Imperial College
London

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2014

M3S2/M4S2/M5S2

Statistical Modelling II

Date: Thursday, 15th May 2014 Time: 10 am – 12 noon

Credit will be given for all questions attempted but extra credit will be given for complete or nearly complete answers.

Calculators may not be used.

1. Consider an exponential family model of the form

$$\log f(y;\theta) = \frac{y\theta}{a(\phi)} - \frac{d(\theta)}{a(\phi)} + h(y,\phi), \quad a(\phi) = \frac{\phi}{w}. \tag{1}$$

(a) Assume without proof that $U_\theta = \frac{\partial \log f(y;\theta)}{\partial\theta}$ satisfies:

$$\mathbb{E}[U_\theta] = 0, \ \mathbb{E}[U_\theta^2] = -\mathbb{E}[\partial U_\theta/\partial\theta]$$

Hence derive expressions for $\mathbb{E}[Y]$ and $\text{Var}[Y]$ in terms of $\theta, d(.)$ and $a(\phi)$.

(b) Consider now the negative binomial distribution over the natural numbers, given by

$$P(Y = y; p, r) = \begin{pmatrix} y + r - 1 \\ y \end{pmatrix} (1-p)^r p^y, \quad \text{where } r > 0, r \in \mathbb{N}.$$

Assume throughout that $r = 4$. Consider the reparameterisation $\theta = \log(p)$, and hence obtain a formula for $d(\theta)$ and $a(\phi)$ from (1).

(c) Using (a) and (b), derive the mean and variance of Y in terms of p.

(d) Using (c), prove that the variance function $V(\mu)$ is in this case given by

$$V(\mu) = \frac{1}{4}\mu^2 + \mu.$$

(e) For a dataset $(y_i, X_i)_{i=1}^n$, consider the (unscaled) response residuals, given by $y_i - \hat{\mu}_i$ for $i = 1, \ldots, n$. Assuming the negative binomial GLM with canonical link is the correct model for this dataset, which of the following would we expect:

(i) larger dispersion of the response residuals for larger fitted values;

(ii) smaller dispersion of the response residuals for larger fitted values; or

(iii) equal dispersion of the response residuals for all fitted values.

You do not need to justify your answer. You can assume that all fitted values are positive.

[END OF QUESTION 1.]

2. (a) We fit a Poisson GLM in R with canonical link to the data in Table 1 using the command:

```
> m.p <- glm(y~X, family='poisson', data=D.A)
```

Let $\mathcal{J}$ denote the Fisher information under this model and $\Sigma = \mathcal{J}^{-1}$ its inverse. Use the Delta method to obtain the following approximate standard error for for the fitted value for the 7th datapoint, $\hat{\mu}_7$:

$$\mathrm{se}(\hat{\mu}_7) \approx \hat{\mu}_7 \sqrt{\Sigma_{11} + 2\Sigma_{12} + \Sigma_{22}}$$

(b) Using the same fit as in (a), consider the following output:

```
> sum(residuals(m.p,type="pearson")^2)
[1] 2.901892
> sum(residuals(m.p,type="response")^2)
[1] 23.91614
```

Using (some of) this output, can you determine whether fitting a quasi-Poisson model to this data would have produced a smaller, or larger standard error for $\hat{\mu}_7$? Briefly justify your answer, stating any results you use from lectures without proof.

(c) Consider the simpler model that only includes an intercept:

```
> m.0 <- glm(y~1,family='poisson',data=D.A)
```

State with a brief justification what the fitted values for each observation will be for this model. Hence confirm that the deviance residual of the 1st observation is given by

$$r_D(1) = -\sqrt{12 - 4\log 4}$$

	y	X
1	2	-1
2	3	-1
3	6	0
4	7	0
5	8	0
6	9	0
7	10	1
8	12	1
9	15	1

Table 1: Dataset D.A used in Question 2

[END OF QUESTION 2.]

3. Consider the dataset D.B given in Table 2 on the next page.

(a) We fit the following three models in R:

```
> m.0.G.GS <- lm(y~ 0 + Group + Group:Size, data=D.B)
> m.1.G <- lm(y~ 1 + Group , data=D.B)
> m.C.G <- lm(y~ 1 + C(Group,sum), data=D.B)
```

where the latter two are identical except in their choice of contrasts. Report the first and last rows of the design matrix for each model.

(b) Rank these three models by their RSS values.

(c) We wish to use model m.1.G from part (a) to assess whether the means of groups 'a' and 'e' differ. Explain how this question can be answered at 95% significance level using the following facts and Table 3:

- The RSS for this model is given by

```
> sum(residuals(m.1.G)^2)
[1] 4.363716
```

- The estimated regression coefficients are given by:

```
     (Intercept) Groupb Groupc Groupd Groupe
[1,]        0.78    1.6   1.98   3.83   4.36
```

- letting X be the design matrix, $(X^T X)^{-1}$ is given by:

```
            (Intercept) Groupb Groupc Groupd Groupe
(Intercept)         0.5   -0.5   -0.5   -0.5   -0.5
Groupb             -0.5    1.0    0.5    0.5    0.5
Groupc             -0.5    0.5    1.0    0.5    0.5
Groupd             -0.5    0.5    0.5    1.0    0.5
Groupe             -0.5    0.5    0.5    0.5    1.0
```

[QUESTION 3 CONTINUED OVERLEAF.]

	y	Group	Size
1	0.37	a	9.30
2	1.18	a	2.10
3	1.16	b	6.50
4	3.60	b	1.30
5	3.33	c	2.70
6	2.18	c	3.90
7	4.49	d	0.10
8	4.74	d	3.80
9	5.58	e	8.70
10	4.69	e	3.40

Table 2: Dataset D.B used in Questions 3 and 4

	$P(\|t\| > 4.36)$
df=3	0.022
df=4	0.012
df=5	0.007

Table 3: p-values of the t-distribution for use in Question 3

[END OF QUESTION 3.]

4. Consider again the dataset D.B given in Table 2 of Question 3. This time we fit a random effects model to this data as follows:

```
> m.r1 <- lmer(y~(1|Group)+1+Size,data=D.B); summary(m.r1)
```

```
Linear mixed model fit by REML
Formula: y ~ (1 | Group) + 1 + Size
   Data: D.B
   AIC   BIC logLik deviance REMLdev
 43.84 45.05 -17.92    34.48   35.84
Random effects:
 Groups   Name        Variance Std.Dev.
 Group    (Intercept) 2.60161  1.61295
 Residual             0.87205  0.93384
Number of obs: 10, groups: Group, 5

Fixed effects:
            Estimate Std. Error t value
(Intercept)   3.6432     0.9188   3.965
Size         -0.1222     0.1164  -1.050

Correlation of Fixed Effects:
     (Intr)
Size -0.529
```

(a) Write down the model fit with the above R code in classical notation.

(b) State the formula for the correlation between observations in the same group using the model notation used in (a). Based on the R output, would the estimate of this correlation be greater than or less than 0.5?

(c) Confirm the reported value of the AIC given in the R output using the reported value for the Log-Likelihood.

(d) Now letting y be the $n \times 1$ row of observations, given by Table 2, we may write down the model in design notation as follows:

$$y = X\beta + Z\gamma + \epsilon \qquad (2)$$

where β are the fixed effects and γ the random effects. Complete this description by specifying the sizes and elements of the two matrices X and Z, as well as the distributions of any random quantities (including any independence assumptions). Finally report the joint distribution of the $n \times 1$ vector y in the form $y \sim N(m, S)$, giving formulae for m and S in terms of the quantities in (2).

(e) Verify that, if $L = I_n - X(X^TX)^{-1}X^T$, then Ly does not depend on β.

[END OF QUESTION 4.]

REVISED DURING MARKING
MINOR TYPO ON p. 3.

EXAMINATION SOLUTIONS 2013-14

Course S2

Question 1

Marks & seen/unseen

Part a) — seen

For the mean, we have:

$$U_\theta = \frac{d\mathcal{L}(\theta; y)}{d\theta} = \frac{y}{\alpha(\phi)} - \frac{d'(\theta)}{\alpha(\phi)} \therefore \mathbb{E}[U_\theta] = \frac{\mathbb{E}[Y]}{\alpha(\phi)} - \frac{d'(\theta)}{\alpha(\phi)} = 0$$

$\therefore\ \mathbb{E}[Y] \equiv \mu = d'(\theta)$, solving for $\mathbb{E}[Y]$. — 2

Squaring the above expression for U_θ, we have

$$\mathbb{E}[U_\theta^2] = \mathbb{E}\left[\frac{1}{\alpha(\phi)^2}\left(Y - d'(\theta)\right)^2\right] =$$

$$= \mathbb{E}\left[\frac{1}{\alpha(\phi)^2}\left(Y - \mathbb{E}_\theta[Y]\right)^2\right] = \frac{1}{\alpha(\phi)^2}\text{Var}[Y]$$

— 2

where the second equality follows because $d'(\theta) = \mathbb{E}_\theta[Y]$. But we also note that $-\mathbb{E}\left[dU_\theta/d\theta\right] = \frac{d''(\theta)}{\alpha(\phi)}$. So using the identity $\mathbb{E}[U_\theta^2] = -\mathbb{E}[dU_\theta/d\theta]$, we get:

$$\text{Var}[Y] = \alpha(\phi)d''(\theta)$$

— 2 = 6

Part b) — Unseen

The log-density breaks down as follows:

$$\log f(y;p) = y\log p + 4\log(1-p) + \log\binom{y+4-1}{y}$$

— 2

Setting $\theta = \log p$, we may rewrite this as:

$$\log f(y;p) = y\theta + 4\log(1-e^\theta) + \log\binom{y+4-1}{y}$$

Consequently, $d(\theta) = -4\log(1-e^\theta)$. — 1=3

Part c) — Unseen

For the mean,

$$d'(\theta) = \frac{4e^\theta}{1-e^\theta},\ \therefore\ \mathbb{E}[Y] = \frac{4p}{1-p}$$

— 2

For the variance, we note that $\alpha(\phi) = 1$ in this case, so

$$\text{Var}[Y] = d''(\theta) = \frac{4e^\theta(1-e^\theta) - 4e^\theta(-e^\theta)}{(1-e^\theta)^2} = \frac{4e^\theta}{(1-e^\theta)^2}$$

— 1

$$\therefore\ \text{Var}[Y] = \frac{4p}{(1-p)^2}$$

— 1 = 4

Setter's initials | Checker's initials | Page number

	EXAMINATION SOLUTIONS 2013-14	Course
Question 1		Marks & seen/unseen
Part d)	To obtain the variance function, we must write $\mathrm{Var}[Y]$ in terms of $\mu = \mathbb{E}[Y]$. Since we have an expression for $\mathrm{Var}[Y]$ in terms of p, we start by writing p as a function of μ. $$\mu = \frac{4p}{1-p} \therefore \frac{4p}{\mu} + p - 1 = 0 \therefore p(4+\mu) = \mu \therefore p = \frac{\mu}{4+\mu}.$$ We will now substitute this expression into the formula for $\mathrm{Var}[Y]$, noting that $1 - p = 1 - \frac{\mu}{4+\mu} = \frac{4}{4+\mu}$: $$\mathrm{Var}[Y] = \frac{4p}{(1-p)^2} = \frac{4\frac{\mu}{4+\mu}}{\frac{4^2}{(4+\mu)^2}} = \frac{1}{4}\mu(4+\mu) = \frac{1}{4}\mu^2 + \mu$$ Since $a(\phi) = 1$, this means that $V(\mu) = \frac{1}{4}\mu^2 + \mu$.	Unseen 2 2 1=5
Part e)	The correct answer is (i). We do not need justification here (because the students tend to write too much when asked for "informal" justification). For the benefit of the marker, the reason is that the dispersion of the response residuals agrees with the variance function, which is monotonic increasing for $\mu > 0$.	Seen 2=2
	Setter's initials Checker's initials	Page number

EXAMINATION SOLUTIONS 2013-14

Course

Question 2

Marks & seen/unseen

Part a)

The Delta method can be written in terms of $\mathcal{J}$ as follows: — Part seen

$$\text{Var}[h(\hat{\beta})] \approx \left(\nabla_{\hat{\beta}} h(\hat{\beta})\right)^T \text{Cov}[\hat{\beta}] \left(\nabla_{\hat{\beta}} h(\hat{\beta})\right) \approx \left(\nabla_{\hat{\beta}} h(\hat{\beta})\right)^T \Sigma \left(\nabla_{\hat{\beta}} h(\hat{\beta})\right)$$ 1

In this case, we want to write $\hat{\mu}_7 = h(\hat{\beta})$, and, since $X_7 = (1, 1)$, we have

$$\hat{\mu}_7 = \exp(X\hat{\beta}) = \exp(\hat{\beta}_1 + \hat{\beta}_2).$$ 2

Therefore,

$$\nabla_{\hat{\beta}} h(\hat{\beta}) = \begin{pmatrix} e^{\hat{\beta}_1+\hat{\beta}_2} \\ e^{\hat{\beta}_1+\hat{\beta}_2} \end{pmatrix} = \begin{pmatrix} \hat{\mu}_7 \\ \hat{\mu}_7 \end{pmatrix}.$$ Seen for $\hat{\mu}_1$. 2

So,

$$\text{Var}[\hat{\mu}_7] \approx \hat{\mu}_7 \Sigma_{11} \hat{\mu}_7 + 2\hat{\mu}_7 \Sigma_{12} \hat{\mu}_7 + \hat{\mu}_7 \Sigma_{22} \hat{\mu}_7 = \hat{\mu}_7^2 (\Sigma_{11} + 2\Sigma_{12} + \Sigma_{22}).$$ 1

Hence, using the fact that $\hat{\mu}_7$ is positive (by the choice of link),

$$\text{se}(\hat{\mu}_7) \approx \hat{\mu}_7 \sqrt{\Sigma_{11} + 2\Sigma_{12} + \Sigma_{22}}.$$ 1 = 7

Part b)

Let $\Sigma^{(\phi)}$ denote the variance-covariance matrix of $\hat{\beta}$ under the quasi-Poisson model. It follows that — Similar seen

$$\Sigma^{(\phi)} = \tilde{\phi}\Sigma$$ 1

where $\tilde{\phi}$ is the estimated dispersion parameter, given by $X^2/(N-p)$ where X^2 is the sum of the squared Pearson residuals. Consequently, using the output given in the question, and the fact that $N - p = 7$ in this case, we have: — 1 (for X^2) 1 (for N-p)

$$\tilde{\phi} = 2.9/7 \approx 0.4 < 1.$$ 1 (for phi <1)

Consequently, $\tilde{\phi} < 1$ (the data are under-dispersed). Note also that:

$$\sqrt{\Sigma_{11}^{(\phi)} + 2\Sigma_{21}^{(\phi)} + \Sigma_{22}^{(\phi)}} = \sqrt{\tilde{\phi}(\Sigma_{11} + 2\Sigma_{12} + \Sigma_{22})}$$ 1

Moreover, the fitted values are exactly the same under the quasi-Poisson model (stated in lectures). We now use the definition of $\hat{\Sigma}^{(\phi)}$ to write:

$$\text{se}_{\text{quasi}}(\hat{\mu}_7) = \sqrt{\tilde{\phi}}\,\text{se}(\hat{\mu}_7) < \text{se}(\hat{\mu}_7)$$ 1 = 6

since $\tilde{\phi} < 1$.

Setter's initials — Checker's initials — Page number

	EXAMINATION SOLUTIONS 2013-14	Course
Question **2**		Marks & seen/unseen

Part c)

Solution	Marks
The model that only contains an intercept will output the same fitted value for all observations, which will coincide with their sample mean:	Part seen
$$\hat{\mu}_i = \frac{2+3+6+7+8+9+10+12+15}{9} = \frac{72}{9} = 8,\ \forall i$$	1
The ith contribution to the deviance is given by:	
$$\begin{aligned} d_i &= -2(\log f(y_i; \mu_i = \hat{\mu}_i) - \log f(y_i; \mu_i = y_i)) \\ &= -2(y_i \log \hat{\mu}_i - \hat{\mu}_i - \log(y_i!) - y_i \log y_i + y_i + \log(y_i!)) \\ &= -2(y_i(\log \hat{\mu}_i - \log y_i) + y_i - \hat{\mu}_i) \\ &= -2(y_i(\log 8 - \log y_i) + y_i - 8) \end{aligned}$$	1 (defn) 1 (cancel factorials) 1
For $i = 1$, $y_1 = 2$, so we would hence get	
$$d_1 = -2(2(\log 8 - \log 2) + 2 - 8) = -2(2\log 4 - 6) = 12 - 4\log 4$$	1
Since $y_1 < \hat{\mu}_1 = 8$, we would hence get:	1 (sign)
$$r_D(1) = \text{sign}(y_1 - \hat{\mu}_1)\sqrt{d_1} = -1 \times \sqrt{12 - 4\log 4} = -\sqrt{12 - 4\log 4}$$	1 = 7

	Setter's initials	Checker's initials	Page number

<table>
<tr><td></td><td>EXAMINATION SOLUTIONS 2013-14</td><td>Course</td></tr>
<tr><td>Question 3</td><td></td><td>Marks & seen/unseen</td></tr>
<tr><td>Part a)</td><td>The rows for <code>M.0.G.GS</code> would be given by:

$$X_1 = (1,0,0,0,0,9.3,0,0,0,0)$$
$$X_{10} = (0,0,0,0,1,0,0,0,0,3.4)$$

and for <code>M.1.G</code>, they would be

$$X_1 = (1,0,0,0,0,)$$
$$X_{10} = (1,0,0,0,1).$$

Finally, for <code>M.C.G</code>, they are

$$X_1 = (1,1,0,0,0)$$
$$X_{10} = (1,-1,-1,-1,-1).$$</td><td>Similar seen
1
1
1
1
1
1
=6</td></tr>
<tr><td>Part b)</td><td>Model <code>m.0.G.GS</code> has 10 parameters and 10 datapoints, and a full rank design matrix so it is in fact the saturated model and will have an RSS of 0. Models <code>m.C.G</code> and <code>m.1.G</code> only differ in their contrasts, so their RSS will be the same. So

$$\mathrm{RSS(m.0.G.GS)} < \mathrm{RSS(m.C.G)} = \mathrm{RSS(m.1.G)}$$</td><td>Similar seen
2
2
=4</td></tr>
<tr><td>Part c)</td><td>We will use model <code>m.1.G</code>, where the intercept encodes group 'a' and $\hat\beta_j$ for $j \in \{2,3,4,5\}$ encode the difference between the mean for group 'a' and the other groups. Hence, the question simply corresponds to testing $\hat\beta_5 = 0$. We know that under the assumption of normality, equal residual variance and independence, the null implies that:

$$t = \frac{1}{(X^TX)^{-1}_{55}\,\hat\sigma}\hat\beta_5 \sim t_{n-p}$$

where in this case $n-p=5$, and X is the design matrix. Since $\left(X^TX\right)^{-1}_{55}$ is 1, and $\hat\sigma^2 = \mathrm{RSS}/(n-p) \approx 4.36/5 < 1$, so $\hat\sigma < 1$ and

$$\mathrm{se}(\hat\beta_5) < 1 \;\therefore\; t > \hat\beta_5 = 4.36.$$

According to the table, the p-value for $t = 4.36$ is 0.007, so it will be equal to at most 0.007 for our test statistic (as it is larger). Hence the difference between the two group means is statistically significant at 95% significance level.</td><td>Similar seen
2
2
1 (n-p=4)
1 (for X^TX)
1 (for s^2<1)
2
1 (for right p)
= 10</td></tr>
<tr><td></td><td>Setter's initials　　　Checker's initials</td><td>Page number</td></tr>
</table>

	EXAMINATION SOLUTIONS 2013-14	Course
Question 4		Marks & seen/unseen
Part a)	We can use the following classical notation to represent this model: $$y_{ij} = \mu + \alpha_j + \epsilon_{ij},\ \alpha_j \sim N(0, \sigma_\alpha^2),\ \epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$ where $i = 1, 2$ indexes the two individuals within each group, and $j = 1, \ldots, 5$ indexes the five groups labelled 'a', ..., 'e' in Table 2.	seen 3 1 = 4
Part b)	The formula for the correlation of observations in the same group is given by $$\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}$$ Since in this case $\hat{\sigma}_\alpha^2 = 2.6767 > 0.8727 = \hat{\sigma}_\epsilon^2$, the estimated correlation will be greater than 0.5.	seen 2 1 = 3
Part c)	The model has 4 parameters, since it needs to estimate σ_α^2, σ_ϵ^2 and the fixed effects $\hat{\beta}_1, \hat{\beta}_2$. Therefore, $$\text{AIC} = -2 \times (-17.92) + 8 = 35.84 + 8 = 43.84$$	similar seen 2 (AIC defn) 1 (right DOFs) =3
Part d)	In design notation, we would write $$y = X\beta + Z\gamma + \epsilon,\ \epsilon \sim N(0, \sigma_\epsilon^2 I_{10}),\ \gamma \sim N(0, \sigma_\alpha^2 I_5), \epsilon \perp \gamma$$ where X is a 10×2 matrix whose first column is all 1s and the second column contains the sizes for each datapoint; and Z is a 10×5 matrix with entries: $$Z_{ij} = \begin{cases} 1, & \text{for } i = 2j \text{ or } 2j - 1 \\ 0 & \text{otherwise.} \end{cases}$$ Therefore, the joint distribution of y is given by: $$y \sim N(X\beta, \sigma_\epsilon^2 I_{10} + \sigma_\alpha^2 ZZ^T)$$	seen 1 2 (for X) 2 (for Z) 2 (for joint) = 7
Part e)	This is easy to show, since $$\begin{aligned} Ly &= L(X\beta + Z\gamma + \epsilon) = LX\beta + L(Z\gamma + \epsilon) \\ \therefore\ LX\beta &= (I_n - X(X^TX)^{-1}X^T)X\beta = X\beta - X(X^TX)^{-1}X^TX\beta \\ &= X\beta - X\beta = 0 \end{aligned}$$ Thus, $Ly = L(Z\gamma + \epsilon)$.	seen 1 2 =3
	Setter's initials Checker's initials	Page number