Mixed Model Cheat Sheet

http://alenxav.wix.com/home

1) Model statement

A linear Gauss-Markov models is presented as:

$$\mathbf{y} = \mathbf{Xb} + \boldsymbol{\varepsilon}$$
$$\mathbf{y} \sim \mathrm{N}(\mathbf{Xb}, \mathbf{V})$$
$$\boldsymbol{\varepsilon} \sim \mathrm{N}(0, \mathbf{V})$$

And independent random terms other than the residuals:

$$\boldsymbol{\varepsilon} = \sum_{\mathrm{I}}^{\mathrm{i}=1} \mathbf{Z}_\mathrm{i}\mathbf{u}_\mathrm{i} + \mathbf{e}$$

And the variance-covariance matrix is defined by

$$\mathrm{E}[\boldsymbol{\varepsilon\varepsilon}'] = \mathbf{V} = \sum_{\mathrm{I}}^{\mathrm{i}=1} \mathbf{Z}_\mathrm{i}\mathbf{K}_\mathrm{i}\mathbf{Z}_\mathrm{i}'\sigma_\mathrm{i}^2 + \mathbf{I}\sigma_\mathrm{e}^2$$

Reshape the model the model into:

$$\mathbf{y} = \mathbf{Xb} + \sum_{\mathrm{I}}^{\mathrm{i}=1} \mathbf{Z}_\mathrm{i}\mathbf{u}_\mathrm{i} + \mathbf{e}$$
$$\mathbf{u} \sim \mathrm{N}(0, \mathbf{Z}_\mathrm{i}\mathbf{K}_\mathrm{i}\mathbf{Z}_\mathrm{i}'\sigma_\mathrm{i}^2)$$
$$\mathbf{e} \sim \mathrm{N}(0, \mathbf{I}\sigma_\mathrm{e}^2)$$

2) Likelihood

Gaussian likelihood is defined as:

$$L = 2\pi^{-0.5\mathrm{n}}\,|\mathbf{V}|^{-0.5}\exp(-0.5(\mathbf{y}-\mathbf{Xb})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{Xb}))$$

Log likelihood:

$$\ln L = -\frac{1}{2}[\ln|\mathbf{V}| + \mathbf{y}'\mathbf{Py}]$$

And restricted log likelihood (log REML) is defined as:

$$\ln L = -\frac{1}{2}[\,\ln|\mathbf{V}| + \ln|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + \mathbf{y}'\mathbf{Py}\,]$$

Where the projection matrix **P** is defined by:

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$$

Properties and equalities of projection matrices (**S** and **P**):

$$\mathbf{SX} = \mathbf{PX} = 0, \qquad \mathbf{Py} = \mathbf{V}^{-1}(\mathbf{y}-\mathbf{Xb}), \qquad \mathbf{PP} = \mathbf{P}$$
$$\mathbf{PVP} = \mathbf{P}, \qquad \mathbf{y}'\mathbf{Py} = \mathbf{y}'\mathbf{e}\,\boldsymbol{\sigma}_\mathbf{e}^{-2}, \qquad \mathbf{P} = \mathbf{V}^{-1}\mathbf{S}$$
$$\mathbf{S} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}, \qquad \mathbf{Sy} = \mathbf{y} - \mathbf{Xb}, \qquad \mathbf{SS} = \mathbf{S}$$
$$\mathbf{ZAZPy} = \mathbf{Zu}\boldsymbol{\sigma}_\mathbf{u}^{-2}, \qquad \mathbf{Py} = (\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})\boldsymbol{\sigma}_\mathbf{e}^{-2}$$

The inverse of **V** can be expressed as follows

$$\mathbf{V}^{-1} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{K}^{-1})^{-1}\mathbf{Z}'$$

3) BLUE and BLUP

The solution for the best linear unbiased estimators (**b**, *BLUE*) and best linear unbiased predictor (**u**, *BLUP*) are commonly defined through the equations:

$$\mathbf{b} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$
$$\mathbf{u} = \sigma_\mathrm{i}^2\mathbf{K}_\mathrm{i}\mathbf{Z}_\mathrm{i}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Xb})$$

Conditional to the random terms, BLUE coefficients can be estimated from solving the system of equation:

$$(\mathbf{X}'\mathbf{X})\,\mathbf{b} = \mathbf{X}'(\mathbf{y} - \textstyle\sum \mathbf{Z}_\mathrm{i}\mathbf{u}_\mathrm{i})$$

And coefficients of the i^{th} random effect, conditional to fixed effects and other random effects, as:

$$(\mathbf{Z}_\mathrm{i}'\mathbf{Z}_\mathrm{i} + \mathbf{K}_\mathrm{i}^{-1}\lambda_\mathrm{i})\,\mathbf{u} = \mathbf{Z}_\mathrm{i}'(\mathbf{y} - \mathbf{Xb} - \textstyle\sum \mathbf{Z}_{-\mathrm{i}}\mathbf{u}_{-\mathrm{i}})$$

Where $\lambda_\mathrm{i} = \sigma_\mathrm{e}^2\sigma_\mathrm{i}^{-2}$.

4) General terms

In a more generalized notation, the mixed model equation is described as:

$$\mathbf{y} = \mathbf{Xb} + \sum_{\mathrm{I}}^{\mathrm{i}=1} \mathbf{Z}_\mathrm{i}\mathbf{u}_\mathrm{i} + \mathbf{e} = \mathbf{Wg} + \mathbf{e}$$

Where the design matrices and coefficients are unified as:

$$\mathbf{W} = \mathbf{X}|\mathbf{Z}_1|\mathbf{Z}_{\dots}|\mathbf{Z}_\mathrm{I}$$
$$\mathbf{g} = \mathbf{b}|\mathbf{u}_1|\mathbf{u}_{\dots}|\mathbf{u}_\mathrm{I}$$

And the mixed model can be written as:

$$\mathbf{y} = \mathbf{Wg} + \mathbf{e}$$
$$\mathbf{y} \sim \mathrm{N}(\mathbf{Wg}, \mathbf{I}\sigma_\mathrm{e}^2)$$

Where the coefficients are, in probabilistic terms, described as a normal distribution with a block diagonal variance:

$$\mathbf{g} \sim \mathrm{MVN}\left(\begin{bmatrix} b \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \infty & 0 & \cdots & 0 \\ 0 & \mathbf{K}_1\sigma_1^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{K}_\mathrm{I}\sigma_\mathrm{I}^2 \end{bmatrix}\right)$$

Thus, the mixed model equation is generalized as:

$$\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma} = \mathbf{W}'\mathbf{y}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \mathbf{K}_1^{-1}\lambda_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{K}_\mathrm{I}^{-1}\lambda_\mathrm{I} \end{bmatrix}$$

And the mixed model equation makes the complex problem a simple "$\mathrm{Ax} = \mathrm{b}$" mathematic problem:

$$\mathbf{C} = \mathbf{gr}$$
$$\mathbf{C} \rightarrow\ \mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}$$
$$\mathbf{r} \rightarrow\ \mathbf{W}'\mathbf{y}$$

5) Coefficients

The conditional expectation of each regression coefficient, assuming $\mathbf{K_i} = \mathbf{I}$, is expressed as:

$$\mathrm{g_j} = \frac{\mathbf{w}_j'\left(\mathbf{y} - \mathbf{W}_{-j}\mathbf{g}_{-j}\right)}{\mathbf{w}_j'\mathbf{w}_j + \lambda_j} = \frac{\mathbf{w}_j'\tilde{\mathbf{y}}}{\mathbf{w}_j'\mathbf{w}_j + \lambda_j}$$

Where $\mathrm{w_j}$ from fixed effect coefficients assume $\lambda = 0$.

Conditional to all other terms via *Gauss-Seidel residual update* (GSRU), coefficients can be efficiently estimated in two steps:

$$\mathrm{g}_j^{t+1} = \frac{\mathbf{w}_j'\mathbf{e}^t + \mathbf{w}_j'\mathbf{w}_j\mathrm{g}_j^t}{\mathbf{w}_j'\mathbf{w}_j + \lambda_j}$$
$$\mathbf{e}^{t+1} = \mathbf{e}^t - \mathbf{w}_j(\mathrm{g}_j^{t+1} - \mathrm{g}_j^t)$$

6) Kernel trick

It is also possible to expand this framework to scenarios where $\mathbf{K_i} \neq \mathbf{I}$, that is through the diagonalization of $\mathbf{K_i}$ by spectral decomposition:

$$\mathbf{K}_i = \mathbf{U}_i\mathbf{D}_i\mathbf{U}_i'$$

Where $\mathbf{U}$ is a matrix of Eigenvectors and $\mathbf{D}$ is a diagonal matrix of Eigenvalues. With properties

$$\mathbf{K^{-1}} = \mathbf{U}\mathbf{D}^{-1}\mathbf{U}'$$
$$\mathbf{K}'\mathbf{K} = \mathbf{U}\mathbf{D}^{2}\mathbf{U}'$$
$$\mathbf{D} = \mathbf{U}'\mathbf{K}\mathbf{U}$$
$$\mathbf{I} = \mathbf{U}'\mathbf{U}$$
$$\mathbf{U}' = \mathbf{U}^{-1}$$
$$\mathrm{tr}(\mathbf{K}) = \mathrm{tr}(\mathbf{D})$$

Such that:

$$\mathbf{u_i} \sim \mathrm{N}(0, \mathbf{Z}_i\mathbf{K}_i\mathbf{Z}_i'\sigma_i^2)$$
$$\mathbf{u_i} \sim \mathbf{U}\,\mathrm{N}(0, \mathbf{Z}_i\mathbf{D}_i\mathbf{Z}_i'\sigma_i^2)$$
$$\mathbf{a_i} \sim \mathrm{N}(0, \mathbf{D}_i\sigma_i^2)$$
$$\mathrm{E}[\mathbf{u_i}] = \mathbf{U}_i\mathbf{a}_i$$
$$\mathbf{Z}_i\mathbf{u}_i = \mathbf{Z}_i\mathbf{U}_i\mathbf{a}_i$$

Assume the notation:

$$\mathbf{V}_i = \mathbf{Z}_i\mathbf{U}_i$$

Then, for i^{th} random effect and j^{th} Eigen-pair, the regression coefficient has expectation (GSRU):

$$\mathrm{g_{ij}} = \frac{\mathbf{v}'\tilde{\mathbf{y}}}{\mathbf{v}'\mathbf{v} + \lambda_i \mathrm{d}_j^{-1}}, \qquad \mathbf{v} = \mathbf{Z}_i\mathbf{U}_{ij}$$

7) Information

$$\boldsymbol{V}_i = \frac{\partial \mathbf{V}}{\partial \sigma_i^2} = \mathbf{Z_i K_i Z_i} \quad \text{and} \quad \boldsymbol{V}_0 = \frac{\partial \mathbf{V}}{\partial \sigma_e^2} = \mathbf{I}$$

$$\text{Information} = \mathrm{P}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathrm{P}\frac{\partial \mathbf{V}}{\partial \sigma_j^2} = \mathrm{PV_iPV_j}$$

8) Variance components

First derivative 1: Expectation-maximization

$$\sigma_e^2 = \frac{\mathbf{y}'\mathbf{P}\mathbf{P}\mathbf{y}}{\mathrm{tr}(\mathbf{P})} = \frac{\mathbf{e}'\mathbf{e} + \mathrm{tr}(\mathbf{W}\mathbf{C}^{-1}\mathbf{W}')\sigma_e^{2\,t-1}}{\mathrm{n}} = \frac{\mathbf{y}'\mathbf{e}}{\mathrm{n}-\mathrm{r}}$$

$$\sigma_i^2 = \frac{\mathbf{y}'\mathbf{P}\mathbf{V}^{-1}\mathbf{P}\mathbf{y}}{\mathrm{tr}(\mathbf{P}\mathbf{Z_i}\mathbf{K_i}\mathbf{Z_i})} = \frac{\mathbf{u}_i'\mathbf{K}_i^{-1}\mathbf{u}_i + \mathrm{tr}\left(\mathbf{K}_i^{-1}\mathbf{C}^{\mathbf{ii}}\right)\sigma_e^2}{\mathrm{q_i}} = \frac{\mathbf{u}_i'\mathbf{K}_i^{-1}\mathbf{u}_i}{\mathrm{q_i} - \lambda\,\mathrm{tr}(\mathbf{K}_i^{-1}\mathbf{C}^{\mathbf{ii}})}$$

Where:

n = number of observations
r = rank of $\mathbf{X}$ (number of columns of $\mathbf{X}$)
$\mathrm{q_i}$ = number of columns of $\mathbf{Z}_i$

First derivative 2: Gibbs sampling (flat priors: $S = 0, \nu = -2$)

$$\sigma_e^2 \sim \frac{\mathbf{e}'\mathbf{e} + S_e\nu_e}{\chi^2_{n+\nu_e}} \quad \text{and} \quad \sigma_i^2 \sim \frac{\mathbf{u}_i'\mathbf{K}_i^{-1}\mathbf{u}_i + S_i\nu_i}{\chi^2_{q_i+\nu_i}}$$

Second derivative 1: Newton-Raphson

$$\mathbf{v}^{t+1} = \mathbf{v}^t - \mathbf{H}^{-1}\mathbf{s}, \qquad \mathbf{v} = \begin{bmatrix}\sigma_u^2\\ \sigma_e^2\end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix}\mathbf{tr(PZKZ'PZKZ')} + 2\mathbf{y'PZKZ'PZKZ'Py} & \mathbf{tr(PZKZ'P)} + 2\mathbf{y'PZKZ'PPy}\\ \mathbf{tr(PZKZ'P)} + 2\mathbf{y'PZKZ'PPy} & \mathbf{tr(PP)} + 2\mathbf{y'PPPy}\end{bmatrix}$$

$$\mathbf{s} = \begin{bmatrix}\mathrm{tr}(\mathbf{PZKZ'}) - \mathbf{y'PZKZ'Py}\\ \mathrm{tr}(\mathbf{P}) - \mathbf{y'PPy}\end{bmatrix}$$

Second derivative 2: Fisher information

$$\mathbf{v}^{t+1} = \mathbf{v}^t - \mathbf{I}^{-1}\mathbf{s}, \qquad \mathbf{v} = \begin{bmatrix}\sigma_u^2\\ \sigma_e^2\end{bmatrix}$$

$$\mathbf{I} = \begin{bmatrix}\mathrm{tr}(\mathbf{PZKZ'PZKZ'}) & \mathrm{tr}(\mathbf{PZKZ'P})\\ \mathrm{tr}(\mathbf{PZKZ'P}) & \mathrm{tr}(\mathbf{PP})\end{bmatrix}$$

$$\mathbf{s} = \begin{bmatrix}\mathrm{tr}(\mathbf{PZKZ'}) - \mathbf{y'PZKZ'Py}\\ \mathrm{tr}(\mathbf{P}) - \mathbf{y'PPy}\end{bmatrix}$$

Second derivative 3: Average information

$$\mathbf{v}^{t+1} = \mathbf{v}^t - \mathbf{AI}^{-1}\mathbf{s}, \qquad \mathbf{v} = \begin{bmatrix}\sigma_u^2\\ \sigma_e^2\end{bmatrix}$$

$$\mathbf{AI} = 0.5 \times \begin{bmatrix}\mathrm{tr}(\mathbf{y'PZKZ'PZKZ'Py}) & \mathrm{tr}(\mathbf{y'PZKZ'P})\\ \mathrm{tr}(\mathbf{y'PZKZ'P}) & \mathrm{tr}(\mathbf{y'PPPy})\end{bmatrix}$$

$$\mathbf{s} = \begin{bmatrix}\mathrm{tr}(\mathbf{PZKZ'}) - \mathbf{y'PZKZ'Py}\\ \mathrm{tr}(\mathbf{P}) - \mathbf{y'PPy}\end{bmatrix}$$

First-Second derivative: EMMA

$$\mathbf{ZKZ}' = \mathbf{UDU}'$$

$$\mathbf{b} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$
$$\sigma_e^2 = \frac{(\mathbf{y} - \mathbf{Xb})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Xb})}{\mathrm{n} - \mathrm{r}}$$

$$\mathrm{argmax}\, L(\lambda) = -0.5\,[\,\ln|\mathbf{V}| + \ln|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + \mathbf{y}'\mathbf{P}\mathbf{y}]$$

$$\ln|\mathbf{V}| = \sum \ln\left(\mathrm{d_j}\lambda^{-1} + 1\right)$$
$$\ln|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| = \ln|\,\mathbf{X}'\mathbf{U}(\mathbf{D}^{-1}\lambda^{-1} + \mathbf{I})\mathbf{U}'\mathbf{X}\,|$$
$$\mathbf{y}'\mathbf{P}\mathbf{y} = \mathbf{y}'\mathbf{e}$$