INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

April 2011 examinations

Subject CT3 — Probability and Mathematical Statistics Core Technical

Introduction

The attached subject report has been written by the Principal Examiner with the aim of helping candidates. The questions and comments are based around Core Reading as the interpretation of the syllabus to which the examiners are working. They have however given credit for any alternative approach or interpretation which they consider to be reasonable.

T J Birse
Chairman of the Board of Examiners

July 2011

General comments

The paper was answered very well and overall performance was satisfactory. Some problems were encountered with specific questions. Many candidates did not attempt Question 6 at all, while for those who did, manipulation of covariance terms often proved problematic. In Question 7 part (ii), a number of candidates failed to use the information given in the question regarding the proportion of claims identified as suspicious (i.e. 0.2) – instead they tried to compute this using the total probability theorem. This erroneously assumes a zero false positive rate for the software. In Question 10 part (iii), the non-standard form of the likelihood caused some poor answers. Also, some candidates inserted the data directly into the likelihood derivation, which resulted to only obtaining the ML estimate rather than also deriving the ML estimator as instructed.

1 (i) 30^{th} and 31^{st} observations in order are both 2 $\Rightarrow$ median = 2

mode = value with highest frequency = 1

$\Sigma x = 1(14) + 2(11) + 3(10) + 4(5) + 5(4) + 6(3) + 7(1) = 131$
$\Rightarrow$ mean = 131/60 = 2.18

(ii) Lower quartile is 15.5^{th} observation counting from below = 1

Upper quartile is 15.5^{th} observation counting from above = 3

$\Rightarrow$ IQR = 2

(iii) $\Sigma x^2 = 1(14) + 4(11) + 9(10) + 16(5) + 25(4) + 36(3) + 49(1) = 485$

$\Rightarrow$ standard deviation = $[(485 - 131^2/60)/59]^{1/2} = 3.3726^{1/2} = 1.84$

2 $$P(\bar{X} > 44.5) = P\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} > \frac{44.5 - 42}{7/\sqrt{36}}\right)$$

$\Rightarrow P(\bar{X} > 44.5) \approx P(Z > 2.143)$, where $Z \sim N(0,1)$,

and from tables,

$$P(\bar{X} > 44.5) = 1 - 0.984 = 0.016$$

(A t_{35} distribution can also be used if a normal distribution is assumed for the data.)

3 If X is the number of voters in the sample voting for party A, we have

$X \sim$ Binomial(200, 0.35) and using the CLT $X \sim N(70, 45.5)$ approximately.

Using continuity correction

$$P(X \geq 80) = P\left(Z > \frac{79.5-70}{\sqrt{45.5}}\right) = P(Z > 1.408)$$

$= 1 - P(Z < 1.408) = 1 - 0.920 = 0.08.$

4 (i) Compound Poisson distribution

(ii) $E[S] = 50 * 1000 = 50{,}000$

$V[S] = 50 * E[S^2] = 50 * \{V[X] + (E[X])^2\} = 50 * \{200^2 + 1000^2\} =$ 52,000,000

$SD[S] = 7{,}211.10$

5 (i) $$E\left[S\right] = 2\lambda \sum_{i=1}^{5} \frac{i}{\lambda} = 30$$

$$V\left[S\right] = 4\lambda^2 \sum_{i=1}^{5} \frac{i}{\lambda^2} = 60$$

(ii) $M_{X_i}(t) = \left(1 - \frac{t}{\lambda}\right)^{-i}$ (from book of formulae)

$$M_S(t) = E\left[\exp(tS)\right] = E\left[\exp\left(2\lambda t \sum_{i=1}^{5} X_i\right)\right] = \prod_{i=1}^{5} E\left[\exp(2\lambda t X_i)\right]$$

$$= \prod_{i=1}^{5} M_{X_i}(2\lambda t) = \prod_{i=1}^{5} (1-2t)^{-i} = (1-2t)^{-15}$$ so $S \sim \chi^2$, with 30 df

(iii) χ^2_{30} has mean 30 and variance 60 , as found in part (i).

6 (i) $Cov[S,D] = Cov[X+Y\,,\,X-Y] = Cov[X,X] - Cov[X,Y] + Cov[Y,X] - Cov[Y,Y]$
$= V[X] - V[Y] = 4V[Y]$

(ii) $V[S] = V[X] + V[Y] + 2Cov[X,Y] = 8V[Y]$

$V[D] = V[X] + V[Y] - 2Cov[X,Y] = 4V[Y]$

$\Rightarrow Corr[S,D] = 4V[Y]/\{8V[Y] \times 4V[Y]\}^{1/2} = +1/\sqrt{2} = +0.707$

7 (i) $1 - P[T1 \cup T2 \cup T3] = 1 - (0.1 + 0.02 + 0.003) = 1 - 0.123 = 0.877$

(ii) (a) $P[T1 \mid S] = \dfrac{P[T1 \cap S]}{P[S]} = \dfrac{P[S \mid T1]P[T1]}{P[S]} = \dfrac{0.5*0.1}{0.2} = 0.25$

(b) $P[T1 \cup T2 \cup T3 \mid S] = \dfrac{1}{P[S]}\ (P[T1 \cap S] + P[T2 \cap S] + P[T3 \cap S])$

$= \dfrac{1}{P[S]}\ (P[S \mid T1]\ P[T1] + P[S \mid T2]\ P[T2] + P[S \mid T3]\ P[T3])$

$= \dfrac{1}{0.2}(0.5*0.1 + 0.7*0.02 + 0.9*0.003) = \dfrac{0.0667}{0.2} = 0.3335$

(iii) $P[\mathrm{T}1 \cup T2 \cup T3 \mid S^C] = \dfrac{1}{0.8}(\mathrm{P}[\mathrm{T}1 \cup T2 \cup T3] - P[\{T1 \cup T2 \cup T3\} \cap S])$

$= \dfrac{1}{0.8}(0.123 - 0.5*0.1 - 0.7*0.02 - 0.9*0.003) = \dfrac{0.0563}{0.8} = 0.0704$

8 (i) $\bar{x}_A = \dfrac{2972}{6} = 495.33 \qquad s_A^2 = \dfrac{1}{5}\{1530284 - \dfrac{2972^2}{6}\} = 11630.67$

$\bar{x}_B = \dfrac{2791}{6} = 465.17 \qquad s_B^2 = \dfrac{1}{5}\{1343205 - \dfrac{2791^2}{6}\} = 8984.97$

$s_P^2 = \dfrac{5(11630.67) + 5(8984.97)}{10} = 10307.82 \qquad \therefore s_P = 101.527$

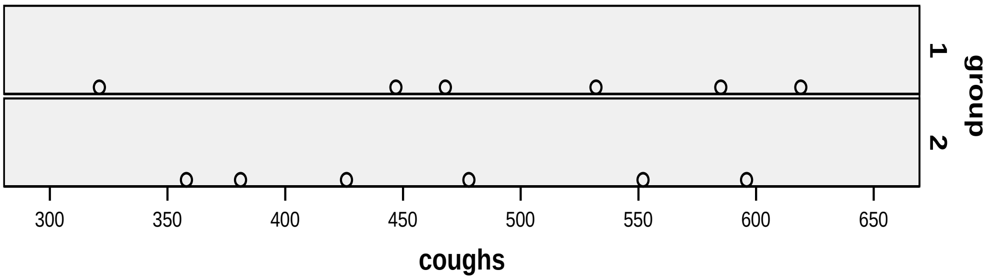
$t = \dfrac{495.33 - 465.17}{101.527\sqrt{\dfrac{1}{6} + \dfrac{1}{6}}} = \dfrac{30.16}{58.62} = 0.51 \quad \text{on } 10 \text{ df}$

without needing to look up tables (although candidates can do so, e.g. $t_{10}(2.5\%) = 2.228$)

there is clearly no evidence of a difference between medications *A* and *B* as regards their effectiveness for the relief of coughing.

(ii) (a) As the 12 patients were split <u>at random</u> into the two groups, the two samples are independent.
(Valid comments on the *need* for this assumption will also receive full credit.)

(b) The most appropriate graphical representation is two dotplots (or boxplots):



These show that there is nothing that suggests lack of normality in each case.
(Valid comments on the *need* for this assumption will also receive full credit.)

(c) $$F = \frac{s_A^2}{s_B^2} = \frac{11630.67}{8984.97} = 1.29 \quad \text{on } 5,5 \text{ df}$$

$F_{5,5}(10\%) = 3.453$. So no evidence against the assumption of equal variances.

(iii) $\Sigma x = 2972 + 2791 + 4151 = 9914$,
$\Sigma x^2 = 1530284 + 1343205 + 2933001 = 5806490$

$$SS_T = 5806490 - \frac{9914^2}{18} = 346079$$

$$SS_B = \frac{1}{6}(2972^2 + 2791^2 + 4151^2) - \frac{9914^2}{18} = 181800$$

$$SS_R = SS_T - SS_B = 164279$$

giving the ANOVA table:

Source of variation	*df*	*SS*	*MSS*
Between groups	2	181800	90900
Residual	15	164279	10952
Total	17	346079	

$$F = \frac{90900}{10952} = 8.30 \quad \text{on } 2,\ 15 \text{ df}$$

$F_{2,15}(5\%) = 3.682$ and $F_{2,15}(1\%) = 6.359$. So P-value < 0.01

So there is very strong evidence of a difference between medications *A* and *B* and the placebo as regards their effectiveness for the relief of coughing.

(iv) It would appear that both medications have a more beneficial effect on the level of coughing as compared to the placebo, but that they are equally beneficial.

9 (i) (a) With n large we use normal approximation to t_{100}.

$$416 \pm 1.96 \frac{72}{\sqrt{101}}$$

$$= 416 \pm 14.04 = (402.0,\ 430.0)$$

(b) Using $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{100}$

a 95% CI for σ^2 is $\frac{(n-1)S^2}{\chi^2_{100}(0.025)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{100}(0.975)}$

which gives $\left(\frac{100 \times 72^2}{129.6}, \frac{100 \times 72^2}{74.22} \right) = (4000,\ 6985).$

95% CI for standard deviation σ is therefore

$\left(\sqrt{4000}, \sqrt{6985} \right) = (63.2,\ 83.6).$

(ii) (a) 95% CI is $\bar{x} \pm 1.96\dfrac{s}{\sqrt{n}}$ and with $s = 72$ we have

$$\frac{1.96 \times 72}{\sqrt{n}} = 10 \Rightarrow n = 199.15$$

So $n \geq 200$.

(b) Taking s = 83.57 gives

$$\frac{1.96 \times 83.57}{\sqrt{n}} = 10 \Rightarrow n = 268.30, \text{ , so } n \geq 269.$$

(iii) Assuming a larger value of s results in a larger standard error, so a larger sample size is required to achieve the same width of confidence interval.

10 (i) $\hat{\mu} = \bar{X} = 1,175$

(ii) (a) $S_{tt} = 6900 - \dfrac{160^2}{4} = 500$

$$S_{xx} = 613379 - \frac{1179^2}{4} = 265,868.75$$

$$S_{tx} = 57515 - \frac{160 * 1179}{4} = 10,355$$

$$Corr(t,x) = \frac{S_{tx}}{\sqrt{S_{tt} * S_{xx}}} = \frac{10355}{\sqrt{500 * 265868.75}} = 0.898114$$

This implies that there is a strong linear relationship between age and number of deaths.

(b) Model: $\hat{x} = \hat{\alpha} + \hat{\beta} t$

$$\hat{\beta} = \frac{S_{tx}}{S_{tt}} = \frac{10355}{500} = 20.71,$$

$$\hat{\alpha} = \bar{x} - \hat{\beta}\bar{t} = \frac{1179}{4} - 20.71 * \frac{160}{4} = -533.65$$

Estimated model: $\hat{x} = 20.71t - 533.65$

(iii) $p(x_i, w, t_i) = \dfrac{\exp(-wt_i)(wt_i)^{x_i}}{x_i!}$

$\log p(x_i, w, t_i) = -wt_i + x_i(\log w + \log t_i) - \log(x_i!)$

$$\frac{\partial}{\partial w}\log p(x_i, w, t_i) = -t_i + \frac{x_i}{w}$$
$$\Sigma\frac{\partial}{\partial w}\log p(x_i, w, t_i) = -\Sigma t_i + \frac{1}{w}\Sigma x_i = 0$$

$$\hat{w} = \frac{\Sigma x_i}{\Sigma t_i}$$

(Second derivative gives $-\sum x_i / w^2 < 0$ which confirms maximum.)

For the observed values we obtain $\hat{w} = \dfrac{1179}{160} = 7.36875$

END OF EXAMINERS' REPORT