is 1 if $p \leq ½$ and is $[(1 - p)/p]^2$ if $p > ½$.

Graduate Diploma, Module 3, 2009. Question 2

(i) A Markov chain is said to be irreducible if it is possible, with non-zero probability, to move from any state in the state space to any other state. A chain is said to be recurrent if, starting from any state in the space, the probability of eventually returning to that state is 1. [These explanations may be put more formally in terms of n-step transition probabilities.]

In the present case, because all the transition probabilities are non-zero, it is clearly possible to move from any state to any other state in one step, so the chain is irreducible. It is a general result that all finite irreducible Markov chains are recurrent.

(ii) The stationary distribution (π_1, π_2, π_3) is given by the solution of the equations

$$(2/5)\pi_1 + (1/5)\pi_2 + (1/5)\pi_3 = \pi_1$$
$$(2/5)\pi_1 + (3/5)\pi_2 + (2/5)\pi_3 = \pi_2$$
$$(1/5)\pi_1 + (1/5)\pi_2 + (2/5)\pi_3 = \pi_3 ,$$

which reduce to

$$3\pi_1 = \pi_2 + \pi_3$$
$$\pi_2 = \pi_1 + \pi_3$$
$$3\pi_3 = \pi_1 + \pi_2 ,$$

together with the normalisation condition $\pi_1 + \pi_2 + \pi_3 = 1$.

It readily follows that the solution is $(\pi_1, \pi_2, \pi_3) = (¼, ½ , ¼)$.

(iii) The probabilities and hence also the proportions in the second generation are given by the terms of the matrix product

$$(2/5 \quad 2/5 \quad 1/5)\begin{pmatrix} 2/5 & 2/5 & 1/5 \\ 1/5 & 3/5 & 1/5 \\ 1/5 & 2/5 & 2/5 \end{pmatrix}$$

which gives the vector of probabilities/proportions (7/25 12/25 6/25).

(iv) The approximate proportions that we would expect to find are the ones given by the stationary distribution $(\pi_1 \quad \pi_2 \quad \pi_3) = (1/4 \quad 1/2 \quad 1/4)$.

The reasoning lying behind this is as follows. Let $p_{ij}^{(n)}$ represent the n-step transition probability from state i to state j; then, for all i and j, $p_{ij}^{(n)} \to \pi_j$ as $n \to \infty$. Hence, after a large number, n, of generations, we would expect that $p_{ij}^{(n)} \approx \pi_j$. In a large population of individuals, each of whom has the same approximate probability π_j of being in state j, π_j is also the approximate proportion of the population who are in state j.

Graduate Diploma, Module 3, 2009. Question 3

(i) Because of the memory-less property of the exponential distribution, how long line 2 has been under repair is statistically independent of how much longer it will take to repair it.

(ii) Define states as follows.

0: no line runs
1: line 1 runs, line 2 under repair
2: line 1 under repair, line 2 runs
3: both lines run.

The instantaneous transition rates are as follows.

transition	rate
$0 \to 1$	1/2
$1 \to 0$	1/10
$1 \to 3$	1/2
$2 \to 0$	1/5
$2 \to 3$	1/2
$3 \to 2$	1/30
$3 \to 1$	1/15

(iii) The equilibrium equations are

$$\begin{aligned}(1/2)\pi_0 &= (1/10)\pi_1 + (1/5)\pi_2\\ (3/5)\pi_1 &= (1/2)\pi_0 + (1/15)\pi_3\\ (7/10)\pi_2 &= (1/30)\pi_3\\ (1/10)\pi_3 &= (1/2)\pi_1 + (1/2)\pi_2 .\end{aligned}$$

These reduce to

$$\begin{aligned}25\pi_0 &= 26\pi_2\\ 5\pi_1 &= 16\pi_2\\ \pi_3 &= 21\pi_2 .\end{aligned}$$

Using the normalisation condition $\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$, it follows that

$$(\pi_0,\ \pi_1,\ \pi_2,\ \pi_3) = k(26/25,\ 16/5,\ 1,\ 21),$$

where $1/k = (26/25) + (16/5) + 1 + 21 = 656/25$.

So $(\pi_0, \pi_1, \pi_2, \pi_3) = (13/328, 5/41, 25/656, 525/656) = (0.04, 0.12, 0.04, 0.80)$.

In particular the long-term proportion of time that the factory is unable to meet the production target is $\pi_0 = 0.04$.

Graduate Diploma, Module 3, 2009. Question 4

(i) The state space is the set of all non-negative integers. The instantaneous transition rates are as follows.

transition	rate	
$i \to i+1$	λ	$(i \geq 0)$
$i \to i-1$	μ	$(i \geq 1)$

(ii) The traffic intensity is defined by $\rho = \lambda/\mu$. A necessary and sufficient condition for an equilibrium distribution to exist is $\rho < 1$, i.e. $\lambda < \mu$.

(iii) The detailed balance equations are $\lambda\pi_{n-1} = \mu\pi_n \quad (n \geq 1)$.

Thus $\pi_n = \rho\pi_{n-1}$ $(n \geq 1)$ and, using this relation recursively, we find $\pi_n = \rho^n\pi_0$ $(n \geq 0)$. Using the normalisation condition $\Sigma\pi_n = 1$, we find, using the formula for the sum of a geometric series (or observing that we are dealing with a geometric distribution), that $\pi_n = (1-\rho)\rho^n$ $(n \geq 0)$, as required.

(iv) The service time distribution, i.e. here the waiting time distribution, for this model is exponential with parameter μ. The pdf is $\mu e^{-\mu t}$ $(t \geq 0)$.

(v) The arriving customer's waiting time is the sum of $n+1$ independently and identically distributed service times, each having an exponential distribution with parameter μ. These are the service times of the customers ahead of him in the queue plus his own (note that, because of the memory-less property of the exponential distribution, the residual service time of the customer being served at the time of arrival of the new customer is also exponential with parameter μ).

Using the note given in the question, the required pdf is $\mu^{n+1}t^n e^{-\mu t}/n!$ $(t \geq 0)$.

(vi) In equilibrium, from part (iii) the probability that an arriving customer finds n customers ahead of him in the queue is given by $\pi_n = (1-\rho)\rho^n$ $(n \geq 0)$. Thus, using the result of part (v), the pdf of his waiting time is given by

$$\sum_{n=0}^{\infty}(1-\rho)\rho^n\mu^{n+1}t^n e^{-\mu t}/n! = (1-\rho)\mu e^{-\mu t}\sum_{n=0}^{\infty}(\rho\mu t)^n/n!$$

$$= (\mu-\lambda)e^{-\mu t}\sum_{n=0}^{\infty}(\lambda t)^n/n! = (\mu-\lambda)e^{-(\mu-\lambda)t},$$

which is the pdf of the exponential distribution with parameter $\mu - \lambda$.

Graduate Diploma, Module 3, 2009. Question 5

(i) The autoregressive characteristic equation is $1 - (3/4)z + (1/8)z^2 = 0$, which has roots $z = 2, 4$. Both the roots are greater than one in modulus, so the stationarity condition is satisfied.

(ii) On substituting, we obtain

$$\sum_{i=0}^{\infty}\psi_i\varepsilon_{t-i} = \frac{3}{4}\sum_{i=0}^{\infty}\psi_i\varepsilon_{t-1-i} - \frac{1}{8}\sum_{i=0}^{\infty}\psi_i\varepsilon_{t-2-i} + \varepsilon_t = \frac{3}{4}\sum_{i=1}^{\infty}\psi_{i-1}\varepsilon_{t-i} - \frac{1}{8}\sum_{i=2}^{\infty}\psi_{i-2}\varepsilon_{t-i} + \varepsilon_t .$$

Equating coefficients of the ε_{t-i}, we obtain the following.

$i = 0$: $\psi_0 = 1$

$i = 1$: $\psi_1 = (3/4)\psi_0 = 3/4$

$i \geq 2$: $\psi_i = (3/4)\psi_{i-1} - (1/8)\psi_{i-2}$.

The last of these provides the required set of recurrence relations, for $i \geq 2$, and the values for ψ_0 and ψ_1 provide the required initial conditions.

(iii) The general solution is of the form $\psi_i = A_1\alpha_1^{\,i} + A_2\alpha_2^{\,i}$ $(i \geq 0)$, where A_1 and A_2 are arbitrary constants and α_1 and α_2 are the roots of the auxiliary equation

$$\alpha^2 = (3/4)\alpha - (1/8).$$

The roots of the auxiliary equation are [the inverses of the roots of the characteristic equation of part (i)] 1/2 and 1/4. Hence the general solution is $\psi_i = A_1(1/2)^i + A_2(1/4)^i$. Using the initial conditions,

$$A_1 + A_2 = 1 \quad \text{and} \quad (1/2)A_1 + (1/4)A_2 = 3/4.$$

Hence $A_1 = 2$ and $A_2 = -1$, and the solution for the ψ_i is as stated in the question.

(iv) Generally, $\mathrm{Var}(Y_t) = \sum_{i=0}^{\infty}\psi_i^2\sigma^2$. In the present case, this gives

$$\mathrm{Var}(Y_t) = \sum_{i=0}^{\infty}\left[2(1/2)^i - (1/4)^i\right]^2\sigma^2 = \sum_{i=0}^{\infty}\left[4(1/4)^i - 4(1/8)^i + (1/16)^i\right]\sigma^2$$

Summing the geometric series in this expression gives

$$\mathrm{Var}(Y_t) = \left(\frac{4}{1-(1/4)} - \frac{4}{1-(1/8)} + \frac{1}{1-(1/16)}\right)\sigma^2 = \frac{64}{35}\sigma^2 .$$

Graduate Diploma, Module 3, 2009. Question 6

(i) If the underlying trend is an exponential one, taking logarithms will transform the trend to a linear one, in which case an ARIMA model is likely to provide a better fit.

If the variability of the series and, in particular, of any seasonal effects increases with increase in the underlying level of the series, taking logarithms will tend to stabilise the variation, and in this case also an ARIMA model is likely to provide a better fit.

(ii) Approximate 95% confidence limits are at $\pm 2/\sqrt{180} = \pm 0.149$. So any autocorrelation outside these limits differs significantly from zero at the 5% level. We see that a number of autocorrelations lie well outside these limits, notably at lags 1, 6, 12, 18, 24, 30, 36. This clearly indicates the presence of seasonality of period 12 months and also suggests the presence of trend.

The purpose of taking differences is to eliminate the trend and the purpose of taking seasonal differences is to eliminate the seasonality.

(iii) Approximate 95% confidence limits are at $\pm 2/\sqrt{167} = \pm 0.155$. So any autocorrelation outside these limits differs significantly from zero at the 5% level. Here the only significant autocorrelations are at lag 1 and at lag 12. This shows that any trend and seasonality have been removed by the differencing to obtain a stationary series and suggests that the stationary series may be modelled by moving average terms at lags 1 and 12. So a seasonal ARIMA$(0,1,1)\times(0,1,1)_{12}$ model is suggested.

(iv) A seasonal ARIMA$(0,1,1)\times(0,1,1)_{12}$ has been fitted. The equation of the fitted model is (see the parameter estimates in the computer output in the question)

$$(1 - L)(1 - L^{12})Y_t = (1 - 0.8759L)(1 - 0.7789L^{12})\varepsilon_t,$$

where L is the lag operator (backward shift operator) and $\{\varepsilon_t\}$ is a white noise process, i.e.

$$(1 - L - L^{12} + L^{13})Y_t = (1 - 0.8759L - 0.7789L^{12} + 0.6822L^{13})\varepsilon_t$$

or

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + \varepsilon_t - 0.8759\varepsilon_{t-1} - 0.7789\varepsilon_{t-12} + 0.6822\varepsilon_{t-13}.$$

(v) None of the p-values of the modified Box-Pierce statistics is significant. So the residuals of the fitted model appear to come from a white noise process – our model appears to give a good fit to the data.

(vi) The forecast and 95% prediction interval for Y_{192} are given by 8.67702 and (8.45985, 8.89420) respectively. The forecast sales and prediction interval are given by taking exponentials of these values. This, correct to the nearest 1000 litres, gives 5867 as the forecast sales for December 1995 and (4721, 7290) as the 95% prediction interval.

Graduate Diploma, Module 3, 2009. Question 7

(i) The updating equations are as follows.

$$L_t = \alpha\left(Y_t / I_{t-p}\right) + \left(1-\alpha\right)\left(L_{t-1} + B_{t-1}\right)$$
$$B_t = \gamma\left(L_t - L_{t-1}\right) + \left(1-\gamma\right) B_{t-1}$$
$$I_t = \delta\left(Y_t / L_t\right) + \left(1-\delta\right) I_{t-p}$$

(ii) $\hat{y}_T(h) = (L_T + hB_T) I_{T-p+h.}$

(iii) We require $\hat{y}_T(1)$ and $\hat{y}_T(12)$.

(a) $\hat{y}_T(1) = (311.44 + 1.70)(0.709) = (313.14)\ (0.709) = 222.02$

$= 222$ to nearest whole number.

(b) $\hat{y}_T(12) = [311.44 + (12)(1.70)](0.970) = (4519.06)(0.970) = 321.88$

$= 322$ to nearest whole number.

(iv) For January 1994, the values are as follows.

Level	$L_t = 0.4(245/0.709) + 0.6(311.44 + 1.70) = 326.11$
Trend	$B_t = 0.1(326.11 - 311.44) + 0.9(1.70) = 3.00$
Index	$I_t = 0.01(245/326.11) + 0.99(0.709) = 0.709$
Fitted	From (iii), $\hat{y}_T(1) = 222.02$
Residual	Deaths – Fitted = 245 – 222.02 = 22.98

(v) Given some appropriately chosen initial values for the level and the trend and for the first twelve seasonal indices, for any given set of values of the smoothing constants α, γ and δ (each between 0 and 1 inclusive, of course) the numerical values of all the quantities in the table may be calculated for each month in the series. The sum of squares of the residuals (or some other appropriate function of the residuals) may be used as a measure of how well the Holt-Winters method with the chosen values of α, γ and δ performs. By looking at a grid of values of α, γ and δ or by carrying out a formal optimisation, the values that minimise the sum of squares of the residuals may be found as the best set of values to use.

Graduate Diploma, Module 3, 2009. Question 8

(i) $\{Y_t\}$ is an ARIMA(0,2,1) process.

(ii) $Y_t = 2Y_{t-1} - Y_{t-2} + \varepsilon_t - \theta\varepsilon_{t-1}$.

(iii) $\hat{y}_T(1) = E(Y_{T+1} \mid H_T) = E(2Y_T - Y_{T-1} + \varepsilon_{T+1} - \theta\varepsilon_T \mid H_T) = 2Y_T - Y_{T-1} - \theta\varepsilon_T$.

[Note that these are *conditional* expectations given the *entire history* of the process up to and including time T. So Y_T and Y_{T-1} are known. Further, ε_T (sometimes called the "innovation" at time T) can be found using the one-step-ahead prediction at time $T-1$ and the observed Y_T – this is as shown in part (iv), replacing $T+1$ by T. ε_{T+1} cannot be found, of course, and so has expectation 0 as a white noise term.]

(iv) $Y_{T+1} - \hat{y}_T(1) = \varepsilon_{T+1}$.

(v) $\hat{y}_T(2) = E(Y_{T+2} \mid H_T) = E(2Y_{T+1} - Y_T + \varepsilon_{T+2} - \theta\varepsilon_{T+1} \mid H_T) = 2\hat{y}_T(1) - Y_T$

$= 3Y_T - 2Y_{T-1} - 2\theta\varepsilon_T$ (substituting from the result of part (iii)).

(vi) For $h \geq 3$, setting $t = T + h$ in the model equation,

$$\hat{y}_T(h) = E(Y_{T+h} \mid H_T) = E(2Y_{T+h-1} - Y_{T+h-2} + \varepsilon_{T+h} - \theta\varepsilon_{T+h-1} \mid H_T)$$

$$= 2E(Y_{T+h-1} \mid H_T) - E(Y_{T+h-2} \mid H_T) + 0 - 0 = 2\hat{y}_T(h-1) - \hat{y}_T(h-2).$$

(vii) The general form of the solution of the difference equation of part (vi) (in the examination, this could be quoted or easily found) is $\hat{y}_T(h) = A + Bh \quad (h \geq 1)$. To determine A and B, the initial conditions of parts (iii) and (v) give

$$A + B = 2Y_T - Y_{T-1} - \theta\varepsilon_T,$$

$$A + 2B = 3Y_T - 2Y_{T-1} - 2\theta\varepsilon_T.$$

Hence $B = b_T = Y_T - Y_{T-1} - \theta\varepsilon_T$ and $A = Y_T$.

(viii) Using the result of part (iv), replacing T by $T-1$, we have $Y_T - \hat{y}_{T-1}(1) = \varepsilon_T$.

Note also that $\hat{y}_{T-1}(1) = Y_{T-1} + b_{T-1}$.

Substituting into the expression for b_T in part (vii),

$$b_T = Y_T - Y_{T-1} - \theta(Y_T - \hat{y}_{T-1}(1)) = Y_T - Y_{T-1} - \theta(Y_T - Y_{T-1} - b_{T-1})$$

$$= (1-\theta)(Y_T - Y_{T-1}) + \theta b_{T-1}.$$