Statistical Methods for Correlated Data

Inference for Random Effects

Michele Guindani

Department of Biostatistics
UCLA

Inference for Random effects

- Not the major focus of inference; however, it may be important to predict individual trajectories (e.g., new patient in a trial) and for model checking

- Several approaches; here we follow an Empirical Bayes approach

- Starting point: "Mixed effects" is a misnomer in a Bayesian approach, we can look at the marginal posterior of b_i

- Consider the LMM,

$$\boldsymbol{y}_i = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

  with $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i$ independent, $\boldsymbol{b}_i|\boldsymbol{D} \sim_{iid} \mathbf{N}_{q+1}(\mathbf{0}, \boldsymbol{D})$ and $\sigma_\epsilon^2 \sim_{ind} \mathbf{N}_{n_i}\left(\mathbf{0}, \sigma_\epsilon^2\mathbf{I}\right)$, $\boldsymbol{\alpha} = \left[\sigma_\epsilon^2, \boldsymbol{D}\right]$.

- Then, the posterior

$$p(\boldsymbol{b}|\boldsymbol{y}) = \iint p(\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\alpha}|\boldsymbol{y})d\boldsymbol{\beta}d\boldsymbol{\alpha}$$

  In the expression within the integral, we focus first on the (full) conditional of b_i,

$$\begin{aligned} p\left(\boldsymbol{b}_i|\boldsymbol{y}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\right) &\propto p\left(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\right) \times \pi\left(\boldsymbol{b}_i|\boldsymbol{\alpha}\right) \\ &\propto \exp\left[-\frac{1}{2\sigma_\epsilon^2}\left(\boldsymbol{y}_i^\star - \boldsymbol{z}_i\boldsymbol{b}_i\right)^{\mathrm{T}}\left(\boldsymbol{y}_i^\star - \boldsymbol{z}_i\boldsymbol{b}_i\right) - \frac{1}{2}\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{D}^{-1}\boldsymbol{b}_i\right] \end{aligned}$$

  where $\boldsymbol{y}_i^\star = \boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta}$, so that

$$\boldsymbol{b}_i\left|\boldsymbol{y}_i, \boldsymbol{\beta}, \boldsymbol{\alpha} \sim \mathbf{N}_{q+1}\left[\mathbf{E}\left(\boldsymbol{b}_i|\boldsymbol{y}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\right), \mathrm{var}\left(\boldsymbol{b}_i|\boldsymbol{y}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\right)\right]\right.$$

One can show

$$\hat{\boldsymbol{b}}_i = \mathrm{E}\left[\boldsymbol{b}_i | \boldsymbol{y}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\right] = \boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\left(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta}\right)$$

and

$$\mathrm{var}\left(\boldsymbol{b}_i | \boldsymbol{y}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\right) = \boldsymbol{D} - \boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\boldsymbol{z}_i\boldsymbol{D}$$

- See Problem 8.4 in the textbook
- $\hat{\boldsymbol{b}}_i$ can be derived under many different formulations
- Here we note that a maximizer (MAP estimate) of $p(\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\alpha} | \boldsymbol{y})$ is also a maximizer of $p(\boldsymbol{b} | \boldsymbol{y})$

Empirical Bayes Approach for prediction of random effects

- We can predict $\boldsymbol{b}_i$ using empirical estimates for $\boldsymbol{\beta}, \boldsymbol{\alpha}$
- For $\boldsymbol{\beta}$ we can take the GLS estimator $\hat{\boldsymbol{\beta}}_G$ (justified as the maximizer of the posterior when $\boldsymbol{\beta}$ is given an improper flat prior)
- Then,

$$\begin{aligned}\hat{\boldsymbol{b}}_i = \mathrm{E}\left[\boldsymbol{b}_i|\boldsymbol{y},\boldsymbol{\alpha}\right] &= \mathrm{E}_{\beta|y,\alpha}\left[\mathrm{E}\left(\boldsymbol{b}_i|\boldsymbol{\beta},\boldsymbol{y},\boldsymbol{\alpha}\right)\right] \\ &= \boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\left(\boldsymbol{y}_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}}_{\mathrm{G}}\right)\end{aligned}$$

Empirical Bayes Approach for prediction of random effects

- We can predict $\boldsymbol{b}_i$ using empirical estimates for $\boldsymbol{\beta}, \boldsymbol{\alpha}$
- For $\boldsymbol{\beta}$ we can take the GLS estimator $\hat{\boldsymbol{\beta}}_G$ (justified as the maximizer of the posterior when $\boldsymbol{\beta}$ is given an improper flat prior)
- Then,

$$
\begin{aligned}
\hat{\boldsymbol{b}}_i = \mathrm{E}\left[\boldsymbol{b}_i|\boldsymbol{y},\boldsymbol{\alpha}\right] &= \mathrm{E}_{\beta|y,\alpha}\left[\mathrm{E}\left(\boldsymbol{b}_i|\boldsymbol{\beta},\boldsymbol{y},\boldsymbol{\alpha}\right)\right] \\
&= \boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\left(\boldsymbol{y}_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}}_{\mathrm{G}}\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}\left(\boldsymbol{b}_i|\boldsymbol{y},\boldsymbol{\alpha}\right) &= \mathbf{E}_{\beta|y,\alpha}\left[\mathrm{var}\left(\boldsymbol{b}_i|\boldsymbol{\beta},\boldsymbol{y},\boldsymbol{\alpha}\right)\right] + \mathrm{var}_{\beta|y,\alpha}\left(\mathrm{E}\left[\boldsymbol{b}_i|\boldsymbol{\beta},\boldsymbol{y},\boldsymbol{\alpha}\right]\right) \\
&= \mathbf{E}_{\beta|y,\alpha}\left[\boldsymbol{D} - \boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\boldsymbol{z}_i\boldsymbol{D}\right] + \mathrm{var}_{\beta|y,\alpha}\left(\boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\left(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta}\right)\right) \\
&= \boldsymbol{D} - \boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\boldsymbol{z}_i\boldsymbol{D} + \boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\boldsymbol{x}_i\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{x}\right)^{-1}\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\boldsymbol{z}_i\boldsymbol{D}
\end{aligned}
$$

- For $\boldsymbol{\alpha}$ we use a consistent estimator, say $\hat{\boldsymbol{\alpha}}$ so that one can substitute $\widehat{\boldsymbol{D}} = \boldsymbol{D}(\widehat{\boldsymbol{\alpha}})$ and $\hat{\boldsymbol{V}} = \boldsymbol{V}(\widehat{\boldsymbol{\alpha}})$ in the previous expressions.

- The implications of the substitution of $\widehat{\boldsymbol{\beta}}_{\mathrm{G}}$ are not consequential, since it is an unbiased estimator and appears in the previous expressions in a linear fashion.
- Given correct specification of the marginal variance model, $\mathrm{var}(\boldsymbol{Y}|\boldsymbol{\alpha}) = \boldsymbol{V}(\boldsymbol{\alpha})$, and a consistent estimator of $\boldsymbol{\alpha}$, then $\widehat{\boldsymbol{b}}_i$ is asymptotically normal
- From a frequentist perspective, $\hat{b}_i$ is known as the best (empirical) linear unbiased predictor (empirical BLUP), where unbiased refers to it satisfying $\mathrm{E}\left[\widehat{\boldsymbol{b}}_i\right] = \mathrm{E}\left[\boldsymbol{b}_i\right]$.

Shrinkage toward the population mean

Shrinkage toward the population mean

- Consider the expression for the predicted response profile of the i-th individual:

$$\begin{aligned}\hat{\boldsymbol{Y}}_i &= \boldsymbol{x}_i\,\hat{\boldsymbol{\beta}} + \boldsymbol{z}_i\,\hat{\boldsymbol{b}}_i \\ &= \boldsymbol{x}_i\,\hat{\boldsymbol{\beta}}_G + \boldsymbol{z}_i\hat{\boldsymbol{D}}\boldsymbol{z}_i^{\mathrm{T}}\hat{\boldsymbol{V}}_i^{-1}\left(\boldsymbol{y}_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}}_{\mathrm{G}}\right) \\ &= (I_{n_i} - \boldsymbol{z}_i\hat{\boldsymbol{D}}\boldsymbol{z}_i^{\mathrm{T}}\hat{\boldsymbol{V}}_i^{-1})\,\boldsymbol{x}_i\hat{\boldsymbol{\beta}}_G + \boldsymbol{z}_i\hat{\boldsymbol{D}}\boldsymbol{z}_i^{\mathrm{T}}\hat{\boldsymbol{V}}_i^{-1}\boldsymbol{y}_i\end{aligned}$$

  Let $\hat{\boldsymbol{V}}_i = \boldsymbol{z}_i\hat{\boldsymbol{D}}\boldsymbol{z}_i^T + \Sigma_i$ with $\hat{\boldsymbol{\Sigma}}_i = \hat{\sigma}^2\,I_{n_i}$ typically or $\Sigma_i = \sigma^2\,R_i$ more in general. The term $\boldsymbol{z}_i\hat{\boldsymbol{D}}\boldsymbol{z}_i^T$ explains the heterogeneity across individuals, while the term Σ_i explains within-individual measurement error and serial correlation. Then,

$$\hat{\boldsymbol{V}}_i\hat{\boldsymbol{V}}_i^{-1} = I_{n_i} = (\boldsymbol{z}_i\hat{\boldsymbol{D}}\boldsymbol{z}_i^T + \hat{\sigma}^2\,I_{n_i})\,\hat{\boldsymbol{V}}_i^{-1} = \boldsymbol{z}_i\hat{\boldsymbol{D}}\boldsymbol{z}_i^T\,\hat{\boldsymbol{V}}_i^{-1} + \hat{\sigma}^2\,\hat{\boldsymbol{V}}_i^{-1}$$

- So,

$$\hat{\boldsymbol{Y}}_i = \hat{\sigma}^2\,\hat{\boldsymbol{V}}_i^{-1}\boldsymbol{x}_i\hat{\boldsymbol{\beta}}_G + \left(I_{n_i} - \hat{\sigma}^2\,\hat{\boldsymbol{V}}_i^{-1}\right)\boldsymbol{Y}_i$$

Shrinkage toward the population mean

$$\hat{\boldsymbol{Y}}_i = \hat{\sigma}^2 \, \hat{\boldsymbol{V}_i}^{-1} \boldsymbol{x}_i \hat{\boldsymbol{\beta}}_G + \left(I_{n_i} - \hat{\sigma}^2 \, \hat{\boldsymbol{V}_i}^{-1} \right) \boldsymbol{Y}_i$$

- The BLUP estimator shrinks the i-the subject response profile toward the population-averaged mean response profile.
- The amount of "shrinkage" toward the population depends on the relative magnitude of $\hat{\sigma}^2$ (more, in general, R_i, within-subject variability) and $\hat{\boldsymbol{V}}_i$ which incorporates both within-subject and between-subject sources of variability.
- If R_i is relatively "large," the within-subject variability is greater than the between-subject variability, more weight is assigned to $X_i \widehat{\beta}_G$, the estimated population-averaged mean response profile, than to the i^{th}-individual's observed responses.
- If the between-subject variability is large relative to the within-subject variability, more weight is given to the i^{th}-subject's observed responses, Y_i.

Shrinkage toward the population mean

- Intuitively, this weighting scheme is quite sensible since greater weight should be given to the i^{th}-individual's observed responses when any within-subject variability in the longitudinal responses (e.g., due to measurement error) is relatively small when compared to the natural heterogeneity in the individual-specific longitudinal response trajectories.

- On the other hand, less weight should be given to the i^{th} individual's observed responses when the within-subject variability is relatively large and the population is relatively homogeneous.

- Finally, the amount of "shrinkage" toward the population mean depends also on n_i the number of observation on the i^{th} subject. In general, there is more shrinkage toward the population mean curve when n_i is small. Intuitively, this is also quite sensible since less weight should be given to the i^{th} individual's observed responses when fewer data points are available.