INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

April 2011 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.

T J Birse
Chairman of the Board of Examiners

July 2011

1 Example 1: Testing sensitivity to parameter variation – we want the results to change as a result of changes to the parameter not as a result of variations in the random numbers.

Example 2: Performance evaluation. When comparing two or more schemes which might be adopted we want differences in results to arise from differences between the schemes rather than as a result of variations in the random numbers.

Alternative Example: The same set of simulations could be used for the numerical evaluation of derivatives

$$\theta'(\alpha) = \frac{\theta(\alpha+\delta) - \theta(\alpha)}{\delta}$$

This question was generally answered well, although weaker candidates explained how to obtain random numbers or perform Monte-Carlo simulation instead of explaining why you would want to use the <u>*same*</u> *random numbers.*

2 (i) $E(s^2(\theta))$ represents the average variability of claim numbers from year to year for a single risk.

$V(m(\theta))$ represents the variability of the average claim numbers for different risks i.e. the variability of the means from risk to risk.

(ii) The credibility factor is given by

$$Z = \frac{n}{n + \dfrac{E(s^2(\theta))}{V(m(\theta))}}$$

We can see that it is the relative values of $E(s2(\theta))$ and $V(m(\theta))$ that matter. In particular, if $E(s2(\theta))$ is high relative to $V(m(\theta))$, this means that there is more variability from year to year than from risk to risk. More credibility can be placed on the data from other risks leading to a lower value of Z.

On the other hand, if $V(m(\theta))$ is relatively higher this means there is greater variation from risk to risk, so that we can place less reliance on the data as a whole leading to a higher value of Z.

This question was generally well answered. Weaker students were not able to give clear, concise descriptions of the quantities in part (i).

3 (i) First note that $f(y_1|\theta) = \frac{1}{\theta}$ for $\theta > y_1$

$$f(\theta|y_1) \propto f(y_1|\theta)\, f(\theta)$$

$$= \frac{1}{\theta}\,\alpha\beta^\alpha\,\theta^{-(1+\alpha)} \text{ for } \theta > \beta \text{ and } \theta > y_1$$

$$\propto \theta^{-(2+\alpha)} \text{ for } \theta > \max(\beta, y_1)$$

$$\propto (\alpha + 1)\,\bar{\beta}^{\alpha+1}\,\theta^{-(1+1+\alpha)} \text{ for } \theta > \bar{\beta} \text{ where } \bar{\beta} = \max(\beta, y_1)$$

Which is of the same form with parameters $\alpha + 1$ and $\bar{\beta}$.

Alternatively, we can derive this formally as:

$$f(\theta|y_1) = \frac{f(y_1|\theta)f(\theta)}{\int_\theta f(y_1|\theta)f(\theta)d\theta}$$

This gives:

$$f(\theta|y_1) = \frac{\frac{1}{\theta}\times\alpha\beta^\alpha\theta^{-(1+\alpha)}}{\int_\beta^\infty \alpha\beta^\alpha\theta^{-\alpha}d\theta} = \frac{\alpha\beta^\alpha\theta^{-(2+\alpha)}}{\frac{\alpha}{\alpha+1}\beta^{-1}} = (\alpha+1)\beta^{\alpha+1}\theta^{-(2+\alpha)}$$

(ii) The posterior distribution has the same form with parameters $\alpha + n$ and $\max(\beta, y_1, \ldots, y_n)$.

This question received a wide range of quality of answers. Only the strongest candidates stated the correct range for the posterior in part (i). A number of candidates assumed a sample of size n *in part (i) and therefore failed to differentiate between parts (i) and (ii).*

4 Let the annual number of claims be denoted by N. Then

$$P(N = k) = \int_0^\infty P(N = k|\mu)\, f(\mu)\, d\mu$$

$$= \int_0^\infty e^{-\mu}\frac{\mu^k}{k!}\lambda e^{-\lambda\mu}d\mu$$

$$= \frac{\lambda}{k!}\int_0^\infty \mu^k e^{-(1+\lambda)\mu}d\mu$$

$$= \frac{\lambda}{k!}\times\frac{\Gamma(k+1)}{(1+\lambda)^{k+1}}\int_0^\infty \frac{(1+\lambda)^{k+1}}{\Gamma(k+1)}\mu^k e^{-(1+\lambda)\mu}d\mu$$

$$= \frac{\lambda}{(1+\lambda)^{k+1}}\times 1$$

Where the final integral is 1 since the integrand is the pdf of a Gamma distribution.

So

$$P(N=k) = \frac{\lambda}{(1+\lambda)^{k+1}} = \frac{\lambda}{1+\lambda} \times \frac{1}{(1+\lambda)^k}, \text{ for k = 0, 1, 2, …}$$

Which means that N has a geometric distribution with parameter $p = \frac{\lambda}{1+\lambda}$. This is equivalent to a Type II negative binomial with k=1

This was the worst answered question on this paper, with few candidates able to write down the first integral. Candidates who attempted the algebra often did not recognise the resultant distribution.

5 Premiums charged to the policyholder are

$$1.5 \times (0.2 \times (0.5 \times 50 + 0.5 \times 20) + 0.4 \times (0.25 \times 100 + 0.5 \times 70 + 0.25 \times 40)) = 52.5$$

The completed table is

		Claims					
Arrangement	*Net Premium*	*None*	*1 L*	*1 S*	*1S 1L*	*2 S*	*2 L*
A	52.5	0	50	20	70	40	100
B	47.5	0	40	10	50	20	80
C	42.5	0	30	20	50	40	60

So the completed table of profits for the insurer is:

	Profit						
	None	*1 L*	*1 S*	*1S 1L*	*2 S*	*2 L*	*E(P)*
Probability	0.4	0.1	0.1	0.2	0.1	0.1	
A	52.5	2.5	32.5	–17.5	12.5	–47.5	17.5
B	47.5	7.5	37.5	–2.5	27.5	–32.5	22.5
C	42.5	12.5	22.5	–7.5	2.5	–17.5	17.5

So the insurer should introduce the policy excess (arrangement **B**)

This question was answered well. Some candidates gave incorrect completed tables without showing any working, and the examiners were therefore unable to give partial credit in these cases.

6 (i) We first note that the distribution function for this double exponential is given by

$$G(x) = \begin{cases} 0.5e^{\lambda x} & x < 0 \\ 0.5(1 - e^{-\lambda x}) + 0.5 & x \geq 0 \end{cases}$$

And so the inverse function is given by

$$G^{-1}(u) = \begin{cases} \dfrac{\log 2u}{\lambda} & u < 0.5 \\ -\dfrac{\log(2(1-u))}{\lambda} & u \geq 0.5 \end{cases}$$

And so our algorithm is:

(A) Generate u from $U(0,1)$

(B) If $u < 0.5$ set $x = \dfrac{\log 2u}{\lambda}$ otherwise set $x = -\dfrac{\log(2(1-u))}{\lambda}$

(ii) We must first find $M = \text{Sup}\dfrac{f(x)}{g(x)}$ where $f(x)$ is the pdf of the $N(0,1)$ distribution.

$$\text{Sup}\frac{f(x)}{g(x)} = \text{Sup}\frac{2}{\lambda\sqrt{2\pi}}e^{-\frac{x^2}{2}+\lambda|x|}$$

And using the symmetry around 0 we can concentrate on positive values of x

$$\text{Sup}\frac{f(x)}{g(x)} = \text{Sup}\frac{2}{\lambda\sqrt{2\pi}}e^{-\frac{x^2}{2}+\lambda x}$$

And the exponential expression is maximised when $-\dfrac{x^2}{2} + \lambda x$ is maximised.

Differentiating, this occurs when $-\lambda + x = 0$ i.e. $x = \lambda$

Hence $M = \text{Sup}\dfrac{f(x)}{g(x)} = \dfrac{2}{\lambda\sqrt{2\pi}}e^{\frac{\lambda^2}{2}}$. So define

$$h(x) = \frac{f(x)}{Mg(x)} = e^{-\frac{x^2}{2}+\lambda|x|-\frac{\lambda^2}{2}}$$

So algorithm is as follows:

(A) Generate x as in part (i)
(B) Generate u from $U(0,1)$
(C) If $u \le h(x)$ then set $y = x$ otherwise return to (A)

This question was poorly answered. Only the strongest candidates treated the modulus in the pdf correctly. There was also difficulty deriving both parts of the inverse function for both the $u \ge 0.5$ and the $u < 0.5$ case.

7 (i) The model is ARIMA(2,0,0) provided that the model is stationary.

(ii) The lag polynomial is $1 - 0.4L - 0.12L^2 = (1 - 0.6L)(1 + 0.2L)$

Since the roots $\frac{1}{0.6}$ and $-\frac{1}{0.2}$ are both greater than one in absolute value the process is stationary.

(iii) Since the process is stationary we know that $E(Y_t)$ is equal to some constant μ independent of t.

Taking expectations on both sides of the equation defining Y_t gives

$E(Y_t) = 0.7 + 0.4E(Y_{t-1}) + 0.12E(Y_{t-2})$

$\mu = 0.7 + 0.4\mu + 0.12\mu$

$$\mu = \frac{0.7}{1-0.4-0.12} = 1.45833333$$

(iv) The auto-covariance function is not affected by the constant term of 0.7 in the equation, and this term can be ignored.

The Yule-Walker equations are

$\gamma_0 = 0.4\gamma_1 + 0.12\gamma_2 + \sigma^2$

$\gamma_1 = 0.4\gamma_0 + 0.12\gamma_1$ (A)

$\gamma_2 = 0.4\gamma_1 + 0.12\gamma_0$ (B)

$\gamma_s = 0.4\gamma_{s-1} + 0.12\gamma_{s-2}$ for $s > 2$

Dividing both sides of (A) by γ_0 and noting that $\rho_s = \frac{\gamma_s}{\gamma_0}$ we have

$\rho_1 = 0.4 + 0.12\rho_1$ so that $\rho_1 = \frac{0.4}{0.88} = 0.45454545$

Substituting this result into (B) we have

$\rho 2 = 0.4 \times 0.45454545 + 0.12 = 0.30181818$

And using the final result we have

$\rho 3 = 0.4 \times 0.3018181818 + 0.12 \times 0.45454545 = 0.1752727$

and

$\rho 4 = 0.4 \times 0.1752727 + 0.12 \times 0.30181818 = 0.10632727$

Expressed as fractions:

$\rho_1 = \frac{5}{11} \quad \rho_2 = \frac{83}{275} \quad \rho_3 = \frac{241}{1375} \quad \rho_4 = \frac{731}{6875}$

This straightforward question was answered well.

8 (i) $MY(t) = \int e^{ty} f(y, \theta, \varphi)\, dy$

$$= \int \exp(ty) \exp\left[\frac{y\theta - b(\theta)}{\varphi} + c(y, \varphi)\right] dy$$

$$= \int \exp\left[\frac{y(\theta + t\varphi) - b(\theta)}{\varphi} + c(y, \varphi)\right] dy$$

$$= \exp\left[\frac{b(\theta + t\varphi) - b(\theta)}{\varphi}\right] \int \exp\left[\frac{y(\theta + \varphi t) - b(\theta + \varphi t)}{\varphi} + c(y, \varphi)\right] dy$$

$$= \exp\left[\frac{b(\theta + t\varphi) - b(\theta)}{\varphi}\right] \times 1$$

using the hint to evaluate the second integral.

(ii) $$\frac{dM_Y(t)}{dt} = \frac{d}{dt}\left(\frac{b(\theta + t\varphi) - b(\theta)}{\varphi}\right) M_Y(t)$$

$$= \frac{\varphi b'(\theta + t\varphi)}{\varphi} M_Y(t)$$

$$= b'(\theta + t\varphi) M_Y(t)$$

And $E(Y) = M_Y'(0) = b'(\theta)M_Y(0) = b'(\theta)\times 1 = b'(\theta)$

$$\frac{d^2M_Y(t)}{dt^2} = \varphi b''(\theta+t\varphi)M_Y(t) + b'(\theta+t\varphi)M_Y'(t)$$

So $E(Y2) = M_Y''(0) = \varphi b''(\theta)M_Y(0) + b'(\theta)M_Y'(0)$

$$= \varphi b''(\theta) + b'(\theta)^2$$

So $\text{Var}(Y) = E(Y2) - E(Y)2 = \varphi b''(\theta) + b'(\theta)^2 - b'(\theta)^2 = \varphi b''(\theta)$

Credit given for alternative approaches (e.g. CGF).

(iii) For the Poisson distribution with parameter μ we have

$$f(y, \theta, \varphi) = \frac{\mu^y e^{-\mu}}{y!} = \exp[y \log \mu - \mu - \log y!]$$

Which is of the form in the question with $\theta = \log \mu$, $\varphi = 1$ and $b(\theta) = e\theta$ and $c(y, \varphi) = -\log y!$

So the result from (i) gives

$$MY(t) = \exp\left[\frac{b(\theta+t\varphi)-b(\theta)}{\varphi}\right]$$

$$= \exp\left[\frac{e^{\log\mu+t} - e^{\log\mu}}{1}\right] = \exp[\mu(et - 1)]$$

which is indeed the MGF of the Poisson distribution as shown on p7 of the tables.

This question was generally well done, with many candidates who could not complete the derivation in part (i) nevertheless able to use the result to score well in parts (ii) and (iii). For part (iii) some candidates calculated the first two moments rather than showing that the MGF of the Poisson distribution has the form given in part (i).

9 (i) The adjustment coefficient is the unique positive solution to

$\lambda MX(R) - \lambda - \lambda(1 + \theta)\, E(X)\, R = 0$

(ii) Cancelling the λ terms we have

(a) $MX(R) = E(eRX) = 1 + (1 + \theta)\, E(X)R$

$$E\left(1 + RX + \frac{R^2X^2}{2} + \ldots\right) = 1 + (1 + \theta)\, E(X)R$$

And truncating the expression we get

$E(1 + RX + R2X2/2) = 1 + (1 + \theta)\, E(X)R$

i.e. $1 + Rm1 + R2m2/2 = 1 + (1 + \theta)\, m1R$

i.e. $R2m2 = 2\theta m1R$

i.e. $R = \dfrac{2\theta m_1}{m_2}$

(b) Once more we have

$$E\left(1 + RX + \frac{R^2X^2}{2!} + \frac{R^3X^3}{3!} + \ldots.\right) = 1 + (1 + \theta)\, E(X)R$$

And truncating the expression we get

$$E\left(1 + RX + \frac{R^2X^2}{2} + \frac{R^3X^3}{6}\right) = 1 + (1 + \theta)\, E(X)R$$

i.e. $1 + Rm_1 + \dfrac{R^2m_2}{2} + \dfrac{R^3m_3}{6} = 1 + (1 + \theta)m1R$

i.e. $3R2m2 + R3m3 = 6\theta m1R$

i.e. $m3R2 + 3Rm2 - 6\theta m1 = 0$

As required

(iii) In this case $m1 = 10$ and $m2 = 200$ and $m3 = 6000$

So the estimate from (ii) (a) is $R = \dfrac{2\theta m_1}{m_2} = \dfrac{2\times 0.3\times 10}{200} = \dfrac{6}{200} = 0.03$

The estimate from (ii) (b) is the solution to $6000R2 + 600R - 18 = 0$

Which is given by

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-600 + \sqrt{600^2 + 4\times 6000\times 18}}{12000} = 0.024161984$$

[The negative root of the equation is –0.12416]

The true value of R is given by the solution to

$MX(R) = E(eRX) = 1 + (1 + \theta)\, E(X)R$

That is $\dfrac{\mu}{\mu - R} = 1 + \dfrac{(1+\theta)R}{\mu}$ where $\mu = {}^{1}\!/_{10}$ is the parameter of the exponential distribution.

And so

$\mu2 = \mu(\mu - R) + (1 + \theta)\, R(\mu - R)$

$\mu2 = \mu2 - \mu R + \mu R - R2 + \mu\theta R - \theta R2$

$0 = \mu\theta - R(1 + \theta)$

$R = \dfrac{\mu\theta}{1+\theta} = \dfrac{0.1\times 0.3}{1.3} = 0.0230769$

So the first estimate gives a greater error than the second (the error is 30% for the first approximation and only about 4.7% for the second). This is as we would expect since we took more terms before truncating.

Candidates generally answered part (i) and part (ii) well, although part (ii) b) caused problems and many candidates did not give sufficient detail. In (iii) many candidates produced an answer wrongly using a denominator of 100, or calculated the estimate but then did not explain the difference in estimates adequately.

10 (i) Denote the insurers profits by Z

Under A:

Premium income = 200 × 40 × 1.4 = 11200

Expected claims = 200 × 40 = 8000

So $E(Z)$ = 11200 − 8000 = 3200

Under B

We need first to calculate the expected loss for the insurer. Denote the insurer's loss by X. Then

$$E(X) = \int_0^{60} 0.025xe^{-0.025x}dx + 60 \times \int_{60}^{\infty} 0.025e^{-0.025x}dx$$

$$= [-xe^{-0.025x}]_0^{60} + \int_0^{60} e^{-0.025x}dx + 60 \times [-e^{-0.025x}]_{60}^{\infty}$$

$$= -60e^{-1.5} + [-40e^{-0.025x}]_0^{60} + 60e^{-1.5}$$

$$= 40 - 40e{-1.5} = 31.07479359$$

So the expected loss for the re-insurer is 40 − 31.07479359 = 8.925206406

Premium income = 11200 − 200 × 1.55 × 8.925206406 = 8433.186014

Expected claims = 200 × 31.07479359 = 6214.958718

So $E(Z)$ = 8433.186014 − 6214.958718 = 2218.227

Under C

Premium income = 200 × 40 × 1.4 − 200 × 40 × 0.25 × 1.45 = 8300

Expected claims = 200 × 40 × 0.75 = 6000

So $E(Z)$ = 8300 − 6000 = 2300

(ii) We now need to find the variance of the total claim amount paid by the insurer. Denote this by Y. Then

<u>Under A</u>

Var(Y) = 200Var(X) + 200E(X)2

= 200 × 402 + 200 × 402 = 640,000 = 8002

So

$$\Pr(Z < 2000) = \Pr(Y > 9200) = \Pr\left(N(0,1) > \frac{9200-8000}{800}\right)$$

= Pr(N(0,1) > 1.5) = (1 – 0.93319) = 0.06681

<u>Under B</u>

We first need to find E(X2) as defined above.

$$E(X2) = \int_0^{60} 0.025x^2 e^{-0.025x}dx + 60^2\int_{60}^{\infty} 0.025e^{-0.025x}dx$$

$$= [-x^2e^{-0.025x}]_0^{60} + \int_0^{60} 2xe^{-0.025x}dx + 3600e^{-1.5}$$

$$= -3600e^{-1.5} + \frac{2}{0.025}\int_0^{60} 0.025xe^{-0.025x}dx + 3600e^{-1.5}$$

$$= \frac{2}{0.025}(E(X) - 60e^{-1.5}) = 1414.958718$$

And so

$\text{Var(X)} = 1414.958718 - 31.07479359^2 = 449.3159219$

And therefore

Var(Y) = 200Var(X) + 200E(X)2

= 200 × 449.3159219 + 200 × 31.074793592 = 282991.7438

Finally

Pr(Z < 2000) = Pr(Y > 6433.186014)

$$= \Pr\left(N(0,1) > \frac{6433.186014 - 6214.958718}{531.97}\right)$$

= Pr(*N*(0,1) > 0.41023) = 1 – 0.65918 = 0.34082

<u>Under C</u>

Var(*Y*) = 200Var(*X*) + 200*E*(*X*)2

Var(*Y*) = 200 × 0.752 × 402 + 200 × (0.75 × 30)2 = 360000 = 6002

So

$$\Pr(Z < 2000) = \Pr(Y > 6300) = \Pr\left(N(0,1) > \frac{6300-6000}{600}\right)$$

= Pr(*N*(0,1) > 0.5) = (1 – 0.69146) = 0.30854

This question received a wide range of quality of answers. Most candidates calculated A and C correctly, but many failed to produce a reasonable answer for B. Common errors for A and C included not re-calculating the variance and using the wrong claim amount when calculating the probability. Some candidates were unable to calculate the normal distribution probability correctly after deriving the correct claim and variance values. For arrangement B many struggled to evaluate the integral correctly.

11 (i) The development factors are:

$$r0,1 = \frac{281.4+320+312.9}{240+260+270} = \frac{914.3}{770} = 1.187403$$

$$r1,2 = \frac{302+322}{281.4+320} = \frac{624}{601.4} = 1.037579$$

$$r2,3 = \frac{305}{302} = 1.009934$$

And the ultimate claims are:

For AY 2008: 322 × 1.009934 = 325.20

For AY 2009: 312.9 × 1.037579 × 1.009934 = 327.88

For AY 2010: 276 × 1.187403 × 1.037579 × 1.009934 = 343.42

So outstanding claims reserve is

325.20 + 327.88 + 343.42 – 322 – 312.9 – 276 = 85.60

(ii) Suppose that $R_{i,i+1} \sim \log N(\mu_i, \sigma_i^2)$ for i = 0, 1, 2.

Then $\log R_{i,i+1} \sim N(\mu_i, \sigma_i^2)$.

So (for example)

$\log R_{i,i+1} R_{i+1,i+2} = \log R_{i,i+1} + \log R_{i+1,i+2} \sim N(\mu_i + \mu_{i+1}, \sigma_i^2 + \sigma_{i+1}^2)$

Which means that the product $R_{i,i+1} R_{i+1,i+2}$ is also log-normally distributed. Since any product of log-normally distributed development factors is also log-normally distributed the development factors to ultimate must also be log-normally distributed.

(iii) Using the results from (ii) the development factor to ultimate for AY 2010 is log-normally distributed with parameters:

$\mu = 0.171251 + 0.035850 + 0.008787 = 0.215889$

$\sigma^2 = 0.0321482 + 0.0456062 + 0.0468532 = 0.0728602$

So an upper 99% confidence limit for the development factor to ultimate is given by $\exp(0.215889 + 0.07286 \times 2.3263) = 1.47018$

So an upper 99% confidence limit for total claims is
$1.47018 \times 276 = 405.77$

So an upper 99% confidence limit for outstanding claims is
$405.77 - 276 = 129.77$

Part (i) was answered very well by the majority of candidates. Parts (ii) and (iii) were answered poorly. Many candidates failed to produce sufficient detail in part (ii) for instance calculating the parameters of the distribution rather than explaining why it was a log-normal. For part (iii) few candidates were able to calculate the parameters of the distribution correctly. A further common mistake was to calculate a two-sided confidence interval.

END OF EXAMINERS' REPORT