EXAMINERS REPORT

April 2010 examinations

Subject CT6 — Statistical Methods 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.

R D Muckart
Chairman of the Board of Examiners

July 2010

Comments

These are given in italics at the end of each question.

1 $$P(p=0.4|4H)=\frac{P(4H|p=0.4)P(p=0.4)}{P(4H)}$$

But $P(4H)$ $= P(4H|p=0.4)P(p=0.4)+P(4H|p=0.75)P(p=0.75)$

$$= \binom{6}{4}0.4^4 0.6^2\times 0.6+\binom{6}{4}0.75^4 0.25^2\times 0.4$$

= 0.082944 + 0.11865

= 0.201596

So $P(p=0.4|4H)=\frac{0.082944}{0.201596}=0.411436$

So the posterior distribution of p is given by $P(p = 0.4) = 0.411436$ and $P(p = 0.75) = 0.588564$

Comment: *This question was intended to be a straightforward application of Bayes' Theorem. However, the question was generally not well answered, with many candidates unable to find the posterior distribution in this slightly unfamiliar scenario.*

2 (i) The link function here is $g(\mu)=\log\mu$.

(ii) (a) The linear predictor is $\alpha_i+\beta x$ where the intercept α_i for $i=1,2$ depends on gender.

(b) The linear predictor is $\alpha_i+\beta_i x$ so that both parameters depend on gender.

Comment: *This straightforward question was generally well answered.*

3 (i) **Model 1**

In general we have $\gamma_k=\alpha\gamma_{k-4}$

Taking covariance with $Y_t,Y_{t-1},Y_{t-2},Y_{t-3}$ we get:

$\gamma_0=\alpha\gamma_4$
$\gamma_1=\alpha\gamma_3$
$\gamma_2=\alpha\gamma_2$
$\gamma_3=\alpha\gamma_1$

For $\alpha\neq 0$ these equations imply that $\rho_k=0$ unless k is divisible by 4.

So we have $\rho_{4k} = \alpha^k$ and all other autocorrelations are zero.

Model 2

Here we have $\gamma_k = 0$ unless *k*=4 and k=0. In these cases

$$\gamma_0 = cov(e_t + \beta e_{t-4}, e_t + \beta e_{t-4}) = (1+\beta^2)\sigma^2$$

$$\gamma_4 = cov(e_t + \beta e_{t-4}, e_{t-4} + \beta e_{t-8}) = \beta\sigma^2.$$

So $\rho_0 = 1$, $\rho_4 = \frac{\beta}{1+\beta^2}$ and all the other autocorrelations are zero.

(ii) Model 1 is preferred in situations where the sample auto-correlation is non-zero and decays exponentially.

Comment: *This question was reasonably well attempted, although a number of candidates dropped marks especially in calculating the ACF of Model 1.*

4 (i) Let *X* represent the distribution of individual claims. Let π denote the probability that an individual claim involves the reinsurer. Then

$$\pi = P(X > M) = \int_M^\infty \lambda e^{-\lambda x} dx$$

$$= \left[-e^{-\lambda x}\right]_M^\infty$$

$$= e^{-\lambda M}$$

(ii) $M_{I_i}(t) = E(e^{tI_i}) = \pi e^t + 1 - \pi = e^{t-\lambda M} + 1 - e^{-\lambda M}$

(iii) Using the results for the moment generating function of a compound distribution, we have

$$M_K(t) = M_N(\log M_{I_i}(t))$$

$$= \left(pM_{I_i}(t) + 1 - p\right)^n$$

$$= \left(p(\pi e^t + 1 - \pi) + 1 - p\right)^n$$

$$= \left(p\pi e^t + p - p\pi + 1 - p\right)^n$$

$$= \left(p\pi e^{t} + 1 - p\pi\right)^{n}$$

$$= \left(pe^{t-\lambda u} + 1 - pe^{-\lambda u}\right)^{n}$$

Which is the MGF of a binomial distribution with parameters n and $p\pi$.

Hence, by the uniqueness of MGFs K has a binomial distribution with parameters n and $p\pi$.

Comment: *This question was well answered.*

5 (i) $E(N) = E\left[E(N|q)\right]$

$$= E\left[1000q\right]$$
$$= 1000 \times \left(0.5 \times 0.001 + 0.5 \times 0.005\right)$$
$$= 1000 \times 0.003 = 3$$

$$Var(N) = Var\left[E(N|q)\right] + E\left[Var\left(N|q\right)\right]$$

$$= Var(1000q) + E(1000q(1-q))$$

Now $E(q) = 0.003$ and $E(q^2) = 0.5 \times 0.001^2 + 0.5 \times 0.005^2 = 0.000013$

So $Var(q) = 0.000013 - 0.003^2 = 0.000004$

$$E(q(1-q)) = 0.5 \times 0.001 \times 0.999 + 0.5 \times 0.005 \times 0.995 = 0.002987$$

So $Var(N) = 1000^2 \times 0.000004 + 1000 \times 0.002987 = 6.987$

(ii) In this case $N \sim B(1000, 0.003)$ and so

$$E(N) = 1000 \times 0.003 = 3$$

and

$$Var(N) = 1000 \times 0.003 \times 0.997 = 2.991$$

(iii) The simplification in (ii) results in the same mean number of deaths, but a very significantly lower variance.

This is because in (i) there is a tendency for deaths to occur at the same time, or not at all (as a result of the weather) whereas in (ii) deaths are genuinely independent.

Comment: *Many good answers here although some candidates struggled to articulate why the variance in (ii) was lower.*

6 (i) We can deduce that $Y_t = at + \sum_{i=1}^{t} e_i$

and so $E(Y_t) = at$ and
$Var(Y_t) = t\sigma^2$.

Since these expressions depend on *t* the process is not stationary.

(ii) As $s < t$ we have

$$Cov(Y_t, Y_{t-s}) = Cov(at + \sum_{i=1}^{t} e_i, as + \sum_{j=1}^{t-s} e_j) = Var(\sum_{j=1}^{t-s} e_j) = (t-s)\sigma^2$$

Which is linear in *s* as required.

(iii) First note that the differenced series:

$$X_t = Y_t - Y_{t-1} = a + e_t$$

is essentially a white noise process. So estimates of *a* and σ^2 can be found by constructing the sample differences series $x_i = y_i - y_{i-1}$ for $i = 1, 2, \ldots, n$ and taking the mean and sample variance(or its square for estimating σ) respectively.

(iv) In this case $\hat{y}_n(1) = \hat{a} + y_n + 0 = \hat{a} + y_n$

And $\hat{y}_n(2) = \hat{a} + \hat{y}_n(1) + 0 = 2\hat{a} + y_n$

Comment: *This was a rather hard question with many candidates confirming the non-stationarity in (i) but not finding the general solution. In part (ii) a good number of answers failed to score full marks. Alternative answers to (i) and (ii) could have been obtained by increasing t iteratively and noticing a pattern developing.*

7 (i) $$1 = \int_0^c ae^{-\lambda x}dx = \left[-\frac{a}{\lambda}e^{-\lambda x}\right]_0^c = \frac{a}{\lambda}\left(1-e^{-\lambda c}\right)$$

So $a = \dfrac{\lambda}{1-e^{-\lambda c}}$

(ii) The distribution function is

$$F(x) = \int_0^x ae^{-\lambda y}dy = \left[-\frac{a}{\lambda}e^{-\lambda y}\right]_0^x$$

$$= \frac{a}{\lambda}\left(1-e^{-\lambda x}\right) = \frac{1-e^{-\lambda x}}{1-e^{-\lambda c}}$$

The required transformation is therefore given by:

$$u = F(x) \Leftrightarrow 1-e^{-\lambda x} = u(1-e^{-\lambda c}) \Leftrightarrow x = -\frac{\log\left(1-u(1-e^{-c\lambda})\right)}{\lambda}$$

So the algorithm is:

- Generate u from U(0, 1)
- Set $x = -\dfrac{\log\left(1-u(1-e^{-c\lambda})\right)}{\lambda}$

(iii) We need to find $M = \underset{0<x<c}{Max}\dfrac{g(x)}{f(x)} = \underset{0<x<c}{Max}\dfrac{(1-e^{-c\lambda})e^{-x^2/2+\lambda x}}{\sqrt{2\pi}[\Phi(c)-0.5]\lambda}$

Let $h(x) = e^{-x^2/2+\lambda x}$ then we can simply find the maximum of *h(x)* since it differs only by a constant.

Then $h'(x) = h(x)\times(-x+\lambda)$

So $h'(x) = 0$ when $x = \lambda$

$$h''(x) = h'(x)(\lambda - x) - h(x)$$

And so $h''(\lambda) = -h(\lambda) < 0$ so we do have a maximum.

Hence $M = \dfrac{(1-e^{-c\lambda})e^{\lambda^2/2}}{\sqrt{2\pi}[\Phi(c)-0.5]\lambda}$

And so $\dfrac{g(x)}{Mf(x)} = e^{-x^2/2+\lambda x-\lambda^2/2}$

So the algorithm is:

- Generate u from U(0,1)
- Set $x = -\dfrac{\log\left(1-u(1-e^{-c\lambda})\right)}{\lambda}$
- Generate v from U(0,1)
- If $v < e^{-x^2/2+\lambda x-\lambda^2/2}$ return x as the random sample, otherwise begin again.

Comment: *This was found particularly hard especially part (iii), where some harder calculations are involved. Many candidates just described the general theory of rejection algorithms without being able to apply it explicitly here.*

8 (i) Adjusting the incremental data for inflation to mid 2008 prices gives:

	Development Year			
Accident Year	*0*	*1*	*2*	*3*
2006	114.3	34.2	29.5	13
2007	93.9	21.4	16	
2008	112.0	35		
2009	132			

Cumulating gives

	Development Year			
Accident Year	*0*	*1*	*2*	*3*
2006	114.3	148.5	178.0	191.0
2007	93.9	115.3	131.3	
2008	112.0	147.0		
2009	132.0			

The development factors are:

Year 0 to year 1 $\dfrac{148.5+115.3+147.0}{114.3+93.9+112.0} = 1.2827$

Year 1 to year 2 $\dfrac{178.0+131.3}{148.5+115.3} = 1.1726$

Year 2 to year 3 $\dfrac{191.0}{178.0} = 1.0730$

The completed table at mid 2008 prices is:

		Development Year		
Accident Year	*0*	*1*	*2*	*3*
2006				
2007				140.9
2008			172.4	185.0
2009		169.3	198.5	213.0

Differencing gives:

		Development Year		
Accident Year	*0*	*1*	*2*	*3*
2006				
2007				9.6
2008			25.4	12.6
2009		37.3	29.2	14.5

And so the total reserve is

$$(37.3 + 25.4 + 9.6) \times 1.03 + (29.2 + 12.6) \times 1.03^2 + 14.5 \times 1.03^3 = 134.7$$

Comment: *This was a straightforward question with many candidates scoring full marks here.*

9 (i) The possibilities are (where H denotes trying a hot chilli and M denotes trying a mild chilli)

$d_1(H) = 4H$ and $d_1(M) = 4H$
$d_2(H) = 4H$ and $d_2(M) = 8H$
$d_3(H) = 8H$ and $d_3(M) = 4H$
$d_4(H) = 8H$ and $d_4(M) = 8H$

(ii) Under 4H we have $P(H) = 0.2$ and $P(M) = 0.8$
Under 8H we have $P(H) = 0.4$ and $P(M) = 0.6$

We can present the game so that the loss to the actuary is what he has to pay for the plate of chillis (i.e. the loss is either 2.5 or 5). Under this approach we have:

$$R(d_1,4H) = P(H|4H)\times L(d_1(H),4H) + P(M|4H)\times L(d_1(M),4H) = 0.2\times 2.5 + 0.8\times 2.5 = 2.5$$
$$R(d_1,8H) = P(H|8H)\times L(d_1(H),8H) + P(M|8H)\times L(d_1(M),8H) = 0.4\times 5 + 0.6\times 5 = 5$$
$$R(d_2,4H) = P(H|4H)\times L(d_2(H),4H) + P(M|4H)\times L(d_2(M),4H) = 0.2\times 2.5 + 0.8\times 5 = 4.5$$
$$R(d_2,8H) = P(H|8H)\times L(d_2(H),8H) + P(M|8H)\times L(d_2(M),8H) = 0.4\times 5 + 0.6\times 2.5 = 3.5$$
$$R(d_3,4H) = P(H|4H)\times L(d_3(H),4H) + P(M|4H)\times L(d_3(M),4H) = 0.2\times 5 + 0.8\times 2.5 = 3$$
$$R(d_3,8H) = P(H|8H)\times L(d_3(H),8H) + P(M|8H)\times L(d_3(M),8H) = 0.4\times 2.5 + 0.6\times 5 = 4$$
$$R(d_4,4H) = P(H|4H)\times L(d_4(H),4H) + P(M|4H)\times L(d_4(M),4H) = 0.2\times 5 + 0.8\times 5 = 5$$
$$R(d_4,8H) = P(H|8H)\times L(d_4(H),8H) + P(M|8H)\times L(d_4(M),8H) = 0.4\times 2.5 + 0.6\times 2.5 = 2.5$$

(iii) The payoff matrix for the player is:

	d_1	d_2	d_3	d_4
4H	2.5	4.5	3	5
8H	5.0	3.5	4	2.5
Expected Loss	3.75	4.0	3.5	3.75

So the Bayes criterion strategy is d_3.

Under this approach, the average price paid is £3.50.

Comment: *Alternative solutions are possible here. The question was not answered as well as those relating to the same material in previous years. In particular, many weaker candidates were unable to fully specify the possible decision functions and therefore made little headway with this question.*

10 (i) Let X denote the individual claim amounts net of re-insurance. Then $X = 100\alpha$ and $M_X(t) = e^{100\alpha t}$.

The insurer's annual net premium income is

$$100\times\lambda\times 1.15 - 100\times(1-\alpha)\times\lambda\times 1.2 = \lambda(120\alpha - 5)$$

So the adjustment coefficient R satisfies

$$\lambda + \lambda(120\alpha - 5)R = \lambda e^{100\alpha R}$$

That is $1 + (120\alpha - 5)R = e^{100\alpha R}$

(ii) Differentiating this equation with respect to α we get

$$120R + (120\alpha - 5)\frac{dR}{d\alpha} = \frac{d}{d\alpha}\left[e^{100\alpha R}\right]$$

and $\frac{d}{d\alpha}\left[e^{100\alpha R}\right] = e^{100\alpha R}\frac{d}{d\alpha}\left[100\alpha R\right]$

$$= e^{100\alpha R}\left(100R + 100\alpha\frac{dR}{d\alpha}\right)$$

So putting these together, we have:

$$(120\alpha - 5)\frac{dR}{d\alpha} + 120R = \left(100R + 100\alpha\frac{dR}{d\alpha}\right)e^{100\alpha R}$$

(iii) Firstly, by Lundberg's inequality the higher the value of R the lower the upper bound on the probability of ruin.

So we wish to choose α so that R is a maximum.

That is, we need $\frac{dR}{d\alpha} = 0$.

(iv) Putting $\frac{dR}{d\alpha} = 0$ in the equation in (ii) we get

$$120R = 100\,\mathrm{R}e^{100\alpha R}$$

i.e. $e^{100\alpha R} = 1.2$

i.e. $R = \frac{\log 1.2}{100\alpha}$

substituting into the equation in (i) gives

$$1 + (120\alpha - 5)\times\frac{\log 1.2}{100\alpha} = 1.2$$

i.e. $100\alpha + (120\alpha - 5)\log 1.2 = 120\alpha$

$$-20\alpha + 120\alpha\log 1.2 = 5\log 1.2$$

So $\alpha = \frac{5\log 1.2}{120\log 1.2 - 20} = 0.48526$

Comment: *This question on material new to the syllabus for 2010 was not answered well with many candidates struggling, especially in the last part.*

11 (i) The overall mean is given by $\bar{Y} = \dfrac{113+93+82+144}{4} = 108$

$$E\left(s^2(\theta)\right) = \frac{1}{4}\sum_{i=1}^{4}\left(\frac{1}{2}\sum_{j=1}^{3}(Y_{ij}-\bar{Y}_i)^2\right) = \frac{362+206+86+224}{8} = 109.75$$

$$Var(m(\theta)) = \frac{1}{3}\sum_{i=1}^{4}\left(\bar{Y}_i - \bar{Y}\right)^2 - \frac{1}{3}E(S^2(\theta))$$

$$= \frac{(113-108)^2+(93-108)^2+(82-108)^2+(144-108)^2}{3} - \frac{109.75}{3}$$

$$= 704.083$$

So the credibility factor is $Z = \dfrac{3}{3+\dfrac{109.75}{704.083}} = 0.950608$

And the estimate for next quarter is

$$0.950608\times 113 + (1-0.950608)\times 108 = 112.75$$

(ii) The average number of pieces of mail is assumed to be growing each year. We need to adjust the data to take account of this. Two approaches are:

- Convert the data into "Year 4" values by increasing by 10% p.a. and then applying the methodology above; OR

- Recognise the lower volume of data in earlier years, by applying a risk volume to each year and using EBCT model 2. If the risk volume for year 4 is 1, then the risk volume for year 3 is 1/1.1 and year 2 is 1/1.21 etc.

(iii) Let the mean number of items in quarter 1 of year 1 be given by λ. Then the likelihood is given by:

$$L \propto e^{-\lambda}\lambda^{Y_{11}}e^{-1.1\lambda}(1.1\lambda)^{Y_{12}}e^{-1.1^2\lambda}(1.1^2\lambda)^{Y_{13}}$$

And so the log likelihood is

$$l = \log L = C - \lambda(1+1.1+1.1^2) + (Y_{11}+Y_{12}+Y_{13})\log\lambda$$

Differentiating $\frac{\partial l}{\partial \lambda} = -(1+1.1+1.1^2) + \frac{Y_{11}+Y_{12}+Y_{13}}{\lambda}$

And setting this equal to zero gives:

$$\hat{\lambda} = \frac{Y_{11}+Y_{12}+Y_{13}}{1+1.1+1.1^2} = \frac{98+117+124}{3.31} = 102.417$$

So the estimate for Q1 in year 4 is $1.1^3 \times \hat{\lambda} = 1.331 \times 102.417 = 136.32$

(iv) The main difference is that the maximum likelihood estimate approach considers the data for Q1 in isolation, whereas the EBCT approach assumes that data from other quarters come from a related distribution and so can tell us something about Q1.

Specifically, the EBCT approach assumes that the mean volume of unsolicited mail for each quarter is itself a sample from a common distribution. Hence whilst each quarter has a different mean, they provide some information about the population from which the mean is drawn.

Comment: *The same comment as in the previous question is valid here. This question was not very well answered question but with various alternative answers in (ii) and (iv).*

END OF EXAMINERS' REPORT