Estimate Null Correlation in MASH (OLD)

Yuxin Zou

November 30, 2018

1 Background

The MASH model is

$$\hat{\boldsymbol{b}}_j|\boldsymbol{b}_j, \hat{\boldsymbol{S}}_j \sim N_R(\boldsymbol{b}_j, \hat{\boldsymbol{S}}_j\boldsymbol{V}\hat{\boldsymbol{S}}_j) \tag{1.1}$$

$$\boldsymbol{b}_j|\boldsymbol{\pi} \sim \sum_{k=1}^{K}\sum_{l=1}^{L}\pi_{kl}N_R(\boldsymbol{0}, \omega_l\boldsymbol{U}_k) \tag{1.2}$$

Let P = KL, $\Sigma_p = \omega_l\boldsymbol{U}_k$. We can re-write 1.2 as

$$\boldsymbol{b}_j|\boldsymbol{\pi} \sim \sum_{p=1}^{P}\pi_p N_R(\boldsymbol{0}, \Sigma_p) \tag{1.3}$$

We want to estimate $\boldsymbol{V}$ and $\boldsymbol{\pi}$ by maximum likelihood.

$$p(\hat{\boldsymbol{B}}) = \prod_{j=1}^{J} p(\hat{\boldsymbol{b}}_j) = \prod_{j=1}^{J}\sum_{p=1}^{P}\pi_p N_R(\hat{\boldsymbol{b}}_j; \boldsymbol{0}, \hat{\boldsymbol{S}}_j\boldsymbol{V}\hat{\boldsymbol{S}}_j + \Sigma_p) \tag{1.4}$$

Specifically, we estimate them by coordinate ascend. Given $\boldsymbol{V}$, we estimate $\boldsymbol{\pi}$ by solving a convex problem. Given $\boldsymbol{\pi}$, we want to estimate $\boldsymbol{V}$ by maximum likelihood.

2 Method

It is hard to estimate $\boldsymbol{V}$ by maximizing log of (1.4), so we use EM algorithm. We augment each $\hat{\boldsymbol{b}}_j$ with corresponding $\boldsymbol{b}_j$. The complete likelihood is

$$p(\hat{\boldsymbol{B}}, \boldsymbol{B}) = \prod_{j=1}^{J} N_R(\hat{\boldsymbol{b}}_j; \boldsymbol{b}_j, \hat{\boldsymbol{S}}_j \boldsymbol{V} \hat{\boldsymbol{S}}_j) \sum_{p=1}^{P} \left[\pi_p N_R(\ vb_j; \boldsymbol{0}, \Sigma_p)\right] \tag{2.1}$$

2.1 E step

Taking expectations of log (2.1), we have

$$\mathbb{E}_{\boldsymbol{B}|\hat{\boldsymbol{B}}} \log p(\hat{\boldsymbol{B}}, \boldsymbol{B}) \tag{2.2}$$

$$= \mathbb{E}_{\boldsymbol{B}|\hat{\boldsymbol{B}}} \left[\sum_{j=1}^{J} \log N_R(\hat{\boldsymbol{b}}_j; \boldsymbol{b}_j, \hat{\boldsymbol{S}}_j \boldsymbol{V} \hat{\boldsymbol{S}}_j) + \log \sum_{p=1}^{P} \left[\pi_p N_R(\boldsymbol{b}_j; \boldsymbol{0}, \Sigma_p)\right] \right] \tag{2.3}$$

$$= \sum_{j=1}^{J} -\frac{1}{2} \log |\boldsymbol{V}| - \log |\hat{\boldsymbol{S}}_j| - \frac{1}{2} \mathbb{E}_{\boldsymbol{b}_j|\hat{\boldsymbol{b}}_j} \left[(\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j)^T \hat{\boldsymbol{S}}_j^{-1} \boldsymbol{V}^{-1} \hat{\boldsymbol{S}}_j^{-1} (\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j) \right] + C \tag{2.4}$$

where C is a constant that does not depend on $\boldsymbol{V}$.

2.2 M step

We maximize (2.2) over $\boldsymbol{V}$. There is a constraint on $\boldsymbol{V}$, the diagonal of $\boldsymbol{V}$ must be 1. Let $\boldsymbol{V} = \boldsymbol{DCD}$, $\boldsymbol{C}$ is the covariance matrix, $\boldsymbol{D} = diag(1/\sqrt{\boldsymbol{C}_{jj}})$.

$$f(\boldsymbol{C}) = \sum_{j=1}^{J} -\frac{1}{2} \log |\boldsymbol{V}| - \frac{1}{2} \mathbb{E}_{\boldsymbol{b}_j|\hat{\boldsymbol{b}}_j} \left[(\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j)^T \hat{\boldsymbol{S}}_j^{-1} \boldsymbol{V}^{-1} \hat{\boldsymbol{S}}_j^{-1} (\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j) \right] \tag{2.5}$$

$$= \sum_{j=1}^{J} -\frac{1}{2} \log |\boldsymbol{DCD}| - \frac{1}{2} \mathbb{E}_{\boldsymbol{b}_j|\hat{\boldsymbol{b}}_j} \left[(\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j)^T \hat{\boldsymbol{S}}_j^{-1} \boldsymbol{D}^{-1} \boldsymbol{C}^{-1} \boldsymbol{D}^{-1} \hat{\boldsymbol{S}}_j^{-1} (\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j) \right] \tag{2.6}$$

$$f(\boldsymbol{C})' = \sum_{j=1}^{J} -\frac{1}{2} \boldsymbol{C}^{-1} + \frac{1}{2} \boldsymbol{C}^{-1} \boldsymbol{D}^{-1} \hat{\boldsymbol{S}}_j^{-1} \mathbb{E} \left((\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j)(\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j)^T | \hat{\boldsymbol{b}}_j \right) \hat{\boldsymbol{S}}_j^{-1} \boldsymbol{D}^{-1} \boldsymbol{C}^{-1} = 0 \tag{2.7}$$

$$\boldsymbol{C} = \frac{1}{J} \boldsymbol{D}^{-1} \left[\sum_{j=1}^{J} \hat{\boldsymbol{S}}_j^{-1} \mathbb{E} \left((\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j)(\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j)^T | \hat{\boldsymbol{b}}_j \right) \hat{\boldsymbol{S}}_j^{-1} \right] \boldsymbol{D}^{-1} \tag{2.8}$$

We can update $\boldsymbol{C}$ and $\boldsymbol{V}$ as

$$\hat{\boldsymbol{C}}_{(t+1)} = \hat{\boldsymbol{D}}_{(t)}^{-1} \frac{1}{J} \left[\sum_{j=1}^{J} \hat{\boldsymbol{S}}_j^{-1} \mathbb{E}\left[(\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j)(\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j)^T | \hat{\boldsymbol{b}}_j \right] \hat{\boldsymbol{S}}_j^{-1} \right] \hat{\boldsymbol{D}}_{(t)}^{-1} \tag{2.9}$$

$$\hat{\boldsymbol{D}}_{(t+1)} = diag(1/\sqrt{\hat{\boldsymbol{C}}_{(t+1)jj}}) \tag{2.10}$$

$$\hat{\boldsymbol{V}}_{(t+1)} = \hat{\boldsymbol{D}}_{(t+1)} \hat{\boldsymbol{C}}_{(t+1)} \hat{\boldsymbol{D}}_{(t+1)} \tag{2.11}$$

The resulting $\hat{\boldsymbol{V}}_{(t+1)}$ is equivalent as

$$\hat{\boldsymbol{C}}_{(t+1)} = \frac{1}{J} \left[\sum_{j=1}^{J} \hat{\boldsymbol{S}}_j^{-1} \mathbb{E}\left[(\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j)(\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j)^T | \hat{\boldsymbol{b}}_j \right] \hat{\boldsymbol{S}}_j^{-1} \right] \tag{2.12}$$

$$\hat{\boldsymbol{D}}_{(t+1)} = diag(1/\sqrt{\hat{\boldsymbol{C}}_{(t+1)jj}}) \tag{2.13}$$

$$\hat{\boldsymbol{V}}_{(t+1)} = \hat{\boldsymbol{D}}_{(t+1)} \hat{\boldsymbol{C}}_{(t+1)} \hat{\boldsymbol{D}}_{(t+1)} \tag{2.14}$$

We notice that updating $\hat{\boldsymbol{V}}$ requires the posterior of $\boldsymbol{b}_j$, which is obtained by mash model.

The algorithm is

Algorithm 1 Estimate Null Correlation

Require: mash data, covariance matrices $\boldsymbol{U}$s, $\boldsymbol{\pi}$, initial value of $\boldsymbol{V}$

1: **repeat**
2: E step: compute the posterior distribution of $\boldsymbol{b}$
3: Update $\boldsymbol{C} \leftarrow$ 2.12
4: Convert $\boldsymbol{C}$ to $\boldsymbol{V} \leftarrow$ 2.14
5: Compute loglikelihood
6: **until** loglikelihood does not change
7: **return** $\boldsymbol{V}$