EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2015

MODULE 3 : Stochastic processes and time series

Time allowed: Three hours

Candidates should answer **FIVE** *questions.*

All questions carry equal marks.
The number of marks allotted for each part-question is shown in brackets.

Graph paper and Official tables are provided.

Candidates may use calculators in accordance with the regulations published in the Society's "Guide to Examinations" (document Ex1).

The notation log *denotes logarithm to base* ***e***.
Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Note also that $\binom{n}{r}$ *is the same as* ${}^nC_r$.

This examination paper consists of 12 printed pages.
This front cover is page 1.
Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1. (a) In the context of a time-homogeneous Markov chain with states i = 0, 1, 2, ..., define the n-step transition probability $p_{ij}^{(n)}$. Derive the Chapman-Kolmogorov equations for the $p_{ij}^{(n)}$ indicating clearly at which step(s) you make use of (i) time-homogeneity, (ii) the Markov property.

(5)

(b) A software development company is conducting competency challenges, comprising a sequence of independent debugging exercises. Each programmer must attempt the exercises successively until awarded a Certificate of Programming Competency, for which either a perfect score on one exercise or a score of at least 90% on two consecutive exercises is required.

This process can be modelled as a Markov chain with three states which depend on the most recent results as follows.

0 : score under 90% on most recent attempt

1 : score between 90% and 99% on most recent attempt, but under 90% on the previous attempt (if one exists)

2 : obtain Certificate

The one-step transition matrix, **P**, can be written in terms of parameters θ and ϕ as

$$\begin{pmatrix} 1-\theta & * & \phi \\ 1-\theta & 0 & \theta \\ 0 & 0 & 1 \end{pmatrix}$$

where $0 < \phi < \theta < 1$.

(i) Write down the interpretations of the parameters θ and ϕ in **P**. Find the missing entry for p_{01} indicated by $*$ and explain the values of p_{11} and p_{22}.

(4)

(ii) For i = 0 or 1, let τ_i be the expected number of transitions needed to reach state 2 given that the chain is currently in state i. By conditioning on the outcome of the next observation, write down a pair of backward equations for τ_0 and τ_1. (4)

(iii) Show that

$$\tau_0 = \frac{1+\theta-\phi}{\theta^2+\phi(1-\theta)}. \qquad (2)$$

(iv) In each exercise, only 5% of programmers obtain 100% and 75% obtain under 90%.

Two programmers score respectively 83% and 96% on their first attempt. Calculate the difference, if any, between the expected number of further attempts each requires to gain the Certificate of Programming Competency.

(5)

2. Two players, Amy and Betsy, compete in a simple game which consists of a sequence of rounds whose results are independent of each other and which continues until one or other of the players loses all her capital, i.e. is ruined. At each play, £1 is transferred from the loser to the winner; Amy has probability γ of winning, where $0 < \gamma < 1$, while Betsy has probability $1 - \gamma$ of winning. Suppose that Amy starts with capital £a and Betsy starts with capital £b, and let the total capital in pounds of both players combined be $T = a + b$.

(i) Let X_A denote Amy's current capital and p_i denote the probability that Amy is eventually ruined, given that $X_A = i$, where $0 \le i \le T$.

(a) Show that the p_i satisfy the difference equation

$$p_i = \gamma p_{i+1} + (1-\gamma)\, p_{i-1} \quad \text{for } i = 1, \ldots, T - 1.$$

State the values of p_0 and p_T, justifying your answers.

(6)

(b) Assuming that $\gamma \neq \frac{1}{2}$, verify that the difference equation in part (a) is satisfied by $p_i = A + B\Phi^i$, where $\Phi = \dfrac{1-\gamma}{\gamma}$.

(3)

(c) Hence solve the difference equation of part (a), assuming that $\gamma \neq \frac{1}{2}$, and deduce that at the start of the game the probability that Amy is eventually ruined is given by

$$\frac{\Phi^T - \Phi^a}{\Phi^T - 1}.$$

(5)

(ii) Suppose instead that Amy starts with capital £a but Betsy has unlimited resources. Find the probability that Amy is eventually ruined, distinguishing between the two cases $0 < \gamma < 0.5$ and $0.5 < \gamma < 1$.

(6)

3. (i) Write down the distribution of $N(t)$, the number of events in time t in a Poisson process of rate λ per unit time. Deduce the cumulative distribution function of T_1, the time to the first event in the process. Identify the distribution of T_1.

(3)

(ii) For $s < t$, write down the joint probability that one event occurs in the interval $[0, s]$ and none in $(s, t]$. Hence show that $P(T_1 < s \mid N(t) = 1) = \frac{s}{t}$, for $s < t$, and deduce that $T_1 \mid N(t) = 1$ is uniformly distributed on $[0, t]$. Write down $E[N(t)]$ and $E[T_1 \mid N(t) = 1]$.

(7)

You are now **given** the result that, under the condition that n events occur in time t, the times $S_1, S_2, \ldots, S_n$ at which these n events occur are jointly distributed as independent uniform random variables on $[0, t]$.

In a factory manufacturing tablet computers, assembled items arrive at the packaging point as a Poisson process of rate λ per unit time. At a fixed time τ after the previous dispatch time, all packaged items are dispatched from the factory.

(iii) Explain why the expected total waiting time of all the items dispatched at time τ is $\frac{1}{2}\lambda\tau^2$.

(3)

(iv) Investment funds are available to introduce two additional dispatch times at u and v respectively, where $0 < u < v < \tau$.

(a) Show that the expected total wait of all items dispatched at times u, v and τ will now be

$$h(u,v) = \frac{\lambda}{2}\left[u^2 + (v-u)^2 + (\tau - v)^2\right].$$

(2)

(b) Deduce the optimal choice of u and v to minimise the expected total wait and show that it does not depend on λ.

(5)

4. (i) An equilibrium M/G/1 queue has arrival rate λ and a random service time whose mean is $\frac{1}{\mu}$ and variance is σ^2. The expected number left in the system by a departing customer is denoted by $E(N_S)$. Explain why the expected time a customer spends in the system from arrival to departure is given by $\frac{1}{\lambda}E(N_S)$.

(3)

(ii) You are given that

$$E(N_S) = \frac{\rho(2-\rho)+\lambda^2\sigma^2}{2(1-\rho)}, \text{ where } \rho = \frac{\lambda}{\mu} < 1.$$

Aircraft operating a round-the-clock airlift of supplies to a famine-stricken area arrive randomly at average intervals of 4 hours at an airfield which has facilities for unloading and refuelling them one at a time. Calculate an aircraft's expected turnround time from arrival to departure under each of the following models.

(a) Unloading and refuelling takes exactly 3 hours in total for each plane.

(3)

(b) The time for unloading plus refuelling is the sum of two independent times each having mean and standard deviation equal to $1\frac{1}{2}$ hours.

(5)

(c) Unloading and refuelling operations happen simultaneously so that the combined time, X, is the maximum of two independent random variables having exponential distributions with mean $1\frac{1}{2}$ hours. In this case you should first show that the probability density function of X is

$$f(x) = \tfrac{4}{3}e^{-\frac{2}{3}x} - \tfrac{4}{3}e^{-\frac{4}{3}x}, \quad x > 0.$$

(9)

[You may use without proof the result

$$\int_0^\infty x^r e^{-x} dx = r!$$

for $r = 0, 1, 2, \dots$.]

5. (i) Define the lag k autocorrelation ρ_k of a time series. (1)

(ii) The approximate formula

$$SE(r_k) \approx \sqrt{\frac{1+2\sum_{j=1}^{k-1} r_j^2}{n}}$$

is commonly used in the identification of time series models. What is r_k? Under what conditions on ρ_k is the above formula valid?

Describe the use of $SE(r_k)$ in the identification of possible models for a time series of length n. (4)

A time series of 100 wind speed measurements (24 hour averages) has been plotted in the computer output **shown on the next page**. The table below gives the sample autocorrelation function (acf) and sample partial autocorrelation function (pacf) for the first 16 lags.

Lag	acf	pacf	Lag	acf	pacf
1	0.654	0.654	9	–0.126	–0.043
2	0.308	–0.210	10	–0.005	0.193
3	0.105	–0.002	11	0.087	0.026
4	0.086	0.123	12	0.094	–0.114
5	–0.038	–0.254	13	0.030	–0.025
6	–0.179	–0.095	14	–0.043	–0.110
7	–0.242	–0.016	15	–0.059	–0.019
8	–0.177	0.013	16	–0.024	0.092

(iii) Calculate $SE(r_k)$ for lags 1, 2 and 3. What standard error should be used for the pacf? Discuss which models of the ARIMA class it might be appropriate to attempt to fit to these data. (6)

(iv) The output also includes some results obtained by fitting two ARIMA models, A and B, to the wind speed data. The software used fits an ARIMA(p, d, q) model to x_t by first calculating y_t by taking dth differences of x_t, and then fitting the model

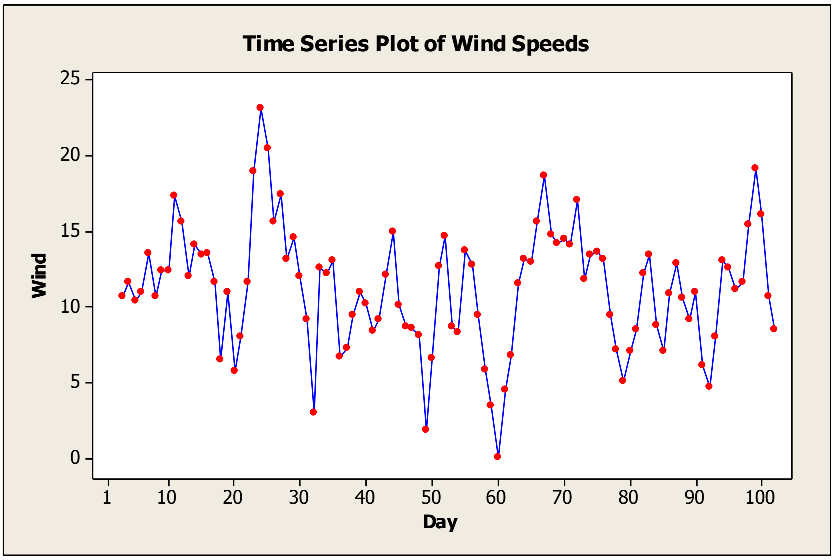
$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \theta_0 + A_t - \sum_{i=1}^{q} \theta_i A_{t-i} \, ,$$

where A_t is white noise.

State which models have been fitted and write down explicitly their model equations. (5)

(v) Discuss what the output tells you about the suitability of each of the models, A and B. (4)

Computer output is on the next page



ARIMA Model A: Wind

Final Estimates of Parameters

```
Type          Coef  SE Coef      T      P
MA   1     -0.8084   0.0897  -9.02  0.000
MA   2     -0.4759   0.0896  -5.31  0.000
Constant   11.1651   0.6649  16.79  0.000
Mean       11.1651   0.6649

Number of observations:  100
Residuals:    SS =  827.243 (backforecasts excluded)
              MS =  8.528  DF = 97

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag             12     24     36     48
Chi-Square    10.1   12.8   30.5   45.3
DF               9     21     33     45
P-Value      0.339  0.915  0.590  0.457
```

ARIMA Model B: Wind

Final Estimates of Parameters

```
Type          Coef  SE Coef      T      P
AR   1     -0.0525   0.0989  -0.53  0.597
AR   2     -0.2091   0.0982  -2.13  0.036
AR   3     -0.2844   0.1008  -2.82  0.006
Constant    0.0133   0.3208   0.04  0.967

Differencing: 1 regular difference
Number of observations:  Original series 100, after differencing 99
Residuals:    SS =  966.857 (backforecasts excluded)
              MS =  10.177  DF = 95

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag             12     24     36     48
Chi-Square    18.4   20.7   45.4   62.7
DF               8     20     32     44
P-Value      0.019  0.416  0.058  0.033
```

6. For daily sales forecasting a forecaster regularly uses the ARIMA(0, 1, 1) model

$$Y_t = Y_{t-1} + A_t - \beta A_{t-1},$$

where $0 < \beta < 1$ and A_t is white noise with variance σ^2.

(i) Show that $\hat{Y}_t(h)$, the minimum mean square error forecast of Y_{t+h} at time t, is constant for all lags $h \geq 1$. Show also that $\hat{Y}_t(1)$ is a function of Y_t and a forecast error a_t. Explain how a_t would be calculated.

(4)

(ii) Show that Y_t can be written in MA(∞) form as

$$Y_t = \sum_{i=0}^{\infty} \psi_i A_{t-i} \text{ with } \psi_0 = 1 \text{ and } \psi_i = 1 - \beta \text{ for } i \geq 1.$$

Deduce that the h-step-ahead forecast error variance is

$$V(h) = \sigma^2\{1 + (h-1)(1-\beta)^2\}, \text{ for } h \geq 1.$$

(7)

(iii) On Monday, in forecasting Tuesday's sales, the forecaster calculated the one-step-ahead (i.e. one-day-ahead) forecast to be 812 with a standard error of 4. At the same time her forecast of Wednesday's sales had a standard error of 5.

(a) Find the value of β used and the variance of the underlying white noise.

(3)

(b) Calculate also the variance of her forecast of the total sales for Tuesday and Wednesday and give a 90% prediction interval for her forecast of this total.

(6)

7. The time series Y_t is second order stationary with mean μ, variance γ_0 and autocorrelation function (acf) ρ_k, for $k = 0, 1, 2, \ldots$.

(i) Define ρ_k and explain what is meant by the phrase *second order stationary*.

(3)

(ii) Find expressions for the mean, variance and acf of $Z_t = Y_t + \omega Y_{t-1}$, where ω is a constant, in terms of μ, γ_0 and the autocorrelations of Y_t. Deduce that Z_t is also second order stationary.

(9)

(iii) Deduce also that the time series $V_t = Y_t - 2Y_{t-1} + Y_{t-2}$ is second order stationary.

(3)

(iv) If the acf of Y_t is α^k, for $k \geq 0$, find the values of ω such that $Y_t + \omega Y_{t-1}$ is white noise.

(5)

8. The spectral density function $f(\omega)$ of a stationary time series model having autocovariance function $\{\gamma_k\}$ (for k = 0, 1, 2, ...) may be written as

$$f(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \gamma_k e^{-i\omega k}, \text{ for } -\pi < \omega < \pi.$$

(i) Show that the normalised spectral density, $g(\omega) = \dfrac{2\pi f(\omega)}{\gamma_0}$, may be written equivalently as

$$g(\omega) = 1 + \sum_{k=1}^{\infty} \rho_k \left(e^{-i\omega k} + e^{i\omega k}\right)$$

where ρ_k is the autocorrelation function of the time series.

(5)

(ii) Deduce that $g(\omega) = g(-\omega)$, for $-\pi < \omega < \pi$.

(2)

(iii) For the time series model $X_t = \alpha X_{t-1} + A_t$ for $|\alpha| < 1$, where A_t is white noise, show that

$$\rho_k = \alpha^k, \text{ for } k = 0, 1, 2, \ldots,$$

and that

$$g(\omega) = \frac{1-\alpha^2}{1-2\alpha\cos\omega + \alpha^2}.$$

(8)

(iv) In the case α = –0.7, sketch $g(\omega)$ for $0 < \omega < \pi$ and compare it with the spectral density function of white noise. Comment on what the shape of $g(\omega)$ tells you about how a time series plot of a realisation of X_t would compare to white noise.

(5)

BLANK PAGE

BLANK PAGE