MASTER IN ACTUARIAL SCIENCE

Risk Models

29/01/2015

Time allowed: 3 hours

Instructions:

1. This paper contains 10 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 10 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 given the following data set for the times of death or right censoring (+) for 25 individuals with the same age after a chirurgical intervention:
2, 3, 3, 3+, 4, 4, 4, 4, 4+, 5+, 6, 6, 7, 7, 7, 7+, 7+, 8, 9, 10, 12+, 13, 13, 14, 16+.
    a. **[10]** Based on the product limit estimator, obtain a 95% linear confidence interval for $S(11)$.
    b. **[15]** Obtain a log-transformed confidence interval for $S(11)$ based on Nelson-Aalen approach.

2. You are given the following grouped data on 100 losses based on underlying loss random variable X:

Interval	$(0,100]$	$(100,200]$	$(200,600]$
Number of losses	a	b	c

a,b and c are non-negative integers. You are also given that the empirical estimate of $E(X)$ is 212.5 and that the empirical estimate of $E(X \wedge 100)$ is 87.5.
    a. **[10]** Find the values of a,b and c.
    b. **[10]** Find $F_{100}(200)$ and $F_{100}(500)$ the empirical estimates of the distribution function value at 200 and 500 respectively. If you were not able to get the values of the constants a,b and c assume that $a=20$ and $b=50$. In this case clearly state that you are assuming these values that constitute an incorrect answer to the previous question.

3. **[15]** Consider a sample where all values are observed without censoring and without truncation. Prove that the empirical cumulative distribution function is an unbiased estimator of the distribution function for all values of x. What is the mean square error of this estimator.

4. **[15]** You are given the following sample with 5 observations (2; 4; 4; 5; 7). Using Kernel estimation (the chosen kernel for data point y is a normal distribution with mean y and variance 1) obtain an estimate for $f(3)$.

5. **[15]** A ground up loss distribution is assumed to follow a Pareto distribution with $\alpha=2$ and unknown θ. A sample of 10 observations based on policy limit amount of 800 has been observed: There are 6 observations below the limit and they total 2350. There are 4 observations of limit amount (censored observations). Using the method of moments, give an estimate of θ.

6. The following random sample was observed from an inverse gamma population

0.7; 0.9; 1.0; 1.4; 2.3; 2.7; 2.9; 5.8; 8.8; 11.6

   a. **[10]** Assuming that α is known obtain the maximum likelihood estimate of θ as a function of α.
   b. **[15]** Now assume that a sample with size $n = 100$ was collected from another inverse gamma population where $\alpha = 2.5$. Using this new setup one get $\hat{\theta} = 2.2658$. As you know a 95% confidence interval for θ can be obtained using an approximation of the observed information (usual approach using the second derivative of the loglikelihood) or Fisher's information $I(\theta)$. Determine the 95% confidence intervals that are obtained using each of these approaches.

7. An insurer has data for 8 independent policies. Each policy has the same ground up **loss** distribution that is Weibull with parameters θ and τ. The policy information is as follows:

Policy	1	2	3	4	5	6	7	8
Payment	10	10	10	10	10	10	10	10
Deductible	none	None	5	5	none	none	5	5
Payment Limit	none	10	none	10	none	10	none	10

Remember that losses below the deductible are not reported and that losses above the deductible are paid in excess of the deductible.

   a. **[15]** Write the log-likelihood function to be maximized.
   b. **[10]** Assume that $\tau = 2$ and determine the mle of θ.

8. Assume that we observed the sample (5.0; 6.3; 5.7) from a population with density function given by $f(x|\theta) = \frac{\theta^3}{2} e^{-\theta x} x^2$, $x > 0$, $\theta > 0$. Using a Bayesian framework, we define our prior as $\pi(\theta) = e^{-\theta}$, $\theta > 0$.
   a. **[20]** Show that the posterior distribution is a gamma distribution with parameters $\alpha = 10$ and $\theta = 1/18$ and compute a Bayes estimate for θ assuming a quadratic loss function.
   b. **[10]** Obtain the predictive distribution.

9. **[15]** The following sample was observed (7.67; 3.78; 2.57; 4.28; 10.44). Using the Kolmogorov-Smirnov test, test if it is acceptable $(\alpha = 0.05)$ to consider that the population is Weibull distributed with $\tau = 2$ and $\theta = 10$.

10. **[15]** A random sample of size $n = 15$ with weekly returns of a financial asset (assumed i.i.d.) has been observed. Our purpose is to estimate the volatility of the returns measured by the standard deviation. However we know that the behavior of the weekly returns is not normally distributed. Explain how we can use the bootstrap technique to approximate the sampling distribution of the sample standard deviation and analyze the mean square error of this estimator.

Solutions

1.

y_j	2	3	4	6	7	8	9	10	13	14
r_j	25	24	21	15	13	8	7	6	4	2
s_j	1	2	4	2	3	1	1	1	2	1
$(r_j - s_j)/r_j$	0.96	0.9167	0.8095	0.8667	0.7692	0.8750	0.8571	0.8333	0.5000	0.5000
$S_n(y_j)$	0.96	0.8800	0.7124	0.6174	0.4749	0.4156	0.3562	0.2968	0.1484	0.0742
s_j / r_j	0.04	0.0833	0.1905	0.1333	0.2308	0.1250	0.1429	0.1667	0.5000	0.5000
$\hat{H}(y_j)$	0.04	0.1233	0.3138	0.4471	0.6779	0.8029	0.9458	1.1124	1.6124	2.1124
$\hat{S}(y_j)$	0.9608	0.8840	0.7307	0.6395	0.5077	0.4480	0.3884	0.3288	0.1994	0.1209
s_j / r_j^2	0.0016	0.0035	0.0091	0.0089	0.0178	0.0156	0.0204	0.0278	0.1250	0.2500

a. Product limit estimate: $S_n(11) = 0.2968$

Greenwood's formula

$$\hat{\text{var}}\left(S_n(11)\right) \approx S_n(11)^2 \times \sum_{i: y_i \le 11} \frac{s_i}{r_i\left(r_i - s_i\right)} = 0.0881 \times 0.1250 = 0.0110$$

The 95% linear CI for $S(11)$ is then $0.2968 \pm 1.96 \times \sqrt{0.0110}$, i.e. (0.0911; 0.5025)

b. Nelson-Aalen estimate of $H(11)$: $\hat{H}(11) = 1.1124$

$$\hat{\text{var}}\left(\hat{H}(11)\right) \approx \sum_{i: y_i \le 11} \frac{s_i}{r_i^2} = 0.0016 + 0.0035 + \ldots + 0.0278 = 0.1046$$

$$U = \exp\left(z_{\alpha/2} \frac{\sqrt{\hat{\text{var}}\left(\hat{H}(11)\right)}}{\hat{H}(11)} \right) = \exp\left(1.96 \frac{\sqrt{0.1046}}{1.1124}\right) = 1.7679$$

Then the 95% CI for $H(11)$ is given by $\left(\hat{H}(11)/U ; \hat{H}(11) \times U\right)$, i.e. (0.6292; 1.9667) and the corresponding interval for $S(11)$ is (0.1399; 0.5330) using $S(x) = e^{-H(x)}$

2.

a. We know that:

$$a+b+c=100$$

$$50\times\frac{a}{100}+150\times\frac{b}{100}+400\times\frac{c}{100}=212.5 \Leftrightarrow 0.5a+1.5b+4c=212.5$$

$$50\times\frac{a}{100}+100\times\frac{b+c}{100}=87.5 \Leftrightarrow 0.5a+b+c=87.5$$

From the 1st equation we get $b+c=100-a$ and applying this result to the 3rd equation leads to $0.5a+100=87.5 \Leftrightarrow a=25$. So $b+c=100-25 \Leftrightarrow c=75-b$ and using this result in the 2nd equation $12.5+1.5b+4(75-b)=212.5 \Leftrightarrow 2.5b=100 \Leftrightarrow b=40$ and then $c=100-25-40=35$

b. Solution using the correct values

$$F_{100}(200)=\frac{a+b}{100}=0.65;$$

$$F_{100}(500)=\frac{100\times F_{100}(200)+300\times F_{100}(600)}{400}=\frac{100\times 0.65+300\times 1}{400}=\frac{365}{400}=0.9125$$

Solution using the given values

$$F_{100}(200)=0.70;\ F_{100}(500)=\frac{100\times 0.7+300\times 1}{400}=\frac{370}{400}=0.925$$

3. Given $x\in\Re$, the ecdf is defined as $F_n(x)=\frac{N_x}{n}$, where n is the sample size and $N_x=\#\{X_i\le x\}$ is the number of observations, in the sample that are less than or equal to x.

As each identically distributed observation in the sample is less than or equal to x with probability $F(x)$ (the population distribution function) and considering that the observations are independent we get $N_x \sim b(n,F(x))$.

Then

$$E\left(F_n(x)\right)=E\left(\frac{N_x}{n}\right)=\frac{E(N_x)}{n}=\frac{nF(x)}{n}=F(x)$$

$$mse\left(F_n(x)\right)=\operatorname{var}\left(F_n(x)\right) \qquad \text{The estimator is unbiased}$$

$$=\operatorname{var}\left(\frac{N_x}{n}\right)=\frac{\operatorname{var}(N_x)}{n^2}=\frac{nF(x)(1-F(x))}{n^2}=\frac{F(x)(1-F(x))}{n}$$

4. $k_y(x)=(2\pi)^{-1/2}\exp\left(-\frac{(x-y)^2}{2}\right) \qquad x>0$

y_j	$p(y_j)$	$k_{y_j}(3)$
2	0.2	$(2\pi)^{-1/2}\exp\left(-\frac{(3-2)^2}{2}\right)=0.24197$
4	0.4	$(2\pi)^{-1/2}\exp\left(-\frac{(3-4)^2}{2}\right)=0.24197$
5	0.2	$(2\pi)^{-1/2}\exp\left(-\frac{(3-5)^2}{2}\right)=0.05399$
7	0.2	$(2\pi)^{-1/2}\exp\left(-\frac{(3-7)^2}{2}\right)=0.00001$

Then

$$\hat{f}(3)=\left(0.2\times 0.24197+0.4\times 0.24197+0.2\times 0.05399+0.2\times 0.00001\right)=0.1560$$

5. $E(X\wedge 800)=\frac{\theta}{\alpha-1}\left(1-\left(\frac{\theta}{\theta+800}\right)^{\alpha-1}\right)=\theta\left(\frac{\theta+800-\theta}{\theta+800}\right)=\frac{800\theta}{\theta+800}$

The empirical estimate of $E(X\wedge 800)$ is

$$\tilde{E}(X\wedge 800)=\frac{2350+4\times 800}{10}=235+320=555$$

The moment estimate of θ is the solution of $555=\frac{800\theta}{\theta+800}$

$$555=\frac{800\theta}{\theta+800}\Leftrightarrow 555\theta+555\times 800=800\theta\Leftrightarrow \theta=\frac{800\times 555}{245}=1812.245$$

6. $f(x)=\theta^{\alpha}\left(\Gamma(\alpha)\right)^{-1}x^{-(\alpha+1)}\,\mathrm{e}^{-\theta/x} \qquad x>0,\ \alpha,\theta>0$

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

$\ell'(\theta)=\sum_{i=1}^{n}\left(\frac{\alpha}{\theta}-\frac{1}{x_i}\right) \qquad \ell''(\theta)=-\sum_{i=1}^{n}\frac{\alpha}{\theta^2}=-\frac{n\alpha}{\theta^2}<0$

$$\ell'(\theta)=0\Leftrightarrow\frac{n\alpha}{\theta}=\sum_{i=1}^{n}\frac{1}{x_i}\Leftrightarrow\theta=\frac{n\alpha}{\sum_{i=1}^{n}(1/x_i)}$$

Then the mle is $\hat{\theta}=\frac{10\alpha}{5.7762}=1.7312\alpha$

b. Using the observed information we obtain the usual result

$$\hat{\text{var}}(\hat{\theta}) = -1/\ell''(\hat{\theta}) = 2.2658^2/250 = 0.020535$$

And the CI is $2.2658 \pm 1.96\sqrt{0.020535}$ → (1.9849; 2.5467)

Now, using Fisher's information

$$I_n(\theta) = -E\left(\ell''(\theta \mid X_1, \cdots, X_n)\right) = E\left(\frac{n\alpha}{\theta^2}\right) = \frac{n\alpha}{\theta} = \frac{250}{\theta}$$

$$\text{var}(\hat{\theta}) \approx \frac{1}{I(\theta)} = \frac{\theta}{n\alpha} = \frac{\theta}{250}$$

$$-1.96 < \frac{\hat{\theta}-\theta}{\sqrt{\theta^2/(n\alpha)}} < 1.961 - \frac{1.96}{\sqrt{n\alpha}} < \frac{\hat{\theta}}{\theta} < 1 + \frac{1.96}{\sqrt{n\alpha}} \Leftrightarrow -\frac{1.96}{\sqrt{n\alpha}} < \frac{\hat{\theta}}{\theta} - 1 < \frac{1.96}{\sqrt{n\alpha}}$$

$$\Leftrightarrow \frac{\hat{\theta}}{1+\frac{1.96}{\sqrt{n\alpha}}} < \theta < \frac{\hat{\theta}}{1-\frac{1.96}{\sqrt{n\alpha}}}$$

And the CI is (2.0159; 2.5864)

7. $f(x) = \tau\theta^{-\tau} x^{\tau-1} \exp(-(x/\theta)^{\tau}) \qquad x > 0,\ \tau, \theta > 0$

$F(x) = 1 - \exp(-(x/\theta)^{\tau}) \qquad S(x) = \exp(-(x/\theta)^{\tau})$

a. For policy 3 the payment is 10 after a deductible of 5, so the ground-up loss is 15. A policy limit of 10 and a deductible of 5 imply that losses greater than 15 are paid 10.

$$L(\theta) = f(10) \times S(10) \times \frac{f(15)}{S(5)} \times \frac{S(15)}{S(5)} \times f(10) \times S(10) \times \frac{f(15)}{S(5)} \times \frac{S(15)}{S(5)}$$

$$= \left(f(10)\right)^2 \times \left(f(15)\right)^2 \times \left(S(10)\right)^2 \times \left(S(15)\right)^2 \times \left(S(5)\right)^{-4}$$

$$\ell(\theta) = 2\ln\left(f(10)\right) + 2\ln\left(f(15)\right) + 2\ln\left(S(10)\right) + 2\ln\left(S(15)\right) - 4\ln\left(S(5)\right)$$

$$= 2\left(\ln\tau - \tau\ln\theta + (\tau-1)\ln 10 - (10/\theta)^{\tau}\right) + 2\left(\ln\tau - \tau\ln\theta + (\tau-1)\ln 15 - (15/\theta)^{\tau}\right) -$$

$$-2(10/\theta)^{\tau} - 2(15/\theta)^{\tau} + 4(5/\theta)^{\tau}$$

$$= 4\ln\tau - 4\tau\ln\theta + 2(\tau-1)(\ln 10 + \ln 15) - 4(10^{\tau} + 15^{\tau} - 5^{\tau})\theta^{-\tau}$$

b.

$$\ell(\theta) = 4\ln 2 - 8\ln\theta + 2\,(\ln 10 + \ln 15) - 4(10^2 + 15^2 - 5^2)\theta^{-2}$$

$$\ell'(\theta) = -8\theta^{-1} + 2400\,\theta^{-3}$$

$$\ell'(\theta) = 0 \Leftrightarrow 8\theta^{-1} = 2400\,\theta^{-3} \Leftrightarrow \theta = \sqrt{300}$$

$$\ell''(\theta) = 8\theta^{-2} - 7200\theta^{-4}$$

Then when $\theta = \sqrt{300}$, $\ell''(\theta) = \frac{8}{300} - \frac{7200}{300^2} = \frac{8}{300} - \frac{24}{300} = -\frac{16}{300} < 0$

and the mle of θ is $\hat{\theta} = \sqrt{300}$

8. $f(x \mid \theta) = \dfrac{\theta^3}{2} e^{-\theta x} x^2$, $x > 0$, $\theta > 0$

a. $L(\theta) = \prod_{i=1}^{n} f(x_i \mid \theta) \propto \theta^{3n} e^{-\theta \sum_{i=1}^{n} x_i} = \theta^9 e^{-17\theta}$, $\theta > 0$

$\pi(\theta) = e^{-\theta}$, $\theta > 0$

$$\pi(\theta \mid x_1, x_2, x_3) \propto \theta^9 e^{-\theta \sum_{i=1}^{n} x_i} e^{-\theta} = \theta^9 e^{-\theta(1+\sum_{i=1}^{n} x_i)} = \theta^9 e^{-18\theta}$$

$$= \theta^9 e^{-\theta/(1/18)} \qquad \theta > 0$$

Then $\theta \mid x_1, x_2, x_3 \sim \mathrm{G}(10; 1/18)$

$\hat{\theta}_B = E(\theta \mid x_1, x_2, x_3) = 10/18$ quadratic loss

b.

$$g(y \mid x_1, x_2, x_3) = \int_0^\infty f(\mathrm{y} \mid \theta)\, \pi(\theta \mid x_1, x_2, x_3)\, d\theta = \int_0^\infty \frac{\theta^3}{2} e^{-\theta y} y^2 \frac{\theta^9 e^{-18\theta}}{\Gamma(10)\,(1/18)^{10}}\, d\theta$$

$$= \frac{y^2}{2\Gamma(10)\,(1/18)^{10}} \int_0^\infty \theta^{12} e^{-\theta(y+18)}\, d\theta \quad \text{core of a gamma } 13; (y+18)^{-1}$$

$$= \frac{18^{10} y^2}{2\Gamma(10)} \Gamma(13) \frac{1}{(y+18)^{13}} = \frac{18^{10} \times 12 \times 11 \times 10}{6} y^2 (y+18)^{-13}$$

$$= 18^{10}\, 660\, y^2 (y+18)^{-13} \qquad y > 0$$

9. $H_0 : F(x) = 1 - \exp(-(x/10)^2)$ $\quad x > 0$, $\tau, \theta > 0$ $\quad H_1 : H_0$ false

x_i	$F(x_i)$	i/n	$(i-1)/n$	D_i
2.57	0.063915	0.2	0	0.136085
3.78	0.133145	0.4	0.2	0.266855
4.28	0.167385	0.6	0.4	0.432615
7.67	0.444723	0.8	0.6	0.355277
10.44	0.663762	1	0.8	0.336238

$D = \max D_i = 0.4326$ Approximate critical value: $1.36/\sqrt{5} = 0.6082$

Conclusion: We do not reject H0.

10.

i. Compute the sample standard deviation of the observed sample

$$s^* = \sqrt{\frac{\sum_{i=1}^{15}(x_i - \overline{x})^2}{14}}$$

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

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

1. Resample from the original sample with replacement. One possible method is to define 15 uniforms, $u_1, u_2, \cdots, u_{15}$ and, for each u_i pick a value y_i using: $y_i = x_k$ if $\frac{k-1}{15} < u_i \le \frac{k}{15}$
2. Compute the pseudo sample standard deviation using $s_j = \sqrt{\frac{\sum_{i=1}^{15}(y_i - \overline{y})^2}{14}}$ and keep it.

iv. Using the set of values s_j, $j = 1, 2, \cdots, NR$, determine the ecdf of the bootstrap distribution of the sample standard deviation.

v. The bootstrap estimate of the mean squared error is given by

$$\frac{\sum_{j=1}^{NR}(s_j - s^*)^2}{NR}$$