Hyperparameter updates for `covdepGE`

*Author: Jacob Helwig*

*Created: 1/13/22*

*Last edit: 01/15/22*

## 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, $\beta_j^l$ 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 $(\beta_j^l, \gamma_j^l)$:

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

## 0.2 CAVI with hyperparameter updates

The joint posterior of $(\beta_j^l, \gamma_j^l)$ is parameterized by $\mu_j^l$, $(s_j^l)^2$, and $\alpha_j^l$. $\mu_{j,k}^l$ 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 $\sigma^2$, $\sigma_\beta^2$, and $\pi$.

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

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

## 0.3 MLE update for $\sigma^2$

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 $\sigma^2$ 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 $\sigma^2$ gives:

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

**Estimating the value of the parameter $\beta_j^l$ using its posterior expected value:

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

***I don't believe that this is the correct reasoning*

## 0.4 MAPE update for $\sigma_\beta^2$

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

$$\sigma_\beta^2 \sim f_0(\sigma_\beta^2) \propto \frac{\exp\left(\frac{-n_0 s_0}{2\sigma_\beta^2}\right)}{(\sigma_\beta^2)^{1+n_0/2}} \tag{0.8}$$

The joint posterior of $(\beta_j^l, \gamma_j^l)$ is given by:

$$p(\beta_j^l, \gamma_j^l | X) \propto L_l^w(j) p_0(\beta_j^l, \gamma_j^l) f_0(\sigma_\beta^2) \tag{0.9}$$

Thus, the log posterior is given by:

$$\begin{aligned}
\log(p(\beta_j^l, \gamma_j^l|X)) &\propto \log(L_l^w(j)) + \log(p_0(\beta_j^l, \gamma_j^l)) && (0.10)\\
&\propto \log(L_l^w(j)) + \log\left\{\prod_{k=1}^{p-1} \delta_{\{0\}}\left(\beta_{j,k}^l\right)^{1-\gamma_{j,k}^l}\left(\frac{1}{\sqrt{\sigma^2}}\right)^{\gamma_{j,k}^l} \pi^{\gamma_{j,k}^l}(1-\pi)^{1-\gamma_{j,k}^l}\right\} && (0.11)\\
&\quad - \frac{1}{2}\log(\sigma_\beta^2)\sum_{k=1}^{p-1}\gamma_{j,k}^l - \frac{1}{2}(\sigma^2\sigma_\beta^2)^{-1}\sum_{k=1}^{p-1}\gamma_{j,k}^l(\beta_{j,k}^l)^2 - \frac{n_0 s_0}{2\sigma_\beta^2} - (1+n_0/2)\log(\sigma_\beta^2) && (0.12)
\end{aligned}$$

Differentiating with respect to $\sigma_\beta^2$ gives:

$$\frac{\partial \log(p(\beta_j^l, \gamma_j^l | X))}{\partial \sigma_\beta^2} \propto -\frac{1}{2}(\sigma_\beta^2)^{-1}\sum_{k=1}^{p-1}\gamma_{j,k}^l + \frac{1}{2\sigma^2}(\sigma_\beta^2)^{-2}\sum_{k=1}^{p-1}\gamma_{j,k}^l(\beta_{j,k}^l)^2 + (\sigma_\beta^2)^{-2}\frac{n_0 s_0}{2} - (1+n_0/2)(\sigma_\beta^2)^{-1} \quad (0.13)$$

Setting equal to 0 and solving for $\sigma_\beta^2$ gives:

$$(\sigma_\beta^2)_{MAPE} = \frac{\sum_{k=1}^{p-1}\gamma_{j,k}^l(\beta_{j,k}^l)^2 + n_0 s_0}{\sigma^2(2 + n_0 + \sum_{k=1}^{p-1}\gamma_{j,k}^l)} \quad (0.14)$$

**Estimating the value of the parameters $\beta_j^l$ and $\gamma_j^l$ using their posterior expected values:

$$(\sigma_\beta^2)_{MAPE} = \frac{\sum_{k=1}^{p-1}\alpha_{j,k}^l(\mu_{j,k}^l)^2 + n_0 s_0}{\sigma^2(2 + n_0 + \sum_{k=1}^{p-1}\alpha_{j,k}^l)} \quad (0.15)$$

***I don't believe that this is the correct reasoning*

## 0.5 `varbvs` hyperparameter updates

Note that these updates are for an unweighted regression of $X$ onto $y$. 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 $\mu_{j,k}^l\alpha_{j,k}^l$

$$\sigma_{MLE}^2 = \frac{\|y - X_r\|_2^2 + \sum_{k=1}^{p-1}\left\{(X_k^\top X_k)\alpha_{j,k}^l\left[(s_{j,k}^l)^2 + (1-\alpha_{j,k}^l)(\mu_{j,k}^l)^2\right]\right\} + \frac{1}{\sigma_\beta^2}\sum_{k=1}^{p-1}\left\{\alpha_{j,k}^l((s_{j,k}^l)^2 + (\mu_{j,k}^l)^2)\right\}}{n + \sum_{k=1}^{p-1}\alpha_{j,k}^l} \quad (0.16)$$

$$(\sigma_\beta^2)_{MAPE} = \frac{s_0 n_0 + \sum_{k=1}^{p-1}\alpha_{j,k}^l((s_{j,k}^l)^2 + (\mu_{j,k}^l)^2)}{n_0 + \sigma^2\sum_{k=1}^{p-1}\alpha_{j,k}^l} \quad (0.17)$$