EXAMINATION

September 2006

Subject CT4 — Models (includes both 103 and 104 parts) Core Technical

EXAMINERS' REPORT

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.

M A Stocker
Chairman of the Board of Examiners

November 2006

Comments

Comments on solutions presented to individual questions for this September 2006 paper are given below.

103 Part

Question A1 *This was reasonably well answered, even by the weaker candidates.*
In part (ii), very few candidates used the information in the question and calculated $\sum_{i=1}^{n}(t_i - s_i)$.

Question A2 *This was reasonably well answered.*
In part (iv), many candidates wrote down a suitable estimate, but failed to provide an explanation as required.

Question A3 *This was reasonably well answered.*
In part (i), many candidates attempted to describe the simple random walk rather than the general case.
In part (ii), very few candidates identified the correct state space for the compound Poisson process or general random walk.
In part (iii), credit was not given if the examples cited were not likely to be encountered by an actuary working in a professional capacity.

Question A4 *This was not well answered overall, but many of the stronger candidates did score highly.*
In part (i), some candidates incorrectly attempted to calculate the long-run probability of being in state B.
Part (ii) was generally well answered.
In part (iv), the stronger candidates provided good answers, but overall candidates did not score well here.

Question A5 *Overall this was poorly answered, although the stronger candidates did well. Many candidates failed to split the two states labelled B and C in the solution, giving instead a 3-state chain. Some marks were still awarded for the long-run probability calculations in part (iv), but such candidates were not able to calculate the required final answer.*

Question A6 *This was poorly answered by most candidates, even though some parts of the question had been asked in previous (103) exams.*
Marks were lost in all parts of the question. Many candidates did not make a serious attempt at part (iii)(c).

104 Part

Question B1 *This was well answered.*
Some candidates assumed a constant force of mortality, for which credit was not given. Some candidates struggled with the second calculation.

Question B2 *This was poorly answered overall, although some of the stronger candidates did manage to score highly.*
In part (ii), the question asked candidates to "derive an expression" and therefore we were looking for clearly set out steps here. Many candidates lost marks by not providing sufficient explanation of their working.

Question B3 *This was well answered overall.*
In parts (i) and (ii), candidates were asked to "estimate" and some indication was required of how the numerical estimate was reached.

Question B4 *This was not well answered overall.*
In part (i), many candidates did not calculate the correct exposed to risk. Marks were frequently lost because of insufficient working combined with an incorrect final answer. Candidates who wrote down the formulae they were using were given credit even if arithmetic slips were made.

Question B5 *This was very well answered by most candidates.*
The most common errors were: inconsistency in the assumed order of death and censoring at ages 51 and 54 $^{3}/_{12}$; and continuation of the estimated survival function after age 55.

Question B6 *This was reasonably well answered overall.*
Parts (i) and (ii) were poorly answered.
In part (iii), the main areas where candidates lost marks were: not correctly stating the null hypothesis; failure to identify the correct degrees of freedom to be used in the chi-squared test; and a failure to state relevant and clear conclusions to the tests.

103 Solutions

A1 (i) If the *i*th component is still working at the end of the test period its contribution to the likelihood is:

$$ {}_{t_i - s_i}p_{s_i} = \exp(-\mu(t_i - s_i)) $$

under the assumption of a constant force of failure.

If the *i*th component fails at time t_i its contribution to the likelihood is:

$$ {}_{t_i - s_i}p_{s_i}.\mu_{t_i} = \exp(-\mu(t_i - s_i)).\mu $$

under the assumption of a constant force of failure.

In both cases the contribution equals:

$$ \exp(-\mu(t_i - s_i)).\mu^{f_i} $$

(ii) Denote the total number of components used in the test by n. The likelihood for n independent components is:

$$ L = \prod_{i=1}^{n} \exp(-\mu(t_i - s_i)).\mu^{f_i} $$

$$ L = \exp(-\mu\sum_{i=1}^{n}(t_i - s_i)).\mu^{\sum_{i=1}^{n} f_i} $$

Now the rig contains 100 components at all times because it is fully loaded and failed components are immediately replaced, so $\sum_{i=1}^{n}(t_i - s_i) = 200(\text{years})$.

So $$ L = \exp\left(-200\mu\right)\cdot\mu^{\sum_{i=1}^{n} f_i} $$

$$ \ln L = -200\mu + \ln\mu.\sum_{i=1}^{n} f_i $$

$$ \frac{\partial \ln L}{\partial\mu} = -200 + \frac{\sum_{i=1}^{n} f_i}{\mu} $$

Setting this to zero the MLE is:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} f_i}{200}$$

To verify this is a maximum we see that:

$$\frac{\partial^2 \ln L}{\partial \mu^2} = -\frac{\sum_{i=1}^{n} f_i}{\mu^2} < 0$$

A2 (i) The generator matrix is

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

(ii) The distribution is exponential in both cases; with parameter σ in state A, ρ in state B.

(iii) The probability that the process stays in A throughout [0, t] is

$$\int_t^{\infty} \sigma e^{-\sigma s} ds = e^{-\sigma t}.$$

For $\sigma = 3$, we get $e^{-3t} = 0.2$
which gives t = - ln (0.2)/3 = 0.54 weeks.

(iv) The time spent in state A before the next visit to B has mean $1/\sigma$.

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

$\hat{\sigma}$ = (Number of transitions from A to B) / (Total time spent in state A up until the last transition from A to B).

[An alternative is to use the maximum likelihood estimator for σ, which is (Number of transitions from A to B)/Total time spent in state A).]

Similarly we can estimate $\hat{\rho}$.

(v) Testing whether the successive holding times are exponential variables and independent would be best. Any procedure which does this test is acceptable.

A3 (i) (a) A Poisson process with rate λ is an integer-valued process N_t, $t \geq 0$ with the following properties:

$N_0 = 0$;
N_t has independent increments;
N_t has stationary increments, each having a Poisson distribution, i.e.

$$P\left[N_t - N_s = n\right] = \frac{\left[\lambda(t-s)\right]^n e^{-\lambda(t-s)}}{n!}, \quad s < t, n = 0, 1, 2, ...$$

(b) Let N_t be a Poisson process, $t \geq 0$ and let $Y_1, Y_2, \ldots, Y_j, \ldots$, be a sequence of i.i.d. random variables. Then a compound Poisson process is defined by

$$X_t = \sum_{j=1}^{N_t} Y_j, \quad t \geq 0.$$

(c) Let $Y_1, Y_2, \ldots, Y_j, \ldots$, be a sequence of independent and identically distributed random variables and define

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

with initial condition $X_0 = 0$. Then $\left\{X_n\right\}_{n=0}^{\infty}$ constitutes a general random walk.

(ii) (a) A Poisson process operates in continuous time and has a discrete state space, the set of nonnegative integers.

(b) 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.

(c) A general random walk operates in discrete time. Again, this has a discrete or continuous state space according to whether the variables Y_j have a discrete or continuous distribution.

(iii) (a) Examples of a Poisson process:

- claims arriving to an insurance company through time
- car accidents reported over time
- arrival of customers at a service point over time

(b) A standard example of a compound Poisson process used by actuaries is for modelling the total amount of claims to an insurance company over time.

(c) Examples of a general random walk:

- modelling share prices daily
- inflation index, measured on say a monthly basis

Other reasonable examples received credit.

A4 (i) Probability that a company is never in state B is:

$$\Pr(A \to D) + \Pr(A \to A \to D) + \Pr(A \to A \to A \to D) + \ldots\ldots$$

$$= 0.03 + 0.92 \times 0.03 + 0.92^2 \times 0.03 + \ldots\ldots$$

$$= 0.03 \times \sum_{i=0}^{\infty} 0.92^i = \frac{0.03}{1 - 0.92} = 0.375$$

(ii) (a)

$$A^2 = \begin{pmatrix} 0.92 & 0.05 & 0.03 \\ 0.05 & 0.85 & 0.1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0.92 & 0.05 & 0.03 \\ 0.05 & 0.85 & 0.1 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0.8489 & 0.0885 & 0.0626 \\ 0.0885 & 0.725 & 0.1865 \\ 0 & 0 & 1 \end{pmatrix}$$

(b) Probability of default within 2 years for an A rated company 6.26%, so 6.26 defaults expected.

(iii) *Either*

Calculate revised transition probabilities based on the rating of bonds held by the investment manager after rebalancing:

$$A' = \begin{pmatrix} 0.97 & 0 & 0.03 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

(state B is unnecessary so this can be shown as 2 × 2 or 3 × 3)

$$A'^2 = \begin{pmatrix} 0.9409 & 0 & 0.0591 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

So the expected number of defaults is 0.0591 × 100 = 5.91.

Or

Required probability is

$$\Pr(A \to D) + \Pr(A \to A) \times \Pr(A \to D) + \Pr(A \to B) \times \Pr(A \to D)$$

$$= 0.03 + 0.92 \times 0.03 + 0.05 \times 0.03 = 0.0591$$

So expected defaults 5.91.

(iv) The expected number of defaults has been reduced by this strategy. (The variance of the number of defaults would also reduce.)

However it is not possible to tell whether the overall return is improved as this depends on the price at which bonds were bought and sold at the end of year 1.

The price of the debt sold may have been depressed by the companies having been downgraded to rating B, and the manager loses out on any increase in price if they recover.

The "downgrade trigger" strategy will incur dealing costs, which should be considered when comparing the returns.

A5 (i) Consider the following four states that the policyholder might be at the end of a year:

- the policyholder has made at least one claim both in the year just ended and the previous one (state A)
- the policyholder has made no claims in the year just ended but s/he made at least one claim during the previous year (state B)
- the policyholder has made at least one claim in the year just ended but not in the previous one (state C)
- the policyholder has made no claim during either the year ended or the previous one (state D)

If the year ended is year n, and X_n denotes the current state of the policyholder, then X_n constitutes a Markov chain.

(ii) The transition matrix is

$$P = \begin{pmatrix} 0.25 & 0.75 & 0 & 0 \\ 0 & 0 & 0.15 & 0.85 \\ 0.15 & 0.85 & 0 & 0 \\ 0 & 0 & 0.10 & 0.90 \end{pmatrix}$$

(iii) The chain has a finite number of states (A,B,C,D). In order to show that it has a stationary distribution, it suffices to show that it is irreducible and aperiodic.

It is apparent from the transition matrix above that any state can be reached from any other; hence the chain is irreducible.

The chain is also aperiodic since for states A, D the state can remain at the same state after one step, while for states B, C the state may return to its current state after 2 or 3 steps.

Hence the chain has a stationary distribution (which is unique).

(iv) The set of equations is given (in matrix from) by $\pi P=\pi$, where $\pi = (\pi_A, \pi_B, \pi_C, \pi_D)$ denotes the stationary distribution.

Using the transition matrix from (ii) above we obtain the equations

$$
\begin{array}{llllll}
0.25\,\pi_A + & & +0.15\,\pi_C & & = \pi_A & (1) \\
0.75\,\pi_A + & & +0.85\,\pi_C & & = \pi_B & (2) \\
 & 0.15\,\pi_B & & +0.10\,\pi_D & = \pi_C & (3) \\
 & 0.85\,\pi_B & & +0.90\,\pi_D & = \pi_D &
\end{array}
$$

Discard the last of these equations and use also that the stationary probabilities must also satisfy

$$\pi_A + \pi_B + \pi_C + \pi_D = 1 \qquad (4)$$

Equation (1) gives

$$0.75\,\pi_A = 0.15\,\pi_C \qquad (5)$$

Or $5\,\pi_A = \pi_C$

Substituting (5) into (2) yields immediately

$$\pi_B = \pi_C$$

and inserting this into (3) we get

$$\pi_D = \frac{17}{2}\pi_B.$$

In view of the above, we obtain now from (4) that

$$\pi_B\left(\frac{1}{5}+1+1+\frac{17}{2}\right)=1 \;\Rightarrow \pi_B = \frac{10}{107}.$$

Hence the other probabilities are

$$\pi_A = \frac{2}{107},\; \pi_C = \frac{10}{107},\; \pi_D = \frac{85}{107}.$$

The proportion of policyholders who, in the long run, make at least one claim in a given year is

$$\pi_A + \pi_B = \frac{12}{107}.$$

A6 (i) The probability that an event occurs during the short time interval between t and $t + h$ is approximately equal to $\lambda(t)\, h$ for small h where $\lambda(t)$ is called the rate of the process. For a time-inhomogeneous process, $\lambda(t)$ depends on the current time t; for a time-homogeneous process it is independent of time.

(ii) (a) Divide the time period into intervals of a suitable size, say one month. Estimate the arrival rate separately for each time period.

See if the observed data match the pattern which would be expected if the model were accurate and if the parameters had their values given by their estimates.

If not, the model should be revised.

(b) A goodness of fit test, such as the chi-squared test, should be carried out for each time period chosen.

Tests for serial correlation [e.g. portmanteau test] should use the whole data set at once.

(iii) (a) This implies that claims are seasonal with period 12 months, and that claims in the peak (presumably winter) are double those at the low point of the year.

This would be reasonable if in a climate where driving conditions are worse in winter.

(b) Kolmogorov forward equations:

$$\frac{\partial}{\partial t} P(s,t) = P(s,t).A(t) \qquad\qquad t \geq s$$

Where:

$$A(t) = \begin{pmatrix} -\lambda(t) & \lambda(t) & & \\ & -\lambda(t) & \lambda(t) & \\ & & -\lambda(t) & \ddots \\ & & & \ddots \end{pmatrix}$$

(c) Consider the case $j > 0$,

$$\frac{\partial}{\partial t} P_{0j}(s,t) = \lambda(t).P_{0,j-1}(s,t) - \lambda(t).P_{0j}(s,t) \qquad\qquad \text{(I)}$$

with $P_{0j}(s,s) = 0$

If solution is of the form

$$P_{0j}(s,t) = \frac{(f(s,t))^j.\exp(-f(s,t))}{j!}$$

LHS of I

$$(j.(f(s,t))^{j-1} - f(s,t)^j).\frac{\exp(-f(s,t))}{j!}.\frac{d}{dt}f(s,t)$$

RHS of I

$$\lambda(t).\frac{f(s,t)^{j-1}}{(j-1)!}.\exp(-f(s,t)) - \lambda(t).\frac{f(s,t)^j.\exp(-f(s,t))}{j!}$$

These are equal if

$$\frac{\partial}{\partial t}f(s,t) = \lambda(t)$$

Now

$$\int_s^t \lambda(v)dv = \int_s^t (3+\cos(2\pi v))dv$$

$$= \left[3v + \frac{1}{2\pi}\sin(2\pi v)\right]_s^t$$

$$= 3(t-s) + \frac{1}{2\pi}\left[\sin(2\pi t) - \sin(2\pi s)\right] \equiv f(s,t)$$

this satisfies the boundary condition.

Consider the case $j = 0$

$$\frac{\partial}{\partial t}P_{00}(s,t) = -\lambda(t).P_{00}(s,t) \qquad \text{(II)}$$

with boundary condition $P_{00}(s,s) = 1$

Need to verify that $P_{00}(s,t) = \exp(-f(s,t))$ satisfies II

LHS of II

$$-\exp(-f(s,t)).\frac{\partial}{\partial t}(f(s,t)) = -P_{00}(s,t).\lambda(t)$$

and $P_{00}(s,s) = 1$

(d) Solution is of the same form, except that for the homogeneous case $f(s,t) = \lambda(t-s)$.

104 Solutions

B1 $_{0.25}p_{80} = 1 - {}_{0.25}q_{80} = 1 - 0.25 \times q_{80}$
under the assumption of a uniform distribution of deaths (UDD) between ages 80 and 81.

From ELT 15, $q_{80} = 0.05961$, so

$$_{0.25}p_{80} = 1 - 0.25 \times 0.05961 = 0.98510$$

ALTERNATIVE 1

Under UDD we have, for $0 \leqslant s < t \leqslant 1$,

$$_{t-s}q_{x+s} = \frac{(t-s)q_x}{1-sq_x}.$$

Putting $t = 0.75$, $s = 0.5$ and $x = 80$, therefore,

$$_{0.75-0.5}q_{80+0.5} = \frac{0.25q_{80}}{1-0.5q_{80}}, \text{ and so}$$

$$_{0.25}p_{80.5} = 1 - \frac{0.25q_{80}}{1-0.5q_{80}}.$$

Using ELT15, this is evaluated as

$$1 - \frac{0.25(0.05961)}{1-0.5(0.05961)} = 1 - \frac{0.01490}{0.97020} = 1 - 0.01536 = 0.98464$$

ALTERNATIVE 2

Using $_{t}p_x = {}_{s}p_x \cdot {}_{t-s}p_{x+s}$,

$$_{0.75}p_{80} = {}_{0.5}p_{80} \cdot {}_{0.25}p_{80.5}$$

Using an assumption of UDD between ages 80 and 81, we have

$$_{0.5}p_{80} = 1 - 0.5 \times 0.05961 = 0.97020$$

$$_{0.75}p_{80} = 1 - 0.75 \times 0.05961 = 0.95529$$

So, $\quad {}_{0.25}p_{80.5} = \dfrac{{}_{0.75}p_{80}}{{}_{0.5}p_{80}} = \dfrac{0.95529}{0.97020} = 0.98463$

B2 (i) (a) The age definition changes 6 months before/after each birthday, so this is a life year rate interval.

(b) Lives are aged x - ½ at the start of the rate interval.

(ii) Under the principle of correspondence the age definition of deaths and census should correspond, which they do here. So we do not need to adjust the census information.

The exposed to risk is given by $E_x^c = \int_0^3 P_x(t)\,dt$.

Assuming $P_x(t)$ is linear over calendar years, we can approximate this to

$$E_x^c = \sum_0^2 \frac{1}{2}\left(P_x(t) + P_x(t+1)\right) \text{, where } t \text{ is measured from 1 January 2002}$$
$$= \left(\frac{1}{2}P_x(0) + P_x(1) + P_x(2) + \frac{1}{2}P_x(3)\right)$$

(iii) The age definitions for deaths and census no longer correspond. So, we need to adjust the census information for those companies who supply details of $P_x^*(t)$.

Assuming birthdays are uniformly distributed over the calendar year, we can approximate $P_x(t) \approx \frac{1}{2}\left(P_{x-1}^*(t) + P_x^*(t)\right)$.

And the exposed to risk is then:

$$E_x^c = \sum_0^2 \frac{1}{2}\left(P_x(t) + P_x(t+1)\right)$$
$$= \sum_0^2 \frac{1}{2}\left(\frac{1}{2}\left(P_{x-1}^*(t) + P_x^*(t)\right) + \frac{1}{2}\left(P_{x-1}^*(t+1) + P_x^*(t+1)\right)\right)$$
$$= \frac{1}{4}\left(P_{x-1}^*(0) + P_x^*(0)\right) + \frac{1}{2}\left(P_{x-1}^*(1) + P_x^*(1) + P_{x-1}^*(2) + P_x^*(2)\right) + \frac{1}{4}\left(P_{x-1}^*(3) + P_x^*(3)\right)$$

B3 (i) The hazard for a female patient is:

$$h_f(t) = h_0(t) \times \exp(0 + \beta_2 z_2)$$

and the hazard for a male patient is:

$$h_m(t) = h_0(t) \times \exp(\beta_1 \times 1 + \beta_2 z_2)$$

Using $\hat{\beta}_i$ to denote our estimate of β_i, we know from A that, if the model is correct,

$$h_m(t) = 1.02 \times h_f(t) \text{, so that:}$$

$$h_0(t) \times \exp\left(\hat{\beta}_1 + \hat{\beta}_2 z_2\right) = 1.02 \times h_0(t) \times \exp\left(\hat{\beta}_2 z_2\right)$$

$$\Rightarrow \exp(\hat{\beta}_1) = 1.02$$

$$\Rightarrow \hat{\beta}_1 = \ln(1.02) = 0.0198$$

And similarly, from B, we know that:

$$h_0(t) \times \exp\left(\hat{\beta}_1 z_1 + 0\right) = 1.05 \times h_0(t) \times \exp\left(\hat{\beta}_1 z_1 + \hat{\beta}_2 z_2\right)$$

$$\Rightarrow 1 = 1.05 \times \exp\left(\hat{\beta}_2\right)$$

$$\Rightarrow \hat{\beta}_2 = \ln\left(1/1.05\right) = -0.0488$$

(ii) The hazard for a male patient who has been given the new treatment is:

$$h_{m,n}(t) = h_0(t) \times \exp(\beta_1 \times 1 + \beta_2 \times 1)$$

$$= h_0(t) \times \exp(0.0198 - 0.0488)$$

$$= h_0(t) \times \exp(-0.029)$$

$$= 0.9714 \times h_0(t)$$

The hazard for a female patient given the existing treatment is the baseline hazard.

Hence, the ratio of the hazard for a male patient who has been given the new treatment to that for a female patient given the existing treatment is:

$$\frac{h_{m,n}(t)}{h_0(t)} = 0.9714$$

ALTERNATIVELY

Candidates may recognise that the proportions given in A and B can be combined to give:

$$\frac{h_{m,n}(t)}{h_{f,e}(t)} = \left[\frac{h_{m,x}(t)}{h_{f,x}(t)}\right] \times \left[\frac{h_{x,n}(t)}{h_{x,e}(t)}\right] = 1.02 \times \frac{1}{1.05} = 0.9714$$

(iii) The probability of death is given by:

$$\begin{aligned} 1 - S_{m,n}(3) &= 1 - \exp\left\{-\int_0^3 h_{m,n}(s)ds\right\} \\ &= 1 - \exp\left\{-\int_0^3 0.9714 \times h_0(s)ds\right\} \\ &= 1 - \exp\left\{0.9714 \times \left(-\int_0^3 h_0(s)ds\right)\right\} \\ &= 1 - \left(e^{-\int_0^3 h_0(s)ds}\right)^{0.9714} \end{aligned}$$

B4 (i) Let the age individual i enters observation be a_i and the age that individual i leaves observation be b_i. Define an indicator variable d_i such that $d_i = 0$ if individual i is not observed to die and $d_i = 1$ if individual i dies.

Measure all ages in years since exact age 60.

The estimate of q_{60} using the Binomial model is:

$$\hat{q}_{60} = \frac{\sum_{i=1}^{10} d_i}{\sum_{i=1}^{10}\left(1 - a_i - \left[(1-d_i)(1-b_i)\right]\right)}.$$

The denominator in this formula shows that for persons who do not die ($d_i = 0$) the exposed to risk is $b_i - a_i$ and for persons who die ($d_i = 1$) the exposed to risk is $1 - a_i$.

Thus the relevant calculations are shown in the table below (all durations are in years).

Person	a_i	b_i	d_i	$1 - a_i$	$1 - b_i$	$1 - a_i - (1 - d_i)(- b_i)$
1	0	6/12	0	1	6/12	6/12
2	1/12	1	0	11/12	0	11/12
3	1/12	3/12	1	11/12	9/12	11/12
4	2/12	1	0	10/12	0	10/12
5	3/12	9/12	1	9/12	3/12	9/12
6	4/12	1	0	8/12	0	8/12
7	5/12	11/12	1	7/12	1/12	7/12
8	7/12	1	0	5/12	0	5/12
9	8/12	10/12	1	4/12	2/12	4/12
10	9/12	1	0	3/12	0	3/12
Totals			4			74/12

Therefore $\hat{q}_{60} = \dfrac{4}{74/12} = 0.6486$.

ALTERNATIVELY

Take the central exposed to risk, $\sum_{1}^{10}(b_i - a_i)$ (in years) and add ½d_{60} to give the initial exposed to risk.

This involves estimating q_{60} using the formula

$$\hat{q}_{60} = \frac{d_{60}}{E_{60}^c + 0.5d_{60}} = \frac{4}{(59/12) + 2} = \frac{4}{83/12} = 0.5783.$$

[This approach is inferior to the first, as it does not use all the information available in the data, and involves the assumption that the deaths take place, on average, half way through the year.]

(ii) Strengths of Binomial model

- avoids numerical solution of equations

- can be generalised to give the Kaplan-Meier estimate

Weaknesses of Binomial model

- need to compute an initial exposed-to-risk is a pointless complication if census-type data are available

- not so easily generalised as two-state or Poisson models to processes with more than one decrement, and not so easily generalised as two-state model to increments

- estimate of q_x has a higher variance than that of the two-state Poisson models (though the difference is very small unless mortality is very high)

B5 (i) There will be Type I censoring of lives that survive to age 55 years.

There will be random censoring of lives that withdraw before age 55 years.

(ii) The calculations are shown in the table below, where durations are measured in years since the 50th birthday.

Using the convention that, when deaths and withdrawals are observed at the same duration, deaths occur first:

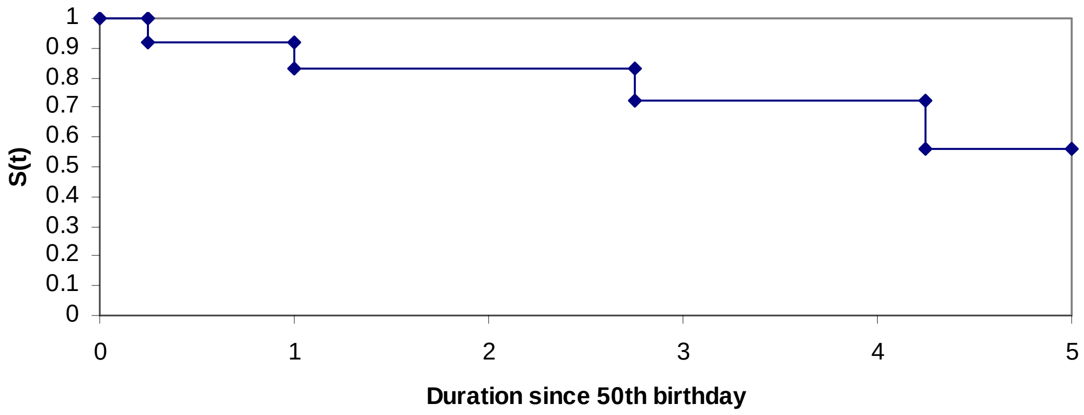
t_j	N_j	d_j	c_j	d_j / N_j	$\hat{\Lambda}_t = \sum_{t_j \le t} (d_j / N_j)$
0	12				
0.25	12	1	1	0.0833	0.0833
1.00	10	1	2	0.1000	0.1833
2.75	7	1	2	0.1429	0.3262
4.25	4	1	3	0.25	0.5762

Since $\hat{S}(t) = \exp(-\hat{\Lambda}_t)$

the estimated survival function is

t	$\hat{S}(t)$
$0 \le t < 0.25$	1.0000
$0.25 \le t < 1.00$	0.9201
$1.00 \le t < 2.75$	0.8325
$2.75 \le t < 4.25$	0.7217
$4.25 \le t < 5.00$	0.5620

(iii)



B6 (i) (a) The general form is

$$\mu_x = (\text{polynomial(1)}) + \exp(\text{polynomial(2)}),$$

where polynomial (1) takes the form

$$\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots$$

and polynomial (2) takes the form

$$\beta_0 + \beta_1 x + \beta_2 x^2 + \ldots.$$

(b) In the case of the Gompertz formula $\mu_x = Bc^x$, then putting

$$B = \exp(\beta_0) \text{ and } c = \exp(\beta_1),$$

we can re-write the formula as

$$\mu_x = \exp(\beta_0)\exp(\beta_1 x) = \exp(\beta_0 + \beta_1 x),$$

which is of the required form if

$$\alpha_i = 0 \text{ for all } i$$

and

$$\beta_i = 0 \text{ for } i = 2, 3, \ldots.$$

Similarly the Makeham formula $\mu_x = A + Bc^x$
can be expressed in the required form by putting

$$A = \alpha_0,\; B = \exp(\beta_0) \text{ and } c = \exp(\beta_1).$$

(ii) (a) The Gompertz formula written

$$\mu_x = \exp(\beta_0 + \beta_1 x)$$

is an exponential function which implies that the rate of increase of mortality with age is constant.

This is often a reasonable assumption for ordinary lives at middle ages and older ages.

In the special case of the impaired lives known to be suffering from a degenerative disease, it is plausible to suppose that the rate of increase of mortality might increase with age.

The term $b_2\left(x+\frac{1}{2}\right)^2$ in the formula can allow for this possibility.

(b) The graduation can be achieved by

maximum likelihood estimation of the parameters

or by ordinary least squares regression

of $\log\left[\hat{\mu}_{x+\frac{1}{2}}\right]$ on $x+\frac{1}{2}$ and $\left(x+\frac{1}{2}\right)^2$.

(iii) (a) The null hypothesis is that there is no difference between the graduated rates and the underlying rates in the population from which the crude rates are derived.

To test overall goodness-of-fit we use the chi-squared test.

$$\sum_x z_x^{\,2} \sim \chi^2_m,$$

where m is the number of degrees of freedom.

In this case, we have 8 ages, but 3 parameters were estimated when performing the graduation, so $m = 5$.

The calculations are shown in the table below.

Age x last birthday	z_x	z_x^2
50	-0.12031	0.01447
51	-0.20055	0.04022
52	-0.24749	0.06125
53	0.11341	0.01286
54	-0.79336	0.62942
55	-0.66436	0.44137
56	-0.44369	0.19686
57	-0.35225	0.12408
Sum		1.52053

The critical value of the chi-squared distribution with 5 degrees of freedom at the 5 per cent level is 11.07.

Since $1.52052 \ll 11.07$, we do not reject the null hypothesis and conclude that the graduation adheres satisfactorily to the data.

(b) To test for bias we use EITHER the Signs Test or the Cumulative Deviations test.

Signs Test

The test statistic, P, is the number of signs that is positive.

Under the null hypothesis, $P \sim \text{Binomial}(8,0.5)$

In this case $P = 1$, and Prob[$P \leq 1$] = 0.0352.

Since this probability > 0.025 (two-tailed test) we do not reject the null hypothesis.

We conclude that the graduated rates are not biased above or below the crude rates.

Cumulative deviations test

The test statistic

$$\frac{\sum_x (\hat{\mu}_{x+\frac{1}{2}} E_x - \overset{o}{\mu}_{x+\frac{1}{2}} E_x)}{\sqrt{\sum_x \overset{o}{\mu}_{x+\frac{1}{2}} E_x}} \sim \text{Normal}(0,1)\,.$$

The calculations are shown in the table below.

Age x *last birthday*	$\hat{\mu}_{x+\frac{1}{2}} E_x - \overset{o}{\mu}_{x+\frac{1}{2}} E_x$	$\overset{o}{\mu}_{x+\frac{1}{2}} E_x$
50	-0.63	27.63
51	-1.06	28.06
52	-1.32	28.32
53	0.61	29.39
54	-4.13	27.13
55	-3.20	23.20
56	-2.03	21.03
57	-1.72	23.72
Sum	-13.48	208.48

The value of the test statistic is therefore

$(-13.48/\sqrt{208.48}) = -0.9335.$

using a two-tailed test, the absolute value of the test statistics is less than 1.96, so we do not reject the null hypothesis.

We conclude that the graduated rates are not biased above or below the crude rates.

(c) To test for the existence of individual ages at which the graduated rates depart greatly from the observed rates we can use the Individual Standardised Deviations Test.

There are no ages at which the absolute value of z_x exceeds 1.96.

Therefore we do not reject the null hypothesis and conclude that there are no outliers.

END OF EXAMINERS' REPORT