Institute of Actuaries of India

Subject CT4 – Models

May 2011 Examinations

INDICATIVE SOLUTIONS

Introduction

The indicative solution has been written by the Examiners with the aim of helping candidates. The solutions given are only indicative. It is realized that there could be other points as valid answers and examiner have given credit for any alternative approach or interpretation which they consider to be reasonable.

Solution 1:

(i)

(a) Let N_t be a Poisson process, $t \geq 0$, and let $Y_1, Y_2, ..., Y_j...$ be a sequence of independent and identically distributed random variables. Then, a Compound Poisson Process is defined by:

$$X_t = \sum_{j=1}^{N_t} Y_j$$

(b) Let $Y_1, Y_2, ..., Y_j...$ be a sequence of independent and identically distributed random variables. Define:

$$X_n = \sum_{j=1}^{n} Y_j$$

with initial condition $X_0 = 0$. Then, X_n constitutes a General Random Walk.

(ii)

(a) A Compound Poisson Process operates in continuous time. It has a discrete or continuous state space depending on whether the variables Y_j are discrete or continuous respectively.

It is important to mention that the state space of a Compound Poisson process depends on the state space of the underlying variables Y_j. Where a candidate does not mention this and simply mentions any one of the state spaces; half mark shall be deducted.

(b) A General Random Walk operates in discrete time. It has a discrete or continuous state space depending on whether the variables Y_j are discrete or continuous respectively.

(iii) An example of Compound Poisson Process is modelling total claims over time. An example of General Random Walk is modelling share prices or modelling inflation indices on a periodic basis.

[Total Marks - 6]

Solution 2:

(i) There is no single solution here. A good solution would cover the following points:

Benefits:

- Time saving would help in quicker response to regulatory actions.
- Freeing up internal staff for other important activities.
- External consultancy may bring better modelling expertise than internally available.
- External consultancy may have specialised modelling staff and may be able to deliver much faster than internal staff.

Disadvantages:

- Can be expensive
- Internal staff would need to spend time to understand the models subsequently and validate them.
- Dependability on external consultancy for future maintenance of models. If these are to be maintained in-house, considerable time may be spent in understanding them.
- Increased costs for any additional scenario testing, sensitivity testing etc.

(ii) There is no single solution here. A good solution would cover the following points:

Documentation should cover:
- Objective of the model
- Basis & Assumptions underlying the model
- Validation of input data
- How the model may be adapted or extended
- References to regulatory requirements, research papers, industry studies etc if used
- Important limitations of the model
- Key results of the model

[Total Marks - 7]

Solution 3:

(i) The null hypothesis is that the industry-wide proportion of causes of break-downs is consistent with the causes of break-downs of Ambassador Cars.

The total number of actual breakdowns of Ambassadors over last year is $3269 + 2627 + 1680 + 1037 + 839 + 415 = 9{,}867$.

We can derive the expected number of break-downs by multiplying total break-downs by expected proportions. Thus we have the following:

Cause of breakdown	Expected proportion of break-downs	Expected number of break-downs	Actual number of break-downs	A - E	$\frac{(A - E)^2}{E}$
Flat / Faulty Battery	32%	3157.44	3,269	111.56	3.94
Flat tyres	27%	2664.09	2,627	-37.09	0.52
Alternator	17%	1677.39	1,680	2.61	0.00
Starter Motor	11%	1085.37	1,037	-48.37	2.16
Fuel	9%	888.03	839	-49.03	2.71
Others	4%	394.68	415	20.32	1.05
		9867.00	**9867.00**		**10.37**

The test statistic, $X = \sum \frac{(A-E)^2}{E}$ is equal to 10.37

Here, we have 6 causes of break-downs. We haven't estimated any parameters but we have calculated the expected numbers by assuming the total is the same as for actual numbers. So the degrees of freedom is 6 – 1 = 5

Large values of X indicate excessive deviations, so we will test X against the upper 5% percentage point of the χ_5^2 distribution and say the test fails if $X > \chi_{5;0.95}^2$

The upper 5% point of the χ_5^2 distribution is 11.07

Since $X = 10.37 < \chi_{5;0.95}^2 = 11.07$ we do not have sufficient evidence to conclude that the causes of Ambassador break-downs are different to the industry-wide proportions.

(ii) The χ^2 test will fail to detect several defects that could be considerable importance for the management:

1. There could be a few large deviations offset by a lot of very small deviations. In other words, the χ^2-test could be satisfied although the data do not satisfy the distributional assumptions that underlie it. This is, in essence, because the χ^2-statistic summarises a lot of information in a single figure. For example, in the above test, there are far more Ambassador Cars breaking down due to flat or faulty battery but this is offset by other causes.
2. The graduation might be biased above or below the data by a small amount. The χ^2-statistic can often fail to detect consistent bias if it is small, but we should still wish to avoid it. Even if the graduation is not biased as a whole, there could be significant runs or clumps over which it is biased up or down. Because the χ^2-test is based on squared deviations, it tells us nothing about the direction of any bias or the nature of any lack of adherence to data

of a graduation, even if the bias is large or the lack of adherence manifest. To ascertain this there is no substitute for an inspection of the experience.

[Total Marks - 7]

Solution 4 :

[1 mark for stating the three methods correctly; 6 marks for bookwork to explain advantages & disadvantages of these methods of graduation; and 2 marks reserved for candidates' ability to comment on the insurance company's circumstances (large; been selling TA for a long time) and adapt the responses appropriately/recommend which method most suitable with appropriate justification for the company]

Three methods of graduation that the life company can use, along with their respective advantages and disadvantages are described below:

Graduation by fitting **Parametric Formula**: we assume that mortality can be modelled using a mathematical formula

Advantages

- Suitable with reasonably large experience data to be able to fit a parametric formula to the crude rates: The insurance company has been selling term assurances for a number of years and is considered 'large' so it is likely that the company would have sufficiently large data to consider this approach
- The rates will automatically be smooth
- It is easy to identify the mortality trends if the same formula is used
- The goodness of fit is usually satisfactory
- Calculations can be computerised
- Can give most weight to the ages where most data was available.

Disadvantages

- It is often difficult to find a single formula that fits over the whole age range
- If the formula used does not include enough parameters, it will not be flexible enough to follow the crude rates closely, which may result in over-graduation. If too many parameters are included, sudden bends may appear in the graduated curve, which may result in under-graduation.
- It can be very time consuming, even with computerisation.
- For practical use, it may not be sufficient to choose and fit a formula using statistical methods alone. It will also be necessary to inspect the results in the light of previous knowledge of mortality experience, especially at very young and very old ages where the data may be scarce. Therefore, it may be necessary to adjust the graduation to obtain a satisfactory final result.

Graduation by reference to a **Standard Table**: we assume that there is a simple relationship between the observed mortality and an appropriate standard table

Advantages

- This will be particularly useful if we do not have much data from experience in which we are interested
- the method can give good results on very scanty data
- you usually do not have to bother testing for smoothness
- knowledge of other tables is automatically brought into the graduation
- there should be little difficulty with the ends of the table, ie amount of extrapolation required is limited

Disadvantages

- reliability of results can be doubtful if there is little data (although this is true for other graduation methods as well)
- it is not always possible to find a suitable standard table (and thus adherence to data would be poor if this were the case)
- any errors in the original table will be repeated.

Graduation by **Graphical Method**: we draw a curve by hand on a graph of the crude estimates.

Advantages

- it can give good results even when data are scanty
- it is easy to make allowance for special features ("intrinsic roughness")
- it naturally allows weight to be given to those ages where most data is available
- it allows scope for individual judgement (experience)
- it can be done quickly, without the need for a computer
- it involves the actuary in the data, which may give him or her a better understanding of the rates.

Disadvantages

- a relatively high degree of skill is required
- the method only gives results to approximately 3 significant figures
- because of the scales involved, it is often necessary to draw the curve in two parts (although a transformation could be used)
- individual judgement can lead to bias and prejudice
- different results can be obtained from the same data
- it can be difficult to achieve a high degree of smoothness
- it is unclear as to how many degrees of freedom should be used when testing the data using a chi square test for overall adherence to the data.

[Total Marks - 9]

Solution 5:

(i)

Define:

- K_x as the curtate number of overs Sachin will play more, given he has already played "x" overs.
- ${}_k p_x$ as the probability of playing "k" overs more given that he has already played "x" overs.
- ${}_k q_x$ as the probability of getting dismissed before next "k" overs given that Sachin has already played "x" overs.
- e_x as the curtate expectation of life, i.e. the expected number of integer overs Sachin will face before getting out, given he has already played x overs.

Therefore we have,

$$e_x = E[K_x]$$

$$= \sum_{k=0}^{[25-x]} k.{}_k p_x . q_{x+k}$$

$$= \begin{array}{l} {}_1 p_x . q_{x+1} \\ + {}_2 p_x . q_{x+2} + {}_2 p_x . q_{x+2} \\ + {}_3 p_x . q_{x+3} + {}_3 p_x . q_{x+3} + {}_3 p_x . q_{x+3} \\ + \ldots \end{array}$$

$$= \sum_{k=1}^{[25-x]} \sum_{j=k}^{[25-x]} {}_j p_x . \; q_{x+j} \qquad \text{(summing columns)}$$

Given that $\sum_{j=k}^{[25-x]} {}_j p_x . q_{x+j}$ represents the probability of playing "k" overs after having played "x" over (${}_k p_x$) and then getting out at *any* time in subsequent (25-x-k) overs. Since the probability of second part is one, we can write this more simply as ${}_k p_x$

$$\therefore e_x = \sum_{k=1}^{[25-x]} {}_k p_x .$$

(iii) (a)

We can calculate this by estimating $\sum_{k=1}^{[25-x]} {}_k p_x$. given x = 18.

Since Sachin has already batted 18 overs, assume that he is at the start of the 19th over.
We can calculate survival probabilities based on the statistics provided.
Sachin has started the 19th over on 40 previous occasions, and on only 3 of these has he got out in the 19th over.
Therefore, ${}_1 p_{18} = \frac{37}{40}$
Similarly, ${}_2 p_{18} = \frac{33}{40}$, ${}_3 p_{18} = \frac{21}{40}$, and so on...

From this we have,

$$\therefore e_x = \sum_{k=1}^{[25-x]} {}_k p_x = 2.80$$

Thus, the expected number of overs that Sachin will face in total is 20.80 given that he has already played 18 overs.

(ii) (b)

The probability that Sachin will get out before completing 25th over is $\frac{37}{40}$ and hence probability that Sachin will still be batting at the start of the last over is $\frac{3}{40}$

(iii) Given x=18

$$Var[K_x] = \sum_{k=0}^{[25-x]} k^2 \cdot {}_k p_x \cdot q_{x+k} - e_x^2$$
$$= \frac{37}{40} \times \frac{4}{37} \times 1 + \frac{33}{40} \times \frac{12}{33} \times 2^2 + \ldots$$
$$= 10.25 - 7.84$$
$$= 2.41$$

Therefore, the variance of the number of overs Sachin is likely to face is 2.41.

Following from this,

Sachin's expected final score is 89 + 2.8 x 86.36 x 6 /100 = 104; and the variance of Sachin's final score is 2.41 x 86.36 x 6 / 100 = 12.49

[Total Marks - 10]

Solution 6:

(i)
Central exposed to risk is defined as the observed waiting time at age x, i.e. the time spent by a given life under observation. Under the given scenario, it will be the duration for which each life insurance policy has been in-force.

To calculate this exactly, for all ages, the following data will be required:

- Exact date of birth of all the policyholders
- Inception date for all policies
- Exact date of exit for all policies

(ii)

Let $P_x(t)$ denote the number of lives at time t aged x next birthday and suppose that time is measured in years from 31 December 2009.
From the survey, we have $P_x(0)$, $P_x(1)$ and $P_x(1.25)$ for all x.

Let S_x be the number of lapses during the investigation, aged x last birthday.
Now, define $P_x^{'}(t)$ as the number of lives at the time t aged x last birthday.

$$E_x^c = \int_0^{1.25} P_x^{'}(t)dt$$

Assume lapses occur uniformly from time t=0 to t=1 and from t=1 to t=1.25. Then,

$$E_x^c = \tfrac{1}{2}\left[P_x^{'}(0) + P_x^{'}(1)\right] + \tfrac{1}{8}\left[P_x^{'}(1) + P_x^{'}(1.25)\right]$$

Note that $P_x^{'}(0)$ is the number of lives at t=0 aged x last birthday, i.e. no. of lives at t=0 aged x+1 next birthday.

$$\Rightarrow P_x^{'}(t) = P_{x+1}(t)$$

Substituting this in the above equation,

$$E_x^c = \tfrac{1}{2}\left[P_{x+1}(0) + P_{x+1}(1)\right] + \tfrac{1}{8}\left[P_{x+1}(1) + P_{x+1}(1.25)\right] = \tfrac{1}{2}P_{x+1}(0) + \tfrac{5}{8}P_{x+1}(1) + \tfrac{1}{8}P_{x+1}(1.25)$$

(iii)

We made the following assumptions in part (ii):

1. We have implicitly assumed that the only decrement out of the population is lapses, since we are considering lapses as the only cause of 'exit' for lives under observation, while deriving the exposed to risk. This will not be true in reality since there will be deaths as well as maturities. However, given that the focus is on younger ages, deaths and maturities may not be significant and it may be appropriate to ignore these for the current investigation. Materiality of deaths and maturities would ultimately depend on product features, demographic characteristics of policyholders and proportion of policies lapsing compared with deaths/maturities.

2. We have assumed uniform distribution of lapses over the year. This may not be true, as lapses may tend to increase towards policy anniversaries, which may be concentrated towards the end of tax year.

[Total Marks – 11]

Solution 7:

(i)

First, we need to draw up a schedule showing new recruits (entries), leavers (exits) and censorings from the investigation. Let passing a paper be an event denoted by t. We then have the number at risk immediately before each t.

No. of papers (t)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15[1]
ENTRY	A E F X Y		B C		W			D					Z			
EXIT						C		X					Y		A	
CENSORED														D[2] B[2]		E F W[3] Z
Number at risk before t	0	5	5	7	7	8	7	7	7	7	7	7	7	7	5	4

Notes :

1. Being employed with SAC after passing all 15 papers is equivalent to "survival" as per definition. Thus, those employees who continue at SAC beyond qualification are censored at the end of the investigation at $t = 15$.
2. B is censored at t=13 since the investigation ends when she still hasn't qualified. D is also censored at t=13 since she is still in the firm (i.e. not an exit but a transfer to another office) though local office is unable to track exam progress.
3. W leaves the local office at t=11 however is still in employment at SAC until qualification. The investigation is clear that it seeks to know whether an associate continues until qualification in the firm (and not necessarily in the local office). Thus, even though there is no exam pass information for W between t=11 and t=15, there is the necessary information that W was at the firm throughout this period (and therefore at risk) as well as that he qualified and at stayed within the firm (i.e. survival). Hence, W is not censored at t=11 but at t=15.

The Kaplan-Meier estimate of the survival is a step function that starts at 1 and steps down every time an exit is observed. So, if we measure qualification in exam passes starting from no papers to 15, then:

$$\hat{S}_{T_{Grad}}(t) = 1 \, for \, 0 \leq t < 5$$

Out of 8 lives at risk at t=5, one is observed to exit. The next exit only occurs at t=7. So,

$$\hat{S}_{T_{Grad}}(t) = 1 - \tfrac{1}{8} = \tfrac{7}{8} \; for \; 5 \le t < 7$$

Just before t=7, there are 7 lives at risk. We observe one entry and one exit at t=7 and we assume that the exit occurs first. There are no more exits till t=12

$$\hat{S}_{T_{Grad}}(t) = \tfrac{7}{8} \times \left(1 - \tfrac{1}{7}\right) = \tfrac{7}{8} \times \tfrac{6}{7} = \tfrac{3}{4} \; for \; 7 \le t < 12$$

Similarly at t=12,

$$\hat{S}_{T_{Grad}}(t) = \tfrac{3}{4} \times \left(1 - \tfrac{1}{7}\right) = \tfrac{3}{4} \times \tfrac{6}{7} = \tfrac{9}{14} \; for \; 12 \le t < 14$$

The last exit is seen at t=14, at which time there are 5 lives at risk:

$$\hat{S}_{T_{Grad}}(t) = \tfrac{9}{14} \times \left(1 - \tfrac{1}{5}\right) = \tfrac{9}{14} \times \tfrac{4}{5} = \tfrac{18}{35} \; for \; 14 \le t < 15$$

Summarising this we have,

$$\hat{S}_{T_{Grad}}(t) = \begin{cases} 1 \; for \; 0 \le t < 5 \\ \tfrac{7}{8} \; for \; 5 \le t < 7 \\ \tfrac{3}{4} \; for \; 7 \le t < 12 \\ \tfrac{9}{14} \; for \; 12 \le t < 14 \\ \tfrac{18}{35} \; for \; 14 \le t < 15 \end{cases}$$

(ii)

The Poisson model assumes that μ is constant. The maximum likelihood estimate of μ is given by:

$$\hat{\mu}_{Grad} = \frac{d}{E^c_{Grad}}$$

Where:

$$E^c_{Grad} = \sum (b_i - a_i)$$

E^c_{Grad} = [14(A) +11(B) +3(C) +6(D) +15(E) +15(F) +11(W) +7(X) +12(Y) +3(Z)]/15 = 97/15 = $6\tfrac{7}{15}$

$$\hat{\mu}_{Grad} = \frac{4}{6\tfrac{7}{15}} = 0.6186$$

The Poisson estimate of q_{Grad} is then given by:

$1 - e^{-\hat{\mu}} = 1 - e^{-0.6186} = 0.4613$

using the invariance property of MLEs.

A 95% confidence interval for μ based on the Poisson model is:

$$\hat{\mu} \pm 1.96\sqrt{\frac{\hat{\mu}}{E^{c}_{Grad}}} = 0.6186 \pm 1.96\sqrt{\frac{0.6186}{6.4667}} = (0.0124, 1.2247)$$

So the 95% confidence interval for q_{Grad} based on this model is:
$(1 - e^{-0.0124}, 1 - e^{-1.2247}) = (0.0123, 0.7062)$

[Total Marks - 13]

Solution 8 :

(i) The score currently stands at 'Tie'. Whosever wins the next point will move into a 'Lead'. If the player in 'Lead' wins the subsequent point as well, he would win the tie-breaker. However, if the player in 'Lead' loses the next point, the score would be back at 'Tie'.

Since the probability of moving to the next state does not depend on the history prior to entering the state, Markov property holds.

The state space is defined as follows:

State	Description
T	Tie
L_F	Federer Leads
L_N	Nadal Leads
G_F	Federer Wins
G_N	Nadal Wins

Alternative solution: It is possible to construct a Markov chain where the terminal state merely indicates end of game without specifying who won. Full marks should be awarded to such alternative solution. State space for such a solution would be:

State	Description
T	Tie
L_S	Federer Leads
L_R	Nadal Leads
G	Game ends

Award two marks for clearly defining the state space.

(ii) The transition matrix is set out below:

$$\begin{bmatrix} 0 & 0.55 & 0.45 & 0 & 0 \\ 0.45 & 0 & 0 & 0.55 & 0 \\ 0.55 & 0 & 0 & 0 & 0.45 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Alternative solution: Transition matrix for the alternative solution stated above is set out below:

$$\begin{bmatrix} 0 & 0.55 & 0.45 & 0 \\ 0.45 & 0 & 0 & 0.55 \\ 0.55 & 0 & 0 & 0.45 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Award two marks for clearly setting out the transition matrix. The transition matrix shall be awarded marks in entirety, and zero marks to be awarded for where the transition matrix set out is incorrect.

(iii) The chain is reducible as it has two absorbing states – G_F and G_N.

Absorbing states have no period and the other three states have a period of 2. Thus, the chain is not a-periodic.

(iv) After two points from the tie, the tie-breaker would either be completed or be back to tie again.

The probability of returning to tie after two points is given by:

Probability of Federer winning the first point x Probability of Nadal winning the second point
+
Probability of Nadal winning the first point x Probability of Federer winning the second point

$= 0.55 \text{ x } 0.45 + 0.45 \text{ x } 0.55$
$= 0.495$

We need to find number of such cycles of returning to tie such that

$0.495^N = 1 - 0.95$

Solving the above equation:

$N = \dfrac{\ln 0.05}{\ln 0.495} = 4.26$

Since the game can finish in cycles of two points, the required number of cycles is 5 i.e. 10 points.

(v) After two points:

a. Nadal may have won the tie-breaker (probability of 0.2025 i.e. 0.45^2); or
b. Federer may have won the tie-breaker (probability of 0.3025 i.e. 0.55^2); or
c. Tie-breaker may have come back to tie (probability of 0.495).

Let F_T be the probability that Federer wins the tie-breaker that is presently tied.

Let N_T be the probability that Nadal wins the tie-breaker that is presently tied.

We have:

$N_T = 0.2025 + 0.495 \text{ x } N_T$

Solving:

$N_T = 0.401$

Probability that Federer eventually wins the tie-breaker is 0.599 ($1-N_T$). This can be verified by:

$F_T = 0.3025 + 0.495 \text{ x } F_T$

Solving:
$F_T = 0.599$

Probability of Nadal winning a point is 0.45. However, in order to win the game, Nadal would need to win at least two consecutive points at some point in the game. The probability of Nadal winning two consecutive points is lower than the probability of him winning a point – this is what one would reasonably expect.

[Total Marks - 16]

Solution 9 :

(i) What happens upon a comet collision is not influenced by previous comet collisions, if any. Thus, the future evolution of the process is independent of the past and Markov property, therefore, holds.

X(t) denotes the number of habitable planets. Process *X(t)* can take values 0,1,2,3,4.

(ii) The two inner and two outer planets can never become habitable. The four central planets, upon becoming habitable, will continue to be habitable regardless of future comet collisions. Absorbing state for the process is therefore, when all the four central planets become habitable i.e. $X(t) = 4$.

(iii) Comet collisions with the planet happen at the rate of 6 every million years. Of these, the comet collisions that can supply the missing elements to the planet happen at the rate of 3 every million years.

The process moves from k (number of habitable planets) to k+1 if a comet strikes one of the central planets that presently is not habitable. Since the total number of planets is 8 and only maximum 4 planets can ever become habitable, the transition rate for moving from state k to state k+1 is $3 \times \frac{4-k}{8}$ every million years.

(iv) Working in time units of million years, the generator matrix for the process is given by:

$$\begin{bmatrix} -12/8 & 12/8 & 0 & 0 & 0 \\ 0 & -9/8 & 9/8 & 0 & 0 \\ 0 & 0 & -6/8 & 6/8 & 0 \\ 0 & 0 & 0 & -3/8 & 3/8 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Please deduct half a mark where a candidate has not mentioned the unit of time applicable to the generator matrix.

Please deduct one mark each for one incorrect row in the generator matrix. For example, if one or more entries in a particular row are incorrect, one mark shall be deducted. If the incorrect entries span out over two rows, two marks shall be deducted and so on.

(v) Holding times are exponentially distributed with mean $\frac{8}{12-3k}$ million years.

(vi) Expected time until n planets become habitable is given by:

$$\sum_{k=0}^{n-1} \frac{8}{12-3k}$$

(a) Expected time until 1 planet becomes habitable is 8/12 (0.67) million years.
(b) Expected time until 2 planets become habitable is 14/9 (1.56) million years.
(c) Expected time until 3 planets become habitable is 26/9 (2.89) million years.
(d) Expected time until 4 planets become habitable is 50/9 (5.56) million years.

(vii) Expected time until the first planet becomes habitable is 0.67 million years. Thereafter, incremental expected time for subsequent planets is 0.89 million years, 1.33 million years and 2.67 million years respectively. Expected time for the first planet is the smallest, since a comet having the required materials could strike any one of the four planets and make it habitable. Thereafter, a comet with the required materials needs to strike one of only three remaining central planets to make one more planet habitable – resulting in an increased expected time. Once three central planets are habitable, a comet with the required materials needs to strike the only remaining planet that is not habitable. Thus, expected time taken to move from 3 habitable planets to 4 habitable planets is considerably longer.

[Total Marks -21]

[Total Marks – 100]