Variable Selection with Sum of Single Effects Model (SuSiE)

- Uses the technique of *sum of single effects* regression
  - The *single effects* model assumes that exactly one of the predictors has a nonzero coefficient
  - Single-effect model: $1 \times p$ vector $\boldsymbol{\beta} = \beta\gamma$ has exactly one nonzero element, where $\gamma \sim Mult(1, \pi)$, $\beta \sim N_1(0, {\sigma_0}^2)$
- **Sum of single effects model**:
  - Introduce L single-effect vectors $\boldsymbol{\beta}_l = \beta_l \gamma_l$, where $\gamma_l \sim Mult(1, \pi)$, $\beta_l \sim N_1(0, {\sigma_{0l}}^2)$
  - Then $\boldsymbol{\beta} = \sum_{l=1}^{L} \boldsymbol{\beta_l}$ gives the sum
  - $L \ll p$ gives standard BVSR
- Approximately equivalent to BVSR, but with two computational advantages
  - Deterministic algorithm for approximate posterior distributions → *variational method*
  - Simple calculation of credible sets
    - *Credible set* of level ρ is a subset of variables with probability ρ or greater of containing at least one effect variable (useful for quantifying uncertainty in which variables to select)
- Given $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{L-1}$, estimating $\boldsymbol{\beta}_L$ reduces to fitting a single effects model
  - Leads to an iterative algorithm the estimates $\boldsymbol{\beta}_l$ with a single effects model given the current estimates for other $\boldsymbol{\beta}_{l'}$, $l' \neq l$
  - Effect of predictor j is $\beta^{(j)} = \sum_{l=1}^{L} \beta_{lj}$
- Model is robust to large L
  - If L is overstated, the uncertainty in the model spreads out extra effects among the covariates and makes little difference overall
  - Key signals remain the same
- **Iterative Bayesian Stepwise Selection (IBSS)**
  - Essentially a variational method
  - Relies on fact (above) that given $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{L-1}$, estimating $\boldsymbol{\beta}_L$ reduces to fitting a single effects model
  - Requires that we initialize number of effects L and hyperparameters, and initialize posterior means
  - Then cycle through each effect vector and estimate it given the others, continue recycling through until convergence criterion satisfied
  - Result: gives an approximation of the posterior of each effect vector
  - **Scales linearly with data size**