INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

September 2013 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

December 2013

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 September 2013 paper

Performance was overall satisfactory, resulting in high pass rate. Candidates that were sufficiently prepared were able to answer all questions, and there was a good proportion of very high marks. As in previous diets, questions that addressed topics that were not recently examined proved to be 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 $= 50/40 = 1.25$

Median = 1.252

Mode = 1.257

(ii) Mean. Distribution is roughly symmetrical with no outliers so no reason to use anything else.

Generally well answered. In part (ii), answers claiming that the median is preferred due to some skewness in the distribution were not penalised.

2 Annual claims ~ Poisson(150) so six-month claims, X ~ Poisson(75)

CLT gives approximate distribution $N(75,75)$

$$P\left(X > 90\right) = P\left(X > 90.5\right) = P\left(\frac{X-75}{\sqrt{75}} > \frac{90.5-75}{\sqrt{75}}\right) = 1-\Phi\left(1.790\right) = 1-0.963 = 0.037$$

[Without continuity correction $1-\Phi\left(1.732\right) = 0.042$]

There were no particular problems with this question. Note that the continuity correction must be applied for full marks.

3 (i)

$$E(X) = \int_0^\theta x\frac{2x}{\theta^2}dx = \left[\frac{2x^3}{3\theta^2}\right]_0^\theta = \frac{2\theta}{3}$$

$$E(X^2) = \int_0^\theta x^2\frac{2x}{\theta^2}dx = \left[\frac{2x^4}{4\theta^2}\right]_0^\theta = \frac{\theta^2}{2}$$

$$\text{So, } \operatorname{var}(X) = E(X^2) - [E(X)]^2 = \frac{\theta^2}{2} - \frac{4\theta^2}{9} = \frac{\theta^2}{18}$$

(ii) $E(\hat{\theta}) = E\left(\frac{3\bar{X}}{2}\right) = \frac{3}{2}E(\bar{X}) = \frac{3}{2}E(X) = \theta$, so estimator is unbiased.

Generally well answered. A few problems were encountered when deriving the variance.

4 (i) First derive expected value:

$$E(X) = \sum_{x=1}^{\infty} -x \frac{1}{\log(1-\theta)} \frac{\theta^x}{x}$$

$$= \sum_{x=0}^{\infty} \left(-\frac{\theta^x}{\log(1-\theta)} \right) + \frac{1}{\log(1-\theta)}$$

$$= -\frac{\theta}{(1-\theta)\log(1-\theta)}$$

$$\bar{X} = E(X) \Rightarrow \bar{X} = -\frac{\tilde{\theta}}{(1-\theta)\log(1-\tilde{\theta})}$$

$$\Rightarrow \bar{X}\left(1-\tilde{\theta}\right)\log\left(1-\tilde{\theta}\right) + \tilde{\theta} = 0$$

(ii) (a) $$L(\theta) = \frac{\theta^{\sum_i x_i}}{(-\log(1-\theta))^n \prod_i x_i}$$

And $l(\theta) = -n\log\left(-\log\left(1-\theta\right)\right) + \sum_i x_i \log(\theta) + C$

(b) MLE given by:

$$\frac{dl(\theta)}{d\theta} = 0 \Rightarrow \frac{n}{\log\left(1-\hat{\theta}\right)\left(1-\hat{\theta}\right)} + \frac{\sum_i x_i}{\hat{\theta}} = 0$$

$$\Rightarrow \bar{X}\left(1-\hat{\theta}\right)\log\left(1-\hat{\theta}\right) + \hat{\theta} = 0$$

(iii) The equation above needs to be solved numerically. Alternatively, the likelihood (or log-likelihood) function can be plotted and the maximum can be identified from the graph.

In part (ii)(a) of the question the log-likelihood was shown as being equal, rather than proportional, to the given expression plus a constant (as given in the solution above). Candidates did not seem to be confused by this, but marking was adjusted in relevant cases.

In general the question was not particularly well answered, mainly due to difficulties in the mathematical operations involved in obtaining the log-likelihood function of non-standard densities. Candidates are advised to practise their calculus skills to deal with such questions.

5 (i) $$E\left(S^2\right)=E\left\{\frac{\left(\sum X_i^2-n\bar{X}^2\right)}{n-1}\right\}=\frac{1}{n-1}\sum E\left(X_i^2\right)-\frac{n}{n-1}E\left(\bar{X}^2\right)$$

and using $E\left(X^2\right)=var\left(X\right)+\left\{E(X)\right\}^2$

$$E\left(S^2\right)=\frac{n}{n-1}\left(\sigma^2+\mu^2\right)-\frac{n}{n-1}\left(\sigma^2/n+\mu^2\right)=\sigma^2$$

(ii) (a) $$\text{var}\left\{\frac{(n-1)S^2}{\sigma^2}\right\}=2(n-1)$$

$$\Rightarrow \text{var}\left(S^2\right)=\frac{2(n-1)\sigma^4}{(n-1)^2}=\frac{2\sigma^4}{(n-1)}$$

(b) Estimator gets better (more accurate) as n increases, as its variance reduces.

(MSE also gets smaller)

This question was generally well answered. There were a few problems with determining the expectation of the sample mean in part (i).

6 (i) H_0 = variances are the same, H_1 = variances are different

$S_2/S_1 \sim F_{24,24}$

Test statistic = 9.21/2.86 = 3.22.

$F_{24,24,0.995}=0.337$ and $F_{24,24,0.005}=2.967$

i.e. reject H_0 at 1% significance level.

(ii) Confidence interval is given by $\left(\frac{(n-1)S^2}{X^2_{0.025,n-1}},\frac{(n-1)S^2}{X^2_{0.975,n-1}}\right)$

$X^2_{0.975,24}=12.40, X^2_{0.025,24}=39.36$

Confidence interval 1 = (1.74, 5.54)

Confidence interval 2 = (5.61, 17.83)

(iii) Confidence intervals don't overlap i.e. agree with result in (i) that variances are different.

Generally well answered. In part (i) some candidates worked with the S_1/S_2 ratio, which of course gives the same conclusion. Part (ii) requires the calculation of two CIs, but some candidates attempted to provide a CI for the ratio.

7 (i)

$$P[\text{claim}] = P[\text{claim}|A]P[A] + P[\text{claim}|B]P[B] + P[\text{claim}|C]P[C]$$
$$= 0.15*0.2 + 0.1*0.2 + 0.05*0.6 = 0.08$$

(ii)

$$P[\text{claim} \cap A] = P[\text{claim}|A]P[A] = 0.15*0.2 = 0.03$$

$$P[\text{claim} \cap B] = P[\text{claim}|B]P[B] = 0.1*0.2 = 0.02$$

$$P\left[\text{claim} \cap \bar{C}\right] = 0.03 + 0.02 = 0.05$$

$$P\left[\text{claim} \mid \bar{C}\right] = \frac{P\left[\text{claim} \cap \bar{C}\right]}{P\left[\bar{C}\right]} = \frac{0.05}{0.4} = 0.125$$

(iii)

$$P[A|\text{claim last year}] = \frac{P[\text{claim} \cap A]}{P[\text{claim}]} = \frac{0.03}{0.08} = 0.375$$

(iv)

$$P[B|\text{claim last year}] = \frac{P[\text{claim} \cap B]}{P[\text{claim}]} = \frac{0.02}{0.08} = 0.25$$

$$P[C|\text{claim last year}] = 1 - 0.375 - 0.25 = 0.375$$

(CLY means "claim last year")

$$P[\text{claim}|\text{claim last year}] = P[\text{claim}|A]P[A|CLY] + \ldots + P[\text{claim}|C]P[C|CLY]$$
$$= 0.15*0.375 + 0.1*0.25 + 0.05*0.375 = 0.1$$

(v) Let Y be the event that a claim is submitted in two consecutive years

$$P[Y] = P[\text{claim in second year}|\text{claim in first year}]P[\text{claim in first year}]$$
$$= 0.1*0.08 = 0.008$$

alternatively:

$$P[Y] = P[Y|A]P[A] + P[Y|B]P[B] + P[Y|C]P[C]$$
$$= 0.15*0.15*0.2 + 0.1*0.1*0.2 + 0.05*0.05*0.6 = 0.008$$

This turned out to be the most challenging question for the majority of candidates, with only a small number of "full mark" answers. Many candidates did not attempt parts (iv) and (v) at *all. The question deals with conditional probability concepts, starting with straightforward parts but building up to more complex calculations.*

8 (i) Mean = (0*40 + 1*25 +2*20+3*10+4*5)/100 = (25+40+30+20)/100 = 1.15

Median = 1

Mode = 0

VAR = [$(-1.15)^2$ *40 + $(-0.15)^2$*25 + 0.85^2*20+ 1.85^2*10 + 2.85^2*5]/99
= 1.4419

STD = 1.2

(ii) The estimate for the expectation of X is $\hat{\lambda}=1.15$
$n\hat{\lambda}=115$ is rather large and we can therefore, use a normal approximation to calculate the confidence interval.

$$1.15 \pm 1.96\sqrt{\frac{1.15}{100}} = \left(0.9398, 1.3602\right)$$

(iii) Total amount of claims = 25*1*1000 + 20*2*1100 +10*3*930 +5* 4*980
= 116,500

Average claim size = 116500/115 = 1,013.043

(iv) Compound Poisson

(v) Using standard results on compound distributions:

Expected value: $E = 100 * \lambda * 1010$ = 115 * 1010 = 116,150.00
Var = 115*120^2 + 115*1010^2 = 118,967,500
STD = 10,907.22

Generally well done, but some mixed performance in parts (ii) and (v). Note that calculations refer to a group of 100 policyholders – some candidates failed to take this into account.

9 (i) Sample sizes are small, therefore, we need a t-test. We need to assume that the variances are equal, although the sample standard deviations are different. Since the sample size is small we can argue that equal variances is a reasonable assumption.

Test statistic $t = (\overline{Y}_A - \overline{Y}_B) / \left(S_P \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \right) \sim t_{n_A+n_B-2}$ under the null hypothesis that expected car usage is equal in both cities.

$$S_P^2 = \frac{9S_A^2 + 9S_B^2}{18} = \frac{1}{2}\left(7.5^2 + 8^2\right) = 60.125$$

$$t = \frac{33-29}{\sqrt{12.025}} = 1.1535$$

Two sided test, critical values are −2.101 and 2.101 from t_{18}.

The null hypothesis of equal car usage is not rejected.

(ii) $X_A = Y_A - Z_A \sim N\left(\mu_A, \sigma_A^2\right)$

$H_0 : \mu_A \le 0$ and $H_1 : \mu_A > 0$ (also full marks for $H_0 : \mu_A = 0$ vs. $H_1 : \mu_A > 0$)

$$t = \frac{33-28.5}{2/\sqrt{10}} = \frac{4.5}{0.6325} = 7.115$$

Critical values from t_9 at 5%: 1.833

This is clear evidence that the null hypothesis is rejected, and therefore, car usage has been reduced significantly in City A.

(iii) $X_B = Y_B - Z_B \sim N\left(\mu_B, \sigma_B^2\right)$

CI: $29 - 28 \pm \frac{t_{9,0.025}\,2.5}{\sqrt{10}} = \left[1 - 2.262 \times 0.79,\ 1 + 2.262 \times 0.79\right] = \left[-0.788, 2.788\right]$

(marking: test statistic 1mark, critical value 1mark, correct answer 1mark)

(iv) Let x_{ij} be the difference in city i household j.

$$\sum_{j=1}^{10} x_{Aj} = 10\left(\bar{y} - \bar{z}\right) = 45,$$

$$\sum_{j=1}^{10} x_{Aj}^2 = \left(10-1\right)2^2 + 10(33-28.5)^2 = 238.5$$

(v) $$\sum_{i=A,B,C}\sum_{j=1}^{10} x_{ij}^2 = \sum_{j=1}^{10} x_{Aj}^2 + \sum_{j=1}^{10} x_{Bj}^2 + \sum_{j=1}^{10} x_{Cj}^2$$

$$SS_T = \left(238.5 + 66.25 + 241\right) - \frac{\left(45+10+40\right)^2}{10+10+10} = 545.75 - \frac{95^2}{30} = 244.92$$

$$SS_B = \left(\frac{45^2}{10}+\frac{10^2}{10}+\frac{40^2}{10}\right)-\frac{95^2}{30}=\frac{3\times 3725-9025}{30}=71.67$$

$$SS_R = 244.92-71.67=173.25$$

$$F=\frac{71.67/2}{173.25/27}=\frac{35.835}{6.417}=5.58 \text{ on (2, 27) degrees of freedom}$$

Critical value at 5%: 3.354

The null hypothesis that reduction in car usage is equal in the three cities is rejected.

There were no particular problems with this question. However a number of candidates failed to justify the assumptions in part (i), while some seemed not to understand fully the different test (or CI) requirements in different parts of the question.

10 (i) There is a positive linear relationship between the two.

(ii) (a)

$$S_{xx}=0.3612-0.101^2/12=0.360$$
$$S_{ff}=0.1710-0.622^2/12=0.139$$
$$S_{xf}=0.1989-0.101*0.622/12=0.194$$

$$r=\frac{S_{xf}}{\sqrt{S_{xx}S_{ff}}}=\frac{0.194}{\sqrt{0.360*0.139}}=0.867$$

(b) Statistic $=\dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}=\dfrac{0.867*\sqrt{10}}{\sqrt{1-0.867^2}}=5.50$

$t_{10,0.995}=3.169$

So reject H_0 that correlation coefficient = 0 at 1% level (2-sided test)

(iii) $\hat{\beta}=S_{xf}/S_{xx}=0.194/0.360=0.539$

$$\hat{\alpha}=\overline{f}-\hat{\beta}\overline{x}=\frac{0.622}{12}-0.539*\frac{0.101}{12}=0.0473$$

(iv) $$\hat{\sigma}^2 = \frac{1}{n-2}\left(S_{ff} - \frac{S_{xf}^2}{S_{xx}}\right) = \frac{1}{10}\left(0.139 - \frac{0.194^2}{0.360}\right) = 0.0034$$

$$t_{10,0.975} = 2.228$$

$$\text{C.I.} = \hat{\beta} \pm t_{10;0.975}\sqrt{\hat{\sigma}^2 / S_{xx}} = 0.539 \pm 2.228\sqrt{0.0034/0.36} = (0.321, 0.757)$$

(v) C.I. does not contain zero. Consistent with correlation coefficient not equal to zero as the test is actually the same. Both suggest that the hedge industry's claim that correlation is low may not be correct.

Very well answered in general. This is a typical regression/correlation question and the only (minor) problems concerned errors with calculators. Note that part (ii)(b) can also be answered using Fisher's transformation, which results in the same conclusion.

END OF EXAMINERS' REPORT