Faculty of Actuaries **Institute of Actuaries**

EXAMINATIONS

April 2001

Subject 104 — Survival Models

EXAMINERS' REPORT

Examiners' Comments:

The last three examinations have seen very high numbers of candidates for Subject 104 together with disappointingly low pass rates and a large number of inadequately prepared candidates obtaining very low marks.

Subject 104 requires a working knowledge of Subjects 101 and 102. The evidence from examinations is that many candidates have but a superficial knowledge of key topics in 104, eg. the stochastic basis of life contingencies, the statistical basis of graduation, proportional hazards models.

We now make some specific comments on the difficulties experienced by candidates in each question in the April 2001 paper.

1 The main cause of lost marks was the failure to distinguish clearly between the events whose probabilities were given in (a).

2 The majority of candidates either had very sketchy knowledge of Thiele's equation, or were unable to apply their knowledge of Thiele's equation to the special case described. Marks were lost in (iii) as a result of using the A67-70 rather the a(55) functions.

3 Marks were mainly lost in (i)(c) and (ii) as a result of incorrectly allowing for the retrospective claims.

4 In (i) many candidates confused assessing if there was a linear relationship between $\overset{\circ}{q}_x$ and q_x^s with the goodness of fit of $\hat{q}_x$ to q_x^s.

In (ii) marks were often lost because an expression for the weighting function was just stated rather than determined.

In (iii) the impact of the linear relationships on the third differences of the rates was rarely stated.

5 There were various common errors including the failure to value benefits payable immediately on death and to allow for changing mortality. The most common mistake was to use

$$\bar{a}_{70} \simeq (1 + \tfrac{1}{2}i)a_{70}$$

6 In (i) the most common mistake was the failure to say <u>which</u> calendar year was the rate interval.

In (ii)(a) only a minority of candidates attempted to find the age definition at <u>the date of the census</u> of $P_x(t)$ t=½, 1½, 2½. In (ii)(b) the assumption of a constant force was rarely mentioned.

7 The main causes of poor marks were:

(i) writing down rather than deriving the sampling distributions of the test
statistic
(ii) not defining the test statistics
(iii) using 'two tailed' rather than 'one tailed' test rejection regions.

Much of the knowledge acquired in Subject 101 had apparently been forgotten.

8 Many candidates had no idea what was a random variable and what was an expected value. Statements like

$$\mathrm{Var}(a_{40:\overline{20}|}) = E[(a_{40:\overline{20}|})^2] - (E[a_{40:\overline{20}|}])^2$$

were common.

Some candidates failed to realise that a temporary immediate annuity is payable in arrears. If the annuity is payable in advance several parts of the question become much more straightforward.

9 (ii)(a) was only attempted by a minority of candidates. Some candidates incorrectly assumed that an estimate of q_{70} was required, calculating E_{70}=4.5 and θ_{70}=3 and then finding the actuarial estimate of q_{70}.

In (ii)(b) many candidates failed to construct the likelihood from the data and derive the maximum likelihood estimate. In many cases the actuarial estimate of q_{70} was given instead of $\hat{q}_{70}$.

10 Simple algebraic errors, eg. logging additive functions detracted from some good attempts. There were few serious attempts at (iii) and (iv).

1 (a) ${}_tp_x^{II}$ represents the probability of the event of being in state I at time/age $x + t$ given that the life was in state I at time/age x. Movements in the interval $(x, x + t)$ are not restricted.

${}_tp_x^{\overline{II}}$ represents the probability of the event of being in state I at time/age $x + t$ given that the life was in state I throughout the interval $(x, x + t)$.

(b) In this model the probabilities are equal because there is no way of returning to state I once a life has left this state.

2 (i) $$\frac{\partial({}_tV)}{\partial t} = \delta\,{}_tV + P - \mu_{x+t}\,({}_tV - {}_tV) = \delta\,{}_tV + P$$

i.e. ${}_tV' - \delta\,{}_tV = P$.

(ii) To solve, multiply by $e^{-\delta t}$ to get:

$$\left(e^{-\delta t}\,{}_tV\right)' = Pe^{-\delta t},$$

$$e^{-\delta t}\,{}_tV = (-P/\delta)\,e^{-\delta t} + c$$

and ${}_0V = 0 \Rightarrow c = P/\delta$

Hence ${}_tV = e^{\delta t}\,(P/\delta)\,(1 - e^{-\delta t})$

$= P(e^{\delta t} - 1) / \delta = P\bar{s}_{\overline{t}|i}$, with $1 + i = e^{\delta}$.

(Alternatively verify directly that $P\bar{s}_{\overline{t}|i}$, satisfies the equation <u>and</u> the initial condition, ${}_0V = 0$.)

(iii) $P\bar{s}_{\overline{20}|0.04} = S = 20{,}000a_{65} = 20{,}000 \times 9.790902$

$\Rightarrow P = 6{,}447.80$

3 (i) (a) $P = A_x / \ddot{a}_{x:\overline{m}|}$

(b) $A_{x+t} - 0$

(c) $\left(P\ddot{a}_{x:\overline{m}|} - A^{1}_{x:\overline{t}|}\right) / v^t \ {}_tp_x$

(iii) $P\ddot{a}_{x:\overline{m}|} = A_x = A^{1}_{x:\overline{t}|} + v^t \ {}_tp_x \ A_{x+t}$

$\Rightarrow A_{x+t} = \left(P\ddot{a}_{x:\overline{m}|} - A^{1}_{x:\overline{t}|}\right) \big/ v^t \ {}_tp_x$, as required.

A solution using commutation functions is also possible.

4 (i) Plot $\hat{q}_x$ against q_x^s and look for an approximate straight line fit.

(ii) At each age there will be a different sample size/exposed to risk, E_x. This will usually be largest at ages where many term assurances are sold e.g. 25 to 50 and smaller at other ages.

The estimation procedure should pay more attention to ages where there are lots of data. These ages should have a greater influence on the choice of a and b than other ages.

Other relevant comments also received marks.

So weights $w_x \ \alpha \ E_x$

Suitable choice is $w_x = \left[\text{Var}\left(\hat{q}_x\right)\right]^{-1}$

$$= \frac{E_x}{q_x(1 - q_x)}$$

$$\simeq \frac{E_x}{q_x} \ \text{ as } q_x \simeq 10^{-2}$$

These weights can be estimated by $\dfrac{E_x}{\hat{q}_x} = \dfrac{E_x^2}{\theta_x}$

(iii) The graduated rates $\mathring{q}_x$ are a linear function of the rates in the standard table q_x^s. The standard table rates will already be smooth.

Smoothness is based on the size of the third differences of the graduated rates $\Delta^3 \mathring{q}_x$, which because the relationship is linear will be equal to $b\Delta^3 q_x^s$. $\Delta^3 q_x^s$ will already be acceptably small because the standard table rates will already be smooth.

5 Expected Present Value of Benefits

$$5{,}000 \frac{D_{65}}{D_{45}} \left\{ \bar{a}_{\overline{5}|4\%} + \frac{D'_{70}}{D'_{65}} \bar{a}'_{70} \right\} = 19{,}910.340$$

where $'$ = a(55)

assuming weekly payments are a good approximation to continuous payment.

Marks for evaluations

$$\bar{a}_{\overline{5}|4\%} = \frac{1-(1.04)^{-5}}{\log_e(1.04)} = 4.54028$$

$$\bar{a}'_{70} \simeq a_{70} + ½ = 8.463$$

$$\frac{D_{65}}{D_{45}} = \frac{2{,}144.1713}{5{,}689.1776} = 0.37689$$

$$\frac{D'_{70}}{D'_{65}} = \frac{43{,}852}{61{,}593} = 0.71196$$

$$10{,}000 \bar{A}_{45:\overline{10}|} \simeq 10{,}000\,(1+½i)\,\frac{(M_{45}-M_{55})}{D_{45}} = 371.4054$$

$$25{,}000 \frac{D_{55}}{D_{45}} \bar{A}_{55:\overline{10}|} \simeq 25{,}000\,(1+½i)\,\frac{(M_{55}-M_{65})}{D_{45}} = 1{,}732.3690$$

Expected Present Value of Premiums. Annual Premium P.

$$P\ddot{a}_{45:\overline{20}|} = 13.488P$$

Equation of Value

$$13.488P = 19{,}910.340 + 371.4054 + 1{,}732.3690 = 22{,}014.114$$

$$P = £1{,}632.126$$

6 (i) x = CY of death – CY of birth
= age, x, on birthday in CY of death
= age next, x, on 1 January in CY of death
= age next, x, on 1 January before date of death

So Calendar Year Rate Interval starting, for lives classified x, on 1 January on which the life is aged x next birthday.

Age range at start of calendar year $x - 1$ to x.

(ii) (a) $P_x(t)$ census at t of those x next on previous 1 January would correspond to the classification of deaths

but ages in the censuses used are ages on 1 July

So $(x - 1, x)$ on 1 January
is $(x - ½, x + ½)$ on 1 July = date of census

So required x in $P_x(½)$, $P_x(1½)$, $P_x(2½)$ is x nearest birthday at date of census

(b) Need Birthdays uniform over calendar year
to get average age at start of rate interval, $x - ½$

Need force constant over $(x - ½, x + ½)$
So $\hat{\mu}_{x+f}$ will be $x + 0$, $f = 0$

7 (i) H_0: The observed rates $\mathring{q}_x$ are consistent with coming from a population in which the premium rate basis q_x^p are the true rates.

(ii) E_x initial exposed to risk at age x in the investigation

Chi-squared

(a) The test statistic is

$$\sum_x \frac{\left(E_x\mathring{q}_x - E_x q_x^p\right)^2}{E_x q_x^p}$$

where E_x is the initial exposed to risk at age x.

(b) If H_0 is true, then

$$\frac{E_x\mathring{q}_x - E_x q_x^p}{\sqrt{E_x q_x^p}} \underset{\text{approx.}}{\sim} N(0,1)$$

for each age x using the Central Limit Theorem and assuming that $1 - q_x^p \approx 1$.

Then if the observations at each of the n ages are independent

$$\sum_x \frac{(E_x\mathring{q}_x - E_x q_x^p)^2}{E_x q_x^p} \underset{\text{approx.}}{\sim} \chi_n^2$$

(c) Reject H_0 if the Observed Value of the test statistic > χ_n^2 $(1 - \alpha)$ at α% level of significance. This is a one-tailed test.

Grouping of Signs

(a) The test statistic is G, the observed number of positive groups of signs among n_1 positive and n_2 negative deviations, where the deviation at age $x = E_x\ \mathring{q}_x - E_x q_x^p$.

(b) If H_0 is true, then the deviations will be allocated randomly to groups.

Then n_1 positive deviations can be divided into g groups in

$$\binom{n_1 - 1}{g - 1} \text{ ways}$$

These g groups of positive signs can be allocated amongst the n_2 negative deviations in

$$\binom{n_2+1}{g} \text{ ways}$$

The unrestricted number of ways of arranging $n_1 + n_2$ positive and negative deviations is

$$\binom{n_1+n_2}{n_1}$$

So the sampling distribution of the test statistic, G, is

$$P[G=g] = \frac{\binom{n_2+1}{g}\binom{n_1-1}{g-1}}{\binom{n_1+n_2}{n_1}} \quad g = 1, 2, 3, \ldots n_2 - 1$$

Alternatively, if $n_1 + n_2 > 20$ (approx.) then

$$G \underset{\text{approx.}}{\sim} N\left(\frac{n_1(n_2+1)}{(n_1+n_2)}, \frac{(n_1 n_2)^2}{(n_1+n_2)^3}\right)$$

using the Central Limit Theorem and E[G] and Var[G].

(c) Reject H_0 if the Observed Value of the test statistic, $G \le k^*$ where k^* is smallest integer such that

$$P[G \le k^*] \ge 0.05$$

(or as similar statement based on the $N(0, 1)$ distribution). This is a one-tailed test.

8 (i) $W = a_{\overline{\min(K_{40},20)}|} \quad K_{40} \geq 0$

or $W = a_{\overline{K_{40}}|} \quad K_{40} = 0, 1, 2, \ldots 20$

$= a_{\overline{20}|} \quad K_{40} > 20$

(ii) (a) $E[W] = a_{40:\overline{20}|}$ A67/70 ultimate 5%

$$= \ddot{a}_{40:\overline{20}|} + \frac{\ell_{60} v^{20}}{\ell_{40}} - 1$$

$$= 12.760 + \frac{30{,}039.787\,(1.05)^{-20}}{33{,}542.311} - 1$$

$$= 12.0975$$

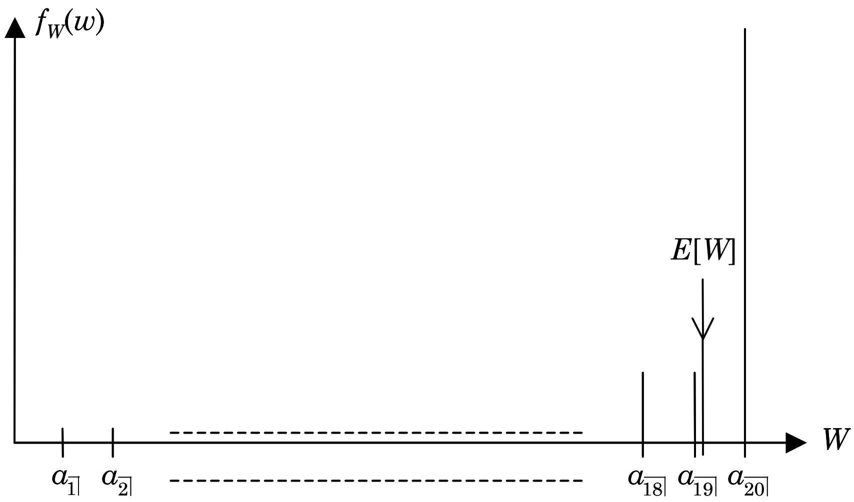
5% tables give: $a_{\overline{19}|} = 12.0853$ $a_{\overline{20}|} = 12.4622$

So $W < 12.0975$ if life dies in (40, 60)

$$= 1 - \frac{\ell_{60}}{\ell_{40}} = 1 - \frac{30{,}039.787}{33{,}542.311}$$

$$= 1 - 0.89558 = 0.10442$$

(b) The probability function of the discrete random variable W is very skew so $P[W < E[W]]$ = the very small "left hand tail area".



(iii) Now $W = \begin{cases} \ddot{a}_{\overline{K_{40}+1}|} - 1 & K_{40} = 0, 1, 2, \ldots 20 \\ \ddot{a}_{\overline{21}|} - 1 & K_{40} > 20 \end{cases}$

$$= \begin{cases} \dfrac{1 - v^{K_{40}+1}}{d} - 1 & K_{40} = 0, 1, 2, \ldots 19, 20 \\ \dfrac{1 - v^{21}}{d} - 1 & K_{40} > 20 \end{cases}$$

$$= \begin{cases} \dfrac{1}{d} - 1 - \dfrac{1}{d} v^{K_{40}+1} & K_{40} = 0, 1, 2, \ldots 19, 20 \\ \dfrac{1}{d} - 1 - \dfrac{1}{d} v^{21} & K_{40} > 20 \end{cases}$$

So $\text{Var}(W) = \dfrac{1}{d^2}$ $\text{Var}(Y)$ where

$$Y = \begin{cases} v^{K_{40}+1} & K_{40} = 0, 1, 2, \ldots, 19, 20 \\ v^{21} & K_{40} > 20 \end{cases}$$

Now $E[Y] = \sum_{k=0}^{20} v^{k+1} {}_{k|}q_{40} + \sum_{k=21}^{k=\infty} v^{21} {}_{k|}q_{40}$

$= A_{40:\overline{21}|}$

Where A is determined at 5% p.a.

$E[Y^2] = \sum_{k=0}^{20} (v^{k+1})^2 {}_{k|}q_{40} + \sum_{k=21}^{k=\infty} (v^{21})^2 {}_{k|}q_{40}$

$= A'_{40:\overline{21}|}$

Where A' is determined at $(1.05)^2 - 1 = 10.25\%$ p.a.

So $\text{Var}(W) = \left(\dfrac{1.05}{0.05}\right)^2 \left\{A'_{40:\overline{21}|} - A^2_{40:21}\right\}$ where $d = \dfrac{0.05}{1.05}$

$= 441\left(A'_{40:\overline{21}|} - A^2_{40:\overline{21}|}\right)$

The result can be derived in a similar way using

$a_{\overline{\min(K_{40},20)}|} = \ddot{a}_{\overline{\min(K_{40}+1,\ 21)}|} - 1$

9 (i) (a) ${}_{b-a}q_{x+a} = 1 - {}_{b-a}p_{x+a}$

Now ${}_{b}p_x = {}_{a}p_x \; {}_{b-a}p_{x+a}$

So ${}_{b-a}q_{x+a} = 1 - \dfrac{{}_{b}p_x}{{}_{a}p_x}$

(b) If deaths are uniformly distributed over $(x, x+1)$ then

$${}_{t}q_x = t \,.\, q_x$$

So $\quad {}_{t}p_x = 1 - t \,.\, q_x$

Then $\quad {}_{b-a}q_{x+a} = 1 - \dfrac{1 - bq_x}{1 - aq_x} = \dfrac{(b-a)q_x}{1 - aq_x}$

(ii) (a) The likelihood for each life is

Life i	1	2	3	4	5	6
Likelihood	p_{70}	${}_{0.6}p_{70.3}$	${}_{0.5}q_{70.5}$	${}_{0.4}p_{70}$	${}_{0.9}q_{70}$	q_{70}

And the total likelihood is the product

$$(1 - q_{70})\,(1 - {}_{0.6}q_{70.3})\;{}_{0.5}q_{70.5}\;(1 - {}_{0.4}q_{70})\;{}_{0.9}q_{70}q_{70}$$

Using (i)(b) we can write

$$(1 - q_{70})\left(1 - \frac{0.6q_{70}}{1 - 0.3q_{70}}\right)\left(\frac{0.5q_{70}}{1 - 0.5q_{70}}\right)(1 - 0.4q_{70})\;0.9q_{70}q_{70}$$

$$= \frac{(1 - q_{70})(1 - 0.9q_{70})\;\;(1 - 0.4q_{70})\;0.45q_{70}^3}{(1 - 0.3q_{70})(1 - 0.5q_{70})}$$

(b) The likelihood for each life is proportional to (assuming constant force $\bar{\mu}_{70}$).

Life i	1	2	3	4	5	6
Likelihood	$e^{-\bar{\mu}_{70}}$	$e^{-0.6\bar{\mu}_{70}}$	$e^{-0.4\bar{\mu}_{70}}\bar{\mu}_{70}$	$e^{-0.4\bar{\mu}_{70}}$	$e^{-0.7\bar{\mu}_{70}}\bar{\mu}_{70}$	$e^{-0.8\bar{\mu}_{70}}\bar{\mu}_{70}$

And the total likelihood is the product

$$L \propto e^{-3.9\bar{\mu}_{70}}(\bar{\mu}_{70})^3$$

Then $\dfrac{\partial L}{\partial \bar{\mu}_{70}} = -\,3.9e^{-3.9\mu_{70}}(\bar{\mu}_{70})^3 + e^{-3.9\bar{\mu}_{70}}\;3\bar{\mu}_{70}^2$

Then $-3.9\hat{\hat{\mu}}_{70} + 3 = 0$

$$\hat{\hat{\mu}}_{70} = \frac{3}{3.9} = 0.7692$$

The same result can be obtained using the log likelihood.

Then ML estimator of q_{70} ; $\hat{q}_{70} = 1 - e^{-0.7692} = 0.5366$

10 (i) Write times in rank order, label groups and label first absences from work.

Determine cumulative probabilities and contributions to partial likelihood from first absences.

2+	F	e^{β}	$7 + 8e^{\beta}$	
4	F	e^{β}	$7 + 7e^{\beta}$	$\dfrac{e^{\beta}}{7+7e^{\beta}}$
6+	M	1	$7 + 6e^{\beta}$	
7	F	e^{β}	$6 + 6e^{\beta}$	$\dfrac{e^{\beta}}{6+6e^{\beta}}$
8+	F	e^{β}	$6 + 5e^{\beta}$	
10+	F	e^{β}	$6 + 4e^{\beta}$	
11	M	1	$6 + 3e^{\beta}$	$\dfrac{1}{6+3e^{\beta}}$
12+	F	e^{β}	$5 + 3e^{\beta}$	
13+	M	1	$5 + 2e^{\beta}$	
15	M	1	$4 + 2e^{\beta}$	$\dfrac{1}{4+2e^{\beta}}$
16+	M	1	$3 + 2e^{\beta}$	
17	F	e^{β}	$2 + 2e^{\beta}$	$\dfrac{e^{\beta}}{2+2e^{\beta}}$
19+	M	1	$2 + e^{\beta}$	
20	M	1	$1 + e^{\beta}$	$\dfrac{1}{1+e^{\beta}}$
21+	F	e^{β}	e^{β}	

Partial Likelihood, L

$$\frac{e^{\beta}}{7(1+e^{\beta})}\cdot\frac{e^{\beta}}{6(1+e^{\beta})}\cdot\frac{1}{3(2+e^{\beta})}\cdot\frac{1}{2(2+e^{\beta})}\cdot\frac{e^{\beta}}{2(1+e^{\beta})}\cdot\frac{1}{1+e^{\beta}}$$

$$= \frac{e^{3\beta}}{84(1+e^{\beta})^4\ 6(2+e^{\beta})^2}$$

So $\text{Log}_e\, L = \ell(\beta) = 3\beta - 4\log_e(1+e^{\beta}) - 2\log_e(2+e^{\beta}) + c$

(ii) Then $\dfrac{\partial\ell}{\partial\beta} = 3 - \dfrac{4e^{\beta}}{1+e^{\beta}} - \dfrac{2e^{\beta}}{2+e^{\beta}}$

Let $x = e^{\beta}$ then $\hat{x}$ is given by

$$3 - \frac{4\hat{x}}{1+\hat{x}} - \frac{2\hat{x}}{2+\hat{x}} = 0$$

$$3(1+\hat{x})(2+\hat{x}) - 4\hat{x}(2+\hat{x}) - 2\hat{x}(1+\hat{x}) = 0$$

$$6 + 9\hat{x} + 3\hat{x}^2 - 8\hat{x} - 4\hat{x}^2 - 2\hat{x} - 2\hat{x}^2 = 0$$

$$3\hat{x}^2 + \hat{x} - 6 = 0$$

$$\hat{x} = \frac{-1 \pm \sqrt{1-4\,(-6.3)}}{6}$$

$$= \frac{-1 \pm \sqrt{73}}{6}$$

Estimate must be +ve, so $\dfrac{\sqrt{73}-1}{6} = 1.25733$

So $\hat{\beta} = \log_e 1.25733 = 0.2290$

(iii) Now $\dfrac{\partial^2\ell}{\partial\beta^2} = -\left(\dfrac{-4e^{\beta}\,.\,e^{\beta} + (1+e^{\beta})\,.\,4e^{\beta}}{(1+e^{\beta})^2}\right) - \left(\dfrac{-2e^{\beta}\,.\,e^{\beta} + (2+e^{\beta})\,.\,2e^{\beta}}{(2+e^{\beta})^2}\right)$

$$= -\left(\frac{4e^{\beta}}{(1+e^{\beta})^2} + \frac{4e^{\beta}}{(2+e^{\beta})^2}\right)$$

At $e^{\hat{\beta}} = 1.25733$, this has value $-(0.98700 + 0.47401) = -1.46101$

So asymptotic standard error $= +\dfrac{1}{\sqrt{1.46101}}$

$= 0.8273$

(iv) 95% Confidence Interval for $\hat{\beta}$ is approximately

$0.2290 \pm 2 \times 0.8273$

i.e. $(-1.43,\ 1.88)$

This interval includes $\beta = 0$, so the model

$$\lambda(t \mid x) = \lambda_0(t)\exp\{0 \times x\} = \lambda_0(t)$$

is compatible with the observed data i.e. same model applies to males and females. There is no significant difference between rates.

OR

Using a one sided alternative hypothesis and testing

H_0: $\beta = 0$ against H_1: $\beta > 0$ we have:

Under the Null hypothesis

$$\frac{0.2290 - 0}{0.8273} = 0.28$$

should be $N(0, 1)$.

This value is not in the critical region $(1.64, \infty)$ and so there is no reason to reject the null hypothesis. Conclusions as above.

OR

A likelihood test can be used

$$-2(\ell(0) - \ell(\hat{\beta})) = -2(-4.9698 + 4.9315) = 0.0765$$

which is χ_1^2 if H_0: $\beta=0$ is true. This value is not in the critical region so there is no reason to reject H_0.