Analytical Uncertainty Quantification for Predicted Random Effects in GLMMs

This note summarizes the work of Booth and Hobert (1998) as it pertains to Rowin's thesis.

1 Model

Our statistical model is a GLM with random effects. We assume that, given the random effects, our response, Y, has an exponential family distribution with the canonical parameterization. That is,

$$f(y) = \exp\left([y\theta - b(\theta)] + c(y)\right) \tag{1}$$

If Y is Bernoulli(p), then $\theta = \log[p/(1-p)]$, $b(\theta) = \log(1 + e^\theta) = -\log(1-p)$, and $c(y) = 0$. We model the parameter θ as depending on two sets of covariates, X and Z, through some fixed effects, β, and random effects, U, respectively. We assume that the observed responses occur in groups, or clusters. Write Y_{ij} for the jth observation from group i. We further assume that each group has its own level for the random effect, U_i, and that these random effects are independent and identically distributed from some common distribution, usually Normal$(0, \Sigma)$[1].

Since we're doing regression, we will focus on the conditional mean of Y given covariates. Since we're also doing mixed-effects modelling, we start by further conditioning on the random effects, U. To this end, write $\mu_{ij} = \mathbb{E}(Y | U = U_i, X = X_{ij}, Z = Z_{ij})$ for the mean of observation j from group i, when the random effects and covariates are held fixed. Following the usual GLM setup, we model $g(\mu_{ij}) = \eta_{ij} = X_{ij}\beta + Z_{ij}U_i$.

Feel free to skip this paragraph, I just wanted to rant about GLMs a bit. I've always had trouble keeping all the different scales which arise in GLMs straight in my head. There's μ, the mean of Y, η, the linear predictor, and θ, the canonical parameter[2]. There are then functions which map between these different scales: the link function, $g(\mu) = \eta$, the mean function, $b'(\theta) = \mu$, and the "variance function", $b''(\theta)$. This last one bothers me because if you include

---

[1] Booth and Hobert use G instead of Σ here.

[2] Note that this canonical parameter often differs from the parameters we're used to thinking about for the distribution in question. See, e.g., the Bernoulli distribution above, or try expressing the Normal(μ, σ) as an exponential family distribution.

nuisance (or overdispersion) parameters in your exponential family then the variance of Y is not equal to the variance function. If you choose the canonical link function though, then $g(\mu) = \theta$ (this is the definition of the canonical link), which really muddies things for me. Okay, rant over.

2 Prediction and UQ

Our current goal is the estimate/predict the values of the random effects within each group. It is common in the literature to use the word "predict" here, because the random effects aren't parameters to be estimated, but instead random variables, which we would usually talk about "predicting". Note that the different groups in our problem are independent, so we can focus on prediction of a single U_i, then apply our method individually to the other groups. Booth and Hobert recommend predicting U_i with its conditional mean given the data from group i, Y_i. Note that this conditional mean depends on the model parameters, β and Σ, which in practice must be estimated. A lot of software (e.g. the `lme4` package in `R`) actually uses the conditional mode instead of the conditional mean, since there is a nice algorithm to compute the mode, but the mean is hard (i.e. slow) in general.

Booth and Hobert discuss prediction of the linear predictor, η_{ij}, but we can just think of this as U_i. They predict the random effects within a group with their conditional mean given the observed responses from that group. See Equations (6) and (7) of Booth and Hobert. They then define the variance of this predicted random effect as its conditional variance given the group's observed responses. See Equation (8). Similarly, they define the conditional mean squared error of prediction (CMSEP) within each group conditional on the observed responses (Equation (9)). They also give a decomposition of this CMSEP in Equation (10). For simplicity, we are just going to focus on one term in this decomposition, v_i. Section 4 of Booth and Hobert goes into detail on how to estimate v_i, culminating in Equation (15). We use Equation (15) as our predictive variance of u_i, the random effect in group i.

3 Interpretation of Equation (15)

There are a bunch of symbols in Equation 15 of Booth and Hobert which we define here. First, Z_i is the observed random effects covariance matrix in group i. Next, z_i is the contrast for the random effects in group i at which we want predictions. E.g. If we want to predict $u_i[1]$ then $z_i = (1, 0, \ldots, 0)$. Next, G is the covariance matrix of u_i. That is, the covariance matrix of the random effects within a group. Note that G must be estimated. Finally, $\bar{W}_i$ is a bit complicated. Formally, it is the matrix of weights for the iteratively re-weighted least squares algorithm, evaluated at the predicted random effects in group i, $\hat{u}_i$. Let's unpack that.

In the bottom paragraph of the first column on page 265, Booth and Hobert

define the "iterative weights" as $w_{ij}/\left[\sigma_0^2 V(\mu_{ij}) g'(\mu_{ij})^2\right]$. For our purposes, both w_{ij} and σ_0^2 can be ignored (i.e. set to 1). Now onto $V(\mu)$ and $g'(\mu)$. The former, called the "variance function", can equivalently be expressed as $b''(\theta)$, where b comes directly from the exponential family density given in (1). Differentiating b twice, then returning to μ-scale, we get that $V(\mu) = \mu(1 - \mu)$. Next, g is the link function, which maps μ to the linear predictor scale. For logistic regression, $g(\mu) = \text{logit}(\mu)$. Differentiating, we get $g'(\mu) = [\mu(1 - \mu)]^{-1}$. Putting this all together, we get that the "iterative weight" for a single observation in the logistic model is $\mu_{ij}(1 - \mu_{ij})$.

Returning now to Equation (15) of Booth and Hobert, we get that $\bar{W}_i$ is a diagonal matrix with jjth element equal to $\hat{\mu}_{ij}(1 - \hat{\mu}_{ij})$. Note that the hat here denotes both estimation of parameters and prediction of u_{ij} by its conditional mean given y_i.

3.1 Interpretation as Covariance Matrix

This subsection is experimental. It's mostly based on pattern-matching formulas; I haven't done the corresponding formal derivations to be sure everything holds-up.

We can actually be a bit more general than Equation (15). Note the pre- and post-multiplication by z_i. This accounts for the chosen contrast, z_i. However, we can get the joint conditional predictive covariance matrix of all the random effects in group i by computing $(Z_i^T \bar{W}_i Z_i + G^{-1})^{-1}$ and not taking its quadratic form with z_i.

4 Implementation

It turns out that all of this has already been implemented in the `lme4` package in `R`. In fact, they focus only on v_i from the CMSEP decomposition in the same way we do.

As with most things in `lme4`, what we're looking for is a bit hidden and not well-documented. First, fit your `lmer` or `glmer` model, then call the `ranef()` function. This gives predicted random effects for each group, but also computes the predictive SDs as given by Equation (15) of Booth and Hobert. Unfortunately, these SDs are hidden in an attribute. The best way to extract them is to call `as.data.frame()` on the output of `ranef()`. This will give a well-organized data frame with columns for (among other things) the predicted random effects (labelled `condval`) and predictive SDs (labelled `condsd`).

References

James G. Booth and James P. Hobert. Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association*, 93(441), 1998.