**BurdenEM: Concomitant variable mixture model for effect size distribution**

Consider the following hierarchical model

Where: (observed/fixed variables; latent variables; inferred variables)

* are observed counts of *de novo* mutations for gene [length of : number of genes]
* is the expected rate of *de novo* mutations for gene [length of : number of genes]
* is the gene-wise rate ratio for *de novo* mutations for gene . The distribution of across genes is the estimand of the method. [length of : number of genes]
* is the number of individuals in the study [scalar]
* is a latent variable that matches genes to mixture components [length of : number of genes]
* specify the parameters of the mixture components by scaling the distribution [length of : number of mixture components]
* specify the mixing weights for the mixture components for gene . Note that these are inferred *indirectly* by inferring [length of : number of mixture components]
* are the *concomitant variables*; the mixing weights are a linear combination of the concomitant variables. Each row of is a convex combination; this requirement could be relaxed with a link function such as the logit. [dimension of : number of genes x number of features]
* are the coefficients that connect the concomitant variables to the mixing weights . [length of : number of concomitant variables. Dimension of coefficient matrix : number of concomitant variables x number of mixture components]

At a high level, this model seeks to fit a flexible mixture distribution to estimate the distribution of gene-wise effect sizes (in units of rate ratio). The mixing weights are allowed to vary as a function of gene-wise annotations (i.e. constraint).

The log likelihood of given de novo counts and mixture assignments is:

The distribution implied by this likelihood is a compound Poisson-Uniform distribution. To simplify, we approximate this compound distribution as a scaled Poisson distribution with parameter related to the mixture component parameter :

The likelihood of each observation, for each component, can be efficiently pre-computed. Note that it is also possible to compute this likelihood as the average across a scaling grid for , which can buffer against misspecification of mixture components.

We denote the matrix of likelihoods, where rows correspond to genes and columns correspond to mixture components, as .

We first initiate some values of .

*E-Step*

In the E-step, we compute posterior probabilities for each gene belonging to each component, conditioning on the current values of .

We first compute the weights from the concomitant variables and current values of . These weights can be interpreted as the probabilities that each gene is drawn from each mixture component.

We then compute the posterior probabilities (i.e., using information about the observed counts) for each observation corresponding to each component.

Where represents the element-wise product. Note that is then normalized such that rows of sum to 1.

*M-step*

In the M-step, we seek to find the values of that maximize the likelihood of the posterior probabilities computed in the E-step. In particular, we seek to solve:

The least-squares minimizing solution is given by:

We repeat until convergence.