MAS474 Extended linear models
2017-18 Exam Solutions

1. (i) (Unseen but routine) `model1` is

$$\text{bwt}_{ij} = \alpha + \beta \times \text{days}_j + b_i + \epsilon_{ij}$$

where $b_i \sim N(0, \sigma_1^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$.

`model2` is

$$\text{bwt}_{ij} = \alpha + \beta \times \text{days}_j + b_i^{(1)} + b_i^{(2)} \times \text{days}_j + \epsilon_{ij}$$

where $b_i^{(1)} \sim N(0, \sigma_1^2)$, $b_i^{(2)} \sim N(0, \sigma_2^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$.

(ii) (Unseen but routine) `model2` as both the slopes and intercepts look as if they differ between guinea pigs.

(iii) (Unseen)Use the generalised likelihood ratio test

$$\Gamma = -2(l(\theta_1) - l(\theta_2)) = -2(-487.5 - -473.1) = 28.8$$

According to Wilk's theorem $\Gamma \sim \chi_1^2$. The 95^{th} percentile of a χ_1^2 is $1.96^2 = 3.84 < 28.8$ (or from tables) and so we reject `model1` in favour of `model2` at the 95% level.

(iv) (Bookwork) Fit model2 to the data to find $\hat{\theta}$. Then

For $i = 1, \ldots, B$

- Simulate fake data from model 2 using parameter $\hat{\theta}$.

$$y_{ij} \sim f(y \mid x, \hat{\theta})$$

- Refit model2 to the simulated data to get parameter estimate $\hat{\theta}^{(i)}$.

Form a 95% confidence interval by looking at the 2.5th and 97.5th percentiles of

$$\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}.$$

(v) (Unseen) The best prediction is

$$107.17 + 10 \times 4.27 = 149.9 \text{ grams}$$

A 95% prediction interval is formed by considering the variance of

$$b^{(1)} + 10b^{(2)} + \epsilon_{ij}$$

which is

$$\sigma_1^2 + 100\sigma_2^2 + \sigma^2 = 1757$$

So a 95% prediction interval is

$$149 \pm 1.96 * \sqrt{1757} = [67.7, 231.9]$$

2. (i) (a) (Bookwork) The data are MAR if

$$f(M \mid Y) = f(M|Y_{obs})\checkmark$$

and NMAR if the $f(M \mid Y)$ depends on the missing values of $Y\checkmark$.

(b) (Bookwork) Given the observed data (Y_{obs}, M), the full likelihood function is

$$\begin{aligned} L_{full}(\theta, \psi \mid Y_{obs}, M) &= f(Y_{obs}, M \mid \theta, \psi) \\ &= \int f(Y_{obs}, Y_{mis} \mid \theta) f(M \mid Y_{obs}, Y_{mis}, \psi) \mathrm{d}Y_{mis}.\checkmark \end{aligned}$$

We can see that if the distribution of M doesn't depend on Y_{mis}, i.e., if it is MAR, then

$$\begin{aligned} L_{full}(\theta, \psi \mid Y_{obs}, M) &= f(M \mid Y_{obs}, \psi) \int f(Y_{obs}, Y_{mis} \mid \theta) \mathrm{d}Y_{mis}\checkmark \\ &= f(M \mid Y_{obs}, \psi) f(Y_{obs} \mid \theta) \\ &= f(M \mid Y_{obs}, \psi) L_{ign}(\theta \mid Y_{obs})\checkmark \end{aligned}$$

and so to learn θ we only need to maximize $L_{ign}(\theta \mid Y_{obs})$.$\checkmark$

(ii) a) Interest here lies in the effect of the choice of fuel. Thus this should be represented as a fixed effect$\checkmark$. The cars used in the study are not of direct interest, thus these are represented as random effects. $\checkmark$

b)

$$y = X\beta + Zb + \epsilon\checkmark$$

where

$$\beta = \begin{pmatrix} \alpha \\ \beta_1 \end{pmatrix}, \qquad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\checkmark$$

and

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix}, \checkmark \qquad Z = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \checkmark$$

c) $\hat{\alpha} = \bar{Y}_{...} = \frac{1}{8} \sum_{ijk} Y_{ijk}\checkmark$

$$
\begin{aligned}
\mathbb{V}\text{ar}(\hat{\alpha}) &= \frac{1}{64}\mathbb{V}\text{ar}(\sum_{ijk} \alpha + \beta_i + b_j + \epsilon_{ijk}) \\
&= \frac{1}{64}\left(\mathbb{V}\text{ar}(4\sum_{j} b_j) + \sum_{ijk} \mathbb{V}\text{ar}(\epsilon_{ijk})\right) \\
&= \frac{1}{64}(16\mathbb{V}\text{ar}(\sum_{j} b_j) + 8\sigma^2) \\
&= \frac{\sigma_1^2}{2} + \frac{\sigma^2}{8}
\end{aligned}
$$

d) Note that

$$
\bar{Y}_{1\cdot\cdot} - \bar{Y}_{2\cdot\cdot} = \frac{1}{4}\sum_{jk}((\alpha + \beta_1 + b_j + \epsilon_{1jk}) - (\alpha + \beta_2 + b_j + \epsilon_{2jk})) = \frac{1}{4}\sum_{jk}(\epsilon_{1jk} - \epsilon_{2jk})
$$

$$
\begin{aligned}
\mathbb{C}\text{ov}(\hat{\alpha}, \hat{\beta}_1) &= \mathbb{C}\text{ov}(\bar{Y}_{\cdots}, \frac{\bar{Y}_{1\cdot\cdot} - \bar{Y}_{2\cdot\cdot}}{2}) \\
&= \mathbb{C}\text{ov}\left(\frac{1}{8}\sum_{ijk} b_j + \epsilon_{ijk}, \frac{1}{8}\sum_{jk}(\epsilon_{1jk} - \epsilon_{2jk})\right) \\
&= \frac{1}{64}\mathbb{C}\text{ov}\left(\sum_{ijk} \epsilon_{ijk}, \sum_{jk}(\epsilon_{1jk} - \epsilon_{2jk})\right) \\
&= \frac{1}{64}\left(\mathbb{C}\text{ov}(\sum_{jk} \epsilon_{1jk}, \sum_{jk} \epsilon_{1jk}) - \mathbb{C}\text{ov}(\sum_{jk} \epsilon_{2jk}, \sum_{jk} \epsilon_{2jk})\right) \\
&= 0
\end{aligned}
$$

3. **This question is a fairly standard problem, and close to the example studied in the lecture notes, albeit with a twist in the information they are provided. As such, it is a mix of bookwork and unseen. Given that we will not spend much time on the EM algorithm this year, I think that is okay.**

(i)

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \prod_{j=1}^{m} \lambda e^{-\lambda X_{j+n}} \checkmark$$

$$= \lambda^{m+n} \exp\left(-\lambda \left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{m} X_{n+j} \right) \right)$$

and so

$$\log L(\theta) = (m+n) \log \lambda - \lambda \left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{m} X_{j+n} \right) \checkmark\checkmark$$

as required.

(ii)

$$\mathbb{P}(X \leq x \mid X \leq h) = \frac{\mathbb{P}(X \leq x)}{\mathbb{P}(X \leq h)} = \frac{1 - e^{-\lambda x}}{1 - e^{-\lambda h}} \checkmark\checkmark$$

(iii) $F(y) = \mathbb{P}(Y \leq y)$ and so $1 - F(y) = \mathbb{P}(Y > y)$ Hence

$$\int_0^\infty 1 - F(y) dy = \int_0^\infty \mathbb{P}(Y > y) dy \checkmark$$

$$= \int_0^\infty \mathbb{E}\mathbb{I}_{Y>y} dy \checkmark$$

$$= \mathbb{E}(\int_0^\infty \mathbb{I}_{Y>y} dy)$$

$$= \mathbb{E}(\int_0^Y 1 dy) \checkmark$$

$$= \mathbb{E}(Y) \checkmark$$

as required.

(iv) Use the result from part (ii)

$$\mathbb{E}(X \mid X \leq h, \lambda) = \int_0^h \mathbb{P}(X > x \mid X \leq h) dx \checkmark$$

$$= \int_0^h 1 - \frac{1 - e^{-\lambda x}}{1 - e^{-\lambda h}} dx$$

$$= h - \frac{1}{1 - e^{-\lambda h}} \left[x + \frac{e^{-\lambda x}}{\lambda} \right]_0^h \checkmark$$

$$= \frac{1}{\lambda} - \frac{h e^{-\lambda h}}{1 - e^{-\lambda h}} \checkmark$$

(v)

$$
\begin{aligned}
Q(\lambda \mid \lambda^{(k)}) &= \mathbb{E}_{\lambda^{(k)}}\left(\log L(\lambda \mid x_1,\ldots,x_n,X_{n+1},\ldots X_{n+m}) \mid x_1,\ldots,x_n, r \text{ of the } X_i > h, \lambda^{(k)}\right) \\
&= (m+n)\log\lambda - \lambda\left(\sum_{i=1}^{n} x_i + (m-r)\mathbb{E}(X \mid X > h) + r\mathbb{E}(X \mid X \le h)\right) \\
&= (m+n)\log\lambda - \lambda\left(n\bar{x} + (m-r)(h + \frac{1}{\lambda^{(k)}}) + r(\frac{1}{\lambda^{(k)}} - \frac{he^{-h\lambda^{(r)}}}{1-e^{-h\lambda^{(r)}}})\right)
\end{aligned}
$$

as $\mathbb{E}(X \mid X > h) = h + \frac{1}{\lambda}$ because X is exponential and has the memoryless property.

(vi) Thus to find $\lambda^{(k+1)}$ we minimize $Q(\lambda \mid \lambda^{(k)})$ wrt λ.

$$
\frac{dQ}{d\lambda} = \frac{m+n}{\lambda} - \left(n\bar{x} + (m-r)(h + \frac{1}{\lambda^{(k)}}) + r(\frac{1}{\lambda^{(k)}} - \frac{he^{-h\lambda^{(k)}}}{1-e^{-h\lambda^{(k)}}})\right)
$$

and so

$$
\lambda^{(k+1)} = \frac{m+n}{n\bar{x} + (m-r)(h + \frac{1}{\lambda^{(k)}}) + r(\frac{1}{\lambda^{(k)}} - \frac{he^{-h\lambda^{(k)}}}{1-e^{-h\lambda^{(k)}}})}
$$