Addendum to `vglmer` for Additional Models

Max Goplerud

September 10, 2022

Abstract

This addendum extends Goplerud (2022) to new algorithms implemented in `vglmer`. At present, results are shown for linear and count (negative binomial) outcomes.

1 Linear Model

Assume that a linear outcome is employed; the generative model is similar to the binomial case; note that the random effect variance depends on σ^2.

$$y_i \sim N(\boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{\alpha}, \sigma^2). \quad \boldsymbol{\alpha}_{j,g} \sim N(\mathbf{0}, \sigma^2\boldsymbol{\Sigma}_j); \boldsymbol{\Sigma}_j \sim p_0(\boldsymbol{\Sigma}_j); \sigma^2 \sim p_0(\sigma^2) \tag{1}$$

The new log-posterior can be expressed as follows, using notation from Goplerud 2022:

$$f(y, \boldsymbol{\theta}, \sigma^2) = \begin{array}{l} \displaystyle\sum_{i=1}^N -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} - \boldsymbol{z}_i^T\boldsymbol{\alpha})^2}{2\sigma^2} + \\ \displaystyle\sum_{j=1}^J -g_j/2\ln(|2\pi\sigma^2\boldsymbol{\Sigma}_j|) - \frac{1}{2\sigma^2}\left(\sum_{g=1}^{G_j} \boldsymbol{\alpha}_{j,g}^T\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\alpha}_{j,g}\right) + \ln p_0(\boldsymbol{\Sigma}_j) + \\ \ln p_0(\sigma^2) \end{array} \tag{2}$$

I adjust the factorization assumption such that $q(\boldsymbol{\theta})q(\sigma^2)$ where $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\alpha}, \{\boldsymbol{\Sigma}_j\}\}$ and their factorization assumption is discussed elsewhere. I assume a conditionally conjugate prior of $p_0(\sigma^2) \sim \text{InverseGamma}(a_0, b_0)$. This implies that $q(\sigma^2)$ is also Inverse-Gamma. The software uses an improper prior of $a_0 = 0, b_0 = 0$, i.e. $p_0(\sigma^2) \propto 1/\sigma^2$.

$$
\begin{aligned}
&q(\sigma^2) \sim \text{InverseGamma}\left(a_\sigma, b_\sigma\right) \\
&a_\sigma = \left(\frac{N + \sum_{j=1}^J d_j G_j}{2}\right) + a_0 \\
&b_\sigma = \frac{\sum_{i=1}^N E_{q(\boldsymbol{\theta})}[(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta} - \boldsymbol{z}_i^T \boldsymbol{\alpha})^2] + \sum_{j=1}^J \sum_{g=1}^{G_j} E_{q(\boldsymbol{\theta})}[\text{tr}\left(\boldsymbol{\Sigma}_j^{-1} \boldsymbol{\alpha}_{j,g} \boldsymbol{\alpha}_{j,g}^T\right)]}{2} + b_0
\end{aligned}
\tag{3}
$$

Further, some relevant expectations for $q(\sigma^2)$ are noted below:

$$
\begin{aligned}
&E_{q(\sigma^2)}[1/\sigma^2] = a_\sigma / b_\sigma \quad E_{q(\sigma^2)}[\ln \sigma^2] = \ln\left(b_\sigma\right) - \psi(a_\sigma) \\
&E_{q(\sigma^2)}[\ln q(\sigma^2)] = a_\sigma + \ln(b_\sigma) + \ln \Gamma(a_\sigma) - (a_\sigma + 1)\psi(a_\sigma)
\end{aligned}
\tag{4}
$$

The updates for $q(\boldsymbol{\beta}, \boldsymbol{\alpha})$ are straightforward. The update for $\boldsymbol{\Sigma}_j$ is adjusted to account for σ^2. Assuming an Inverse-Wishart prior $p_0(\boldsymbol{\Sigma}_j) \sim \text{IW}(\nu_0, \boldsymbol{\Phi}_0)$, the update is as follows given the independence between σ^2 and $\boldsymbol{\alpha}$ in the variational approximation:

$$
\tilde{\nu}_j = \nu_j + g_j; \quad \tilde{\boldsymbol{\Phi}}_j = \boldsymbol{\Phi}_j + E_{q(\sigma^2)}[1/\sigma^2] \left(\sum_{g=1}^{g_j} E_{q(\boldsymbol{\alpha})}[\boldsymbol{\alpha}_g \boldsymbol{\alpha}_g^T]\right)
\tag{5}
$$

2 Negative Binomial

The generative model is standard and ensures that $E[y_i] = \exp\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)$ while $Var(y_i) = E[y_i]\left(1 + E[y_i]/r\right)$. Thus, r is interpretable in the usual way as an dispersion parameter where $r \to \infty$ recovers the original Poisson model. This differs from other negative binomial implementations (e.g. Pillow and Scott 2012; Zhou et al. 2012) but matches standard practice in applied social scientific research. This definition of r agrees with implementations in `R` (e.g. `glm.nb`) and with the "reciprocal dispersion" from `rstanarm`.

$$y_i \sim \mathrm{NB}\left(r, 1-p_i\right) \quad p_i = \frac{\exp(\psi_i)}{1+\exp(\psi_i)} \quad \psi_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{\alpha} - \ln r \tag{6a}$$

$$f(y_i) = \frac{\Gamma(y_i + r)}{\Gamma(y_i+1)\Gamma(r)}\frac{\exp(\psi_i)^{y_i}}{(1+\exp(\psi_i))^{y_i+r}} \tag{6b}$$

The same hierarchical structure is placed on $\boldsymbol{\alpha}$ as described in Goplerud (2022). Applying Polya-Gamma augmentation yields the following augmented likelihood:

$$f(y_i, \omega_i) = \frac{\Gamma(y_i+r)}{\Gamma(y_i+1)\Gamma(r)} 2^{-(y_i+r)} \exp\left([y_i - r]/2\psi_i - \omega_i\psi_i^2\right) f_{PG}(\omega_i | y_i + r, 0) \tag{7}$$

Using $\boldsymbol{\theta}$ to collect $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\{\boldsymbol{\Sigma}_j\}$, and $\{\omega_i\}$, the complete data log-posterior can be expressed as follows where $p_0(r)$ is the prior placed on r and $p_0(\boldsymbol{\Sigma}_j)$ is the prior on $\boldsymbol{\Sigma}_j$:

$$\begin{gathered} f(\boldsymbol{y}, \boldsymbol{\theta}, r) = \sum_{i=1}^N \ln\Gamma(y_i + r) - \ln\Gamma(r) - \ln\Gamma(y_i+1) - (y_i + r)\ln(2) + \\ \sum_{i=1}^N (y_i - r)/2\left(\boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{\alpha} - \ln r\right) - \omega_i/2\left(\boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{\alpha} - \ln r\right)^2 + \ln f_{PG}(\omega_i | y_i + r, 0) \\ \left[\sum_{j=1}^J -G_j/2\ln(|2\pi\boldsymbol{\Sigma}_j|) - \frac{1}{2}\left(\sum_{g=1}^{G_j} \boldsymbol{\alpha}_{j,g}^T \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\alpha}_{j,g}\right) + \ln p_0(\boldsymbol{\Sigma_j})\right] + \ln p_0(r) \end{gathered} \tag{8}$$

It is useful to characterize the ELBO in two ways; first, the main objective ELBO that depends on $q(\boldsymbol{\theta})$ and $q(r)$. Second, a "conditional" ELBO (ELBO^c) that holds r fixed and depends thus only on $q(\boldsymbol{\theta})$.

$$
\begin{aligned}
\text{ELBO}(q(\boldsymbol{\theta}, q(r)) &= E_{q(r)}\left[\text{ELBO}^c(q(\boldsymbol{\theta}); r) - \ln q(r)\right] \\
\text{ELBO}^c(q(\boldsymbol{\theta}); r) &= E_{q(\boldsymbol{\theta})}\left[f(\boldsymbol{y}, \boldsymbol{\theta}, r) - \ln q(\boldsymbol{\theta})\right]
\end{aligned}
\tag{9}
$$

Considering first the conditional ELBO, it is clear this has updates of a nearly identical form to those in Goplerud (2022). Thus, for any fixed r, the conditional ELBO can be maximized. When the expectation is taken over r, however, the main ELBO becomes intractable because of both the log-gamma terms involving r but also the log-Polya-Gamma density ($\ln f_{PG}(\omega_i|y_i + r)$). To proceed, I outline two strategies.

First, one can assume that $q(r)$ has a degenerate point-mass distribution and thus perform variational EM with r in the 'M-Step' and all other parameters in the variational E-Step. One proceeds by optimizing $q(\boldsymbol{\theta})$ given r and then maximizing $\text{ELBO}^c(q(\boldsymbol{\theta}; r)$ over r. The CAVI updates for ELBO^c are straightforward and involve only slight adjustments to the updates for the logistic case. All of the variational distributions maintain the same form. Updating r is more challenging insofar as ELBO^c involves an intractable expectation over $q(\omega_i)$. Note, however, that this can be evaluated and optimized after profiling out the Polya-Gamma parameters, i.e. noting that $\tilde{b}_i$ and $\tilde{c}_i$ are functions of the other variational parameters and r and thus can be substituted out. The profiled ELBO (PELBO^c) is shown below:

$$
\begin{aligned}
&\mathrm{PELBO}^c(q(\boldsymbol{\theta}, r)) = \sum_{i=1}^{N} \ln\Gamma(y_i + r) - \ln\Gamma(r) - \ln\Gamma(y_i + 1) - (y_i + r)\ln(2) + \\
&\frac{1}{2}(\boldsymbol{y} - r)^T[\boldsymbol{X}\tilde{\boldsymbol{\mu}}_\beta + \boldsymbol{Z}\tilde{\boldsymbol{\mu}}_\alpha - \ln r] + E_{q(\boldsymbol{\alpha})q\left(\{\boldsymbol{\Sigma}_j\}_{j=1}^{J}\right)}[\ln p(\boldsymbol{\alpha})] + \sum_{j=1}^{J} E_{q(\boldsymbol{\Sigma}_j)}\left[\ln p(\boldsymbol{\Sigma}_j)\right] + \ln p_0(r) \\
&\frac{1}{2}\ln\left[2\pi e|\tilde{\boldsymbol{\Lambda}}_{\alpha-\beta}|\right] + \sum_{i=1}^{N} -(y_i + r)\ln\cosh\left[\frac{1}{2}\sqrt{\begin{array}{c}\left[\boldsymbol{x}_i^T\tilde{\boldsymbol{\mu}}_\beta + \boldsymbol{z}_i^T\tilde{\boldsymbol{\mu}}_\alpha - \ln r\right]^2 + \\ {[\boldsymbol{x}_i^T, \boldsymbol{z}_i^T]}\tilde{\boldsymbol{\Lambda}}_{\beta-\alpha}\begin{bmatrix}\boldsymbol{x}_i \\ \boldsymbol{z}_i\end{bmatrix}\end{array}}\right] + \\
&\sum_{j=1}^{J} E_{q(\boldsymbol{\Sigma}_j)}[-\ln q(\boldsymbol{\Sigma}_j)]
\end{aligned}
\tag{10}
$$

This objective monotonically increases at each iteration and thus can be monitored for convergence. The limitation of this approach is that it does not quantify uncertainty in r nor propagate it through to the other parameters. Future work might extend this to allow for quantification of uncertainty in r.