Hyperparameter updates for `covdepGE`

Author: Jacob Helwig

Created: 1/13/22

Last edit: 01/13/22

0.1 Notation

For data $X \in \mathbb{R}^{n\times(p+1)}$, let X_j be the j-th column of X, and X_{-j} be X with the j-th column removed. Fix $l \in \{1, .., n\}$ and the j-the variable as the response X_j, and let $w_{i,l}$ be the weight of individual i with respect to individual l. Then, β_j^l denotes the weighted regression effect of X_{-j} on X_j, and $\gamma_{j,k}^l$ is a Bernoulli random variable that takes on 1 if $\beta_{j,k}^l$ is non-zero.

It is assumed that:

$$x_{i,j}|\sigma^2, \beta_j^l \sim \mathcal{N}(x_{i,-j}^\top \beta_j^l, \sigma^2/w_{i,l}) \qquad \gamma_{j,k}^l \sim \text{Bern}(\pi) \tag{0.1}$$

Thus, the weighted, conditional likelihood of X_j for the l-th individual is given by:

$$L_l^w(j) \propto \prod_{i=1}^n (\sigma^2)^{-1/2} \exp\left(-\frac{(x_{i,j} - x_{i,-j}^\top \beta_j^l)^2 w_{i,l}}{2\sigma^2}\right) \tag{0.2}$$

With the joint prior of (β_j^l, γ_j^l):

$$p_0(\beta_j^l, \gamma_j^l) = \prod_{k=1}^{p-1} \delta_{\{0\}}\left(\beta_{j,k}^l\right)^{1-\gamma_{j,k}^l} \mathcal{N}\left(\beta_{j,k}^l; 0, \sigma^2\sigma_\beta^2\right)^{\gamma_{j,k}^l} \pi^{\gamma_{j,k}^l}(1-\pi)^{1-\gamma_{j,k}^l} \tag{0.3}$$

0.2 CAVI with hyperparameter updates

The joint posterior of (β_j^l, γ_j^l) is parameterized by μ_j^l, $(s_j^l)^2$, and α_j^l. $\mu_{j,k}^l$ and $(s_{j,k}^l)^2$ are the mean and variance of the Gaussian part of the marginal posterior distribution of $\beta_{j,k}^l$, and $\alpha_{j,k}^l$ is the probability that $\beta_{j,k}^l$ is non-zero. To produce a variational approximation to these parameters, CAVI is performed for fixed values of the hyperparameters σ^2, σ_β^2, and π.

The following sections detail how, for a fixed value of π, σ^2 and σ_β^2 may be updated during each CAVI iteration so that they are fit to the data. The update for σ^2 uses MLE, while the update for σ_β^2 uses maximum a posteriori estimation.

These updates, as specified by `varbvs`, are provided in the last section.

0.3 MLE update for σ^2

From (0.2), the weighted conditional log-likelihood is given by:

$$\log(L_l^w(j)) \propto \sum_{i=1}^{n} -\log(\sigma^2) - \frac{(x_{i,j} - x_{i,-j}^\top \beta_j^l)^2 w_{i,l}}{\sigma^2} \tag{0.4}$$

Differentiating with respect to σ^2 gives:

$$\frac{\partial(\log(L_l^w(j)))}{\partial\sigma^2} \propto \sum_{i=1}^{n} -\frac{1}{\sigma^2} + \frac{(x_{i,j} - x_{i,-j}^\top \beta_j^l)^2 w_{i,l}}{(\sigma^2)^2} \tag{0.5}$$

Setting equal to 0 and solving for σ^2 gives:

$$\sigma^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - x_{i,-j}^\top \beta_j^l)^2 w_{i,l} \tag{0.6}$$

As this is the conditional weighted likelihood**, conditioning on $\beta_j^l = \mu_j^l$ gives:

***I am not sure if this is the correct or complete reasoning*

$$\sigma^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - x_{i,-j}^\top \mu_j^l)^2 w_{i,l} \tag{0.7}$$

0.4 MAPE update for σ_β^2

The joint posterior of (β_j^l, γ_j^l) is given by:

$$p(\beta_j^l, \gamma_j^l | X) \propto L_l^w(j) p_0(\beta_j^l, \gamma_j^l) \tag{0.8}$$

Thus, the log posterior is given by:

$$\begin{aligned}
\log(p(\beta_j^l, \gamma_j^l|X)) &\propto \log(L_l^w(j)) + \log(p_0(\beta_j^l, \gamma_j^l)) && (0.9)\\
&\propto \log(L_l^w(j)) + \log\left\{\prod_{k=1}^{p-1} \delta_{\{0\}}\left(\beta_{j,k}^l\right)^{1-\gamma_{j,k}^l} \pi^{\gamma_{j,k}^l}(1-\pi)^{1-\gamma_{j,k}^l}\right\} && (0.10)\\
&+ \sum_{k=1}^{p-1}\gamma_{j,k}^l \log(\frac{1}{\sqrt{2\pi\sigma^2}}) - \frac{1}{2}\log(\sigma_\beta^2)\sum_{k=1}^{p-1}\gamma_{j,k}^l - \frac{1}{2}(\sigma^2\sigma_\beta^2)^{-1}\sum_{k=1}^{p-1}\gamma_{j,k}^l(\beta_{j,k}^l)^2 && (0.11)
\end{aligned}$$

Differentiating with respect to σ_β^2 gives:

$$\frac{\partial \log(p(\beta_j^l, \gamma_j^l|X))}{\partial\sigma_\beta^2} \propto -\frac{1}{2}(\sigma_\beta^2)^{-1}\sum_{k=1}^{p-1}\gamma_{j,k}^l + \frac{1}{2\sigma^2}(\sigma_\beta^2)^{-2}\sum_{k=1}^{p-1}\gamma_{j,k}^l(\beta_{j,k}^l)^2 \tag{0.12}$$

Setting equal to 0 and solving for σ_β^2 gives:

$$(\sigma_\beta^2)_{MAPE} = \frac{\sum_{k=1}^{p-1} \gamma_{j,k}^l (\beta_{j,k}^l)^2}{\sigma^2 \sum_{k=1}^{p-1} \gamma_{j,k}^l} \tag{0.13}$$

Approximating the value of the unknown parameters by their expected values** (that is, $\gamma_{j,k}^l$ with $\alpha_{j,k}^l$ and $\beta_{j,k}^l$ with $\mu_{j,k}^l$):

***Note: I am certain that this is not the correct reasoning for plugging in the variational approximations*

$$(\sigma_\beta^2)_{MAPE} = \frac{\sum_{k=1}^{p-1} \alpha_{j,k}^l (\mu_{j,k}^l)^2}{\sigma^2 \sum_{k=1}^{p-1} \alpha_{j,k}^l} \tag{0.14}$$

0.5 MAPE update for σ_β^2 with hyperprior

`varbvs` imposes a scaled inverse chi-squared hyperprior on σ_β^2 with degrees of freedom n_0 and scale parameter s_0. That is:

$$\sigma_\beta^2 \sim f_0(\sigma_\beta^2) \propto \frac{\exp\left(\frac{-n_0 s_0}{2\sigma_\beta^2}\right)}{(\sigma_\beta^2)^{1+n_0/2}} \tag{0.15}$$

This section is a copy of the previous section with the addition of this hyperprior.

The joint posterior of (β_j^l, γ_j^l) is given by:

$$p(\beta_j^l, \gamma_j^l | X) \propto L_l^w(j) p_0(\beta_j^l, \gamma_j^l) f_0(\sigma_\beta^2) \tag{0.16}$$

Thus, the log posterior is given by:

$$\begin{aligned}
\log(p(\beta_j^l, \gamma_j^l | X)) &\propto \log(L_l^w(j)) + \log(p_0(\beta_j^l, \gamma_j^l)) && (0.17)\\
&\propto \log(L_l^w(j)) + \log\left\{\prod_{k=1}^{p-1} \delta_{\{0\}}\left(\beta_{j,k}^l\right)^{1-\gamma_{j,k}^l} \pi^{\gamma_{j,k}^l}(1-\pi)^{1-\gamma_{j,k}^l}\right\} && (0.18)\\
&\quad + \sum_{k=1}^{p-1} \gamma_{j,k}^l \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2}\log(\sigma_\beta^2)\sum_{k=1}^{p-1}\gamma_{j,k}^l - \frac{1}{2}(\sigma^2\sigma_\beta^2)^{-1}\sum_{k=1}^{p-1}\gamma_{j,k}^l(\beta_{j,k}^l)^2 && (0.19)\\
&\quad - \frac{n_0 s_0}{2\sigma_\beta^2} - (1+n_0/2)\log(\sigma_\beta^2) && (0.20)
\end{aligned}$$

Differentiating with respect to σ_β^2 gives:

$$\frac{\partial \log(p(\beta_j^l, \gamma_j^l | X))}{\partial \sigma_\beta^2} \propto -\frac{1}{2}(\sigma_\beta^2)^{-1}\sum_{k=1}^{p-1}\gamma_{j,k}^l + \frac{1}{2\sigma^2}(\sigma_\beta^2)^{-2}\sum_{k=1}^{p-1}\gamma_{j,k}^l(\beta_{j,k}^l)^2 + (\sigma_\beta^2)^{-2}\frac{n_0 s_0}{2} - (1+n_0/2)(\sigma_\beta^2)^{-1} \tag{0.21}$$

Setting equal to 0 and solving for σ_β^2 gives:

$$(\sigma_\beta^2)_{MAPE} = \frac{\sum_{k=1}^{p-1} \gamma_{j,k}^l (\beta_{j,k}^l)^2 + n_0 s_0}{\sigma^2(2 + n_0 + \sum_{k=1}^{p-1} \gamma_{j,k}^l)} \tag{0.22}$$

Approximating the value of the unknown parameters by their expected values** (that is, $\gamma_{j,k}^l$ with $\alpha_{j,k}^l$ and $\beta_{j,k}^l$ with $\mu_{j,k}^l$):

***Note: I am certain that this is not the correct reasoning for plugging in the variational approximations*

$$(\sigma_\beta^2)_{MAPE} = \frac{\sum_{k=1}^{p-1} \alpha_{j,k}^l (\mu_{j,k}^l)^2 + n_0 s_0}{\sigma^2(2 + n_0 + \sum_{k=1}^{p-1} \alpha_{j,k}^l)} \tag{0.23}$$

0.6 `varbvs` hyperparameter updates

Note that these updates are for an unweighted regression. These updates can be found in this `varbvs` script, under sections 2d and 2e. Unfamiliar helper functions, such as `dot` and `betavar` can be found in this `varbvs` script.

Let $X_r = Xr$, where r is a vector whose k-th entry is $\mu_{j,k}^l \alpha_{j,k}^l$

$$\sigma_{MLE}^2 = \left\{ \|y - X_r\|_2^2 + \sum_{k=1}^{p-1} (X_k^\top X_k) \alpha_{j,k}^l \left[(s_{j,k}^l)^2 + (1 - \alpha_{j,k}^l)(\mu_{j,k}^l)^2 \right] + \frac{1}{\sigma_\beta^2} \sum_{k=1}^{p-1} \alpha_{j,k}^l ((s_{j,k}^l)^2 + (\mu_{j,k}^l)^2) \right\} / (n + \sum_{k=1}^{p-1} \alpha_{j,k}^l) \tag{0.24}$$

$$(\sigma_\beta^2)_{MAPE} = \frac{s_0 n_0 + \sum_{k=1}^{p-1} \alpha_{j,k}^l ((s_{j,k}^l)^2 + (\mu_{j,k}^l)^2)}{n_0 + \sigma^2 \sum_{k=1}^{p-1} \alpha_{j,k}^l} \tag{0.25}$$