Institute of Actuaries of India

Subject CT4 – Models

November 2010 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:

Deterministic modeling uses a single set of input parameters and produces the results of calculations reflecting one scenario.

Stochastic model involves at least one input parameter varying in accordance with an assumed probability distribution. Therefore, the output of a stochastic model would also be a distribution of results reflecting a distribution of scenarios, rather than one single point estimate.

Deterministic model produces a single result, though sensitivity analysis can be carried out to gain more insights. Stochastic model produces a range of results, and therefore gives more insights. In particular, stochastic models not only produce a central estimate; but also capture the variation around the central estimate.

Deterministic models are generally easy to model and explain compared to stochastic models. Stochastic models are relatively difficult to build and explain, and often require substantial volume of data to fit probability distributions to the input parameters being varied.

[Total 4]

Solution 2:

(i) A process X_t is said to have independent increments if for all t and every u>0 the increment $(X_{t+u} - X_t)$ is independent of all the past of the process $\{X_s : 0 \leq s \leq t\}$

(ii) A stochastic process with a discrete state space has the Markov property if:

$P(X_t = a|X_{s_1} = x_1, X_{s_2} = x_2 \ldots X_{s_n} = x_n, X_s = x) = P(X_t = a|X_s = x)$ for $s \leq t$

(iii) Proof:

Consider $P(X_t \in A|X_{s_1} = x_1, X_{s_2} = x_2 \ldots X_{s_n} = x_n, X_s = x)$

Since $X_s = x$; the above probability equates to
$= P(X_t - X_s + x \in A|X_{s_1} = x_1, X_{s_2} = x_2 \ldots X_{s_n} = x_n, X_s = x)$

Applying Independent Increments property:
$= P(X_t - X_s + x \in A|X_s = x)$

Again using the fact that $X_s = x$:
$= P(X_t \in A|X_s = x)$

Thus, any process with independent increments has the Markov property.

[Total 5]

Solution 3:

In order to calculate the probability that the hazard rate observed by the scientist is greater than 0.6, we need to use the distribution of the Maximum Likelihood Estimator $\hat{\mu}$. We can use the Central Limit Theorem to find the asymptotic distribution of $\hat{\mu}$.

By Central Limit Theorem, we have:

$$\hat{\mu} \sim N\left(\mu, \frac{\mu}{E(V)}\right)$$

The expected waiting time for ith kitten, assuming a hazard rate of 0.05 is arrived at by multiplying the waiting time with the corresponding probability. For the kittens that die, we need to integrate – waiting time * probability of survival to the waiting time * death at the waiting time – over the range of possible waiting times. For the surviving kittens, the waiting time is 1 and we simply need to multiply it with the probability of survival.

$$E(V_i) = \int_0^1 0.05te^{-0.05t}\,dt + 1\,x\,e^{-0.05t}$$
$$\therefore E(V_i) = -e^{-0.05} + \frac{1}{0.05}(1 - e^{-0.05}) + e^{-0.05}$$
$$\therefore E(V_i) = 0.97541$$

Total expected waiting time, therefore, is 975.41 years.

Therefore:

$$\hat{\mu} \sim N\left(0.05, \frac{0.05}{975.41}\right)$$
$$\therefore \hat{\mu} \sim N(0.05, 0.0071596^2)$$
$$\therefore P(\hat{\mu} > 0.06) = 1 - \Phi\left(\frac{0.06 - 0.05}{0.0071596}\right)$$
$$\therefore P(\hat{\mu} > 0.06) = 1 - \Phi(1.40)$$
$$\therefore P(\hat{\mu} > 0.06) = 0.085$$

[Total 5]

Solution 4:

Right Censoring
Data are right censored if the censoring mechanism cuts short observations in progress e.g. ending of a mortality investigation before all the lives being observed have died.

Left Censoring
Data are left censored if the censoring mechanism prevents us from knowing when the entry into the state which we are observing took place like the medical studies involving regular examination, the discovery of a condition tells us only that the onset fell in the period since the previous examination.

Interval Censoring
Data are interval-censored if the observation allows us only to say that an event of interest fell within some interval of time e.g. mortality investigation where we might know only calendar year of death.

Random Censoring
If a censoring is random, then the time Ci, at which the observation of the i^{th} life is censored is a random variable.
If Ci < Ti where Ti = lifetime of the ith life, then the observation is censored e.g. in a pension scheme the employee leaving one employer and joining another.

[Total 6]

Solution 5:

(i) $P(D = 0) = e^{-(0.5*0.1)} = 0.951229$
$P(D = 1) = 1- e^{-(0.5*0.1)} = 0.04877$
(ii) $E(D) = 0*0.951229 + 1*0.04877 = 0.04877$

(iii) The probability density/mass function of V is:

$$f(V) = {}_{v}p_{2.5}\mu_{2.5+v} \qquad V < 0.50$$
$$= {}_{0.5}p_{2.5} \qquad V = 0.50$$

V is a random variable indicating waiting time. For the parrots that die, V is between 0 and 0.5. For parrots that survive, V is 0.5. Therefore, there is a probability mass at the point V=0.5.

(iv) $E(V) = \int_0^{0.5} 0.1te^{-0.1t}dt + 0.5 * e^{-0.1*0.5}$

Integrating by parts, the integral is:

$$[-te^{-0.1t}]_0^{0.5} - \int_0^{0.5} -e^{-0.1t}dt$$

$= -0.5e^{-0.05} - 10(e^{-0.05} - 1)$
$= 0.0121$

$E(V) = 0.0121 + 0.9513*0.5 = 0.4877$

(v) Surviving parrot contributes exactly 0.5 years to the waiting time. Dying parrots contribute less to the waiting time, depending on the time of the death. Assuming that parrots die roughly uniformly, they contribute about ¼ year to the waiting time. Therefore, expected waiting time is approximately equal to survival probability x ½ years + death probability x ¼ years. This translates to 0.4878, very close to the answer calculated above.

[Total 8]

Solution 6:

(i) Age x = [age last birthday at 1st April prior to start of a policy] + [no of 31st mach passed]
= age x last birthday on 1st April prior to death.

As age changes on 1 April of a calendar year (1st April to 31st March) each year, so rate year is calendar year starting on 1 April.

Age range at start of calendar year = x to x + 1

(ii)

(a) <u>Principle of Correspondence:</u>
A life alive at time t should be included in E_x^c if and only if, were that life to die immediately that life to be included in the death data θx.

(b) E_x^c= total number of years for which lives were exposed to the risk of dying whilst aged x last birthday on the immediately preceding 1st April.

Or

total number of years for which lives were exposed to the risk of dying whilst aged x + 1 next birthday on the immediately preceding 1st April during the Period of the Investigation.

x + f is the mean age of lives half way through the rate interval assuming force of mortality is constant over the calendar year rate interval.

Assuming birthdays are uniformly distributed over the calendar year, the average age at the start of the rate interval is x + ½ and halfway through the rate interval is x + 1.

So, x + f = x + 1.

[Total 8]

Solution 7:

(i) The generator matrix at time t is as follows:

$$A(t) = \begin{bmatrix} -0.15t & 0.1t & 0.05t & 0 \\ 0.3t & -0.65t & 0.15t & 0.2t \\ 0 & 0 & -0.6t & 0.6t \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

(ii) The matrix form of backward differential equations is:

$$\frac{\partial}{\partial s}P(s,t) = -A(s)P(s,t)$$

For this model:

$$\frac{\partial}{\partial s}P_{33}(s,t) = -[-0.6sP_{33}(s,t)]=0.6sP_{33}(s,t)$$

Similarly:

$$\frac{\partial}{\partial s}P_{13}(s,t) = -[-0.15sP_{13}(s,t) + 0.1sP_{23}(s,t) + 0.05sP_{33}(s,t)]$$

$$\therefore \frac{\partial}{\partial s}P_{13}(s,t) = [0.15sP_{13}(s,t) - 0.1sP_{23}(s,t) - 0.05sP_{33}(s,t)]$$

(iii) Separating the variables gives:

$$\frac{\frac{\partial}{\partial s}P_{33}(s,t)}{P_{33}(s,t)} = 0.6s$$

Changing the variable from *s* to *u*:

$$\frac{\partial}{\partial s}\ln P_{33}(u,t) = 0.6u$$

Integrating both sides w.r.t. u between the limits of u=s and u=t, we get:

$$[\ln P_{33}(u,t)]_s^t = \int_s^t 0.6u\,du = [0.3u^2]_s^t$$

$$\ln P_{33}(t,t) - \ln P_{33}(s,t) = 0.3(t^2 - s^2)$$

$$Since\ P_{33}(t,t) = 1, \ln P_{33}(t,t) = 0.\ Thus:$$

$$-\ln P_{33}(s,t) = 0.3(t^2 - s^2)$$

$$\therefore P_{33}(s,t) = e^{-0.3(t^2-s^2)}$$

[Total 9]

Solution 8:

i.

$$h(z,t) = h_o(t) * \exp(\beta, z_i^t)$$

Where h (z, t) is the hazard at duration t; $h_0(t)$ is the baseline hazard zi are the covariates and β is the vector of regression parameters

ii. $z1 = 1$ medical coaching, 0 otherwise ; $\beta1 = 0.08$
$z2 = 1$ new tuition method, 0 otherwise; $\beta2 = -0.05$
$z3 = 1$ boys, 0 otherwise; $\beta3 = 0.02$

iii. The baseline hazard refers to a girl child following a traditional method and attending the engineering coaching classes.

iv. The Covariate Confidence Interval consists of mean (-0.05) and standard deviation (0.05).

The parameter associated with the new tuition method is -0.05. Because the parameter is negative, the hazard of dropping out is reduced by the new tuition method.

Therefore the new tuition method does appear to improve the chances of a child continuing with his or her tuition classes.

The 95% confidence interval for the parameter new tuition method is
-0.05 + (1.96* 0.05) = (-0.148, 0.048), which includes zero.

So at the 5% significance level it is not possible to conclude that the new tuition method has improved the chances of children continuing to attend the classes.

[Total 9]

Solution 9:

(i) q_x is the probability that a life aged exactly x will die before reaching exact age x+1, and is called the initial rate of mortality.

m_x is called the central rate of mortality and represents the probability that a life alive between the ages of x and x+1 dies.

The relationship between m_x and q_x can be defined as:

$$m_x = \frac{q_x}{\int_0^1 {}_t p_x dt}$$

Or $$mx = \frac{qx}{1-qx/2}$$

(ii)

a. Uniform distribution of deaths

$_tq_x = t*q_x$

b. Balducci assumption

$_{1-t}q_{x+t} = (1-t)*q_x$

a. <u>Uniform Distribution</u>

$$\int_0^1 {}_tp_x dt = \int_0^1 (1-t*q_x)dt = \int_0^1 (1-t*(0.06))dt = 1-\left[\frac{t^2}{2}\right]_0^1 *(0.06)$$

=1-0.03=0.97

So

mx = 0.06/0.97 = 0.061856

<u>Balducci</u>

For consistency purpose, consider that ${}_1p_x = {}_tp_x \cdot {}_{1-t}p_{x+t}$

$${}_tp_x = \frac{{}_1p_x}{{}_{1-t}p_{x+t}} = \frac{1-0.06}{1-{}_{1-t}q_{x+t}} = \frac{0.94}{0.94+0.06*t}$$

$$\int_0^1 {}_tp_x dt = \int_0^1 \frac{0.94}{0.94+0.06t}dt = \frac{0.94}{0.06}*\left[\ln(0.94+0.06t]_0^1\right] = 15.6667 *(-)\ln(0.94)$$

= -15.6667*-0.06188=0.96938

mx = 0.06/0.96938 = 0.61895

b. The Balducci assumption implies a decreasing mortality rate over [x, x+1] and Uniform distribution gives an increasing mortality rate.

For a given number of deaths over the period, the estimated exposure would be highest if we assumed an increasing mortality rate.

We would expect the central rate to be highest for that with the lowest estimate exposure; hence mx under Balducci is greater than Uniform distribution.

[Total 13]

Solution 10:

(i)

"Undergraduation" occurs when too much emphasis is given to goodness of fit. Undergraduated rates adhere closely to the crude rates, but the resulting rates do not show a smooth progression from age to age.

"Overgraduation" occurs when too much emphasis is given to smoothness. Overgraduated rates show a smooth progression from age to age, but the resulting rates do not adhere closely to the crude rates.

(ii)

The chi-squared test is for the overall fit of the graduated rates to the data.
The test statistics is $\sum z_x^2$, where

$$z_x = \frac{(\theta_x - E_x \overset{o}{q}_x)}{\sqrt{E_x \overset{o}{q}_x (1-\overset{o}{q}_x)}}$$

However, considering the fact that q_x is very small, we redefined the z_x as below

$$z_x \approx \frac{(\theta_x - E_x \overset{o}{q}_x)}{\sqrt{E_x \overset{o}{q}_x}}$$

The calculations are given in the table below:

Average Age	**Θx**	**qx**	**qx**	**Expected Deaths**	**Zx**	**zx^2**
23	2	0.00889	0.00220	1.98	0.01421	0.00020
28	4	0.00833	0.00240	2.88	0.65997	0.43556
33	5	0.00923	0.00260	3.38	0.88116	0.77645
38	7	0.01000	0.00360	5.40	0.68853	0.47407
43	8	0.01091	0.00540	5.94	0.84523	0.71441
48	9	0.01500	0.00900	7.20	0.67082	0.45000
53	9	0.01385	0.01500	9.75	-0.24019	0.05769
58	5	0.01429	0.02300	8.05	-1.07498	1.15559

$$\sum z_x^2 = 4.06397$$

The test statistic has a chi-squared distribution with degrees of freedom given by number of age groups less 1 for the parametric function and further reduction for using the standard table.

The critical value of chi-squared distribution with 6 degree freedom at 5% level is 12.59.

Since 4.06397 < 12.59, there is no evidence to reject the null hypothesis that the graduated rates are the true rates underlying the crude rates.

(iii)

Signs test

a. The Signs test looks for overall bias

b. If the null hypothesis is true, the number of positive signs is distributed Binomial (8, 0.5).
From the table, we observe that there are 6 positive signs.
Prob (observed number of positive signs <=6) = 1- Prob(positive signs >6)

$= 1 - \{\binom{8}{7} + \binom{8}{8}\} * (0.5)^8$ = 1- 0.035156 = 0.965

This is greater than 0.025 (two tailed test)

c. We cannot reject the null hypothesis and conclude that the graduated rates are not systematically higher or lower than the crude rates.

Grouping of Signs test

a. The grouping of signs test looks for run or clumping of deviations with the same sign for overgraduation.

b. We have total 8 age groups with 6 positive signs and 2 negative signs. There is only one run in this analysis.
Pr (one positive run) =

$$\frac{\left\{\binom{5}{0}\binom{3}{1}\right\}}{\binom{8}{6}}$$ =3/28=0.10714

This is greater than 0.05 (using one-tailed test).

c. We accept the null hypothesis that graduated rates are true underlying the crude rates.

[Total 14]

Solution 11:

(i) Projections based on existing position:

Particulars	2011	2012	2013
Existing Customers	100,000	105,000	110,000
New Customers	5,000	5,000	5,000
Total Customers	105,000	110,000	115,000
Average Travel (KMs)	3,000	3,000	3,000

Total Travel (Million KMs)	315	330	345
Profits per KM	5	5	5
Projected Profits (Rs in Million)	1,575	1,650	1,725

(ii) Projected customer base

Year 2011

Customer base as at end of 2010 is:
$[65000 \quad 20000 \quad 15000]$

After allowing for additional 10,000 customers, the customer base for 2011 will be:
$[75000 \quad 20000 \quad 15000]$

Year 2012

Customer base as at end of 2011 is:

$$[75000 \quad 20000 \quad 15000] \times \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.5 & 0.3 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} = [67000 \quad 27000 \quad 16000]$$

After allowing for additional 10,000 customers, the customer base for 2012 will be:
$[77000 \quad 27000 \quad 16000]$

Year 2013

Customer base as at end of 2012 is:

$$[77000 \quad 27000 \quad 16000] \times \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.5 & 0.3 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} = [72200 \quad 29900 \quad 17900]$$

After allowing for additional 10,000 customers, the customer base for 2012 will be:
$[82200 \quad 29900 \quad 17900]$

(iii) Projected profits

Particulars	2011	2012	2013
Standard Class			
No of Customers	75,000	77,000	82,200
Average Travel	1,700	1,700	1,700
Total Travel (Million KMs)	127.50	130.90	139.74
Profit per KM	5	5	5
Total Profit (Rs in Millions)	**637.50**	**654.50**	**698.70**
Gold Class			

Particulars	2011	2012	2013
No of Customers	20,000	27,000	29,900
Average Travel	3,850	3,850	3,850
Total Travel (Million KMs)	77.00	103.95	115.12
Profit per KM	4.5	4.5	4.5
Total Profit (Rs in Millions)	**346.50**	**467.78**	**518.02**
Platinum Class			
No of Customers	15,000	16,000	17,900
Average Travel	9,375	9,375	9,375
Total Travel (Million KMs)	140.63	150.00	167.81
Profit per KM	4	4	4
Total Profit (Rs in Millions)	**562.50**	**600.00**	**671.25**
Total Profit	**1,546.50**	**1,722.28**	**1,887.97**

(iv) Markov chain can be constructed by altering the state space in the following manner:

Standard –Customers who travelled <2,500 KMs in the previous year
Gold – Customers who travelled >=2,500 KMs and < 5,000 KMs in the previous year
Platinum (I) – Customers who travelled >=2,500 KMs and < 5,000 KMs in the previous year and were Platinum (II) customers in the previous year
Platinum (II) – Customers who travelled >=5,000 KMs in the previous year

Note that division of Platinum into two state spaces is merely for application of Markov model and the customers will still continue to see only three classes as before.

The transition matrix is as follows

$$\begin{bmatrix} 0.7 & 0.2 & 0 & 0.1 \\ 0.5 & 0.3 & 0 & 0.2 \\ 0.3 & 0.4 & 0 & 0.3 \\ 0.3 & 0 & 0.4 & 0.3 \end{bmatrix}$$

(v) Various points that can be discussed are:

- ✓ How much confidence can be placed upon the validity of external research agency – is an internal marketing research required?
- ✓ Expenses of administration of this promotional scheme, especially IT systems to track and monitor customer travels
- ✓ What if the higher growth estimates upon execution of the scheme do not turn out to be true?
- ✓ Another form of promotional schemes – example, family discounts or off-season discounts
- ✓ Allowance for time value of money, taxation etc

- ✓ Estimation of transition probabilities for subsequent modifications proposed by Marketing Director

Additional information that can be furnished is sensitivity analysis on various parameters such as growth estimates, average distance travelled etc.

[Total 19]

****