M&M ASH model and inference

October 8, 2016

Notations

Notation	Definition
J	Number of conditions
N	Sample size
$Y \mapsto \mathrm{R}^{N\times J}$	Phenotypes
$X \mapsto \mathrm{R}^{N\times P}$	Genotypes (SNPs)
$B \mapsto \mathrm{R}^{P\times J}$	Genotypic effect size of P SNPs in J conditions
$\beta_p \mapsto \mathrm{R}^{J}$	Genotypic effect size of one SNP (a row in B)
$\Sigma \mapsto \mathrm{R}^{J\times J}$	Residual covariance
$U \mapsto \mathrm{R}^{K\times J\times J}$	Prior matrices (candidate models)
$\omega \mapsto \mathrm{R}^{L}$	Grid values, scales of U

M&M ASH model

Consider a multivariate, multiple regression problem

$$\begin{aligned} Y &= XB + E \\ E &\sim \mathcal{MN}(0, I_N, \Sigma) \end{aligned}$$

Or equivalently,

$$Y \mid X, B, \Sigma \sim \mathcal{MN}(XB, I_N, \Sigma) \tag{1}$$

The goal is to make inference on effect size B. We use a unimodal mixture prior (`ash` prior) for B

$$B \mid \pi, U, \omega \sim \sum_k \sum_l \pi_{k,l} \mathcal{MN}(0, I_p, \omega_l U_k) \tag{2}$$

where ω_l are given grid values, U_k are given matrices and the mixture components $\pi_{k,l}$ are learned from data. This is the multivariate + multiple regression extension of the `ash` model (Stephens 2016).

Variational inference

We use variational inference to solve the model. The objective function to minimize is the K-L divergence:

$$\begin{aligned} F &= E_q \log \frac{q(B)}{p(B)p(Y|X,B)} && (3) \\ &= E_q \log q(B) - E_q \log p(B) - E_q \log p(Y|X,B) && (4) \end{aligned}$$

where

$$B = \left\{ \begin{array}{c} \beta_1^T \\ \beta_2^T \\ \dots \\ \beta_P^T \end{array} \right\}$$

We take a variational approach assuming that

$$q(B) = \prod_p q(\beta_p) \tag{5}$$

where

$$q(\beta_p) = \sum_t \alpha_{pt} N(\mu_{pt}, S_{pt}) \tag{6}$$

We then minimize F and estimate model parameters.

Known Σ

For starters we assume Σ is known; thus without loss of generality we set $\Sigma = I_J$. This is because if we scale B by $\Sigma^{-\frac{1}{2}}$ we can consider the equivalent model:

$$\begin{array}{rcl} Y\Sigma^{-\frac{1}{2}} & = & XB\Sigma^{-\frac{1}{2}} + E\Sigma^{-\frac{1}{2}} \\ E\Sigma^{-\frac{1}{2}} & \sim & \mathcal{MN}(0, I_N, I_J) \end{array}$$

and the equivalent model

$$Y \mid X, B, \Sigma \sim \mathcal{MN}(XB, I_N, I_J) \tag{7}$$

For (2), let

$$V_t = \omega_l U_k$$

we re-parameterize the prior

$$B \mid \pi, W \sim \sum_t \pi_t \mathcal{MN}(0, I_p, V_t) \tag{8}$$

We denote that

$$r_p = E_q \beta_p = \sum_t \alpha_{pt} \mu_{pt}$$

and

$$Y^T X = \left\{ \begin{array}{c} \phi_1 \\ \phi_2 \\ \dots \\ \phi_P \end{array} \right\}$$

We work out terms in (3)

$$
\begin{aligned}
E_q \log p(Y|X,B) &= -\frac{NJ}{2}\log 2\pi - \frac{1}{2}E_q\{tr[(Y-XB)^T(Y-XB)]\} &(9)\\
&= -\frac{NJ}{2}\log 2\pi - \frac{1}{2}E_q\{tr[Y^TY - Y^TXB - B^TX^TY + B^TX^TXB]\} &(10)\\
&= c_1 + tr[E_q(B^T)X^TY] - \frac{1}{2}E_q[tr(\sum_{i=1}^P\sum_{p=1}^P\sum_{k=1}^N \beta_i X_{ki}X_{kp}\beta_p^T)] &(11)\\
&= c_1 + tr(\sum_p r_p\phi_p^T) - \frac{1}{2}tr[\sum_{i=1}^P\sum_{p=1}^P\sum_{k=1}^N X_{ki}X_{kp}E_q(\beta_i\beta_p^T)] &(12)\\
&= c_1 + tr(\sum_p r_p\phi_p^T) - \frac{1}{2}tr(\sum_{i=1}^P\sum_{p=1}^P\sum_{k=1}^N X_{ki}X_{kp}r_ir_p^T)\\
&+ \frac{1}{2}tr(\sum_{p=1}^P\sum_{k=1}^N X_{kp}X_{kp}r_pr_p^T) - \frac{1}{2}tr[\sum_{p=1}^P\sum_{k=1}^N X_{kp}X_{kp}E_q(\beta_p\beta_p^T)] &(13)\\
&= c_1 + tr(\sum_p r_p\phi_p^T) - \frac{1}{2}tr(\sum_{i=1}^P\sum_{p=1}^P\sum_{k=1}^N X_{ki}X_{kp}r_ir_p^T) + \frac{1}{2}tr(\sum_{p=1}^P\sum_{k=1}^N X_{kp}X_{kp}r_pr_p^T)\\
&- \frac{1}{2}tr[\sum_{p=1}^P\sum_{k=1}^N X_{kp}X_{kp}\sum_t \alpha_{pt}(\mu_{pt}\mu_{pt}^T + S_{pt})] &(14)
\end{aligned}
$$

$$
E_q \log p(B) = \sum_p\sum_t \alpha_{pt}\{\log\pi_t - \frac{J}{2}\log 2\pi - \frac{1}{2}\log|V_t| - \frac{1}{2}tr[V_t^{-1}(\mu_{pt}\mu_{pt}^T + S_{pt})]\} \qquad (15)
$$

$$
E_q \log q(B) = \sum_p\sum_t \alpha_{pt}(\log\alpha_{pt} - \frac{J}{2}\log 2\pi - \frac{1}{2}\log|S_{pt}| - \frac{J}{2}) \qquad (16)
$$

The first derivatives

$$
\begin{aligned}
\frac{\partial F}{\partial S_{pt}} &= \frac{1}{2}\alpha_{pt}(d_pI - S_{pt}^{-1} + V_t^{-1}) &(17)\\
\frac{\partial F}{\partial \mu_{pt}} &= \alpha_{pt}(-\phi_p + \sum_{i=1}^P\sum_{k=1}^N X_{ki}X_{kp}r_i - d_pr_p + d_p\mu_{pt} + V_t^{-1}\mu_{pt}) &(18)\\
\frac{\partial F}{\partial \alpha_{pt}} &= -\mu_{pt}^T\phi_p + \sum_{i=1}^P\sum_{k=1}^N X_{ki}X_{kp}\mu_{pt}^Tr_i - d_p\mu_{pt}^Tr_p + \frac{1}{2}d_ptr(\mu_{pt}\mu_{pt}^T + S_{pt})\\
&+ \log\frac{\alpha_{pt}}{\pi_t} - \frac{1}{2}\log\frac{|S_{pt}|}{|V_t|} + \frac{1}{2}tr[V_t^{-1}(\mu_{pt}\mu_{pt}^T + S_{pt})] - \frac{J}{2} + 1 &(19)
\end{aligned}
$$

where $d_p = \sum_k X_{kp}X_{kp}$. The solutions are

$$
\begin{aligned}
S_{pt} &= (d_pI + V_t^{-1})^{-1} &(20)\\
\mu_{pt} &= S_{pt}(\phi_p - \sum_i [X^TX]_{ip}r_i + [X^TX]_{pp}r_p) &(21)\\
\alpha_{pt} &\propto \pi_t\sqrt{\frac{|S_{pt}|}{|V_t|}}\exp\{\frac{1}{2}\mu_{pt}^TS_{pt}^{-1}\mu_{pt}\} &(22)\\
\sum_t \alpha_{pt} &= 1
\end{aligned}
$$

or in another notation

$$
\begin{aligned}
S_{pt} &= ([X^TX]_{pp}I + V_t^{-1})^{-1} && (23)\\
\mu_{pt} &= S_{pt}([Y^TX]_p - \sum_{i\neq p}[X^TX]_{ip}r_i) && (24)\\
\alpha_{pt} &\propto \pi_t\sqrt{\frac{|S_{pt}|}{|V_t|}}\exp\{\frac{1}{2}\mu_{pt}^TS_{pt}^{-1}\mu_{pt}\} && (25)\\
\sum_t \alpha_{pt} &= 1
\end{aligned}
$$

We iterate this procedure until F converges.

Unknown Σ

We now treat $\Sigma_{J\times J}$ unknown and estimate it in the VB procedure. When J is large, for example $J = 50$, we will have an additional 2,500 parameters to estimate at each iteration; this is likely problematic. There are two simplifications to this: 1) Assuming Σ is diagonal, or 2) Assuming Σ is low rank.

When Σ is Low rank

Let

$$\Sigma = \sigma^2I + WW^T \quad (26)$$

where W is low rank. Then removing genetic effect the model equivalent to (1) is

$$Z = Y - XB \sim \mathcal{MN}(0, I_N, \sigma^2I + WW^T) \quad (27)$$

Using plug-in estimate of B, (27) may be solved in the VB procedure along the lines of Tipping and Bishop (1999) (their EM updates with respect to W).