Instituto Superior de Economia e Gestão
UNIVERSIDADE TÉCNICA DE LISBOA

MASTER IN ACTUARIAL SCIENCE

Risk Models

12/01/2012

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 of 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 parametrization used for the different distributions is that of the distributed formulary.

1. For a survival study with censored and truncated data, you are given

Time(t)	Risk set at time t	Failures at time t
1.3	30	5
7.4	27	9
9.3	32	6
10.2	25	5
12.5	20	4

   a) Estimate the survival function at time 10 and give a log-transformed confidence interval (95%) for $S(10)$. [marks 15]

   b) Estimate ${}_8q_2$ and also estimate a 95% conditional confidence interval for ${}_8q_2$. What is the conditioning behind the variance used to compute the confidence interval? [marks 15]

2. A survival study gave (0.283; 1.267) as the symmetric linear confidence interval for $H(5)$ (level of confidence $1-\alpha$). Determine the symmetric linear confidence interval for $S(5)$ with the same confidence level as before. [marks 15]

3. You are given the kernel

$$k_y(x)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-(x-y)^2} & y-1\le x\le y+1\\ 0 & \text{otherwise}\end{cases}.$$

   You also are given the following sample 1; 2.5; 3; 3; 5. Determine the kernel density estimator at $x=3$. [marks 20]

4. A sample $(x_1,x_2,\cdots,x_{10})$ is drawn from a distribution with probability density function: $f(x\mid\alpha,\beta)=\dfrac{1}{2}\left(\dfrac{1}{\alpha}e^{-x/\alpha}+\dfrac{1}{\beta}e^{-x/\beta}\right)$, $x>0$, $\alpha>\beta>0$.

   From the sample we get $\sum_{i=1}^{n}x_i=150$ and $\sum_{i=1}^{n}x_i^2=5000$. Estimate α and β using the method of moments. [marks 15]

5. You are given the following 20 bodily injury losses (before the deductible is applied)

Loss	Nº of observations	Deductible	Policy limit
750	3	200	∞
200	3	0	10000
300	4	0	20000
>10000	6	0	10000
400	4	300	∞

Past experience indicates that these losses follow a Pareto distribution with parameters α and $\theta = 10000$. Determine the maximum likelihood estimate of α.

[marks 20]

6. Let us assume that, for a given risk, losses follow a Weibull distribution with unknown parameters $\tau = \theta_1$ and $\theta = \theta_2$ (Loss Models parametrization). All policies have a deductible of 50 (losses lesser than the deductible are not reported to the insurance company and losses greater than the deductible are fully reimbursed). The intent of the company is to analyze the effects of the elimination of the deductible. To get an answer an actuary used R software and obtained the following output

```
> summary(y)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 50.84   77.52  112.70  116.70  142.10  265.10
> length(y)
[1] 80
mloglik1=function(theta){
 -sum(dweibull(y,theta[1],theta[2],log=TRUE)+(t/theta[2])^theta[1])
 }
> theta.start=c(1,mean(y))
>out= nlm(mloglik1,theta.start,hessian=TRUE)
> out
$minimum
[1] 513.7865
$estimate
[1]   1.993743 116.432904
$gradient
[1] -1.138726e-04 -3.776773e-06
$hessian
          [,1]       [,2]
[1,] 33.8551949 -0.73708638
[2,] -0.7370864  0.02930704
$code
[1] 1
$iterations
[1] 15
> solve(out$hessian)
          [,1]      [,2]
[1,] 0.06528645  1.641986
[2,] 1.64198604 75.418241
```

a) Define a 95% confidence interval for the parameter τ. [marks 10]
b) Estimate the probability that a loss is reported when the deductible of 50 is in force. [marks 10]

7. Let us assume a Bernoulli population with parameter θ where, from a Bayesian point of view, our beliefs about θ are modeled using a beta distribution with parameters α and β, (see Formulae and Tables for Examinations) i.e.

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad 0<\theta<1, \quad \alpha,\beta>0$$

Let us also assume that we observed a sample of size n, $(x_1, x_2, \cdots, x_n)$.

a) Show that the posterior distribution is a beta distribution with parameters $t+\alpha$ and $n-t+\beta$ where $t=\sum_{i=1}^{n} x_i$. [marks 15]

b) Assume that $\alpha=\beta=1$.

I. Is the corresponding distribution an improper prior? [marks 5]

II. Compare the Bayes estimate using a zero-one loss function with the maximum likelihood estimate. [marks 5]

c) Now assume that $\alpha=3$, $\beta=2$ and that we observed $\sum_{i=1}^{95} x_i = 60$ and $n=95$. Using Bayes central limit theorem get a credibility interval for θ. [marks 10]

8. You observed the following sample of claim payments: 1.67; 2.35; 2.55; 3.35; 4.89. An actuary claims that the sample came from a Pareto population with parameters $\alpha=2$ and $\theta=10$. Use a Kolmogorov-Smirnov test to discuss the actuary's claim. [marks 25]

9. An actuary needs to estimate the expected value of a claim amount for a given risk. He knows from past experience that the distribution of this kind of claim amounts is highly skewed and then the sample mean can be influenced by extreme observations. To overcome this situation he decides to use a 10% trimmed mean of a sample with 30 observations, i.e. he will use the sample mean after rejecting the 3 largest and the 3 smallest observations. Assume that the claims follow a Pareto distribution ($\alpha=2.1$ and $\theta=100$). Using simulation explain how to get an estimate of the mean squared error of the trimmed mean as an estimator of the mean of the population. [marks 20]

SOLUTION

1.

y_j	r_j	s_j	$(r_j - s_j)/r_j$	$\prod (r_j - s_j)/r_j$
1.3	30	5	25/30	0.8333
7.4	27	9	18/27	0.5556
9.3	32	6	26/32	0.4514
10.2	25	5	20/25	0.3611
12.5	20	4	16/20	0.2889

a) $S_n(10) = 0.4514$;

$$\hat{\mathrm{var}}\, S_n(10) = \left(S_n(10)\right)^2 \times \left(\frac{5}{30 \times 25} + \frac{9}{27 \times 18} + \frac{6}{32 \times 26} \right) = 0.006601$$

$$U = \exp\left(1.96 \frac{\sqrt{\hat{\mathrm{var}}\, S_n(10)}}{S_n(10) \times \ln S_n(10)} \right) = 0.641769$$

$\left(0.4514^{1/0.641769}; 0.4514^{0.641769}\right)$, i.e. $(0.289561; 0.600217)$

b) $_8q_2 = \frac{S(2) - S(10)}{S(2)} = 1 - \frac{S(10)}{S(2)}$; $S_n(2) = 0.8333$; $_8\hat{q}_2 = 0.458298$

$$\hat{\mathrm{var}}\left({}_8\hat{q}_2 \mid S(2) = S_n(2) \right) = \left(\frac{0.4514}{0.8333} \right)^2 \left(\frac{9}{27 \times 18} + \frac{6}{32 \times 26} \right) = 0.00755$$

$$_8\hat{q}_2 \pm 1.96 \sqrt{\hat{\mathrm{var}}\left({}_8\hat{q}_2 \right)} \to (0.28799; 0.628607)$$

The confidence interval is conditioned by $S(2) = S_n(2)$

2. $\hat{H}(5) = \frac{0.283 + 1.267}{2} = 0.775$; $z_{\alpha/2} \sqrt{\hat{\mathrm{var}}\left(\hat{H}(5)\right)} = 1.267 - 0.775 = 0.492$ and then

$\hat{\mathrm{var}}\left(\hat{H}(5)\right) = \left(0.492 / z_{\alpha/2}\right)^2$

$S(5) = e^{-H(5)}$ $\quad$ $\hat{S}(5) = e^{-\hat{H}(5)}$ $\quad$ $\mathrm{var}\left(\hat{S}(5)\right) \approx \left(-e^{-H(5)}\right)^2 \mathrm{var}\left(\hat{H}(5)\right)$

Estimates

$\hat{S}(5) = e^{-0.775} = 0.4607$

$\hat{\mathrm{var}}\left(\hat{S}(5)\right) \approx \left(0.4607 \times 0.492 / z_{\alpha/2}\right)^2$

$$\hat{S}(5) \pm z_{\alpha/2} \times \sqrt{\hat{\mathrm{var}}\left(\hat{S}(5)\right)} = 0.4607 \pm z_{\alpha/2} \times \left(0.4607 \times 0.492 / z_{\alpha/2}\right)$$

$$= 0.4607 \pm \left(0.4607 \times 0.492\right) = 0.4607 \pm 0.2267$$

The confidence interval is then (0.2340; 0.6874)

3. At $x=3$ the only sample values y_j for which $k_{y_j}(3)>0$ are $y_2=2.5$ and $y_3=3$.

   Then

   $$\hat{f}(3)=p(2.5)\times k_{2.5}(3)+p(3)\times k_3(3)=0.2\times\frac{2}{\pi}\times\sqrt{1-(2.5-3)^2}+0.4\times\frac{2}{\pi}\times\sqrt{1-(3-3)^2}$$

   $$=\frac{0.4\times\sqrt{0.75}+0.8}{\pi}=0.364914$$

4. We must equate the first two moments of the population to the corresponding moments of the sample. Note that the given density is a mixture of 2 exponential densities.

   $$E(X)=\frac{1}{2}\alpha+\frac{1}{2}\beta=\frac{\alpha+\beta}{2}$$

   $$E(X^2)=\frac{1}{2}(2\alpha^2)+\frac{1}{2}(2\beta^2)=\alpha^2+\beta^2$$

   The system to be solved is then

   $$\begin{cases}\dfrac{\alpha+\beta}{2}=\dfrac{150}{10}\\ \alpha^2+\beta^2=\dfrac{5000}{10}\end{cases}\Leftrightarrow\begin{cases}\alpha+\beta=30\\ \alpha^2+\beta^2=500\end{cases}\Leftrightarrow\begin{cases}\beta=30-\alpha\\ \alpha^2+(30-\alpha)^2=500\end{cases}$$

   Solving the second equation we get $\alpha=15\pm5$ and then $\beta=15\mp5$. As $\alpha>\beta$, the solution is $\tilde{\alpha}=20$ and $\tilde{\beta}=10$.

5.

$$L(\alpha)=\left(\frac{f(750\,|\,\alpha)}{1-F(200\,|\,\alpha)}\right)^3\times\left(f(200\,|\,\alpha)\right)^3\times\left(f(300\,|\,\alpha)\right)^4\times\left(1-F(10000\,|\,\alpha)\right)^6\times\left(\frac{f(400\,|\,\alpha)}{1-F(300\,|\,\alpha)}\right)^4$$

$$\ell(\alpha)=3\ln f(750\,|\,\alpha)-3\ln\left(1-F(200\,|\,\alpha)\right)+3\ln f(200\,|\,\alpha)+4\ln f(300\,|\,\alpha)+6\ln\left(1-F(10000\,|\,\alpha)\right)+$$
$$+4\ln f(400\,|\,\alpha)-4\ln\times\left(1-F(300\,|\,\alpha)\right)$$

$$\ell(\alpha)=3(\ln\alpha+\alpha\ln10000-(\alpha+1)\ln10750)-3(\alpha\ln10000-\alpha\ln10200)+$$
$$3(\ln\alpha+\alpha\ln10000-(\alpha+1)\ln10200)+4(\ln\alpha+\alpha\ln10000-(\alpha+1)\ln10300)$$
$$+6(\alpha\ln10000-\alpha\ln20000)+$$
$$+4(\ln\alpha+\alpha\ln10000-(\alpha+1)\ln10400)$$
$$-4(\alpha\ln10000-\alpha\ln10300)$$

$$\ell(\alpha)=14\ln\alpha+13\alpha\ln10000-3\alpha\ln10750-6\alpha\ln20000-4\alpha\ln10400$$
$$-3\ln10200-4\ln10300-4\ln10400-3\ln10750$$

$$\ell'(\alpha)=\frac{14}{\alpha}+13\ln 10000-3\ln 10750-6\ln 20000-4\ln 10400$$

$$\ell'(\alpha)=0 \Leftrightarrow \alpha=\frac{14}{3\ln 10750+6\ln 20000+4\ln 10400-13\ln 10000}=3.088648$$

$\ell''(\alpha)=-14/\alpha^2<0$ then $\hat{\alpha}=3.088648$

6.

a) $1.993743 \pm 1.96\times\sqrt{0.06528645}$ → (1.493; 2.495)

b) $\hat{\Pr}(Y>50)=\exp\left(-(50/116.432904)^{1.993743}\right)=0.83078$.

7.

a. $X\,|\,\theta \sim Ber(\theta)$ then $L(\theta\,|\,\mathbf{x})=\prod_{i=1}^{n}\theta^{x_i}(1-\theta)^{1-x_i}=\theta^t(1-\theta)^{n-t}$, where $t=\sum_{i=1}^{n}x_i$ and $0<\theta<1$.

Prior: $\theta\sim \text{beta}(\alpha,\beta)$, i.e. $\pi(\theta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$, $\alpha,\beta>0$, $0<\theta<1$

Simplifying, $\pi(\theta)\propto\theta^{\alpha-1}(1-\theta)^{\beta-1}$, $\alpha,\beta>0$, $0<\theta<1$.

Then

$\pi(\theta\,|\,\mathbf{x})\propto\theta^{\alpha-1}(1-\theta)^{\beta-1}\times\theta^t(1-\theta)^{n-t}=\theta^{t+\alpha-1}(1-\theta)^{n-t+\beta-1}$, $\alpha,\beta>0$, $0<\theta<1$

Which is the core of a beta distribution with parameters $t+\alpha$ and $n-t+\beta$.

b. $\alpha=\beta=1$

i. No. It is a Uniform distribution between 0 and 1.

ii. The posterior is $\pi(\theta\,|\,\mathbf{x})\propto\theta^{t+1-1}(1-\theta)^{n-t+1-1}=\theta^t(1-\theta)^{n-t}$ and then the value of θ that maximizes the posterior (Bayes estimate using a 0-1 loss function) is the same as the value that maximizes the likelihood ($\theta=\bar{x}$).

c. $\alpha=3$, $\beta=2$ and $\sum_{i=1}^{95}x_i=60$, $n=95$.

$\theta\,|\,\mathbf{x}\sim\text{beta}(63;37)$ then $E(\theta\,|\,\mathbf{x})=\frac{63}{63+37}=0.63$ and

$$\text{var}(\theta\,|\,\mathbf{x})=\frac{63\times37}{(63+37)^2(63+37+1)}=0.002308$$

Bayes central limit theorem: $\theta\,|\,\mathbf{x}\overset{\circ}{\sim} n\left(0.63;\sqrt{0.002308}\right)$

95% Credibility interval: $0.63\pm1.96\times\sqrt{0.002308}$ →(0.5358; 0.7242)

Note: using the beta (symmetric interval) we get (0.5336; 0.7215)

8. You observed the following sample of claim payments: 1.67; 2.35; 2.55; 3.35; 4.89. An actuary claims that the sample came from a Pareto population with

parameters $\alpha=2$ and $\theta=10$. Use a Kolmogorov-Smirnov test to discuss the actuary claim.

$$H_0 : F(x) = 1 - \left(\frac{10}{10+x}\right)^2, \quad x > 0$$

$(i-1)/5$	$i/5$	$F(x_{(i)})$	D^-	D^+
0	0.2	0.2657	0.2657	0
0.2	0.4	0.3444	0.1444	0.0556
0.4	0.6	0.3651	0	0.2349
0.6	0.8	0.4389	0	0.3611
0.8	1	0.5490	0	0.4510

$$D = \max(D^-, D^+) = 0.4510$$

Approximate critical value= $1.36/\sqrt{5} = 0.6082$

Then we do not reject H_0.

9.

i. Define NR as the number of replicas to be used in the simulation analysis;

ii. For each replica, $j = 1, 2, \cdots, NR$ do:

  a. Generate 30 random numbers $(u_1, u_2, \cdots, u_n)$
  b. Using the inversion method get 30 Pareto distributed random numbers calculating $x_i = 100 \times \left((1-u_i)^{-1/2.1} - 1\right)$
  c. Sort the x_i sample → $(x_{(1)}, x_{(2)}, \cdots, x_{(30)})$ and calculate $\bar{x}_j^{tr} = (1/24)\sum_{i=4}^{27} x_{(i)}$

iii. Our estimate of the trimmed mean mean squared error is $1/(NR-1)\sum_{j=1}^{NR}\left(\bar{x}_j^{tr} - 100/1.1\right)^2$