Hyperparameter specification for `covdepGE`

Author: Jacob Helwig & Debdeep Pati

Created: 1/11/22

Last edit: 01/12/22

0.1 `varbvs` approach to hyperparameter specification

For $X' \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$, the default algorithm for `varbvs` begins by initializing hyperparameters as follows:

$$\sigma^2 = \text{var}(y) \text{ (the error term variance)}$$
$$\sigma^2_\beta = 1 \text{ (the slab variance)}$$
$$\underset{\sim}{\frac{\pi}{1-\pi}} \text{ (vector of 20 candidates for the prior odds of inclusion)}$$

The components of $\underset{\sim}{\frac{\pi}{1-\pi}}$ are equally spaced and range from $\frac{\pi_1}{1-\pi_1} = -\log_{10}(p)$ to $\frac{\pi_{20}}{1-\pi_{20}} = -1$.

Next, for the j-th value in $\underset{\sim}{\frac{\pi}{1-\pi}}$, $\frac{\pi_j}{1-\pi_j}$, CAVI is performed to obtain a variational estimate for μ_j (the posterior mean of the non-zero regression coefficients) and α_j (the posterior inclusion probabilities). At each step of the CAVI, after the variational update is performed, σ^2_j and $\sigma^2_{\beta_j}$ are both updated using the current values of the variational parameters and hyperparameters. The update for σ^2_j is performed using maximum likelihood estimation and the update for $\sigma^2_{\beta_j}$ is performed using maximum a posteriori estimation. Finally, after the CAVI has concluded, the lower bound to the marginal log-liklihood given the hyperparameters, $\log(w_j)$, is calculated.

This process is repeated for each of the components of $\underset{\sim}{\frac{\pi}{1-\pi}}$, resulting in a σ^2_j and a $\sigma^2_{\beta_j}$ corresponding to each $\frac{\pi_j}{1-\pi_j}$ and having been fit to the data. Additionally, to each $\frac{\pi_j}{1-\pi_j}$ there will be a $\log(w_j)$. Then, $\log(w_k)$ is chosen such that it maximizes the lower bound over $\underset{\sim}{\frac{\pi}{1-\pi}}$. CAVI is then repeated for all of the values in $\underset{\sim}{\frac{\pi}{1-\pi}}$, however, this time, the initial values for each σ^2_j and $\sigma^2_{\beta_j}$ are not $\text{var}(y)$ and 1, but rather σ^2_k and $\sigma^2_{\beta_k}$, respectively. Similarly, the starting values for each μ_j and α_j are μ_k and α_k. As with the first CAVI, the values of σ^2_j and $\sigma^2_{\beta_j}$ are updated at each iteration using maximum liklihood estimation and maximum a posteriori estimation.

Upon completing CAVI for each of the $\frac{\pi_j}{1-\pi_j}$, updated values are returned as $\underset{\sim}{\sigma^2}$ and $\underset{\sim}{\sigma^2_\beta}$, where the j-th value of these vectors corresponds to $\frac{\pi_j}{1-\pi_j}$. Finally, the posterior inclusion probabilities are calculated as $\underset{\sim}{\alpha}\underset{\sim}{w}$, where $\underset{\sim}{w}$ is the vector of normalized lower bounds and the j-th column of $\underset{\sim}{\alpha}$ is α_j.

0.2 Our approach to hyperparameter specification

In our algorithm, we define y as the i-th column of X and X' as X with the i-th column removed. We then rely on `varbvs` to obtain our hyperparameters for the i-th variable by first running the `varbvs` algorithm on X' and y as described in the previous section. To select hyperparameters, we set $\sigma^2 = \bar{\underset{\sim}{\sigma}}^2$ (sample mean of $\underset{\sim}{\sigma}^2$). We transform $\underset{\sim}{\frac{\pi}{1-\pi}}$ to $\underset{\sim}{\pi}$ and then set $\pi = \bar{\underset{\sim}{\pi}}$. We then define a vector $\underset{\sim}{\sigma}_\beta^2$ of slab variance candidates independent of the results obtained from `varbvs` and independent of the data. We perform CAVI for each value in $\underset{\sim}{\sigma}_\beta^2$, selecting the final value for the i-th variable to be the component of $\underset{\sim}{\sigma}_\beta^2$ that maximizes the ELBO.

This approach seems inefficient for several reasons. First, it opts to reduce the values of $\underset{\sim}{\sigma}^2$ to $\bar{\underset{\sim}{\sigma}}^2$ and $\underset{\sim}{\pi}$ to $\bar{\underset{\sim}{\pi}}$ instead of exploring the different elements of the hyperparameter space $(\sigma_j^2, \sigma_{\beta_j}^2, \pi_j)$ that have been fit to the data in conjunction with one another. It also ignores entirely the values of $\underset{\sim}{\sigma}_\beta^2$ that have been fit to the data and instead generates a grid independent of the data and the other two hyperparameters. Thus, a modification that could perhaps make our approach more effective would be to perform CAVI for each of the elements of the hyperparameter space $(\sigma_j^2, \sigma_{\beta_j}^2, \pi_j)$ estimated by `varbvs` for the i-th variable and select the element $(\sigma_k^2, \sigma_{\beta_k}^2, \pi_k)$ that maximizes the ELBO.

0.3 Conclusion and next steps

I believe that this modified approach would not only result in `covdepGE` utilizing more sensible values for $(\sigma_j^2, \sigma_{\beta_j}^2, \pi_j)$, but also would result in faster convergence, since the hyperparameters are fit to the data. Although it is true that these hyperparameters have been fit to the data under the assumption that the conditional dependence structure is homogeneous across all of the individuals, it seems that our current choice of hyperparameters is informed by the data even less.

To explore this hypothesis, I will perform an experiment where I generate the continuous data multiple times. I will then call `covdepGE` twice, first using the hyperparameter specification scheme that we are currently using, and then using the specification scheme that I suggest here. I will compare the distribution of the difference in the total model ELBO using each scheme within each data generation.