



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

MASTER IN ACTUARIAL SCIENCE

Risk Models

16/01/2018

Time allowed: 3 hours

Instructions:

1. This paper contains **9** questions and comprises **5** 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.

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

Time(t)	Risk set at time t	Failures at time t
2.3	30	5
10.4	25	9
12.5	32	6
15.3	27	5
18.1	15	4

- a) [10] Estimate the survival function at time 13 and give a log-transformed confidence interval (95%) for $S(13)$.
- b) [5] Assuming that 7 policies enter the portfolio between time 2.3 and time 10.4 how many surrenders have been observed during this time interval?
- c) [10] Estimate ${}_{10}q_3$ and also estimate a 95% conditional confidence interval for ${}_{10}q_3$. What is the conditioning behind the variance used to compute the confidence interval?

2. [10] For a survival study, you are given:

- (i) Deaths occurred at times $y_1, y_2, \dots, y_8, \dots$.
- (ii) The Nelson-Aalen estimates of the cumulative hazard function at y_6, y_7 and y_8 are respectively $\hat{H}(y_6) = 1.99$, $\hat{H}(y_7) = 2.59$ and $\hat{H}(y_8) = 3.09$
- (iii) The estimated variances of the estimators in (ii) are $\hat{\text{var}}(\hat{H}(y_6)) = 0.1101$ and $\hat{\text{var}}(\hat{H}(y_7)) = 0.2301$
- (iv) There are no entries and no surrenders in the portfolio between times y_7 and y_8 .

Determine the number of deaths at y_8 .

3. [15] You study five lives to estimate the time from the onset of a disease to death. The times to death are: 2; 3; 3; 7; 7. Using a Weibull ($\theta = \frac{2y}{\sqrt{\pi}}$ and $\tau = 2$) kernel, estimate the **distribution function** at 2.5.

4. [10] You are given:

- (i) Losses are uniformly distributed on $(0, \theta)$ with $\theta > 200$
- (ii) The policy limit is 200
- (iii) A sample of payment is: 14, 83, 112, 154, 170, 185, 200, 200, 200

Estimate θ by matching the average sample payment to the expected payment per loss.

5. A random sample with size $n = 50$ has been observed from a population with density function $f(x|\theta) = \frac{x}{\theta} \exp\left(-\frac{x^2}{2\theta}\right)$, $x > 0$, $\theta > 0$. We know that

$$\bar{x} = 114.57 \text{ and } s^2 = 3148.49.$$

- a. **[10]** Determine the maximum likelihood **estimator** for θ and the maximum likelihood **estimate**.

If you are unable to solve this question you can use, if necessary, $\hat{\theta} = 8100$ and $\ell''(\hat{\theta}) = -50/8100^2$ (which are not the correct figures) to answer the next questions

- b. **[10]** Using the asymptotic distribution of the maximum likelihood estimator determine a 95% confidence interval for θ ? Comment about the use of the asymptotic distribution for this case.
- c. **[15]** Using the asymptotic distribution of the maximum likelihood estimator obtain a 95% confidence interval for $P(X > 100)$.
- d. **[10]** Is the maximum likelihood estimator unbiased (you can take advantage that $Y = X^2$ follows an exponential distribution with mean 2θ)?

6. You are given the following 6 losses before the deductible is applied, in case any deductible is in force (" $>$ " means that the loss exceeds the reported value)

Loss	Deductible
750	200
200	None
300	None
>10000	None
400	300
>5000	500

Assume that the density function of the losses is given by $f(x|\alpha, \theta)$ and the survival function by $S(x|\alpha, \theta)$.

- a. **[10]** Write the likelihood (using the density and the survival functions) function that needs to be maximized to estimate the unknown parameters. Also write the log-likelihood.
- b. **[15]** Now assume that the losses are Pareto distributed with $\theta = 10000$. Obtain a maximum likelihood estimate for α and also obtain a maximum likelihood estimate for the probability that a loss is reported if a deductible of 1000 is in force.

7. Assume that, given θ , X follows an exponential distribution with mean $1/\theta$ and that the random sample (0.3, 0.5, 1.2, 2.0) was observed. From a Bayesian point of view we define that the prior for θ is an exponential distribution with mean 0.25.
- [10]** Show that the posterior distribution is a gamma and obtain the values of the parameters.
 - [10]** Obtain the Bayes estimate using a quadratic loss function and compute the posterior probability that θ is greater than 0.5.
(Hint: $X \sim G(a, b) \Rightarrow \frac{2X}{b} \sim \chi^2_{(2a)}$)
 - [10]** Show that, given the sample, the predictive distribution of y is a Pareto distribution with parameters 5 and 8.

8. You are given:

- A random sample of 100 losses from a Gamma distribution
- To get maximum likelihood estimates of α and θ an actuary gets the following output from R. Hint: Remember that using the R function **nlm** we are minimizing minus the log-likelihood

```
> minusloglikgamma=function(param){
+ alpha=param[1]; theta=param[2]
+ ll=dgamma(x,shape=alpha,scale=theta,log=T)
+ return(-sum(ll))
+ }
>
> param.start=c(3,150)
>
> out=nlm(minusloglikgamma,param.start,hessian=T)
> out
$minimum
[1] 665.4673
$estimate
[1] 3.179853 117.994595
$gradient
[1] -5.934870e-06 -7.515237e-08
$hessian
      [,1] [,2]
[1,] 36.8973349 0.84745408
[2,] 0.8474541 0.02283019
$code
[1] 1
$iterations
[1] 20
```

```

> V=solve(out$hessian)
> V
      [,1] [,2]
[1,] 0.1838256 -6.823583
[2,] -6.8235829 297.092230

```

- a) **[15]** Determine a maximum likelihood estimate for $\tau = \alpha\theta$ and using the asymptotic distribution of the maximum likelihood estimators define a 95% confidence interval for θ .
- b) **[10]** Now the actuary needs to test $H_0 : \alpha = 3$ against $H_1 : \alpha \neq 3$ using the likelihood ratio test. To do so, she uses R again and assuming that $\alpha = 3$ she gets the following output

```

> minusloglikgamma1=function(theta){
+ alpha=3;
+ ll=dgamma(x,shape=alpha,scale=theta,log=T)
+ return(-sum(ll))
+ }
>
> theta.start=300
>
> out1=nlm(minusloglikgamma1,theta.start)
> out1
$minimum
[1] 665.5591
$estimate
[1] 125.0684
$gradient
[1] -2.662454e-06
$code
[1] 1
$iterations
[1] 91

```

Complete the test and conclude.

9. **[15]** An actuary observed a sample with size $n = 15$ from a given population. Using a non-parametric bootstrap approach he wants to analyze the behavior of the interquartile range (IQR=Q3-Q1 where Q1 is the 1st quartile and Q3 the 3rd quartile) of the sample as an estimate of the IQR of the population. Explain how to proceed and how to evaluate if this estimator is unbiased.

SOLUTION

1.

j	y_j	r_j	s_j	$\frac{r_j - s_j}{r_j}$	$\prod \left(\frac{r_j - s_j}{r_j} \right)$
1	2.3	5	30	25/30	0.8333
2	10.4	9	25	16/25	0.5333
3	12.5	6	32	26/32	0.4333
4	15.3	5	27	22/27	0.3531
5	18.1	4	15	11/15	0.2589

a.

$$S_n(13) = 0.4333$$

$$\hat{\text{var}} S_n(13) \approx 0.4333^2 \left(\frac{5}{30 \times 25} + \frac{9}{25 \times 16} + \frac{6}{32 \times 26} \right) = 0.006830$$

$$U = \exp \left(1.96 \frac{\sqrt{0.006737}}{0.4333 \times \ln(0.4333)} \right) = 0.6395$$

The 95% log-transformed confidence interval is then given by

$$\left(S_n(13)^{1/U}; S_n(13)^U \right) \text{ i.e. } (0.2704; 0.5857)$$

b.

Let ns be the number of surrenders between times $y_1 = 2.3$ and $y_2 = 10.4$. As we know that $r_2 = r_1 - 5 + 7 - ns$ we get $ns = r_1 - 5 + 7 - r_2 = 30 - 5 + 7 - 25 = 7$.

c.

$${}_{10}q_3 = \frac{S(3) - S(13)}{S(3)} \text{ and then } {}_{10}\hat{q}_3 = \frac{S_n(3) - S_n(13)}{S_n(3)} = \frac{0.8333 - 0.4333}{0.8333} = 0.48$$

$$\hat{\text{var}}({}_{10}\hat{q}_3 | S(3) = S_n(3)) = \left(\frac{0.4333}{0.8333} \right)^2 \left(\frac{9}{25 \times 16} + \frac{6}{26 \times 32} \right) = 0.008033$$

The 95% conditional confidence interval is then (0.3271; 0.6329)

The condition behind the conditional variance is that the true value of the survival function at time 3 is equal to the estimated one.

2.

As $\hat{H}(y_6) = 1.99$ and $\hat{H}(y_7) = 2.59$, we know that $\frac{s_7}{r_7} = 2.59 - 1.99 = 0.6$.

Similarly, $\frac{s_8}{r_8} = \hat{H}(y_8) - \hat{H}(y_7) = 3.09 - 2.59 = 0.5$

As $\hat{\text{var}}(\hat{H}(y_6)) = 0.1101$ and $\hat{\text{var}}(\hat{H}(y_7)) = 0.2301$ we get

$$\frac{s_7}{r_7^2} = 0.2301 - 0.1101 = 0.12$$

As there are no entries and no surrenders in the portfolio between times y_7 and y_8 we have $r_8 = r_7 - s_7$

As $\frac{s_7}{r_7^2} = 0.12$ and $\frac{s_7}{r_7} = 0.6$ we get $r_7 = \frac{0.6}{0.12} = 5$ and $s_7 = r_7 \times 0.6 = 5 \times 0.6 = 3$

Consequently $r_8 = 5 - 3 = 2$ and then $s_8 = 1$.

3.

$$K_y(x) = 1 - e^{-(x/\theta)^2} = 1 - \exp\left(-\left(\frac{x\sqrt{\pi}}{2y}\right)^2\right) = 1 - \exp\left(-\left(\frac{x^2\pi}{4y^2}\right)\right) = 1 - e^{-0.7854x^2/y^2}$$

$$K_y(2.5) = 1 - e^{-0.7854 \times 6.25/y^2}$$

$$\begin{aligned}\hat{F}(2.5) &= \sum_{j=1}^3 p(y_j) \times K_{y_j}(2.5) \\ &= 0.2 \times \left(1 - e^{-0.7854 \times 6.25/4}\right) + 0.4 \times \left(1 - e^{-0.7854 \times 6.25/9}\right) + 0.4 \times \left(1 - e^{-0.7854 \times 6.25/49}\right) \\ &= 0.2 \times 0.7069 + 0.4 \times 0.4204 + 0.4 \times 0.0953 = 0.1414 + 0.1682 + 0.0381 \\ &= 0.3477\end{aligned}$$

4.

$$E(X \wedge 200) = \int_0^{200} x \frac{1}{\theta} dx + \int_{200}^{\theta} 200 \frac{1}{\theta} dx = \frac{200^2}{2\theta} + \frac{200(\theta - 200)}{\theta} = 200 - \frac{200^2}{2\theta}$$

The average payment is 146.44 and then one should solve the equation

$$200 - \frac{200^2}{2\theta} = 146.44 \Leftrightarrow 53.56 = \frac{20000}{\theta} \Leftrightarrow \theta = \frac{20000}{53.56} = 373.44 \text{ and then } \tilde{\theta} = 373.44$$

5.

a.

$$\ell(\theta) = \sum_{i=1}^n \ln\left(\frac{x_i}{\theta} \exp\left(-\frac{x_i^2}{2\theta}\right)\right) = \sum_{i=1}^n \left(\ln(x_i) - \ln(\theta) - \frac{x_i^2}{2\theta}\right)$$

$$\ell'(\theta) = \sum_{i=1}^n \left(-\frac{1}{\theta} + \frac{x_i^2}{2\theta^2}\right) \text{ then } \ell'(\theta) = 0 \Leftrightarrow \frac{n}{\theta} = \frac{\sum_{i=1}^n x_i^2}{2\theta^2} \Leftrightarrow \theta = \frac{\sum_{i=1}^n x_i^2}{2n}$$

$$\ell''(\theta) = \sum_{i=1}^n \left(\frac{1}{\theta^2} - \frac{x_i^2}{\theta^3}\right) = \frac{n}{\theta^2} - \frac{\sum_{i=1}^n x_i^2}{\theta^3} = \frac{n}{\theta^2} \left(1 - \frac{\sum_{i=1}^n x_i^2}{n\theta}\right)$$

So $\ell''(\theta = \frac{\sum_{i=1}^n x_i^2}{2n}) = \frac{n}{\theta^2} \left(1 - \frac{\sum_{i=1}^n x_i^2}{n \frac{\sum_{i=1}^n x_i^2}{2n}} \right) = -\frac{n}{\theta^2} < 0$ and the mle estimator is

$$\hat{\theta} = \frac{\sum_{i=1}^n X_i^2}{2n}$$

To get the m.l. estimate we must obtain $\sum_{i=1}^n x_i^2$.

As $s^2 = 3148.49$ we get $m^2 = \frac{n-1}{n} s^2 = 3085.52$ and as $m^2 = 3085.52 = \frac{\sum_{i=1}^{50} x_i^2}{50} - \bar{x}^2$

we finally get $\sum_{i=1}^{50} x_i^2 = 50(3085.52 + \bar{x}^2) = 50 \times 16211.81 = 810590.3$.

Then the mle estimate is $\hat{\theta} = 810590.3/100 = 8105.903$

b.

As $\ell''(\hat{\theta}) = -\frac{n}{\hat{\theta}^2} = -\frac{50}{8105.903^2}$ we get $\text{var}(\hat{\theta}) = -1/\ell''(\hat{\theta}) = \frac{8105.903^2}{50} = 1314113$

The 95% CI is then given by $\hat{\theta} \pm 1.96\sqrt{\text{var}(\hat{\theta})}$, i.e. (5859.061; 10351.74)

Comment about using asymptotic results when the sample size is only $n = 50$

c.

$$P(X > 100) = \int_{100}^{\infty} \frac{x}{\theta} \exp\left(-\frac{x^2}{2\theta}\right) dx = \left(-\exp\left(-\frac{x^2}{2\theta}\right) \right) \Big|_{100}^{\infty} = \exp\left(-\frac{100^2}{2\theta}\right)$$

$$\hat{P}(X > 100) = \exp\left(-\frac{100^2}{2\hat{\theta}}\right) = 0.53965$$

$$g(\theta) = \exp\left(-\frac{100^2}{2\theta}\right); \quad g'(\theta) = \frac{100^2}{2\theta^2} \exp\left(-\frac{100^2}{2\theta}\right);$$

$$g'(\hat{\theta}) = \frac{100^2}{2\hat{\theta}^2} \exp\left(-\frac{100^2}{2\hat{\theta}}\right) = 0.00004107$$

$$\text{var}(\hat{P}(X > 100)) = (g'(\hat{\theta}))^2 \text{var}(\hat{\theta}) = 0.00004107^2 \times 1314113 = 0.02216$$

The 95% CI is then given by $\hat{P}(X > 100) \pm 1.96\sqrt{\text{var}(\hat{P}(X > 100))}$, i.e.

$0.53965 \pm 1.96 \times 0.04708$ and then we get (0.4474, 0.6319)

d.

$E(\hat{\theta}) = E\left(\frac{\sum_{i=1}^n X_i^2}{2n}\right) = \frac{\sum_{i=1}^n E(X_i^2)}{2n} = \frac{\sum_{i=1}^n E(Y)}{2n} = \frac{n(2\theta)}{2n} = \theta$ then $\hat{\theta}$ is an unbiased estimator of θ .

6.

a.

$$L(\alpha, \theta) = \frac{f(750|\alpha, \theta)}{S(200|\alpha, \theta)} \times f(200|\alpha, \theta) \times f(300|\alpha, \theta) \times S(10000|\alpha, \theta) \times \frac{f(400|\alpha, \theta)}{S(300|\alpha, \theta)} \times \frac{S(5000|\alpha, \theta)}{S(500|\alpha, \theta)}$$
$$\ell(\alpha, \theta) = \ln f(750|\alpha, \theta) - \ln S(200|\alpha, \theta) + \ln f(200|\alpha, \theta) + \ln f(300|\alpha, \theta) + \ln S(10000|\alpha, \theta) + \ln f(400|\alpha, \theta) - \ln S(300|\alpha, \theta) + \ln S(5000|\alpha, \theta) - \ln S(500|\alpha, \theta)$$

b.

$$\theta = 10000$$

$$\ln f(x|\alpha) = \ln \alpha + \alpha \ln 10000 - (\alpha + 1) \ln(x + \theta)$$

$$\ln S(x|\alpha) = \alpha \ln 10000 - \alpha \ln(x + \theta)$$

$$\begin{aligned} \ell(\alpha) &= \ln \alpha + \alpha \ln 10000 - (\alpha + 1) \ln(10750) - \alpha \ln 10000 + \alpha \ln(10200) \\ &\quad + \ln \alpha + \alpha \ln 10000 - (\alpha + 1) \ln(10200) + \ln \alpha + \alpha \ln 10000 - (\alpha + 1) \ln(10300) \\ &\quad + \alpha \ln 10000 - \alpha \ln(20000) + \ln \alpha + \alpha \ln 10000 - (\alpha + 1) \ln(10400) - \alpha \ln 10000 + \alpha \ln(10300) \\ &\quad + \alpha \ln 10000 - \alpha \ln(15000) - \alpha \ln 10000 + \alpha \ln(10500) \end{aligned}$$

$$\begin{aligned} \ell(\alpha) &= 4 \ln \alpha + 3\alpha \ln 10000 - \alpha \ln(10750) - \alpha \ln(10400) - \alpha \ln(15000) + \alpha \ln(10500) - \alpha \ln(20000) \\ &\quad - \ln(10200) - \ln(10300) - \ln(10750) - \ln(10400) \end{aligned}$$

$$\begin{aligned} \ell'(\alpha) &= \frac{4}{\alpha} + 3 \ln 10000 - \ln(10750) - \ln(10400) - \ln(15000) + \ln(10500) - \ln(20000) \\ &= \frac{4}{\alpha} - 1.16136 \end{aligned}$$

$$\ell''(\alpha) = -\frac{4}{\alpha^2} < 0$$

Then the maximum likelihood estimate of α is the solution of $\ell'(\alpha) = 0$, i.e.

$$\ell'(\alpha) = 0 \Leftrightarrow \frac{4}{\alpha} - 1.16136 = 0 \Leftrightarrow \hat{\alpha} = \frac{4}{1.16136} = 3.444$$

The required probability is

$$\hat{P}(X > 1000) = \left(\frac{10000}{11000}\right)^{3.444} = 0.7202$$

7.

a) $f(x|\theta) = \theta e^{-\theta x}, x > 0, \theta > 0$ $L(\theta|x) = \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right) = \theta^n e^{-4\theta}$

$\pi(\theta) = (1/0.25) e^{-\theta/0.25} = 4 e^{-4\theta}, \theta > 0$

$\pi(\theta|x) \propto e^{-4\theta} \times \theta^4 e^{-4\theta} = \theta^4 e^{-8\theta}$ i.e. the posterior is a gamma distribution with parameters 5 and $1/8=0.125$

b) Bayes estimate using a quadratic loss function \rightarrow expected value of the posterior, i.e. $\theta_{quad}^B = 5 \times 0.125 = 0.625$

$\theta | \mathbf{x} \sim G(5; 0.125)$ then $\frac{2\theta}{0.125} | \mathbf{x} \sim \chi_{(10)}^2$

$\Pr(\theta > 0.5 | \mathbf{x}) = \Pr(16\theta > 8 | \mathbf{x}) = 1 - 0.3712 = 0.6288$

c)

$$\begin{aligned} f_{Y|\mathbf{x}}(y|\mathbf{x}) &= \int_0^\infty \theta e^{-\theta y} \frac{\theta^4 e^{-8\theta}}{(1/8)^5 \Gamma(5)} d\theta = \frac{8^5}{\Gamma(5)} \int_0^\infty \theta^5 e^{-(y+8)\theta} d\theta \text{ core of a gamma with par. 6 and } \frac{1}{y+8} \\ &= \frac{8^5}{\Gamma(5)} \frac{\Gamma(6)}{(y+8)^6} \int_0^\infty \frac{(y+8)^6}{\Gamma(6)} \theta^4 e^{-(y+25)\theta} d\theta \\ &= \frac{5 \times 8^5}{(y+8)^6} \quad y > 0 \end{aligned}$$

which is the density of a Pareto with parameters $\alpha^* = 5$ and $\theta^* = 8$

8.

a.

As $\hat{\alpha} = 3.179853$ and $\hat{\theta} = 117.994595$, using the invariance property of the m.l.e., one get $\hat{\tau} = \hat{\alpha} \times \hat{\theta} = 375.2055$

The 95% CI for θ will be given by $\hat{\theta} \pm 1.96 \sqrt{\text{var}(\hat{\theta})}$, i.e. (84.21; 151.78) as

$\text{var}(\hat{\theta}) = 297.09223$

b.

$H_0 : \alpha = 3$ against $H_1 : \alpha \neq 3$

The asymptotic test statistic is $Q = -2 \left(\ell(\alpha = 3, \hat{\theta}) - \ell(\hat{\alpha}, \hat{\theta}) \right) \overset{\circ}{\sim} \chi_{(1)}^2$

$Q_{obs} = -2(-665.5591 - (-665.4673)) = 0.1837$

p-value=0.6682 so we do not reject the null hypothesis, i.e. we do not reject that $\alpha = 3$.

9.

To obtain the sampling distribution of the IQR of the sample: (The first step is only mandatory to analyze the unbiasedness)

- i. Calculate the IQR of the observed sample, IQR^* . We sort the sample – $(x_{(1)}, x_{(2)}, \dots, x_{(15)})$ is the sorted sample and use $\tilde{Q}_1 = x_{(4)}$, $\tilde{Q}_3 = x_{(12)}$ and consequently $IQR^* = x_{(12)} - x_{(4)}$
- ii. Define B as the number of bootstrap samples to be generated.
- iii. For each bootstrap generation, $j = 1, 2, \dots, B$ do:
 - a. Resample (with replacement) from the original sample; Generate 15 pseudo random numbers, $u_i, i = 1, 2, \dots, 15$, and replace each u_i by x_k such that $\frac{k-1}{15} < u_i \leq \frac{k}{15}, k = 1, 2, \dots, 15$.
 - b. Calculate the IQR of the bootstrap sample, IQR_j , using the same procedure as before (step i);
- iv. Now, using the B bootstrap observations we can approximate the sampling distribution by means of a histogram or using the empirical cumulative distribution function.

To evaluate if the estimator is unbiased, compute the bootstrap estimate of the bias as

$$\hat{bias} = \frac{\sum_{j=1}^B IQR_j}{B} - IQR^*.$$

If the estimator is unbiased we should obtain an estimate close to zero.