INSTITUTE AND FACULTY OF ACTUARIES

Curriculum 2019

SPECIMEN SOLUTIONS

Subject CS2A – Risk Modelling and Survival Analysis

© Institute and Faculty of Actuaries
Sample path A – continuous time, discrete state process
Sample path B – discrete time, continuous state process
Sample path C – discrete time, discrete state process
Sample path D – continuous time, continuous state process

2 In a proportional hazards model the hazard of experiencing an event may be factorised into two components:
   one depending only on duration since some start event,
   which is known as the baseline hazard,
   and the other depending only on a set of covariates and associated parameters.

   The hazards for any two lives are therefore in the same proportion at all durations.

3 Advantages

   The method takes into account the evolution of age patterns of mortality over time (it is a two-factor model)
   It has been programmed in several statistical packages.
   It is stochastic, which allows users to assess the degree of uncertainty in parameter estimates, and the extent of random error in the mortality forecasts.
   It is simple to understand.
   It can easily be extended and adapted in various ways to suit particular contexts, for example by smoothing the age patterns of mortality using penalised regression splines.

Disadvantages

   It is calibrated using past data, which may reflect particular period events in the recent past. Unless care is used, these events could unduly influence mortality forecasts.
   As mortality rates for each age are forecast independently, it can produce Implausible results, such as mortality rates which are forecast to decrease with age.
   Unless observed rates are used for the forecasting, it can produce “jump-off” effects.
It assumes that the ratio of the rates of mortality change at different ages remains constant over time, when there is empirical evidence that this is not so. 

4

(i) (a) To protect itself from the risk of large claims. +1

(b) Excess of loss reinsurance where the reinsurer pays any amount of a claim above the retention. +1

Proportional reinsurance where the reinsurer pays a fixed proportion of any claim. +1

(ii) We must first find the parameters $\alpha$ and $\lambda$ of the Pareto distribution.

\[
\frac{\lambda}{\alpha - 1} = 270 \quad \text{and} \quad \frac{\alpha \lambda^2}{(\alpha - 1)^2 (\alpha - 2)} = 340^2
\]

\[
\frac{\alpha}{\alpha - 2} \times \frac{\lambda^2}{(\alpha - 1)^2} = 340^2
\]

so \[
\frac{\alpha}{\alpha - 2} = \frac{340^2}{270^2} = 1.585733882
\]

so \[
\alpha = \frac{2 \times 1.585733882}{1.585733882 - 1}
\]

= 5.4145

and \[
\lambda = 270 \times 4.4145 = 1191.920375
\]

We need to find $M$ such that $P(X > M) = 0.05$

i.e. \[
\left( \frac{\lambda}{\lambda + M} \right)^\alpha = 0.05
\]

\[
\frac{\lambda}{\lambda + M} = 0.05 \frac{1}{\alpha}
\]

\[
\lambda = 0.05^\alpha (\lambda + M)
\]
\[
M = \frac{\lambda \left[1 - 0.05^\alpha \right]}{0.05^\alpha}
\]

\[
1191.920375 \times \left[1 - 0.05^{5.4145} \right]
\]

\[
= \frac{1}{0.05^{5.4145}}
\]

\[= 880.8 \quad \text{[4]}\]

[Total 7]

5 (i) The conventional approach in machine learning is to divide the data into two. One part of the data (usually the majority) is used to train the algorithm to choose the ‘best’ hypothesis. The other is used to test the chosen hypothesis on data that the algorithm has not seen before.

In practice, the ‘training’ data is often split into a part used to estimate the parameters of the model, and a part used to validate the model.

It involves three data sets:

a **training data set**: the sample of data used to fit the model

a **validation data set**: the sample of data used to provide an unbiased evaluation of model fit on the training dataset while tuning model hyper-parameters

a **test data set**: the sample of data used to provide an unbiased evaluation of the final model fit on the training data set.

Typically, the training data set will use around 60% of the data available, the validation data set about 20% and the test data set about 20%

but these percentages may vary, and a key criterion in dividing the data is that the validation and test data sets should have enough data to be fit for purpose.

(ii) The number of covariates which may be used to model risk prices could be vary large.

Using conventional approaches to finding the ‘best’ model (the ‘best’ combination of covariates) can be very time consuming and computationally infeasible.
Penalised regression constrains the sum of the estimated parameter values to be less than a certain value. We exact a penalty which may be written $\lambda P(\beta_1, \ldots, \beta_j)\), where $\beta_1, \ldots, \beta_j$ are the parameters $P()$ is some increasing function.  

$\frac{1}{2}$

Need sufficient data to find best choice of $\lambda$ (done e.g. through cross-validation).  

$\frac{1}{2}$

Advantages for risk pricing:

- model building is automated
- no debates about borderline decisions
- quicker speed to market
- models can be continuously updated as new data appear.

$+1$

Disadvantages:

- data sets can be very large.
- throwing all your data at penalised regression does not work.
- data are not equally relevant (e.g. should we use three years’ data or just two years’ data?)

$+1$

*Other relevant points could score credit.*

$max. 4$

$[Total 8]$

$i)$ $\mu$ is the hazard for a man who does not drink beer.

$+1$

$\beta$ measures the impact on the hazard of a one glass increase in the daily amount of beer drunk.

$+1$

$ii)$ The hazard for a man who drinks two glasses of beer a day is 

$$0.03 \exp(0.2 \times 2) = 0.0448.$$ 

$+1$

$iii)$ (a) The probability that a man aged 60 years who drinks three glasses of beer a day will survive to his 70th birthday is 

$$\exp \left( - \int_0^{10} 0.03 \exp(0.2 \times 3) dt \right) = \exp(-0.547) = 0.579.1$$ 

$+1$

(b) Because the hazard of death is constant, the expectation of life at age 60 years is given by 

$$\frac{1}{h(t)} = \frac{1}{\mu \exp(\beta x)} = \frac{1}{0.03 \exp(0.2 \times 3)} = 18.3 \text{ years.}$$ 

$+1$
(iv) The owner’s total revenue $R$ is proportional to the average number of glasses of beer drunk per day multiplied by the man's expectation of life:

$$R \propto \frac{x}{0.03 \exp(0.2x)}.$$  

We maximise $R$ with respect to $x$.

$$\frac{dR}{dx} = \frac{0.03 \exp(0.2x) - x(0.03 \times 0.2) \exp(0.2x)}{[0.03 \exp(0.2x)]^2} = \frac{1 - 0.2x}{0.03 \exp(0.2x)}. $$  

This is zero when $1 - 0.2x = 0$, or when $x = 5$.

THEN EITHER

$$\frac{d^2R}{dx^2} = \frac{-0.2[0.03 \exp(0.2x)] - (1 - 0.2x)[0.03 \times 0.2 \exp(0.2x)]}{[0.03 \exp(0.2x)]^2}$$  

$$= \frac{-0.2 - 0.2(1 - 0.2x)}{[0.03 \exp(0.2x)]} = \frac{0.2(0.2x - 2)}{[0.03 \exp(0.2x)]}$$

which is negative when $x = 5$,

so we have a maximum.  

OR

When $x$ is 0, total revenue is 0.

When $x$ is strictly positive, total revenue is always greater than zero.

Therefore we must have a maximum.

So the owner should sell the man five glasses of beer per day.

Various alternatives can be awarded credit, e.g. using the logarithm of the likelihood, or a demonstration by calculating predicted revenue for $x = 0, 1, 2, 3...$ that $x = 5$ marks a maximum.

[Total 9]
(i) ALTERNATIVE 1

If the probabilities do not depend solely upon the length of the time interval $t-s$, the process is time-inhomogeneous. +1

OR ALTERNATIVE 2

The transition rates vary with time. +1

OR ALTERNATIVE 3

The transition rates depend upon the start and end times. +1

(ii) A model with time-inhomogeneous rates has more parameters, and there may not be sufficient data available to estimate these parameters. +1

Also, the solutions to Kolmogorov’s equations may not be easy (or even possible) to find analytically. +1

Time-inhomogeneous processes are computationally harder to simulate. +1

[max. 2]

(iii) We need $P_{GG}(t)$, i.e. probability the process remains in $G$ throughout time $0$ to $t$.

This satisfies

$$
\frac{d}{dt} P_{GG}(t) = (-0.2 - 0.04t) P_{GG}(t).
$$

Hence

$$
\frac{1}{P_{GG}(t)} \frac{d}{dt} P_{GG}(t) = (-0.2 - 0.04t);
$$

$$
\frac{d}{dt} [\ln P_{GG}(t)] = (-0.2 - 0.04t).
$$

Integrate both sides:

$$
\ln P_{GG}(t) \bigg|_{t=0}^{t=s} = -0.2s - 0.02s^2 \bigg|_{s=0}^{t}.
$$

$$
P_{GG}(0) = 1,
$$
so \( P_{GG}(t) = \exp(-0.2t - 0.02t^2) \).\[3\]

(iv) Occurs when \( P_{GG}(t) = 0.5 \), so we have

\[
0.5 = \exp(-0.2t - 0.02t^2),
\]

\[
0.02t^2 + 0.2t - 0.69315 = 0,
\]

and solving using the quadratic equation formula produces

\[
t = 2.724 \text{ or } t = -12.724.
\]

The answer lies between 0 and 8, so we require \( t = 2.724 \).\[2\]

(v) \[
\frac{d}{dt} P_G(t) = (-0.2 - 0.04t) P_G(t) + (0.4 - 0.04t) P_N(t)
\]

But \( P_N(t) = 1 - P_G(t) \),

So \[
\frac{d}{dt} P_G(t) = (-0.2 - 0.04t) P_G(t) + (0.4 - 0.04t) (1 - P_G(t)),
\]

OR

\[
\frac{d}{dt} P_G(t) = 0.4 - 0.04t - 0.6P_G(t).
\]

8 (i) We can deduce that \( Y_t = at + \sum_{i=1}^{t} e_i \)

and so \( E(Y_t) = at \) and \( Var(Y_t) = t\sigma^2 \).

Since these expressions depend on \( t \) the process is not stationary.\[3\]

(ii) As \( s < t \) we have

\[
Cov(Y_t, Y_{t-s}) = Cov(at + \sum_{i=1}^{t} e_i, as + \sum_{j=1}^{t-s} e_j) = Var(\sum_{j=1}^{t-s} e_j) = (t-s)\sigma^2
\]

+1½
(i) The force of mortality, $\mu_{x+t}$ at age $x + t$ is defined by the expression

$$\mu_{x+t} = \lim_{dt \to 0} \frac{1}{dt} \Pr[T \leq x + t + dt | T > x + t]$$

(ii) $5 p_0 = e^{-5\mu}$

(iii) $5 p_5 = e^{-5\lambda}$

$15 p_5 = e^{-15\lambda}$

But $5 p_5 = 2_{15} p_5$

Hence $e^{-5\lambda} = 2 e^{-15\lambda}$

$-5\lambda = \log_e 2 - 15\lambda = 0.693 - 15\lambda$  

so that $\lambda = 0.0693$.

(iii) First, note that the differenced series:

$$X_t = Y_t - Y_{t-1} = a + e_t$$

is essentially a white noise process.

So estimates of $a$ and $\sigma^2$ can be found by constructing the sample differences series $x_i = y_i - y_{i-1}$ for $i = 1, 2, \ldots, n$ and taking the mean and sample variance (or its square for estimating $\sigma$) respectively.

(iv) In this case $\hat{y}_n(1) = \hat{a} + y_n + 0 = \hat{a} + y_n$

and $\hat{y}_n(2) = \hat{a} + \hat{y}_n(1) + 0 = 2\hat{a} + y_n$
(iv) If $\lambda = \mu = 0.0693$, then $e_0 = \frac{1}{\mu} = \frac{1}{0.0693} = 14.43$ years.

(v) If $\lambda \neq \mu$, then

$$e_0 = \int_0^\infty e^{-\mu t} dt + e^{-5\mu} \int_0^\infty e^{-\lambda s} ds.$$ 

Evaluating the integrals gives

$$e_0 = \left[ -\frac{1}{\mu} e^{-\mu x} \right]_0^5 + e^{-5\mu} \left[ -\frac{1}{\lambda} e^{-\lambda x} \right]_0^\infty$$

$$= \frac{1}{\mu} e^{-5\mu} + \frac{1}{\mu} \left( 1 - e^{-5\mu} \right)$$

and putting $\lambda = 0.0693$ gives

$$e_0 = \frac{1}{0.0693} e^{-5\mu} + \frac{1}{\mu} \left( 1 - e^{-5\mu} \right)$$

$$= 14.43 e^{-5\mu} + \frac{1}{\mu} \left( 1 - e^{-5\mu} \right).$$

OR

If $\lambda \neq \mu$, then

$$e_0 = \int_0^\infty t p_\lambda \mu dt.$$ 

$$e_0 = \int_0^5 t e^{-\mu t} dt + e^{-5\mu} \int_5^\infty \int_5^\infty t e^{-\lambda t} dt$$

Integrating by parts

$$= \int_0^5 \left( -\frac{\partial}{\partial t} e^{-\mu t} \right) dt + e^{-5(\mu-\lambda)} \int_5^\infty \left( -\frac{\partial}{\partial t} e^{-\lambda t} \right) dt$$

$$= \left[ -te^{-\mu} \right]_0^5 + \int_0^5 e^{-\mu t} dt + e^{-5(\mu-\lambda)} \left( \left[ -te^{-\lambda t} \right]_5^\infty + \int_5^\infty e^{-\lambda t} dt \right)$$
(i) The lag polynomial here is \(1 - 0.6L - 0.16L^2 = (1 - 0.8L)(1 + 0.2L)\) with roots 1.25 and -5 therefore it is stationary.

Hence an ARMA(2,0) process.

(ii) From the stationarity condition then

\[
E(Y_t) = \mu = 1 + 0.6\mu + 0.16\mu + 0
\]

\[
\mu = \frac{1}{1 - 0.76} = \frac{1}{0.24} = 5/7
\]

(iii) From the Yule-Walker equations for autocorrelation function values and for lags 1, 2, 3, … we have that \(\rho_k = 0.6\rho_{k-1} + 0.16\rho_{k-2}\).

In particular, for \(k = 1\),

\[
\rho_1 = 0.6\rho_0 + 0.16\rho_1 = 0.6 + 0.16\rho_1
\]

or

\[
\rho_1 = \frac{0.6}{1 - 0.16} = \frac{0.6}{0.84} = 0.7143 = 5/7
\]

For \(k = 2\) we have that

\[
\rho_2 = 0.6\rho_1 + 0.16\rho_0 = 0.16\rho_1 + 0.6 = \frac{0.6^2}{0.84 + 0.16} = 0.5886 = 103/175.
\]
\[ \rho_3 = 0.6\rho_2 + 0.16\rho_1 = 0.6 \times 0.5885714 + 0.16 \times 0.7142857 \\
= 0.4674 = \frac{409}{875} \]

For \( k = 4 \)

\[ \rho_4 = 0.6\rho_3 + 0.16\rho_2 = 0.6 \times 0.4674286 + 0.16 \times 0.5884714 \\
= 0.3746 = \frac{1639}{4375} \]

For the partial autocorrelation function we have that

\[ \psi_1 = \rho_1 = 0.7143 = \frac{5}{7} \]

\[ \psi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = 0.16 = \frac{4}{25} \]

and \( \psi_3 = \psi_4 = 0 \) since \( Y_t \) is AR(2).

11 (i) **Formula for the Gumbel copula function**

The inverse of \( x = \psi(y) = (-\ln y)^{\alpha} \) is \( y = \psi^{-1}(x) = \exp\left(-x^{1/\alpha}\right) \).

So the copula function is:

\[ C(u, v) = \psi^{-1}\left(\psi(u) + \psi(v)\right) = \exp\left\{-\left((-\ln u)^{\alpha} + (-\ln v)^{\alpha}\right)^{1/\alpha}\right\} \]

(ii) **Lower tail dependence**

Using the formula given and the copula function from part (i):
\[
\lambda_L = \lim_{{u \to 0^+}} \frac{C(u, u)}{u} = \lim_{{u \to 0^+}} \frac{1}{u} \exp \left\{ -\left( (-\ln u)^\alpha + (-\ln u)^{1/\alpha} \right)^{1/\alpha} \right\}
\]
\[
= \lim_{{u \to 0^+}} \frac{1}{u} \exp \left\{ \left( 2(-\ln u)^\alpha \right)^{1/\alpha} \right\}
\]
\[
= \lim_{{u \to 0^+}} \frac{1}{u} \exp \left\{ -2^{\frac{1}{\alpha}} (-\ln u) \right\}
\]
\[
= \lim_{{u \to 0^+}} \frac{1}{u} \exp \left\{ \ln u \times 2^{\frac{1}{\alpha}} \right\}
\]
\[
= \lim_{{u \to 0^+}} \frac{1}{u} u^{\frac{1}{\alpha}} - 1
\]
\[
= \lim_{{u \to 0^+}} u^{\frac{1}{\alpha}} - 1 + 1^{\frac{1}{2}}
\]

Since \(1 < \alpha < \infty\), we know that \(0 < 1/\alpha < 1\) and \(0 < 2^{\frac{1}{\alpha}} - 1 < 1\).
So \(\lambda_L = 0\).

(iii) **Extreme values**

(a) In this context the extreme values are the claims in the right-hand tail of the claims distribution. +1/2

These claims have a low probability but are for large amounts. +1/2

(b) Because the payments for these claims are large, they can have a significant financial impact on the insurance company. +1/2

The probabilities for the right-hand tail cannot be calculated accurately using the usual techniques used for the main body of the distribution. +1/2

(iv) **Probability that the individual claims exceed £10 million**

Using the distribution functions give:

\[
P(X > 10) = 1 - F_X (10) = 1 - \exp \left\{ - \exp \left[ - \left( \frac{10 - 2.5}{5} \right) \right] \right\}
\]
\[
= 1 - e^{-e^{-1.5}} = 1 - 0.8000 = 0.2000 + 1
\]
\[ P(Y > 10) = 1 - F_Y(x) = 1 - \exp \left\{ - \exp \left[ - \frac{10 - 5}{7.5} \right] \right\} \]
\[ = 1 - e^{-0.6667} = 1 - 0.5984 = 0.4016 \quad +1 \]

(v) **Probability that both claims exceed £10 million**

(a) **… using the product copula**

The product copula implies that \( X \) and \( Y \) are independent.  +½

So with this copula:

\[ P(X > 10, Y > 10) = P(X > 10) \times P(Y > 10) = 0.2000 \times 0.4016 = 0.0803 \quad +½ \]

(b) **… using the Gumbel copula**

With the Gumbel copula we have:  \( u = P(X \leq 10) = 0.8000 \) and  \( v = P(Y \leq 10) = 0.5984 \).

So

\[ P(X \leq 10, Y \leq 10) = \exp \left\{ - \left( -\ln 0.8000 \right)^2 + (-\ln 0.5984)^2 \right\}^{0.5} = 0.5713 \quad +1 \]

Using the hint given:

\[ P(X > 10, Y > 10) = 1 - P(X \leq 10) - P(Y \leq 10) + P(X \leq 10, Y \leq 10) \]
\[ = 1 - 0.8000 - 0.5984 + 0.5713 \]
\[ = 0.1729 \quad +1 \]

(vi) **Comment**

With the Gumbel copula the probability that both claims will exceed £10 million (17.29%) is more than twice what it would be if the claims were independent (8.03%).  +½

So, if the insurance company treated the claims from these two policies as independent, it would significantly underestimate the risk.  +½

[Total 12]
(i) We require each row of the transition matrix to sum to 1. Here this holds for all values of $\beta$. We require each of the following to lie between 0 and 1 inclusive:

$\beta, \beta^2, 2\beta, 1-3\beta, 1-\beta-\beta^2$.

The first two require that $0 \leq \beta \leq 1$. The third requires that $0 \leq \beta \leq 1/2$. The fourth that $1/3 \leq \beta \leq 1$. The fifth implies $\beta \leq \frac{\sqrt{5}-1}{2}$ as the negative root is not viable. So overall $0 \leq \beta \leq \frac{1}{3}$.

(ii) If $\beta > 0$ then it can reach any other state (so it is irreducible) and it has a loop on each state (so it is aperiodic). However if $\beta = 0$ it can never leave its current state so it is reducible.
(iv) In this case the matrix is \( P = \begin{pmatrix} 0.89 & 0.1 & 0.01 \\ 0.1 & 0.7 & 0.2 \\ 0.01 & 0.1 & 0.89 \end{pmatrix} \) +\( \frac{1}{2} \)

The stationary distribution satisfies \( \pi = \pi P \).

We have:

\[
\begin{align*}
\pi_0 &= 0.89\pi_0 + 0.1\pi_20 + 0.01\pi_40 \\
\pi_20 &= 0.1\pi_0 + 0.7\pi_20 + 0.1\pi_40 \\
\pi_40 &= 0.01\pi_0 + 0.2\pi_20 + 0.89\pi_40
\end{align*}
\]

and

\[
\pi_0 + \pi_20 + \pi_40 = 1.
\]

The first and third equations give

\[
\begin{align*}
0.11\pi_40 - 0.22\pi_0 &= 0.01\pi_0 - 0.02\pi_40 \\
\pi_40 &= \frac{0.23}{0.13} \pi_0 = 1.769\pi_0 \\
\pi_20 &= \frac{0.1 + 0.1769}{0.3} \pi_0 = 0.923\pi_0
\end{align*}
\]

so \((1 + 0.923 + 1.769)\pi_0 = 1\), and hence

\[
\begin{align*}
\pi_0 &= \frac{13}{48} \\
\pi_20 &= \frac{1}{4} \\
\pi_40 &= \frac{23}{48}
\end{align*}
\]

are the long term proportion of taxpayers at each marginal rate.

(v) Looking for the rates two years’ later, these are given by \( P^2 \), which is

\[
\begin{pmatrix} 0.89 & 0.1 & 0.01 \\ 0.1 & 0.7 & 0.2 \\ 0.01 & 0.1 & 0.89 \end{pmatrix} \times \begin{pmatrix} 0.89 & 0.1 & 0.01 \\ 0.1 & 0.7 & 0.2 \\ 0.01 & 0.1 & 0.89 \end{pmatrix} = \begin{pmatrix} 0.8022 & 0.16 & 0.0378 \\ 0.161 & 0.52 & 0.319 \\ 0.0278 & 0.16 & 0.8122 \end{pmatrix}
\]
So the required probabilities are:

(a) 0.161
(b) 0.52
(c) 0.319.

END OF MARKING SCHEDULE