MASTERS IN ACTUARIAL SCIENCE

Risk Models

18/01/2021

1st part of the exam

Time allowed: 2 hours

Instructions:

1. This paper contains **6** 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: 140.
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. An insurance company employs agents on a commission basis. It claims that in their first-year agents will earn a mean commission of at least $40,000. In a random sample of nine agents we observed, for commission in the first year, $\sum_{i=1}^{9} x_i = 333$ and $\sum_{i=1}^{9}(x_i - \bar{x})^2 = 312$ where x_i is measured in thousands of dollars. We assume that the population distribution is normal.
    a. **[5]** Test, at the 5% level, the null hypothesis that the population mean is at least 40 thousand US$.
    b. **[5]** Test, at the 5% level, the hypothesis that the population standard deviation is no more than 6 thousand US$
    c. **[5]** Compute a 95% confidence interval for the population mean.

2. You are given the following information from a random sample of claims (in a given monetary unit)

Claim size	$(0\,;1]$	$(1\,;2.5]$	$(2.5\,;10]$	$(10\,;20]$	$(20\,;50]$	$(50;\ \infty)$	Total
Nº of claims	5	15	25	a	3	0	n

You also know that $F_n(12) = 16/21$.
    a. **[10]** Calculate a and n. Note: If and only if you are unable to answer this question you can use $n = 65$ (which is not the correct answer) in the next questions.
    b. **[10]** Estimate the survival function at 10 and compute a 95% asymptotic confidence interval for $S(10)$.
    c. **[5]** Estimate the probability that a claim is greater than 10 given that it is greater than 1.

3. From a population having density function f you are given the following sample (2,3,5,3,7,5).
    a. **[15]** Calculate the kernel density estimate of $f(6)$ using the kernel $k_y(x) = \left(\frac{3}{32}\right)(4 - (x - y)^2)$, $y - 2 < x < y + 2$.
    b. **[5]** Write the corresponding kernel function that should be used to estimate $F(5)$. Note that only the kernel is required. No need to estimate $F(5)$.

4. **[15]** Let X be a population with distribution function given by $F(x) = 1 - \frac{\theta}{x}$, $x > \theta, \theta > 0$. A sample of 20 losses resulted in the following

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

Calculate the maximum likelihood estimate for θ.

5. You are given a random sample with $n = 50$ observations from a binomial population with parameters 3 and θ , i.e. $f_X(x|\theta) = \frac{3!}{x!(3-x)!}\theta^x(1-\theta)^{3-x}$, $x = 0,1,2,3$ and $0 < \theta < 1$, where θ is an unknown parameter.
You also know that $t = \sum_{i=1}^{n} x_i = 15$.
    a. **[10]** Show that $\hat{\theta} = \frac{\bar{X}}{3}$ is the maximum likelihood estimator for θ.
    b. **[10]** Is $\hat{\theta}$ an UMVUE estimator for θ? Justify.
    c. **[5]** Using the asymptotic distribution of the maximum likelihood estimator, obtain a 95% confidence interval for θ.

d. Now estimate θ using a Bayesian framework and choosing $\pi(\theta) = 1, 0 < \theta < 1$, as the prior.

   i. **[5]** Is the prior an improper distribution? Justify

   ii. **[15]** Obtain the posterior and also obtain a point estimate for θ using a 0-1 loss function. Comment.

   iii. **[10]** What is your estimate for the probability that the next observation's value is 0?

6. **[10]** You are given the following random sample (0.65, 1.20, 1.61, 1.98, 2.30). Using the Kolmogorov-Smirnov test, test if it is acceptable to consider that the distribution function of the population is given by a Burr with parameters 2, 2 and 3, i.e. $F(x) = 1 - \frac{64}{(8+x^3)^2}$, $x \geq 0$.

Solutions

1.

a.

$H_0: \mu \geq 40$ against $H_1: \mu < 40$

Test Statistic: $T = \frac{(\bar{X}-40)}{\frac{S}{3}} \sim t(8)$

$T_{obs} = \frac{333/9-40}{\frac{\sqrt{312/8}}{3}}$= -1.44115 p-value= $P(T < T_{obs}) = 0.094$

As the p-value is greater than the significance level we do not reject the null and consequently we have not enough statistical evidence to reject the claim.

b.

$H_0: \sigma^2 \leq 6^2$ against $H_1: \sigma^2 > 6^2$

Test Statistic: $Q = \frac{8\,S^2}{36} \sim \chi^2(8)$

$Q_{obs} = \frac{8\times(\frac{312}{8})}{36} = 8.6667$ p-value= $P(Q > Q_{obs}) = 0.371$

As the p-value is greater than the significance level we do not reject the null and consequently we have not enough statistical evidence to reject that the population's standard deviation is smaller than or equal to $6000.

c.

Pivotal Quantity: : $T = \frac{\bar{X}-\mu}{\frac{S}{3}} \sim t(8)$ $t_{\alpha/2} = 2.306$

The 95% CI is then (32.20; 41.80) $\frac{333}{9} \pm 2.306\,\frac{\sqrt{312/8}}{3}$

2.

a.

We know that $n = 48 + a$

Using the ogive, $F_n(12) = F_n(10) + \frac{2}{10}\left(F_n(20) - F_n(10)\right) = \frac{45}{n} + \frac{2}{10}\left(\frac{45+a}{n} - \frac{45}{n}\right) = \frac{450+2a}{10\,n}$

Then we must solve the system

$$\begin{cases} n = 48 + a \\ \dfrac{450 + 2a}{10\,n} = \dfrac{16}{21} \end{cases} \Leftrightarrow \begin{cases} a = n - 48 \\ \dfrac{450 + 2n - 96}{10\,n} = \dfrac{16}{21} \end{cases} \Leftrightarrow \begin{cases} a = n - 48 \\ 354 \times 21 + 42n = 160n \end{cases} \Leftrightarrow \begin{cases} a = n - 48 \\ n = \dfrac{7434}{118} = 63 \end{cases}$$

Getting $n = 63$ and $a = 15$.

b.

$\widehat{S_n}(10) = \frac{63-45}{63} = \frac{18}{63}$ $\widehat{var}\left(\widehat{S_n}(10)\right) = \frac{\left(\frac{18}{63}\right)\times\left(1-\frac{18}{63}\right)}{63} = \frac{810}{250047} = 0.00324$

The 95% CI is then given by $\frac{18}{63} \pm 1.96\,\sqrt{0.00324}$ and we obtain (0.1742; 0.3973)

c.

$P(X > 10|X > 1) = \frac{P(X>10)}{P(X>1)} = \frac{S(10)}{S(1)}$

Then, the estimate is

$\hat{P}(X > 10|X > 1) = \frac{S_n(10)}{S_n(1)} = \frac{18}{58} = 0.310345$

3.

a.

$$\hat{f}(6) = \frac{1}{6}\,k_2(6) + \frac{2}{6}\,k_3(6) + \frac{2}{6}\,k_5(6) + \frac{1}{6}\,k_7(6)$$

$$\hat{f}(6) = \frac{1}{6} \times 0 + \frac{2}{6} \times 0 + \frac{2}{6}\left(\frac{3}{32}\right)(4-(6-5)^2) + \frac{1}{6}\left(\frac{3}{32}\right)(4-(6-7)^2) = \frac{3}{32} + \frac{3}{64} = \frac{9}{64} = 0.1406$$

b.

The kernel will be given by the corresponding distribution function

$$K_y(x) = \begin{cases} 0 & x < y-2 \\ \int_{y-2}^{x} k_y(t)\,dt & y-2 \le x < y+2 \\ 1 & x \ge y+2 \end{cases}$$

$$= \begin{cases} 0 & x < y-2 \\ \left(\frac{1}{32}\right)(12(x-y)-(x-y)^3+16) & y-2 \le x < y+2 \\ 1 & x \ge y+2 \end{cases}$$

As

$$\int_{y-2}^{x} k_y(t)\,dt = \int_{y-2}^{x}\left(\frac{3}{32}\right)(4-(t-y)^2)\,dt = \left(\frac{3}{32}\right)\left(4t - \frac{(t-y)^3}{3}\right]_{y-2}^{x}$$

$$= \left(\frac{3}{32}\right)\left(\left(4x - \frac{(x-y)^3}{3}\right) - \left(4(y-2) - \frac{(y-2-y)^3}{3}\right)\right)$$

$$= \left(\frac{3}{32}\right)\left(4x - \frac{(x-y)^3}{3} - 4y + 8 - \frac{8}{3}\right) = \left(\frac{1}{32}\right)(12(x-y)-(x-y)^3+16)$$

4.

Let k be the number of intervals ($k = 3$) and let us consider that interval j is limited by c_{j-1} and c_j and that it includes n_j observations. Let us define $c_0 = \theta$ (then $F(c_0) = 0$) and $c_k = \infty$ (then $F(c_k) = 1$).

$$\ell(\theta) = \sum_{j=1}^{k} n_j \ln\left(F(c_j) - F(c_{j-1})\right) = n_1 \ln F(c_1) + \sum_{j=2}^{k-1} n_j \ln\left(\frac{\theta}{c_{j-1}} - \frac{\theta}{c_j}\right) + n_k \ln(1 - F(c_{k-1}))$$

$$= n_1 \ln\left(1 - \frac{\theta}{c_1}\right) + \sum_{j=2}^{k-1} n_j \ln\left(\frac{\theta\,(c_j - c_{j-1})}{c_{j-1}\,c_j}\right) + n_k \ln\left(\frac{\theta}{c_{k-1}}\right)$$

$$= n_1 \ln\left(1 - \frac{\theta}{c_1}\right) + \sum_{j=2}^{k-1} n_j\left(\ln\theta + \ln(c_j - c_{j-1}) - \ln(c_{j-1}\,c_j)\right) + n_k(\ln\theta - \ln c_{k-1})$$

Then

$$\ell'(\theta) = n_1 \frac{(-1)}{(c_1-\theta)} + \sum_{j=2}^{k-1} \frac{n_j}{\theta} + \frac{n_k}{\theta} = -\frac{n_1}{c_1-\theta} + \frac{n-n_1}{\theta}$$

$$\ell'(\theta) = 0 \Leftrightarrow \frac{n_1}{c_1-\theta} = \frac{n-n_1}{\theta} \Leftrightarrow \frac{c_1-\theta}{\theta} = \frac{n_1}{n-n_1} \Leftrightarrow \frac{c_1}{\theta} = \frac{n_1}{n-n_1} + 1 \Leftrightarrow \theta = \frac{c_1(n-n_1)}{n}$$

As

$\ell''(\theta) = -\frac{n_1}{(c_1-\theta)^2} - \frac{n-n_1}{\theta^2} < 0$

The ml estimate is $\hat{\theta} = \frac{c_1(n-n_1)}{n} = \frac{10 \times 11}{20} = \frac{11}{2}$

5.

a.

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

$$\ell'(\theta) = \sum_{i=1}^{n}\left(\frac{x_i}{\theta} - \frac{3-x_i}{1-\theta}\right) = \frac{t}{\theta} - \frac{3n-t}{1-\theta}$$
$$\ell'(\theta) = 0 \Leftrightarrow \frac{t}{\theta} = \frac{3n-t}{1-\theta} \Leftrightarrow \frac{1-\theta}{\theta} = \frac{3n-t}{t} \Leftrightarrow \frac{1}{\theta} = \frac{3n}{t} \Leftrightarrow \theta = \frac{t}{3n} = \frac{\bar{x}}{3}$$

As $\ell''(\theta) = -\frac{t}{\theta^2} - \frac{3n-t}{(1-\theta)^2} < 0$, the maximum likelihood estimator will be $\hat{\theta} = \frac{\bar{X}}{3}$.

b.

As $E(\hat{\theta}) = E\left(\frac{\bar{X}}{3}\right) = \frac{E(X)}{3} = \frac{3\theta}{3} = \theta$, $\hat{\theta}$ is an unbiased estimator for θ.

$$var(\hat{\theta}) = var\left(\frac{\bar{X}}{3}\right) = \frac{var(\bar{X})}{9} = \frac{var(X)}{9n} = \frac{3\theta(1-\theta)}{9n} = \frac{\theta(1-\theta)}{3n}$$

Fisher's information quantity is

$$I_{X_1,X_2,\dots,X_n}(\theta) = -E(\ell''(\theta|X_1,X_2,\dots,X_n) = -E\left(-\frac{\sum_{i=1}^{n}X_i}{\theta^2} - \frac{3n-\sum_{i=1}^{n}X_i}{(1-\theta)^2}\right) = \frac{3n}{\theta^2} + \frac{3n-3n}{(1-\theta)^2}$$
$$= \frac{3n}{\theta} + \frac{3n}{1-\theta} = \frac{3n}{\theta\,(1-\theta)},$$

And then Cramer-Rao lower bound is equal to $I_{X_1,X_2,\dots,X_n}(\theta)^{-1} = \frac{\theta\,(1-\theta)}{3n}$.

As the maximum likelihood estimator is unbiased and its variance is equal to Cramer-Rao lower bound it is an UMVUE estimator for θ.

c.

The 95% asymptotic confidence interval is given by $\hat{\theta} \pm 1.96\sqrt{\widehat{var}(\hat{\theta})} = 0.1 \pm 1.96\sqrt{\frac{0.1\times 0.9}{150}}$, i.e. (0.052; 0.148)

d.

i.

No, the prior is a proper distribution as $\int_0^1 \pi(\theta)\, d\theta = 1$.

ii.

$\pi(\theta) = 1, 0 < \theta < 1$

$L(\theta) = \prod_{i=1}^{n}\frac{3!}{x_i!(3-x_i)!}\theta^{x_i}(1-\theta)^{3-x_i} \propto \prod_{i=1}^{n}\theta^{x_i}(1-\theta)^{3-x_i} = \theta^t(1-\theta)^{3n-t}, 0 < \theta < 1$

$\pi(\theta|\boldsymbol{x}) \propto \theta^t(1-\theta)^{3n-t}, 0 < \theta < 1$

The posterior is then a beta distribution with parameters $t+1$ and $3n-t+1$, i.e. 16 and 136.

The point estimate against a 0-1 loss function is the mode of the posterior and then we will obtain the same estimate as we obtained using maximum likelihood.

iii.

Let us denote the next observation by Y.

$$f_Y(y) = \int_0^1 f_X(y|\theta)\,\pi(\theta|\boldsymbol{x})\,d\theta = \int_0^1 \frac{3!}{y!(3-y)!}\theta^y(1-\theta)^{3-y}\,\frac{\Gamma(3n+2)}{\Gamma(t+1)\,\Gamma(3n-t+1)}\theta^t(1-\theta)^{3n-t}\,d\theta$$
$$f_Y(y) = \frac{3!}{y!(3-y)!}\,\frac{\Gamma(3n+2)}{\Gamma(t+1)\,\Gamma(3n-t+1)}\int_0^1 \theta^{t+y}(1-\theta)^{3n-t+3-y}\,d\theta$$

As $\theta^{t+y}(1-\theta)^{3n-t+3-y}$ is the core of a beta distribution we can easily compute the integral:

$$\int_0^1 \theta^{t+y}(1-\theta)^{3n-t+3-y}\,d\theta = \frac{\Gamma(t+y+1)\,\Gamma(3n-t-y+4)}{\Gamma(3n+3)}\int_0^1 \frac{\Gamma(3n+5)}{\Gamma(t+y+1)\,\Gamma(3n-t-y+4)}\,\theta^{t+y}(1-\theta)^{3n-t+3-y}\,d\theta$$
$$= \frac{\Gamma(t+y+1)\,\Gamma(3n-t-y+4)}{\Gamma(3n+5)}$$

And then

$$f_Y(y) = \frac{3!}{y!(3-y)!} \frac{\Gamma(3n+2)}{\Gamma(t+1)\,\Gamma(3n-t+1)} \frac{\Gamma(t+y+1)\,\Gamma(3n-t-y+4)}{\Gamma(3n+5)}, y = 0, 1,2,3$$

So

$$f_Y(0) = \frac{\Gamma(3n+2)}{\Gamma(t+1)\,\Gamma(3n-t+1)} \frac{\Gamma(t+1)\,\Gamma(3n-t+4)}{\Gamma(3n+5)} = \frac{(3n-t+3)(3n-t+2)(3n-t+1)}{(3\mathrm{n}+4)\,(3\mathrm{n}+3)\,(3\mathrm{n}+2)}$$

$$= \frac{138 \times 137 \times 136}{154 \times 153 \times 152} = 0.7179$$

6.

$H_0: F(x) = 1 - \frac{64}{(8+x^3)^2}$, $x \geq 0$ against $H_1: H_0$ is false

x_i	$F(x_i)$	$\frac{i-1}{n}$	$\frac{i}{n}$	$D_i = \max(D_i^-, D_i^+)$
0.65	0.065	0	0.2	0.135
1.20	0.324	0.2	0.4	0.124
1.61	0.568	0.4	0.6	0.168
1.98	0.742	0.6	0.8	0.142
2.30	0.843	0.8	1	0.157

$D = \max(D_i) = 0.168$ Critical asymptotic value $(\alpha = 0.05) = \frac{1.36}{\sqrt{5}} = 0.6082$

We do not reject the null as the statistical evidence is not strong enough to reject the null.