INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Class: M.Sc. SEM 3

Subject: Statistical and Risk Modelling 3

Chapter: Unit 2 Chapter 1

Chapter Name: Markov Chains – Practice questions

5 Modelling using Markov Chains

Testing the Markov assumption

We can test the Markov assumption using a triplets test. The formula for the test statistic is:

$$\sum_{i,j,k} \frac{\left(n_{ijk} - n_{ij}\hat{p}_{jk}\right)^2}{n_{ij}\hat{p}_{jk}}$$

The null hypothesis is:

H_0 : the process has the Markov property

The alternative hypothesis is:

H_1 : the process does not have the Markov property

Under the null hypothesis, $N_{ijk} \sim Binomial(n_{ij}, p_{jk})$.

This is of the familiar form $\sum_{i,j,k} \frac{(O-E)^2}{E}$ where O represents the observed frequency and E represents the expected frequency. If $N_{ijk} \sim Binomial(n_{ij}, p_{jk})$, then $E(N_{ijk}) = n_{ij}p_{jk}$. When calculating the expected frequencies, n_{ij} is calculated as $\sum_k n_{ijk}$ to ensure that the observed and expected frequencies tally.

Question

A 3-state process has been observed over a period of time and the sequence of states occupied is as follows:

1,3,2,2,1,3,3,2,3,1,2,3,2,1,1,2,2,1,3,3

(i) Calculate the values of n_{ijk} , n_{ij} and n_i .

(ii) Estimate the one-step transition probabilities.

(iii) State the null and alternative hypotheses for the triplets test.

(iv) Calculate the test statistic for the triplets test (without combining triplets with small expected frequencies).

Solution

1,3,2,2,1,3,3,2,3,1,2,3,2,1,1,2,2,1,3,3

(i) ***Values***

The values of n_{ijk} are shown in the matrices below:

$$\left(n_{1jk}\right)=\begin{pmatrix}0&1&0\\0&1&1\\0&1&2\end{pmatrix}\qquad\left(n_{2jk}\right)=\begin{pmatrix}1&0&2\\2&0&0\\1&1&0\end{pmatrix}\qquad\left(n_{3jk}\right)=\begin{pmatrix}0&1&0\\1&1&1\\0&1&0\end{pmatrix}$$

The first row of the matrix $\left(n_{1jk}\right)$ contains the entries n_{111}, n_{112} and n_{113}; the second row consists of n_{121}, n_{122} and n_{123}, *etc*.

The value of n_{ij} is the ij th entry of the following matrix:

$$\left(n_{ij}\right)=\begin{pmatrix}1&2&3\\3&2&2\\1&3&2\end{pmatrix}$$

and the n_i values are the row sums of the matrix $\left(n_{ij}\right)$:

$$\left(n_i\right)=\begin{pmatrix}6\\7\\6\end{pmatrix}$$

$$\left(n_{ij}\right)=\begin{pmatrix}1 & 2 & 3\\ 3 & 2 & 2\\ 1 & 3 & 2\end{pmatrix}$$

and the n_i values are the row sums of the matrix $\left(n_{ij}\right)$:

$$\left(n_i\right)=\begin{pmatrix}6\\ 7\\ 6\end{pmatrix}$$

(ii) ***One-step transition probabilities***

Using the formula $\hat{p}_{ij}=\dfrac{n_{ij}}{n_i}$, we obtain the following estimates:

$$\begin{pmatrix}1/6 & 2/6 & 3/6\\ 3/7 & 2/7 & 2/7\\ 1/6 & 3/6 & 2/6\end{pmatrix}=\begin{pmatrix}1/6 & 1/3 & 1/2\\ 3/7 & 2/7 & 2/7\\ 1/6 & 1/2 & 1/3\end{pmatrix}=\begin{pmatrix}0.167 & 0.333 & 0.5\\ 0.429 & 0.286 & 0.286\\ 0.167 & 0.5 & 0.333\end{pmatrix}$$

(iv) ***Chi-squared test***

The formula for the test statistic is:

$$\sum_{i,j,k} \frac{\left(n_{ijk} - n_{ij}\hat{p}_{jk}\right)^2}{n_{ij}\hat{p}_{jk}}$$

where n_{ij} is calculated using the formula $n_{ij} = \sum_k n_{ijk}$. This gives rise to the same values for n_{ij} as shown in the matrix in (i), except for n_{33}, as the string of observations ends with 3,3. Using this formula gives $n_{33} = 1$. The final 3,3 in the list of observations is not counted here because this is unable to give rise to an observation of the form 3,3,*k*.

We have:

ijk	Observed frequency n_{ijk}	n_{ij}	$\hat{p}_{jk}$	Expected frequency $n_{ij}\,\hat{p}_{jk}$	$\dfrac{(n_{ijk}-n_{ij}\hat{p}_{jk})^2}{n_{ij}\hat{p}_{jk}}$
111	0	1	0.167	0.167	0.167
112	1	1	0.333	0.333	1.333
113	0	1	0.500	0.500	0.500
121	0	2	0.429	0.857	0.857
122	1	2	0.286	0.571	0.321
123	1	2	0.286	0.571	0.321
131	0	3	0.167	0.500	0.500
132	1	3	0.500	1.500	0.167
133	2	3	0.333	1.000	1.000
211	1	3	0.167	0.500	0.500
212	0	3	0.333	1.000	1.000
213	2	3	0.500	1.500	0.167
221	2	2	0.429	0.857	1.524
222	0	2	0.286	0.571	0.571
223	0	2	0.286	0.571	0.571
231	1	2	0.167	0.333	1.333
232	1	2	0.500	1.000	0.000
233	0	2	0.333	0.667	0.667
311	0	1	0.167	0.167	0.167
312	1	1	0.333	0.333	1.333
313	0	1	0.500	0.500	0.500
321	1	3	0.429	1.286	0.063
322	1	3	0.286	0.857	0.024
323	1	3	0.286	0.857	0.024
331	0	1	0.167	0.167	0.167
332	1	1	0.500	0.500	0.500
333	0	1	0.333	0.333	0.333

The observed value of the test statistic is the sum of the numbers in the final column, *ie*:

$$\sum_{i,j,k}\frac{(n_{ijk}-n_{ij}\hat{p}_{jk})^2}{n_{ij}\hat{p}_{jk}}=14.61$$

Recall that, when carrying out a chi-squared test, the expected frequencies should ideally all be 5 or more. If this is not the case, the validity of the test is questionable.

Question

2. CT4 April 2011 Q4

Children at a school are given weekly grade sheets, in which their effort is graded in four levels : 1"Poor", 2 "Satisfactory", 3"Good" and 4 "Excellent". Subject to a maximum level of Excellent and a minimum level of Poor, between each week and the next, a child has :

- a 20 per cent chance of moving up one level
- a 20 per cent chance of moving down one level
- a 10 per cent chance of moving up two level
- a 10 per cent chance of moving down two level

moving up or down three levels in a single week is not possible.

(i) Write down the transition matrix of his process.

Children are graded on Friday afternoon in each week. On Friday of the first week of the school year, as there is little evidence on which to base an assessment, all children are graded "Satisfactory".

(ii) Calculate the probability distribution of the process after the grading on Friday of the third week of the school year.

Solution

(i) $$\begin{pmatrix} 0.7 & 0.2 & 0.1 & 0 \\ 0.3 & 0.4 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.4 & 0.3 \\ 0 & 0.1 & 0.2 & 0.7 \end{pmatrix}$$

There were two common errors on this question. The first was to assume that if a child could not move up or down two levels, he or she would not move at all. The phrase in the question "[s]ubject to a maximum level of Excellent and a minimum level of Poor" was intended to indicate that children could not move beyond these limits in either direction, but would move as far as they could. Thus a child at level "Good", who had a 20% chance of moving up one level and a 10% chance of moving up two levels, would have a 30% chance of moving to level Excellent, as the 10% who would have moved up two levels will only be able to move up one level. The second error was to use ΠM^3 *in part (ii). Candidates who made the first error were penalised in part (i) but could gain full credit for part (ii) if they followed through correctly.*

(i) $\begin{pmatrix} 0.7 & 0.2 & 0.1 & 0 \\ 0.3 & 0.4 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.4 & 0.3 \\ 0 & 0.1 & 0.2 & 0.7 \end{pmatrix}$

(ii) If the probability distribution in the first week is Π, and the transition matrix is M, then the probability distribution at the end of the third week is

$$\Pi M^2 = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.7 & 0.2 & 0.1 & 0 \\ 0.3 & 0.4 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.4 & 0.3 \\ 0 & 0.1 & 0.2 & 0.7 \end{pmatrix} \begin{pmatrix} 0.7 & 0.2 & 0.1 & 0 \\ 0.3 & 0.4 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.4 & 0.3 \\ 0 & 0.1 & 0.2 & 0.7 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.56 & 0.24 & 0.15 & 0.05 \\ 0.35 & 0.27 & 0.21 & 0.17 \\ 0.17 & 0.21 & 0.27 & 0.35 \\ 0.05 & 0.15 & 0.24 & 0.56 \end{pmatrix}$$

so that there is a probability of

35% that a child will be graded Poor',
27% that a child will be graded Satisfactory,
21% that a child will be graded Good and
17% that a child will be graded Excellent..

Question

4. CT4 September 2012 Q5

A no claims discount system operates with three levels of discount, 0%, 15% and 40%. If a policyholder makes no claim during the year he moves up a level of discount (or remains at the maximum level). If he makes one claim during the year he moves down one level of discount (or remains at the minimum level) and if he makes two or more claims he moves down to, or remains at, the minimum level.

The probability for each policyholder of making two or more claims in a year is 25% of the probability of making only one claim.

The long-term probability of being at the 15% level is the same as the long-term probability of being at the 40% level.

(i) Derive the probability of a policyholder making only one claim in a given year. [4]

(ii) Determine the probability that a policyholder at the 0% level this year will be at the 40% level after three years. [2]

(iii) Estimate the probability that a policyholder at the 0% level this year will be at the 40% level after 20 years, without calculating the associated transition matrix. [3]

[Total 9]

Solution

(i) Let $x = \frac{5}{4}c$

where c is the probability of exactly one claim in a year and x is the probability of one or more claims in a year.

The transition matrix is

$$\begin{pmatrix} x & 1-x & 0 \\ x & 0 & 1-x \\ \frac{c}{4} & c & 1-x \end{pmatrix}$$

Using $\pi = \pi P$ we get

$$\pi_1 = x\pi_1 + x\pi_2 + {}^{c}\!/_{4}\,\pi_3$$
$$\pi_2 = (1-x)\pi_1 + c\pi_3$$
$$\pi_3 = (1-x)\pi_2 + (1-x)\pi_3$$

The equation for π_3 gives

$$\pi_2(1-x) = \pi_3\{1-(1-x)\} = \pi_3 x$$

$$\pi_2 = \pi_3 \frac{x}{1-x}$$

Long term prob $\pi_2 = \pi_3$

So $x = 1 - x$ from which $x = 0.5$ and $c = 0.4$

So the probability of exactly one claim in any given year is 0.4.

(ii) EITHER

Using the transition matrix

$$M = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}$$

$$M^2 = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}\begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0.1 & 0.4 & 0.5 \end{pmatrix} = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.3 & 0.45 & 0.25 \\ 0.3 & 0.25 & 0.45 \end{pmatrix}$$

The required probability is therefore

$(0.5 \times 0.25) + (0.5 \times 0.25) + (0 \times 0.45) = 0.25$

OR

We require the probability of no claims in either of years 2 and 3 (since only this will leave the policyholder at the 40% level at the end of year 3).

The probability of one or more claims is 0.5 (from the solution to part (i)).

So the probability of no claims is 0.5, and the probability of no claims in years 2 and 3 is $0.5 \times 0.5 = 0.25$.

(iii) After 20 years the probabilities of being at any level will be close to the stationary probability distribution

From part (i) we know that $\pi_2 = \pi_3$.

Using $\pi = \pi P$ we get

$$0.5\pi_1 + 0.5\pi_2 + 0.1\pi_3 = \pi_1,$$

so $\pi_2 = \dfrac{5}{6}\pi_1$.

Since $\pi_1 + \pi_2 + \pi_3 = 1$,+½

we have $\pi_1 = \dfrac{3}{8}$, $\pi_2 = \pi_3 = \dfrac{5}{16}$.

So the probability of being at the 40% level after 20 years is estimated as 0.3125.

This question proved more difficult for candidates that the Examiners had envisaged, and answers were disappointing. Various alternative specifications of the matrix in part (i) were acceptable. In all three parts of this question some indication of how each result was obtained was required. Candidates who just wrote down the numerical answers did not score full credit. The solution to part (ii) could be found by drawing a diagram and tracing the possible routes through: this is perfectly valid and is arguably the quickest way to the correct answer. In part (iii) some indication that the answer is an estimate was required. This could be provided by saying, for example, that after 20 years the probabilities of being at any level will be close to the stationary probability distribution.

Question

CT4 September 2013 Q2

The two football teams in a particular city are called United and City and there is intense rivalry between them. A researcher has collected the following history on the results of the last 20 matches between the teams from the earliest to the most recent, where:

U indicates a win for United;

C indicates a win for City;

D indicates a draw.

UCCDDUCDCUUDUDCCUDCC

The researcher has assumed that the probability of each result for the next match depends only on the most recent result. He therefore decides to fit a Markov chain to this data.

(i) Estimate the transition probabilities for the Markov chain. [3]

(ii) Estimate the probability that United will win at least two of the next three matches against City. [3]

[Total 6]

Solution

(i) Need to rearrange data as tally chart of next states:

Previous state	*Number where next state is:* *U*	*C*	*D*
U	1	11	111
C	11	111	11
D	11	111	1

So the transition probabilities are estimated as:

From/To	*U*	*C*	*D*
U	1/6	1/3	1/2
C	2/7	3/7	2/7
D	1/3	1/2	1/6

(ii) The possible sequences with at least 2 wins for United are:

UUU, UUC, UUD, DUU, CUU, UDU, UCU

The probabilities if the last match was won by City are:

UUU = 2/7*1/6*1/6 = 1/126
UUC = 2/7*1/6*1/3 = 1/63
UUD = 2/7*1/6*1/2 = 1/42
DUU = 2/7*1/3*1/6 = 1/63
CUU = 3/7*2/7*1/6 = 1/49
UDU = 2/7*1/2*1/3 = 1/21
UCU = 2/7*1/3*2/7 = 4/147

OR (quicker)

UUX = 2/7*1/6 = 1/21
DUU = 2/7*1/3*1/6 = 1/63
CUU = 3/7*2/7*1/6 = 1/49
UDU = 2/7*1/2*1/3 = 1/21
UCU = 2/7*1/3*2/7= 4/147

where X refers to any result

Total = 140/882 = 10/63 = 0.15873

Answers to this question were generally disappointing. In both parts (i) and (ii) the question said "estimate" so some explanation of where the answer is coming from was required for full credit (e.g. in part (i) a statement that n_{ij}/n_i is needed, or a suitable diagram were acceptable). A common error was to use 8 as the denominator for the C row. A more serious error was to use 19 as the denominator for all the transition probabilities. Many candidates did not take account of the fact that City had won the last match in the string given and thus only used pairs, rather than triplets, of probabilities.

Question

CT4 September 2015 Q10

A profession has examination papers in two subjects, A and B, each of which is marked by a team of examiners. After each examination session, examiners are given the choice of remaining on the same team, switching to the other team, or taking a session's holiday.

In recent sessions, 10% of subject A's examiners have elected to switch to subject B and 10% to take a holiday. Subject B is more onerous to mark than subject A, and in recent sessions, 20% of subject B's examiners have elected to take a holiday in the next session, with 20% moving to subject A.

After a session's holiday, the profession allocates examiners equally between subjects A and B. No examiner is permitted to take holiday for two consecutive sessions.

(i) Sketch the transition graph for the process. [2]

(ii) Determine the transition matrix for this process. [2]

(iii) Calculate the proportion of the profession's examiners marking for subjects A and B in the long run. [4]

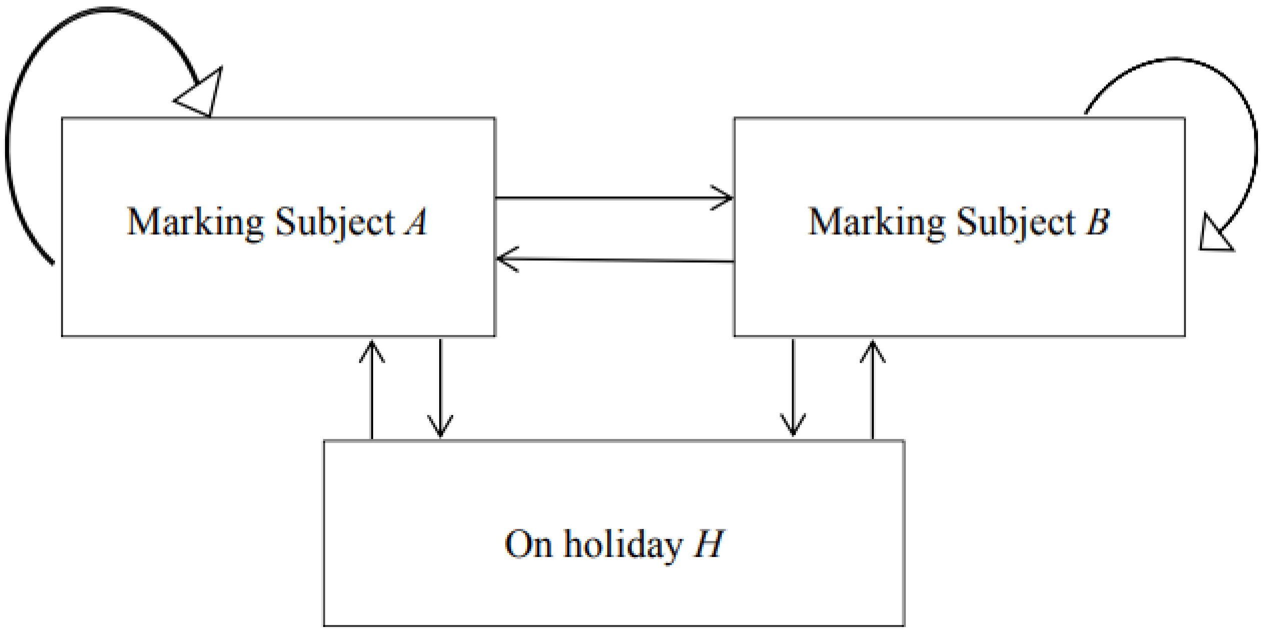
The profession considers that in future, an equal number of examiners is likely to be required for each subject. It proposes to try to ensure this by adjusting the proportion of those examiners on holiday who, when they return to marking, are allocated to subjects A and B.

(iv) Calculate the proportion of examiners who, on returning from holiday, should be allocated to subject B in order to have an equal number of examiners on each subject in the long run. [4]

[Total 12]

Solution

(i)



(ii)

$$\begin{array}{r} \text{Subject A} \\ \text{Subject B} \\ \text{Holiday} \end{array} \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.5 & 0.5 & 0 \end{bmatrix}.$$

(iii) We have $\pi P = \pi$.

The stationary distribution of examiners can be found as the solution of the set of equations

$$\pi_A = 0.8\pi_A + 0.2\pi_B + 0.5\pi_H \quad (1)$$
$$\pi_B = 0.1\pi_A + 0.6\pi_B + 0.5\pi_H \quad (2)$$
$$\pi_H = 0.1\pi_A + 0.2\pi_B \quad (3)$$

(1) gives

$$0.2\pi_A = 0.2\pi_B + 0.5\pi_H$$
$$0.4\pi_A = 0.4\pi_B + \pi_H$$
$$0.4\pi_B = 0.4\pi_A - \pi_H$$

(2) gives

$$0.4\pi_B = 0.1\pi_A + 0.5\pi_H$$

so

$$0.4\pi_A - \pi_H = 0.1\pi_A + 0.5\pi_H$$
$$0.3\pi_A = 1.5\pi_H$$
$$\pi_A = 5\pi_H$$

In (2) this gives

$$\pi_B = 0.5\pi_H + 0.6\pi_B + 0.5\pi_H$$
$$0.4\pi_B = \pi_H$$
$$\pi_B = 2.5\pi_H$$

So, since $\pi_A + \pi_B + \pi_H = 1$,

the stationary distribution is $\{5\pi_H, 2.5\pi_H, \pi_H\}$

and hence

$$\pi_A = \frac{10}{17}$$
$$\pi_B = \frac{5}{17}$$
$$\pi_H = \frac{2}{17}.$$

So in the long run 58.8% of examiners are marking subject A and 29.4% are marking subject B.

(iv) Let the new transition probability from H to A be x, and that from H to B be $1-x$.

The proportion we require is thus just x. The new transition matrix is

$$\begin{array}{r} \text{Subject A} \\ \text{Subject B} \\ \text{Holiday} \end{array} \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ x & 1-x & 0 \end{bmatrix}.$$

The stationary probability distribution is given by the three equations

$$\vartheta_A = 0.8\vartheta_A + 0.2\vartheta_B + x\vartheta_H \qquad (1)$$
$$\vartheta_B = 0.1\vartheta_A + 0.6\vartheta_B + (1-x)\vartheta_H \qquad (2)$$
$$\vartheta_H = 0.1\vartheta_A + 0.2\vartheta_B \qquad (3)$$

We also have $\vartheta_A = \vartheta_B$.

EITHER

From (3) $\vartheta_H = 0.3\vartheta_B = 0.3\vartheta_A$

Therefore the new stationary probability distribution is $\left\{\frac{10}{23}, \frac{10}{23}, \frac{3}{23}\right\}$.

In (1) we have

$10 = 8 + 2 + 3x$

Hence $x = 0$.

OR

If $\vartheta_A = \vartheta_B$ then in (1)

$$\vartheta_A = 0.8\vartheta_A + 0.2\vartheta_B + x\vartheta_H = 0.8\vartheta_A + 0.2\vartheta_A + x\vartheta_H = \vartheta_A + x\vartheta_H$$

Hence $x = 0$.

AND HENCE

All those returning from holiday will have to be allocated to subject B.

There were a gratifying number of completely correct answers to this question, and many candidates scored full marks on parts (i)–(iii). Part (iv) was more demanding, and required candidates to invert the usual question and establish what transition matrix would give rise to a particular stationary distribution.

CT4 April 2014 Q10

Question

An industrial kiln is used to produce batches of tiles and is run with a standard firing cycle. After each firing cycle is finished, a maintenance inspection is undertaken on the heating element which rates it as being in Excellent, Good or Poor condition, or notes that the element has Failed.

The probabilities of the heating element being in each condition at the end of a cycle, based on the condition at the start of the cycle are as follows:

START	*END* *Excellent*	*Good*	*Poor*	*Failed*
Excellent	0.5	0.2	0.2	0.1
Good		0.5	0.3	0.2
Poor			0.5	0.5
Failed				1

(i) Write down the name of the stochastic process which describes the condition of a single heating element over time. [1]

(ii) Explain whether the process describing the condition of a single heating element is:

(a) irreducible.
(b) periodic. [2]

(iii) Derive the probability that the condition of a single heating element is assessed as being in Poor condition at the inspection after two cycles, if the heating element is currently in Excellent condition. [2]

If the heating element fails during the firing cycle, the entire batch of tiles in the kiln is wasted at a cost of £1,000. Additionally a new heating element needs to be installed at a cost of £50 which will, of course, be in Excellent condition.

(iv) Write down the transition matrix for the condition of the heating element in the kiln at the start of each cycle, allowing for replacement of failed heating elements. [2]

(v) Calculate the long term probabilities for the condition of the heating element in the kiln at the start of a cycle. [4]

The kiln is fired 100 times per year.

(vi) Calculate the expected annual cost incurred due to failures of heating elements. [2]

The company is concerned about the cost of ruined tiles and decides to change its policy to replace the heating element if it is rated as in Poor condition.

(vii) Evaluate the impact of the change in replacement policy on the profitability of the company. [6]

[Total 19]

Solution

(i) Markov chain.

(ii) (a) It is not irreducible

because a heating element cannot move to a state of being in better condition.

(b) It is not periodic

because it can remain in each state (or any other suitable reason).

(iii) EITHER

The second order transition matrix is:

0.25	0.2	0.26	0.29
0	0.25	0.3	0.45
0	0	0.25	0.75
0	0	0	1

Hence probability in Poor condition at the second inspection is 0.26.

OR

The required probability is equal to

Prob [Excellent to Excellent to Poor] +
Prob [Excellent to Good to Poor] +
Prob [Excellent to Poor to Poor]

which is $(0.5 \times 0.2) + (0.2 \times 0.3) + (0.2 \times 0.5) = 0.26$.

(iv)

	Excellent	*Good*	*Poor*
Excellent	0.6	0.2	0.2
Good	0.2	0.5	0.3
Poor	0.5	0	0.5

(v) Long-term probabilities satisfy $\pi = \pi P$.

$$0.6\pi_E + 0.2\pi_G + 0.5\pi_P = \pi_E \quad (1)$$
$$0.2\pi_E + 0.5\pi_G = \pi_G \quad (2)$$
$$0.2\pi_E + 0.3\pi_G + 0.5\pi_P = \pi_P \quad (3)$$

Also $\pi_E + \pi_G + \pi_P = 1$.

(2)–(3) gives:

$$\pi_G = \frac{5}{8}\pi_P.$$

So $\pi_E = \frac{25}{16}\pi_P$.

Hence $\left(\frac{25}{16} + \frac{5}{8} + 1\right)\pi_P = 1$.

Stationary distribution $\pi_E = \frac{25}{51}, \pi_G = \frac{10}{51}, \pi_P = \frac{16}{51}$.

(vi) The expected number of failures of heating elements is:

$$(0.1\pi_E + 0.2\pi_G + 0.5\pi_P)*100 \quad = 24.51.$$

The cost of each failure is £1,050 so the expected cost over a year is £25,735.

(vii) The transition matrix for the condition of the element at the start of each cycle will now be:

	Excellent	*Good*
Excellent	0.8	0.2
Good	0.5	0.5

The revised stationary distribution satisfies $\rho = \rho P$

$$0.8\rho_E + 0.5\rho_G \ = \rho_E \quad (1)$$
$$0.2\rho_E + 0.5\rho_G = \rho_G \quad (2)$$

$$0.2(1-\rho_G) + 0.5\rho_G = \rho_G$$

$$\rho_E = \frac{5}{7}, \rho_G = \frac{2}{7}$$

Expected cost of failures is now:

$$(0.1\rho_E + 0.2\rho_G)*100*1050 \quad = £13,500.$$

But we also now have extra heating element replacement costs of:

$$(0.2\rho_E + 0.3\rho_G)*100*50 \quad = £1,143.$$

So overall profits have improved by:

£25,735 – £13,500 – £1,143 = £11,092.

Question

22. CT4 September 2018 Q1

Explain why each of the following matrices is, or is not, a valid transition matrix for a Markov chain.

(a) $\begin{pmatrix} 0 & 0 \\ 0.5 & 0.5 \end{pmatrix}$

(b) $\begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.3 & 0.1 & 0.6 \end{pmatrix}$

(c) $\begin{pmatrix} 0 & 0.5 & 0.5 \\ 0.2 & 0 & 0.8 \\ 0.4 & 0.6 & 0 \end{pmatrix}$

(d) $\begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & -0.1 & 0.7 \end{pmatrix}$

Solution

(a) Not a valid transition matrix, [+½]
because the first row sums to 0 not 1. [+½]

(b) Not a valid transition matrix, [+½]
because it is not square. [+½]

(c) This is a valid transition matrix. [+½]

Each row sums to one, and each entry is between 0 and 1 inclusive, and the matrix is square. [+½]

(d) Not a valid transition matrix, [+½]
because there is an entry which is less than 0. [+½]
[Total 4]

Question

CS2A September 2020 Q9

An insurance company offers annual home insurance policies in partnership with a bank. The distribution deal involves taking part in the bank's loyalty scheme called '1234'. Under '1234', a customer gets a discount when buying or renewing the policy according to how many bank accounts they hold, as follows:

Number of bank accounts	*Discount*
One	5%
Two	10%
Three	15%
Four or more	25%

An analysis of the data suggests that the transition matrix for the number of bank accounts held at annual intervals is as follows:

$$\begin{array}{l} \text{One} \\ \text{Two} \\ \text{Three} \\ \text{Four+} \end{array} \begin{pmatrix} 0.5 & 0.2 & 0.2 & 0.1 \\ 0.2 & 0.4 & 0.3 & 0.1 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0 & 0.2 & 0.2 & 0.6 \end{pmatrix}$$

A customer takes out a policy in January 2017 at the 10% discount level.

(i) Calculate the probability that the customer remains at the 10% discount level for all of their renewals up to and including 2020. [2]

(ii) Calculate the probability that the customer receives a discount of at least 15% in 2019. [2]

(iii) Which one of the following options represents the correct stationary distribution, π, of the transition matrix above? [3]

A	$\pi_1 = \frac{34}{45}\pi_3,$	$\pi_2 = \frac{8}{9}\pi_3,$	$\pi_3,$	$\pi_4 = \frac{41}{45}\pi_3$
B	$\pi_1 = \frac{34}{45}\pi_3,$	$\pi_2 = \frac{8}{9}\pi_3,$	$\pi_3,$	$\pi_4 = \pi_3$
C	$\pi_1 = \pi_3,$	$\pi_2 = \frac{8}{9}\pi_3,$	$\pi_3,$	$\pi_4 = \frac{41}{45}\pi_3$
D	$\pi_1 = \pi_3,$	$\pi_2 = \pi_3,$	$\pi_3,$	$\pi_4 = \pi_3$

(iv) Which one of the following options represents the correct average long-term level of discount? [3]

A 13.60%
B 13.75%
C 14.19%
D 14.45%

(v) Comment on the commercial implications for the insurer of the bank's loyalty scheme. [3]
[Total 13]

Solution

(i) For this to be the case, the customer's 2018, 2019 and 2020 renewals all have to be at the 10% discount level.

The probability of remaining in state with 10% discount is 0.4. [½]

Therefore, the required probability is 0.4^3 = 0.064. [1½]

(ii) This is the probability of being in states Three or Four or more in 2019.

We can obtain these probabilities from the matrix *P*^2:

$$\begin{pmatrix} 0.5 & 0.2 & 0.2 & 0.1 \\ 0.2 & 0.4 & 0.3 & 0.1 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0 & 0.2 & 0.2 & 0.6 \end{pmatrix} \cdot \begin{pmatrix} 0.5 & 0.2 & 0.2 & 0.1 \\ 0.2 & 0.4 & 0.3 & 0.1 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0 & 0.2 & 0.2 & 0.6 \end{pmatrix} = \begin{pmatrix} 0.33 & 0.24 & 0.26 & 0.17 \\ 0.24 & 0.28 & 0.3 & 0.18 \\ 0.22 & 0.24 & 0.3 & 0.24 \\ 0.08 & 0.24 & 0.26 & 0.42 \end{pmatrix}$$

Given that the customer starts in 10% discount level in 2017, the probability of a 15% discount in 2019 is 0.3 and a 25% discount in 2019 is 0.18.

OR:

Given that the customer starts in 10% discount level in 2017, the probability of a 15% discount in 2019 is:
0.2 * 0.2 + 0.4 * 0.3 + 0.3 * 0.4 + 0.1 * 0.2 = 0.3

Given that the customer starts in 10% discount level in 2017, the probability of a 25% discount in 2019 is:

0.2 * 0.1 + 0.4 * 0.1 + 0.3 * 0.2 + 0.1 * 0.6 = 0.18 [1]

Required probability is 0.48. [1]

(iii) Answer: **A** [3]

For this we require the long-term probabilities for the chain.

Stationary distribution π satisfies $\pi = \pi P$.

$$\frac{1}{2}\pi_1 + \frac{1}{5}\pi_2 + \frac{1}{5}\pi_3 = \pi_1 \qquad (1)$$

$$\frac{1}{5}\pi_1 + \frac{2}{5}\pi_2 + \frac{1}{5}\pi_3 + \frac{1}{5}\pi_4 = \pi_2 \qquad (2)$$

$$\frac{1}{5}\pi_1+\frac{3}{10}\pi_2+\frac{2}{5}\pi_3+\frac{1}{5}\pi_4=\pi_3 \qquad (3)$$

$$\frac{1}{10}\pi_1+\frac{1}{10}\pi_2+\frac{1}{5}\pi_3+\frac{3}{5}\pi_4=\pi_4 \qquad (4)$$

from (2)-(3)

$$\frac{1}{10}\pi_2-\frac{1}{5}\pi_3=\pi_2-\pi_3$$

$$\pi_2=\frac{8}{9}\pi_3$$

substitute in (1)

$$\frac{1}{2}\pi_1=\frac{8}{45}\pi_3+\frac{1}{5}\pi_3=\frac{17}{45}\pi_3$$

$$\pi_1=\frac{34}{45}\pi_3$$

substitute in (4)

$$\frac{2}{5}\pi_4=\left(\frac{34}{450}+\frac{8}{90}+\frac{1}{5}\right)\pi_3$$

$$\pi_4=\frac{41}{45}\pi_3$$

(iv) Answer: **C** [3]

As $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$

$$\left(\frac{34}{45} + \frac{40}{45} + \frac{45}{45} + \frac{41}{45}\right)\pi_3 = 1$$

$$\pi_3 = \frac{9}{32} = 0.28125$$

Hence:

$$\pi_1 = \frac{17}{80} = 0.2125$$

$$\pi_2 = \frac{1}{4} = 0.25$$

$$\pi_4 = \frac{41}{160} = 0.25625$$

Hence the average discount is:

$$0.2125*5\% + 0.25*10\% + 0.28125*15\% + 0.25625*25\% = 14.19\%$$

(v) The level of discount the insurance company gives in its home insurance is not completely within its control. [1]

The arrangement could lead to significant volumes of business for the insurance company through access to the bank's customers, especially if the loyalty scheme is popular. [1]

Customers may open a number of small accounts in order to benefit from a higher discount. [1]

Bank customers with lots of accounts may be more loyal and this might result in better persistency on the home insurance, or even better claims experience. [1]

The insurer might need to raise its "full" prices to compensate for the introduction of the discount, and hence might become uncompetitive for those purchasing insurance outside the discount scheme. [1]

If the insurer does not raise its "full" prices to compensate for the introduction of the discount, it could find large numbers of poorer risks opening a number of bank accounts to take advantage of the larger discounts. This could lead to solvency issues for the insurer. [1]

[Marks available 6, maximum 3]

[Total 13]

Part (i) was poorly answered. The two most common mistakes were for candidates to mistakenly suggest that four years of renewals were required rather than three and/or for candidates to determine the second row and second column entry of the transition matrices, P^3 or P^4.

Part (ii) was slightly better answered. Again some candidates mistakenly suggested that three years of renewals were required rather than two.

A number of candidates lost marks for not showing their workings in parts (i) and (ii).

Parts (iii) and (iv) were very well-answered.

Part (v) was very poorly answered with many candidates commenting on the commercial implications for the bank rather than the insurer. Candidates are reminded of the need to read the question carefully. Alternative comments that were clear, distinct and relevant to the context of the question were also awarded credit.

Question

CS2A September 2021 Q9

In a small country, there are only four authorised car insurance companies A, B, C and D. All car owners take out car insurance from an authorised insurance company. All policies provide cover for a period of 1 calendar year.

The probabilities of car owners transferring between the four companies at the end of each year are believed to follow a Markov chain with the following transition matrix:

$$\begin{array}{c} A \\ B \\ C \\ D \end{array} \begin{pmatrix} 1-2\alpha-\alpha^2 & \alpha & \alpha & \alpha^2 \\ \alpha & \frac{1}{2}-\alpha & \frac{1}{2}-\alpha & \alpha \\ \alpha & \frac{1}{2}-\alpha & \frac{1}{2}-\alpha & \alpha \\ 0 & \alpha & \alpha & 1-2\alpha \end{pmatrix}$$

for some parameter α.

(i) Determine the range of values of α for which this is a valid transition matrix. [4]

(ii) Explain whether the Markov chain is irreducible, including whether this depends on the value of α. [3]

The value of α has been estimated as 0.2.

Mary has just bought her policy from Company D for the first time.

(iii) Determine the probability that Mary will be covered by Company D for at least 4 years before she transfers to another insurance company. [3]

James took out a policy with Company A in January 2018. Sadly, James' car was stolen on 23 December 2020.

(iv) Determine the probability that a different company, other than Company A, covers James' car at the time it was stolen. [3]

Company A makes an offer to buy Company D. It bases its purchase price on the assumption that car owners who would previously have purchased policies from Company A or Company D would now buy from the combined company, to be called ADDA.

(v) Write down the transition matrix that will apply after the takeover if Company A's assumption about car owners' behaviour is correct. [2]

(vi) Comment on the appropriateness of Company A's assumption. [5]
[Total 20]

Solution

(i)

For the transition matrix to be valid each row should sum to 1 [½]
This holds for all values of alpha [½]
All entries of the matrix should lie between 0 and 1 inclusive [½]
Therefore:
The entries of alpha and alpha^2 require 0 <= alpha <= 1 [½]
The entries ½ - alpha and 1 - 2 * alpha require alpha <= ½ as alpha must be >=0 from above [½]
The entry 1 - 2 * alpha - alpha^2 requires alpha <= -1 + sqrt(2) as alpha must be >=0 from above [1]
Hence, overall 0 <= alpha <= sqrt(2) - 1 [½]

(ii)

If 0 < alpha <= sqrt(2) - 1 [½]

then any state can be reached from any other state and so the chain is irreducible [1]

If alpha = 0 [½]

then it's not possible to leave states A or D and so the chain is reducible [1]

(iii)

Transition matrix is:

A	0.56	0.2	0.2	0.04
B	0.2	0.3	0.3	0.2
C	0.2	0.3	0.3	0.2
D	0	0.2	0.2	0.6

[1]

For company D to provide cover to Mary for at least four years before she changes provider, Mary must renew her policy with company D at least three times [1]

The probability of renewing three times with company D is

0.6^3 = 0.216 (or 27/125) [1]

(iv)

The company covering the car on 23 December 2020 will be that securing James' business at the second renewal [1]

The probability of James being with Company A for the second renewal is the first element of the second order transition matrix, which is: [1]

$0.56 * 0.56 + 0.2 * 0.2 + 0.2 * 0.2 + 0.04 * 0 = 0.3936$ [½]

and hence the probability of James being with a different company for the second renewal is 0.6064 [½]

Original

Transition matrix is:

A	0.56	0.2	0.2	0.04
B	0.2	0.3	0.3	0.2
C	0.2	0.3	0.3	0.2
D	0	0.2	0.2	0.6

New- post ADDA

(v) Transition matrix is:

ADDA	0.6	0.2	0.2
B	0.4	0.3	0.3
C	0.4	0.3	0.3

[2]

(vi)

Observe that currently the probability of customers going from Company D to Company A is zero [1]

which suggests that there may be reasons customers of Company D do not want to use Company A [½]

There may also be reasons customers of Company A do not want to use Company D [½]

ADDA might merge its pricing system. This would change the relative pricing of an individual's cover from the different companies. To the extent that pricing is a driver of the likelihood of customers moving this might change the probabilities [1]

Economies of scale may lead to lower premiums. To the extent that pricing is a driver of the likelihood of customers moving this might change the probabilities [1]

It is not clear whether the products sold by ADDA would be the same as those previously sold by Company A or Company D. This might change the probabilities [1]

To the extent that customer service is a driver, it is not clear what the customer service of ADDA would be relative to Company A or Company D. This might change the probabilities [1]

Reduction in competition might encourage a new entrant [1]

It might be a valid assumption that customer behaviour continues unaltered after the merger [½]

[Marks available 7½, maximum 5]

[Total 20]

Additional Practice

Question

CS2A April 2022 Q8

The number of customers, N_t, in a queue at each integer time t is modelled using a Markov chain model.

At the start of each time interval, a number of customers following a Poisson distribution with parameter p join the queue, subject to the constraint that the number of customers in the queue can be no more than N_{max}.

At the end of each time interval, a number of customers following a Poisson distribution with parameter q, where q > p, are served and leave the queue, subject to the constraint that the number of customers in the queue cannot be negative.

(i) Comment on the limitations of this model. [2]

An analyst sets the model's parameters to N_{max} = 2, p = 0.5 and q = 1.

(ii) Verify, by separately considering the two transition matrices of customers joining the queue and customers leaving the queue, respectively, that the transition matrix of N_t is:

$$\begin{array}{c} 0 \\ 1 \\ 2 \end{array}\begin{pmatrix} 0.82207 & 0.14475 & 0.03318 \\ 0.48737 & 0.36788 & 0.14475 \\ 0.26424 & 0.36788 & 0.36788 \end{pmatrix}$$

(iii) Determine the stationary distribution of N_t. [8]
[Total 18]

Solution

(i)

The Poisson assumption for the distribution of the number of customers joining the queue might be violated e.g. if customers join in groups rather than individually [1]

The Poisson assumption for the distribution of the number of customers being served might be violated e.g. if there is a lower limit on the time taken for a customer to be served [1]

The assumption of time-homogeneity might be violated e.g. if more customers join the queue at certain times of day [1]

Customers might leave the queue without being served [1]

In practice, customers will join and leave the queue continuously rather than at integer times [1]

In practice, there will not be a fixed upper limit to the number of customers in the queue [1]

[Marks available 6, maximum 2]

(ii)

Transition matrix, A, for customers joining the queue:

	0	**1**	**2**
0	exp(-0.5) = 0.60653	0.5*exp(-0.5) = 0.30327	1–0.60653–0.30327 = 0.09020
1	0	exp(-0.5) = 0.60653	1–0.60653 = 0.39347
2	0	0	1

[3]

Transition matrix, B, for customers leaving the queue:

	0	**1**	**2**
0	1	0	0
1	1–0.36788 = 0.63212	exp(-1) = 0.36788	0
2	1–0.36788–0.36788 = 0.26424	1*exp(-1) = 0.36788	exp(-1) = 0.36788

[3]

Hence, transition matrix for $N\_t = (A)(B)$: [1]

	0	**1**	**2**
0	0.82207	0.14475	0.03318
1	0.48737	0.36788	0.14475
2	0.26424	0.36788	0.36788

[1]

(iii)

The probabilities pi(i) for the stationary distribution satisfy:

0.82207 * pi(0) + 0.48737 * pi(1) + 0.26424 * pi(2) = pi(0) (1)

0.14475 * pi(0) + 0.36788 * pi(1) + 0.36788 * pi(2) = pi(1) (2)

0.03318 * pi(0) + 0.14475 * pi(1) + 0.36788 * pi(2) = pi(2) (3) [1]

Also pi(0) + pi(1) + pi(2) = 1 (4) [1]

Substituting (4) into (1) and (2):

0.82207 * (1 – pi(1) – pi(2)) + 0.48737 * pi(1) + 0.26424 * pi(2)

= 1 – pi(1) – pi(2) [½]

0.14475 * (1 – pi(1) – pi(2)) + 0.36788 * pi(1) + 0.36788 * pi(2) = pi(1) [½]

Giving:

0.66530 * pi(1) + 0.44217 * pi(2) = 0.17793 (5) [1]
0.77687 * pi(1) – 0.22313 * pi(2) = 0.14475 (6) [1]

From 0.22313 * (5) + 0.44217 * (6), 0.49196 * pi(1) = 0.10371 [½]
i.e. pi(1) = 0.2108 [½]

Substituting in (5), 0.66530 * 0.2108 + 0.44217 * pi(2) = 0.17793 [½]
i.e. pi(2) = 0.0852 [½]

Substituting in (4), pi(0) = 1 – 0.2108 – 0.0852 = 0.7040 [1]
Hence, the required stationary distribution is (0.7040, 0.2108, 0.0852)

[Total 18]

Part (i) was surprisingly poorly answered, despite the fact that a wide range of limitations would have gained credit. Many candidates stated assumptions of the model, with no reason given as to why they may be violated, which did not constitute limitations. Candidates are reminded of the need to read the question carefully. Alternative comments that were clear, distinct and relevant to the context of the question were also awarded credit.

Part (ii) was very poorly answered. Many candidates obtained the correct probabilities from the Poisson distributions but were unable to insert the probabilities into the correct cells of the transition matrices. Many candidates also failed to recognise that, since the rows of a transition matrix must sum to 1, the last column of the matrix A and the first column of the matrix B should be calculated as 1 less the sum of the other columns.

Part (iii) was fairly well answered, although some candidates lost marks for not showing sufficient workings to demonstrate that a valid method had been used.