1 The Model

Let $i = 1, ..., n$ denote subjects, $j = 1, ..., p$ cell subsets and $l = 1, ..., L_{ij}$ the observations we have for each subject/subset combination. Assume the following model:

$$\nu \sim \mathrm{N}_p(0, \Sigma), \qquad z \in \{0, 1\} \sim \mathrm{Ising}(\Theta),$$

$$\mathrm{logit}(\mu_{ijl}) = X_{ijl}\beta_j + z_{ij}T_{ijl}\tau_j + \nu_j,$$

$$p_{ijl} \sim \mathrm{Beta}(M_j\mu_{ijl}, M_j(1 - \mu_{ijl})), \qquad y_{ijl} \sim \mathrm{Binom}(N_{il}, \mu_{ijl}).$$

In the model X_{ijl} corresponds to fixed effects that are not dependent on the treatment e.g. age, gender etc.. T_{ijl} is a vector of covariates corresponding to the treatment effect, in the simplest case this would just be an indicator for stimulation but in more interesting cases T_{ijl} may be a binary vector for which stimulation was introduced to a sample or indicators for time varying effects.

Conditional on ν_i and z_i, y_{ijl} follows a beta-binomial distribution and so, it has expectation and variance:

$$E(y_{ijl}/N_{il}|\nu_i, z_i) = \mu_{ijl},$$

$$Var(y_{ijl}/N_{il}|\nu_i, z_i) = \frac{\mu_{ijl}(1 - \mu_{ijl})}{N_{il}} + \frac{\mu_{ijl}(1 - \mu_{ijl})}{M_j + 1}.$$

And so, the beta-binomial model captures the fact that regardless of how many cells we get to see (N_{il}), we cannot obtain perfect information regarding the data generating process based on a single blood sample.

2 Computation

2.1 An EM Formulation

Estimating the mixed mixture model is difficult because the likelihood involves both a high-dimensional integral and a summation over 2^p possible response assignments:

$$\mathcal{L}(\beta, \tau, \Sigma, \Theta) = \prod_{i=1}^{n} \left\{ \sum_{z \in \{0,1\}^p} \int_{\mathcal{R}^p} P(z)\varphi(\nu; \Sigma) \prod_{j=1}^{p} \prod_{l=1}^{L_{ij}} f(y_{ijl}|z, \nu) d\nu \right\}.$$

A common method for maximizing such complex likelihood functions is the EM algorithm. In the EM algorithm we maximize the complete-information log-likelihood where we replace the unknown quantities with their expectation conditional on the observed data as dictated by the current parameter estimates.

$$Q\left(\{\beta, \tau, \Sigma, \Theta\}^t \mid |\{\beta, \tau, \Sigma, \Theta\}^{t-1}\right) = \sum_{i=1}^{n} E(\log f(y, z, \nu)|y)$$

$$= \sum_{i=1}^{n} \sum_{z \in \{0,1\}^p} P(z|y) E\left[\log f(y, \nu, z) | z, y\right]$$

$$= \sum_{i=1}^{n} \sum_{z \in \{0,1\}^p} P(z|y) \int_{\mathcal{R}^p} f(\nu|y, z) \log f(y, \nu, z) d\nu.$$

with

$$\log f(y_i, z_i, \nu_i) = \log P(z_i) + \log \varphi(\nu_i; \Sigma) + \sum_{j=1}^{p} \sum_{l=1}^{L_{ij}} \log f(y_{ijl}|z, \nu).$$

The EM complete-data log-likelihood is still intractable because it involves the same high-dimensional integrals and summations, but they are suggestive of the possibility of approximating the full-information log-likelihood using Monte-Carlo integration. Let $(\nu_i^*, z_i^*)_1, ..., (\nu_i^*, z_i^*)_M$ be joint samples from the posterior distribution of ν and z given y. Then, the complete information log-likelihood can be approximated with:

$$\frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{n} \log f(y, \nu_{im}^*, z_{im}^*).$$

Replacing the expectation step with a posterior sampling step yields a Stochastic-EM (SEM) algorithm. We discuss the implementation of the SEM algorithm in our setting next.

2.2 Posterior Sampling for the Stochastic-EM

We sample from the posterior joint distribution of ν_i and z_i in two stages. First, we sample from the marginal posterior distribution of z_i and then sample $\nu_i|z_i$. We start by describing a Gibbs sampler for sampling z_i. For an arbitrary index $j \in \{1, ..., p\}$, the posterior probability of response in the jth subset can be written as:

$$P(z_j = 1|z_{-j}) = \frac{P(z_j = 1|z_{-j}) \int f(y, \nu|z_{-j}, z_j = 1) \varphi(\nu; \Sigma) d\nu}{\sum_{k=0}^{1} P(z_j = k|z_{-j}) \int f(y, \nu|z_{-j}, z_j = k) \varphi(\nu; \Sigma) d\nu}$$

we perform numerical integration using importance sampling.

Conditionally on z_i, we sample ν_i using a Metropolis-Hastings, sampling a proposal r_{ij} for ν_{ij} from:

$$r_{ij} \sim N(\nu_{ij}, c_j \Sigma_{j,j})$$

and keep the proposal with probability:

$$P(r_{ij}) = \min \left\{ \frac{f(y_{ij}|z_i, r_{ij}) \varphi(r_{ij}|\nu_{i,-j})}{f(y_{ij}|z_i, \nu_{ij}) \varphi(\nu_{ij}|\nu_{i,-j})}, 1 \right\}.$$

in the proposal, c_j is tuned on the fly to obtain an acceptance rate of about 0.234.

2.3 M-Step

2.3.1 Regression Model

With posterior samples on hand, it is straightforward to obtain updated parameter estimates for $\{\beta, \tau, \Sigma, \Theta\}$. We estimate β and τ using standard regression analysis, our implementation includes the following options:

- Binomial regression using the **glm** functions.
- Firth regression for data with separation using the **brglm2** package.
- Sparse regression using the **glmnet** package.
- Robust regression using the **robustbase** package.

In practice, robust regression seems to work best for flow-cytometry data.

2.3.2 Ising Model

We estimate the parameters of the Ising model Θ using neighborhood selection, using a similar methodology as in the **IsingFit** package with some adjustments for handling cell-subsets with very low response. Specifically, for each cell-subset separately we estimate the model:

$$\text{logit}(P(z_j = 1|z_{-j})) = \theta_0 + \sum_{t \neq j} \theta_t z_t,$$

By maximizing the ℓ_1 penalized log-likelihood:

$$\hat{\theta}_j(\lambda) = \arg\max_{\theta_j} \ell(\theta_j) - \lambda \|\theta_j\|_1$$

where λ is selected in such a way as to maximize the EBIC:

$$\begin{aligned}\lambda^* &= \arg\max_{\lambda} EBIC(\lambda) \\ &= \arg\max_{\lambda} \ell(\hat{\theta}_j(\lambda)) - \frac{1}{2}\|\hat{\theta}_j(\lambda)\|_0 \log n - 2\gamma \|\hat{\theta}_j(\lambda)\|_0 \log(p-1), \qquad \gamma \in [0,1].\end{aligned}$$

Once a neighborhood model was estimated for each cell-subset, we use an AND or an OR rule to decide which off-diagonal elements of Θ should be set to non-zero values and average the non-diagonal elements to obtain a symmetric adjacency matrix.

2.3.3 Covariance of Random Effects

We provide three methods for estimating the covariance of the random effects.

- Estimating a diagonal covariance (independence model).
- Estimating a (naive) dense covariance matrix.
- Estimating a sparse covariance using the **PDSCE** package.

2.4 Technicalities

2.4.1 Initialization

The regression coefficients are initialized based on fitting a naive regression model which assumes that all subjects are responders and that there are no random effects. So,

$$\text{logit}(\mu_{ijl}^0) = X_{ijl}\beta_j^0 + T_{ijl}\tau_j^0$$

2.4.2 Averaging Parameter Estimates

Given that our fitting algorithm is a stochastic one, it makes sense to average our final parameter estimates across iterations. Let $T \geq 1$ be the total number of SEM iterations we perform and $0 \leq D < T$ be a pre-specified parameter. Further, let $\{\tilde{\beta}, \tilde{\tau}, \tilde{\Theta}, \tilde{\Sigma}\}^t$ be the parameter estimates obtained from the M-step at the tth iteration. Define $w^t = \max(1, T - D)^{-1}$. At the tth iteration we set:

$$\{\hat{\beta}, \hat{\tau}, \hat{\Theta}, \hat{\Sigma}\}^t = (1 - w^t)\{\hat{\beta}, \hat{\tau}, \hat{\Theta}, \hat{\Sigma}\}^{t-1} + w^t\{\tilde{\beta}, \tilde{\tau}, \tilde{\Theta}, \tilde{\Sigma}\}^t$$

The rationale behind only averaging the last $T - D$ iterations is that the EM algorithm may take several iterations before converging to the vicinity of a local a maxima.

Similarly, we maintain estimates of the posterior probabilities for of response for each subject. Let $p_{ij} = P(z_{ij} = 1)$. Our estimate for the the posterior response probability of for the ith subject is:

$$\hat{p}_i = \frac{1}{T - D} \sum_{t=D+1}^{T} \frac{1}{M} \sum_{m=1}^{M} z_{im}^t.$$

We maintain a running average of the random effect estimates in a similar manner.

2.4.3 Gradual Tuning of the Dispersion Parameters