INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

April 2014 examinations

Subject CT3 – Probability and Mathematical Statistics Core Technical

Introduction

The Examiners' Report is written by the Principal Examiner with the aim of helping candidates, both those who are sitting the examination for the first time and using past papers as a revision aid and also those who have previously failed the subject.

The Examiners are charged by Council with examining the published syllabus. The Examiners have access to the Core Reading, which is designed to interpret the syllabus, and will generally base questions around it but are not required to examine the content of Core Reading specifically or exclusively.

For numerical questions the Examiners' preferred approach to the solution is reproduced in this report; other valid approaches are given appropriate credit. For essay-style questions, particularly the open-ended questions in the later subjects, the report may contain more points than the Examiners will expect from a solution that scores full marks.

The report is written based on the legislative and regulatory context pertaining to the date that the examination was set. Candidates should take into account the possibility that circumstances may have changed if using these reports for revision.

D C Bowie
Chairman of the Board of Examiners

June 2014

General comments on Subject CT3

Some of the questions in this paper admit alternative solutions from these presented in the marking schedule, or different ways in which the provided answer can be determined. All mathematically correct and valid alternative solutions or answers received credit as appropriate. Rounding errors were not penalised, unless excessive rounding led to significantly different answers. In cases where the same error was carried forward to later parts of the answer, candidates were only penalised once. In questions where comments were required, reasonable comments that were different from those provided in the solutions also received full credit.

Comments on the April 2014 paper

The performance was generally good. The pass rate was in line with previous diets. Candidates that were sufficiently prepared were able to answer all questions and the best candidates scored close to full marks. As in previous diets, questions that covered topics that were not recently examined proved to be more challenging for less well prepared candidates.

The comments on individual questions that follow cover important frequent errors, and specific parts that were not answered well.

1 (i) Mean = $\dfrac{1585}{20} = 79.25$

Median = 70

(ii) Var = $\dfrac{142{,}127 - \dfrac{1585^2}{20}}{19} = 869.25$, SD $= \sqrt{869.25} = 29.48$

Well answered.

2 We want to find a and b for $y = a + bx$ such that

$\bar{y} = a + b\bar{x} = 50$ and $s_y = |b| s_x$

These give $b = 2$ and $a = 50 - 124 = -74$
or, $b = -2$ and $a = 50 + 124 = 174$

Generally well answered.

3 Consider the following events:

A: Driver has had additional education
B: Driver has *not* had additional education
C: Driver has *not* had accident in the first year.

(a) P(C) = P(C|A) Pr(A) + P(C|B) Pr(B) = 0.95*0.6 + 0.91*0.4

= 0.934

(b) $$P(\text{A} \mid \text{C}) = \frac{P(\text{C} \mid \text{A})P(\text{A})}{P(\text{C})} = \frac{0.95 \times 0.6}{0.934} = 0.610$$

Reasonably well answered. Some candidates did not realise that the answer from part (a) could be used in part (b).

4 (i) $$M_X(t) = E(e^{tX}) = \frac{1}{2}\int_{-\infty}^{0} e^{(t+1)x}dx + \frac{1}{2}\int_{0}^{\infty} e^{(t-1)x}dx$$

$$= \frac{1}{2}\left[\frac{e^{(t+1)x}}{t+1}\right]_{-\infty}^{0} + \frac{1}{2}\left[\frac{e^{(t-1)x}}{t-1}\right]_{0}^{\infty}$$

and for $|t| < 1$

$$M_X(t) = \frac{1}{2}\left(\frac{1}{t+1} - \frac{1}{t-1}\right) = \frac{1}{1-t^2}$$

(ii) $M'_X(t) = \left\{(1-t^2)^{-1}\right\}' = -(1-t^2)^{-2}(-2t) = 2t(1-t^2)^{-2}$

$\Rightarrow E(X) = M'_X(0) = 0$

$M''_X(t) = \left\{2t(1-t^2)^{-2}\right\}' = 2(1-t^2)^{-2} + 2t(-2)(1-t^2)^{-3}(-2t)$

$= 2(1-t^2)^{-2} + 8t^2(1-t^2)^{-3}$

$\Rightarrow E(X^2) = M''_X(0) = 2$

$V(X) = E(X^2) - E^2(X) = 2$

(Alternatively, based on a series expansion:

$M_X(t) = 1 + t^2 + t^4 + \ldots \Rightarrow E(X) = 0$ and $E(X^2) = 2$ and the variance follows.)

Generally well answered. In part (ii) most candidates were familiar with the method, but some showed poor differentiation skills.

5 (i) $X \sim N\left(\mu, \sigma^2\right)$ with $\mu = 10 \times 4 = 40$ and $\sigma^2 = 10 \times (4)^2 = 160$

(ii) X is symmetric so $P[X < 40] = 0.5$

(iii) The exact distribution of X is gamma(10, ¼)

$P[X < 40] = P\left[2 \times \frac{1}{4} \times X < 20\right] = P[Y < 20]$ where Y has a χ^2 distribution with 20 d.f.

$P[X < 40] = P[Y < 20] = 0.5421$

(iv) Although the sample size here is small, the CLT gives an answer which is close to the exact probability.

Mostly well answered. There were a few problems with the distributions in part (iii). In part (iv) comments should refer to the use of CLT with small samples for full marks.

6 (i) $$P(B>0.5)=P\left(\frac{B-0.47}{\sqrt{0.47*\frac{1-0.47}{100}}}>\frac{0.5-0.47}{\sqrt{0.47*\frac{1-0.47}{100}}}\right)=1-\Phi(0.601)=0.274$$

(ii) H_0 = Towns have the same voting intentions

Actual	Candidate	A	B	C	Sum
	Town 1	32	47	21	100
	Town 2	57	56	37	150
		89	103	58	250

Expected	Candidate	A	B	C	
	Town 1	35.6	41.2	23.2	100
	Town 2	53.4	61.8	34.8	150
		89	103	58	250

$\frac{(f-e)^2}{e}$			
	0.364	0.817	0.209
	0.243	0.544	0.139

Test statistic = 2.315

Degrees of freedom = (3 – 1)*(2 – 1) = 2. Approximate p-value of X_2^2 distribution is between 0.30 and 0.32 (0.314 from interpolation.)

Therefore we fail to reject H_0 that towns have the same voting intentions.

The wording in part (i) of the question was not entirely clear, as the question should in fact refer to the probability that the candidate will get more than 50% of the vote in a different sample of the same size. However, there was very little evidence that candidates were confused by this, and marking was generous in cases where answers seemed to be affected.

In general the question was very well answered. Answers including a continuity correction were also given full marks in part (i).

7 (i) $$E\left[E[Y|X]\right]=\int E[Y|x]\,f_X(x)\,dx=\int(\int yf_{y|x}(y)\,dy)\,f_X(x)\,dx$$

$$=\int\int yf(x,y)\,dydx=E[Y]$$

(ii) $$E[S]=E\left[E[S|N]\right]=E\left[E[X_1+\ldots+X_N|N]\right]=E[N\alpha/\lambda]=\mu\alpha/\lambda$$

(iii) As S is compound Poisson,

$$V[S] = \mu E\left[X^2\right]$$

$$= \mu\left(\frac{\alpha}{\lambda^2} + \left(\frac{\alpha}{\lambda}\right)^2\right)$$

$$= 0.15 * \left(\frac{100}{0.1^2} + \left(\frac{100}{0.1}\right)^2\right)$$

$$= 0.15 * (1{,}010{,}000)$$

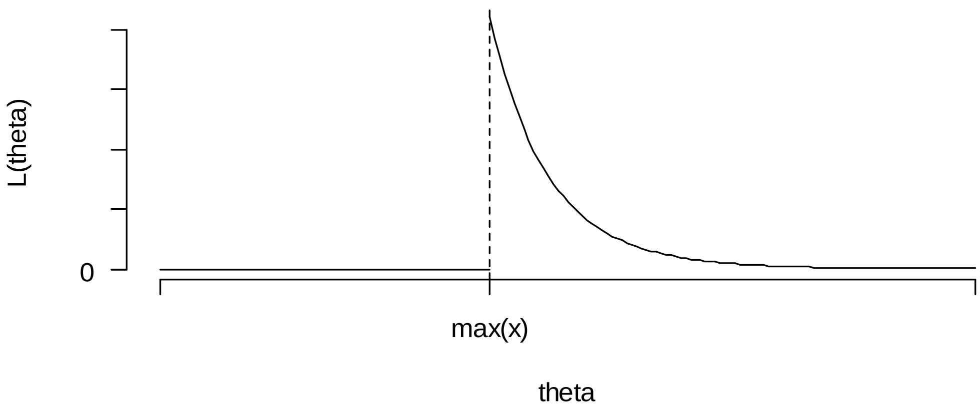
$$= 151{,}500$$

Some mixed performance in part (i), which suggests that some candidates struggled with basic integration skills. The rest of the question was answered well. Use of alternative formulae was given full credit where correct.

8 (i) (a) Likelihood is given as

$$L(\theta; \underline{x}) = \begin{cases} \Pi_{i=1}^{n} f(x_i;\theta) = \frac{2^n x_1 x_2 \cdots x_n}{\theta^{2n}} & \text{if } \theta \geq x_{(n)} = \max\{x_1, x_2, \cdots, x_n\} \\ 0 & \text{if } \theta < x_{(n)}. \end{cases}$$

Its graph is given below:



(b) From the graph, the likelihood is maximised at

$$\theta = x_{(n)} = \max\{x_1, x_2, \ldots, x_n\}.$$

(ii) (a) $F_{X_{(n)}}(x) = P\left\{X_{(n)} < x\right\} = P\{X_1 < x, X_2 < x, \ldots, X_n < x\}$

$= P(X_1 < x)P(X_2 < x)\cdots P(X_n < x)$ as X_i are independent

$= P(X_1 < x)^n$ since X_i are identically distributed

$$= \left\{\int_0^x \frac{2u}{\theta^2}\,du\right\}^n = \left\{\left[\frac{u^2}{\theta^2}\right]_0^x\right\}^n = \left(\frac{x}{\theta}\right)^{2n}$$

(b) Differentiating we obtain

$$f_{X_{(n)}}(x) = \begin{cases} \dfrac{2nx^{2n-1}}{\theta^{2n}} & \text{if } 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$$

(c) $$E\left(X_{(n)}\right) = \int_0^\theta \frac{2nx^{2n}}{\theta^{2n}}\,dx = \frac{2n\theta}{2n+1}$$

$$E\left(X_{(n)}^2\right) = \int_0^\theta \frac{2nx^{2n+1}}{\theta^{2n}}\,dx = \frac{n\theta^2}{n+1}$$

$$V\left(X_{(n)}\right) = E\left(X_{(n)}^2\right) - \left\{E\left(X_{(n)}\right)\right\}^2 = \frac{n\theta^2}{n+1} - \left(\frac{2n\theta}{2n+1}\right)^2 = \frac{n\theta^2}{(n+1)(2n+1)^2}$$

(d) $$E\left\{\frac{2n+1}{2n}X_{(n)}\right\} = \frac{2n+1}{2n}E\left(X_{(n)}\right) = \frac{2n+1}{2n}\,\frac{2n\theta}{2n+1} = \theta$$

(iii) (a) $$MSE\left\{\frac{2n+1}{2n}X_{(n)}\right\} = V\left\{\frac{2n+1}{2n}X_{(n)}\right\}$$

$$= \left(\frac{2n+1}{2n}\right)^2 V\left\{X_{(n)}\right\} = \left(\frac{2n+1}{2n}\right)^2 \frac{n\theta^2}{(n+1)(2n+1)^2} = \frac{\theta^2}{4n(n+1)}$$

(b) We have $MSE \to 0$ as $n \to \infty$, therefore the estimator is consistent.

This question was not well answered, and there were some poor efforts especially in part (i). In many cases, the plotted graph revealed inadequate understanding of the likelihood concept, with some candidates attempting to draw it as a function of x. Note that for full marks the likelihood needs to include the range of the parameter and the graph must indicate *the value of max(x) on the x-axis. In parts (ii) and (iii) some candidates did not cope well with the algebra.*

9 (i) Overall average is (200 + 170 + 155)/3 = 175 since sample sizes are all equal.

(ii) $\bar{X} \pm t_{0.025,24}\dfrac{30}{5} = [200 - 2.064*6, 200 + 2.064*6] = [187.62,\ 212.38]$

(iii) $\bar{Z} \pm t_{0.025,24}\dfrac{20}{5} = [155 - 2.064*4, 155 + 2.064*4] = [146.74,\ 163.26]$

(iv) $\left[\dfrac{S_X^2}{S_Z^2} \times \dfrac{1}{F_{24,24}}, \dfrac{S_X^2}{S_Z^2} \times F_{24,24}\right] = \left[\dfrac{2.25}{2.269}, 2.25 \times 2.269\right] = [0.992, 5.105]$

(v) The ratio 1 is contained in the confidence interval, therefore the null hypothesis $\sigma_X^2 = \sigma_Z^2$ cannot be rejected.

(vi) Pooled variance: $s_p^2 = \dfrac{24 \times \left(30^2 + 20^2\right)}{48} = 650$.

Difference: 200 − 155 = 45

$$\left[45 - t_{0.025,48}\sqrt{650}\sqrt{\frac{2}{25}},\ 45 + t_{0.025,48}\sqrt{650}\sqrt{\frac{2}{25}}\right]$$

$$[45 - 2.01 \times 7.21,\ 45 + 2.01 \times 7.21] = [30.51, 59.49]$$

where we have used the approximation $t_{0.025,48} = 2.01$ (see tables, value for $t_{0.025,50} = 2.009$)

We made the assumption $\sigma_X^2 = \sigma_Z^2$ which is justified by the result in parts (iv) and (v).

(vii) The confidence interval does not contain 0, so there is a difference.

(viii) $SS_R = 24 \times \left(30^2 + 30^2 + 20^2\right) = 52800$

Alternative solution possible

$$SS_B = 25 \times \left[(200 - 175)^2 + (170 - 175)^2 + (155 - 175)^2\right] = 26250$$

$$\frac{SS_B / 2}{SS_R / 72} = \frac{13125}{733.33} = 17.9$$

This is clearly a very large value compared to $F_{2,72} < F_{2,60} = 4.977$ at the 1% level, so the age of the child has an impact on childcare cost.

Generally well answered. In part (iv) calculation of the ratio of the variance of Z over the variance of X was given full credit. In part (viii) many candidates attempted to calculate the SS values using the original data, rather than the "quick" formulae given in the answer. This was given full marks where appropriate, but was not the best use of time in the exam.

10 (i) There appears to be a positive linear relationship

(ii) (a) $$S_{ss} = \sum s_i^2 - \left(\left(\sum s_i\right)^2\right) / n = 397499.8 - (2843.7)^2 / 33 = 152450.4$$

$$S_{vv} = \sum v_i^2 - \left(\left(\sum v_i\right)^2\right) / n = 689.37 - (115.34)^2 / 33 = 286.24$$

$$S_{vs} = \sum v_i s_i - \left(\left(\sum v_i \sum s_i\right)\right) / n = 15417.75 - (2843.7 \times 115.34) / 33 = 5478.6$$

$$r = \frac{S_{vs}}{\sqrt{S_{ss} S_{vv}}} = \frac{5478.6}{\sqrt{152450.4 \times 286.24}} = 0.8294$$

(ii) (b) $H_0 : r = 0, H_1 : r \neq 0$

Test statistic $= r \dfrac{\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.8294\sqrt{33-2}}{\sqrt{1-0.8294^2}} = 8.266$

At 0.5% level $t_{31} = 2.744$ which << test statistic

So reject H_0.

(iii) $$\beta = \frac{S_{vs}}{S_{ss}} = \frac{5478.6}{152450.4} = 0.0359$$

$$\alpha = \bar{v} - \beta\bar{s} = \frac{115.34}{33} - 0.0359\frac{2843.7}{33} = 0.398$$

$$v_i = 0.398 + 0.0359 s_i$$

(iv) Testing whether β is significantly different from zero is mathematically the same as testing whether the correlation coefficient is significantly different from zero.

As H_0 was rejected in (ii)(b), we can conclude testing $H_0 : \beta = 0$ would give the same result.

(v) It is true that extreme observations can determine the strength of a linear relationship. However, there are many more bonds in the central part of the data and we would consequently expect a greater range of value traded.

Generally well answered. In part (ii)(b) Fisher's z transformation method was also allowed. In part (v) other possible reasonable comments were given credit.

END OF EXAMINERS' REPORT