![ISEG logo]

Instituto Superior de Economia e Gestão

UNIVERSIDADE TÉCNICA DE LISBOA

MASTER IN ACTUARIAL SCIENCE

Risk Models

29/01/2013

Time allowed: 3 hours

Instructions:

1. This paper contains 9 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 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	Number of deaths
1.5	50	7
6.3	40	9
10.2	30	9
14.6	20	11

a) You know that exactly 5 policies enter the study between time 1.5 and 6.3. How many policies surrender during this time interval? [marks 5]

b) Using Nelson-Aalen approach, estimate the survival function at time 11.0 and determine a log-transformed confidence interval (95%) for $H(11)$ and get a confidence interval for $S(11)$. [marks 15]

c) Using Kaplan-Meier estimator obtain a 95% confidence interval (direct method) for $_2q_9$. Explain why you should use a conditional variance instead of an unconditional one. [marks 20]

2. A random sample of claim amounts from a portfolio of policies resulted in the following:

Claim Size	Number of Claims
(0 ; 100]	6
(100 ; 250]	14
(250 ; 500]	20
(500 ; 1000]	a
(1000 ; 5000]	2
over 5000	0
Total	n

You also know that, using the ogive, $F_n(600) = 0.832$. Calculate a and give an estimate for the probability that a claim amount is between 250 and 300.

[marks 10]

3. From a population having density function f, you are given the following sample: 2.0, 3.0, 3.0, 4.0, 4.0

Calculate the kernel density estimate of $f(4)$, using the gamma kernel with $\alpha = 2$, i.e. $k_y(x) = 4 y^{-2} x e^{-2x/y}$, $x > 0$. [marks 15]

4. You are given:

(i) Losses are uniformly distributed on $(0; \theta)$ with $\theta > 150$.

(ii) The policy limit is 150.

(iii) A sample of payments is:

14, 33, 72, 94, 120, 135, 150, 150

Estimate θ by matching the average sample payment to the expected payment per loss. [marks 15]

5. For a group of policies, you are given:

(i) Losses follow the distribution function $F(x) = 1 - \dfrac{\theta}{x}$, $x > 0$, $\theta > 0$

(ii) A sample of 20 losses resulted in the following:

Interval	Nº of losses
$x \leq 10$	9
$10 < x \leq 25$	6
$x > 25$	5

Calculate the maximum likelihood estimate of θ. [marks 15]

6. You have observed the following three loss amounts: 186 91 66
Seven other amounts are known to be **less than or equal to** 60. Losses follow an inverse exponential with distribution function $F(x) = e^{-\theta/x}$, $x > 0$, $\theta > 0$.
Let $p = \Pr(X \leq 60)$.

a) Show that the maximum likelihood estimate for θ is 20.245 and show that the asymptotic standard error of $\hat{\theta}$ is 11.688. [marks 20]

b) Obtain the maximum likelihood estimate for p and, using the delta method, calculate the standard error of $\hat{p}$. [marks 15]

c) Using a non-parametric framework, i.e. without assuming any distribution for the claim amounts, compute a 95% confidence interval for p. [marks 15]

7. Assume that we observed $x = 4$ from a Pareto distributed variable with $\alpha = 3$ and unknown parameter θ. Using a Bayesian framework we assume that our prior for θ is given by $\pi(\theta) = 8\theta^{-3}$, $\theta > 2$.

a) Obtain the posterior distribution of θ and calculate the posterior probability that θ is greater than 3. [marks 15]

b) Write the expression needed to obtain the predictive distribution of a new observation (you do not need to solve any integral, just write it) [marks 5]

8. You observed the following random sample (29; 64; 90; 135; 182). Test, using the Kolmogorov-Smirnov test, if it is acceptable (α =0.05) to consider that the corresponding population is exponentially distributed with mean 100. [marks 15]

9. Assume that we observed a random sample with 20 observations of the random variable X, price of an apartment with 150 m2 in a given country. Explain how to use bootstrap to obtain a 95% confidence interval for the expected value of such apartments. [marks 20]

Blank page

Solutions – 29/01/2013

1.

a) Using Loss Models notation, $r_2 = r_1 + d_1 - x_1 - u_1 \Leftrightarrow 40 = 50 + 5 - 7 - u_1 \Leftrightarrow u_1 = 8$

b) $\hat{H}(11) = \sum_{i:t_i \le 11} \dfrac{s_i}{r_i} = \dfrac{7}{50} + \dfrac{9}{40} + \dfrac{9}{30} = 0.665$

$\hat{S}(11) = e^{-\hat{H}(11)} = 0.5143$

$\text{vâr}\left(\hat{H}(11)\right) = \sum_{i:t_i \le 11} \dfrac{s_i}{r_i^2} = 0.0028 + 0.0056 + 0.01 = 0.0184$

$U = \exp\left(\dfrac{1.96 \times \sqrt{0.0184}}{0.665}\right) = 1.4919$

95% CI for $H(11)$: $\left(U^{-1}\hat{H}(11); U\hat{H}(11)\right)$, i.e (0.4457; 0.9921)

95% CI for $S(11)$: (0.3708 ; 0.6404)

c) $_2q_9 = \Pr\left(T \le 11 | T > 9\right) = \dfrac{S(9) - S(11)}{S(9)}$

$S_n(9) = \dfrac{43}{50} \times \dfrac{31}{40} = 0.6665$; $\quad S_n(11) = S_n(9) \times \dfrac{21}{30} = 0.4666$

$_2\hat{q}_9 = \dfrac{0.6665 - 0.4666}{0.6665} = 0.3$

$\text{vâr}\left(_2\hat{q}_9 | S(9) = S_n(9)\right) = \dfrac{S_n(11)^2}{S_n(9)^2} \times \dfrac{s_3}{r_3(r_3 - s_3)} = \left(\dfrac{0.4666}{0.6665}\right)^2 \times \dfrac{9}{30 \times 21} = 0.007$

95% CI for $_2q_9 \rightarrow {}_2\hat{q}_9 \pm z_{\alpha/2}\sqrt{\text{vâr}\left(_2\hat{q}_9 | S(9) = S_n(9)\right)}$, i.e (0.1360; 0.4640)

We should use a conditional variance because the variance of $_2\hat{q}_9$ does not exist as $S_n(9)$ (estimator) can be 0 with a positive probability.

2.
We know that
$n = 42 + a$;
$F_n(500) = 40/n$; $\quad F_n(1000) = (40 + a)/n$;
$F_n(600) = \dfrac{1000 \times F_n(500) - 500 \times F_n(1000)}{(1000 - 500)} + \dfrac{F_n(1000) - F_n(500)}{(1000 - 500)} \times 600$

Then we must solve the equation
$0.832 = \dfrac{1000 \times 40 - 500 \times (40 + a)}{500 \times (42 + a)} + \dfrac{(40 + a) - 40}{500 \times (42 + a)} \times 600$

$\Leftrightarrow 17472 + 416a = 20000 - 500a + 600a \Leftrightarrow 316a = 2528 \Leftrightarrow a = 8$

Then $n = 50$; $F_n(250) = 20/50 = 0.4$; $F_n(500) = 40/50 = 0.8$

$F_n(300) = \dfrac{500 \times 0.4 - 250 \times 0.8}{250} + \dfrac{0.4 \times 300}{250} = 0.48$

The estimate is $0.48 - 0.40 = 0.08$

3.

y_j	$p(y_j)$	$k_{y_j}(4)$
2.0	0.2	0.0733
3.0	0.4	0.1235
4.0	0.4	0.1353

$$\hat{f}(4) = \sum_{j=1}^{3} p(y_j) \times k_{y_j}(4) = 0.1182$$

4.

The expected payment per loss is given by

$$E(X \wedge 150) = \int_0^{150} x \frac{1}{\theta} dx + \int_{150}^{\theta} 150 \frac{1}{\theta} dx = \frac{\left(x^2/2\right]_0^{150}}{\theta} + \frac{150}{\theta}(\theta - 150)$$

$$= \frac{150^2/2}{\theta} + 150 - \frac{150^2}{\theta} = 150 - \frac{150^2}{2\theta}$$

Then we must solve

$$\bar{x} = 150 - \frac{150^2}{2\theta} \Leftrightarrow 96 = 150 - \frac{150^2}{2\theta} \Leftrightarrow \frac{150^2}{2\theta} = 54 \Leftrightarrow \theta = \frac{150^2}{108} = 208.3$$

Our estimate will be $\tilde{\theta} = 208.3$

5.

$$L(\theta) = \left(F(10 \mid \theta)\right)^9 \left(F(25 \mid \theta) - F(10 \mid \theta)\right)^6 \left(1 - F(25 \mid \theta)\right)^5$$

$$= \left(1 - \frac{\theta}{10}\right)^9 \left(\theta\left(\frac{1}{10} - \frac{1}{25}\right)\right)^6 \left(\frac{\theta}{25}\right)^5$$

$$\ell(\theta) = 9\ln(10 - \theta) - 9\ln(10) + 6\ln(\theta) + 6\ln\left(\frac{1}{10} - \frac{1}{25}\right) + 5\ln(\theta) - 5\ln(25)$$

$$\ell'(\theta) = -\frac{9}{10 - \theta} + \frac{11}{\theta}$$

$$\ell'(\theta) = 0 \Leftrightarrow \frac{9}{10 - \theta} = \frac{11}{\theta} \Leftrightarrow \frac{10 - \theta}{\theta} = \frac{9}{11} \Leftrightarrow \theta = 5.5$$

$$\ell''(\theta) = -9 \times (-1) \times (10 - \theta)^{-2} \times (-1) - 11 \times \theta^{-2} = -9 \times (10 - \theta)^{-2} - 11 \times \theta^{-2} < 0$$

Then the maximum likelihood estimate is $\hat{\theta} = 5.5$

6.

a) $F(x \mid \theta) = e^{-\theta/x}$ $\qquad f(x \mid \theta) = \theta x^{-2} e^{-\theta/x}$ $\qquad p = F(60 \mid \theta) = e^{-\theta/60}$

$$\ell(\theta) = \ln f(186 \mid \theta) + \ln f(91 \mid \theta) + \ln f(66 \mid \theta) + 7 \times \ln F(60 \mid \theta)$$

$$\ell(\theta) = \ln\left(\theta 186^{-2} e^{-\theta/186}\right) + \ln\left(\theta 91^{-2} e^{-\theta/91}\right) + \ln\left(\theta 66^{-2} e^{-\theta/66}\right) + 7 \times \ln\left(e^{-\theta/60}\right)$$

$$= 3\ln\theta - 2(\ln 186 + \ln 91 + \ln 66) - \theta\left(\frac{1}{186} + \frac{1}{91} + \frac{1}{66} + \frac{7}{60}\right)$$

$$\ell'(\theta) = \frac{3}{\theta} - \left(\frac{1}{186} + \frac{1}{91} + \frac{1}{66} + \frac{7}{60}\right)$$

$$\ell'(\theta) = 0 \Leftrightarrow \frac{3}{\theta} = \left(\frac{1}{186} + \frac{1}{91} + \frac{1}{66} + \frac{7}{60}\right) \Leftrightarrow \theta = 20.24$$

$$\ell''(\theta) = -\frac{3}{\theta^2} < 0$$

Then the ML estimate is $\hat{\theta} = 20.245$ and the estimate of the (asymptotic)

variance of the ML estimator is $\text{vâr}(\hat{\theta}) \approx -\frac{1}{\ell''(\hat{\theta})} = \frac{\hat{\theta}^2}{3} = \frac{20.245^2}{3} = 136.62$

The asymptotic standard error is then $\sqrt{\text{vâr}(\hat{\theta})} = 20.245/\sqrt{3} = 11.688$

b) Remind that $p = g(\theta) = e^{-\theta/60}$ then $\hat{p} = e^{-\hat{\theta}/60} = 0.7137$

$$g'(\theta) = -\frac{1}{60}e^{-\theta/60}$$

$$\text{var}(\hat{p}) = \text{var}\left(e^{-\hat{\theta}/60}\right) \approx \frac{1}{3600}e^{-2\hat{\theta}/60}\,\text{var}(\hat{\theta})$$

$$\text{vâr}(\hat{p}) \approx \frac{136.62}{3600}e^{-2\times20.245} = 0.0193 \text{ and } \sqrt{\text{vâr}(\hat{p})} \approx 0.1389$$

c) Using a non-parametric framework one can use $\tilde{p} = \frac{\#\{X_i \le 60\}}{10} = \frac{N}{10}$

where N is the number of observation, in a random sample with size 10, that are less than or equal to 60.

The estimate is $\tilde{p} = \frac{7}{10} = 0.7$

As it is well-known, $N \sim b(10; p)$ where $p = \Pr(X \le 60)$

$$\text{vâr}(\tilde{p}) = \text{vâr}\left(\frac{N}{10}\right) = \frac{10 \times \tilde{p} \times (1-\tilde{p})}{10^2} = 0.021$$

95% confidence interval for $p \rightarrow \tilde{p} \pm 1.96\sqrt{0.021}$, i.e. (0.416; 0.984)

7.

a) $L(\theta \mid x = 4) \propto \theta^3 (4+\theta)^{-4}$

$\pi(\theta) \propto \theta^{-3}, \; \theta > 2$

$\pi(\theta \mid x = 4) \propto \theta^3 (4+\theta)^{-4} \theta^{-3} = (4+\theta)^{-4}, \; \theta > 2$

As the integral of the posterior between 2 and infinity has to be equal to 1,

$$k = \int_2^\infty (4+\theta)^{-4}\, d\theta = \left(\frac{(4+\theta)^{-3}}{-3}\right]_2^\infty = \frac{0 - 6^{-3}}{-3} = \frac{1}{648} \text{ . Then}$$

$\pi(\theta \mid x = 4) = 648(4+\theta)^{-4}, \; \theta > 2$.

$$\hat{\Pr}(\theta > 3 \mid x = 4) = \int_3^\theta 648(4+\theta)^{-4}\, d\theta = 648\left(\frac{(4+\theta)^{-3}}{-3}\right]_3^\infty = 648\frac{7^{-3}}{3} = 0.6297$$

b) $f_{Y|X}(y \mid 3) = \int_2^\infty \frac{3\theta^3}{(y+\theta)^4} \times \frac{648}{(4+\theta)^4}\, d\theta = \int_2^\infty \frac{1944\,\theta^3}{(y+\theta)^4 \times (4+\theta)^8}\, d\theta$

8.

i	x_i	F(x_i)	F_n(x_i) =i/n	F_n(x_i-) =(i-1)/n	D_i
1	29	0.251736432	0.2	0	0.251736432
2	64	0.472707576	0.4	0.2	0.272707576
3	90	0.59343034	0.6	0.4	0.19343034
4	135	0.740759739	0.8	0.6	0.140759739
5	182	0.837974249	1	0.8	0.162025751

D= 0.272707576

Critical asymptotic value $(\alpha = 0.05) : 1.36/\sqrt{5} = 0.6082$

We do not reject the hypothesis that the observations are exponentially distributed with mean 100.

9.

First we need to generate the bootstrap sampling distribution of the sample mean.

- Determine the number of bootstrap replicas, NR, to be used
- For replica i, $i = 1, 2, \cdots, NR$
  - Generate a bootstrap sample by resampling with replacement from the original sample. Each bootstrap sample has the size of the original sample.
  - Calculate the average of the bootstrap sample and keep this value as entry i of the array y.
- Array y contains our bootstrap approximation to the sampling distribution of the sample mean.

Now we can use the percentile method to get a 95% confidence interval for the population mean by calculating empirical percentiles 0.025 and 0.975. These percentiles define the confidence interval.