1a) • The pdf can be written as

$$f_Y(y) = \lambda e^{-\lambda y} = e^{-\lambda y + \ln \lambda}$$

$$= e^{\theta y - b(\theta)}$$

where $\theta = -\lambda$ is the natural parameter and $b(\theta) = -\ln(-\theta)$.

• Using formulas for the mean and variance

$$EY = b'(\theta)$$

and

$$\mathrm{Var}\,Y = \frac{\phi}{w} b''(\theta)$$

• we find

$$EY = +\frac{1}{-\theta} = \frac{1}{\lambda}$$

and

$$\mathrm{Var}\,Y = \frac{1}{\theta^2} = \frac{1}{(-\lambda)^2} = \frac{1}{\lambda^2}$$

• Given the cumulant generating function (with $w = \phi = 1$)

$$K_Y(u) = b(\theta + u) - b(\theta) = -\ln(-\theta - u) + \ln(-\theta)$$

we have

$$K'_Y(u) = +\frac{1}{-\theta - u}$$

$$\vdots$$

$$K''_Y(u) = \frac{1}{(-\theta - u)^2}$$

$$K'''_Y(u) = \frac{2}{(-\theta - u)^3}$$

the third central moment is

$$E\left[(Y - EY)^3\right] = K'''_Y(0) = \frac{2}{(-\theta)^3} = \frac{2}{\lambda^3}$$

1b) • Using the canonical link the natural parameter $\theta_i = -\lambda_i$ is equated to the linear predictor $\eta_i = \underline{x}_i^T \beta$ so the expected value $\mu_i = E(y_i) = \frac{1}{\lambda_i}$ is related to η_i via

$$\mu_i = \frac{1}{\lambda_i} = \frac{1}{-\theta_i} = \frac{1}{-\eta_i}$$

or

$$-\frac{1}{\mu_i} = \eta_i$$

so the link function $g(\mu_i) = -\frac{1}{\mu_i}$.

• The log likelihood is

$$\ell(\underline{\beta}) = \sum_{i=1}^{n} \eta_i y_i - b(\eta_i)$$

and the score function

$$\underline{s}(\underline{\beta}) = \sum_{i=1}^{n} (y_i - \mu_i) \frac{\partial \eta_i}{\partial \underline{\beta}} = \sum_{i=1}^{n} \left(y_i - \frac{1}{\lambda_i}\right) \underline{x}_i$$

• The Fisher information can be found via

$$F(\underline{\beta}) = Var(\underline{s}(\underline{\beta})) = \sum_{i=1}^{n} Var\left(\underline{x}_i \left(y_i - \frac{1}{\lambda_i}\right)\right)$$

$$= \sum_{i=1}^{n} \underline{x}_i \frac{1}{\lambda_i^2} \underline{x}_i^T = \sum_{i=1}^{n} \frac{1}{\lambda_i^2} \underline{x}_i \underline{x}_i^T$$

• The constraint $\lambda_i > 0$ translates to

$$-\frac{1}{\eta_i} > 0$$

$$\eta_i < 0$$

$$x_i^T \beta < 0$$

or assuming the the model includes an intercept

$$\beta_0 < -\sum_{j=1}^{k} \beta_j x_{ij}$$

for all observations $i = 1, 2, \ldots, n$, i.e.

$$\beta_0 < \min\left(-\sum_{j=1}^{k} \beta_j x_{ij}\right).$$

Thus, the initial value $\beta = (0, .., 0)^T$ would not be on the interior of the parameter space.

- A more "natural" choice would be the log link

$$\ln(\mu_i) = \eta_i$$

since this ensures that

$$\frac{1}{\lambda_i} = \mu_i = e^{\eta_i} > 0 \text{ for any } \beta \in \mathbb{R}^p.$$

2a) - The counts $Y_1, Y_2, \ldots, Y_n$ are independent,

$$Y_i = (Y_{iA}, Y_{iB}, \ldots, Y_{iF}) \sim \text{Multinom}(n_i, \pi_i),$$

for $i = 1, 2, \ldots, n$ and

$$\text{logit } P(Y_i \leq r) = \text{logit}(\pi_{i1} + \ldots + \pi_{ir})$$
$$= \theta_r + x_i^T \beta.$$

- $\hat{\theta}_1 = -2.64$, $\hat{\theta}_2 = -1.44, \ldots, \hat{\theta}_5 = 1.28$,

and $\hat{\beta} = (-0.26, \ldots, -1.58)$

2b) • The logit (log odds) for a pass ($y_i \leq 5$) change by -0.26 when comparing 2005 to 2004 (the reference category), so the odds of passing changes by

$$e^{-0.26} = 0.77,$$

i.e. a 23% decrease in odds of passing.

• Similarly, comparing resit to ordinary exams, the odds of passing change by a factor of

$$e^{-1.58} = 0.21$$

that is, the odds is 79% smaller at a resit exam, regardless of the value of other covariates.

• P(getting a B in ordinary exam in 2008)

$$= P(y_i \leq 2) - P(y_i \leq 1)$$

$$= \frac{1}{1 + e^{-(\hat{\theta}_2 + \hat{\beta}_{2008})}} - \frac{1}{1 + e^{-(\hat{\theta}_1 + \hat{\beta}_{2008})}}$$

$= 0.155$

Based on the observed data $\hat{\pi}_{iB} = \frac{Y_{iB}}{n_i} = \frac{264}{113 + \ldots + 305}$

$= 0.169$

So a reasonable agreement to the fitted model.

2c) • Under the saturated model all π_{is} are free parameters. Given the constraint that $\sum_{s=1}^{6} \pi_{is} = 1$, we have a total of

$$5 \times 10 = 50$$

parameters.

• From the multinomial likelihood, the deviance becomes

$$D = 2 \sum_{i=1}^{10} \sum_{s=1}^{6} Y_{is} \ln \frac{Y_{is}/n_i}{\hat{\pi}_{is}}.$$

• Since this is chi-square with $p_1 - p_0$

$= 50 - 10$ d.f. $E(D) = 40$.

- The critical value of the goodness-of-fit test is the quantile

$$\chi^2_{0.05,40} \approx 55.75.$$

Given the observed deviance D = 424 we reject the hypothesis that the fitted model is true.

2d) - In the model with an interaction between year and cont, the effect of an exam being a resit is different between the different years in terms of the log of the cumulative odds.

- The odds of passing in 2005 when comparing resit to ordinary exam now change by a factor

$$e^{-1.64 + 0.14} = 0.22$$

i.e. a 78% reduction in odds of passing (instead of 79% in 2b).

- The change in deviance between the models with and without the interaction is

$$D_0 - D_1 = 424.25 - 417.02 = 7.23$$

The critical value is

$$\chi^2_{\alpha, p_1 - p_0} = \chi^2_{0.05, 14-10} = 9.487.$$

Thus, the null hypothesis of no interaction cannot be rejected.

3a) - Letting y_{ij} and x_{ij} denote the birthweigth and sex (encoded as 0 and 1) for ~~litter~~ $i = 1, \dots, m$ and pup $j = 1, 2, \dots, n_i$, we assume that

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_{0,i} + \gamma_{1,i} x_{ij} + \varepsilon_{ij}$$

where $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$ and

$$\begin{bmatrix} \gamma_{0,i} \\ \gamma_{1,i} \end{bmatrix} \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix} \right).$$

- The estimates are

$$\hat{\beta}_0 = 6.40, \quad \hat{\beta}_1 = -0.39$$

$$\tau_0^2 = 0.42, \quad \tau_1^2 = 0.02$$

and

$$\hat{\rho} = \frac{\hat{\tau}_{01}}{\sqrt{\hat{\tau}_0^2 \hat{\tau}_1^2}} = -1.0$$

- The REML estimate of τ_{01} is thus

$$\hat{\tau}_{01} = -\sqrt{0.42 \cdot 0.02} = -0.0955$$

3b) - To test $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$ (an effect of sex on birth weight), the the likelihood ratio stat based on the ML fits is

$$LRT = 2\left(-206.82 - (-233.06)\right) = 54$$

The critical value is

$$\chi^2_{0.05, 1} = 3.86.$$

We thus reject H_0 in favour of H_1.

- To test if the random intercept is significant we can use either the REML or the ML fits. As only the latter is available, we get

$$LRT = 2\left(-206.82 - (-309.51)\right) = 205.$$

Since $H_0: \tau_0^2 = 0$ is on the boundary of the parameter space, the LRT statistic is distributed as a 50% : 50% mixture of chi-squares with 0 and 1 d.f. Hence the critical value is at $\alpha = 0.05$

$$\chi^2_{0.70,\, 1} = 2.7.$$

We can thus reject H_0.

3c) • When testing if the random slope is significant we have

$$H_0 : \tau_1^2 = 0, \quad \tau_{01}^2 = 0$$

vs.

$$H_1 : \tau_1^2 > 0, \quad |\tau_{01}| < \sqrt{\tau_0^2 \tau_1^2}$$

- We are thus estimating two additional parameters under H_1 (τ_1^2 and τ_{01}).
- Constraints
- Asymptotically, the LRT-statistic is a 50% : 50% mixture of chi-squares with 1 and 2 d.f.
- The observed value (based on the REML fits) is

$$LRT = 2(-207.16 - (-210.21)) = 6.1$$

This gives a p-value of

$$p = P(LRT > 6.1) \approx 0.5\, P(\chi_1^2 > 6.1) + 0.5\, P(\chi_2^2 > 6.1)$$

$$= 0{,}03$$

Thus we reject H_0.

3 d) For two pups j and k within the same litter i,

$$\mathrm{Cov}(Y_{ij}, Y_{ik})$$

$$= \mathrm{Cov}(\beta_0 + \beta_1 x_{ij} + \gamma_{0,i} + \gamma_{1,i} x_{ij} + \varepsilon_{ij},\ \beta_0 + \beta_1 x_{ik} + \gamma_{0,i} + \gamma_{1,i} x_{ik} + \varepsilon_{ik})$$

$$= \mathrm{Cov}(\gamma_{0,i} + \gamma_{1,i} x_{ij},\ \gamma_{0,i} + \gamma_{1,i} x_{ik})$$

$$= \tau_0^2 + \tau_1^2 x_{ij} x_{ik} + \tau_{01}(x_{ij} + x_{ik})$$

Comparing a male $(x_{ij} = 0)$ and female pup $(x_{ik} = 1)$ an estimate of this covariance is

$$\hat{\tau}_0^2 + \hat{\tau}_{01} = 0.42 + (-0.0955) = 0.33.$$