EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2011

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 10 printed pages, **each printed on one side only**.
This front cover is page 1.
Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1. Let $\{X_n\}$ $(n \geq 0)$ represent a branching process, where X_n denotes the population size in the nth generation. The initial population size is 1, i.e. $X_0 = 1$, and in each generation the number of offspring produced by each individual that survive to the next generation follows the offspring distribution $\{p_i\}$ $(i \geq 0)$ with associated probability generating function $G(z)$. The numbers of surviving offspring produced by different individuals are statistically independent of each other.

Let $G_n(z)$ denote the probability generating function of the number of individuals in the population in the nth generation $(n \geq 1)$.

(i) Define $G(z)$ in terms of the distribution $\{p_i\}$ $(i \geq 0)$.

(1)

(ii) By conditioning on the number of individuals in the first generation, prove that

$$G_n(z) = G(G_{n-1}(z)) \qquad (n \geq 2).$$

(4)

(iii) Let $\theta_n = P(X_n = 0)$ $(n \geq 1)$, the probability that the population has become extinct by the nth generation. Using the relationship of part (ii), find a recurrence relationship for the θ_n.

(2)

(iv) Let $\theta = \lim_{n\to\infty} \theta_n$, the probability of ultimate extinction of the population. From the result of part (iii), deduce that θ satisfies the equation $\theta = G(\theta)$.

(2)

Consider now the special case where each individual produces exactly three offspring, each of which survives to the next generation independently of each other with probability $\frac{1}{2}$. In this case, the offspring distribution is a binomial distribution with parameters 3 and $\frac{1}{2}$.

(v) Find $G(z)$.

(1)

(vi) Find θ_1, the probability that the population becomes extinct at the first generation.

(1)

(vii) Find as a fraction the probability θ_2 that the population has become extinct by the second generation.

(3)

(viii) Show that the probability θ of ultimate extinction of the population is given by

$$\theta = \sqrt{5} - 2.$$

(6)

[**Note.** You may assume that θ is given by the smallest positive root of the equation $\theta = G(\theta)$.]

Turn over

2. Consider a random walk with a reflecting barrier at the origin, namely a Markov chain $\{X_n\}$ ($n \geq 0$) with state space the set of all non-negative integers and transition probabilities $p_{i,j}$ given by

$$p_{0,0} = 1 - \theta$$
$$p_{i,i+1} = \theta \qquad (i \geq 0)$$
$$p_{i,i-1} = 1 - \theta \qquad (i \geq 1)$$

where θ is a parameter that satisfies $0 < \theta < 1$.

(i) Explain what is meant in general by the statement that a Markov chain is *irreducible* and prove that, in the present case, the chain is irreducible.

(5)

(ii) Write down the equations for the equilibrium distribution $\{\pi_j\}$ ($j \geq 0$) of $\{X_n\}$, including the normalisation condition.

(3)

(iii) Investigate the solution of the equations of part (ii), distinguishing between the cases $\theta = \frac{1}{2}$, $\theta < \frac{1}{2}$ and $\theta > \frac{1}{2}$.

Show that an equilibrium distribution exists if and only if $\theta < \frac{1}{2}$ and that in this case

$$\pi_j = \frac{1-2\theta}{1-\theta}\left(\frac{\theta}{1-\theta}\right)^j \qquad (j \geq 0).$$

(12)

Turn over

3. (i) Pests immigrate into a habitat according to a Poisson process with rate λ. What is the distribution of the number of pests which arrive during a time period of length t?

(2)

(ii) The pests are exterminated at random instants of time, where the exterminations take place according to a Poisson process with rate μ. So if $N(t)$ represents the number of pests present in the habitat at time t, then the process $\{N(t)\}$ $(t \geq 0)$ is modelled as a continuous time Markov chain model with state space the set of all non-negative integers and the following transition rates.

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

(a) Write down the balance equations for the equilibrium distribution $\{\pi_j\}$ $(j \geq 0)$ together with the normalisation condition.

Find the equilibrium distribution, showing that it exists for all values of the parameters $\lambda > 0$ and $\mu > 0$.

(7)

(b) Define $p_j(t) = P(N(t) = j \,|\, N(0) = 0)$ $(j \geq 0)$. Obtain the forward equations and, using the condition $\sum_{j=0}^{\infty} p_j(t) = 1$ $(t \geq 0)$, show that they simplify to

$$\frac{dp_0}{dt} = \mu - (\lambda + \mu)\, p_0(t),$$

$$\frac{dp_j}{dt} = -(\lambda + \mu)\, p_j(t) + \lambda p_{j-1}(t) \qquad (j \geq 1).$$

(7)

(c) Show that a general solution of the above differential equation for $p_0(t)$ is given by

$$p_0(t) = \frac{\mu}{\lambda + \mu} + Ae^{-(\lambda+\mu)t} \qquad (t \geq 0),$$

where A is an arbitrary constant. Deduce that, in the present case,

$$p_0(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t} \qquad (t \geq 0).$$

(4)

4. Consider an M/M/2 queue with arrival rate λ and service rate μ per server.

(i) For the corresponding continuous time Markov chain, $\{N(t)\}$ $(t \geq 0)$, specify the state space and write down the instantaneous transition rates.

(4)

(ii) Define the *traffic intensity* ρ and state the necessary and sufficient condition for an equilibrium distribution to exist.

(2)

(iii) Write down the detailed balance equations and, assuming that the condition for an equilibrium distribution to exist is satisfied, find the equilibrium distribution.

(10)

(iv) Deduce that the probability that in equilibrium an arriving customer is delayed, i.e. has to wait before his service time starts, is given by

$$\frac{2\rho^2}{1+\rho}.$$

(4)

5. (i) Explain what is meant by a *white noise process* $\{\varepsilon_t\}$.

(2)

(ii) Write down the equations defining the AR(1) and MA(1) models for a stationary process $\{Y_t\}$, explaining all terms used.

(6)

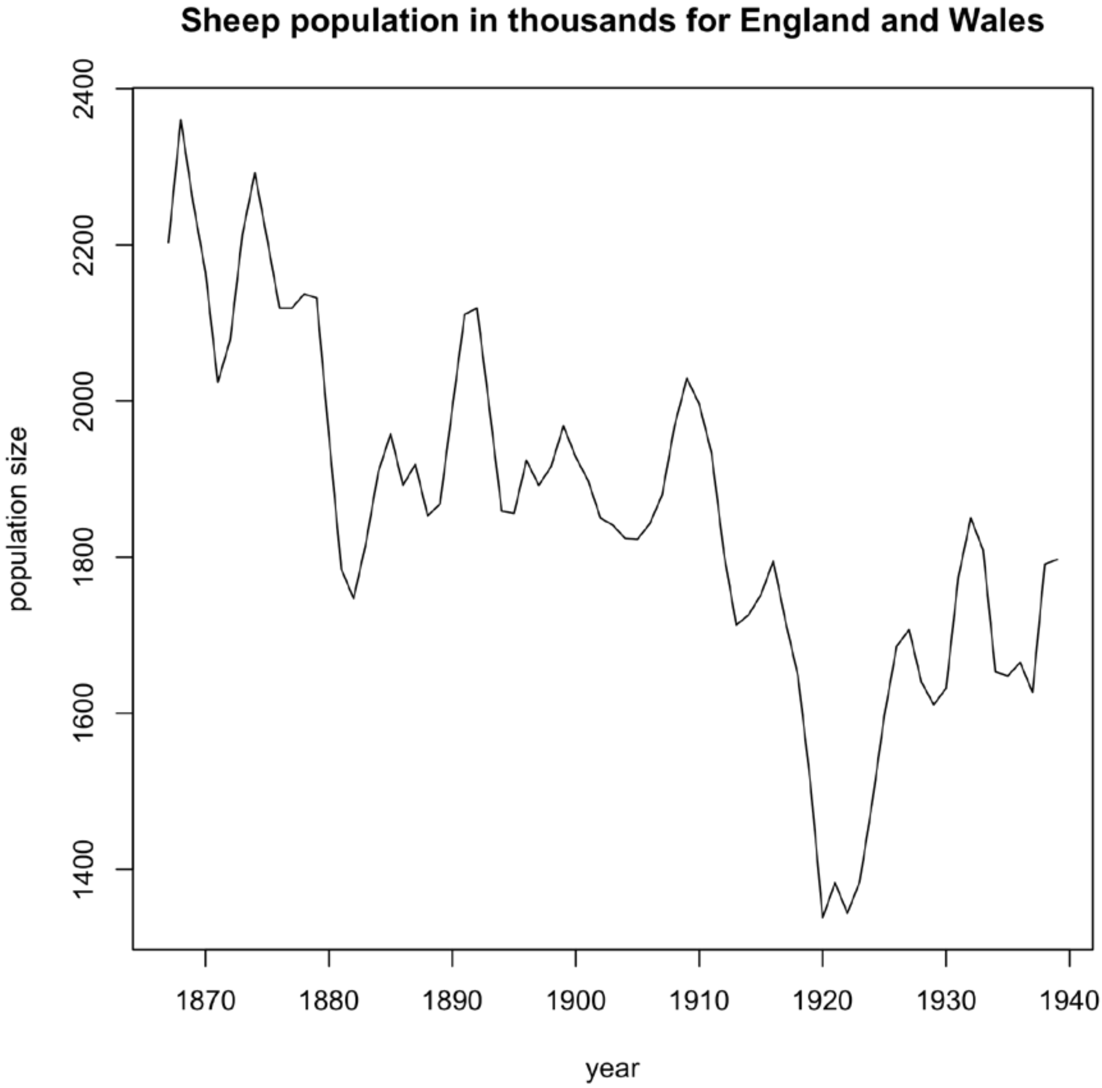
(iii) Specifying the necessary condition on the parameters, show how an AR(1) model can be written as an infinite moving average.

(5)

(iv) Using the model equation, derive the autocorrelation function for the MA(1) model.

(7)

6. A time series of the annual sheep population (in thousands) of England and Wales is plotted for the years 1867 to 1939 in the diagram below.



The autocorrelation function (acf) of the raw time series data is listed in the table **on the next page**, together with the autocorrelation function and partial autocorrelation function (pacf) of the first differences of the data.

Question continued on next page

Turn over

raw data		*differenced data*	
lag	*acf*	*acf*	*pacf*
1	0.913	0.355	0.355
2	0.760	–0.145	–0.311
3	0.634	–0.408	–0.291
4	0.573	–0.270	–0.053
5	0.563	0.068	0.096
6	0.537	0.162	–0.078
7	0.477	0.078	–0.056
8	0.399	–0.027	0.024
9	0.330	–0.061	0.011
10	0.276	–0.062	–0.072
11	0.228	–0.042	–0.046
12	0.185	–0.079	–0.100
13	0.153	–0.107	–0.120
14	0.153	–0.084	–0.104
15	0.178	–0.009	–0.057
16	0.207	0.082	–0.013
17	0.214	0.273	0.243
18	0.173	0.089	–0.123

(i) Comment on the plot, on the acf of the raw data, and on why differencing has been carried out.

(3)

(ii) From careful examination of the acf and pacf of the differenced data, state with reasons which of the family of ARIMA models might be fitted to the data.

(5)

(iii) In the extract from a computer output shown below, one of the family of ARIMA models has been fitted to the data.

```
Call:
arima(x = sheep.ts, order = c(3, 1, 0))

Coefficients:
      ar1      ar2      ar3
      0.4210  –0.2018  –0.3044
s.e.  0.1193   0.1363   0.1243

sigma^2 estimated as 4783
```

State which ARIMA model has been fitted to the data and write out explicitly the equation of the fitted model in terms of the observed values x_t of the sheep population.

(4)

(iv) Explain the meaning of the statement "sigma^2 estimated as 4783".

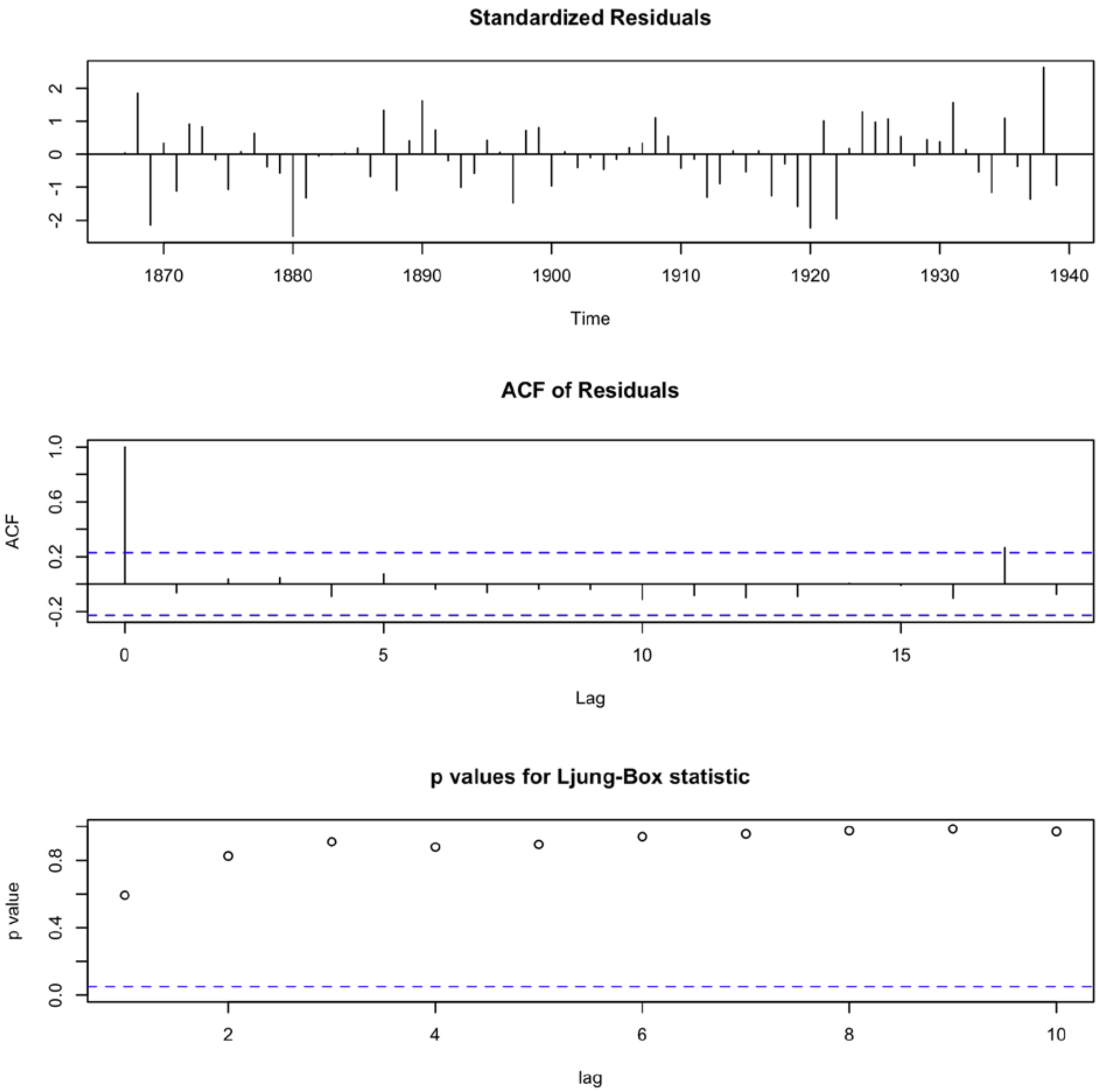
(2)

Question continued on next page

(v) In order to check the adequacy of the fitted model, the set of diagrams below has been produced.

Comment on what each of the diagrams represents and discuss the adequacy of the model.

(6)



Turn over

7. The Holt-Winters forecasting procedure is to be used for a time series with multiplicative seasonal variation of period p.

(i) Let Y_t denote the observed value of the series at time t, L_t the local level, B_t the trend and I_t the seasonal index at time t. If α, γ and δ denote the smoothing constants for L_t, B_t and I_t respectively, write down the updating equations for L_t, B_t and I_t. (3)

(ii) Write down an expression for the forecast $\hat{y}_T(h)$ at time T for lead time h, where $1 \le h \le p$. (2)

The United States Energy Information Administration produces monthly figures for total electricity net generation (in billions of kilowatt-hours). The data have been analysed using Holt-Winters forecasting with multiplicative seasonal variation of period 12. In the table below, the quantity of electricity generated and other quantities that have been calculated month by month, using the smoothing constants $\alpha = 0.3$, $\gamma = 0.01$ and $\delta = 0.3$, are shown for the two years 2006 and 2007.

Month	*Year*	*Generated*	*Fitted/ Forecast*	*Level*	*Trend*	*Seasonal Index*	*Residual/ Error*
Jan	2006	328.658	356.748	331.610	0.431	1.033	–28.090
Feb	2006	307.333	307.346	332.037	0.431	0.926	–0.013
Mar	2006	318.730	320.477	331.924	0.425	0.963	–1.747
Apr	2006	297.858	299.508	331.799	0.420	0.900	–1.650
May	2006	330.616	327.298	333.230	0.430	0.987	3.318
Jun	2006	364.260	359.556	334.969	0.443	1.081	4.704
Jul	2006	410.421	398.728	338.363	0.472	1.196	11.693
Aug	2006	407.763	401.327	340.465	0.489	1.188	6.436
Sep	2006	332.055	348.985	335.992	0.439	1.013	–16.930
Oct	2006	321.567	318.635	337.359	0.448	0.949	2.932
Nov	2006	309.159	308.903	337.892	0.449	0.915	0.256
Dec	2006	336.283	344.262	335.988	0.426	1.013	–7.979
Jan	2007	353.531	347.382	338.200	0.444	1.036	6.149
Feb	2007	323.230	313.455	341.812	0.475	0.932	9.775
Mar	2007	320.471	329.565	339.454	0.447	0.957	–9.094
Apr	2007	303.129	305.959	338.958	0.437	0.898	–2.830
May	2007	330.203	335.077	337.914	0.423	0.984	–4.874
Jun	2007	362.755	365.594	337.548	0.415	1.079	–2.839
Jul	2007	393.226	404.213	335.207	0.387	1.189	–10.987
Aug	2007	421.797	398.821	341.394	0.445	1.203	22.976
Sep	2007	355.394	346.274	344.540	0.472	1.019	9.120
Oct	2007	332.615	327.393	346.664	0.489	0.952	5.222
Nov	2007	314.103	317.503	346.037	0.478	0.913	–3.400
Dec	2007	346.290	350.851	345.163	0.464	1.010	–4.561

(iii) Given the data as at December 2007, calculate forecasts (to 3 decimal places, in billions of kilowatt-hours) of the quantity of electricity generated for (a) January 2008 and (b) December 2008. (4)

(iv) The quantity of electricity generated for January 2008 turned out to be 362.142. Given this fact, calculate the values of all the remaining figures in the corresponding row of the table. (8)

(v) Using the whole historical run of the above monthly series of total electricity generated, a statistical package may be used to find an optimal set of values for the smoothing constants, α, γ and δ. Outline a method by which this might be done. (3)

8. Let $\{Y_t\}$ be any ARMA process with autocorrelation function $\{\rho_\tau\}$. The corresponding spectral density function $f(\omega)$ may be written as

$$f(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \rho_\tau e^{-i\omega\tau} \quad (0 \le \omega \le \pi) .$$

(i) How you would interpret the spectral density function of any given ARMA process?

(2)

(ii) Write down $f(\omega)$ when $\{Y_t\}$ is a white noise process $\{\varepsilon_t\}$, and comment.

(3)

(iii) If $\{Y_t\}$ is an AR(1) process with model equation $Y_t = \varphi Y_{t-1} + \varepsilon_t$, where $-1 < \varphi < 1$, find the autocorrelation function.

(6)

(iv) Deduce that the spectral density function for the above AR(1) process is given by

$$f(\omega) = \frac{1}{2\pi} \frac{1-\varphi^2}{1+\varphi^2 - 2\varphi\cos\omega} \quad (0 \le \omega \le \pi) .$$

(5)

(v) Comment on how the shape of the spectral density function depends on the sign of φ and what this implies about the behaviour of the process, especially in comparison with a white noise process.

(4)