INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

September 2010 examinations

Subject CT4 — Models
Core Technical

Introduction

The attached subject report has been written by the Principal Examiner with the aim of helping candidates. The questions and comments are based around Core Reading as the interpretation of the syllabus to which the examiners are working. They have however given credit for any alternative approach or interpretation which they consider to be reasonable.

T J Birse
Chairman of the Board of Examiners

December 2010

Question 1

One or both of the runs (the original or the new) may have been incorrect as, for example, the second trainee may not have been fully aware of the set-up (for example he or she may not have followed the procedure correctly, or may have used different assumptions)

The difference between the two runs may not have only been the parameter change, for example the two runs may have used different random seeds, or the second run may have had fewer simulations.

The expectation that the model was not sensitive to this parameter could have been incorrect.

Other valid points were given credit, for example that some parameters might be linked to live data, which will necessarily have changed; or that there may have been other amendments to the data in the meantime. However, the maximum number of marks attainable on this question was 3.

Question 2

A deterministic model is a model which does not contain any random components.

The output is determined once the fixed inputs and the relationships between inputs and outputs have been defined.

A stochastic model is one that recognises the random nature of the input components.

The inputs to a stochastic model are random variables, and hence for any given values of the inputs the outputs are an estimate of the characteristics of the model.

Several independent iterations of the model are required for each set of inputs to study their implications.

The output of a stochastic model gives the distribution of relevant results for a distribution of scenarios.

A deterministic model can be seen as a special case of a stochastic model.

The output of a stochastic model can be reproduced if the same random seed is used.

The output of a deterministic model is only a snap shot or an estimate of the characteristics of the model for a given set of inputs.

Full marks could be obtained for rather less than is written above. The maximum number of marks attainable was 4 even if all the above points were made.

Question 3

Define the objectives of the model – what aspects of mortality are to be analysed (e.g. average mortality rates, split male/female, analysis of trends over 50 years).

Plan the model.

Establish what data are available – collect data

Evaluate the accuracy of data and the consistency of the data over time (e.g. there may have been changes to the way deaths and census data were recorded over a 50-year period)

Try to identify the main features of the mortality, and measure them.

Involve experts – e.g. there may be a national census office or Government department who can advise.

Decide between simulation package or general purpose language, or use of spreadsheet package.

Set up computer program and input data.

Debug program.

Test the output for reasonableness – is the model faithful to the actual mortality.experience of the island over the required time frame?

Check the sensitivity of model to small changes to input parameters.

Analyse the model output.

Communicate and document the results.

This question was generally well answered, though many candidates simply reproduced the list in the Core Reading, Unit 1, pages 2 and 3, without any reference to the specific problem in the question – the analysis of mortality. These candidates did not gain full credit. A minority of candidates interpreted this question as being about a mortality investigation, making reference to the estimation of mortality rates and their subsequent graduation. Credit was given to such candidates.

Question 4

The null hypothesis is poorly expressed – should be "underlying rates are the graduated rates" or similar.

The test statistic is incorrect – the denominator should be expected deaths.

Cannot comment on figures in table as no access to workings.

Number of ages is 6 not 5.

However fewer than 6 degrees of freedom is appropriate because should deduct 1 for estimated parameter and some for choice of standard table

This is a one-tailed test not two-tailed.

Even if it were two-tailed, multiplying test statistic by 2 is inappropriate.

The trainee has not stated the level of significance to which he or she is working (presumably 5 per cent)

Does not explain that the reason for conclusion is 12.833 > 5.32826.

The null hypothesis should never be "accepted" rather it is "not rejected".

The trainee has not stated his or her conclusion in terms of the null hypothesis

All the graduated rates are above the crude rates so although the graduation has been accepted it is suspect.

This question was reasonably well answered.

Question 5

(i) (a) ${}_tq_x = t \times q_x$

(b) EITHER ${}_tq_x = 1 - e^{-\mu t}$ OR ${}_tq_x = 1 - (1 - q_x)^t$

(c) $${}_tq_x = \frac{tq_x}{1 - (1-t)q_x}$$

(ii) (a) ${}_{½}q_{60} = 0.025$

therefore ${}_{½}p_{60} = 0.975000$

(b) $1 - e^{-\mu} = 0.05$

so $-\mu = \ln 0.95$ and $\mu = 0.051293$

${}_{½}p_{60} = e^{-0.5\mu} = 0.974679$

(c) $${}_{\frac{1}{2}}q_{60} = \frac{\frac{1}{2}(0.05)}{1-\frac{1}{2}(0.05)}$$

$= 0.025641025$

so ${}_{½}p_{60} = 0.974359$

(iii) The Balducci assumption has the smallest value, and the uniform distribution of deaths (UDD) the largest value

This is because the UDD implies an increasing force of mortality over the year of age, whereas the Balducci assumption implies a decreasing force and a constant force is clearly constant.

The higher the force of mortality in the second half of the year of age relative to its magnitude in the first half of the year of age, the higher the probability of survival to age 60.5 years

The difference between the three values of ${}_{0.5}q_{60}$ is very small in this case.

Most candidates answered the parts relating to the uniform distribution of deaths and the constant force of mortality correctly. Far fewer correctly worked out the formula for ${}_tq_x$ under the Balducci assumption. Instead, many candidates simply wrote down ${}_{1-t}q_{x+t} = (1-t)q_x$ in answer to (i)(c), which was not given credit, as it is not a formula for ${}_tq_x$ and hence is not answering the question set. However, credit was given to such candidates in (ii)(c) if they calculated the correct numerical value for ${}_{½}p_{60}$. Some candidates did not calculate the quantities in (ii) to six decimal places, and this was penalised.

Question 6

(i) Graphical graduation might be used when EITHER a quick visual impression OR a rough estimate is all that is required,

This is useful when the data are scanty and
EITHER
there is very little prior knowledge about the class of lives being analysed so that a suitable standard table cannot be found
OR
the experience of a professional person can be called upon

(ii) Plot the crude data,

preferably on a logarithmic scale.

If data are scanty, group ages together,

choosing evenly spaced groups and making sure there are a reasonable number of deaths (e.g. at least 5) in each group.

Plot approximate confidence limits or error bars around the plotted crude rates.

Draw the curve as smoothly as possible, trying to capture the overall shape of the crude rates.

Test the graduation for goodness-of-fit and
EITHER test for smoothness OR examine third differences

If the graduation fails the test, re-draw the curve.

"Hand polishing" individual ages may be necessary to ensure adequate smoothness.

Many answers to this question were very sketchy and missed several of the points listed above. In (i) simply saying "when data are scanty" was not sufficient for credit, as graduation with reference to a standard table can also be used with scanty data sets provided a suitable standard table can be found. In (ii) credit was given for additional points, including noting that the curve can go outside the 95% confidence intervals at one out of every 20 or so ages, and mentioning that the analyst might want to look at obvious outliers before drawing the curve, as these may indicate data errors. A maximum of 5 marks was available for (ii).

Question 7

We adjust the exposed to risk to correspond to the deaths data.

Deaths are recorded on an "age nearest birthday" basis. Let the number of deaths to persons aged x in countries A and B respectively in year t be $\theta^A_{x,t}$ and $\theta^B_{x,t}$.

This means that the estimated rate μ_x will apply to exact age x, no further adjustment being required.

Let the populations recorded in the censuses of the two countries as being aged x in the census on 1 January in year t be $P^A_{x,t}$ and $P^B_{x,t}$.

A central exposed to risk for each country for year t which corresponds to the deaths data is

$$E^{cA}_{x,t} = \int_{s=0}^{s=1} P^{*A}_{x,t+s}\, ds \quad \text{and} \quad E^{cB}_{x,t} = \int_{s=0}^{s=1} P^{*B}_{x,t+s}\, ds\,,$$

where $P^{*A}_{x,t+s}$ and $P^{*B}_{x,t+s}$ are the populations aged x nearest birthday in countries A and B at time $t + s$.

This central exposed to risk can be approximated by

$$E^{cA}_{x,t} = \frac{1}{2}(P^{*A}_{x,t} + P^{*A}_{x,t+1})$$

and

$$E^{cB}_{x,t} = \frac{1}{2}(P^{*B}_{x,t} + P^{*B}_{x,t+1})\,,$$

assuming the population varies linearly between census dates.

But in country A the census does not collect $P^{*A}_{x,t}$, but $P^{A}_{x,t}$, the population aged x last birthday.

Assuming birthdays are evenly distributed across the calendar year, however, we can write

$$P^{*A}_{x,t} = \frac{1}{2}(P^{A}_{x,t} + P^{A}_{x-1,t})\,.$$

We also know that $P^{*B}_{x,t} = P^{B}_{x,t}$.

Therefore an exposed to risk for the two countries combined which corresponds to the deaths data is

$$E^{cA}_{x,t} + E^{cB}_{x,t} = \frac{1}{2}(P^{*A}_{x,t} + P^{*A}_{x,t+1}) + \frac{1}{2}(P^{*B}_{x,t} + P^{*B}_{x,t+1})$$

$$= \frac{1}{2}\left(\frac{1}{2}(P^{A}_{x,t} + P^{A}_{x-1,t}) + \frac{1}{2}(P^{A}_{x,t+1} + P^{A}_{x-1,t+1})\right) + \frac{1}{2}(P^{B}_{x,t} + P^{B}_{x,t+1})\,,$$

and hence the combined age specific death rate can be estimated as

$$\mu_x = \frac{\theta^{A}_{x,t} + \theta^{B}_{x,t}}{\frac{1}{4}P^{A}_{x,t} + \frac{1}{4}P^{A}_{x-1,t} + \frac{1}{4}P^{A}_{x,t+1} + \frac{1}{4}P^{A}_{x-1,t+1} + \frac{1}{2}P^{B}_{x,t} + \frac{1}{2}P^{B}_{x,t+1}}\,.$$

The solution above assumes that the estimates of μ_x are to be made using a single calendar year t. Additional credit was given to candidates who stated that the appropriate time over which the estimates are to be made should be defined at the outset, and that if this period is longer than one year the deaths and exposed to risk for all relevant calendar years should be summed and the total deaths divided by the total exposed to risk (subject to a maximum of 8 marks being available). The Examiners were looking for understanding of the process that must be gone through in order to obtain the required estimates. Answers consisting mainly of "disembodied" statements, without a coherent argument received limited credit.

Question 8

(i) The null hypothesis, H_0, is that the climate – or the underlying (long-run average) temperature – in Rocky Bay in August is the same as that in the Mediterranean.

EITHER

Signs Test

Let P be the number of days for which the maximum temperature in Rocky Bay is greater than that expected in the Mediterranean.

Under H_0, $P \sim \text{Binomial}(31, 0.5)$.

THEN EITHER NORMAL APPROXIMATION

Using the Normal approximation as we have more than 20 days,

$$P \sim \text{Normal}\left(\frac{31}{2}, \frac{31}{4}\right)$$

In the observations $P = 7$,

The value of the test statistic is therefore

$$Z = \frac{7 - 15.5}{\sqrt{7.75}} = -3.05$$

Since $|Z| > 1.96$ we reject H_0 at the 5 per cent level of significance

OR EXACT CALCULATION

We have $P = 7$

The probability of obtaining 7 or fewer positive signs is

$$\binom{31}{7}0.5^{31}+\binom{31}{6}0.5^{31}+\ldots+\binom{31}{0}0.5^{31}$$

which is 0.00122 + 0.00034 + 0.00008 + ... + 0.00000 = 0.00166

since this is less than 0.025 (two-tailed test)

we reject H_0 at the 5 per cent level of significance

and conclude that the climate of Rocky Bay is not the same as that in the Mediterranean.
OR

Grouping of Signs Test

Let P be the number of days for which the maximum temperature in Rocky Bay is greater than that expected in the Mediterranean.

Let Q (= 31 – P) be the number of days for which the maximum temperature in Rocky Bay is less than that expected in the Mediterranean.

To test the null hypothesis, we need to calculate the maximum number

of positive runs, g, for which $\displaystyle\sum_{t=1}^{g}\frac{\binom{P-1}{t-1}\binom{Q+1}{t}}{\binom{P+Q}{P}}<0.05$.

since P = 7 and Q = 24,

THEN EITHER

using the table on p. 189 of the *Formulae and Tables for Examinations*,

we find that g = 3.

OR

using the normal approximation we have

G ~ Normal(5.64, 0.95),

so, using a one-tailed test, the critical value at the 5% level is 5.645 – 1.645*√0.947 = 4.04.

Since we only have 2 positive runs in the data we reject H_0 at the 5 per cent level of significance and conclude that the climate of Rocky Bay is not the same as that in the Mediterranean.

(ii) Runs of consecutive days with the same sign are likely since the weather tends to be determined by atmospheric conditions lasting more than one day.

The Mediterranean averages are averages for the month of August 2009, not long-run averages.

August 2009 might have been an unusually hot month in the Mediterranean region.

Maximum temperature is not the only measure of climate, also consider mean temperature, hours of sunshine, windiness, etc.

Choice of locations used for Mediterranean data could be important.

Also tests just look at whether one is higher or lower – the difference in each case could be negligible (e.g. 25.001 degrees vs 25.002 degrees)

A non-standard measurement method might have been used in Rocky Bay, which confounds the comparison.

For the signs test the continuity correction was not required, but if done has to be correct. Candidates were given credit for a one-sided signs test in (i) provided that they set the null hypothesis up correctly – i.e. that the average maximum temperature in Rocky Bay in August is ***no lower*** *than that in the Mediterranean. In (ii) other sensible comments were given credit, and the maximum score of 2 marks could be obtained for making four sensible points – not all the points listed above were required.*

Question 9

(i) Type I (right censoring) of patients who survive to duration 5 years.

Random censoring of patients who withdraw from the study.

(ii) Since $S(t) = \exp(-\Lambda_t)$ where $\Lambda_t = \sum_{t_j \le t} \left(\frac{d_j}{n_j} \right)$

$\Lambda_t = -\ln [S(t)]$

So

Duration since operation t (years)	*S(t)*	Λ_t
$0 \le t < 1$	1	0
$1 \le t < 3$	0.9355	0.0667
$3 \le t < 4$	0.7122	0.3394
$4 \le t < 5$	0.6285	0.4644

Let d_j and n_j be the number of deaths and the number in the risk set at the jth point at which events occur.

Consider $t = 1$

$$\frac{d_1}{n_1} = 0.0667$$

Since there can be no more than 16 patients at risk at $t = 1$, the only possible combination is $d_1 = 1$ and $n_1 = 15$

Consider $t = 3$ (the second point at which events occur)

$$\frac{d_2}{n_2} = 0.3394 - 0.0667 = 0.2727$$

Recognising this as 3/11, and that there are at most 14 patients at risk, this implies that $d_2 = 3$ and $n_2 = 11$.

Consider $t = 4$ (the third point at which events occur)

$$\frac{d_3}{n_3} = 0.4644 - 0.0667 - 0.2727 = 0.125$$

Recognising this as 1/8, and that there are at most 11 patients at risk, this implies that $d_3 = 1$ and $n_3 = 8$.

So the answer is:

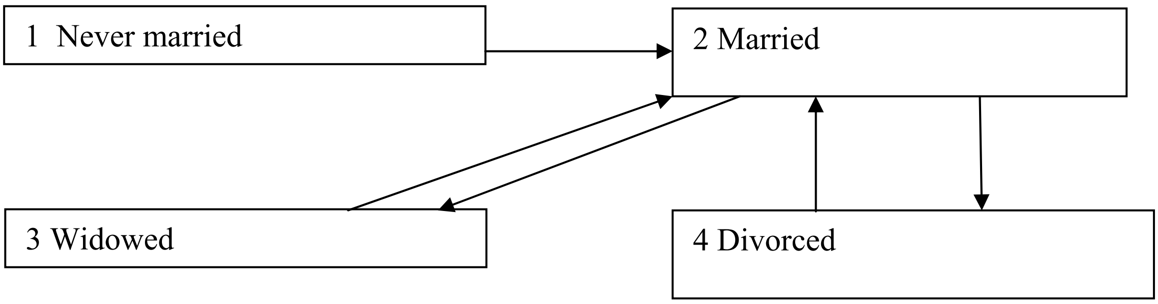
1 death at duration	1 year
3 deaths at duration	3 years
1 death at duration	4 years

(iii) Patients either die or are censored. As the total number of patients is 16 and 5 die the number censored is 16 – 5 = 11.

This was the best answered question on the examination paper. In (i) "right" censoring was awarded credit, as was some explanation of whether the censoring was informative or non-informative. In (ii) a common error was to state the durations as ranges (i.e. 1 death at durations between 1 and 3 years, 3 deaths at durations between 3 and 4 years, and 1 death at durations over 4 years). This reveals a misunderstanding of the estimator, and was penalised by the loss of 1 mark. Candidates who calculated an incorrect number of deaths in (ii) were given credit for (iii) if their answer to (iii) was consistent with their answer to (ii).

Question 10

(i)



(ii) Using the numbering of the states above, let the probability that a women who is in state *i* at time *x* will be in state *j* at time *x* + *t* be ${}_t p_x^{ij}$.

Using the <u>Markov property</u>, and conditioning on the state occupied at time *x* + *t*,

and noting that for first marriages return from the widowed or divorced state is not possible, we can write

$${}_{t+dt}p_x^{12} = {}_t p_x^{11}\, {}_{dt}p_{x+t}^{12} + {}_t p_x^{12}\, {}_{dt}p_{x+t}^{22}$$

Using the law of total probability, ${}_{dt}p_{x+t}^{22} = 1 - {}_{dt}p_{x+t}^{23} - {}_{dt}p_{x+t}^{24}$,

so that

$${}_{t+dt}p_x^{12} = {}_t p_x^{11}\, {}_{dt}p_{x+t}^{12} + {}_t p_x^{12}(1 - {}_{dt}p_{x+t}^{23} - {}_{dt}p_{x+t}^{24})$$

Let the transition rate from state *i* to state *j* at time *x*+*t* be μ_{x+t}^{ij}.

Assume that ${}_{dt}p_{x+t}^{ij} = \mu_{x+t}^{ij}dt + o(dt),\ \ i \neq j$

where $\lim_{dt \to 0^+} \frac{o(dt)}{dt} = 0$.

Substituting for the ${}_t p_x^{ij}$ in the equation above produces

$${}_{t+dt} p_x^{12} = {}_t p_x^{11} \mu_{x+t}^{12} dt + {}_t p_x^{12} (1 - \mu_{x+t}^{23} dt - \mu_{x+t}^{24} dt) + o(dt)$$

Therefore

$${}_{t+dt} p_x^{12} - {}_t p_x^{12} = {}_t p_x^{11} \mu_{x+t}^{12} dt - {}_t p_x^{12} \mu_{x+t}^{23} dt - {}_t p_x^{12} \mu_{x+t}^{24} dt + o(dt)$$

and, taking limits, we have

$$\lim_{dt \to 0^+} \frac{{}_{t+dt} p_x^{12} - {}_t p_x^{12}}{dt} = {}_t p_x^{11} \mu_{x+t}^{12} - {}_t p_x^{12} \mu_{x+t}^{23} - {}_t p_x^{12} \mu_{x+t}^{24}$$

So

$$\frac{d}{dt} {}_t p_x^{12} = {}_t p_x^{11} \mu_{x+t}^{12} - {}_t p_x^{12} \mu_{x+t}^{23} - {}_t p_x^{12} \mu_{x+t}^{24}$$

(iii) Let the waiting time in state i be v_i,

and the number of transitions from state i to state j be d_{ij}

and the transition intensity from state i to state j *is* μ_{ij}

Then the likelihood, L, may be written

$$L = K \exp[-v_1\mu_{12} - v_2(\mu_{23} + \mu_{24}) - v_3\mu_{32} - v_4\mu_{42}]\mu_{12}{}^{d_{12}}\mu_{23}{}^{d_{23}}\mu_{24}{}^{d_{24}}\mu_{32}{}^{d_{32}}\mu_{42}{}^{d_{42}}.$$

(iv) The logarithm of the likelihood is

$$\log_e L = \log_e K - v_1\mu_{12} - v_2(\mu_{23} + \mu_{24}) - v_3\mu_{32} - v_4\mu_{42}$$
$$+ d_{12}\log_e \mu_{12} + d_{23}\log_e \mu_{23} + d_{24}\log_e \mu_{24} + d_{32}\log_e \mu_{32} + d_{42}\log_e \mu_{42}$$

Differentiating with respect to μ_{12} gives

$$\frac{\partial L}{\partial \mu_{12}} = -v_1 + \frac{d_{12}}{\mu_{12}}.$$

Setting this equal to 0 and solving for μ_{12} gives

$$\hat{\mu}_{12} = \frac{d_{12}}{v_1}.$$

This is a maximum because $\dfrac{\partial^2 L}{\partial \mu_{12}^2} = -\dfrac{d_{12}}{\mu_{12}^2}$ which is negative.

Answers to (i), (iii) and (iv) were generally good. In (i) the arrows from Widowed to Married and Divorced to Married were not required for full marks, as the question is about first marriages. Answers to (ii) were more disappointing, with many candidates omitting steps in the argument. In (ii), some candidates included the extra terms $+{}_t p_x^{13} \mu_{x+t}^{32} + {}_t p_x^{14} \mu_{x+t}^{42}$. Since it is just about possible to interpret the question in a way such that these should be included, this was not heavily penalised.

Question 11

(i) A Markov jump process is a continuous-time Markov process with a discrete state space.

For a process to be Markov, the future development of the process must depend only on its current state.

This is the case here, as the future of the process depends only on the number of passengers currently in the front taxi.

The number of passengers in the front taxi also has a discrete state space {0, 1, 2, 3}. (Note that immediately a fourth passenger arrives the taxi will depart so the front taxi in the queue will never have four passengers in it.)

(ii) (a) The generator matrix A is

$$\begin{pmatrix} -\beta & \beta & 0 & 0 \\ 0 & -\beta & \beta & 0 \\ 0 & 0 & -\beta & \beta \\ \beta & 0 & 0 & -\beta \end{pmatrix}$$

(b) Kolmogorov's forward equations can be written in compact form as

$$\frac{d}{dt} P(t) = P(t) A,$$

Which are, for $j = 0$

$$\frac{d}{dt} p_{i0}(t) = \beta p_{i3}(t) - \beta p_{i0}(t)$$

and, for j = 1,2,3

$$\frac{d}{dt} p_{ij}(t) = \beta p_{i,j-1}(t) - \beta p_{ij}(t) .$$

(iii) Since the waiting times under a Poisson process are exponential the expected waiting time between the arrival of passengers at the terminus is $\frac{1}{\beta}$ minutes.

Successive waiting times are independent, therefore the expected waiting time for a passenger arriving at the terminus is

$$E[t] = \sum_{i=0}^{3} p_i \frac{3-i}{\beta},$$

where p_i is the probability that the front taxi has exactly i previous passengers waiting in it when the passenger arrives.

Since the p_is are all equal for i = 0, 1, 2, 3

$$E[t] = 0.25\left(\frac{3}{\beta}+\frac{2}{\beta}+\frac{1}{\beta}+\frac{0}{\beta}\right) = \frac{3}{2\beta} \text{ minutes.}$$

(iv) The transition matrix, P, is

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

(v) The expected waiting time if the front taxi is a three-passenger model is

$$E[t \mid 3-\text{passenger model}] = \sum_{i=0}^{2} p_i \frac{2-i}{\beta} = \frac{1}{3}\left(\frac{2}{\beta}+\frac{1}{\beta}+\frac{0}{\beta}\right) = \frac{1}{\beta}$$

The expected waiting time if the front taxi is a five-passenger model is

$$E[t \mid 5-\text{passenger model}] = \sum_{i=0}^{4} p_i \frac{4-i}{\beta} = \frac{1}{5}\left(\frac{4}{\beta}+\frac{3}{\beta}+\frac{2}{\beta}+\frac{1}{\beta}+\frac{0}{\beta}\right) = \frac{2}{\beta}.$$

But 5-passenger models must expect to wait $\frac{5}{3}$ times as long at the front of the queue than do 3-passenger models.

So when a passenger arrives at the terminus, $\frac{5}{8}$ of the time the taxi at the front of the queue will be a five-passenger model and only $\frac{3}{8}$ of the time will is be a three-passenger model.

So the overall expected waiting time in minutes is

$$\frac{3}{8}(E[t \mid 3-\text{passenger model}]) + \frac{5}{8}(E[t \mid 5-\text{passenger model}]) = \frac{13}{8\beta}.$$

As this is longer than $\frac{3}{2\beta}$, the service provided to the passengers has deteriorated.

Many candidates struggled with this question. A common error in (ii) was to draw a matrix with five states rather than four, failing to recognise that taxis with four passengers in do not wait at the front of the queue, but depart as soon as the fourth passenger arrives. Most candidates who attempted (iv) wrote down a generator matrix, whereas the question asked for a transition matrix.

Question 12

(i)

Start previous day	*Start morning* 1	2	3	4
1	0.4	0	0	0.6
2	0.4	0.4	0	0.2
3	0.2	0.4	0.4	0
4	0	0.2	0.4	0.4

(ii) If stationary distribution is $\underline{\pi} = \begin{pmatrix}\pi_1 & \pi_2 & \pi_3 & \pi_4\end{pmatrix}$

Then $\underline{\pi}A = \underline{\pi}$ where A is the matrix in (i)

$0.4\pi_1 + 0.4\pi_2 + 0.2\pi_3 = \pi_1$ (a)
$0.4\pi_2 + 0.4\pi_3 + 0.2\pi_4 = \pi_2$ (b)
$0.4\pi_3 + 0.4\pi_4 = \pi_3$ (c)
$0.6\pi_1 + 0.2\pi_2 + 0.4\pi_4 = \pi_4$ (d)

From (c) $\pi_3 = 0.666666\pi_4$
From (b) $\pi_2 = 0.7778\pi_4$
From (a) $\pi_1 = 0.7407\pi_4$

$\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1 = (0.7407 + 0.7778 + 0.6666 + 1)\pi_4$

Implies $\pi_1 = 0.2325$, $\pi_2 = 0.2442$, $\pi_3 = 0.2093$, $\pi_4 = 0.31395$

OR $\pi_1 = \frac{10}{43}$, $\pi_2 = \frac{21}{86}$, $\pi_3 = \frac{9}{43}$, $\pi_4 = \frac{27}{86}$.

(iii) Probability of restocking is 0.6 if in π_1 and 0.2 if in π_2

So long term rate = 0.6 * 0.2325 + 0.2 * 0.2442 = 0.1884 per trading day

(iv) Probability of losing a sale is 0.2 if in π_1

So expected lost sales per day = 0.2 * 0.2325 = 0.0465

(v) If restock when fewer than two in stock then transition matrix changes to:

Start previous day	*Start morning* 2	3	4
2	*0.4*	*0*	*0.6*
3	0.4	0.4	0.2
4	0.2	0.4	0.4

Label stationary distribution $\underline{\lambda}$. Then

$0.4\lambda_2 + 0.4\lambda_3 + 0.2\lambda_4 = \lambda_2$ (b1)

$0.4\lambda_3 + 0.4\lambda_4 = \lambda_3$ (c1)

$0.6\lambda_2 + 0.2\lambda_3 + 0.4\lambda_4 = \lambda_4$ (d1)

From (c1) $\lambda_3 = 0.666666\lambda_4$

From (b1) $\lambda_2 = 0.7778\lambda_4$

$\lambda_2 = 0.3182$ OR 7/22

$\lambda_3 = 0.2727$ OR 3/11

$$\lambda_4 = \frac{1}{(1+2/3+7/9)} = 0.4091 \quad \text{OR} \quad 9/22$$

As no more than two snakes sell per day,

there are no lost sales.

Probability of restocking 0.6 if in λ_2 and 0.2 in λ_3 = 0.2455

(vi) Restocking at two or more snakes would not result in fewer lost sales than restocking at 1.

Because the probability of selling more than 2 snakes is zero.

It would, however, result in more restocking charges than restocking at 1.

Therefore it must result in lower profits than restocking at 1 so is not optimal.

(vii) Costs if restock at 0

$0.1884C + 0.0465P$

Costs if restock at 1

$0.24546C$

So should change restocking approach if

$0.24546C < 0.1884C + 0.0465P$

$C < 0.8148P$

In this question, many candidates answered (i) and (ii), but made no further progress. Candidates who wrote down the wrong matrix in (i) but evaluated the stationary distribution correctly for the matrix they had written down were given full credit for (ii), and gained credit in (iii) and (iv) if these parts were answered correctly given the matrix which had been written down in (i). A common error was to write down a five-state model in (i). Few candidates attempted the later sections of this question.

END OF EXAMINERS' REPORT