

Supporting Information

January 30, 2026

S.1 Model

For each variant, we model its phenotypic effects as shown in Figure 1. We assume that the effect of the variant on a single endophenotype (E) determines its effects on a number of other traits (T_1, T_2, \dots, T_q). The effect on each trait T_i then determines its measured effect on that trait O_i . Furthermore, conditional on observing the true effects T_i on all traits, the observed trait effects O_i are conditionally independent. Finally, conditional on observing the effect on the endophenotype E , the true effects T_i are conditionally independent of each other. Dependencies among traits not captured by E could be added, but we omit these for now.

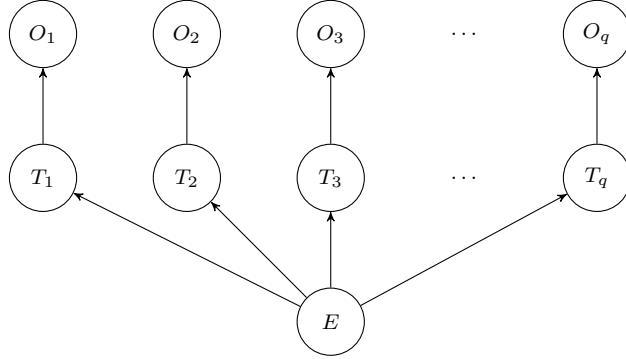


Figure 1: Model for phenotypic effects of a variant

We then specify conditional probability distributions as:

$$E \sim \mathcal{N}(\mu, \tau^2)$$

$$T_i \sim \mathcal{N}(\beta_i E, \sigma_i^2)$$

$$O_i \sim \mathcal{N}(T_i, s_i^2)$$

The intuition behind the parameters is as follows. The β_i control the relative values of the effects of the endophenotype on each of the traits, or in other words the “phenotypic profile”. They are relative values; if a variant has a two-fold effect on the endophenotype relative to another variant, it will have two-fold stronger effects on each of the traits, but the relative effects across traits will be the same between the two variants. The σ_i parameters control how important each trait is in the phenotypic profile; small values will insist that variants with the same endophenotype effects E have very similar effects on the trait, while larger values will allow variants with similar endophenotype effects to have discrepant trait effects.

μ and τ control the prior population distribution of E . Many times μ will be set to 0; however, in some cases (such as modeling variants from within a disease gene) it may be desirable to have $\mu \neq 0$. τ controls the “prevalence” of the endophenotype; small values of τ will bias more variants to have small effects E on the endophenotype.

The s_i values control the sampling distribution of the observed association statistics. They will depend on the frequency of the variant and the population in which it is tested for association.

Specification of these distribution allows us to answer several questions. For example:

1. **What is the posterior distribution of effects on one trait T_i , given observed effect on other traits $O_1, \dots, O_{i-1}, O_{i+1}, \dots, O_q$?** We would run the query $\Pr(T_i | O_2, \dots, O_q)$. The means of this posterior could be used for weights in an aggregate test, or the distributions could be used to filter variants for inclusion. Alternatively, this could be used as a prior in a Bayesian association analysis.
2. **What variants affect specified endophenotype?** For each variant, we would compute $\Pr(E | T_1, \dots, T_q)$.

Fitting the model

To use this model, we need to learn the parameters $\theta = \mu, \tau, \beta_i, \sigma_i$, and s_i .

For s_i , we assume that the variance of the observed coefficient O_i is equivalent under the null distribution ($T_i = 0$) and the alternate distribution. Thus, we can use the value of s_i fit under the null model, for example as determined by an association test. s_i will vary differ from variant to variant.

For fitting the remaining parameters, we consider two options. First, any number of the parameters θ could be externally specified. For example, to model an endophenotype based on lipodystrophy, relative values for β_i , as well as their variances in the population σ_i , could be measured epidemiologically.

Additionally, unspecified (or “unconstrained”) parameters can be fit given a collection of training variants $\hat{\mathbf{O}} = \hat{\mathbf{O}}^1, \dots, \hat{\mathbf{O}}^j$ where each training variant $\hat{\mathbf{O}}^j$ is a tuple of observed effect sizes $\hat{o}_1^j, \dots, \hat{o}_q^j$. The training algorithm will then learn values for unconstrained parameters such that the likelihood of the training variants is maximized. Specifically, we seek to maximize

$$L(\theta; \hat{\mathbf{O}}) = \int_{\hat{\mathbf{T}}^{\hat{E}}} L(\theta; \hat{\mathbf{O}}, \hat{\mathbf{T}}, \hat{E})$$

where $\hat{\mathbf{T}} = \hat{\mathbf{T}}^1, \dots, \hat{\mathbf{T}}^N$ and $\hat{E} = \hat{E}^1, \dots, \hat{E}^N$, or the values of T_i and E for each training sample j , are treated as unobserved latent data.

As is customary, we employ the EM algorithm. In the E-step, we estimate

$$\begin{aligned}
Q(\theta | \theta^{(t)}) &= \text{E}_{\hat{\mathbf{T}}, \hat{E} | \hat{\mathbf{O}}, \theta^{(t)}} \left[\log L(\theta; \hat{\mathbf{O}}, \hat{\mathbf{T}}, \hat{E}) \right] \\
&= \text{E}_{\hat{\mathbf{T}}, \hat{E} | \hat{\mathbf{O}}, \theta^{(t)}} \left[\log \left(\Pr(\hat{\mathbf{O}} | \hat{\mathbf{T}}) \Pr(\hat{\mathbf{T}} | \hat{E}) \Pr(\hat{E}) \right) \right] \\
&= \text{E}_{\hat{\mathbf{T}}, \hat{E} | \hat{\mathbf{O}}, \theta^{(t)}} \left[\log \Pr(\hat{\mathbf{O}} | \hat{\mathbf{T}}) + \log \Pr(\hat{\mathbf{T}} | \hat{E}) + \log \Pr(\hat{E}) \right] \\
&= \sum_j \text{E}_{\hat{\mathbf{T}}^j, \hat{E}^j | \hat{\mathbf{O}}^j, \theta^{(t)}} \left[\sum_i \log \left(\frac{1}{s_i^j \sqrt{2\pi}} e^{-\frac{(o_i^j - \hat{T}_i^j)^2}{2(s_i^j)^2}} \right) + \sum_i \log \left(\frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(\hat{T}_i^j - \beta_i \hat{E}^j)^2}{2\sigma_i^2}} \right) \right. \\
&\quad \left. + \log \left(\frac{1}{\tau \sqrt{2\pi}} e^{-\frac{(\hat{E}^j - \mu)^2}{2\tau^2}} \right) \right] \\
&= \sum_j \text{E}_{\hat{\mathbf{T}}^j, \hat{E}^j | \hat{\mathbf{O}}^j, \theta^{(t)}} \left[-\sum_i \frac{(o_i^j - T_i^j)^2}{2(s_i^j)^2} - \sum_i \frac{(T_i^j - \beta_i E^j)^2}{2\sigma_i^2} - \frac{(E^j - \mu)^2}{2\tau^2} \right. \\
&\quad \left. - \frac{1}{2} \sum_i \log \sigma_i^2 - \frac{1}{2} \log \tau^2 + C \right]
\end{aligned}$$

In the M-step, we compute values for

$$\theta^{(t+1)} = (\beta_1^{(t+1)}, \dots, \beta_q^{(t+1)}, \sigma_1^{(t+1)}, \dots, \sigma_q^{(t+1)}, \mu^{(t+1)}, \tau^{(t+1)})$$

such that

$$\theta^{(t+1)} = \arg \max_{\theta} Q(\theta | \theta^{(t)})$$

Differentiating, we obtain the following equations:

$$\begin{aligned}
\beta_i &= \frac{\sum_j \text{E}[T_i^j E^j]}{\sum_j \text{E}[(E^j)^2]} \\
\sigma_i^2 &= \frac{1}{N} \sum_j \left(\text{E}\left[\left(T_i^j\right)^2\right] - 2\beta_i \text{E}[T_i^j E^j] + \beta_i^2 \text{E}[(E^j)^2] \right) \\
\mu &= \frac{\sum_j \text{E}[E^j]}{N} \\
\tau^2 &= \frac{1}{N} \sum_j \left(\text{E}\left[(E^j)^2\right] - 2\mu \text{E}[E^j] + N\mu^2 \right)
\end{aligned}$$

where all expectations are conditional upon the current parameter estimates $\theta^{(t)}$ and the observed data $\hat{\mathbf{O}}$. Thus, in order to update the parameters, we need to compute the sufficient statistics

$$\text{E}[E^j], \text{E}[(E^j)^2], \text{E}[T_i^j E^j], \text{E}\left[\left(T_i^j\right)^2\right]$$

which we can compute either analytically or via probabilistic inference in the Bayesian network using the current parameters $\theta^{(t)}$.

S.2 Analytic derivation of needed expectations

In the EM algorithm, the M-step updates require the sufficient statistics

$$\mathbb{E}[E^j], \quad \mathbb{E}[(E^j)^2], \quad \mathbb{E}[T_i^j E^j], \quad \mathbb{E}\left[\left(T_i^j\right)^2\right],$$

where all expectations are conditional on the observed data $\hat{\mathbf{O}}$ and on the current parameter estimates $\theta^{(t)}$. In this section we derive analytic expressions for these quantities for the single-endophenotype model without edges between traits.

For a fixed training variant j , recall the conditional distributions

$$E^j \sim \mathcal{N}(\mu, \tau^2), \quad T_i^j | E^j \sim \mathcal{N}(\beta_i E^j, \sigma_i^2), \quad O_i^j | T_i^j \sim \mathcal{N}(T_i^j, (s_i^j)^2).$$

We first collapse over T_i^j to obtain the marginal distribution of O_i^j given E^j :

$$O_i^j | E^j \sim \mathcal{N}\left(\beta_i E^j, \sigma_i^2 + (s_i^j)^2\right).$$

Define

$$v_i^j = \sigma_i^2 + (s_i^j)^2.$$

Then conditional on E^j , the observations O_1^j, \dots, O_q^j are independent with

$$\Pr\left(\hat{\mathbf{O}}^j = \mathbf{o}^j | E^j\right) = \prod_i \frac{1}{\sqrt{2\pi v_i^j}} \exp\left(-\frac{(o_i^j - \beta_i E^j)^2}{2v_i^j}\right).$$

Combining the Gaussian prior with this likelihood yields the conjugate posterior

$$E^j | \mathbf{o}^j \sim \mathcal{N}\left(m_E^j, V_E^j\right),$$

where

$$\begin{aligned} V_E^j &= \left(\frac{1}{\tau^2} + \sum_i \frac{\beta_i^2}{v_i^j}\right)^{-1} \\ m_E^j &= V_E^j \left(\frac{\mu}{\tau^2} + \sum_i \frac{\beta_i o_i^j}{v_i^j}\right). \end{aligned}$$

Therefore,

$$\mathbb{E}[E^j] = m_E^j, \quad \mathbb{E}[(E^j)^2] = V_E^j + (m_E^j)^2.$$

Next, for each trait i , the conditional posterior of T_i^j given E^j and o_i^j is also Gaussian:

$$T_i^j | E^j, o_i^j \sim \mathcal{N}\left(m_{T,i}^j(E^j), V_{T,i}^j\right),$$

with

$$\begin{aligned} V_{T,i}^j &= \left(\frac{1}{\sigma_i^2} + \frac{1}{(s_i^j)^2}\right)^{-1} = \frac{\sigma_i^2 (s_i^j)^2}{\sigma_i^2 + (s_i^j)^2} \\ m_{T,i}^j(E^j) &= V_{T,i}^j \left(\frac{\beta_i E^j}{\sigma_i^2} + \frac{o_i^j}{(s_i^j)^2}\right). \end{aligned}$$

Define the weights

$$a_i^j = \frac{\sigma_i^2}{\sigma_i^2 + (s_i^j)^2}, \quad b_i^j = \frac{(s_i^j)^2}{\sigma_i^2 + (s_i^j)^2} \beta_i,$$

so that $m_{T,i}^j(E^j) = a_i^j o_i^j + b_i^j E^j$.

Using the law of total expectation and the posterior moments of E^j derived above, we obtain

$$\begin{aligned} \mathbb{E}[T_i^j E^j] &= \mathbb{E}\left[\mathbb{E}\left[T_i^j | E^j, o_i^j\right] E^j\right] \\ &= \mathbb{E}\left[\left(a_i^j o_i^j + b_i^j E^j\right) E^j\right] \\ &= a_i^j o_i^j \mathbb{E}[E^j] + b_i^j \mathbb{E}\left[(E^j)^2\right] \\ &= a_i^j o_i^j m_E^j + b_i^j \left(V_E^j + (m_E^j)^2\right). \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbb{E}\left[\left(T_i^j\right)^2\right] &= \mathbb{E}\left[\text{Var}\left(T_i^j | E^j, o_i^j\right)\right] + \mathbb{E}\left[\left(\mathbb{E}\left[T_i^j | E^j, o_i^j\right]\right)^2\right] \\ &= V_{T,i}^j + \mathbb{E}\left[\left(a_i^j o_i^j + b_i^j E^j\right)^2\right] \\ &= V_{T,i}^j + \left(a_i^j o_i^j\right)^2 + 2a_i^j o_i^j b_i^j \mathbb{E}[E^j] + \left(b_i^j\right)^2 \mathbb{E}\left[(E^j)^2\right] \\ &= V_{T,i}^j + \left(a_i^j o_i^j\right)^2 + 2a_i^j o_i^j b_i^j m_E^j + \left(b_i^j\right)^2 \left(V_E^j + (m_E^j)^2\right). \end{aligned}$$

These expressions provide the analytic sufficient statistics needed in the M-step equations derived in the previous section.

S.3 Extension to multiple endophenotypes

We can generalize the model to allow each variant to act on multiple endophenotypes. Let there be K endophenotypes E_1, \dots, E_K . For variant j , let E_k^j denote its effect on endophenotype k , and let $\mathbf{E}^j = (E_1^j, \dots, E_K^j)^T$ denote the vector of endophenotype effects. The corresponding graphical model is shown in Figure ??.

We assume the K endophenotype effects are independent *a priori* (and note that a full covariance prior could be substituted without changing the inference strategy). We specify conditional probability distributions as:

$$E_k \sim \mathcal{N}(\mu_k, \tau_k^2) \quad (k = 1, \dots, K)$$

$$T_i \sim \mathcal{N}\left(\sum_{k=1}^K \beta_{ik} E_k, \sigma_i^2\right) \quad (i = 1, \dots, q)$$

$$O_i \sim \mathcal{N}(T_i, s_i^2) \quad (i = 1, \dots, q)$$

Here, β_{ik} is the contribution of endophenotype k to trait i ; equivalently, it is the strength of the edge $E_k \rightarrow T_i$. In an implementation, the structure (which edges are permitted) can be specified in a configuration file; if the edge $E_k \rightarrow T_i$ is absent, we fix $\beta_{ik} = 0$. As before, the σ_i control how

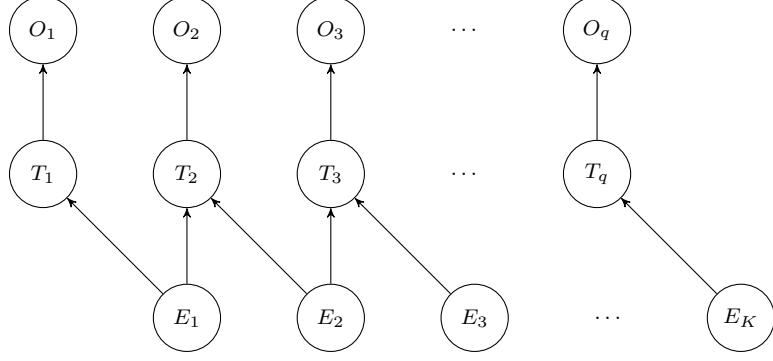


Figure 2: Model for phenotypic effects of a variant with multiple endophenotypes

tightly trait T_i follows the endophenotype-driven profile, and s_i controls the sampling distribution of the observed association statistics. The parameters μ_k and τ_k describe the prior population distribution of endophenotype k .

For convenience, define the $q \times K$ matrix $\mathbf{B} = (\beta_{ik})$, the prior mean vector $\mu = (\mu_1, \dots, \mu_K)^T$, and the prior covariance $\Sigma_E = \text{diag}(\tau_1^2, \dots, \tau_K^2)$. For each variant j , let $\mathbf{o}^j = (o_1^j, \dots, o_q^j)^T$ denote the observed effects. Collapsing over T_i as before yields

$$O_i^j \mid \mathbf{E}^j \sim \mathcal{N}\left(\mathbf{b}_i^T \mathbf{E}^j, \sigma_i^2 + (s_i^j)^2\right),$$

where \mathbf{b}_i^T is the i th row of \mathbf{B} , and s_i^j is the (variant-specific) standard error for O_i^j .

Let $v_i^j = \sigma_i^2 + (s_i^j)^2$ and define the diagonal weight matrix

$$\mathbf{W}^j = \text{diag}\left(\frac{1}{v_1^j}, \dots, \frac{1}{v_q^j}\right).$$

Then the posterior distribution of \mathbf{E}^j given the observations \mathbf{o}^j is multivariate normal:

$$\begin{aligned} \mathbf{E}^j \mid \mathbf{o}^j &\sim \mathcal{N}\left(\mathbf{m}_E^j, \mathbf{V}_E^j\right) \\ \mathbf{V}_E^j &= (\Sigma_E^{-1} + \mathbf{B}^T \mathbf{W}^j \mathbf{B})^{-1} \\ \mathbf{m}_E^j &= \mathbf{V}_E^j (\Sigma_E^{-1} \mu + \mathbf{B}^T \mathbf{W}^j \mathbf{o}^j). \end{aligned}$$

For fitting the model using EM, the M-step requires expectations with respect to the posterior. The basic sufficient statistics for variant j are

$$E[\mathbf{E}^j], \quad E[\mathbf{E}^j (\mathbf{E}^j)^T], \quad E[T_i^j \mathbf{E}^j], \quad E[(T_i^j)^2].$$

From the posterior above,

$$E[\mathbf{E}^j] = \mathbf{m}_E^j, \quad E[\mathbf{E}^j (\mathbf{E}^j)^T] = \mathbf{V}_E^j + \mathbf{m}_E^j (\mathbf{m}_E^j)^T.$$

Additionally, conditional on \mathbf{E}^j , the posterior of T_i^j given o_i^j remains univariate normal:

$$\begin{aligned} T_i^j \mid \mathbf{E}^j, o_i^j &\sim \mathcal{N}\left(m_{T,i}^j(\mathbf{E}^j), V_{T,i}^j\right) \\ V_{T,i}^j &= \left(\frac{1}{\sigma_i^2} + \frac{1}{(s_i^j)^2}\right)^{-1} = \frac{\sigma_i^2(s_i^j)^2}{\sigma_i^2 + (s_i^j)^2} \\ m_{T,i}^j(\mathbf{E}^j) &= V_{T,i}^j \left(\frac{\mathbf{b}_i^T \mathbf{E}^j}{\sigma_i^2} + \frac{o_i^j}{(s_i^j)^2}\right). \end{aligned}$$

Define

$$a_i^j = \frac{\sigma_i^2}{\sigma_i^2 + (s_i^j)^2}, \quad b_i^j = \frac{(s_i^j)^2}{\sigma_i^2 + (s_i^j)^2}.$$

Then $m_{T,i}^j(\mathbf{E}^j) = a_i^j o_i^j + b_i^j \mathbf{b}_i^T \mathbf{E}^j$. Using the law of total expectation, we obtain

$$\begin{aligned} \mathbb{E}[T_i^j \mathbf{E}^j] &= a_i^j o_i^j \mathbf{m}_E^j + b_i^j \left(\mathbf{V}_E^j + \mathbf{m}_E^j \left(\mathbf{m}_E^j\right)^T\right) \mathbf{b}_i \\ \mathbb{E}\left[\left(T_i^j\right)^2\right] &= V_{T,i}^j + \left(a_i^j o_i^j\right)^2 + 2a_i^j o_i^j b_i^j \mathbf{b}_i^T \mathbf{m}_E^j + \left(b_i^j\right)^2 \mathbf{b}_i^T \left(\mathbf{V}_E^j + \mathbf{m}_E^j \left(\mathbf{m}_E^j\right)^T\right) \mathbf{b}_i. \end{aligned}$$

Differentiating the expected complete-data log-likelihood yields M-step updates analogous to the single-endophenotype case. In particular, the update for the K -vector of coefficients for trait i , \mathbf{b}_i , is

$$\mathbf{b}_i = \left(\sum_j \mathbb{E}\left[\mathbf{E}^j (\mathbf{E}^j)^T\right]\right)^{-1} \left(\sum_j \mathbb{E}\left[T_i^j \mathbf{E}^j\right]\right),$$

with the understanding that if the configuration fixes certain entries of \mathbf{b}_i to zero, the corresponding rows/columns should be removed from the linear system and the fixed entries restored after solving. The update for σ_i^2 is

$$\sigma_i^2 = \frac{1}{N} \sum_j \mathbb{E}\left[\left(T_i^j - \mathbf{b}_i^T \mathbf{E}^j\right)^2\right],$$

and the updates for μ_k and τ_k^2 are

$$\mu_k = \frac{1}{N} \sum_j \mathbb{E}\left[E_k^j\right], \quad \tau_k^2 = \frac{1}{N} \sum_j \mathbb{E}\left[\left(E_k^j - \mu_k\right)^2\right].$$

Finally, we note that inference can also be performed by Gibbs sampling. In a Gibbs sampler, we alternate sampling $T_i^j \mid \mathbf{E}^j, o_i^j$ using the univariate normal conditional above, and sampling $\mathbf{E}^j \mid \mathbf{T}^j$ using the multivariate normal conditional

$$\mathbf{E}^j \mid \mathbf{T}^j \sim \mathcal{N}\left(\mathbf{m}_{E|T}^j, \mathbf{V}_{E|T}\right),$$

where

$$\mathbf{V}_{E|T} = \left(\boldsymbol{\Sigma}_E^{-1} + \mathbf{B}^T \text{diag}\left(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_q^2}\right) \mathbf{B}\right)^{-1}, \quad \mathbf{m}_{E|T}^j = \mathbf{V}_{E|T} \left(\boldsymbol{\Sigma}_E^{-1} \mu + \mathbf{B}^T \text{diag}\left(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_q^2}\right) \mathbf{t}^j\right).$$

S.4 Extension to allow edges between traits

We can extend the model to allow dependencies among traits not captured by the endophenotypes by adding directed edges between the true trait effects T_1, \dots, T_q . Let $\text{Pa}(i)$ denote the set of parent traits of T_i in a user-specified directed acyclic graph (DAG). For each directed edge $T_p \rightarrow T_i$, introduce a coefficient α_{ip} which captures the contribution of the true effect on trait p to the true effect on trait i , beyond what is explained by the endophenotypes. If an edge $T_p \rightarrow T_i$ is absent from the graph, we fix $\alpha_{ip} = 0$. The resulting graphical model is shown in Figure ??.

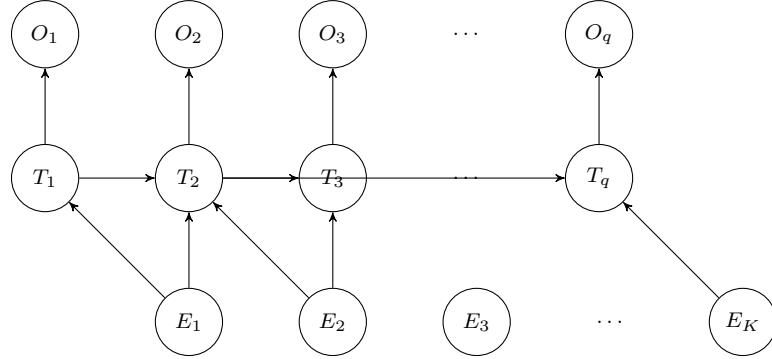


Figure 3: Model for phenotypic effects of a variant with multiple endophenotypes and directed edges between traits

We then specify

$$T_i \sim \mathcal{N} \left(\mathbf{b}_i^T \mathbf{E} + \sum_{p \in \text{Pa}(i)} \alpha_{ip} T_p, \sigma_i^2 \right), \quad O_i \sim \mathcal{N} (T_i, s_i^2),$$

together with the prior on \mathbf{E} from the previous section. Because the trait graph is required to be a DAG, there exists a topological ordering of traits under which each trait depends only on traits earlier in the ordering; if a cycle is present, the model does not define a valid Bayesian network and inference should be refused.

It is convenient to write this extension in matrix form. Let \mathbf{A} be the $q \times q$ matrix with entries $(\mathbf{A})_{ip} = \alpha_{ip}$, and define $\mathbf{L} = \mathbf{I} - \mathbf{A}$. Under the DAG assumption, \mathbf{L} is invertible. Let $\mathbf{D} = \text{diag}(\sigma_1^2, \dots, \sigma_q^2)$. Then the trait model can be written as the linear structural equation

$$\mathbf{LT}^j = \mathbf{BE}^j + \boldsymbol{\varepsilon}^j, \quad \boldsymbol{\varepsilon}^j \sim \mathcal{N} (\mathbf{0}, \mathbf{D}),$$

and hence

$$\begin{aligned} \mathbf{T}^j | \mathbf{E}^j &\sim \mathcal{N} (\mathbf{ME}^j, \Sigma_T) \\ \mathbf{M} &= \mathbf{L}^{-1} \mathbf{B} \\ \Sigma_T &= \mathbf{L}^{-1} \mathbf{D} (\mathbf{L}^{-1})^T. \end{aligned}$$

The observation model remains conditionally independent given \mathbf{T}^j . Let $\mathbf{S}^j = \text{diag}((s_1^j)^2, \dots, (s_q^j)^2)$. Then, collapsing over \mathbf{T}^j yields

$$\mathbf{O}^j | \mathbf{E}^j \sim \mathcal{N} (\mathbf{ME}^j, \mathbf{V}^j), \quad \mathbf{V}^j = \Sigma_T + \mathbf{S}^j.$$

As above, the posterior distribution of \mathbf{E}^j given \mathbf{o}^j is multivariate normal:

$$\begin{aligned}\mathbf{E}^j | \mathbf{o}^j &\sim \mathcal{N} \left(\mathbf{m}_E^j, \mathbf{V}_E^j \right) \\ \mathbf{V}_E^j &= \left(\boldsymbol{\Sigma}_E^{-1} + \mathbf{M}^T (\mathbf{V}^j)^{-1} \mathbf{M} \right)^{-1} \\ \mathbf{m}_E^j &= \mathbf{V}_E^j \left(\boldsymbol{\Sigma}_E^{-1} \mu + \mathbf{M}^T (\mathbf{V}^j)^{-1} \mathbf{o}^j \right).\end{aligned}$$

In addition, the posterior mean of the true trait effects can be computed in closed form. Since $\mathbf{T}^j | \mathbf{E}^j$ and $\mathbf{O}^j | \mathbf{T}^j$ are both multivariate normal, we have

$$E[\mathbf{T}^j | \mathbf{o}^j] = \mathbf{M} \mathbf{m}_E^j + \boldsymbol{\Sigma}_T (\mathbf{V}^j)^{-1} \left(\mathbf{o}^j - \mathbf{M} \mathbf{m}_E^j \right).$$

For fitting the model using EM, the sufficient statistics can be obtained from the joint multivariate normal posterior of $(\mathbf{E}^j, \mathbf{T}^j) | \mathbf{o}^j$, and the M-step updates for $(\mathbf{b}_i, \{\alpha_{ip}\}_{p \in \text{Pa}(i)})$ correspond to a linear regression of T_i on the predictors $(\mathbf{E}, \mathbf{T}_{\text{Pa}(i)})$ using posterior expected cross-products.

As in the previous section, inference can also be performed by Gibbs sampling. One convenient Gibbs scheme alternates the block updates

$$\mathbf{T}^j | \mathbf{E}^j, \mathbf{o}^j \sim \mathcal{N} \left(\mathbf{m}_{T|E}^j, \boldsymbol{\Sigma}_{T|E}^j \right), \quad \mathbf{E}^j | \mathbf{T}^j \sim \mathcal{N} \left(\mathbf{m}_{E|T}^j, \mathbf{V}_{E|T} \right),$$

where

$$\begin{aligned}\mathbf{m}_{T|E}^j &= \mathbf{M} \mathbf{E}^j + \boldsymbol{\Sigma}_T (\mathbf{V}^j)^{-1} (\mathbf{o}^j - \mathbf{M} \mathbf{E}^j) \\ \boldsymbol{\Sigma}_{T|E}^j &= \boldsymbol{\Sigma}_T - \boldsymbol{\Sigma}_T (\mathbf{V}^j)^{-1} \boldsymbol{\Sigma}_T.\end{aligned}$$

For the $\mathbf{E}^j | \mathbf{T}^j$ update, define $\mathbf{y}^j = \mathbf{L} \mathbf{T}^j$. Then $\mathbf{y}^j | \mathbf{E}^j \sim \mathcal{N}(\mathbf{B} \mathbf{E}^j, \mathbf{D})$, and hence

$$\begin{aligned}\mathbf{V}_{E|T} &= (\boldsymbol{\Sigma}_E^{-1} + \mathbf{B}^T \mathbf{D}^{-1} \mathbf{B})^{-1} \\ \mathbf{m}_{E|T}^j &= \mathbf{V}_{E|T} (\boldsymbol{\Sigma}_E^{-1} \mu + \mathbf{B}^T \mathbf{D}^{-1} \mathbf{y}^j).\end{aligned}$$

Alternatively, one can update the trait nodes T_i one-at-a-time using their Markov blanket in the trait DAG; the resulting full conditionals are univariate normal.