INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

September 2012 examinations

Subject CT4 – Models
Core Technical

Introduction

The Examiners' Report is written by the Principal Examiner with the aim of helping candidates, both those who are sitting the examination for the first time and using past papers as a revision aid and also those who have previously failed the subject.

The Examiners are charged by Council with examining the published syllabus. The Examiners have access to the Core Reading, which is designed to interpret the syllabus, and will generally base questions around it but are not required to examine the content of Core Reading specifically or exclusively.

For numerical questions the Examiners' preferred approach to the solution is reproduced in this report; other valid approaches are given appropriate credit. For essay-style questions, particularly the open-ended questions in the later subjects, the report may contain more points than the Examiners will expect from a solution that scores full marks.

D C Bowie
Chairman of the Board of Examiners

December 2012

General comments on Subject CT4

Subject CT4 comprises five main sections: (1) a study of the properties of models in general, and their uses for actuaries, including advantages and disadvantages (and a comparison of alternative models of the same processes); (2) stochastic processes, especially Markov chains and Markov jump processes; (3) models of a random variable measuring future lifetime; (4) the calculation of exposed to risk and the application of the principle of correspondence; (5) the reasons why mortality (or other decrement) rates are graduated, and a range of statistical tests used both to compare a set of rates with a previous experience and to test the adherence of a graduated set of rates to the original data. Throughout the subject the emphasis is on estimation and the practical application of models. Theory is kept to the minimum required in order usefully to apply the models to real problems.

Different numerical answers may be obtained to those shown in these solutions depending on whether figures obtained from tables or from calculators are used in the calculations but candidates are not penalised for this. However, candidates may be penalised where excessive rounding has been used or where insufficient working is shown.

Comments on the September 2012 paper

The general performance was slightly better than in the September 2010 or September 2011 sessions, but substantially inferior to that in April 2012. Nevertheless, well-prepared candidates scored highly across the whole paper. The comments that follow the questions concentrate on areas where candidates could have improved their performance.

1

Benefits

Systems with long time frames can be studied in compressed time

Complex systems with stochastic elements can be studied (especially by simulation modelling).

Different future policies or possible actions can be compared to see which best suits the requirements of a user.

Models allow control over experimental conditions, so that we can reduce the variance of the results output without upsetting the mean values.

Limitations

Model development requires a lot of time and expertise, and hence can be costly.

Models more useful for comparing the results of input variations than for optimising outputs.

Models can look impressive, but can lull the user into a false sense of security. Impressive output is not a substitute for validity and close imitation of the real world.

Models rely heavily on the data input. If this is poor or lacking in credibility the output is likely to be flawed.

Models rely heavily on the assumptions used, poor assumptions can invalidate the model output.

Users need to understand the model sufficiently well to be able to know when it is appropriate to apply it.

Interpretation of models can be difficult.

Models cannot take into account all possible future events, e.g. changes in legislation.

Many candidates scored full marks on this question. The question asked for TWO benefits and TWO limitations, so credit was given for the most fully described two of each. Extra marks for the benefits could not be transferred to the limitations to make up a shortfall, and vice versa.

2

The nature of the existing sickness data the company possesses. The model can only be as complex as the data will allow it to be.

Whether the company has made any previous attempts to model sickness rates among its employees, and how successful they were.

The complexity of the model – e.g. whether it should be stochastic or deterministic. More complex models will be costlier to prepare and run, but eventually there may be diminishing returns to additional complexity.

General trends in sickness at the national level may need to be built in.

The definition of sickness and level of benefits payable under the scheme.

Does the company plan to change the characteristics of the employees? For example, does it plan to recruit more mature persons?

The ease of communication of the model.

The budget and resources available for the construction of the model.

Capability of staff. Will outside consultants be required?

By whom will the model be used? Will they be capable of understanding and using it?

Does the model need to interface with models of other aspects of the company's business (e.g. taking data from other systems)?

The independence of sickness rates should be taken into account e.g. in the event of an epidemic claims cannot be considered independent.

Other relevant points were given credit. The Examiners were looking for comments which made reference to the scenario proposed in the question. Many candidates simply reproduced one of the lists in the Core Reading (unit 1 page 4 and unit 1 page 6) without relating to the scenario in the question. Answers along these lines scored limited credit.

3

(i) The principle of correspondence states that a life should be included in the denominator of the rate at time t if and only if, were that life to die at time t, his or her death would be counted in the numerator.

(ii) In order for the exposed to risk to correspond to the deaths data, it needs to be on an age next birthday basis.

The exposed to risk at age x next birthday may be approximated using the census approximation.

$$E_x^c = \int_0^2 P_{x,t}\,dt$$

Using the trapezium rule (i.e. assuming the population varies linearly between "census" dates) this may be evaluated as

$$E_x^c = \frac{1}{2}(P_{x,1/1/09} + P_{x,1/1/10}) + \frac{1}{4}(P_{x,1/1/10} + P_{x,1/7/10}) + \frac{1}{4}(P_{x,1/7/10} + P_{x,1/1/11})$$
$$= \frac{1}{2}P_{x,1/1/09} + \frac{3}{4}P_{x,1/1/10} + \frac{1}{2}P_{x,1/7/10} + \frac{1}{4}P_{x,1/1/11}$$

where $P_{x,t}$ is the population aged x next birthday at time t.

But, in this case, we have data on an age last birthday basis.

If $P^*_{x,t}$ is the population aged x last birthday at time t, then

$$P_{x,t} = P^*_{x-1,t}$$

and the exposed to risk becomes

$$E_x^c = \frac{1}{2}P^*_{x-1,1/1/09} + \frac{3}{4}P^*_{x-1,1/1/10} + \frac{1}{2}P^*_{x-1,1/7/10} + \frac{1}{4}P^*_{x-1,1/1/11}$$

So, using the data given, the exposed to risk we need at age 50 is

$$E_x^c = \frac{1}{2}(2{,}000) + \frac{3}{4}(2{,}100) + \frac{1}{2}(2{,}300) + \frac{1}{4}(2{,}500) = 4{,}350$$

and the estimated force of mortality at age 50 next birthday is

$$\hat{\mu}_{50} = \frac{200 + 225}{4{,}350} = 0.0977$$

(iii) The estimate $\hat{\mu}_{50}$ applies to the middle of the rate interval,

which is exact age 49.5 years.

In part (ii) the question said "estimate" so some indication of how the answer was arrived at was required for full credit. The correct numerical answer on its own was insufficient. Some candidates noted that a correct exposed-to-risk could be calculated without using the July 2010 population figures. This was given full credit, provided a valid explanation of why the July population figures were not needed was given. In part (iii) the question said "state" so the full mark was awarded for 49.5. In part (iii) for full credit the answer had to be consistent with what the candidate had done in part (ii).

4

(i) We do not need to know the general shape of the hazard/distribution.

(ii) $h(t, z_i) = h_0(t)\exp\left(\beta z_i^T\right)$

$h(t, z_i)$ 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

(iii) Baseline hazard refers to a male sold a whole life policy by the direct sales force.

(iv) For the male policy

the probability still in force is 0.4.

Sum of parameters for male is 0.4

$$0.4 = \exp\left\{-\int_0^5 h_0(t)\exp(0.4)dt\right\} = \exp\left\{-1.49\int_0^5 h_0(t)dt\right\}$$

So $$\int_0^5 h_0(t)dt = \frac{\ln 0.4}{-1.49}$$

And for the female policy sum of parameters is -0.1

THEN EITHER

We therefore want $\exp\left\{-\int_0^5 h_0(t)\exp(-0.1)dt\right\} = \exp\left\{-0.905 * \frac{\ln 0.4}{-1.49}\right\}$

$= 0.57364$

OR

We therefore want $\left\{0.4^{e^{-0.4}}\right\}^{e^{-0.1}} = 0.4^{e^{-0.5}} \qquad = 0.57364$

Parts (i)-(iii) of this question were well answered. Answers to part (iv) were variable. Common errors included working with the probability of having lapsed (i.e. 1 minus the probability of still being in force), and omission of the integral.

5

(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

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

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

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 $\pi_2 = \frac{5}{6}\pi_1$.

Since $\pi_1 + \pi_2 + \pi_3 = 1$,+½

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

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.

6

(i) Let S be the state space. We say that $\{\pi_j \mid j \in S\}$ is a stationary probability distribution for a Markov chain with transition matrix P if the following hold for all $j \in S$:

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}\,, \quad \text{OR } \pi = \pi\, \text{P}$$

$$\sum_{j \in S} \pi_j = 1\,.$$

$$\pi_i \geq 0$$

(ii) With state space {Working, Broken}

Transition matrix $A = \begin{pmatrix} 0.95 & 0.05 \\ 0.6 & 0.4 \end{pmatrix}$

(iii) This requires the stationary distribution π which satisfies

$$\pi A = \pi$$

$$0.95\pi_W + 0.6\pi_B = \pi_W$$
$$0.05\pi_W + 0.4\pi_B = \pi_B$$

and $\quad \pi_W + \pi_B = 1$

$$\pi_W = 12\pi_B$$
$$\pi_W = 12/13$$
$$\pi_B = 1/13$$

So 1 in 13 games is cancelled.

OR

The average number of games for which the lights work before breaking down is 1/0.05 = 20 games.

Once they have broken down the expected number of games for which they will be out of action is 1/0.6 = 5/3 games.

Therefore the proportion of games for which the lights are out of action is

$$\frac{5/3}{20+(5/3)} = \frac{5}{65} = \frac{1}{13}$$

So 1 in 13 games is cancelled.

(iv) First we need to find the new stationary probabilities.

Transition matrix $A' = \begin{pmatrix} 0.95 & 0.05 \\ 0.8 & 0.2 \end{pmatrix}$

$$0.95\pi_{W'} + 0.8\pi_{B'} = \pi_{W'}$$

and $\pi_{W'} + \pi_{B'} = 1$

Giving $\pi_{W'} = 16/17 \quad \pi_{B'} = 1/17$

Lost income (including fees to repair company):

with Floodwatch: ($10,000 + $1,000)*1/13
with Light Fantastic: ($10,000 + X)*1/17 where X is fee to be negotiated.

So need: 11000/13 = (10000 + X)/17

X = $4,384.62 per day.

In part (i) no credit was given for wordy description of "what happens in the long run". In part (iii) the question said "derive", so we needed an explanation of where the answer came from: only limited credit was given for writing down a numerical answer (even if correct) without explanation. Moreover, in part (iii) calculating the stationary distribution was not sufficient for full credit: we were looking for the correct element to be identified and its value indicated i.e. an explicit statement that "1 in 13 games is cancelled". Parts (i), (ii) and (iii) of this question were well answered, and many candidates also evaluated the proportion of games that would be cancelled under the new floodlight regime in part (iv). Few were able to compute the daily saving, however.

7

(i) With state space {L, H} we have generator matrix

$$A = \begin{pmatrix} -\mu & \mu \\ \rho & -\rho \end{pmatrix}$$

(ii) The holding times are exponentially distributed with parameter μ in state *L* and ρ in state *H.*

(iii) EITHER

The time spent in state *L* before the next visit to *H* has mean $1/\mu$.

Therefore a reasonable estimate for μ is the reciprocal of the mean length of each visit:

= (Number of transitions from *L* to *H*) / (Total time spent in state *L*)

Similarly estimate for ρ is the reciprocal of the mean length of each visit:

= (Number of transitions from *H* to *L*) / (Total time spent in state *H*)

OR

Using the maximum likelihood estimator for μ, we have:

(Number of transitions from L to H)/Total time spent in state L).

Similarly, the MLE of ρ is

(Number of transitions from H to L)/Total time spent in state H).

(iv) $$\frac{\partial}{\partial t}\,{}_tP_s^{\overline{LL}} = -\mu\,{}_tP_s^{\overline{LL}}$$

$$\frac{\partial}{\partial t}\,{}_tP_s^{LL} = -\mu\,{}_tP_s^{LL} + \rho\,{}_tP_s^{LH}$$

$$\frac{\partial}{\partial t}\,{}_tP_s^{LH} = \mu\,{}_tP_s^{LL} - \rho\,{}_tP_s^{LH}$$

(v) $$\frac{\partial}{\partial t}\,{}_tP_s^{\overline{LL}} = -\mu\,{}_tP_s^{\overline{LL}}$$

so ${}_tP_0^{\overline{LL}} = \exp(-\mu t)$

Looking for time when ${}_tP_0^{\overline{LL}} = 1/2$

$1/2 = \exp(-\mu T)$
$T = \ln(2)/\mu$

(vi) Observe that ${}_tP_0^{LL} + {}_tP_0^{LH} = 1$

so, substituting, we have

$$\frac{\partial}{\partial t}\,{}_tP_0^{LL} = -\mu\,{}_tP_0^{LL} + \rho(1 - {}_tP_0^{LL})$$

$$\frac{\partial}{\partial t}\left[\exp((\mu+\rho)t)\,{}_tP_0^{LL}\right] = \rho\exp((\mu+\rho)t)$$

$$\exp((\mu+\rho)t)\,{}_tP_0^{LL} = \frac{\rho}{\mu+\rho}\exp((\mu+\rho)t) + \text{constant}$$

And in state L at time zero so $const = \dfrac{\mu}{\mu+\rho}$

$${}_tP_0^{LL} = \frac{\rho}{\mu+\rho} + \frac{\mu}{\mu+\rho}\exp(-(\mu+\rho)t)$$

Few candidates scored highly on this question. In particular, very few made a serious attempt at parts (v) and (vi). In part (iv), there was confusion among some candidates between ${}_tp_0^{LL}$ and ${}_tp_0^{\overline{LL}}$ and a common error was to write down $\frac{\partial}{\partial t}\,{}_tP_s^{\overline{LL}} = \exp(-\mu t)$.
Many candidates did not attempt part (v) even though is is relatively straightforward.
In part (vi) working through with ${}_tP_0^{LH}$ then at the end taking one minus the answer is a valid approach .

8

(i) When preparing standard tables OR when graduating data from a large industrywide scheme, or a national population

because there will be lots of data available.

(ii) (a) EITHER Graphical graduation OR Graduation with reference to a standard table

(b) EITHER
Graphical graduation may be suitable for a analysis of a newly discovered insect (as data will be scanty and an existing table will not exist)

OR

Graduation with reference to a standard table is useful if data are scanty and a suitable standard table exists (e.g. for female pensioners from a small scheme).

(iii) To test for overall goodness of fit we use the χ^2 test.

The null hypothesis is that the graduated rates are the same as the true underlying rates in the block of business.

The test statistic $\sum_x z_x^2 \approx \chi_m^2$ where *m* is the degrees of freedom.

Age	*Exposed to risk*	*Observed deaths*	*Graduated rates* $(\hat{q}_s)$	*Expected deaths*	z_x	z_x^2
40	1,284	4	.00240	3.0816	0.5232	0.2737
41	2,038	4	.00266	5.4211	-0.6103	0.3725
42	1,952	12	.00297	5.7974	2.5760	6.6360
43	2,158	7	.00332	7.1646	-0.0615	0.0038
44	2,480	11	.00371	9.2008	0.5932	0.3518
45	1,456	7	.00415	6.0424	0.3896	0.1518
46	2,100	12	.00464	9.7440	0.7227	0.5223
47	1,866	16	.00519	9.6845	2.0294	4.1184
48	1,989	15	.00577	11.4765	1.0401	1.0818
49	1,725	10	.00642	11.0745	-0.3229	0.1043
				Total	6.8794	13.6163

The observed test statistic is 13.62

The number of age groups is 10, but we lose an unknown number of degrees for the graduation, perhaps 2. So *m* = 8, say.

The critical value of the chi-squared distribution with 8 degrees of freedom at the 5% level is 15.51.

Since 13.62 < 15.51

we do not reject the null hypothesis.

(iv) It is not necessary to test for smoothness if the graduation was performed using a parametric formula or a standard table, provided that a small number of parameters were used in the formula, or in the function linking to the rates in the standard table.

It will be necessary to test for smoothness if the graduation was performed graphically but this is unlikely to be the case with data from a large insurance company.

(v) The null hypothesis is that the graduated rates are the same as the true underlying rates in the block of business. (i.e the same as part (iii))

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

and therefore only 1 in 20 $z_x s$ should have absolute magnitude greater than 1.96 (or none should be outside −3 to +3)

Looking at the $z_x s$ we see that the largest one is 2.576 and the next is 2.0294

Since they are both greater in magnitude than 1.96

we have sufficient evidence to reject the null hypothesis.

In part (ii)(b) credit was given either for a valid reason or an appropriate example: both are not required. In part (iii) some candidates combined ages 40 and 41, on the basis that the expected deaths at age 40 are fewer than 5, and the statement in the Core Reading at the bottom of Unit 11, p. 10. This was fine. The relevant numbers for the combined 40-41 year age group will be

Observed deaths	*8*
Expected deaths	*8.5027*
z_x	*0.1724*
z_x^2	*0.0297*
Chi-squared	*12.9998*

Because we now only have 9 age groups, we should test against the chi-squared distribution with fewer than 9 degrees of freedom. In part (iv) no credit was given for performing a test for smoothness. Very limited credit was given for impressionistic comments on the putative smoothness or otherwise of the data given. In part (v) few candidates specified the null hypothesis or the distribution of the individual deviations under the null hypothesis.

9

(i) Interval
No. We are counting in days and we know which day each event occurred.

Right
Yes. The end of the course at day 30 cut short the investigation when not all candidates had qualified.

Informative
Possible. Those who left during the 30 days will probably take longer to qualify than those who stayed.

(ii) The data can be re-arranged as shown below.

Day	*Candidate*	*Event*
1	G	Qualified
5	B	Qualified
10	L	Left
12	E	Qualified
12	I	Qualified
15	K	Qualified
19	D	Qualified
19	H	Left
21	C	Left
24	M	Qualified
30	A	Left
30	F	Left
30	J	Left

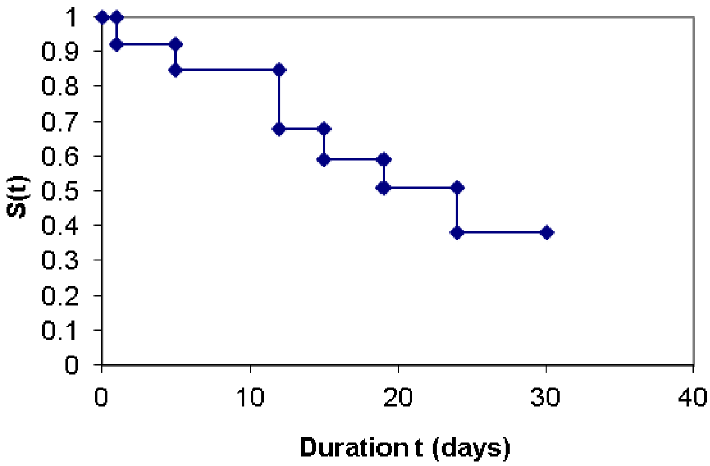
The Kaplan-Meier Estimate is $\hat{S}(t) = \prod_{t_j \le t} 1 - \frac{d_j}{n_j}$

t_j	N_j	D_j	C_j	$\frac{D_j}{N_j}$	$1 - \frac{D_j}{N_j}$
0	13	0	0		
1	13	1	0	1/13	12/13
5	12	1	1	1/12	11/12
12	10	2	0	2/10	8/10
15	8	1	0	1/8	7/8
19	7	1	2	1/7	6/7
24	4	1	0	1/4	3/4

Then the Kaplan-Meier estimate of the survival function is

t	$\hat{S}(t)$
$0 \le t < 1$	1.000
$1 \le t < 5$	0.923
$5 \le t < 12$	0.846
$12 \le t < 15$	0.677
$15 \le t < 19$	0.592
$19 \le t < 24$	0.508
$24 \le t \le 30$	0.381

(iii) A suitable graph is shown below.



(iv) Since qualifications are assumed to happen before censorships, swapping *D* and *H* will have no effect at all, as the order of the two events will simply be reversed, OR is equivalent to re-labelling D as H and H as D, which clearly does not affect the calculations.

Swapping *B* and *L* will reduce the value of N_j at time 10 (replacing time 5)

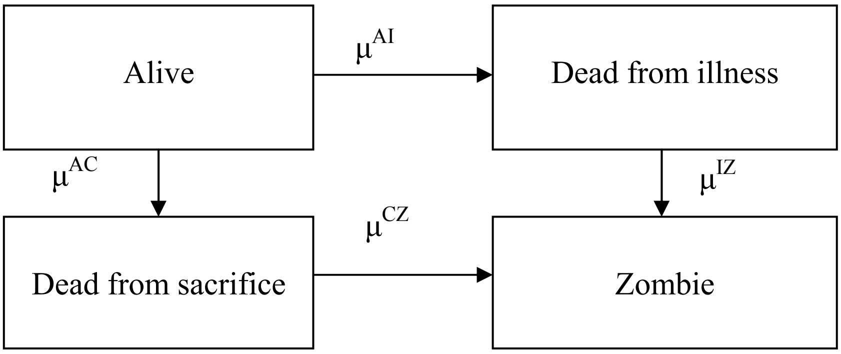
which will increase the value of the hazard at duration 10 compared with that previously at duration 5

This will increase $S(t)$ from durations 5 to immediately before duration 10 and reduce it at durations 10 and over.

In part (i) candidates who gave alternative answers as to whether the form of censoring is present were given credit if the reason was sensible and consistent. So, for example, candidates who stated that interval censoring was present because we do not know exactly when within the day events occurred were given credit. The most common error in part (ii) was performing the calculations with "leaving" as the event. The question asked for the "non-qualification function". This means the survival function with qualification as the event, by analogy with an analysis of mortality in which the survival function can be described as a "non-death" function when death is the event.

10

(i)



(ii) Let the states be labelled as follows:

Alive, A
Dead from illness, I
Dead from sacrifice, C
Zombie, Z

Let the number of transitions observed between states i and j be d^{ij}

and let the transition rate between states i and j be μ^{ij}.

Let the observed waiting time in state i be v^i

The likelihood of the data can be written as follows:

$$L \propto \exp[(-\mu^{AI} - \mu^{AC})v^A]\exp(-\mu^{CZ}v^C)\exp(-\mu^{IZ}v^I)(\mu^{AI})^{d^{AI}}(\mu^{AC})^{d^{AC}}(\mu^{CZ})^{d^{CZ}}(\mu^{IZ})^{d^{IZ}}$$

(iii) Taking logarithms of the likelihood we have:

$$\log_e L = -\mu^{AI}v^A + d^{AI}\log(\mu^{AI}) + \text{terms not depending on } \mu^{AI}.$$

Differentiating this with respect to μ^{AI} gives:

$$\frac{d(\log_e L)}{d\mu^{AI}} = -v^A + \frac{d^{AI}}{\mu^{AI}},$$

and setting the derivative equal to zero produces the maximum likelihood estimate of μ^{AI}:

$$\hat{\mu}^{AI} = \frac{d^{AI}}{v^A}.$$

This is a maximum as the second derivative

$$\frac{d^2(\log_e L)}{(d\mu^{AI})^2} = -\frac{d^{AI}}{(\mu^{AI})^2}$$

is necessarily negative.

(iv) Using the census formula, we estimate v^A as follows

$$v^A = 0.5(P_0 + P_{10}) = 0.5(3{,}189 + 2{,}811) = 3{,}000\,.$$

assuming the population of aliens varies linearly over the ten years between the censuses.

The estimated annual death rate from illness is therefore

$$\frac{369}{3000} = 0.123\,,$$

and the estimated rate of death through sacrifice over the ten years is

$$\frac{231}{3000} = 0.077.$$

(v) (a) The probability that an alien is still alive in ten years' time is given by the formula

$$_{10}p_x^{AA} = \exp\left[-\int_0^{10} (\mu^{AI} + \mu^{AC})du\right] = \exp[-(0.077 + 0.123)10]$$

$$= \exp(-2) = 0.135.$$

(b) Since we are only interested in whether the alien is dead, not what cause (s)he died from,

and since the rate at which aliens become zombies does not depend on cause of death, we can combine the two states "Dead from illness" and "Dead from sacrifice", into a single state "Dead".

For an alien to be Dead in 10 years time (s)he must have survived for u alien years ($0 < u < 10$), died at time u, and then survived in the Dead state (i.e. not become a Zombie) for a duration equal to 10-u alien years.

The probability density of this happening for any given value of u is

exp $[-0.2u]$ survival Alive for a period u
$\times$
0.2 du (here we ignore $o(du)$)
$\times$
exp$[-0.1(10-u)]$ survival Dead for a period $10 - u$

which is

$$0.2 * \exp[-(1+0.1u)]du = 0.2 * \exp(-1)\exp(-0.1u)du$$

$$= 0.0736\exp(-0.1u)du$$

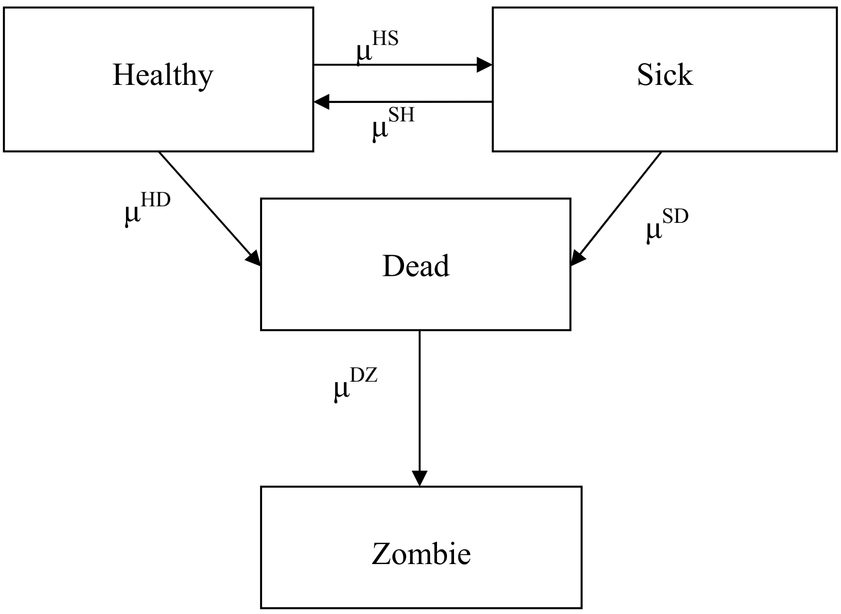
The required probability is obtained by integrating this expression over all values of u from 0 to 10.

This is

$$\int_0^{10} 0.0736\exp(-0.1u)du = \frac{0.0736}{-0.1}\left[\exp(-0.1u)\right]_0^{10}$$

$$= \frac{0.0736}{0.1}\left[1 - 0.3678)\right] = 0.465$$

Parts (i), (ii) and (iii) of this question were well answered by most candidates, but there were few good attempts at parts (iv), (v) and (vi). A minority of candidates produced an alternative transition diagram in part (i) as follows:



Full credit was given for this, and for answers to parts (ii) and (iii) which were consistent with it. In part (iii) some candidates derived the maximum likelihood estimate by applying the correct method to the wrong transition. In part (v)(b) it was possible to write the integral as follows:

$$\int_0^{10} \exp[-0.2(10-w)]*0.2*\exp(-0.1w)\,dw.$$

The evaluation is:

$$0.2*\int_0^{10} \exp(-2+0.2w)\exp(-0.1w)\,dw$$
$$=0.2\int_0^{10} \exp(-2+0.1w)\,dw$$
$$=0.2\exp(-2)\int_0^{10} \exp(0.1w)\,dw$$
$$=\frac{0.2}{0.1}\exp(-2)[\exp(1)-\exp(0)]$$
$$=2*0.1353*(2.718-1)=0.465$$

END OF EXAMINERS' REPORT