EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2014

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* ${}^{n}C_{r}$.

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 branching process starts with a single individual in generation zero. In each generation, each individual independently produces a number of offspring given by the random variable Z having probability generating function (pgf) $G(s)$.

(i) Prove that $G_{n+1}(s) = G_n(G(s))$, where $G_n(s)$ is the pgf of X_n, the number of individuals in the nth generation.

(5)

Consider the case where $G(s) = \dfrac{1+4s}{6-s}$.

(ii) Use the result in part (i) to show by induction that

$$G_n(s) = \frac{n-(n-5)s}{5-n(s-1)}, \quad n = 0, 1, 2, \ldots .$$

(6)

(iii) Find the expected size of the nth generation.

(3)

(iv) By considering the relevant probability for the nth generation, confirm that the process will eventually become extinct.

(2)

(v) Let T be the generation when the process becomes extinct. Show that the distribution of T is given by

$$P(T = n) = \frac{5}{(n+5)(n+4)}, \quad \text{for } n = 1, 2, 3, \ldots .$$

(4)

2. In the Ultrastratos civilisation on planet Anachronista, the population is divided into four strata which, in order of status, are labelled Alpha, Beta, Gamma and Delta. By the traditions of the civilisation, no child can have a status more than one different from its parents. As examples, in each generation 20% of the children of Alphas grow up to be Betas, the rest remaining Alphas, while of the Beta offspring 50% remain Betas while 10% become Alphas and the rest Gammas.

A Markov chain describes the status of members of the population at successive generations. Its transition matrix is given by **P**, defined below, in which some entries labelled * have been omitted.

$$\mathbf{P} = \begin{array}{r} \textit{Alpha} \\ \textit{Beta} \\ \textit{Gamma} \\ \textit{Delta} \end{array} \begin{pmatrix} * & 0.2 & * & 0 \\ 0.1 & 0.5 & * & * \\ * & * & 0.5 & 0.4 \\ * & 0 & * & 0.8 \end{pmatrix}$$

(i) Use the information above to fill in the missing elements of **P** and calculate the two-step-ahead transition matrix.

(4)

(ii) What is the probability that the child of a Beta becomes a Gamma? Show that the grandchild of a Gamma is twice as likely to become a Delta as the grandchild of a Delta is to become a Gamma.

(3)

(iii) Calculate the probability that, after four generations, a descendant of an Alpha is a Delta.

(3)

(iv) Explain how you can tell that the chain consists of a single irreducible class. Are any of the states transient? Give your reason.

(3)

(v) Find the stationary distribution of the chain. If the population initially has no Deltas, what will be the proportion of Deltas after a large number of generations?

(7)

3. (i) In a simple M/M/1 queue with arrival rate λ and service rate μ the equilibrium distribution of the number in the system, i.e. queueing or being served, is given as

$$\pi_n = \rho^n(1-\rho) \text{ for } n = 0, 1, 2, \ldots \text{ .}$$

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

(2)

(ii) If a customer arrives to find that the server is busy, state with reasons the distribution of the remaining service time of the customer currently being served.

(2)

(iii) A customer's waiting time is defined to be the length of time from when he arrives to when he leaves the queue, his service time having been completed. Customers are served on a "first come, first served" basis. A customer arrives to find n customers ahead of him in the queue. Briefly explain why his waiting time has the gamma distribution with probability density function

$$f(t\,|\,n) = \frac{\mu^{n+1}t^n e^{-\mu t}}{n!}, \text{ for } t > 0 \text{ .}$$

(3)

(iv) If a customer arrives to join the queue when it is in equilibrium, show that his waiting time distribution is exponential with parameter $\mu(1-\rho)$.

(5)

(v) Now consider a computer system in which jobs arrive at a high definition graph plotter randomly at average intervals of 2 minutes. If the plotter processes jobs at a rate of θ per hour, what is the expected waiting time of a job?

(2)

(vi) The cost of buying and operating a plotter which processes jobs at a rate of θ per hour is divided by its intended lifetime to give an hourly cost of £5θ. A job which has a waiting time of t minutes is assumed to incur a cost of £8t. Calculate the expected total cost and find the optimal choice of θ.

(6)

4. Suppose that the probability of an accident in $(t, t + \delta t]$ given that n accidents have occurred in $(0, t]$ is $(2 + n)\delta t + o(\delta t)$, regardless of exactly when the n accidents occurred. The probability of more than one accident in $(t, t + \delta t]$ is $o(\delta t)$.

(i) Show that, for $n \geq 1$, the probability $p_n(t)$ of n accidents in $(0, t]$ satisfies the differential equation

$$\frac{dp_n(t)}{dt} = -(2+n)\,p_n(t) + (2+n-1)\,p_{n-1}(t)\,.$$

Derive also the corresponding differential equation for $p_0(t)$.

(6)

The corresponding probability generating function is defined as

$$\Pi(s,t) = \sum_{n=0}^{\infty} p_n(t)s^n\ .$$

(ii) Using the results in part (i), or otherwise, show that $\Pi(s, t)$ satisfies the partial differential equation

$$\frac{\partial \Pi}{\partial t} = 2(s-1)\Pi + s(s-1)\frac{\partial \Pi}{\partial s}$$

with boundary condition $\Pi(s, 0) = 1$.

(8)

(iii) Verify that the function

$$\Pi(s,t) = e^{-2t}[1-(1-e^{-t})s]^{-2}$$

satisfies the partial differential equation in part (ii) with its boundary condition.

(6)

5. An M/G/1 queue in equilibrium has arrival rate λ and random service times X having mean $\frac{1}{\mu}$ and variance σ^2. Customers are served in the order in which they arrive.

(i) Let W_Q be the expected time a customer spends waiting for his service to commence. Assuming the basic identity

$$W_Q = \lambda\{E(X)W_Q + \tfrac{1}{2}E(X^2)\},$$

show that

$$W_Q = \frac{\lambda\left(\sigma^2 + \frac{1}{\mu^2}\right)}{2(1-\rho)}, \text{ where } \rho = \frac{\lambda}{\mu}.$$

(3)

(ii) Deduce the Pollaczek-Khintchine steady-state formula for the expected number of customers in the queue (including any currently being served):

$$L_S = \frac{\rho(2-\rho) + \lambda^2\sigma^2}{2(1-\rho)}.$$

(3)

(iii) In what sense can an M/G/1 queue with deterministic service times be said to be optimal?

(2)

(iv) A system with random arrivals averaging one every 10 minutes currently has one human server, the service times being independent and exponentially distributed with mean 5 minutes and variance 25 minutes^2. It is possible to automate the system so that all services take exactly c minutes. Find the maximum value of c that would give a reduction in the mean time a customer spends waiting for service to commence.

(12)

6. (i) The time series $X_t = 1.4X_{t-1} - 0.45X_{t-2} + A_t$, where A_t is a white noise series with variance σ^2, can be described as an ARIMA(p, d, q) model. What are the values of p, d and q? Verify that X_t is both stationary and invertible.

(6)

(ii) Define the *lag k autocorrelation* ρ_k and show that the autocorrelations satisfy the equations

$$\rho_k = 1.4\rho_{k-1} - 0.45\rho_{k-2} \text{ for } k \geq 1.$$

Deduce that $\rho_1 = 1.4 - 0.45\rho_1$.

(5)

(iii) Calculate and sketch ρ_k for k = 1, 2, 3, 4. What do these autocorrelations tell you about the likely appearance of the series?

(4)

(iv) Consider the following ARIMA(p, d, q) model for a time series W_t:

$$W_t = 2.4W_{t-1} - 1.85W_{t-2} + 0.45W_{t-3} + A_t.$$

Find the values of p, d and q and state whether W_t is

(a) stationary,

(b) invertible.

(4)

(v) Briefly describe the behaviour you would expect to see in realisations of the model in part (iv).

(1)

7. The updating equations of the Holt-Winters forecasting procedure for a time series y_t with multiplicative seasonal variation of period p are given by

$$\ell_t = \alpha\left(\frac{y_t}{s_{t-p}}\right) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \gamma(\ell_t - \ell_{t-1}) + (1-\gamma)b_{t-1}$$

$$s_t = \delta\left(\frac{y_t}{\ell_t}\right) + (1-\delta)s_{t-p}$$

where α, γ and δ are the smoothing constants.

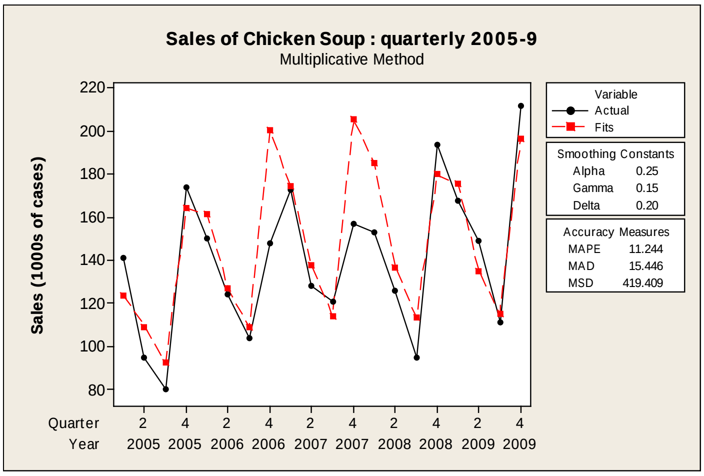
(i) Explain what properties of the series are described by the three quantities ℓ_t, b_t and s_t. Why might Holt-Winters be preferable to a forecasting method based on fixed parameter regression?

(4)

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

(2)

The quarterly sales of chicken soup were recorded over the five years 2005 to 2009 and analysed using the Holt-Winters method with multiplicative seasonal variation of period 4 and smoothing constants α = 0.25, γ = 0.15 and δ = 0.2. The results are shown in the following graph.



The table on the next page shows the sales and other quantities that have been calculated quarter by quarter for the year 2009.

Question continued on the next page

Year	*Quarter*	*Sales (1000s)*	*Level*	*Trend*	*Seasonal*	*Residual*
2009	Q1	168	157.3	1.80875	1.09687	–7.5271
2009	Q2	149	163.2	2.43148	0.86100	14.0831
2009	Q3	111	164.1	2.19638	0.69236	–4.3657
2009	Q4	212	169.5	2.68347	1.19603	15.3588

(iii) Showing your working, calculate forecasts of sales for

(a) 1st quarter 2010,

(b) 4th quarter 2011.

(4)

(iv) The sales in the 1st quarter of 2010 turned out to be 170. Given this fact, calculate the values of Level, Trend, Seasonal and Residual for this quarter. What now is your forecast for the 4th quarter of 2011?

(8)

(v) The graph on the previous page includes, as an Accuracy Measure, the mean absolute percentage error defined as

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\frac{|e_t|}{y_t}, \text{ where } e_t = y_t - \hat{y}_{t-1}(1).$$

Why would MAPE be of particular relevance when the multiplicative form of Holt-Winters forecasting is appropriate?

(2)

8. (a) State the Yule-Walker equations relating the lag k autocorrelation ρ_k of a stationary autoregressive time series process of order m to ρ_{k-i} for $i = 1, 2, \ldots, m$.

(2)

(b) The estimated variance, c_0, and the first two estimated autocovariances, c_1 and c_2, from a set of 250 values y_t of a time series are given by

$$c_0 = 200, \;\; c_1 = -96, \;\; c_2 = 16.$$

(i) Estimate the first two autocorrelations and use the Yule-Walker equations to calculate the first two sample partial autocorrelations of the series.

(6)

(ii) A time series analyst has examined plots of the autocorrelation and partial autocorrelation functions of y_t and concluded that a first order moving average process of the form

$$Y_t = A_t - \theta A_{t-1},$$

where A_t is white noise, might be appropriate. Explain why your calculations in part (b)(i) would have ruled out either an MA(2) or an AR(1) model. Describe what else he would have observed to justify his suggestion of an MA(1) model.

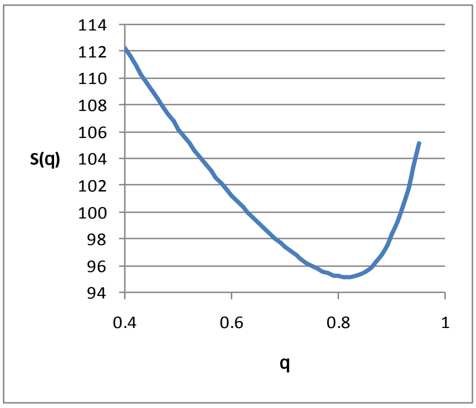
(4)

(iii) Give an initial estimate for the parameter θ in that model.

(5)

(iv) The analyst proceeded to evaluate the conditional sum of squares function $S(\theta)$ for a range of values of the parameter θ, resulting in the plot shown below. Explain how he would have done this and interpret the plot.

(3)



BLANK PAGE

BLANK PAGE