Instituto Superior de Economia e Gestão

UNIVERSIDADE TÉCNICA DE LISBOA

MASTER IN ACTUARIAL SCIENCE

Risk Models

18/01/2013

Time allowed: 3 hours

Instructions:

1. This paper contains 8 questions and comprises 3 pages including the title page.
2. Enter all requested details on the cover sheet.
3. You have 10 minutes reading time. You must not start writing your answers until instructed to do so.
4. Number the pages of the paper where you are going to write your answers.
5. Attempt all 8 questions.
6. Begin your answer to each of the 8 questions on a new page.
7. Marks are shown in brackets. Total marks: 200.
8. Show calculations where appropriate.
9. An approved calculator may be used.
10. The distributed formulary and the Formulae and Tables for Actuarial Examinations (the 2002 edition) may be used. Note that the parametrization used for the different distributions is that of the distributed formulary.

1. For observation *i* of a survival study:

- d_i is the left truncation point
- x_i is the observed value if not right censored
- u_i is the observed value if right censored

You are given:

Observation (i)	d_i	x_i	u_i
1	0	0.9	–
2	0	–	1.3
3	0	1.4	–
4	0	–	1.5
5	0	–	1.5
6	0	1.7	–
7	0.8	–	1.7
8	1.3	2.1	–
9	1.5	1.7	–
10	1.6	–	2.3

a) Determine the Kaplan-Meier Product-Limit estimate, $S_n(1.6)$. Is the Kaplan-Meier approach a parametric or a non–parametric one? Explain. [marks 15]

b) Determine a log-transformed confidence interval (90%) for S(1.6) using Kaplan-Meier approach. [marks 10]

c) Using the Kaplan-Meier approach, determine a confidence interval (95%) for $_{0.9}q_{1.1}$. [marks 15]

d) Using the Nelson-Aalen approach give another estimate for S(1.6). [marks 5]

e) Assuming that the random variable X (without censoring and without truncation) follows a known distribution with unknown parameter θ write the log-likelihood function needed to estimate this parameter in terms of $f(x|\theta)$ and $F(x|\theta)$ using available information. [marks 10]

f) An actuary fitted a Weibull distribution ($\alpha = 5$ and θ unknown) to this data set in order to estimate θ. He used R and got the following output (remember that function nlm minimizes minus the loglikelihood function):

```
> param.start=c(1.5)
> out2=nlm(minusloglik,param.start,hessian=TRUE)
> out2
    $minimum
    [1] 3.117363
    $estimate
    [1] 2.214719
    $gradient
    [1] 1.40362e-09
    $hessian
         [,1]
    [1,] 15.30339
```

```
$code
[1] 1
$iterations
[1] 14
```

i. What is your estimate for $S(1.6)$? [marks 5]

ii. Using the asymptotic distribution of the maximum likelihood estimators, determine a 95% confidence interval for θ and obtain the corresponding confidence interval for $S(1.6)$. [marks 10]

2. For the data set 100; 200; 300; 400; X you are given

   (i) $k = 4$ – Number of unique values of x in the sample
   (ii) $s_2 = 1$
   (iii) $r_4 = 1$
   (iv) The Nelson-Åalen Estimate $\hat{H}(410) > 2.15$

   Determine *X*. [marks 10]

3. From a population having density function f, you are given the following sample: 2.0, 3.3, 3.3, 4.0, 4.0, 4.7, 4.7, 4.7

   Estimate $f(3.5)$ using the kernel $k_y(x) = 0.75 - 0.75 \times (x-y)^2$, $y-1 < x < y+1$.

   [marks 15]

4. You are given:

   (i) Losses follow a loglogistic distribution with cumulative distribution function: $F(x) = \dfrac{(x/\theta)^\gamma}{1+(x/\theta)^\gamma}$

   (ii) The sample of losses is:

   10; 35; 80; 86; 90; 120; 158; 180; 200; 210; 1500

   Calculate the estimate of θ by percentile matching, using the 40th and 80th empirically smoothed percentile estimates. [marks 15]

5. Assume that the value of a claim amount can be considered as a random variable, X, with density function $f(x|\theta) = 3\theta^{-1} x^2 e^{-x^3/\theta}$, $x > 0$, $\theta > 0$.
   From this population we collected a random sample with n observations.

   a) Show that the maximum likelihood estimator for θ is given by

   $$\hat{\theta} = \frac{\sum_i^n X_i^3}{n}.$$ [marks 15]

   b) Is the maximum likelihood estimator for θ unbiased? What can you say about consistency? (If necessary, you can use the fact that $Y = X^3$ follows an exponential distribution with expected value θ). [marks 15]

6. Consider that we observed a random sample $\left(x_1, x_2, \cdots, x_n\right)$ from a population following a Poisson distribution with parameter λ. Using a Bayesian framework we assume that our prior for λ is a gamma distribution with parameters α and θ.

   a) Show that the posterior distribution of λ is a gamma distribution with parameters $\alpha + \sum_{i=1}^{n} x_i$ and $\theta / (n\theta + 1)$ respectively. [marks 10]

   b) Show that, for this situation, the Bayes estimate using a zero-one loss function is always smaller than the Bayes estimate using a squared loss function (assume that $\alpha + \sum_{i=1}^{n} x_i > 1$). [marks 15]

7. You fit a Pareto distribution to a sample of 200 claim amounts and use the likelihood ratio test to test the hypothesis that $\alpha = 1.5$ and $\theta = 7.8$.

   You are given:

   (i) The maximum likelihood estimates are $\hat{\alpha} = 1.4$ and $\hat{\theta} = 7.6$.

   (ii) The natural logarithm of the likelihood function evaluated at the maximum likelihood estimates is −817.92.

   (iii) $\sum_{i=1}^{200} \ln\left(x_i + 7.8\right) = 607.64$

   Determine the result of the test. [marks 15]

8. Assume that the number of claims of a given policy, N, follows a Poisson distribution with parameter 0.12 and that the claims amounts, $\{X_i\}$, $i = 1, 2, \cdots$, can be considered independent exponential random variables with mean 1200 and are also independent of N.
   Explain how to use simulation to approximate the sampling distribution of the aggregate claim amount, $Y = \sum_{i=0}^{N} X_i$, $X_0 \equiv 0$. Also explain how to estimate $\Pr(Y > 1000)$ and the 95th percentile of the distribution of Y. [marks 20]

Solutions – **18/01/2013**

1.

j	y_j	r_j	s_j	$\hat{\pi}_j$	$S_n(y_j)$	$\hat{H}(y_j)$	$\hat{S}(y_j)$	$\sqrt{\hat{\text{var}}(S_n(y_j))}$
1	0.9	7=7-0-0	1	0.8571	0.8571	0.1429	0.8669	0.132
2	1.4	6=8-1-1	1	0.8333	0.7143	0.3095	0.7338	0.171
3	1.7	5=10-2-3	2	0.6	0.4286	0.7095	0.4919	0.187
4	2.1	2=10-4-4	1	0.5	0.2143	1.2095	0.2983	0.176

a) $S_n(1.6)=S_n(1.4)=\frac{(7-1)}{7}\times\frac{(6-1)}{6}=0.7143$. The Kaplan-Meier approach is a non-parametric estimation procedure since we do not assume the knowledge of any functional form for the distribution function.

b) Using Greenwood's formula we get

$\hat{\text{var}}\left(S_n(1.6)\right)\approx 0.7143^2\times\left(\frac{1}{7\times 6}+\frac{1}{6\times 5}\right)=0.0292$, i.e., $\sqrt{\hat{\text{var}}\left(S_n(1.6)\right)}\approx 0.1707$

$$U=\exp\left(1.645\times\frac{0.1707}{0.7143\times\ln(0.7143)}\right)=\exp(-1.1686)=0.3108$$

Then the 95% confidence interval is given by $\left(0.7143^{1/.3108};0.7143^{0.3108}\right)$, i.e., $\left(0.3387;0.9007\right)$

c) $S_n(1.1)=S_n(1.0)=\frac{(7-1)}{7}=0.8571$;

$$S_n(2.0)=S_n(1.7)=\frac{(7-1)}{7}\times\frac{(6-1)}{6}\times\frac{(5-2)}{5}=0.4286$$

$${}_{0.9}\hat{q}_{1.1}=\frac{S_n(1.1)-S_n(2)}{S_n(1.1)}=\frac{0.8571-0.4286}{0.8571}=0.5$$

$$\hat{\text{var}}\left({}_{0.9}\hat{q}_{1.1}\,|\,S(1.1)=S_n(1.1)\right)=\frac{1}{S_n(1.1)^2}\hat{\text{var}}\left(S_n(2)\,|\,S(1.1)=S_n(1.1)\right)$$

$$=\frac{S_n(2.0)^2}{S_n(1.1)^2}\times\left(\frac{1}{6\times 5}+\frac{2}{5\times 3}\right)=0.0417$$

The 95% CI is then given by ${}_{0.9}\hat{q}_{1.1}\pm 1.96\sqrt{\hat{\text{var}}\left({}_{0.9}\hat{q}_{1.1}\,|\,S(1.1)=S_n(1.1)\right)}$, i.e. (0.0999; 0.9001)

d) $H_n(1.6)=\frac{1}{7}+\frac{1}{6}=0.3095$

$$S_n(1.6)=\exp\left(-H_n(1.6)\right)=e^{-0.3095}=0.7338$$

e) Distribution function $F(x\,|\,\theta)$ and density $f(x\,|\,\theta)$

$$L(\theta)=f(0.9|\theta)\times\left(1-F(1.3|\theta)\right)\times f(1.4|\theta)\times\left(1-F(1.5|\theta)\right)^2\times f(1.7|\theta)\times$$
$$\times\frac{\left(1-F(1.7|\theta)\right)}{\left(1-F(0.8|\theta)\right)}\times\frac{f(2.1|\theta)}{\left(1-F(1.3|\theta)\right)}\times\frac{f(1.7|\theta)}{\left(1-F(1.5|\theta)\right)}\times\frac{\left(1-F(2.3|\theta)\right)}{\left(1-F(1.6|\theta)\right)}$$

$$\ell(\theta) = \ln f(0.9|\theta) + \ln\left(1 - F(1.3|\theta)\right) + \ln f(1.4|\theta) + 2\ln\left(1 - F(1.5|\theta)\right) + \ln f(1.7|\theta) +$$
$$+ \ln\left(1 - F(1.7|\theta)\right) - \ln\left(1 - F(0.8|\theta)\right) + \ln f(2.1|\theta) - \ln\left(1 - F(1.3|\theta)\right) +$$
$$+ \ln f(1.7|\theta) - \ln\left(1 - F(1.5|\theta)\right) + \ln\left(1 - F(2.3|\theta)\right) - \ln\left(1 - F(1.6|\theta)\right)$$

f) i. $\hat{\theta} = 2.214719$ then $\hat{S}(1.6) = \exp\left(-(1.6/2.215)^5\right) = 0.8214$

ii. $\hat{\text{var}}(\hat{\theta}) \approx -\left(-1/15.30339\right) = 0.065345$

95% confidence interval for θ → $2.215 \pm 1.96 \times \sqrt{0.065345}$, (1.714; 2.716)

95% confidence interval for $S(1.6)$

Solution 1: $\left(\exp\left(-(1.6/1.714)^5\right); \exp\left(-(1.6/2.716)^5\right)\right)$ → (0.492; 0.931)

Solution 2: Delta method

$S(1.6) = \exp\left(-(1.6/\theta)^5\right) = g(\theta)$

$g'(\theta) = 5 \times 1.6^5 \times \theta^{-6} \times \exp\left(-(1.6/\theta)^5\right)$

$g'(\hat{\theta}) = 5 \times 1.6^5 \times \hat{\theta}^{-6} \times \exp\left(-(1.6/\hat{\theta})^5\right) = 0.364916$

$\hat{\text{var}}\left(\hat{S}(1.6)\right) = 0.364916^2 \times \hat{\text{var}}\left(\hat{\theta}\right) = 0.008702$

CI → $0.8214 \pm 1.96\sqrt{0.008702}$, i.e. (0.6385; 1.004) →(0.6385; 1)

2.

$k = 4$, then X has to be one of the other observed values

$s_2 = 1$, then X cannot be 200

$r_4 = 1$, then X cannot be 400

So, X can be equal to 100 or to 300.

Let us try $X = 100$

$$\hat{H}(410) = \frac{2}{5} + \frac{1}{3} + \frac{1}{2} + \frac{1}{1} = 0.4 + 0.33 + 0.5 + 1 = 2.23 > 2.15$$

Now try $X = 300$

$$\hat{H}(410) = \frac{1}{5} + \frac{1}{4} + \frac{2}{3} + \frac{1}{1} = 0.2 + 0.25 + 0.67 + 1 = 2.12 < 2.15$$

Consequently, $X = 100$.

3.

$k_y(x) = 0.75 - 0.75\times(x-y)^2$, $y-1 < x < y+1$

y_j	$p(y_j)$	$k_{y_j}(3.5)$
2.0	0.125	3.5≥2+1 then $k_{2.0}(3.5)=0$
3.3	0.25	$2.3 \le 3.5 < 4.3$ then $k_{3.3}(3.5) = 0.75 - 0.75\times(3.5-3.3)^2 = 0.72$
4.0	0.25	$3.0 \le 3.5 < 5.0$ then $k_{4.0}(3.5) = 0.75 - 0.75\times(3.5-4)^2 = 0.5625$
4.7	0.375	3.5<4.7-1 then $k_{4.7}(3.5)=0$

Then $\hat{f}(3.5) = 0.125\times 0 + 0.25\times 0.72 + 0.25\times 0.5625 + 0.375\times 0 = 0.32$

4.

$$F(x) = \frac{(x/\theta)^\gamma}{1+(x/\theta)^\gamma} = \frac{1}{(x/\theta)^{-\gamma}+1}$$

$$g = \frac{1}{(\pi/\theta)^{-\gamma}+1} \Leftrightarrow \left(\frac{\pi}{\theta}\right)^{-\gamma} + 1 = \frac{1}{g} \Leftrightarrow \frac{\theta}{\pi} = \left(\frac{1}{g}-1\right)^{1/\gamma} \Leftrightarrow \pi = \theta\left(\frac{1-g}{g}\right)^{-1/\gamma} \Leftrightarrow \pi = \theta\left(\frac{g}{1-g}\right)^{1/\gamma}$$

Empirical percentiles

$\hat{\pi}_{0.4}$: $12\times 0.4 = 4.8$ then $\hat{\pi}_{0.4} = 0.2\times 86 + 0.8\times 90 = 89.2$

$\hat{\pi}_{0.8}$: $12\times 0.8 = 9.6$ then $\hat{\pi}_{0.8} = 0.4\times 200 + 0.6\times 210 = 206.0$

We need to solve

$$\begin{cases} 89.2 = \theta\left(\dfrac{0.4}{1-0.4}\right)^{1/\gamma} \\ 206 = \theta\left(\dfrac{0.8}{1-0.8}\right)^{1/\gamma} \end{cases} \Leftrightarrow \begin{cases} 89.2 = \theta\left(\dfrac{2}{3}\right)^{1/\gamma} \\ 206 = \theta\times 4^{1/\gamma} \end{cases} \Leftrightarrow \begin{cases} \dfrac{206}{89.2} = \left(\dfrac{4}{2/3}\right)^{1/\gamma} \\ \theta = 4^{-1/\gamma}\times 206 \end{cases} \Leftrightarrow \begin{cases} \dfrac{206}{89.2} = 6^{1/\gamma} \\ \theta = 4^{-1/\gamma}\times 206 \end{cases}$$

$$\begin{cases} \dfrac{206}{89.2} = 6^{1/\gamma} \\ \theta = 4^{-1/\gamma}\times 206 \end{cases} \Leftrightarrow \begin{cases} \ln\left(\dfrac{206}{89.2}\right) = \dfrac{1}{\gamma}\ln(6) \\ \theta = 4^{-1/\gamma}\times 206 \end{cases} \Leftrightarrow \begin{cases} \gamma = \dfrac{\ln(6)}{\ln\left(\dfrac{206}{89.2}\right)} \\ \theta = 4^{-1/\gamma}\times 206 \end{cases}$$

And we get

$\tilde{\gamma} = 2.141$ and $\tilde{\theta} = 107.8$

5.

a) $$\ell(\theta) = \sum_{j=1}^{n} \ln\left(3\theta^{-1} x_j^2 e^{-x_j^3/\theta}\right) = \sum_{j=1}^{n}\left(\ln 3 - \ln\theta + 2\ln x_j - \frac{x_j^3}{\theta}\right)$$

$$\ell'(\theta) = \sum_{j=1}^{n}\left(-\frac{1}{\theta} + \frac{x_j^3}{\theta^2}\right) = -\frac{n}{\theta} + \frac{\sum_{j=1}^{n} x_j^3}{\theta^2}$$

$$\ell'(\theta)=0 \Leftrightarrow \frac{\sum_{j=1}^{n} x_j^3}{\theta^2}=\frac{n}{\theta} \Leftrightarrow \theta=\frac{\sum_{j=1}^{n} x_j^3}{n}$$

$$\ell''(\theta)=\frac{n}{\theta^2}-2\frac{\sum_{j=1}^{n} x_j^3}{\theta^3}=\frac{n}{\theta^2}\left(1-\frac{2}{\theta}\frac{\sum_{j=1}^{n} x_j^3}{n}\right)$$ and then $\ell''(\hat{\theta})=-\frac{n}{\hat{\theta}^2}<0$

Then the ML estimator is $\hat{\theta}=\frac{\sum_{j=1}^{n} X_j^3}{n}$

b) $E(\hat{\theta})=E\left(\frac{\sum_{i=1}^{n} X_i^3}{n}\right)=\frac{\sum_{i=1}^{n} E\left(X_i^3\right)}{n}=\frac{n\theta}{n}=\theta$

$\hat{\theta}$ is an unbiased estimator for θ
As the estimator is unbiased we only need to check the variance of the estimator.

$$\operatorname{var}\hat{\theta}=\operatorname{var}\left(\frac{\sum_{i=1}^{n} X_i^3}{n}\right)=\frac{\sum_{i=1}^{n} \operatorname{var}(Y_i)}{n^2}=\frac{n\theta^2}{n^2}=\frac{\theta^2}{n}$$

As $\lim_{n\to\infty} \operatorname{var}(\hat{\theta})=\lim_{n\to\infty}\frac{\theta^2}{n}=0$, $\hat{\theta}$ is a consistent estimator for θ

6.

a) $X \mid \lambda \sim Po(\lambda)$

$\lambda \sim \gamma(\alpha,\theta)$

$$L(\lambda \mid x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}\frac{\lambda^{x_i}e^{-\lambda}}{(x_i!)} \propto \lambda^{\sum x_i} e^{-n\lambda}$$

$$\pi(\lambda)=\frac{\lambda^{\alpha-1}e^{-\lambda/\theta}}{\theta^{\alpha}\Gamma(\alpha)} \propto \lambda^{\alpha-1}e^{-\lambda/\theta}$$

$$\pi(\lambda \mid x_1,x_2,\cdots,x_n) \propto \lambda^{\alpha-1}e^{-\lambda/\theta}\lambda^{\sum x_i}e^{-n\lambda}=\lambda^{\alpha-1+\sum x_i}e^{-\lambda(n\theta+1)/\theta}$$

Then

$$\lambda \mid x_1,x_2,\cdots,x_n \sim \gamma\left(\alpha+\sum_{i=1}^{n} x_i;\theta/(n\theta+1)\right)$$

b) Bayes estimate using a squared loss function: mean of the gamma distribution

$$\hat{\lambda}_B=\frac{\theta}{1+n\theta}\times(\alpha+\sum_{i=1}^{n} x_i)$$

Bayes estimate using a zero-one loss function: mode of the gamma distribution

$$\hat{\lambda}_{0-1}=\frac{\theta}{1+n\theta}\times(\alpha+\sum_{i=1}^{n} x_i-1)=\hat{\lambda}_B-\frac{\theta}{1+n\theta}<\hat{\lambda}_B$$

7.

$H_0 : \alpha = 1.5 \wedge \theta = 7.8$ against $H_1 : \alpha \neq 1.5 \vee \theta \neq 7.8$

Likelihood ratio test

$$\Lambda = -2\ln\left(\frac{L(1.5,7.8)}{L(\hat{\alpha},\hat{\theta})}\right) = -2\left(\ell(1.5,7.8) - \ell(\hat{\alpha},\hat{\theta})\right) \overset{\circ}{\sim} \chi^2_{(2)}$$

Pareto distribution

$$f(x \mid \alpha,\theta) = \alpha\theta^{\alpha}\left(x+\theta\right)^{-(\alpha+1)}$$

$$\ell(\alpha,\theta) = \sum_{i=1}^{n}\left(\ln\alpha + \alpha\ln\theta - (\theta+1)\ln(x_i+\theta)\right) = n\ln\alpha + n\alpha\ln\theta - (\alpha+1)\sum_{i=1}^{n}\ln(x_i+\theta)$$

$$\ell(1.5,7.8) = 200\ln 1.5 + 300\ln 7.8 - 2.5\times 607.64 = -821.77$$

$$\ell(\hat{\alpha},\hat{\theta}) = -817.92$$

$\lambda_{\text{obs}} = -2\times\left(-821.77 + 817.92\right) = 7.70$ p-value less than 0.0224 Reject H_0

8.

- Determine the number of replicas, NR. Usually a large value like 50000.
- For each replica
  - Generate the number of claims, N according to a Poisson with parameter 0.1. We can generate a random number and use the inverse transformation method to generate the Poisson (other methods are possible)
  - If we generate $N = 0$ we get $Y = 0$ otherwise we need to generate N claim amounts. For each claim amount we generate a pseudo random number and we use the inverse transformation method to generate an exponentially distributed random variable. We sum the claim amounts to get the generated value for Y
  - Keep the Y value in an array, called z, with NR elements.
- The NR elements of array z are used to approximate the distribution of Y.
- To estimate $\Pr(Y > 1000)$ just count how many values in z are greater than 1000 and divide by NR.
- To estimate the 95th percentile of the distribution of Y, compute the corresponding empirical percentile which will be our estimate.