EXAMINERS' REPORT

April 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.

R D Muckart
Chairman of the Board of Examiners

July 2010

Comments

Comments on solutions presented to individual questions for the April 2010 examination paper are given below. In general, those using this report should be aware that in the case of non-numerical answers full credit could often be obtained for rather less than is given in the solutions which follow. The solutions are meant as a guide to the various points which could have been made and considered relevant.

Questions without comments in this section were generally well answered, and no specific issues were identified.

Q3 *A common error was to confuse a Markov Jump Chain with a Markov Jump Process. A Markov Jump Chain has a discrete time set, whereas the corresponding Markov Jump Process has a continuous time set.*

Q4 *This was poorly answered. In part (i), many candidates merely gave definitions of the terms "periodic" and "irreducible", rather than applying them to the question. In part (ii), many candidates simply drew the three states with arrows denoting all possible transitions between them.*

Q5 *Answers to this question were disappointing. Many candidates simply wrote down general lists of the advantages and disadvantages of models, without reference to the problem and the modelling strategy described in the question. Such attempts were given little credit.*

Q6 *This was a fairly difficult exposed-to-risk question and many candidates found it challenging. A common error was to include only the first quarter of each employee's tour of duty in the exposed-to-risk. Many candidates assumed that all accidental deaths happened at the end of each quarter. This seems unrealistic and was penalised, though credit was given for computations of the exposed-to-risk that were correct given this assumption. In part (ii), a large number of candidates made no sensible attempt to analyse their own assumptions made in part (i).*

Q8 *Parts (iii) and (iv) of this troubled most candidates. Only a minority realised that, since the Kaplan-Meier estimator is a step function, the point at which S(t) attains the value must lie on one of the "risers" of the steps and therefore be at one of the event durations. In part (iv), many candidates realised that the median estimated in part (iii) included the candidates who had not qualified by the last session of 2009, whereas the median in part (i) did not, but were unable to argue coherently that this meant that the median in part (i) was biased, and under-estimated the true median.*

Q9 *In part (iv), a disturbingly large number of candidates (the majority) wrote that since the parameter was positive, the life expectancy must have increased. In fact the opposite is the case. The positive parameter increases the hazard, which leads to a greater risk of death and hence a decline in the birds' life expectancy in the new enclosure. In part (v) a common error was to assume that 0.1 was the probability of survival, rather than the probability of being killed.*

Q10 *In part (iv) a common error was to assume that every member makes exactly one flight. This produces a profit per member of £2.67 compared with the true profit of at least £0.67, and more if some members make more than two flights per year.*

Q11 *This difficult question was a challenge for almost all candidates. In part (iv) many candidates simply wrote down the probability rather than deriving it. Credit was given for attempts to part (v) which made use of an integrating factor.*

Q12 *In part (iii), many candidates chose the Signs Test. Since from the answer to part (ii) there are five consecutive negative signs followed by five consecutive positive signs it is clear by inspection that the experience will "pass" the Signs Test and so carrying it out is not appropriate. Hence no credit was given for the Signs Test in part (iii)(a). It is much more sensible to conduct a Grouping of Signs Test. As has been the case in previous examinations, the numerical aspects of the tests were generally well performed, but the descriptions of the tests and explanations of what was being done and why were less consistent.*

1 *Sex*
Age
Type of policy
Smoker/non-smoker status
Level of underwriting OR lifestyle/participation in dangerous sports
Duration in force
Sales channel
Policy size
Occupation of policyholder
Known impairments
Post code OR region/county/country OR address

Marks were given for up to four factors from the list above.

2 $E[T_x] = \dot{e}_x = \int_0^{\omega-x} {}_t p_x dt$ OR $E[T_x] = \dot{e}_x = \int_0^{\omega-x} t\, {}_t p_x \mu_{x+t} dt$

$$\text{Var}[T_x] = \left\{\int_0^{\omega-x} t^2\, {}_t p_x \mu_{x+t} dt\right\} - \dot{e}_x{}^2$$

The upper limits to the integrals can also be anything above ω-x, for example ω or ∞, since any age above ω-x just adds zero to the summation.

3

	State Space	*Time Set*
Counting Process	Discrete	Discrete or Continuous
General Random Walk	Discrete or Continuous	Discrete
Compound Poisson Process	Discrete or Continuous	Continuous
Poisson Process	Discrete	Continuous
Markov Jump Chain	Discrete	Discrete

4 (i) As periodic and irreducible then all states are periodic, hence probability of staying in any state is zero.

By law of total probability, $P_{AA} + P_{AB} + P_{AC} = 1$.

But $P_{AB} = P_{AC}$ and $P_{AA} = 0$ so $P_{AB} = P_{AC} = 0.5$.

To be irreducible at least one of P_{BA} or P_{CA} must be greater than zero.

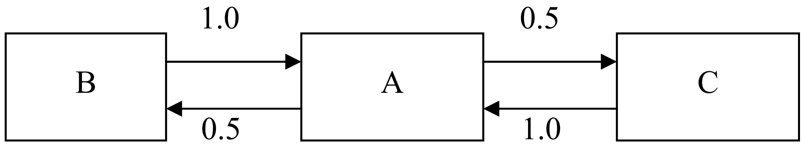
If $P_{BA} > 0$ then to be periodic must have $P_{CB} = 0$,

and to be irreducible $P_{CA} > 0$,

and if $P_{CA} > 0$ then to be periodic must have $P_{BC} = 0$, and to be irreducible $P_{BA} > 0$.

So must have $P_{BC} = P_{CB} = 0$ and $P_{BA} = P_{CA} = 1$.

(ii)



5 **Advantages**

The model is simple to understand and to communicate.

The model takes account of one major source of variation in consumption rates, specifically age.

The model is easy and cheap to implement.

The past data on consumption rates by age are likely to be fairly accurate.

The model can be adapted easily to different projected populations OR takes into account future changes in the population.

Disadvantages

Past trends in consumption by age may not be a good guide to future trends.

Extrapolation of past age-specific consumption rates may be complex or difficult and can be done in different ways.

Consumption of chocolate may be affected by the state of the economy, e.g. whether there is a recession.

Factors other than age may be important in determining consumption, e.g. expenditure on advertising.

Consumption may be sensitive to pricing, which may change in the future.

A rapid increase in consumption rates is unlikely to be sustained for a long period as there is likely to be an upper limit to the amount of Scrummy Bars a person can eat.

The projections of the future population by age may not be accurate, as they depend on future fertility, mortality and migration rates.

The proposed strategy does not include any testing of the sensitivity of total demand to changes in the projected population, or variations in future consumption trends from that used in the model.

Unforeseen events such as competitors launching new products, or the nation becoming increasingly health-aware, may affect future consumption.

The consumption of Scrummy Bars may vary with cohort rather than age, and the model does not capture cohort effects.

Not all the points listed above were required for full credit. Other advantages, for example those related to business prospects, were also given credit.

6 (i) A central exposed to risk for each quarter in person-quarters can be constructed as follows.

In the first quarter there are 90 employees in the first three months of their six-month tour of duty. Of these 10 will die during the quarter, and these contribute 0.5 each to the exposed to risk.

Therefore the total exposed to risk for the first quarter is $80 + (10 \times 0.5) = 85$ person-quarters.

This assumes that accidental deaths occur on average half way through the quarter in which they were reported. OR that accidental deaths are uniformly distributed across quarters.

In the second quarter there are 80 new employees in the first three months of their six-month tour, and 80 (90 minus the 10 who have died) employees in the second three months of their six-month tour. Of these 8 die during the quarter, and these contribute 0.5 each to the exposed to risk.

Therefore the total exposed to risk for the second quarter is $152 + (8 \times 0.5) = 156$ person-quarters

In the third quarter there are 114 new employees in the first three months of their six-month tour, and 76 (the 80 who were new on 1 April 2009 minus half of the 8 who died in the second quarter) employees in the second three months of their six-month tour. Of these 10 die during the quarter, and these contribute 0.5 months each to the exposed to risk.

This assumes that accidental deaths are equally likely for employees in the first quarter of their tour of duty, and those in the second quarter of their tour of duty.

Therefore the total exposed to risk for the third quarter is $180 + (10 \times 0.5) = 185$ person-quarters

Finally, in the fourth quarter there are 126 new employees in the first three months of their six-month tour, and 108 (the 114 who were new on 1 April 2009 minus a proportion equal to $114/(114 + 76) = 0.6$ of the 10 who died in the third quarter) employees in the second three months of their six-month tour.

Of these 13 died during the quarter, and these contribute 0.5 quarters each to the exposed to risk.

Therefore the total exposed to risk for the fourth quarter is $221 + (13 \times 0.5) = 227.5$ person-quarters.

We assume there are no deaths apart from accidental deaths.

These calculations are summarised in the table below.

Quarter beginning	*Employees in first quarter of tour*	*Employees in second quarter of tour*	*Less 0.5 × accidental deaths*	*Central exposed to risk in quarters*
1 January	90	0	5	85
1 April	80	80	4	156
1 July	114	76	5	185
1 October	126	108	6.5	227.5

The quarterly rates of being bitten are therefore as follows:

Quarter beginning	*Spider bites*	*Exposed to risk*	*Rate of being bitten*	
1 January	15	85	15/85	= 0.176
1 April	25	156	25/156	= 0.160
1 July	30	185	30/185	= 0.162
1 October	40	227.5	40/227.5	= 0.176

We assume that all spider bites are treated.

(ii) The assumption that there are no deaths apart from accidental deaths is unlikely to be true, and probably the company would have data on these which could be included in the calculations.

Accidental deaths may be more likely among employees in their first quarter than their second, as those in their second quarter have more experience.

Accidental deaths may be more likely at the beginning of a quarter, when there are newly arrived employees.

The experience of the quarter beginning 1 January may be different from that of other quarters because that is the first quarter that any employees are stationed in the swamp, and they may not know about the spiders when they arrive. In subsequent quarters they may be able to adjust their arrangements to reduce the possibility of being bitten.

Several alternatives to part (i) were also given credit. For example assuming spider bites are all fatal produces the following solution to part (i):

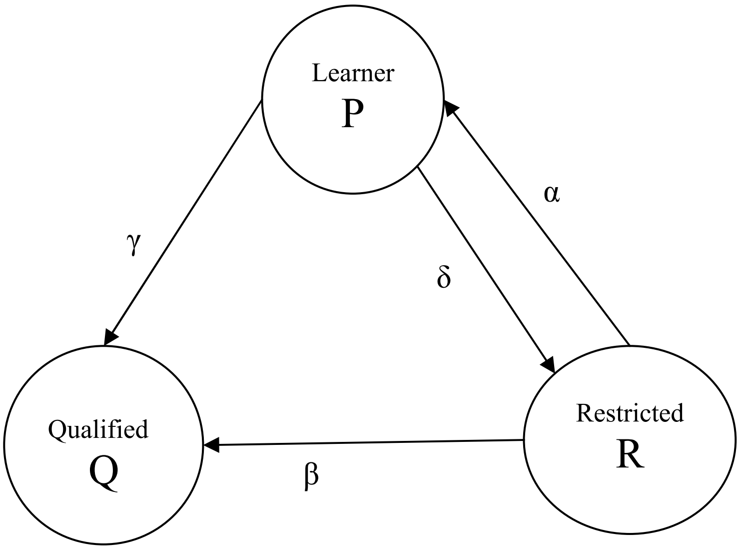
Quarter beginning	*Employees in first quarter of tour*	*Employees in second quarter of tour*	*Less 0.5 × total deaths*	*Central exposed to risk in quarters*
1 January	*90*	*0*	*12.5*	*77.5*
1 April	*80*	*65*	*16.5*	*128.5*
1 July	*114*	*62*	*20*	*156.0*
1 October	*126*	*88*	*26.5*	*187.5*

The quarterly rates of being bitten are therefore as follows:

Quarter beginning	*Spider bites*	*Exposed to risk*	*Rate of being bitten*
1 January	*15*	*77.5*	*15/77.5 = 0.194*
1 April	*25*	*128.5*	*25/128.5 = 0.195*
1 July	*30*	*156*	*30/156 = 0.192*
1 October	*40*	*187.5*	*40/187.5 = 0.213*

In part (ii) credit was only given if the points made related to one of the assumptions stated in the answer to part (i).

7 (i)



(ii) Let α be the transition rate R to P
β be the transition rate R to Q
γ be the transition rate P to Q
δ be the transition rate P to R

Let P be the time spent in Learner state
R be the time spent in Restricted state

Let a be the number of transitions from Restricted to Learner
b be the number of transitions from Restricted to Qualified
c be the number of transitions from Learner to Qualified
d be the number of transitions from Learner to Restricted

$$L \propto \exp\{(-\delta-\gamma)P\}\exp\{(-\alpha-\beta)R\}\,\alpha^a\beta^b\gamma^c\delta^d$$

(iii) Take the logarithm of the likelihood

$$\log_e L = k - P(\delta+\gamma) - R(\alpha+\beta) + a\ln\alpha + b\ln\beta + c\ln\gamma + d\ln\delta$$

Differentiate with respect to α

$$\frac{d\log_e L}{d\alpha} = -R + \frac{a}{\alpha}$$

Set equal to zero to get estimator:

$$\frac{a}{\alpha} = R \qquad \hat{\alpha} = a/R.$$

Differentiate a second time:

$$\frac{d \ln L}{d^2 \alpha} = -\frac{a}{\alpha^2}.$$

which is always negative, so that we have a maximum.

Thus $\hat{\alpha} = 230/1940 = 0.1186$

8 (i) 11 students qualified during the period of observation, so the median is the number of sessions taken to qualify by the sixth student to qualify.

This is 9 sessions.

(ii) Define t as the number of sessions which have taken place since 1 Jan 2003.

Stopped studying implies recorded *after* the session number reported.

t_j	N_j	D_j	C_j	$\frac{D_j}{N_j}$	$1-\frac{D_j}{N_j}$
0	23	0	2	–	1
6	21	1	0	1/21	20/21
8	20	2	1	2/20	18/20
9	17	3	0	3/17	14/17
11	14	2	1	2/14	12/14
13	11	3	0	3/11	8/11

The Kaplan-Meier estimate is given by product of $1-\frac{D_j}{N_j}$

Then the Kaplan-Meier estimate of the survival function is

t	$\hat{S}(t)$
$0 \le t < 6$	1
$6 \le t < 8$	0.9524
$8 \le t < 9$	0.8571
$9 \le t < 11$	0.7059
$11 \le t < 13$	0.6050
$13 \le t < 14$	0.4400

(iii) The median time to qualify as estimated by the Kaplan-Meier estimate is the first time at which $S(t)$ is below 0.5.

Therefore the estimate is 13 sessions.

(iv) The estimate based on students qualifying during the period is a biased estimate because it does not contain information about students still studying at the end of the period, or about those who dropped out (stopped studying without qualifying).

The students still studying at the end of 2009 have (by definition) a longer period to qualification than those who qualified in the period.

Hence the Kaplan-Meier estimate is higher than the median using only students who qualified during the period.

In part (i) the question said "determine" so some explanation of where the answer comes from was required for full credit. In part (ii) the question said "calculate" so the correct S(t) and associated ranges of t scored full marks.

9 (i) $h(z, t) = h_0(t) \exp(\beta z_i^T)$

$h(z,t)$ is the hazard at time t (or just $h(t)$ is OK)
$h_0(t)$ is the baseline hazard

z_i are covariates
β is a vector of regression parameters

(ii) The baseline hazard refers to a female chicken in the old enclosure

(iii) The 95 per cent confidence interval for a parameter β is given by the formula

$$\beta \pm 1.96(\mathrm{SE}[\beta]) = \beta \pm 1.96\sqrt{Var(\beta)}\ ,$$

where SE[β] is the standard error of the parameter β.

Thus, for the covariate z_1 =1 if Duck 0 otherwise, we have

95 per cent confidence interval =
$$-0.210 \pm 1.96\sqrt{0.002} = -0.210 \pm 0.088 = \{-0.298, -0.122\}$$

	95% C.I.
z_1 = 1 if Duck 0 otherwise	β_1 = (–0.298, -0.122)
z_2 = 1 if Goose 0 otherwise	β_2 = (–0.049, 0.199)
z_3 = 1 if New enclosure 0 otherwise	β_3 = (0.049, 0.201)
z_4 = 1 if Male 0 otherwise	β_4 = (0.100, 0.300)

(iv) The parameter for the new enclosure is 0.125 so the ratio of the hazard for two otherwise identical birds is exp(0.125) = 1.133.

So the hazard appears to have got worse.

The 95% confidence interval is entirely positive OR does not include zero

so at the 95% level the deterioration is statistically significant.

(v) ALTERNATIVE 1

Hazard for a Male, Goose in the Old enclosure is
$h_0(t)$ exp (0.2 + 0.075 + 0) = $h_0(t)$ exp (0.275)

Therefore the probability of still being alive in 6 months is

$$S_{\text{Goose}} = \exp\left[-\int_0^6 h_0(t)\exp(0.275)\,dt\right]$$

$$= \exp\left[-1.31653\int_0^6 h_0(t)\,dt\right]$$

This is equal to 0.9 so

$$\tfrac{\ln 0.9}{1.31653} = -\int_0^6 h_0(t)\,dt$$

$$\int_0^6 h_0(t)\,dt = 0.080028951$$

Hazard of a Female, Duck in the New enclosure is
$h_0(t)\exp(0-0.210+0.125) = h_0(t)\exp(-0.085)$

So, the probability she is alive after 6 months is

$$S_{\text{Duck}} = \exp\left[-\int_0^6 h_0(t)\exp(-0.085)\,dt\right]$$

$$= \exp\{-0.080028951(0.918512284)\}$$

$$= \exp\{-0.073507574\}$$

$$= 0.929129$$

So the probability she's dead is 0.07087

ALTERNATIVE 2

Hazard for a Male, Goose in the Old enclosure is
$h_0(t)$ exp (0.2 + 0.075 + 0) = $h_0(t)$ exp (0.275)

Therefore the probability of still being alive in 6 months is

$$S_{\text{Goose}} = \exp\left[-\int_0^6 h_0(t)\exp(0.275)\,dt\right]$$

Similarly, the probability of still being alive in 6 months for
A Female Duck in the New enclosure is

$$S_{\text{Duck}} = \exp\left[-\int_0^6 h_0(t)\exp(-0.085)\,dt\right].$$

Therefore we can write

$$\frac{S_{\text{Goose}}}{S_{\text{Duck}}} = \frac{\exp\left[-\int_0^6 h_0(t)\exp(0.275)\,dt\right]}{\exp\left[-\int_0^6 h_0(t)\exp(-0.085)\,dt\right]},$$

whence

$$\frac{\log_e S_{\text{Goose}}}{\log_e S_{\text{Duck}}} = \frac{-\int_0^6 h_0(t)\exp(0.275)\,dt}{-\int_0^6 h_0(t)\exp(-0.085)\,dt} = \frac{\exp(0.275)}{\exp(-0.085)}.$$

Hence

$$\log_e S_{\text{Duck}} = \frac{\log_e S_{\text{Goose}}[\exp(-0.085)]}{\exp(0.275)}.$$

Since $S_{\text{Goose}} = 0.9$, therefore

$$\log_e S_{\text{Duck}} = \frac{\log_e 0.9[\exp(-0.085)]}{\exp(0.275)} = -0.07351$$

So $S_{\text{Duck}} = 0.929129$

So the probability she's dead is 0.07087

10 (i) (a) The state space is discrete (with four states: O – ordinary passenger, B – bronze member, S – silver member and G – gold member)

The probability that a passenger has a particular membership status next year depends only on their membership status in the current year (i.e. the status in previous years is not relevant).

Therefore the process is Markov.

(b) The state space is finite and therefore there is at least one stationary probability distribution.

Since any state can be reached from any other state, the Markov chain is irreducible.

Therefore the stationary probability distribution is unique.

(ii) The transition matrix P is:

$$\begin{array}{c} O \\ B \\ S \\ G \end{array} \begin{pmatrix} p_0 + p_1 & p_{2+} & 0 & 0 \\ p_0 & p_1 & p_{2+} & 0 \\ 0 & p_0 & p_1 & p_{2+} \\ 0 & 0 & p_0 & p_1 + p_{2+} \end{pmatrix}$$

where the probability that a passenger books i flights in a year is p_i.

(iii) Let the probability that a passenger is in state j according to the stationary distribution be π_j ($j = O, B, S, G$).

The π_j are given by the general formula

$\pi = \pi P$.

With $p_0 = 0.4$, $p_1 = 0.4$ and $p_{2+} = 0.2$, we therefore have the equations

$$\pi_O = 0.8\pi_O + 0.4\pi_B \qquad (1)$$
$$\pi_B = 0.2\pi_O + 0.4\pi_B + 0.4\pi_S \qquad (2)$$
$$\pi_S = 0.2\pi_B + 0.4\pi_S + 0.4\pi_G \qquad (3)$$
$$\pi_G = 0.2\pi_S + 0.6\pi_G \qquad (4)$$

We also know that

$\pi_O + \pi_B + \pi_S + \pi_G = 1$.

Using equation (1) we have

$$0.2\pi_O = 0.4\pi_B$$

so that

$$\pi_O = 2\pi_B\,.$$

Substituting in equation (2) this yields

$$\pi_B = 0.2(2\pi_B) + 0.4\pi_B + 0.4\pi_S\,,$$

so that

$$0.2\pi_B = 0.4\pi_S$$

and hence

$$\pi_S = 0.5\pi_B\,.$$

Finally, substituting in equation (3) yields

$$0.5\pi_B = 0.2\pi_B + 0.4(0.5\pi_B) + 0.4\pi_G\,,$$

so that

$$0.1\pi_B = 0.4\pi_G$$

and hence

$$\pi_G = 0.25\pi_B\,.$$

We therefore have

$$2\pi_B + \pi_B + 0.5\pi_B + 0.25\pi_B = 1\,,$$

whence

$$\pi_B = \frac{1}{3.75} = \frac{4}{15} = 0.2667\,,$$

and the stationary distribution is

$$\pi_O = \frac{8}{15} = 0.5333$$

$$\pi_B = \frac{4}{15} = 0.2667$$

$$\pi_S = \frac{2}{15} = 0.1333$$

$$\pi_G = \frac{1}{15} = 0.0667$$

(iv) EITHER

The expected cost of the scheme per member per year is
$(0 \times 0.5333) + (£10 \times 0.2667) + (£20 \times 0.1333) + (£30 \times 0.0667) = £7.33$

For the scheme to be worth running, therefore, the average profit per member per year must exceed £7.33.

The profit per member is 0 for the 40% who book no flights, £10 for the 40% who book one flight, and £10m for the 20% who book two or more flights, where m is the average number of flights booked by those in the latter category.

For there to be a profit, we must have

$(0.4 \times 0) + (0.4 \times £10) + (0.2 \times £10m) > 7.33$

or

$4 + 2m > 7.33$

$2m > 3.33$

$m > 1.67$

This must be the case since m cannot be less than 2.

Therefore the airline makes a profit on the members of the scheme.

OR

Assuming that the distribution of the number of flights taken is the same for all membership statuses, then for an Ordinary member the expected profit is

$(0.4 \times 0) + (0.4 \times 10) + (0.2 \times 20) = £8$

Similarly for the other classes of member the expected profit is

Bronze:	$(0.4 \times -10) + (0.4 \times 0) + (0.2 \times 10)$	$= £ -2$
Silver:	$(0.4 \times -20) + (0.4 \times -10) + (0.2 \times 0)$	$= £ -12$
Gold:	$(0.4 \times -30) + (0.4 \times -20) + (0.2 \times -10)$	$= £ -22$

In any one year, the proportions of members in each category are given by the stationary distribution,

so the expected profit per member is

$$\frac{8}{15}(£8)+\frac{4}{15}(£-2)+\frac{2}{15}(£-12)+\frac{1}{15}(£-22)=£0.667$$

This assumes no member makes more than 2 flights per year, so is a minimum estimate of the profit.

This minimum estimate is positive, so the airline makes a profit.

11 (i) A time inhomogeneous model should be used.

Because transition probabilities out of the "Suspended" state between times s and t may depend not only on the time difference $t - s$ but on the the duration s the policy has been in that state (e.g. the probability of remaining in the suspended state for $t = 0.75$ and $s = 0.25$ is $\exp(-0.025)$), but the probability for $t = 1.25$ and $s = 0.75$ is 0.

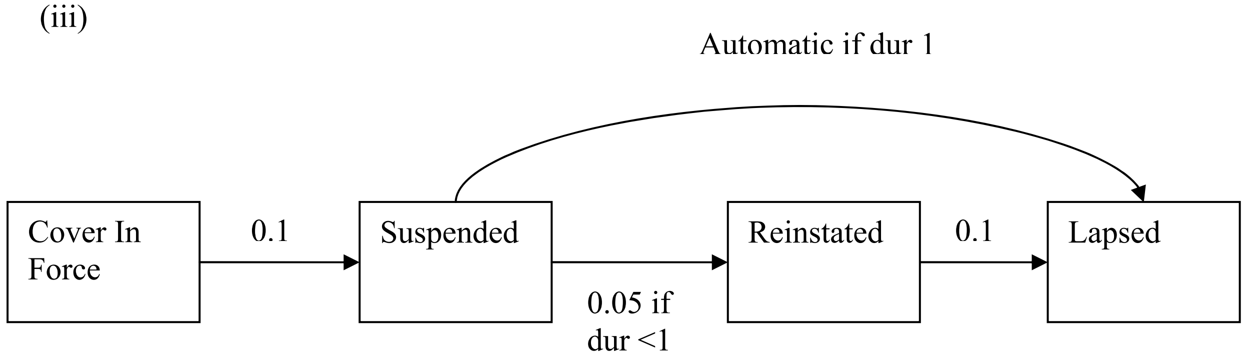
(ii) (a) A model with this state space would not satisfy the Markov property because a policy can only be reinstated once,

so if in state Cover in Force we would need to know if the policy has previously been Suspended.

(b) A Markov model could be obtained by expanding the state space to {Cover In Force, Suspended, Reinstated, Lapsed}.

In this case the future transitions will depend only on the state currently occupied and duration, irrespective of previous states.

(iii)



(iv) Labelling states as C, S, R and L.

$P_{CC}(0,t) = P_{\overline{CC}}(0,t)$ as no return to this state

$$\frac{d}{dt} P_{\overline{CC}}(0,t) = -0.1 * P_{\overline{CC}}(0,t)$$

$$\frac{1}{P_{\overline{CC}}(0,t)} \frac{d}{dt} P_{\overline{CC}}(0,t) = \frac{d}{dt}(\ln P_{\overline{CC}}(0,t)) = -0.1$$

$$\ln(P_{\overline{CC}}(0,t)) = -0.1t + \text{Constant}$$

$P_{CC}(0,t) = \exp(-0.1t)$ with const = 0 as $P_{CC}(0,0) = 1$

(v) To be in S at time t, must have remained in state C until some time w, then transitioned to S at time w, then remained in state S until t time.

(or express in terms of conditioning)

Probabilities are $P_{\overline{CC}}(0,w)$, $0.1dw$, and $P_{\overline{SS}}(w,t)$ respectively.

Integrating over the possible values of w:

$$P_{CS}(0,t) = \int_{t-1}^{t} P_{CC}(0,w) * 0.1 * P_{SS}(w,t)dw$$

As probability of remaining in S if $t - w > 1$ is zero.

If $t - w < 1$

$$P_{SS}(w,t) = \exp(-0.05(t-w))$$

By natural extension from (iv).

Substituting

$$P_{CS}(0,t) = \int_{t-1}^{t} \exp(-0.1w) * 0.1 * \exp(-0.05(t-w))dw$$

$$P_{CS}(0,t) = 0.1\exp(-0.05t) \int_{t-1}^{t} \exp(-0.05w)dw$$

$$P_{CS}(0,t) = -2\exp(-0.05t) \left| \exp(-0.05w) \right|_{t-1}^{t}$$

$P_{CS}(0,t) = -2\exp(-0.05t)(\exp(-0.05t) - \exp(-0.05t).\exp(0.05))$

$= 2(\exp(0.05) - 1)\exp(-0.1t)$

OR

$0.1025\exp(-0.1t)$

ALTERNATIVELY

This assumes that can remain in state 'Suspended' for more than 1 time period (after which permanently suspended)

To be in S at time t, must have remained in state C until some time w, then transitioned to S at time w, then remained in state S until t time.

(or express in terms of conditioning)

Probabilities are $P_{\overline{CC}}(0,w)$, 0.1dw, and $P_{\overline{SS}}(w,t)$ respectively.

Integrating over the possible values of w:

$$P_{CS}(0,t) = \int_0^t P_{CC}(0,w) * 0.1 * P_{SS}(w,t)dw$$

As transition probability out of state S if $t - w > 1$ is zero.

If $t - w < 1$

$P_{SS}(w,t) = \exp(-0.05(t-w))$

By natural extension from part (iv).

Splitting the integral into the parts for t - $w > 1$ and $t - w < 1$

$$P_{CS}(0,t) = \int_{t-1}^{t} \exp(-0.1w) * 0.1 * \exp(-0.05(t-w))dw + \int_0^{t-1} \exp(-0.1w) * 0.1 * \exp(-0.05((w+1)-w))dw$$

$$P_{CS}(0,t) = 0.1\exp(-0.05t)\int_{t-1}^{t} \exp(-0.05w)dw + 0.1\exp(-0.05)\int_0^{t-1} \exp(-0.1w)dw$$

$$P_{CS}(0,t) = -2\exp(-0.05t)\left|\exp(-0.05w)\right|_{t-1}^{t} - \exp(-0.05)\left|\exp(-0.1w)\right|_0^{t-1}$$

$$P_{CS}(0,t) = -2\exp(-0.05t)(\exp(-0.05t) - \exp(-0.05t).\exp(0.05)) + \exp(0.05) - \exp(0.05).\exp(-0.1t)$$

$= (\exp(0.05) - 2)\exp(-0.1t) + \exp(-0.05)$

In part (iii) the label on the arrow going directly from "Suspended" to "Lapsed" is not needed, provided that the label on the arrow going from the "Suspended" to "Reinstated" indicates that the rate of 0.05 only applies if the duration is less than 1. If the label on the arrow going from "Suspended" to "Reinstated" does not indicate this, then we need an indication that movement from "Suspended" to "Lapsed" is automatic if duration = 1

12 (i) By reference to a standard table – appropriate if data are scanty or a table of similar lives exists.

Graphical graduation – appropriate if a "quick and dirty" result needed OR for scanty data where no other method is appropriate

By parametric formula, if the experience is large.

(ii) Standard table data

Age x	*Number of survivors*	p_x	q_x
50	32,669	0.99522	0.00478
51	32,513	0.99462	0.00538
52	32,338	0.99397	0.00603
53	32,143	0.99325	0.00675
54	31,926	0.99245	0.00755
55	31,685	0.99154	0.00846
56	31,417	0.99058	0.00942
57	31,121	0.98952	0.01048
58	30,795	0.98831	0.01169
59	30,435	0.98699	0.01301
60	30,039		

Calculations:

Age last	*Exposed to risk*	*Expected deaths (E)*	*Observed Deaths (O)*	*O-E*	$(O-E)^2/E$	$\frac{(O-E)^2}{E(1-q)}$
50	2,381	11.3697	16	4.6303	1.8857	1.8948
51	3,177	17.1001	21	3.8999	0.8894	0.8942
52	3,460	20.8640	22	1.1360	0.0619	0.0622
53	1,955	13.1984	15	1.8016	0.2459	0.2476
54	3,122	23.5671	24	0.4329	0.0080	0.0080
55	3,485	29.4770	29	–0.4770	0.0077	0.0078
56	2,781	26.2016	26	–0.2016	0.0016	0.0016
57	3,150	32.9970	31	–1.9970	0.1209	0.1221
58	3,651	42.6810	39	–3.6810	0.3175	0.3212
59	3,991	51.9282	48	–3.9282	0.2972	0.3011
				Total	3.8356	3.8606

The null hypothesis is that the data come from a population where the mortality is that represented by the standard table.

The test statistic $\sum \frac{(O-E)^2}{E}$ is distributed χ^2.

There are 10 age groups.

No degrees of freedom lost for choice of table, parameters or constraints on data.

So we use 10 degrees of freedom.

This is a one-tailed test.

The upper 5% point of the χ^2 with 10 degrees of freedom is 18.31.

The observed test statistic is 3.84.

Since 3.84 < 18.31.

We have insufficient evidence to reject the null hypothesis.

(iii) ALTERNATIVE 1

(a) The data easily pass the chi squared test, but there does seem to be a gradual drift of $(O - E)$ figures from strongly positive to strongly negative. I would do a grouping of signs test to see if the data display runs or "clumps" of deviations of the same sign.

(b) G = Number of groups of positive zs = 1

m = number of deviations = 10
n_1 = number of positive deviations = 5
n_2 = number of negative deviations = 5

THEN EITHER

We want k^* the largest k such that

$$\sum_{t=1}^{k} \frac{\binom{n_1-1}{t-1}\binom{n_2+1}{t}}{\binom{m}{n_1}} < 0.05$$

The test fails at the 5% level if $G \leq k^*$.

From the Gold Book $k^* = 1$, so we reject the null hypothesis.

OR

For $t = 1$

$$\binom{n_1 - 1}{t-1} = \binom{4}{0} = 1 \quad \text{and} \quad \binom{n_2 + 1}{t} = \binom{6}{1} = 6 \quad \text{and} \quad \binom{m}{n_1} = \binom{10}{5} = 252$$

So Pr[$t = 1$] if the null hypothesis is true is 6/252 = 0.0238, which is less than 5% so we reject the null hypothesis.

ALTERNATIVE 2

(a) The data easily pass the chi squared test, but there does seem to be a gradual drift of ($O - E$) figures from strongly positive to strongly negative. I would do a serial correlation test to see if the data displays runs or clumps" of deviations of the same sign.

(b) The calculations are shown in the table below

x	z_x	z_{x+1}	$A = z_x - \bar{z}$	$B = z_{x+1} - \bar{z}$	AB	A^2	B^2
50	1.373	0.943	0.908	0.570	0.517	0.824	0.325
51	0.943	0.249	0.478	–0.125	–0.060	0.228	0.016
52	0.249	0.496	–0.217	0.123	–0.027	0.047	0.015
53	0.496	0.089	0.031	–0.284	–0.009	0.001	0.081
54	0.089	0.088	–0.376	–0.286	0.107	0.142	0.082
55	0.088	0.039	–0.378	–0.334	0.126	0.143	0.112
56	0.039	0.348	–0.426	–0.026	0.011	0.181	0.001
57	0.348	0.563	–0.118	0.190	–0.022	0.014	0.036
58	0.536	0.545	0.098	0.172	0.017	0.010	0.029
59	0.545						
$\bar{z}$	0.465	0.373		Sum	0.661	1.589	0.695

$0.661/(1.589*0.695)^{0.5} = 0.629$

Test $0.629\ (9^{0.5}) = 1.887$ against Normal (0,1), and, since

$0.629\ (9^{0.5}) = 1.887 > 1.645$, we reject the null hypothesis.

ALTERNATIVE 3

(a) Do the signs test to detect overall bias.

(b) Under the null hypothesis, the number of positive signs amongst the z_xs is distributed Binomial (10, ½).

We observe 5 positive signs.

The probability of obtaining 5 or more positive signs is 0.623

OR

The probability of obtaining exactly 5 positive signs is 0.246

Since this is greater than 0.025 (two-tailed test), we cannot reject the null hypothesis.

Note that because this test is not really appropriate in a case where there are five negative and five positive deviations, no marks were awarded for part (a) to candidates who chose the Signs Test unless earlier errors meant that the number of negative and positive signs were unequal.

ALTERNATIVE 4

(a) Do the cumulative deviations test to detect overall bias.

(b) The test statistic is $\dfrac{\sum_x \left(\theta_x - E_x \overset{o}{q}_x\right)}{\sqrt{\sum_x E_x \overset{o}{q}_x}} \sim \text{Normal}(0,1)$

Age x	θ_x	$E_x \overset{o}{q}_x$	$\theta_x - E_x \overset{o}{q}_x$
50	16	11.37	4.63
51	21	17.10	3.90
52	22	20.86	1.14
53	15	13.20	1.80
54	24	23.57	0.43
55	29	29.48	–0.48
56	26	26.20	–0.20
57	31	33.00	–2.00
58	39	42.68	–3.68
59	48	51.93	–3.93
	$\sum$	269.38	1.62

So the value of the test statistic is $\dfrac{1.62}{\sqrt{269.38}} = 0.09846$.

Using a 5% level of significance,
we see that $-1.96 < 0.09846 < 1.96$.

We do not reject the null hypothesis.

ALTERNATIVE 5

(a) To check for outliers we do the individual standardised deviations test.

(b) If the standard table rates were the true rates underlying the observed rates

we would expect the individual deviations to be distributed Normal (0,1)

and therefore only 1 in 20 z_x s should have absolute magnitudes greater than 1.96

OR

none should lie outside the range (–3, +3)

OR

or diagram showing split of deviations actual versus expected.

Looking at the z_x s we see that the largest individual deviation is 1.373.

Since this is less in absolute magnitude than 1.96 we cannot reject the null hypothesis.

In part (ii) credit was only given for the null hypothesis if the wording used by the candidate indicates that (s)he understands that it is the mortality underlying the observed data that is not significantly different from that in the standard table, or that the standard table "represents" the mortality in the observed data. The null hypothesis is not that the mortality in the observed data is the same as that in the standard table – as it will normally not be.

END OF EXAMINERS' REPORT