Correction to the final exam

Question 1.

a) **[Seen, 5 points]** See notes, slides 27 onwards

b) **[Seen, 2 points]** Ditto

c) **[Unseen, 3 points]** This is of the form in (a) with $\eta_j = x_j^{\mathrm{T}}\beta$ and $\ell_j(\varepsilon) = -\rho(\varepsilon)$. Now just work out the details. Initial values might be sought using the least squares estimates.

Question 2.

a) **[Seen, 2 points]** See notes, slide 41 onwards

b) **[Seen, 3 points]** See notes, Example 12

c) **[Seen, 1 point]** See notes, slide 48

d) **[Unseen, 4 points]** The log-likelihood of the model is

$$\ell(\beta) = y^{\mathrm{T}}X\beta - \sum_{j=1}^{n}\log\left(1 + e^{x_j^{\mathrm{T}}\beta}\right) = (X^{\mathrm{T}}y)^{\mathrm{T}}\beta + \sum_{j=1}^{n}\log\left(1 - \pi_j\right).$$

The score equation for β equals

$$X^{\mathrm{T}}y - \sum_{j=1}^{n} x_j^{\mathrm{T}}\frac{e^{x_j^{\mathrm{T}}\beta}}{1 + e^{x_j^{\mathrm{T}}\beta}} = X^{\mathrm{T}}y - X^{\mathrm{T}}\pi = 0,$$

so $X^{\mathrm{T}}y = X^{\mathrm{T}}\widehat{\pi}$. Now the deviance is $D = 2\left\{\ell(\widetilde{\beta}) - \ell(\widehat{\beta})\right\}$, and

$$\ell(\widetilde{\beta}) = \sum_{i=1}^{n}\left\{y_j\log\pi_j + (1 - y_j)\log(1 - \pi_j)\right\} = 0,$$

because if $y_j = 1$, then $\tilde{\pi}_j = 1$, and if $y_j = 0$, then $\tilde{\pi}_j = 0$. Hence

$$D = 2\left\{0 - (X^{\mathrm{T}}y)^{\mathrm{T}}\widehat{\beta} - \sum_{j=1}^{n}\log\left(1 - \widehat{\pi}_j\right)\right\} = -2\left\{(X^{\mathrm{T}}\widehat{\pi})^{\mathrm{T}}\widehat{\beta} + \sum_{j=1}^{n}\log\left(1 - \widehat{\pi}_j\right)\right\},$$

which is a function only of $\widehat{\beta}$ and $\widehat{\pi} = \pi(\widehat{\beta})$, and therefore contains no information to contrast fit of the data and the model. Hence this deviance is not useful to assess fit.

Question 3.

a) **[Seen, 3 points]** See notes, slide 113 onwards.

b) **[Seen/unseen, 5 points]** This is based on slide 123 of the notes, but the calculations were not done previously. This is a model with nested random effects, and one assumes that $\alpha_p \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\alpha^2)$

independent of $\beta_{pl} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_\beta)$ independent of $\varepsilon_{pls} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$. The degrees of freedom between plants are $A = 3$, and those for leaf within plant are $C = 4 \times 3 \times (2-1) = 12$. We can write

$$y_{p,l,s} = \overline{y}_{\cdots} + (\overline{y}_{p\cdot\cdot} - \overline{y}_{\cdots}) + (\overline{y}_{p,l\cdot} - \overline{y}_{p\cdot\cdot}) + (y_{p,l,s} - \overline{y}_{p,l\cdot}),$$

where the terms correspond to the grand mean, the variation of plants, the variation of leaves within plants, and that of sites within plants and leaves. The leaf within plant sum of squares is therefore based on

$$\overline{y}_{p,l\cdot} - \overline{y}_{p\cdot\cdot} = (\mu + \alpha_p + \beta_{p,l} + \overline{\varepsilon}_{p,l\cdot}) - (\mu + \alpha_p + \overline{\beta}_{p\cdot} + \overline{\varepsilon}_{p\cdot\cdot}) = \eta_{p,l} - \overline{\eta}_{p\cdot},$$

where $\eta_{p,l} = \beta_{p,l} + \overline{\varepsilon}_{p,l\cdot} \sim \mathcal{N}(0, \sigma^2_\alpha + \sigma^2/2)$, and equals

$$\sum_{p,l.s} (\overline{y}_{p,l\cdot} - \overline{y}_{p\cdot\cdot})^2 = 2\sum_{p,l} (\eta_{p,l} - \overline{\eta}_{p\cdot})^2 \overset{D}{=} 2(\sigma^2_\alpha + \sigma^2/2)\chi^2_{4(3-1)} \overset{D}{=} (\sigma^2 + 2\sigma^2_\alpha)\chi^2_8.$$

In the notation of the question, $B = \sigma^2 + 2\sigma^2_{\text{leaf}}$.

c) **[Unseen, 2 points]** Just solve the equations $M = \mathrm{E}(M)$, giving $\widehat{\sigma} = 0.082$, $\widehat{\sigma}_{\text{leaf}} = 0.40$, and $\widehat{\sigma}_{\text{plant}} = 0.73$. The plant- and leaf-level components of variation are appreciably larger than the site component.

Question 4.

a) **[Seen, 2 points]** See slide 161

b) **[Seen/unseen, 4 points]** See slide 162 for the derivation of $S_\alpha = (I_n + \alpha\Delta)^{-1}$. Standard manipulations then give the form of the trace, where $d_1, \ldots, d_n$ are the eigenvalues of Δ, which are of the form $d_1 = d_2 = 0 < d_3 \leqslant \cdots \leqslant d_n$ because of the stated properties of Δ. Hence $\mathrm{tr}(S_\alpha)$ decreases monotonically as a function of α, with $\mathrm{tr}(S_0) = n$ and $\mathrm{tr}(S_\infty) = 2$, corresponding to a line that passes through all the points and a straight line. Its interpretation is as equivalent degrees of freedom.

c) **[Seen/unseen, 4 points]** See Lemma 26. The interpretation is in terms of variance and square bias. Since S_α is symmetric and semi-positive definite, the spectral decomposition $S_\alpha = UQU^{\mathrm{T}}$, with Q diagonal and U orthogonal, gives that its trace is $\mathrm{tr}(S_\alpha^{\mathrm{T}} S_\alpha) = \sum_{j=1}^n (1 + \alpha d_j)^{-2} < \mathrm{tr}(S_\alpha)$, since $1 + d_j\alpha \geq 1$, with some terms strictly larger than unity. This implies that the fitted model with $S_\alpha^{\mathrm{T}} S_\alpha$ will be smoother, because its equivalent degrees of freedom are smaller.

Question 5.

a) **[Seen, 4 points]** See notes, slide 36 onwards.

b) **[Unseen, 6 points]** A: the residuals seem to be from a Poisson regression model, which have this characteristic banding pattern (corresponding to values 0, 1, 2, etc.) working from the bottom to the top in the figure), but there are two or three outliers, shown by the residuals larger than 3 or so. Drop these observations and try to fit again.

B: Residuals from a binary regression model, since there are negative residuals corresponding to the 0s and positive ones corresponding to the 1s. There seem to be no obvious problems here.

C: These look like residuals from a Poisson regression, for the reasons given in A, with no obvious problems.

D: These are residuals for count data, for the reasons given above, but they seem to be overdispersed (they are spread from -2.5 to $+6$, without any obvious outliers, unlike in A), so perhaps fitting using a quasilikelihood is indicated.