# Supplementary Notes for "Genome-wide identification of the genetic basis of amyotrophic lateral sclerosis"

## 1 Mathematical foundation of RefMap

Here, we provide a mathematical theory to justify Eq. 1 in the Method section of the main text. To facilitate the development of the theory, we first describe a universal discriminative framework that models the relationship between the genotype and phenotype, and then deduce a general distribution over summary statistics from this framework. Based on this result, a flexible probabilistic model that characterizes summary statistics with various prior structures can be developed, which generalizes multiple previous studies [1–4, 6–8]. In particular, Equation 1 of RefMap follows directly after assuming a linear relation between the genotype and phenotype. In the following, we will develop the framework in both cases of quantitative trait and case-control studies.

### 1.1 Quantitative trait studies

We start from considering a general genotype-phenotype model for continuous traits, i.e.,

$$y_n = F(\boldsymbol{x}_n, \boldsymbol{w}) + \epsilon_n, \ n = 1, \cdots, N, \tag{1}$$

in which $N$ is the sample size, $\boldsymbol{x}_n$ and $y_n$ are the genotypes and phenotype for the $n$th sample, respectively, $F$ is an unknown (usually non-linear) function with parameters $\boldsymbol{w}$ determining personal phenotype from his/her genotypes, and $\epsilon_n$ is the random noise following

$$\epsilon_n \sim \mathcal{N}(0, \sigma_\epsilon^2). \tag{2}$$

Note that as a routine procedure, genotypes are first standardized by

$$x_{ni} = \frac{g_{ni} - 2p_i}{\sqrt{2p_i(1-p_i)}}, \ i = 1, \cdots, M, \tag{3}$$

where $M$ is the number of alleles, $g_{ni}$ is the genotype of the $i$th allele for the $n$th sample, and $p_i$ is the frequency of the $i$th allele in the study cohort. After standardization, the sample mean and sample variance of each allele are 0 and 1, respectively. Moreover, we adopt a general setting and treat both genotypes and function parameters as random variables, yielding

$$y_n \mid \boldsymbol{x}_n, \boldsymbol{w}, \sigma_\epsilon \sim \mathcal{N}(F(\boldsymbol{x}_n, \boldsymbol{w}), \sigma_\epsilon^2). \tag{4}$$

Following the conventional annotation in the genome-wide association study (GWAS), the estimated effect sizes $\hat{\beta}_i$ for individual alleles are the most widely-used summary statistics, which are closely related to $\chi^2$ and $Z$-score. Given the genotype standardization, we have

$$\hat{\beta}_i = \frac{\boldsymbol{x}_i^\intercal \boldsymbol{y}}{N}, \tag{5}$$

where $\boldsymbol{x}_i$ is the genotype vector for the $i$th allele and $\boldsymbol{y} = y_{1:N}$. With matrix representation, we have

$$\hat{\boldsymbol{\beta}} = \frac{1}{N}\boldsymbol{X}^\intercal \boldsymbol{y} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n y_n, \tag{6}$$

where $\boldsymbol{X} = (x_{ni}) \in \mathbb{R}^{N\times M}$. Indeed, we have the following theorem characterizing the asymptotic distribution of $\sqrt{N}\hat{\boldsymbol{\beta}}$.

**Theorem 1.** *Given the definitions in Eqs. 1, 2 and 5, when the sample size $N$ is large enough, we have*

$$\sqrt{N}\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}, \boldsymbol{w}, \sigma_\epsilon \sim \mathcal{N}\left(\sqrt{N}\boldsymbol{\mu}(\boldsymbol{X}, F, \boldsymbol{w}), \sigma_\epsilon^2 \boldsymbol{\Sigma}_{\mathsf{LD}}\right), \tag{7}$$

*where $\boldsymbol{\Sigma}_{\mathsf{LD}}$ is the in-sample linkage disequilibrium (LD) matrix quantifying SNP correlations, and $\boldsymbol{\mu}(\boldsymbol{X}, F, \boldsymbol{w})$ is a quantity depending on the genotypes and the discriminative function $F$.*

*Proof.* We first show that $\sqrt{N}\hat{\boldsymbol{\beta}}$ follows a normal distribution asymptotically. In fact, according to Eq. 6, given the genotypes and the discriminative function, $\sqrt{N}\hat{\boldsymbol{\beta}}$ can be computed by the sum of $\boldsymbol{x}_n y_n$, which are independent with each other but with different expectations. On the other hand, the variance of $\boldsymbol{x}_n y_n$ is given by

$$\begin{aligned}
\mathrm{Var}\left[\boldsymbol{x}_n y_n \mid \boldsymbol{x}_n, \boldsymbol{w}, \sigma_\epsilon\right] &= \mathrm{Var}\left[\boldsymbol{x}_n (F(\boldsymbol{x}_n, \boldsymbol{w}) + \epsilon_n) \mid \boldsymbol{x}_n, \boldsymbol{w}, \sigma_\epsilon\right] \\
&= \mathrm{Var}\left[\boldsymbol{x}_n \epsilon_n \mid \boldsymbol{x}_n, \boldsymbol{w}, \sigma_\epsilon\right] \\
&= \mathbb{E}\left[\epsilon_n^2 \boldsymbol{x}_n \boldsymbol{x}_n^\intercal \mid \boldsymbol{x}_n, \boldsymbol{w}, \sigma_\epsilon\right] \\
&= \boldsymbol{x}_n \boldsymbol{x}_n^\intercal \sigma_\epsilon^2,
\end{aligned} \tag{8}$$

yielding

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_n^\intercal \sigma_\epsilon^2 &= \lim_{N\to\infty} \frac{1}{N}\boldsymbol{X}^\intercal \boldsymbol{X} \cdot \sigma_\epsilon^2 \\
&= \sigma_\epsilon^2 \hat{\boldsymbol{\Sigma}}_{\mathsf{LD}} \\
&\approx \sigma_\epsilon^2 \boldsymbol{\Sigma}_{\mathsf{LD}},
\end{aligned} \tag{9}$$

in which the estimated LD matrix $\hat{\boldsymbol{\Sigma}}_{\mathsf{LD}} = (\hat{r}_{ij})$ is given by

$$\hat{r}_{ij} = \frac{\boldsymbol{x}_i^\intercal \boldsymbol{x}_j}{\sqrt{\boldsymbol{x}_i^\intercal \boldsymbol{x}_i}\sqrt{\boldsymbol{x}_j^\intercal \boldsymbol{x}_j}} = \frac{1}{N}\boldsymbol{x}_i^\intercal \boldsymbol{x}_j, \tag{10}$$

and the last approximation is guaranteed by $\mathbb{E}[\hat{r}_{ij}] = r_{ij} = \mathbb{E}[x_i x_j]$. Therefore, according to the multivariate Lindeberg-Feller central limit theorem (CLT), we conclude that $\sqrt{N}\hat{\boldsymbol{\beta}} =$

$1/\sqrt{N}\sum_{i=1}^{N}\boldsymbol{x}_n y_n$ asymptotically follows a normal distribution with covariance $\sigma_\epsilon^2\boldsymbol{\Sigma}_{\mathsf{LD}}$, whose expectation is given by

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\mathbb{E}\left[\boldsymbol{x}_n y_n \mid \boldsymbol{x}_n, \boldsymbol{w}, \sigma_\epsilon\right] &= \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\mathbb{E}\left[\boldsymbol{x}_n(F(\boldsymbol{x}_n, \boldsymbol{w}) + \epsilon_n) \mid \boldsymbol{x}_n, \boldsymbol{w}, \sigma_\epsilon\right] \\ &= \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\boldsymbol{x}_n F(\boldsymbol{x}_n, \boldsymbol{w}) \\ &= \sqrt{N}\boldsymbol{\mu}(\boldsymbol{X}, F, \boldsymbol{w}),\end{aligned} \tag{11}$$

where $\boldsymbol{\mu}(\cdot)$ is defined as

$$\boldsymbol{\mu}(\boldsymbol{X}, F, \boldsymbol{w}) = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n F(\boldsymbol{x}_n, \boldsymbol{w}). \tag{12}$$

This completes the proof. □

Note that if we use $Z$-scores computed by GWAS as the approximation of $\sqrt{N}\hat{\boldsymbol{\beta}}/\sigma_\epsilon$, i.e., dividing $\hat{\beta}_i$ by its estimated standard error, we have

$$\hat{\boldsymbol{z}} \mid \boldsymbol{X}, \boldsymbol{w} \sim \mathcal{N}\left(\sqrt{N}\boldsymbol{\mu}(\boldsymbol{X}, F, \boldsymbol{w}), \boldsymbol{\Sigma}_{\mathsf{LD}}\right), \tag{13}$$

in which $\sigma_\epsilon$ is absorbed into $\boldsymbol{\mu}(\cdot)$ for annotation brevity.

### 1.2 Case-control studies

We state the analysis for case-control studies using a Bernoulli distribution over case-control status, i.e.,

$$y_n \mid \pi_n \sim \mathsf{Bernoulli}(\pi_n),\ n = 1, \cdots, N, \tag{14}$$

whose logit is defined similarly as Eq. 1 but without random noise, i.e.,

$$\log\frac{\pi_n}{1 - \pi_n} = F(\boldsymbol{x}_n, \boldsymbol{w}). \tag{15}$$

After a few calculations we can easily get

$$\pi_n = \sigma(F(\boldsymbol{x}_n, \boldsymbol{w})), \tag{16}$$

where $\sigma(\cdot)$ is the sigmoid function defined by $\sigma(x) = 1/(1 + \exp(-x))$.

To facilitate the following analysis, here we illustrate the standardization procedure in more detail, i.e.,

$$x_{ni} = \frac{g_{ni} - 2\hat{p}_i}{\sqrt{2\hat{p}_i(1 - \hat{p}_i)}}, \tag{17}$$

where $g_{ni}$ is the genotype coded by 0, 1 and 2, and $\hat{p}_i$ is the in-sample allele frequency. Therefore, suppose we have the same number ($N/2$) of cases and controls in the study cohort, the widely-used $Z$-scores for case-control studies defined as

$$\hat{z}_i = \frac{\sqrt{N}(\hat{p}_i^+ - \hat{p}_i^-)}{\sqrt{2\hat{p}_i(1 - \hat{p}_i)}}\ (i = 1, \cdots, M) \tag{18}$$

can be written as

$$\hat{\boldsymbol{z}} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (2\mathbb{1}\{y_n = 1)\} - 1)\boldsymbol{x}_n. \tag{19}$$

Again, utilizing the multivariate Lindeberg-Feller CLT, we can derive the asymptotic conditional distribution of $\hat{\boldsymbol{z}}$, which is approximately the same as that in the quantitative trait studies (Eq. 13). In particular, we have the following result.

**Theorem 2.** *Given the definitions in Eqs. 14, 15 and 19, when the sample size $N$ is large enough, we have*

$$\hat{\boldsymbol{z}} \mid \boldsymbol{X}, \boldsymbol{w} \sim \mathcal{N}\left(\sqrt{N}\boldsymbol{\mu}(\boldsymbol{X}, F, \boldsymbol{w}), \boldsymbol{\Sigma}_{\mathsf{LD}}\right), \tag{20}$$

*where $\boldsymbol{\Sigma}_{\mathsf{LD}}$ is the in-sample LD matrix, and $\boldsymbol{\mu}(\boldsymbol{X}, F, \boldsymbol{w})$ is a quantity depending on the genotypes and the discriminative function.*

*Proof.* Conditioned on $\boldsymbol{X}$ and $\boldsymbol{w}$, the variance of $(2\mathbb{1}\{y_n = 1)\} - 1)\boldsymbol{x}_n$ can be calculated as

$$\begin{aligned}
\mathrm{Var}\left[(2\mathbb{1}\{y_n = 1\} - 1)\boldsymbol{x}_n\right] &= \boldsymbol{x}_n\boldsymbol{x}_n^{\intercal} - \mathbb{E}\left[2\mathbb{1}\{y_n = 1\} - 1\right]^2 \boldsymbol{x}_n\boldsymbol{x}_n^{\intercal} \\
&= 4\boldsymbol{x}_n\boldsymbol{x}_n^{\intercal}\mathbb{P}\left[y_n = 1\right]\left(1 - \mathbb{P}\left[y_n = 1\right]\right) \\
&= 4\boldsymbol{x}_n\boldsymbol{x}_n^{\intercal}\mathrm{Var}\left[y_n\right],
\end{aligned} \tag{21}$$

where the conditions are omitted for brevity. In fact, as $\mathrm{Var}\left[y_n \mid \boldsymbol{x}_n, \boldsymbol{w}\right] < 1$, we conclude that the average of variance $1/N \sum_{n=1}^{N} \mathrm{Var}\left[(2\mathbb{1}\{y_n = 1\} - 1)\boldsymbol{x}_n \mid \boldsymbol{x}_n, \boldsymbol{w}\right]$ converges as $N \to \infty$, whose limit is denoted as $\boldsymbol{\Sigma}_\infty$. According to the multivariate Lindeberg-Feller CLT, the asymptotic conditional distribution of $\hat{\boldsymbol{z}}$ is a normal distribution with covariance matrix $\boldsymbol{\Sigma}_\infty$.

To get a clearer structure of $\boldsymbol{\Sigma}_\infty$, we now apply a few approximations for Eq. 21. In particular, we have

$$\begin{aligned}
\boldsymbol{\Sigma}_\infty &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathrm{Var}\left[(2\mathbb{1}\{y_n = 1\} - 1)\boldsymbol{x}_n \mid \boldsymbol{x}_n, \boldsymbol{w}\right] \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 4\boldsymbol{x}_n\boldsymbol{x}_n^{\intercal}\mathrm{Var}\left[y_n \mid \boldsymbol{x}_n, \boldsymbol{w}\right] \\
&= \mathbb{E}\left[4\boldsymbol{x}_n\boldsymbol{x}_n^{\intercal}\mathrm{Var}\left[y_n \mid \boldsymbol{x}_n, \boldsymbol{w}\right]\right] \\
&\approx 4\boldsymbol{\Sigma}_{\mathsf{LD}}\mathbb{E}\left[\mathrm{Var}\left[y_n \mid \boldsymbol{x}_n, \boldsymbol{w}\right]\right] \\
&= 4\boldsymbol{\Sigma}_{\mathsf{LD}}\left(\mathrm{Var}\left[y_n \mid \boldsymbol{w}\right] - \mathrm{Var}\left[\mathbb{E}\left[y_n \mid \boldsymbol{x}_n, \boldsymbol{w}\right]\right]\right) \\
&< 4\boldsymbol{\Sigma}_{\mathsf{LD}}\mathrm{Var}\left[y_n \mid \boldsymbol{w}\right] \\
&\approx 4\boldsymbol{\Sigma}_{\mathsf{LD}}\mathbb{E}\left[\mathrm{Var}\left[y_n \mid \boldsymbol{w}\right]\right] \\
&= 4\boldsymbol{\Sigma}_{\mathsf{LD}}\left(\mathrm{Var}\left[y_n\right] - \mathrm{Var}\left[\mathbb{E}\left[y_n \mid \boldsymbol{w}\right]\right]\right) \\
&< 4\boldsymbol{\Sigma}_{\mathsf{LD}}\mathrm{Var}\left[y_n\right] \\
&= \boldsymbol{\Sigma}_{\mathsf{LD}},
\end{aligned} \tag{22}$$

in which the third and the fourth "=" come from the law of total variance, the first and the second "<" are implied by the positivity of variance, and $\boldsymbol{\Sigma}_{\mathsf{LD}}$ is the in-sample LD matrix. For the last "=", we argue that the expectation and variance in Eq. 22 are taken over the sampling space in case-control studies, rather than the general population. Under the assumption of equal number of cases and controls, the sampling disease prevalence is 0.5, yielding $\mathrm{Var}[y_n] = 0.25$.

Furthermore, the expectation of the asymptotic conditional distribution can be calculated as

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\mathbb{E}\left[(2\mathbb{1}\{y_n=1\}-1)\boldsymbol{x}_n \mid \boldsymbol{x}_n,\boldsymbol{w}\right] &= \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\boldsymbol{x}_n\left(2\sigma(F(\boldsymbol{x}_n,\boldsymbol{w}))-1\right) \\ &= \sqrt{N}\boldsymbol{\mu}(\boldsymbol{X},F,\boldsymbol{w}),\end{aligned} \tag{23}$$

where we define

$$\boldsymbol{\mu}(\boldsymbol{X},F,\boldsymbol{w}) = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n\left(2\sigma(F(\boldsymbol{x}_n,\boldsymbol{w}))-1\right). \tag{24}$$

This completes the proof. □

### 1.3 A linear model for RefMap

We consider a linear model that underlies the design of RefMap. Specifically, in the quantitative trait studies, we define

$$F(\boldsymbol{x}_n,\boldsymbol{w}) = w_0 + \sum_{i=1}^{M} w_i x_{ni}. \tag{25}$$

Note that this linear model has been widely used in traditional GWAS studies [1–3], and $w_i$ is called the *effect size* of the $i$th allele. The linear model for case-control studies can be developed similarly by considering the approximation of sigmoid function using its Taylor expansion. Therefore, the expectation of the asymptotic distribution of $Z$-scores can be calculated as

$$\begin{aligned}\sqrt{N}\boldsymbol{\mu}(\boldsymbol{X},F,\boldsymbol{w}) &= \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\boldsymbol{x}_n\left(\boldsymbol{x}_n^{\intercal}\boldsymbol{w}+w_0\right) \\ &= \sqrt{N}\hat{\boldsymbol{\Sigma}}_{\mathsf{LD}}\boldsymbol{w},\end{aligned} \tag{26}$$

indicating that the expected $Z$-score for each allele is determined by its effect size as well as its strongly-associated neighbors. By absorbing $\sqrt{N}$ into $\boldsymbol{w}$, we eventually get Eq. 1 in the RefMap model.

## 2 Inference for RefMap

The RefMap model was defined in Eqs. 1 to 18 in the Method section of the main text. Here, we are interested in the posterior $p(\boldsymbol{T} \mid \boldsymbol{Z},\boldsymbol{S})$, whose exact calculation is intractable. Therefore, we seek for approximate inference based on the mean-field variational inference (MFVI). Basically, we first assume that the approximate posterior over latent variables factorizes, indicating conditional independence across latent variables, and then perform approximate inference by optimizing the *evidence lower bound* (ELBO) with respect to factorized proposal distributions, i.e.,

$$q\left(\lambda_{j,k}, \lambda, \boldsymbol{\tau}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{M}, \boldsymbol{T}, \boldsymbol{U}, \boldsymbol{\Lambda}\right) = \max_q \mathbb{E}_q \left[\log\left(\frac{p\left(\boldsymbol{Z}, \lambda_{j,k}, \lambda, \boldsymbol{\tau}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{M}, \boldsymbol{T}, \boldsymbol{U}, \boldsymbol{\Lambda} \mid \boldsymbol{S}\right)}{q\left(\lambda_{j,k}, \lambda, \boldsymbol{\tau}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{M}, \boldsymbol{T}, \boldsymbol{U}, \boldsymbol{\Lambda}\right)}\right)\right], \tag{27}$$

which can be shown to be equivalent to minimizing the Kullback-Leibler (KL) divergence between the true posterior and its proposal.

In the following, we will first introduce several specific techniques we used in MFVI, and then summarize the update rules for different variational parameters. At last, a coordinate ascent-based VI algorithm will be given.

## 2.1 Rectification nonlinearity

We impose non-negativity on $v_{-1}$ and $v_{+1}$ using the technique of rectification nonlinearity proposed in [5]. This technique relaxes the sparsity constraint over factors and meanwhile enjoys tractable variational inference.

We first note that the approximate posterior $q(r_{-1})$ from MFVI follows the free-form solution

$$q(r_{-1}) = \frac{1}{\tilde{Z}_{-1}} \prod_{k=1}^{K}\prod_{j=1}^{J_k} \mathcal{N}\left(\mathbb{E}[m_{j,k}] \mid -v_{-1}, \mathbb{E}[\tau_{-1}]^{-1}\right)^{\mathbb{E}\left[t_{j,k}^{(-1)}\right]} \times \mathcal{N}\left(r_{-1} \mid \mathbb{E}[m_{-1}], \mathbb{E}[\lambda_{-1}]^{-1}\right), \tag{28}$$

where $\tilde{Z}_{-1}$ is the normalization term to be computed later. Moreover, it can be easily shown that Eq. 28 can be written as $q(r_{-1}) = q_p(r_{-1}) + q_n(r_{-1})$ with the form

$$q_p(r_{-1}) = \frac{\tilde{w}_p^{(-1)}}{\tilde{Z}_{-1}} \mathcal{N}\left(r_{-1} \mid \tilde{\mu}_p^{(-1)}, \left(\tilde{\lambda}_p^{(-1)}\right)^{-1}\right) u(r_{-1}), \tag{29}$$

$$q_n(r_{-1}) = \frac{\tilde{w}_n^{(-1)}}{\tilde{Z}_{-1}} \mathcal{N}\left(r_{-1} \mid \tilde{\mu}_n^{(-1)}, \left(\tilde{\lambda}_n^{(-1)}\right)^{-1}\right) u(-r_{-1}), \tag{30}$$

in which

$$\tilde{\mu}_p^{(-1)} = \left(-\mathbb{E}[\tau_{-1}] \sum_{k=1}^{K}\sum_{j=1}^{J_k} \mathbb{E}\left[t_{j,k}^{(-1)}\right] \mathbb{E}[m_{j,k}] + \mathbb{E}[\lambda_{-1}]\mathbb{E}[m_{-1}]\right) \left(\tilde{\lambda}_p^{(-1)}\right)^{-1}, \tag{31}$$

$$\tilde{\mu}_n^{(-1)} = \mathbb{E}[m_{-1}], \tag{32}$$

$$\tilde{\lambda}_p^{(-1)} = \mathbb{E}[\tau_{-1}] \sum_{k=1}^{K}\sum_{j=1}^{J_k} \mathbb{E}\left[t_{j,k}^{(-1)}\right] + \mathbb{E}[\lambda_{-1}], \tag{33}$$

$$\tilde{\lambda}_n^{(-1)} = \mathbb{E}[\lambda_{-1}], \tag{34}$$

and $u(\cdot)$ is the standard step function. With Eqs. 31 to 34, $\tilde{w}_p^{(-1)}$ and $\tilde{w}_n^{(-1)}$ can be computed by integrating Eqs. 28, 29 and 30 with respect to $r_{-1}$. Then the normalization term is given by

$$\tilde{Z}_{-1} = \frac{\tilde{w}_n^{(-1)}}{2}\mathrm{erfc}\left(\tilde{\mu}_n^{(-1)}\sqrt{\tilde{\lambda}_n^{(-1)}/2}\right) + \frac{\tilde{w}_p^{(-1)}}{2}\mathrm{erfc}\left(-\tilde{\mu}_p^{(-1)}\sqrt{\tilde{\lambda}_p^{(-1)}/2}\right). \tag{35}$$

The moments for posteriors are obtained by

$$\mathbb{E}[r_{-1}] = \tilde{M}_p^{(-1)} + \tilde{M}_n^{(-1)}, \tag{36}$$

$$\mathbb{E}[r_{-1}^2] = \tilde{M}_p^{(-2)} + \tilde{M}_n^{(-2)}, \tag{37}$$

$$\mathbb{E}[v_{-1}] = \tilde{M}_p^{(-1)}, \tag{38}$$

$$\mathbb{E}[v_{-1}^2] = \tilde{M}_p^{(-2)}, \tag{39}$$

where

$$\tilde{M}_p^{(-0)} = \frac{\tilde{w}_p^{(-1)}}{2\tilde{Z}_{-1}}\mathrm{erfc}\left(-\tilde{\mu}_p^{(-1)}\sqrt{\tilde{\lambda}_p^{(-1)}/2}\right), \tag{40}$$

$$\tilde{M}_p^{(-1)} = \frac{\tilde{w}_p^{(-1)}}{2\tilde{Z}_{-1}}\left\{\mathrm{erfc}\left(-\tilde{\mu}_p^{(-1)}\sqrt{\tilde{\lambda}_p^{(-1)}/2}\right)\tilde{\mu}_p^{(-1)} + \sqrt{\frac{2}{\pi\tilde{\lambda}_p^{(-1)}}}\frac{1}{\exp\left(\tilde{\lambda}_p^{(-1)}\left(\tilde{\mu}_p^{(-1)}\right)^2/2\right)}\right\}, \tag{41}$$

$$\tilde{M}_p^{(-2)} = \frac{\tilde{w}_p^{(-1)}}{2\tilde{Z}_{-1}}\left\{\mathrm{erfc}\left(-\tilde{\mu}_p^{(-1)}\sqrt{\tilde{\lambda}_p^{(-1)}/2}\right)\left(\left(\tilde{\mu}_p^{(-1)}\right)^2 + \frac{1}{\tilde{\lambda}_p^{(-1)}}\right) + \sqrt{\frac{2}{\pi\tilde{\lambda}_p^{(-1)}}}\frac{\tilde{\mu}_p^{(-1)}}{\exp\left(\tilde{\lambda}_p^{(-1)}\left(\tilde{\mu}_p^{(-1)}\right)^2/2\right)}\right\}, \tag{42}$$

$$\tilde{M}_n^{(-0)} = \frac{\tilde{w}_n^{(-1)}}{2\tilde{Z}_{-1}}\mathrm{erfc}\left(\tilde{\mu}_n^{(-1)}\sqrt{\tilde{\lambda}_n^{(-1)}/2}\right), \tag{43}$$

$$\tilde{M}_n^{(-1)} = \frac{\tilde{w}_n^{(-1)}}{2\tilde{Z}_{-1}}\left\{\mathrm{erfc}\left(\tilde{\mu}_n^{(-1)}\sqrt{\tilde{\lambda}_n^{(-1)}/2}\right)\tilde{\mu}_n^{(-1)} - \sqrt{\frac{2}{\pi\tilde{\lambda}_n^{(-1)}}}\frac{1}{\exp\left(\tilde{\lambda}_n^{(-1)}\left(\tilde{\mu}_n^{(-1)}\right)^2/2\right)}\right\}, \tag{44}$$

$$\tilde{M}_n^{(-2)} = \frac{\tilde{w}_n^{(-1)}}{2\tilde{Z}_{-1}}\left\{\mathrm{erfc}\left(\tilde{\mu}_n^{(-1)}\sqrt{\tilde{\lambda}_n^{(-1)}/2}\right)\left(\left(\tilde{\mu}_n^{(-1)}\right)^2 + \frac{1}{\tilde{\lambda}_n^{(-1)}}\right) - \sqrt{\frac{2}{\pi\tilde{\lambda}_n^{(-1)}}}\frac{\tilde{\mu}_n^{(-1)}}{\exp\left(\tilde{\lambda}_n^{(-1)}\left(\tilde{\mu}_n^{(-1)}\right)^2/2\right)}\right\}. \tag{45}$$

Similar to $q(r_{-1})$, the posterior $q(r_{+1})$ also follows a free-form solution given by

$$q(r_{+1}) = \frac{1}{\tilde{Z}_{+1}}\prod_{k=1}^{K}\prod_{j=1}^{J_k}\mathcal{N}\left(\mathbb{E}[m_{j,k}] \mid v_{+1}, \mathbb{E}[\tau_{+1}]^{-1}\right)^{\mathbb{E}\left[t_{j,k}^{(+1)}\right]} \times \mathcal{N}\left(r_{+1} \mid \mathbb{E}[m_{+1}], \mathbb{E}[\lambda_{+1}]^{-1}\right), \tag{46}$$

where $\tilde{Z}_{+1}$ is the normalization term. Equation 46 can also be written as $q(r_{+1}) = q_p(r_{+1}) +$

$q_n(r_{+1})$ with the form

$$q_p(r_{+1}) = \frac{\tilde{w}_p^{(+1)}}{\tilde{Z}_{+1}} \mathcal{N}\left(r_{+1} \mid \tilde{\mu}_p^{(+1)}, \left(\tilde{\lambda}_p^{(+1)}\right)^{-1}\right) u(r_{+1}), \tag{47}$$

$$q_n(r_{+1}) = \frac{\tilde{w}_n^{(+1)}}{\tilde{Z}_{+1}} \mathcal{N}\left(r_{+1} \mid \tilde{\mu}_n^{(+1)}, \left(\tilde{\lambda}_n^{(+1)}\right)^{-1}\right) u(-r_{+1}), \tag{48}$$

in which

$$\tilde{\mu}_p^{(+1)} = \left(\mathbb{E}[\tau_{+1}] \sum_{k=1}^{K} \sum_{j=1}^{J_k} \mathbb{E}\left[t_{j,k}^{(+1)}\right] \mathbb{E}[m_{j,k}] + \mathbb{E}[\lambda_{+1}]\mathbb{E}[m_{+1}]\right) \left(\tilde{\lambda}_p^{(+1)}\right)^{-1}, \tag{49}$$

$$\tilde{\mu}_n^{(+1)} = \mathbb{E}[m_{+1}], \tag{50}$$

$$\tilde{\lambda}_p^{(+1)} = \mathbb{E}[\tau_{+1}] \sum_{k=1}^{K} \sum_{j=1}^{J_k} \mathbb{E}\left[t_{j,k}^{(+1)}\right] + \mathbb{E}[\lambda_{+1}], \tag{51}$$

$$\tilde{\lambda}_n^{(+1)} = \mathbb{E}[\lambda_{+1}]. \tag{52}$$

After computing $\tilde{w}_p^{(+1)}$ and $\tilde{w}_n^{(+1)}$, the normalization term is given by

$$\tilde{Z}_{+1} = \frac{\tilde{w}_n^{(+1)}}{2} \mathrm{erfc}\left(\tilde{\mu}_n^{(+1)} \sqrt{\tilde{\lambda}_n^{(+1)}/2}\right) + \frac{\tilde{w}_p^{(+1)}}{2} \mathrm{erfc}\left(-\tilde{\mu}_p^{(+1)} \sqrt{\tilde{\lambda}_p^{(+1)}/2}\right). \tag{53}$$

The moments for posteriors are obtained by

$$\mathbb{E}[r_{+1}] = \tilde{M}_p^{(+1)} + \tilde{M}_n^{(+1)}, \tag{54}$$

$$\mathbb{E}[r_{+1}^2] = \tilde{M}_p^{(+2)} + \tilde{M}_n^{(+2)}, \tag{55}$$

$$\mathbb{E}[v_{+1}] = \tilde{M}_p^{(+1)}, \tag{56}$$

$$\mathbb{E}[v_{+1}^2] = \tilde{M}_p^{(+2)}, \tag{57}$$

in which

$$\tilde{M}_p^{(+0)} = \frac{\tilde{w}_p^{(+1)}}{2\tilde{Z}_{+1}} \mathrm{erfc}\left(-\tilde{\mu}_p^{(+1)} \sqrt{\tilde{\lambda}_p^{(+1)}/2}\right), \tag{58}$$

$$\tilde{M}_p^{(+1)} = \frac{\tilde{w}_p^{(+1)}}{2\tilde{Z}_{+1}} \left\{ \mathrm{erfc}\left(-\tilde{\mu}_p^{(+1)} \sqrt{\tilde{\lambda}_p^{(+1)}/2}\right) \tilde{\mu}_p^{(+1)} + \sqrt{\frac{2}{\pi \tilde{\lambda}_p^{(+1)}}} \frac{1}{\exp\left(\tilde{\lambda}_p^{(+1)} \left(\tilde{\mu}_p^{(+1)}\right)^2/2\right)} \right\}, \tag{59}$$

$$\tilde{M}_p^{(+2)} = \frac{\tilde{w}_p^{(+1)}}{2\tilde{Z}_{+1}} \left\{ \mathrm{erfc}\left(-\tilde{\mu}_p^{(+1)} \sqrt{\tilde{\lambda}_p^{(+1)}/2}\right) \left(\left(\tilde{\mu}_p^{(+1)}\right)^2 + \frac{1}{\tilde{\lambda}_p^{(+1)}}\right) + \sqrt{\frac{2}{\pi \tilde{\lambda}_p^{(+1)}}} \frac{\tilde{\mu}_p^{(+1)}}{\exp\left(\tilde{\lambda}_p^{(+1)} \left(\tilde{\mu}_p^{(+1)}\right)^2/2\right)} \right\}, \tag{60}$$

$$\tilde{M}_n^{(+0)} = \frac{\tilde{w}_n^{(+1)}}{2\tilde{Z}_{+1}} \mathrm{erfc}\left(\tilde{\mu}_n^{(+1)} \sqrt{\tilde{\lambda}_n^{(+1)}/2}\right), \tag{61}$$

$$\tilde{M}_n^{(+1)} = \frac{\tilde{w}_n^{(+1)}}{2\tilde{Z}_{+1}} \left\{ \operatorname{erfc}\left( \tilde{\mu}_n^{(+1)} \sqrt{\tilde{\lambda}_n^{(+1)}/2} \right) \tilde{\mu}_n^{(+1)} - \sqrt{\frac{2}{\pi \tilde{\lambda}_n^{(+1)}}} \frac{1}{\exp\left( \tilde{\lambda}_n^{(+1)} \left( \tilde{\mu}_n^{(+1)} \right)^2 /2 \right)} \right\}, \tag{62}$$

$$\tilde{M}_n^{(+2)} = \frac{\tilde{w}_n^{(+1)}}{2\tilde{Z}_{+1}} \left\{ \operatorname{erfc}\left( \tilde{\mu}_n^{(+1)} \sqrt{\tilde{\lambda}_n^{(+1)}/2} \right) \left( \left( \tilde{\mu}_n^{(+1)} \right)^2 + \frac{1}{\tilde{\lambda}_n^{(+1)}} \right) - \sqrt{\frac{2}{\pi \tilde{\lambda}_n^{(+1)}}} \frac{\tilde{\mu}_n^{(+1)}}{\exp\left( \tilde{\lambda}_n^{(+1)} \left( \tilde{\mu}_n^{(+1)} \right)^2 /2 \right)} \right\}. \tag{63}$$

## 2.2 Local variational method

We adopt the local variational method to tackle the intractability of MFVI for $\boldsymbol{w}$ due to the introduction of the sigmoid function (Eq. 16 in the Method section). In particular, we have the following result regarding Eq. 15 in the Method section:

$$\begin{aligned}
(0.5\pi_{j,k})^{t_{j,k}^{(-1)}} (1-\pi_{j,k})^{t_{j,k}^{(0)}} (0.5\pi_{j,k})^{t_{j,k}^{(+1)}} &\propto \pi_{j,k}^{t_{j,k}^{(-1)}+t_{j,k}^{(+1)}} (1-\pi_{j,k})^{t_{j,k}^{(0)}} \\
&= \exp\left\{ \boldsymbol{w}^\intercal \boldsymbol{s}_{j,k} \left( t_{j,k}^{(-1)} + t_{j,k}^{(+1)} \right) \right\} \sigma\left( -\boldsymbol{w}^\intercal \boldsymbol{s}_{j,k} \right) \\
&\geq \exp\left\{ \boldsymbol{w}^\intercal \boldsymbol{s}_{j,k} \left( t_{j,k}^{(-1)} + t_{j,k}^{(+1)} \right) \right\} \sigma(\xi_{j,k}) \exp\left\{ -\frac{1}{2}\left( \boldsymbol{w}^\intercal \boldsymbol{s}_{j,k} + \xi_{j,k} \right) - \chi(\xi_{j,k}) \left( (\boldsymbol{w}^\intercal \boldsymbol{s}_{j,k})^2 - \xi_{j,k}^2 \right) \right\},
\end{aligned} \tag{64}$$

where

$$\chi(\xi) = \frac{1}{2\xi} \left( \sigma(\xi) - \frac{1}{2} \right). \tag{65}$$

Then we can perform standard MFVI with respect to the lower bound of Eq. 64, which yields

$$\begin{aligned}
\ln q(\boldsymbol{w}) &\propto \mathbb{E}_{-\boldsymbol{w}} \left[ \sum_{k=1}^{K} \sum_{j=1}^{J_k} \boldsymbol{w}^\intercal \boldsymbol{s}_{j,k} \left( t_{j,k}^{(-1)} + t_{j,k}^{(+1)} \right) - \frac{1}{2} \boldsymbol{w}^\intercal \boldsymbol{s}_{j,k} - \chi(\xi_{j,k}) (\boldsymbol{w}^\intercal \boldsymbol{s}_{j,k})^2 - \frac{1}{2} \boldsymbol{w}^\intercal \boldsymbol{\Lambda} \boldsymbol{w} \right] \\
&= -\frac{1}{2} \boldsymbol{w}^\intercal \left( \mathbb{E}[\boldsymbol{\Lambda}] + 2 \sum_{k=1}^{K} \sum_{j=1}^{J_k} \chi(\xi_{j,k}) \boldsymbol{s}_{j,k} \boldsymbol{s}_{j,k}^\intercal \right) \boldsymbol{w} + \boldsymbol{w}^\intercal \sum_{k=1}^{K} \sum_{j=1}^{J_k} \boldsymbol{s}_{j,k} \left( \mathbb{E}\left[ t_{j,k}^{(-1)} \right] + \mathbb{E}\left[ t_{j,k}^{(+1)} \right] - \frac{1}{2} \right).
\end{aligned} \tag{66}$$

This indicates that $q(\boldsymbol{w})$ follows a normal distribution given by

$$q\left( \boldsymbol{w}; \tilde{\boldsymbol{\mu}}_w, \tilde{\boldsymbol{\Lambda}}_w \right) = \mathcal{N}\left( \tilde{\boldsymbol{\mu}}_w, \tilde{\boldsymbol{\Lambda}}_w \right), \tag{67}$$

in which

$$\tilde{\boldsymbol{\mu}}_w = \tilde{\boldsymbol{\Lambda}}_w^{-1} \sum_{k=1}^{K} \sum_{j=1}^{J_k} \boldsymbol{s}_{j,k} \left( \mathbb{E}\left[ t_{j,k}^{(-1)} \right] + \mathbb{E}\left[ t_{j,k}^{(+1)} \right] - \frac{1}{2} \right), \tag{68}$$

$$\tilde{\boldsymbol{\Lambda}}_w = \mathbb{E}[\boldsymbol{\Lambda}] + 2 \sum_{k=1}^{K} \sum_{j=1}^{J_k} \chi(\xi_{j,k}) \boldsymbol{s}_{j,k} \boldsymbol{s}_{j,k}^\intercal. \tag{69}$$

### 2.3 Update rules for other variational parameters

For other latent variables in RefMap besides $v_{-1}$, $v_{+1}$ and $\boldsymbol{w}$, we carry out the naive MFVI and obtain

$$q\left(\boldsymbol{u}_k; \tilde{\boldsymbol{\mu}}_k, \tilde{\boldsymbol{\Lambda}}_k\right) = \mathcal{N}\left(\boldsymbol{u}_k; \tilde{\boldsymbol{\mu}}_k, \tilde{\boldsymbol{\Lambda}}_k^{-1}\right), \tag{70}$$

$$q\left(m_{j,k}; \tilde{\mu}_{j,k}, \tilde{\lambda}_{j,k}\right) = \mathcal{N}\left(m_{j,k}; \tilde{\mu}_{j,k}, \tilde{\lambda}_{j,k}^{-1}\right), \tag{71}$$

$$q(\lambda_{j,k}; \tilde{a}_{j,k}, \tilde{b}_{j,k}) = \mathsf{Gamma}\left(\lambda_{j,k}; \tilde{a}_{j,k}, \tilde{b}_{j,k}\right), \tag{72}$$

$$q\left(\tau_{-1}; \tilde{a}_{-1}, \tilde{b}_{-1}\right) = \mathsf{Gamma}\left(\tau_{-1}; \tilde{a}_{-1}, \tilde{b}_{-1}\right), \tag{73}$$

$$q\left(\tau_{+1}; \tilde{a}_{+1}, \tilde{b}_{+1}\right) = \mathsf{Gamma}\left(\tau_{+1}; \tilde{a}_{+1}, \tilde{b}_{+1}\right), \tag{74}$$

$$q\left(\tau_0; \tilde{a}_0, \tilde{b}_0\right) = \mathsf{Gamma}\left(\tau_0; \tilde{a}_0, \tilde{b}_0\right), \tag{75}$$

$$q\left(m_{-1}, \lambda_{-1}; \tilde{\mu}_{-1}, \tilde{\beta}_{-1}, \tilde{c}_{-1}, \tilde{d}_{-1}\right) = \mathcal{N}\left(m_{-1}; \tilde{\mu}_{-1}, \left(\tilde{\beta}_{-1}\lambda_{-1}\right)^{-1}\right) \mathsf{Gamma}\left(\lambda_{-1}; \tilde{c}_{-1}, \tilde{d}_{-1}\right), \tag{76}$$

$$q\left(m_{+1}, \lambda_{+1}; \tilde{\mu}_{+1}, \tilde{\beta}_{+1}, \tilde{c}_{+1}, \tilde{d}_{+1}\right) = \mathcal{N}\left(m_{+1}; \tilde{\mu}_{+1}, \left(\tilde{\beta}_{+1}\lambda_{+1}\right)^{-1}\right) \mathsf{Gamma}\left(\lambda_{+1}; \tilde{c}_{+1}, \tilde{d}_{+1}\right), \tag{77}$$

$$q\left(\boldsymbol{t}_{j,k}; \tilde{\boldsymbol{\pi}}_{j,k}\right) = \tilde{\boldsymbol{\pi}}_{j,k}^{\boldsymbol{t}_{j,k}}, \tag{78}$$

$$q\left(\boldsymbol{\Lambda}; \tilde{\boldsymbol{W}}_\Lambda, \tilde{\nu}_\Lambda\right) = \mathcal{W}\left(\tilde{\boldsymbol{W}}_\Lambda, \tilde{\nu}_\Lambda\right), \tag{79}$$

in which

$$\tilde{\boldsymbol{\mu}}_k = \tilde{\boldsymbol{\Lambda}}_{u_k}^{-1}\left(\sqrt{N}\boldsymbol{z}_k + \mathbb{E}[\boldsymbol{\Lambda}_k]\mathbb{E}[\boldsymbol{m}_k]\right), \tag{80}$$

$$\tilde{\boldsymbol{\Lambda}}_k = N\boldsymbol{\Sigma}_k + \mathbb{E}[\boldsymbol{\Lambda}_k], \tag{81}$$

$$\tilde{\mu}_{j,k} = \left(\mathbb{E}[\lambda_{j,k}]\sum_{i=1}^{I_{j,k}}\mathbb{E}[u_{i,j,k}] - \mathbb{E}[v_{-1}]\mathbb{E}[\tau_{-1}]\mathbb{E}\left[t_{j,k}^{(-1)}\right] + \mathbb{E}[v_{+1}]\mathbb{E}[\tau_{+1}]\mathbb{E}\left[t_{j,k}^{(+1)}\right]\right)\tilde{\lambda}_{j,k}^{-1}, \tag{82}$$

$$\tilde{\lambda}_{j,k} = I_{j,k}\mathbb{E}[\lambda_{j,k}] + \mathbb{E}\left[t_{j,k}^{(-1)}\right]\mathbb{E}[\tau_{-1}] + \mathbb{E}\left[t_{j,k}^{(0)}\right]\mathbb{E}[\tau_0] + \mathbb{E}\left[t_{j,k}^{(+1)}\right]\mathbb{E}[\tau_{+1}], \tag{83}$$

$$\tilde{a}_{j,k} = a_0 + \frac{I_{j,k}}{2}, \tag{84}$$

$$\tilde{b}_{j,k} = b_0 + \frac{1}{2}\sum_{i=1}^{I_{j,k}}\mathbb{E}\left[u_{i,j,k}^2\right] + \frac{I_{j,k}}{2}\mathbb{E}\left[m_{j,k}^2\right] - \mathbb{E}\left[m_{j,k}\right]\sum_{i=1}^{I_{j,k}}\mathbb{E}\left[u_{i,j,k}\right], \tag{85}$$

$$\tilde{a}_{-1} = a_0 + \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{J_k}\mathbb{E}\left[t_{j,k}^{(-1)}\right], \tag{86}$$

$$\tilde{b}_{-1} = b_0 + \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{J_k}\mathbb{E}\left[t_{j,k}^{(-1)}\right]\left(\mathbb{E}\left[m_{j,k}^2\right] + \mathbb{E}\left[v_{-1}^2\right] + 2\mathbb{E}\left[m_{j,k}\right]\mathbb{E}\left[v_{-1}\right]\right), \tag{87}$$

$$\tilde{a}_{+1} = a_0 + \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{J_k}\mathbb{E}\left[t_{j,k}^{(+1)}\right], \tag{88}$$

$$\tilde{b}_{+1} = b_0 + \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{J_k}\mathbb{E}\left[t_{j,k}^{(+1)}\right]\left(\mathbb{E}\left[m_{j,k}^2\right] + \mathbb{E}\left[v_{+1}^2\right] - 2\mathbb{E}\left[m_{j,k}\right]\mathbb{E}\left[v_{+1}\right]\right), \tag{89}$$

$$\tilde{a}_0 = a_0 + \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{J_k}\mathbb{E}\left[t_{j,k}^{(0)}\right], \tag{90}$$

$$\tilde{b}_0 = b_0 + \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{J_k}\mathbb{E}\left[m_{j,k}^2\right]\mathbb{E}\left[t_{j,k}^{(0)}\right], \tag{91}$$

$$\tilde{\mu}_{-1} = \frac{\beta_0\mu_0 + \mathbb{E}[r_{-1}]}{\beta_0 + 1}, \tag{92}$$

$$\tilde{\beta}_{-1} = \beta_0 + 1, \tag{93}$$

$$\tilde{c}_{-1} = a_0 + \frac{1}{2}, \tag{94}$$

$$\tilde{d}_{-1} = b_0 + \frac{1}{2}\beta_0\mu_0^2 + \frac{1}{2}\mathbb{E}\left[r_{-1}^2\right] - \frac{1}{2(\beta_0 + 1)}\left(\beta_0\mu_0 + \mathbb{E}[r_{-1}]\right)^2, \tag{95}$$

$$\tilde{\mu}_{+1} = \frac{\beta_0\mu_0 + \mathbb{E}[r_{+1}]}{\beta_0 + 1}, \tag{96}$$

$$\tilde{\beta}_{+1} = \beta_0 + 1, \tag{97}$$

$$\tilde{c}_{+1} = a_0 + \frac{1}{2}, \tag{98}$$

$$\tilde{d}_{+1} = b_0 + \frac{1}{2}\beta_0\mu_0^2 + \frac{1}{2}\mathbb{E}\left[r_{+1}^2\right] - \frac{1}{2(\beta_0 + 1)}\left(\beta_0\mu_0 + \mathbb{E}[r_{+1}]\right)^2, \tag{99}$$

$$\tilde{\pi}_{j,k}^{(i)} = \frac{\exp\left\{\tilde{\rho}_{j,k}^{(i)}\right\}}{\exp\left\{\tilde{\rho}_{j,k}^{(-1)}\right\} + \exp\left\{\tilde{\rho}_{j,k}^{(0)}\right\} + \exp\left\{\tilde{\rho}_{j,k}^{(+1)}\right\}} \;\; (i = -1, 0, +1), \tag{100}$$

$$\tilde{\nu}_\Lambda = \nu_0 + 1, \tag{101}$$

$$\tilde{\boldsymbol{W}}_\Lambda = \left(\boldsymbol{W}_0^{-1} + \mathbb{E}\left[\boldsymbol{w}\boldsymbol{w}^\intercal\right]\right)^{-1}, \tag{102}$$

and we define

$$\tilde{\rho}_{j,k}^{(-1)} = \frac{1}{2}\mathbb{E}[\ln\tau_{-1}] - \frac{1}{2}\mathbb{E}[\tau_{-1}]\left(\mathbb{E}\left[m_{j,k}^2\right] + \mathbb{E}\left[v_{-1}^2\right] + 2\mathbb{E}[m_{j,k}]\mathbb{E}[v_{-1}]\right) + \mathbb{E}[\ln\pi_{j,k}] - \ln 2, \tag{103}$$

$$\tilde{\rho}_{j,k}^{(+1)} = \frac{1}{2}\mathbb{E}[\ln\tau_{+1}] - \frac{1}{2}\mathbb{E}[\tau_{+1}]\left(\mathbb{E}\left[m_{j,k}^2\right] + \mathbb{E}\left[v_{+1}^2\right] - 2\mathbb{E}[m_{j,k}]\mathbb{E}[v_{+1}]\right) + \mathbb{E}[\ln\pi_{j,k}] - \ln 2, \tag{104}$$

$$\tilde{\rho}_{j,k}^{(0)} = \frac{1}{2}\mathbb{E}[\ln\tau_0] - \frac{1}{2}\mathbb{E}[\tau_0]\mathbb{E}\left[m_{j,k}^2\right] + \mathbb{E}[\ln(1 - \pi_{j,k})]. \tag{105}$$

**Algorithm 1:** MFVI for RefMap

**Input** : $Z$-scores $z_{i,j,k}$, epigenome features $\boldsymbol{s}_{j,k}$ and LD matrices $\boldsymbol{\Sigma}_k$.
**Output** : Posteriors $q$ and local variational parameters $\xi_{j,k}$.

Initialize variational parameters.
**while** *not converged* **do**
  Update global variational parameters based on Eqs. 31, 32, 33, 34, 49, 50, 51, 52, 68, 69, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, and 102.
  Update local variational parameters based on Eq. 107.
  Calculate ELBO (details omitted).
**end**

### 2.4 Update rules for local variational parameters

One needs to maximize the lower bound on marginal likelihood in Eq. 64 with respect to $\xi_{j,k}$ to rationalize the local variational inference. In particular, we have the following optimization problem

$$Q\left(\boldsymbol{\xi}, \boldsymbol{\xi}^{\text{old}}\right) \propto \sum_{k=1}^{K} \sum_{j=1}^{J_k} \ln \sigma(\xi_{j,k}) - \frac{1}{2}\xi_{j,k} - \chi(\xi_{j,k}) \left((\boldsymbol{w}^\intercal \boldsymbol{s}_{j,k})^2 - \xi_{j,k}^2\right). \tag{106}$$

Solving the above problem with respect to each $\xi_{j,k}$ gives its update rule

$$\xi_{j,k}^{\text{new}} = \sqrt{\boldsymbol{s}_{j,k}^\intercal \mathbb{E}\left[\boldsymbol{w}\boldsymbol{w}^\intercal\right] \boldsymbol{s}_{j,k}}. \tag{107}$$

### 2.5 Coordinate ascent algorithm for MFVI

With the above update rules we can construct a coordinate ascent algorithm to update variational parameters iteratively until convergence (i.e., the change of ELBO falls below a threshold which was set to be $10^{-6}$ in our study). The inference algorithm is summarized in Algorithm 1.

## References


[1] B. Bulik-Sullivan, H. K. Finucane, V. Anttila, A. Gusev, F. R. Day, P.-R. Loh, R. Consortium, P. G. Consortium, G. C. for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, L. Duncan, J. R. B. Perry, N. Patterson, E. B. Robinson, M. J. Daly, A. L. Price, and B. M. Neale. An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, 47:1236, Sept. 2015. URL `https://doi.org/10.1038/ng.3406`.

[2] B. K. Bulik-Sullivan, P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang, S. W. G. of the Psychiatric Genomics Consortium, N. Patterson, M. J. Daly, A. L. Price, and B. M. Neale. Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, 47:291, Feb. 2015. URL `https://doi.org/10.1038/ng.3211`.

[3] H. K. Finucane, B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P.-R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, S. Ripke, F. R. Day, R. Consortium, S. W. G. of the Psychiatric Genomics Consortium, T. R. A. C. I. Consortium, S. Purcell, E. Stahl, S. Lindstrom, J. R. B. Perry, Y. Okada, S. Raychaudhuri, M. J. Daly, N. Patterson, B. M. Neale, and A. L. Price. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, 47:1228, Sept. 2015. URL `https://doi.org/10.1038/ng.3404`.

[4] B. Han, H. M. Kang, and E. Eskin. Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. *PLOS Genetics*, 5(4):e1000456, Apr. 2009. doi: 10.1371/journal.pgen.1000456. URL `https://doi.org/10.1371/journal.pgen.1000456`.

[5] M. Harva and A. Kabán. Variational learning for rectified factor analysis. *Signal Process.*, 87(3):509–527, Mar. 2007. ISSN 0165-1684. doi: 10.1016/j.sigpro.2006.06.006. URL `https://doi.org/10.1016/j.sigpro.2006.06.006`.

[6] J. W. J. Joo, F. Hormozdiari, B. Han, and E. Eskin. Multiple testing correction in linear mixed models. *Genome Biology*, 17(1):62, Apr. 2016. ISSN 1474-760X. URL `https://doi.org/10.1186/s13059-016-0903-6`.

[7] G. Kichaev, W.-Y. Yang, S. Lindstrom, F. Hormozdiari, E. Eskin, A. L. Price, P. Kraft, and B. Pasaniuc. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLOS Genetics*, 10(10):e1004722, Oct. 2014. doi: 10.1371/journal.pgen.1004722. URL `https://doi.org/10.1371/journal.pgen.1004722`.

[8] B. Pasaniuc and A. L. Price. Dissecting the genetics of complex traits using summary association statistics. *Nature Reviews Genetics*, 18:117, Nov. 2016. URL `https://doi.org/10.1038/nrg.2016.142`.