Statistical and Risk Modelling
Assignment-1

Tushar Khandelwal
39

1) $uqx = \Pr[T_x \leq u = \Pr[K_x = 0 \; and \; S_x \leq u] = \Pr[K_x = 0] * \Pr[S_x \leq u]]$

   $since \; K_x \; and \; S_x \; are \; independent.$

   $\Pr[S_x \leq u] = \int_0^u 1dx = u, since \; uniform \; distribution.$

   $Thus, uqx = u * q_x \; since \Pr[K_x = 0] = q_x$

2)

   a) Central exposed to risk

      Period of exposure is 1-6-2000 to 25-10-2000

      = 30 + 31 + 31 + 30 + 25 = 147 days

      =147/7 = 21 weeks

   b) Initial exposed to risk

      Period of exposure is 1-6-2000 to 31-5-2000 = 52 weeks

3)

   a) Left Censoring

      Data in this study would be left censored if the censoring mechanism prevent us

      from knowing when the policyholder joined the company.

      This is not present because the policy issue date is given.

   b) Right Censoring

      Data would be right censored if the censoring mechanism cuts short observations in

      progress, so that we are not able to discover if and when the policy is surrendered.

Data in this study would be right censored if the policy is terminated before the
maturity date for reasons than surrender.

c) Interval Censoring
Data in this study would be interval censored if the observational plan only allows us
to say that the duration of policy at the time of surrender fell within some interval of
time.
Here we know the calendar year of surrender and the policy issue date, so we will
know that the duration of the policy falls within one year rate interval. Interval
censoring is present.

d) Informative Censoring
Censoring in this study would be informative if the censoring event divided
individuals into two groups whose subsequent experience was thought to be
different.
Here the censoring event of surrendering the policy might be suspected to be
informative, as those who are likely to surrender the policy to be in better health
than those who do not surrender the policy.

4)

a) $Complete\ expectation\ of\ life, e_x$

$$e_x = E[T_x] = \int_0^{\omega - x} {}_t p_x \, dt$$

This represents the integral of the probability of survival at each future age, i.e., the expected

future lifetime of a life currently aged x. In other words, this is the expectation of life at age x or

how many years a life is expected to live given that it is currently x years old.

b) The curtate expectation of life

$$e_x = \sum_{k=1}^{\infty} kp_0 = \sum_{k=1}^{\infty} e^{-0.0325k} = \frac{e^{-0.0325}}{1 - e^{-0.0325}} = 30.27$$

c) The probability that a life aged exactly 36 will survive to age 45.

$$9p36 = \exp\left[-\int_0^9 0.0325dt\right] = e^{-0.2925} = 0.7464 \approx 75\%$$

d) The exact age x representing the median of the life-time T of a new born baby.

The median of the life-time T implies that the probability, xp0 = 0.5

$$Thus, xp_0 = 0.5 \Rightarrow \exp(x - 0.0325) = 0.5 \Rightarrow x = -\left(\frac{\log 0.5}{0.0325}\right) = 21.33$$

5)

i) Gompertz law is suitable for human mortality for middle to older ages ie. Between ages 35 to 90 years.

ii) We know that

$$tp_x = \exp\left(-\int_0^t \mu_{(x+s)}ds\right) = \exp\left(-\int_0^t Bc^{x+s}ds\right)$$

$We\ can\ write\ c^{x+s}\ as\ c^x e^{s*logc}, so\ that\ the\ integral\ becomes$:

$$\int_0^t B * c^x * e^{s*\log c}ds = \frac{Bc^x}{\log c}\left[e^{s*\log c}\right]_0^t = \frac{Bc^x}{\log c}[c^s]_0^t = \frac{Bc^x}{\log c}[c^t - 1]$$

$If\ we\ introduce\ the\ auxiliary\ parameter\ g\ defined\ by\ logg$

$$= -\frac{B}{\log c}, the\ value\ of\ the\ integral\ is - logg$$

$$* c^x(c^t - 1)\ and\ we\ find\ that:$$

$$tp_x = \exp\left[\log g * c^x(c^t - 1) = \left(e^{\log g}\right) * c^x * (c^t - 1) = gc^x(c^t - 1)\right.$$

6)

i) Female smoker aged 30 at entry.

ii) $\frac{h_j(t)}{h_i(t)} = \frac{\exp -0.05}{\exp 0.1} = 0.86070$

Where j is male smoker aged 30 at entry and i is female smoker aged 40 at entry.

$$But\ s(t) = \exp\left(-\int_0^t h(s)ds\right) hence$$

$$s_j(t) = \left(s_i(t)\right)^{0.86070}$$

Which implies that

$$s_j(t) > s_i(t)\ for\ all\ t > 0$$

iii) $\frac{h_j(t)}{h_i(t)} = \frac{\exp 0.2}{\exp 0.05} = 1.161$

Where j is male smoker aged 30 at entry and i is male smoker aged 40 at entry

$$But\ s(t) = \exp\left(-\int_0^t h(s)ds\right) hence$$

$$s_j(t) = \left(s_i(t)\right)^{1.161}$$

Which implies that

$$s_j(t) < s_i(t)\ for\ all\ t > 0$$

7)

i) The most appropriate rate interval to use (for lives classified x) is the policy year rate interval starting on the policy anniversary where lives are aged x next birthday.

The reason is that this corresponds to the definition of the deaths and the rate is more sensitive to errors in approximation of the numerator than the denominator.

The average age at the start of the rate interval is x – ½ assuming that birthdays are uniformly distributed over the policy year.

ii) We will use the following symbols:

$P_{x,t}$ to represent the in force at time t from the 1 January 1997 classified x next birthday on policy anniversary nearest to time t

$\theta_{x.t}$ to represent the deaths in the calendar year 1997 aged x next birthday on policy anniversary (= age next birthday at entry plus curtate duration at date of death) before death

E_x E_x^c to represent the initial and central exposed to risk respectively of lives age x last birthday on previous policy anniversary.

$P_x(t)$ to represent the in force at time t from the 1 January 1997 classified x next birthday on the policy anniversary preceding time t.

$Now\ P_x(t) = \frac{1}{2}\left(P_{x,t} + P_{x+1,t}\right)$ assuming that policy anniversaries are uniformly distributed over the calendar year.

$E_x^c = \int_0^{10} P_x(t)dt = \frac{1}{2}\sum_{t=0}^{9}\left(P_x(t) + P_x(t+1)\right)$ assuming that the in-force population varies linearly between the dates of the investigation.

$E_x = E_x^c + \frac{1}{2}\sum_{t=0}^{9}\theta_{x,t}$ assuming that in aggregate the deaths occur on average halfway through the policy year.

8)

i) Types of censoring presents:

- Type I censoring present because the study ends at a predetermined duration of 45 days.
- Type II censoring is not present because the study did not end after a predetermined number of patients had died.

- Random censoring is present because the duration at which a patient left hospital before the study ended can be considered as a random variable.
- Right Censoring is present for those lives that exit before the end of investigation period.

ii) The censoring is likely to be informative.

The patients who died were probably recovering less well that patient who discharged from the hospital.

If they had not died, they would likely to remain in the hospital for longer than those who were not censored.

iii) The Kaplan-Meier estimate of the survival function is estimated as follows:

T	n	d	c	d/n	(1-d/n)	S(t)
0	13					
5	13	1	0	0.0769	0.9231	0.92
7	12	1	0	0.0833	0.9167	0.85
14	11	1	2	0.0909	0.9091	0.77
28	8	1	2	0.1250	0.8750	0.67
35	5	1		0.2000	0.8000	0.54

So, the value survival function at end of investigation period is 0.54

Assumptions:

- The censoring happens just after the death.
- Ignoring the discharge on any other ground except recovery from illness.
- Ignore any admission period before the start of investigation.

iv) Comments:

- The survival of a patient from the infection who given treatment is around 50% in light of the answer in c) above.
- However, the hospital excluded the number of deaths who died within two weeks of observation period.
- It also ignores the admission pre investigation period
- It is assuming that the censored patient at the end of investigation will survive for sure.
- Also ignoring the patients being discharged on any other ground like shifting to another hospital etc.
- It claims that 8 out of 10 patients who responded the treatment beyond two weeks would survive.
- So, the claims have to be viewed with respect to above considerations.

9)

i)

a) Under the uniform distribution of deaths assumption:

$$\int_0^1 tPxdt = \int_0^1 (1 - tqx)dt = [t - 0.5t^2 q_x]_0^1$$

$Since\ q_x = 0.3, we\ have$

$$m_x = \frac{0.3}{1 - 0.15} = 0.352941$$

b) Under the constant force of mortality:

$$q_x = 1 - e^{-\mu}$$

$$\int_0^1 tPxdt = \int_0^1 e^{-\mu}dt = \frac{1}{\mu} * (1 - e^{-\mu}) = \frac{q_x}{\mu}$$

$$So, m_x = \mu = -\ln(1 - q_x) = -\ln 0.7 = 0.356675$$

10)

i) Under the Cox model each individual's hazard is proportional to the baseline hazard, with the constant of proportionality

depending on certain measurable quantities called co-variates. Hence the model is also called a proportional hazards model.

ii) (t) 0(t)exp (F* F M * M D * D), where (t) is the estimated hazard and 0 (t) is the baseline hazard.

iii) The baseline hazard refers to annual policy taken through the Online channel and where premiums are paid by direct debit.

iv) The results imply that

exp [(βD *1)]/ exp [(βD *1) + βF*1 + βM*1] = 0.75

exp (βF + βM) = 4/3

exp (βD*1) / exp [(βF *1)] = 1

exp (βM*1) / exp [(βD *2)] = 0.75

Substituting from (2) into (1) gives

exp (βD + βM) = 4/3

exp(βD) * exp(βM) = 4/3

From Eqn 3

(Exp(βD)) ^2*0.75 = exp(βM)

So

Substituting in Eqn 4

exp(βD) * (exp(βD))2*0.75 = 4/3

(Exp(βD)) ^3= 1.7778

exp(βD) = 1.2114

βD = 0.19179

βF = 0.19179

βM = 0.0959

11)

ii)

T	S(t)	Λ(t)	nt	dt	ct
0	1	0	12	0	

1	0.9167	0.0833	12	1	2
3	0.7130	0.22	9	2	2
6	0.4278	0.4	5	2	3

iii) Summing up the number of deaths we have total deaths = d1+d3+d6= 1+2+2= 5. Since we started with 12 insects, the remaining 7 insects' histories were right censored.

12)

i) Gompertz Law:

Gompertz Law is an exponential function, and it is often a reasonable assumption for middle and older ages. It can be expressed as follows:

$\lambda_x = Bc^x$; $where, \lambda_x$ $is\ a\ force\ of\ mortality\ at\ age\ x$

ii) Substituting, BB = exp (β0+β1X1 + β2X2) ; into the Gompertz model,

λx = exp (β0+β1X1 + β2X2). c^x; defining x as duration since 50th birthday.

The hazard can therefore be factorized into two parts:

exp (β0+β1X1 + β2X2), which depends only on the values of the covariates, and

$ccxx$, which depends only on duration.

So, the ration of between the hazards for any two persons with different characteristics does

not depend on duration, and so the model is a proportional hazards model.

iii) The baseline hazard in this model relates to a non-smoker female

iv) For a female cigarette smoker, we have

X1 = 0 and X2 = 1 and x = 4

Therefore, the hazard at age 54 is given by

λx = exp ($\beta\beta$0 + $\beta\beta$1. 0 + $\beta\beta$2. 1). c^4

= exp (-4+0.65) x 1.05^4

= 0.0351x1.2155

= 0.04266

v) The hazard for a non-smoker at duration, 's' is given by the formula

λs = exp(β0+β1X1). c^s

The hazard for a smoker at duration, 't' is given by the formula

λt = exp (β0+β1X1 + 0.65). c^t

If the smoker's and non-smoker's hazards are the same, then

$\lambda s = \lambda t$

i.e., exp(β0+β1X1). c^s = exp (β0+β1X1 + 0.65). c^t

i.e., c^s = exp (0.65). c^t

i.e., c^($s-t$) = exp (0.65) = 1.9155

Since, c = 1.05

Hence, 1.05^($s-t$) = 1.9155

So, s-t = ln (1.9155)/ln (1.05) = 0.65/0.04879

s-t = 13.32

Hence, when the two hazards are equal, the non-smoker is approximately 13 years older than the smoker.

13)

i) (Let P'x(t) be the number of policies in force aged x nearest birthday at time t.

Also, let Px(t) be the number of policies in force aged x last birthday at time t.

Let Ex^C refers to the central exposed to risk at age label x respectively.

$$E_x^C = \int_{t=0}^{2} P'x(t)dt$$

Assuming that P'56(t) is linear over the year (2015,2016) and (2016,2017), we can
approximate the exposure as follows

E56^c = ½*(P'56(2015) + P'56(2016)) + ½*(P'56(2016) + P'56(2017))

= ½*P'56(2015) + P'56(2016) +½*P'56(2017)

Since, the number of policyholders aged label 56 nearest birthday will be between 55.5 and 56.5 i.e., between age label 55 last birthday and 56 last birthdays. Assuming that the birthdays are uniformly distributed over the calendar year:

P'56(2015) = ½*(P55(2015) + P56(2015))

= 20050

Similarly,

P'56(2016) = ½*(P55(2016) + P56(2016))

= 20800

And,

P'56(2017) = ½*(P55(2017) + P56(2017))

= 19250

E56^c = ½*20050+20800+1/2*19250

= 40450

μ56 = d56/ E56^c

= 1380/40450

= 0.0341

Deriving the force of mortality for age 57 as above:

P'57(2015) = ½*(P56(2015) + P57(2015))

= 19850

Similarly,

P'57(2016) = ½*(P56(2016) + P57(2016))

= 20900

And,

P'57(2017) = ½*(P56(2017) + P57(2017))

= 17500

E57c = ½*19850+20900+1/2*17500

= 39575

μ57 = d57/ E57^c

= 1420/39575

= 0.03588

dx is deaths aged x nearest birthday on the date of death. So, the age label at death changes

with reference to life year. Therefore, the age at the middle of life year is x and estimates μx.

ii) We can estimate the initial rates of mortality using the estimated values of μ from part (i) and the following formula

q55.5 = 1- exp(-μ56)

= 0.0335

And

q56.5 = 1- exp(-μ57)

= 0.0352

14)

15)

i) Two advantages of central exposed to risk over initial exposed to risk are:

a) The central exposed to risk is simpler to calculate from the data typically available compared to the initial exposed to risk. Moreover, central exposed to risk has an intuitive appeal as the total observed waiting time and is easier to understand than the initial exposed to risk.

b) It is difficult to interpret initial exposed to risk in terms of the underlying process being modelled if the number of decrements under study increase or the situations become

more elaborate. On the contrary, the central exposed to risk is more versatile and it is easy to extend the concept of central exposed to risk to cover more elaborate situations.

ii) Calculation of exposed to risk:

Rita

Rita turned 30 on 1 October 2009, when she was already married. She died on 1 January 2010, 3 months after her 30th birthday.

Thus, Rita's contribution to central exposed to risk = 3 months

And contribution to initial exposed to risk = 1 year

Sita

Sita turned 30 on 1 September 2011, when she was already married. Time spent under investigation, aged 30 last birthdays by Sita was 1 September 2011 – 31 August 2012.

Thus, Sita's contribution to both central and initial exposed to risk is 1 year.

Nita

Nita turned 30 on 1 December 2009 and married 2 months later. Therefore, she joined the investigation of married women on 1 February 2010. She divorced 9 months later, when she would be censored from the investigation of married women.

Thus, Nita's contribution to both central and initial exposed to risk is 9 months.

Gita

Gita got married on 1 June 2011, at which time she was already past her 31st birthday. Therefore, she has spent no time during the investigation period as a married woman at age 30 last birthday.

Thus, her contribution to both central and initial exposed to risk is nil.

iii) Total exposed to risk:

Hence, total exposed to risk is:

Central exposed to risk = 0.25 + 1 + 0.75 + 0 = 2 years.

Initial exposed to risk = 1 + 1 + 0.75 + 0 = 2.75 years

From the results above, it can be seen that the central exposed to risk is 2 years and the initial exposed to risk is 2.75 years. The approximation would suggest that the initial exposed to risk should be 2.5 years. However, this is not a good approximation for the data provided as the approximation is based on the assumption that deaths would be evenly spread and thus can be assumed to occur half way through the year, on average. This also relies on an implicit assumption of a reasonably large data set. In the data above, there were only 4 lives, which is not statistically significant. Moreover, there was only one death, which occurred 3 months after the 30thbirthday. As a result of the statistical sparseness in the data, the approximation is seen not to work very well.