INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

UNIVERSIDADE DE LISBOA

MASTER IN ACTUARIAL SCIENCE

Risk Models

17/01/2014

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 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 questions.
6. Begin your answer to each of the 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 (Loss Models book).

1. The following is a sample of 10 payments:

   4 4 5+ 5+ 5+ 8 10+ 10+ 12 15

   where "+" indicates that a loss exceeded the policy limit.

   a. Using Kaplan-Meier estimator and Greenwood's approximation get a 95% confidence interval for $S(11)$. [marks 15]

   b. Using the same approach get a 95% confidence interval for ${}_5q_6$ [marks 15]

   c. Now assume that the claim amounts follow an exponential distribution with mean θ.

      i. Get a maximum likelihood estimate for θ [marks 15]

      ii. Using the asymptotic distribution of the maximum likelihood estimator, present a 95% confidence interval for θ and for $S(11)$ [marks 15]

2. A survival study gave (1.63; 2.55) as the 95% linear confidence interval for the cumulative hazard function at time $t = t_0, H(t_0)$. Calculate the 95% log-transformed confidence interval for $H(t_0)$. [marks 15]

3. From a population having density function f, you are given the following sample (0.11; 0.16; 0.16; 0.16; 0.95; 0.95; 2.75;2.75). Calculate the kernel density estimate of $f(1)$, using a Pareto kernel, $k_y(x) = 2\,y^2\left(y+x\right)^{-3},\ x>0$. [marks 15]

4. You observed the following sample (0.49; 0.51; 0.66; 1.82; 3.71; 5.20; 7.62; 12.66). Using the percentile matching method (percentiles 40th and 80th) estimate the parameters of an inverse Weibull distribution. [marks 15]

5. You are given a random sample with 6 payments. You also know, for each policy in the sample, if a deductible and a policy limit on the losses are in force. We assume that losses below the deductible are not reported and that all the losses above the deductible are reported. Also assume that the payment is net of the deductible (loss minus the deductible). Let $F(x\,|\,\theta)$ be the distribution function of the losses (not the payments) and $f(x\,|\,\theta)$ their density function. The available information is the following:

Observation	1	2	3	4	5	6
Policy limit	none	100	100	none	120	150
Deductible	0	0	10	10	20	0
Payment	25	32	5	15	100	150

   a. Using the available information write the likelihood function for θ in terms of $F(.\,|\,\theta)$ and $f(.\,|\,\theta)$. Do not forget to replace each x_i by its value. [marks 15]

   b. Assuming that $\hat{\theta}$ is the maximum likelihood estimate of θ, give an estimate, in terms of $F(.\,|\,\theta)$ and $f(.\,|\,\theta)$, for the probability that a claim is reported when a deductible of value 50 is in force. Also give an estimate for the probability that a reported claim is censored when a deductible of value 50 and a policy limit of value 100 are in force. [marks 10]

6. Assume that, given θ, the number of claims follows a Poisson distribution with parameter θ. Also assume that the prior distribution for θ is a continuous uniform distribution between 0 and 5. You have observed the random sample $(x_1, x_2, \cdots, x_n)$.

    a. Explaining your deduction, show that the posterior density for θ is proportional to $e^{-n\theta}\,\theta^{n\bar{x}}$, $0 < \theta < 5$. [marks 10]

    b. Assuming that we have observed the sample (0; 0; 0), determine the posterior density for θ. [marks 10]

    c. Assuming again we have observed the sample (0; 0; 0), obtain an estimate for θ using an absolute loss function. [marks 10]

    d. Assuming that the Bayes estimate is $\hat{\theta}_B = 0.3$, estimate the probability that the next value to be observed is also equal to 0. [marks 5]

7. You observed a random sample $(x_1, x_2, \cdots, x_n)$ from a Weibull population. You know that $n = 150$, $\sum_{i=1}^{n} \ln x_i = 657.5$ and $\sum_{i=1}^{n} x_i^2 = 1491956$. At the maximum likelihood estimates of θ and τ, you got $\sum \ln\left(f\left(x_i\right)\right) = -767.10$. You also computed that, when $\tau = 2$, the maximum likelihood estimate of θ is 99.73. Use the likelihood ratio test to test the hypothesis $H_0 : \tau = 2$ against $H_1 : \tau \neq 2$. [marks 15]

8. You need an approximation for the distribution of the discounted value of the sum of 2 future payments for a given insurance contract. You know that time between consecutive payments is exponentially distributed (mean 1/3) and that payments are independent and can be modeled by a Pareto distribution ($\alpha = 3$ and $\theta = 1000$). The force of interest is 0.1. Explain how to use simulation to get this approximate distribution and to estimate the probability that this discounted value is larger than 600. [marks 20]

Solutions: Exam 17/01/2014

1.

y_j	r_j	s_j	$\frac{r_j - s_j}{r_j}$	$S_n(y_j)$
4	10	2	0.8	0.8
8	5	1	0.8	0.64
12	2	1	0.5	0.32
15	1	1	0	0

a)

$S_n(11) = 0.64$

$$\hat{\text{var}}\left(S_n(11)\right) \approx 0.64^2\left(\frac{2}{10\times 8}+\frac{1}{5\times 4}\right)=\frac{0.64^2\times 3}{40}=0.03072$$

95% Confidence Interval (asymptotic) for $S(11) \rightarrow$ (0.2965; 0.9835)

$0.64 \pm 1.96\sqrt{0.03072}$

Or a log-transformed interval can be computed

$$U=\exp\left(1.96\frac{\sqrt{0.03072}}{0.64\times\ln(0.64)}\right)=\exp(-1.20274)=0.30037$$

95% Confidence Interval (asymptotic) for $S(11) \rightarrow$ (0.2263; 0.8745)

$0.64^{1/U}; 0.64^{U}$

b)

$$_5\hat{q}_6=\frac{S_n(6)-S_n(11)}{S_n(6)}=\frac{0.8-0.64}{0.8}=0.2$$

$$\hat{\text{var}}\left(_5\hat{q}_6 \mid S(6)=S_n(6)\right)=\frac{\left(S_n(11)\right)^2}{\left(S_n(6)\right)^2}\left(\frac{1}{5\times(5-1)}\right)=\frac{0.64^2}{0.8^2\times 20}=0.032$$

95% Confidence Interval (asymptotic) for $_5q_6 \rightarrow$ (-0.151; 0.551), i.e (0; 0.551)

$0.2 \pm 1.96\sqrt{0.032}$

c)

$$L(\theta)=f(4\mid\theta)\times(1-F(5\mid\theta))^3\times f(8\mid\theta)\times(1-F(10\mid\theta))^2\times f(12\mid\theta)\times f(15\mid\theta)$$

$$=\theta^{-5}\exp\left(-\frac{2\times 4+3\times 5+8+2\times 10+12+15}{\theta}\right)=\theta^{-5}\exp\left(-\frac{78}{\theta}\right)$$

$$\ell(\theta)=-5\ln\theta-\frac{78}{\theta}$$

$$\ell'(\theta)=-\frac{5}{\theta}+\frac{78}{\theta^2} \qquad \ell'(\theta)=0 \Leftrightarrow \frac{5}{\theta}=\frac{78}{\theta^2} \Leftrightarrow \theta=\frac{78}{5}=15.6$$

$\ell''(\theta) = \dfrac{5}{\theta^2} - \dfrac{2\times 78}{\theta^3}$ **As** $\ell''(15.6) = -0.02055 < 0$ **we get** $\hat{\theta} = 15.6$ **and**

$\hat{\text{var}}\left(\hat{\theta}\right) \approx -\dfrac{1}{\ell''(\hat{\theta})} = 48.672$

95% Confidence Interval (asymptotic) for θ →(1.9260; 29.2740)

$15.6 \pm 1.96\sqrt{48.672}$

95% Confidence Interval (asymptotic) for $S(11) = e^{-11/\theta}$ →(0.0033; 0.6868)

Or the delta method can be used

$\hat{S}(11) = e^{-11/\hat{\theta}} = 0.494$

$g(\theta) = e^{-11/\theta}$ $\quad g'(\theta) = (11/\theta^2)\, e^{-11/\theta}$

$\hat{\text{var}}(\hat{S}(11)) = \hat{\text{var}}(e^{-11/\hat{\theta}}) \approx \left((11/\hat{\theta}^2)\, e^{-11/\hat{\theta}}\right)^2 \hat{\text{var}}(\hat{\theta}) \approx 0.02427$

95% Confidence Interval (asymptotic) for $S(11)$ → (0.1887; 0.7993)

$0.494 \pm 1.96\sqrt{0.02427}$

2.

The given CI (1.63; 2.55) is obtained using $\hat{H}(t_0) \pm z_{\alpha/2}\sqrt{\hat{\text{var}}(\hat{H}(t_0)}$. Then

$\hat{H}(t_0) = \dfrac{1.63 + 2.55}{2} = 2.09$ and $z_{\alpha/2}\sqrt{\hat{\text{var}}(\hat{H}(t_0))}$ =2.55-2.09=0.46.

Then $U = \exp\left(\dfrac{z_{\alpha/2}\sqrt{\hat{\text{var}}(\hat{H}(t_0))}}{\hat{H}(t_0)}\right)$ and consequently the log-transformed interval is

$\left(\hat{H}(t_0)\times U^{-1}; \hat{H}(t_0)\times U\right)$, i.e. (1.677; 2.605)

3.

y_j	$p(y_j)$	$k_{y_j}(1)$	$p(y_j)\,k_{y_j}(1)$
0.11	0.125	0.0177	0.0022
0.16	0.375	0.0328	0.0123
0.95	0.25	0.2434	0.0609
2.75	0.25	0.2868	0.0717

$\hat{f}(1) = \sum_{j=1}^{4} p(y_j)\, k_{y_j}(x) = 0.147074$

4.

Inverse Weibull $\quad F(x\,|\,\theta,\tau)=\exp\left(-\left(\frac{\theta}{x}\right)^{\tau}\right)$

Empirical percentiles:
$n=8$
Percentile 0.4: $9\times 0.4=3.6$ $\quad \tilde{\pi}_{0.4}=0.4\times 0.66+0.6\times 1.82=1.356$
Percentile 0.8: $9\times 0.8=7.2$ $\quad \tilde{\pi}_{0.8}=0.8\times 7.62+0.2\times 12.66=8.628$
We need to solve the system

$$\begin{cases} 0.4=\exp\left(-\left(\frac{\theta}{1.356}\right)^{\tau}\right) \\ 0.8=\exp\left(-\left(\frac{\theta}{8.628}\right)^{\tau}\right) \end{cases} \Leftrightarrow \begin{cases} \ln(0.4)=-\left(\frac{\theta}{1.356}\right)^{\tau} \\ \ln(0.8)=-\left(\frac{\theta}{8.628}\right)^{\tau} \end{cases} \Leftrightarrow \begin{cases} \ln(0.4)=-\left(\frac{\theta}{1.356}\right)^{\tau} \\ \frac{\ln(0.4)}{\ln(0.8)}=\left(\frac{8.628}{1.356}\right)^{\tau} \end{cases}$$

$$\Leftrightarrow \begin{cases} \ln(0.4)=-\left(\frac{\theta}{1.356}\right)^{\tau} \\ \ln\left(\frac{\ln(0.4)}{\ln(0.8)}\right)=\tau\ln\left(\frac{8.628}{1.356}\right) \end{cases} \Leftrightarrow \begin{cases} \tau=\dfrac{\ln\left(\frac{\ln(0.4)}{\ln(0.8)}\right)}{\ln\left(\frac{8.628}{1.356}\right)}=0.763328 \\ \theta=1.356\times\left(-\ln(0.4)\right)^{1/\tau}=1.209265 \end{cases}$$

5.
Losses are obtained adding the payment with the deductible, i.e

Observation	1	2	3	4	5	6
x_i	25	32	15	25	120+	150+

a) $L(\theta)=f(25\,|\,\theta)\times f(32\,|\,\theta)\times\frac{f(15\,|\,\theta)}{1-F(10\,|\,\theta)}\times\frac{f(25\,|\,\theta)}{1-F(10\,|\,\theta)}\times\frac{1-F(120\,|\,\theta)}{1-F(20\,|\,\theta)}\times\left(1-F(150\,|\,\theta)\right)$

b) Deductible of 50. A claim is reported when $X>50$.
$\hat{\Pr}(X>50)=1-F(50\,|\,\hat{\theta})$
Deductible of 50 and policy limit of 100. We want
$\hat{\Pr}(X>100\,|\,X>50)=\dfrac{1-F(100\,|\,\hat{\theta})}{1-F(50\,|\,\hat{\theta})}$

6.
$X\,|\,\theta\sim Po(\theta)$ and $\theta\sim U(0;5)$

a. $L(\theta)\propto e^{-n\theta}\,\theta^{\sum x_i}=e^{-n\theta}\,\theta^{n\bar{x}}$
$\pi(\theta)\propto 1 \qquad 0<\theta<5$
$\pi(\theta\,|\,\underline{x})\propto e^{-n\theta}\,\theta^{n\bar{x}} \qquad 0<\theta<5$ since the condition is included in the prior

b. $\pi(\theta\,|\,0,0,0)\propto e^{-3\theta}\,\theta^{0}=e^{-3\theta} \qquad 0<\theta<5$
$\pi(\theta\,|\,0,0,0)\propto c\,e^{-3\theta} \qquad 0<\theta<5$

The constant of proportionality is

$$c=\frac{1}{\int_0^5 e^{-3\theta}\,d\theta}=\frac{1}{\left(-(1/3)e^{-3\theta}\right]_0^5}=\frac{1}{(1/3)\times\left(-e^{-15}+1\right)}=\frac{3}{\left(1-e^{-15}\right)} \qquad (\approx 3)$$

Then, the posterior is

$$\pi(\theta\,|\,0,0,0)=\frac{3}{1-e^{-15}}e^{-3\theta} \qquad 0<\theta<5 \qquad (\approx 3e^{-3\theta})$$

The constraint $0<\theta<5$ is, in practical terms, ineffective .

c. Using an absolute loss function the estimate is the median of the posterior distribution.

$1-e^{-3\theta}=0.5 \Leftrightarrow 3\theta=-\ln(0.5) \Leftrightarrow \theta=0.231$, i.e. $\hat{\theta}_B=0.231$

d. $\hat{\Pr}(X=0\,|\,0,0,0)=e^{-\hat{\theta}}=0.74$

or you can use the predictive distribution, assuming that we observed the sample $(0,0,0)$,

$$\hat{\Pr}(X=0\,|\,0,0,0)=\int_0^5 \frac{e^{-\theta}\theta^0}{0!}\frac{3}{1-e^{-15}}e^{-3\theta}\,d\theta=\frac{3}{1-e^{-15}}\int_0^5 e^{-4\theta}\,d\theta=\frac{3}{1-e^{-15}}\left(\frac{e^{-4\theta}}{-4}\right]_0^5=0.75$$

7.

$H_0:\tau=2$ against $H_1:\tau\neq 2$

The test statistic is $Q=2\times\left(\ell(\hat{\theta},\hat{\tau})-\ell(\hat{\theta}^*,\tau_0)\right)\sim\chi^2_{(1)}$

Where $\ell(\hat{\theta},\hat{\tau})$ is the loglikelihood evaluated at the ML estimators and $\ell(\hat{\theta}^*,\tau_0)$ is the loglikelihood evaluated at $\tau=2$ and at $\hat{\theta}^*$ which is the optimized value of θ given that $\tau=2$.

We know that, for the observed sample, $\ell(\hat{\theta},\hat{\tau})=-767.10$ and that $\hat{\theta}^*=99.73$

As the loglikelihood function is

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

$$\ell(\hat{\theta}^*,\tau_0)=\sum_{i=1}^n\left(\ln 2+\ln(x_i)-2\ln(99.73)-(x_i/99.73)^2\right)$$
$$=150\times\ln 2+657.5-1380.74-\frac{1491956}{99.73^2}=-769.27$$

$Q_{obs}=2\times\left(-767.10+769.27\right)=4.34$; p-value=0.0372

We reject at 5%.

8.

- Define NR, the number of replicas
- For each replica, $j = 1, 2, \cdots, NR$
  - Generate 2 exponentially distributed variables t_i, time to payment 1 and time between payment 1 and payment 2. To generate each t_i: generate u_i as a Uniform(0,1) random variable and use the inverse method, i.e. compute $t_i = -\frac{\ln(1-u_i)}{3}$.
  - Generate 2 Pareto r.v., x_i, representing the value of each payment. To generate each x_i: generate v_i as a Uniform(0,1) random variable and use the inverse method, i.e. compute $x_i = \theta\left((1-v_i)^{-1/\alpha} - 1\right)$
  - Now compute the discounted value for replica j, y_j, and keep it.
    $y_j = x_1 \times e^{-0.1 t_1} + x_2 \times e^{-0.1(t_1+t_2)}$
- The NR elements of array y are used to get an approximation of the required distribution (histogram, kernel density estimation, ...)
  To get an estimate the probability just count how many values of y_j are larger than 600 and divide this number by NR