ACSC/STAT 4703, Actuarial Models II

Fall 2015

Toby Kenney

Homework Sheet 3

Model Solutions

Basic Questions

1. *An insurance company collects the following claim data (in thousands):*

i	d_i	x_i	u_i	i	d_i	x_i	u_i	i	d_i	x_i	u_i
1	0	0.4	-	8	1.0	-	15	15	2.0	-	10
2	0	1.6	-	9	1.0	4.6	-	16	2.0	-	10
3	0	-	20	10	1.0	-	15	17	2.0	2.6	-
4	0	1.8	-	11	1.0	1.3	-	18	2.0	-	20
5	0	-	10	12	1.5	-	10	19	2.0	14.6	-
6	0.5	1.9	-	13	1.5	6.8	-	20	5.0	-	15
7	0.5	1.6	-	14	1.5	1.9	-	21	5.0	8.4	-

Using a Kaplan-Meier product-limit estimator:

(a) estimate the probability that a random loss exceeds 17.3.

x_i	s_i	r_i	$S(x_i)$
0.4	1	5	0.8
1.3	1	10	0.72
1.6	2	12	0.6
1.8	1	10	0.54
1.9	2	9	0.42
2.6	1	12	0.385
4.6	1	11	0.35
6.8	1	12	0.3208333
8.4	1	11	0.2916667
14.6	1	6	0.24305556

So the estimated survival probability is 0.24305556.

(b) estimate the median of the distribution.

From the above table, we see that the survival probability is 0.54 from 1.8 to 1.9, but 0.42 from 1.9 to 2.6, so the median is 1.9.

(c) Use a Nelson-Åalen estimator to estimate the median of the distribution.

The Nelson-Åalen estimator is an estimator for the cumulative hazard rate. That is $H(x) = -\log(S(x))$. The median is the solution to $S(x) = \frac{1}{2}$, which gives $H(x) = -\log\left(\frac{1}{2}\right) = \log(2) = 0.6931472$. We therefore need to adapt the table from part (a) to calculate the cumulative hazard rate, and find the value where it first exceeds 0.6931472.

x_i	s_i	r_i	$H(x_i)$
0.4	4	5	0.2
1.3	9	10	0.3
1.6	10	12	0.4667
1.8	9	10	0.5667
1.9	7	9	0.7889

So again 1.9 is the estimate for the median.

2. *An insurance company observes the following claim history:*

Number of claims	*Frequency*
0	*2,846*
1	*701*
2	*360*
3	*202*
4	*114*
5	*56*
6	*12*
7	*0*
8	*2*

Use a Nelson-Åalen estimate to obtain a 95% confidence interval for the probability that a random individual makes more than 5 claims.

A Nelson-Åalen estimate gives $H(5) = \frac{2846}{4293} + \frac{701}{1447} + \frac{360}{746} + \frac{202}{386} + \frac{114}{184} + \frac{56}{70} = 3.572845$.

The variance of this estimator is approximately $\frac{2846}{4293^2} + \frac{701}{1447^2} + \frac{360}{746^2} + \frac{202}{386^2} + \frac{114}{184^2} + \frac{56}{70^2} = 0.01728762$

A 95% confidence interval is therefore

$H(x) \in 3.572845 \pm 1.96\sqrt{0.01728762} = [3.315144, 3.830546]$.

The corresponding interval for the survival function is

$$S(x) \in [e^{-3.830546}, e^{-3.315144}] = [0.02169777, 0.03632882]$$

3. *For the data in Question 1, use Greenwood's approximation to obtain a 95% confidence interval for the probability that a random loss exceeds 17.3, based on the Kaplan-Meier estimator.*

(a) Using a normal approximation

Recall from Question 1, that the estimated survival function is $S(17.3) = 0.24305556$. The variance of the Kaplan-Meier estimator is

$0.24305556^2 \left(\left(1 + \frac{1-0.8}{5\times0.8}\right)\left(1 + \frac{0.8-0.72}{10\times0.72}\right)\left(1 + \frac{0.72-6}{12\times0.6}\right)\left(1 + \frac{0.6-0.54}{10\times0.54}\right)\left(1 + \frac{0.54-0.42}{9\times0.42}\right)\left(1 + \frac{0.42-0.385}{12\times0.385}\right)\left(1 + \frac{0.385-0.35}{11\times0.35}\right)\left(1 + \frac{0.35-0}{12\times0.}\right.\right.$
0.01198086

The confidence interval is therefore $0.24305556 \pm 1.96\sqrt{0.01198086} = [0.02852356, 0.45758756]$.

Using Greenwood's approximation for the product, the variance of the Kaplan-Meier estimator is

$$0.24305556^2 \left(\frac{1}{5\times4} + \frac{1}{10\times9} + \frac{2}{12\times10} + \frac{1}{10\times9} + \frac{2}{9\times7} + \frac{1}{12\times11} + \frac{1}{11\times10} + \frac{1}{12\times11} + \frac{1}{11\times10} + \frac{1}{6\times5}\right) = 0.01106503$$

which gives a confidence interval of $[0.03688607, 0.44922505]$.

(b) Using a log-transformed confidence interval.

Using the product formula, we have $U = e^{\frac{1.96\sqrt{0.01198086}}{0.24305556\log(0.24305556)}} = 0.5357895$, and the confidence interval is $[0.24305556^{\frac{1}{U}}, 0.24305556^{U}] = [0.07136381, 0.46867036]$.

Using the variance from Greenwood's approximation, we have $U = e^{\frac{1.96\sqrt{0.01106503}}{0.24305556\log(0.24305556)}} = 0.548982$, and the confidence interval is $[0.24305556^{\frac{1}{U}}, 0.24305556^{U}] = [0.07603786, 0.46000591]$.

4. *An insurance company records the following data in a mortality study:*

entry	*death*	*exit*	*entry*	*death*	*exit*	*entry*	*death*	*exit*
60.4	-	*64.4*	*61.6*	-	*64.2*	*62.1*	-	*63.9*
62.7	-	*63.7*	*60.8*	-	*63.8*	*62.9*	-	*64.5*
63.4	-	*64.4*	*64.3*	-	*66.3*	*61.8*	*63.7*	-
61.2	-	*63.2*	*63.3*	-	*66.3*	*60.2*	*60.6*	-
62.2	-	*65.2*	*62.8*	-	*64.8*	*63.8*	*65.2*	-
60.9	-	*62.9*	*61.3*	-	*63.3*	*62.1*	*63.4*	-
63.0	-	*65.6*	*62.1*	-	*65.1*			

Estimate the probability of an individual currently aged exactly 63 dying within the next year using:

(a) the exact exposure method.

Using the exact exposure, the total exposure for lives currently aged 63 is $1 + 0.7 + 0.6 + 0.2 + 1 + 0 + 1 + 1 + 0.8 + 0 + 0.7 + 1 + 0.3 + 1 + 0.9 + 1 + 0.7 + 0 + 0.2 + 0.4 = 12.5$. The number of deaths aged 63 is 2, so the hazard rate is $\frac{2}{12.5} = 0.16$. The probability of dying aged 63 is therefore $1 - e^{-0.16} = 0.1478562$.

(b) the actuarial exposure method.

Using the actuarial exposure method, the exposure is $1 + 0.7 + 0.6 + 0.2 + 1 + 0 + 1 + 1 + 0.8 + 0 + 0.7 + 1 + 0.3 + 1 + 0.9 + 1 + 1 + 0 + 0.2 + 1 = 13.4$. The number of deaths aged 63 is 2, so the mortality probability is $\frac{2}{13.4} = 0.1492537$.

Standard Questions

5. *An insurance company collects the following claim data (in thousands):*

i	d_i	x_i	u_i	i	d_i	x_i	u_i	i	d_i	x_i	u_i
1	*0*	*0.7*	-	*8*	*1.0*	-	*20*	*15*	*2.0*	*4.1*	-
2	*0*	*1.3*	-	*9*	*1.0*	*4.2*	-	*16*	*2.0*	-	*15*
3	*0*	-	*10*	*10*	*1.0*	-	*10*	*17*	*2.0*	*2.9*	-
4	*0*	*11.8*	-	*11*	*1.0*	*1.5*	-	*18*	*2.0*	*8.6*	-
5	*0.5*	-	*15*	*12*	*1.0*	-	*10*	*19*	*5.0*	-	*10*
6	*0.5*	-	*15*	*13*	*1.5*	*4.8*	-	*20*	*5.0*	-	*15*
7	*1.0*	*3.6*	-	*14*	*1.5*	*2.9*	-	*21*	*5.0*	*18.4*	-

It is attempting to price a new policy with a deductible of 1.0. Using a Kaplan-Meier estimator, calculate the probability that a random claim on a policy with a deductible of 1.0 exceeds 5.0.

We have the following table. (Since we are calculating the conditional survival, we do not need to consider claim values less than 1.0.

x_i	s_i	r_i	$S_1(x_i)$
1.3	1	11	0.909
1.5	1	10	0.818
2.9	2	15	0.709
3.6	1	13	0.655
4.1	1	12	0.6
4.2	1	11	0.545
4.8	1	10	0.491

so the probability is 0.4909 [or $\frac{27}{55}$].

6. *An insurance company has historical data from 2,861 claims. It finds that 1,830 are less than \$5,000, 793 are between \$5,000 and \$20,000, 168 are between \$20,000 and \$100,000, and the remaining 70 are more than \$100,000. Calculate a 95% confidence interval for the probability that a random claim is more than \$30,000.*

   The empirical estimates are $F_n(20000) = \frac{2623}{2861}$ and $F_n(100000) = \frac{2791}{2861}$, so the ogive gives

   $$F_n(30000) = \frac{70000}{80000} \times \frac{2623}{2861} + \frac{100000}{80000} \times \frac{2791}{2861} = \frac{2644}{2861}$$

   The variance of $F_n(20000)$ is $\frac{2623\times 238}{2861^3}$ and the variance of $F_n(100000) - F_n(20000)$ is $\frac{168\times 2693}{2861^3}$, and the covariance is $-\frac{168\times 2623}{2861^3}$, so the variance of $F_n(20000) + \frac{1}{8}(F_n(100000) - F_n(20000))$ is $\frac{2623\times 238}{2861^3} + \frac{1}{64} \times \frac{168\times 2693}{2861^3} - \frac{2}{8} \times \frac{168\times 2623}{2861^3} = \frac{4085481}{8\times 2861^3}$

   The confidence interval for $F_n(30000)$ is therefore $\frac{2644}{2861} \pm 1.96\sqrt{\frac{4085481}{8\times 2861^3}} = [0.9149996, 0.9333052]$ so the confidence interval for $S_n(30000)$ is $[0.06669477, 0.08500045]$.

7. *An insurance company observes the following claims (in thousands):*

   0.8 2.3 5.7 4.2 11.6 8.7 3.0 7.4 1.5 15.2 9.3 2.5 3.8

   using a kernel density estimate with a triangular kernel with bandwidth 1, estimate the expected loss per claim if the company introduces a deductible of 2.0 on each policy.

   We sort the sample to get

   ```
   0.8 1.5 2.3 2.5 3.0 3.8 4.2 5.7 7.4 8.7 9.3 11.6 15.2
   ```

   Under the kernel density distribution, the probability that a random loss exceeds 2.0 is $\frac{0+0.125+0.755+0.875+1+1+1+1+1+1+1+1+1}{13} = \frac{10.755}{13} = 0.8273$

   The conditional expected loss of a claim given that the claim exceeds 2.0 is therefore

$$\frac{\int_2^{2.5} x(2.5-x)\,dx + \int_2^{2.3} x(x-1.3)\,dx + \int_{2.3}^{3.3} x(3.3-x)\,dx + \int_2^{2.5} x(x-1.5)\,dx + \int_{2.5}^{3.5} x(3.5-x)\,dx + 69.9}{10.755}$$
$$= \frac{69.9 + 1.25(2.5^2-2^2) - \frac{2.5^3-2^3}{3} + \frac{2.3^3-2^3}{3} - 0.65(2.3^2-2^2) + 3.3(3.3^2-2.3^2) - \frac{3.3^3-2.3^3}{3} + \frac{2.5^3-2^3}{3} - 0.75(2.5^2-2^2) + 1.75(3.5^2-2.5^2) - \frac{3.5^3-2.5^3}{3}}{10.755}$$
$$= \frac{69.9 + 1.25\times 2.25 - \frac{7.625}{3} + \frac{4.167}{3} - 0.65\times 1.29 + 3.3\times 5.6 - \frac{23.77}{3} + \frac{7.625}{3} - 0.75\times(2.25) + 1.75\times 6 - \frac{27.25}{3}}{10.755}$$
$$= \frac{69.9 + 2.8125 - 0.8385 + 18.48 - 1.6975 + 10.5 + \frac{4.167}{3} - \frac{23.77}{3} - \frac{27.25}{3}}{10.755}$$
$$= \frac{99.1565 + -\frac{46.853}{3}}{10.755}$$
$$= 7.767442$$