LISBOA SCHOOL OF ECONOMICS & MANAGEMENT

MASTER IN ACTUARIAL SCIENCE

Risk Models

12/01/2015

Time allowed: 3 hours

Instructions:

1. This paper contains 9 questions and comprises 4 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 9 questions.
6. Begin your answer to each of the 9 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 parameterization used for the different distributions is that of the distributed formulary.

1. You are performing a survival analysis where, for observation i, d_i is the left truncation point, x_i is the time of death and u_i corresponds to a right censoring point.

Observation	1	2	3	4	5	6	7	8	9	10
d_i	0	0	0	0	0.1	0.1	0.3	0.7	0.9	1.3
x_i	0.8	1.2	-	1.5	1.2	-	0.6	1.2	1.5	2.1
u_i	-	-	1.4	-	-	1.4	-	-	-	-

   a. **[15]** Obtain an estimate of $S(1.4)$ using the product-limit estimator and also obtain another estimate of $S(1.4)$ using the Nelson-Aalen approach.
   b. **[10]** Obtain a 95% approximate log-transformed confidence interval for $S(1.4)$ using the Kaplan-Meier approach.

2. **[15]** Product limit estimation is being applied to a data set which includes some right-censored and some left-truncated data. Successive death times are $y_1, y_2, \cdots, y_n$. The following estimated values are given:
   - Product limit estimate of $S(y_2)$ is 0.912
   - Product limit estimate of $S(y_3 \mid X > y_1)$ is 0.874
   - Nelson-Aalen estimate of $H(y_2) - H(y_1)$ is 0.05

   Find the product limit estimate of $S(y_3)$.

3. **[15]** A random sample of size 4 is (1; 2; 2; 3). Kernel smoothing is applied to the data using a Pareto Kernel ($\alpha = 2$ and $\theta = y$). Find the Kernel estimate of $S(2)$.

4. The percentile matching method is applied to the estimation of parameter θ for the random variable with density function $f(x \mid \theta) = (\theta + 1)x^{\theta}$, $0 < x < 1$, $\theta > 0$.
   a. **[15]** Assuming that we observed (0.31; 0.51; 0.61; 0.65; 0.86; 0.94; 0.97) use the 80th sample percentile to obtain an estimate of θ.
   b. **[10]** Now using another sample and based on the sample median, one obtained the estimate 0.75647 for $E(X)$. Find the value of the sample median.

5. Let X be a random variable with a beta distribution where a is unknown and $b=2$, i.e. $f(x\,|\,a)=(a+1)\,a\,x^{a-1}(1-x)$, $0<x<1$, $a>0$. For a random sample of size $n=100$, it is found that $\bar{x}=0.4009$, $s=0.23521$, $t=\sum_{i=1}^{100}\ln x_i=-113.6571$.
   a. **[15]** Find the maximum likelihood estimate of a and determine a 95% confidence interval for a.
   b. **[15]** Using the maximum likelihood estimator determine a 95% confidence interval for $\mu=E(X)$.
   c. **[10]** Without taking advantage of the maximum likelihood estimator determine another 95% confidence interval for $\mu=E(X)$.

6. **[15]** A Pareto distribution is known to have a value of $\theta=100$. A random sample of 100 observations is taken from the distribution. Thirty of the observations were right censored at 50 (all thirty values are over 50, but the actual values are unknown). The maximum likelihood estimate of α is 3.5. Suppose that of the thirty data points censored at 50, twenty five of them are greater than 60, and the others are known to be 52, 55, 56, 56 and 58. Using this information find the revised mle of α.

7. Let us assume that we observed a sample of size $n=50$, $(x_1,x_2,\cdots,x_{50})$, where $\bar{x}=0.3$, from a Bernoulli population with parameter θ. Let us also assume a Bayesian approach and that our prior is a beta distribution with parameters 2 and 4, i.e., $\pi(\theta)=20\,\theta\left(1-\theta\right)^3$, $0<\theta<1$.
   a. **[15]** Show that the posterior is a beta distribution with parameters 17 and 39.
   b. **[10]** Obtain a Bayes estimate using a quadratic loss function.
   c. **[10]** Obtain a HPD interval for θ using the Bayesian Central Limit theorem.

8. A random sample of size $n=100$, $(x_1,x_2,\cdots,x_{100})$, was observed. You are given the following output from R software (remember that, using the function nlm, we minimize minus the log-likelihood)

```
> m=mean(x); m
[1] 9.531913
> param_start=c(1,m)
> out=nlm(minusloglik_weibull,param_start)
> out
$minimum
[1] 322.6313
```

```
$estimate
[1]  1.215457 10.177031
$gradient
[1]  1.089674e-05 -1.010969e-06
$code
[1] 1
$iterations
[1] 13
```

Using the likelihood ratio test you want to test if the population is exponential versus Weibull.

a. **[5]** Explain why the likelihood ratio test is appropriate.
b. **[10]** Perform the test ($\alpha = 0.05$) and write down all the likelihood functions used.

9. **[15]** Explain how to use simulation to get an approximation of the sampling distribution of the median of a sample of 6 observations from an exponential population with mean 10. Using this distribution explain how to estimate the probability that the sample median is greater than the sample mean?

Solutions

1.

y_j	0.6	0.8	1.2	1.5	2.1
r_j	7	7	7	3	1
s_j	1	1	3	2	1
$(r_j - s_j)/r_j$	0.857143	0.857143	0.571429	0.333333	0
$S_n(y_j)$	0.857143	0.734694	0.419825	0.139942	0
s_j / r_j	0.142857	0.142857	0.428571	0.666667	1
$\hat{H}(y_j)$	0.142857	0.285714	0.714286	1.380952	2.380952
$\hat{S}(y_j)$	0.866878	0.751477	0.489542	0.251339	0.092462

a. Product limit estimate: $S_n(1.4) = 0.4198$

Nelson-Aalen estimate: $\hat{S}(1.4) = 0.4895$

b. Greenwood's formula:

$$\hat{\text{var}}\left(S_n(1.4)\right) \approx S_n(1.4)^2 \times \sum_{i: y_i \le 1.4} \frac{s_i}{r_i\left(r_i - s_i\right)}$$

$$= 0.4198^2 \left(\frac{1}{7\times 6} + \frac{1}{7\times 6} + \frac{3}{7\times 4}\right) = 0.027274$$

$$U = \exp\left(z_{\alpha/2} \frac{\sqrt{\hat{\text{var}}\left(S_n(1.4)\right)}}{S_n(1.4)\times \ln S_n(1.4)}\right) = \exp\left(1.96 \frac{\sqrt{0.027274}}{0.4198\times \ln(0.4198)}\right) = 0.411337$$

Then the 95% CI is given by $\left(\left(S_n(1.4)\right)^{1/U}; \left(S_n(1.4)\right)^{U}\right)$, i.e.

$(0.1212; 0.6998)$

2.

$$S_n(y_2) = \frac{r_1 - s_1}{r_1} \times \frac{r_2 - s_2}{r_2} = 0.912 \qquad (1)$$

$$S(y_3 \mid X > y_1) = \Pr(X > y_3 \mid X > y_1) = \frac{\Pr(X > y_3)}{\Pr(X > y_1)} = \frac{S(y_3)}{S(y_1)}$$ as $y_3 > y_1$.

Then, as $S_n(y_3 \mid X > y_1) = 0.874$,

$$S_n(y_3) = 0.874 \times S_n(y_1) \Leftrightarrow \frac{r_2 - s_2}{r_2} \times \frac{r_3 - s_3}{r_3} = 0.874 \quad (2)$$

As $\hat{H}(y_2) - \hat{H}(y_1) = \frac{s_1}{r_1} + \frac{s_2}{r_2} - \frac{s_1}{r_1} = \frac{s_2}{r_2}$ we get $\frac{s_2}{r_2} = 0.05 \quad (3)$

Combining (1) and (3),

$$\frac{r_1 - s_1}{r_1}\times(1-0.05)=0.912 \Leftrightarrow \frac{r_1 - s_1}{r_1}=\frac{0.912}{0.95} \Leftrightarrow S_n(1)=0.96$$

And then , using (2), $S_n(y_3)=0.874\times S_n(y_1)=0.874\times 0.96=0.839$

3. $K_y(x)=1-\left(\frac{y}{x+y}\right)^2 \quad x>0$

y_j	$p(y_j)$	$K_{y_j}(2)$
1	.25	$1-(1/3)^2=8/9\approx 0.8889$
2	.50	$1-(2/4)^2=0.75$
3	.25	$1-(3/5)^2=0.64$

Then

$$\hat{S}(2)=1-\hat{F}(2)=1-\left(0.25\times 8/9+0.5\times 0.75+0.25\times 0.64\right)=1-0.757222=0.2428$$

4.

a. The percentile matching estimate will be the solution of the equation $F(\tilde{\pi}_{0.8})=0.8$ in order to θ where $\tilde{\pi}_{0.8}$ is the 80th empirical percentile. $(n+1)\times 0.8=6.4$, then $\tilde{\pi}_{0.8}=0.6\times 0.94+0.4\times 0.97=0.952$.

As $F(x)=\int_0^x(\theta+1)u^\theta\,du=x^{\theta+1}$ (for $0<x<1$) the equation is $0.952^{\theta+1}=0.8$. The solution is $\theta=\frac{\ln 0.8}{\ln 0.952}-1=3.536$ and then the estimate will be $\tilde{\theta}=3.536$

b. Let $\mu=E(X)=\int_0^1 x(\theta+1)x^\theta\,dx=\int_0^1(\theta+1)x^{\theta+1}\,dx=\frac{\theta+1}{\theta+2}$. Then, as the percentile estimate of μ is $\tilde{\mu}=\frac{\tilde{\theta}+1}{\tilde{\theta}+2}=0.75647$, we can deduce that the percentile estimate of θ based on the median is $\tilde{\theta}=\frac{2\times 0.75647-1}{1-0.75647}=2.10627$.

If we denote by m the sample median, we know that $F(m\,|\,\theta=2.10627)=0.5$ and we solve this equation in order to m.

$$F(m\,|\,\theta=2.10627)=0.5 \Leftrightarrow m^{3.10627}=0.5 \Leftrightarrow m=0.5^{1/3.10627}=0.8$$

5.

a. $f(x\mid a)=(a+1)\,a\,x^{a-1}(1-x)$, $0<x<1$, $a>0$

$$\ell(a)=\ln L(a)=\sum_{i=1}^{n}\left(\ln(a+1)+\ln a+(a-1)\ln x_i+\ln(1-x_i)\right)$$

$$\ell'(a)=\sum_{i=1}^{n}\left(\frac{1}{a+1}+\frac{1}{a}+\ln x_i\right)=\frac{n}{a+1}+\frac{n}{a}+t \quad \text{where } t=\sum_{i=1}^{n}\ln x_i$$

$$\ell''(a)=-\frac{n}{(a+1)^2}-\frac{n}{a^2}<0$$

$$\ell'(a)=0\Leftrightarrow \frac{n}{a+1}+\frac{n}{a}+t=0\Leftrightarrow \frac{an+an+n+ta^2+ta}{a(a+1)}=0$$

$$\Leftrightarrow a^2t+a(2n+t)+n=0$$

$$a=\frac{-(2n+t)\pm\sqrt{(2n+t)^2-4tn}}{2t}$$, i.e. $a=1.391827$ or $a=-0.6321474$.

As $a>0$, the MLE is $\hat{a}=1.3918$.

$$\widehat{\text{var}}(\hat{a})\approx -1/\ell''(\hat{a})=0.01447149$$

95% CI for a: $1.3918\pm1.96\times\sqrt{0.01447149}$ →(1.1560; 1.6276)

b. $\mu=a/(a+2)$ then $\mu=\hat{a}/(\hat{a}+2)=0.4103$

$$\text{var}(\hat{\mu})\approx\left(\frac{d}{da}\left(\frac{a}{(a+2)}\right)\right)^2\text{var}(\hat{a})=\left(\frac{2}{(a+2)^2}\right)^2\text{var}(\hat{a})$$

$$\widehat{\text{var}}(\hat{\mu})\approx\left(\frac{2}{(\hat{a}+2)^2}\right)^2\widehat{\text{var}}(\hat{a})=0.0004373597$$

95% CI for μ: $0.4103\pm1.96\times\sqrt{0.0004374}$ →(0.3694; 0.4513)

c. Using an extension of the CLT, $\frac{\bar{X}-\mu}{S/\sqrt{n}}\overset{\circ}{\sim} n(0;1)$ and then

95% CI for μ: $0.4009\pm1.96\times(0.23521/10)$ →(0.3548; 0.4470)

6. $\theta=100$ $\quad F(x\mid\alpha)=1-\theta^{\alpha}(x+\theta)^{-\alpha}$ $\quad f(x\mid\alpha)=\alpha\theta^{\alpha}(x+\theta)^{-(\alpha+1)}$

Let us assume that the observations are ranked in ascending order.

30 right censored observations at 50:

$$\ell_1(\alpha)=\sum_{i=1}^{70}\left(\ln\alpha+\alpha\ln\theta-(\alpha+1)\ln(x_i+\theta)\right)+30\left(\alpha\ln\theta-\alpha\ln(50+\theta)\right)$$

$$\ell_1'(\alpha)=\frac{70}{\alpha}+70\ln100-\sum_{i=1}^{70}\ln(x_i+100)+30\ln100-30\ln150$$

$$=\frac{70}{\alpha}-\sum_{i=1}^{70}\ln(x_i+100)+100\ln100-30\ln150$$

As $\hat{\alpha} = 3.5$ is the mle, $\ell_1'(3.5) = 0$, i.e.

$$0 = \frac{70}{3.5} - \sum_{i=1}^{70} \ln(x_i + 100) + 100\ln 100 - 30\ln 150 \Leftrightarrow \sum_{i=1}^{70} \ln(x_i + 100) = 330.198$$

Now, using additional information:

$$\ell_2(\alpha) = \sum_{i=1}^{75} \left(\ln\alpha + \alpha\ln\theta - (\alpha+1)\ln(x_i + \theta)\right) + 25\left(\alpha\ln\theta - \alpha\ln(60+\theta)\right)$$

$$\ell_2'(\alpha) = \frac{75}{\alpha} + 75\ln 100 - \sum_{i=1}^{75} \ln(x_i + 100) + 25\ln 100 - 25\ln 160$$

$$= \frac{75}{\alpha} - \sum_{i=1}^{75} \ln(x_i + 100) + 100\ln 100 - 25\ln 160$$

$$\ell_2'(\alpha) = 0 \Leftrightarrow 0 = \frac{75}{\alpha} - \sum_{i=1}^{75} \ln(x_i + 100) + 100\ln 100 - 25\ln 160$$

$$\Leftrightarrow 0 = \frac{75}{\alpha} - (330.198 + \ln 152 + \ln 155 + 2\ln 156 + \ln 158) + 100\ln 100 - 25\ln 160$$

$$\Leftrightarrow 0 = \frac{75}{\alpha} - (330.198 + \ln 152 + \ln 155 + 2\ln 156 + \ln 158) + 100\ln 100 - 25\ln 160$$

$$\Leftrightarrow 0 = \frac{75}{\alpha} - 21.7899 \Leftrightarrow \alpha = \frac{75}{21.7899} = 3.44196$$

Then the mle of α is 3.4420.

7.

a. $L(\theta \mid x_1, x_2, \cdots, x_n) = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i}$

$\pi(\theta) = 20\theta(1-\theta)^3 \propto \theta(1-\theta)^3$

$\pi(\theta \mid x_1, x_2, \cdots, x_n) \propto \mathrm{L}(\theta) \times \pi(\theta) = \theta^{1+\sum_{i=1}^n x_i} (1-\theta)^{n+3-\sum_{i=1}^n x_i}$

Then $\theta \mid x_1, x_2, \cdots, x_n \sim beta\left(2 + \sum_{i=1}^n x_i, 4 + n - \sum_{i=1}^n x_i\right)$

b. Bayes estimate against a quadratic loss function:

$$\hat{\theta}_B = E(\theta \mid x_1, x_2, \cdots, x_n) = \frac{2 + \sum_{i=1}^n x_i}{2 + \sum_{i=1}^n x_i + 4 + n - \sum_{i=1}^n x_i} = \frac{2 + \sum_{i=1}^n x_i}{6+n} = \frac{17}{56}$$

c. Bayesian central limit theorem $\theta \mid x_1, \cdots, x_n \overset{\circ}{\sim} normal$

As $\theta \mid x_1, x_2, \cdots, x_n \sim beta\left(2 + \sum_{i=1}^n x_i, 4 + n - \sum_{i=1}^n x_i\right)$, i.e.

$\theta \mid x_1, x_2, \cdots, x_{50} \sim beta(17;39)$ we know that (see Loss models appendix)

$E(\theta \mid x_1, x_2, \cdots, x_n) = \frac{17}{17+39} = \frac{17}{56}$ and $E(\theta^2 \mid x_1, x_2, \cdots, x_n) = \frac{17 \times 18}{56 \times 57}$

And then $\text{var}(\theta \mid x_1, x_2, \cdots, x_n) = \frac{17\times 18}{56\times 57} - \frac{17^2}{56^2} = 0.003709$

The HPD interval is then

$\left(\frac{17}{56} - 1.96\sqrt{0.003709}; \frac{17}{56} + 1.96\sqrt{0.003709}\right)$ i.e. (0.1842; 0.4229)

8.

a. The likelihood ratio test is adequate because the exponential distribution is nested in the Weibull family, i.e. the exponential distribution is a Weibull with $\tau = 1$.

b. $H_0 : \tau = 1$ against $H_1 : \tau \neq 1$

Likelihood ratio test: $\Lambda = -2\ln\left(\frac{L(1,\hat{\theta}^*)}{L(\hat{\tau},\hat{\theta})}\right) = -2\left(\ell(1,\hat{\theta}^*) - \ell(\hat{\tau},\hat{\theta})\right) \overset{\circ}{\sim} \chi^2_{(1)}$

Where $\hat{\theta}^*$ is the mle of an exponential distribution (Weibull with $\tau = 1$) and $\hat{\tau}$ and $\hat{\theta}$ are the mle of a Weibull distribution.

Log-likelihood of the Weibull distribution:

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

Using R output we get $\hat{\tau} = 1.2155$, $\hat{\theta} = 10.1770$ and

$\ell(\hat{\tau},\hat{\theta}) = -322.6313$

Log-likelihood of the Exponential distribution:

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

$$\ell'(\theta) = -\frac{n}{\theta} + \frac{n\,\overline{x}}{\theta^2} \qquad \ell'(\theta) = 0 \Leftrightarrow \theta = \overline{x}$$

$$\ell''(\theta) = \frac{n}{\theta^2} - 2\frac{n\,\overline{x}}{\theta^3} \qquad \ell''(\overline{x}) = \frac{n}{\overline{x}^2} - 2\frac{n\,\overline{x}}{\overline{x}^3} = -\frac{n}{\overline{x}^2} < 0$$

Then $\hat{\theta}^* = \overline{x}$ and

$\ell(1,\hat{\theta}^*) = -n\ln\overline{x} - n = -100\ln 9.5319 - 100 = -325.4645$

$\Lambda_{obs} = -2(-325.4645 + 322.6313) = 5.6664$

As the critical value is 3.841 we reject H0 and conclude that the Weibull is a better model.

9.

a. Define NR, the number of replicas to be used (large value)

b. For each replica $j = 1, 2, \cdots, NR$

   i. Generate 6 uniforms, $u_1, u_2, \cdots, u_6$ and using the inversion method determine the corresponding exponential observations, $x_1, x_2, \cdots, x_6$. $x_i = -\theta \ln(1 - u_i) = -10 \ln(1 - u_i)$

   ii. Sort the sample $x_{(1)} \le x_{(2)} \le \cdots \le x_{(6)}$ and compute the median for replica j as $m_j = \dfrac{x_{(3)} + x_{(4)}}{2}$

c. Use the ecdf or a kernel approach to get an approximation to the sampling distribution of the sample median.

To estimate the probability that the sample median is greater than the sample mean, define for each replica a dummy variable assuming value 1 when this happens. Then estimate the probability computing the average value of the dummy variable.