EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2012

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 9 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 has the binomial distribution with parameters 2 and $\frac{1}{2}$, so that the offspring distribution $\{p_n\}$ is given by $p_0 = \frac{1}{4},\ p_1 = \frac{1}{2},\ p_2 = \frac{1}{4}$. The numbers of surviving offspring produced by different individuals are statistically independent of each other.

(i) Write down an expression for the probability generating function $G(z)$ of the offspring distribution.

(1)

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

(ii) By conditioning on X_1, prove that

$$G_n(z) = \tfrac{1}{4}[1 + G_{n-1}(z)]^2 \qquad (n \geq 2).$$

(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 relation for the θ_n and deduce the value of $\theta = \lim_{n\to\infty} \theta_n$, the probability of ultimate extinction of the population.

(5)

(iv) Let μ_n denote the mean population size in the nth generation $(n \geq 1)$. By differentiating the relationship of part (ii), find a recurrence relation for the μ_n and deduce that $\mu_n = 1$ $(n \geq 1)$.

(5)

(v) Let σ_n^2 denote the variance of the population size in the nth generation $(n \geq 1)$. By differentiating the relationship of part (ii) twice, find a recurrence relation for the σ_n^2 and deduce that

$$\sigma_n^2 = \frac{n}{2} \quad (n \geq 1).$$

(5)

2. A Markov chain model for death from a particular infection of an animal from some population has four states as follows.

state 1: animal alive and free from infection

state 2: animal alive and infected

state 3: animal dead from the infection

state 4: animal dead from another cause

The unit of time is taken to be one year, and the transition matrix **P** is given by

$$\mathbf{P} = \begin{pmatrix} \frac{5}{8} & \frac{1}{4} & 0 & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

(i) Identify, with reasons, the states that are transient and those that are recurrent.

(4)

(ii) If in a given year the animal is in state 1, calculate the probability that two years later it is in state i, for each of $i = 1, 2, 3, 4$.

(5)

(iii) Let x_i ($i = 1, 2$) denote the probability that the animal eventually dies from the infection, given that it starts in state i. Write down and solve a pair of backward equations for the x_i.

(6)

(iv) Let y_i ($i = 1, 2$) denote the expected number of years until the animal dies (either from the infection or from another cause), given that it starts in state i. Write down and solve a pair of backward equations for the y_i.

(5)

Turn over

3. In a Markov chain model for the state of a machine in a factory, X_n denotes the state in week n, where the state space is $\{0, 1, 2, \ldots, M\}$ for some positive integer M. The state 0 represents a machine that is in perfect working order and the states $1, 2, \ldots, M$ represent successively worse states of operation. When the machine reaches state M then it is serviced so that it returns to state 0 the following week. The transition rates $p_{i,j}$ of the model are specified as follows:

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

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

(i) Write down a set of equations satisfied by the stationary distribution $\{\pi_i\}$ $(0 \le i \le M)$ and deduce expressions for the probabilities π_i $(0 \le i \le M)$ in terms of θ and M.

(12)

(ii) If c, K are constants such that the cost per week of running the machine is given by ci whenever the machine is in state i $(0 \le i \le M)$ and K is the cost of servicing the machine, show that the long-term average cost per week of running and servicing the machine is given by

$$\frac{\frac{1}{2}cM(M-1)+\theta(cM+K)}{M+\theta}.$$

(8)

4. Consider a continuous time Markov chain model $\{N(t): t \geq 0\}$ for a car park with M spaces, where $N(t)$ denotes the number of spaces occupied at time t. It is assumed that cars arrive at the car park according to a Poisson process with rate λ and that if there are any free spaces then an arriving car is parked, but if all spaces are occupied then the arriving car leaves. It is also assumed that if at any time there are i spaces occupied then the probability that one of the parked cars leaves in the next small interval of time of length h is $\mu ih + o(h)$.

(i) Write down the state space for the model and specify the transition rates.

(4)

(ii) Define

$$p_n(t) = P(N(t) = n \mid N(0) = \nu) \qquad (0 \leq n \leq M),$$

where ν represents the number of cars in the car park at time 0. Derive the forward equations for the $p_n(t)$, distinguishing between the cases $n = 0$, $1 \leq n \leq M - 1$ and $n = M$.

(8)

(iii) By first writing down the detailed balance equations (or otherwise), obtain the equilibrium distribution $\{\pi_n\}$ $(0 \leq n \leq M)$ for the chain. Show in particular that the long-term proportion of time that the car park is full is given by

$$\frac{\frac{\rho^M}{M!}}{\sum_{n=0}^{M} \frac{\rho^n}{n!}}, \text{ where } \rho = \frac{\lambda}{\mu}.$$

(8)

5. In a single-server queue, customers arrive according to a Poisson process with rate λ and service times have a gamma distribution with parameters μ and 2, so that the probability density function of the service times is

$$\mu^2 t e^{-\mu t} \qquad (t \geq 0).$$

(i) Find the mean of the service time distribution. Define the *traffic intensity* ρ of a queue and, for the present case, write down an expression for ρ in terms of λ and μ. State the condition for an equilibrium distribution to exist.

(5)

(ii) Conditional upon the length t of a service time being given, write down the distribution of the number of customers who arrive during the service time and show that the probability generating function (pgf) $G(z)$ of this distribution is given by

$$G(z) = e^{-\lambda t(1-z)}.$$

(4)

(iii) Deduce that, unconditionally, the pgf $K(z)$ of the number of customers who arrive during a service time is given by

$$K(z) = \left(\frac{\mu}{\mu + \lambda - \lambda z} \right)^2.$$

(5)

(iv) Deduce an expression for the expected number of customers who arrive during a service time and discuss how your expression relates to the results of part (i).

(6)

Turn over

6. Consider the AR(2) model

$$Y_t = \tfrac{3}{10}Y_{t-1} + \tfrac{1}{10}Y_{t-2} + \varepsilon_t \qquad (-\infty < t < \infty)$$

for a process $\{Y_t\}$, where $\{\varepsilon_t\}$ is a white noise process.

(i) Find the roots of the autoregressive characteristic equation and verify that the stationarity condition is satisfied.

(4)

(ii) The infinite moving average representation of Y_t can be written

$$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} .$$

By substituting this expression into the model equation, find a set of recurrence relations satisfied by the ψ_i together with appropriate initial conditions.

(6)

(iii) Solve these recurrence equations to show that

$$\psi_i = \tfrac{5}{7}\left(\tfrac{1}{2}\right)^i + \tfrac{2}{7}\left(-\tfrac{1}{5}\right)^i \qquad (i \geq 0).$$

(6)

(iv) Deduce that

$$\text{Var}(Y_t) = \tfrac{25}{22}\sigma^2 ,$$

where σ^2 is the variance of the white noise process $\{\varepsilon_t\}$.

(4)

Turn over

7. A time series consists of the mean lake surface elevation in feet at a particular location on Lake Huron in July for the 127 years from 1860 to 1986.

(a) The sample autocorrelation function (acf) and sample partial autocorrelation function (pacf) are tabulated below for the first 20 lags.

lag	acf	pacf	lag	acf	pacf
1	0.831	0.831	11	0.216	–0.012
2	0.643	–0.157	12	0.194	0.071
3	0.513	0.083	13	0.212	0.111
4	0.436	0.062	14	0.207	–0.115
5	0.403	0.097	15	0.187	0.038
6	0.364	–0.033	16	0.170	0.002
7	0.340	0.082	17	0.118	–0.107
8	0.309	–0.027	18	0.085	0.007
9	0.310	0.133	19	0.077	0.051
10	0.273	–0.141	20	0.077	0.006

(i) Explain in general how the acf and pacf may be used to determine whether an observed time series appears to be stationary and, if so, which of the family of ARMA models it might be appropriate to attempt to fit to the series. (4)

(ii) Draw conclusions in the present case. (4)

(b) An edited version of the output from a statistical package for fitting one of the ARMA models to the Lake Huron data is given below.

```
ARMA 'level'
Final Estimates of Parameters

Type           Coef    SE Coef        T        P
AR   1       0.8660     0.0470    18.43    0.000
Constant    77.6617     0.0632  1228.99    0.000
Mean        579.503      0.472

Number of observations: 127
Residuals:    SS = 62.0777  (backforecasts excluded)
              MS = 0.4966   DF = 125

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

Lag            12       24       36       48
Chi-Square   12.9     26.4     36.3     46.9
DF             10       22       34       46
P-Value     0.229    0.236    0.364    0.434
```

(i) State which of the family of ARMA models has been fitted to the data and write down explicitly the equation of the fitted model, including the white noise term. (4)

(ii) Discuss how the above output may be used to determine whether the fitted model provides an adequate fit to the data, and draw conclusions. (4)

(iii) Given that the mean level in 1986 was 581.27, calculate a 95% prediction interval for the mean level in 1987. (4)

Turn over

8. Let $y_1, y_2, \ldots, y_t, \ldots$ denote the observed values of a time series starting from time 1.

(i) Let L_t denote the smoothed value (the level) of the series at time t obtained using simple exponential smoothing. If α denotes the smoothing constant, write down the updating equation for L_t in terms of L_{t-1} and y_t. State what is the corresponding forecast $\hat{y}_t(h)$ at time t for lead time h ($h \geq 1$).

(3)

(ii) Assume that the value L_1 is taken to be equal to y_1. By applying the updating equation iteratively, find an explicit expression, in as simple terms as you can, for L_t in terms of $y_1, y_2, \ldots, y_t$.

(4)

(iii) Denote by e_t the one-step-ahead forecast error, $e_t = y_t - \hat{y}_{t-1}(1)$ $(t \geq 2)$. Rewrite the updating equation for L_t in terms of L_{t-1} and e_t.

(3)

The monthly sales of TV sets in a retail store have been recorded over 24 successive months and the method of simple exponential smoothing is applied. The sales for the first four months, t = 1, 2, 3, 4, are 30, 32, 30, 39, respectively.

(iv) Using the initial value y_1 of the series as the initial smoothed value L_1, and taking the smoothing constant α to be 0.2, calculate (to 2 decimal places) the smoothed value and forecast error for t = 2, 3, 4, respectively.

(4)

In order to explore what might be the best value of the smoothing constant to use, the method of simple exponential smoothing has been applied to the whole run of the series using a variety of values of the smoothing constant. In each case the values of the mean absolute deviation (MAD) and the error sum of squares (SSE) have been calculated. The results are tabulated below.

α	MAD	SSE
0.10	3.47	383.98
0.20	3.13	335.95
0.25	3.05	333.47
0.30	3.00	336.22
0.35	3.03	342.38
0.40	3.07	351.11
0.50	3.22	374.78

(v) Define the MAD and the SSE.

(2)

(vi) Given the tabulated results, discuss what value of α you would recommend.

(4)