A Bayesian Linear Mixed Models

In this section, we explain in more detail the Bayesian take on the linear model

$$\bar{\mathbf{y}}_{g,c} = \sum_{t=1}^{T} \mathbf{m}_{t,g} \omega_{t,c} + \text{noise}, \tag{1}$$

where we model the normalized gene expression signal $\bar{\mathbf{y}}_{g,c} = \mathbf{y}_{g,c} - \bar{\mathbf{y}}_g - \bar{\mathbf{y}}_c$ as a linear combination of the motif scores $\mathbf{M}_{T,G}$, which are each, dependent on the sample c and the TF t, weighted with $\omega_{T,C}$. The term 'noise' represents all signal that cannot be explained by the model, i.e. the linear combination of the motif scores $\mathbf{M}_{T,G}$. This can be any technical noise, motif influence for which the linear assumption might be too simplistic, but also any other source that drives the gene expression $\mathbf{y}_{g,c}$ and is not modeled. The model was originally introduced by The Fantom Consortium and the Riken Omics Science Center (2009) and subsequently expanded by Balwierz *et al.* (2014).

The main idea behind a Bayesian approach is to include some prior knowledge into the data, called the prior. Here, we model $\omega_{t,c}$, the influence of motif $t \in \{1, \ldots, T\}$ in condition $c \in \{1, \ldots, C\}$, as a normal distributed prior with mean zero. Its marginal distributions are

$$\omega_C \sim \mathcal{N}\left(\mathbf{0}, \mathbf{V}_C\right),\tag{2}$$

where \mathbf{V}_{C} is the covariance in motif activity over all conditions and

$$\omega_T \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_T\right),\tag{3}$$

with $\sigma^2 \mathbf{I}_T$ being the covariance over all motifs. Hence, we assume independence between motifs. Assuming dependence between motifs with covariance Ψ can easily be implemented in the model. Making use of the vectorization representation of multivariate normal distribution, we can write the multivariate Normal distribution of $\omega_{T,C}$ as follows:

$$\operatorname{vec}\left(\omega_{T,C}\right) \sim \mathcal{N}\left(\mathbf{0}, \sigma^{2}\mathbf{V}_{C} \otimes \mathbf{I}_{T}\right).$$
(4)

Analogously, we can rewrite Eq. 1 in matrix-vector notation:

$$\operatorname{vec}\left(\mathbf{Y}_{G,C}\right) \sim \mathcal{N}\left(\mathbf{M}_{T,G}^{\mathsf{T}}\omega_{T,C}, \delta \boldsymbol{\Sigma}_{C} \otimes \mathbf{I}_{G}\right),\tag{5}$$

with Σ_C being the covariance conditions and $\delta \mathbf{I}_G$ covariance over genes. In the following, we use Bayes' rule on marginal and conditional Gaussians (see ?, Chapter 2.3, p. 93). We first list the general formulas here, which we copy from ?: The marginal distribution of \mathbf{x} and conditional distribution of \mathbf{y} given \mathbf{x} are given as follows:

$$p(\mathbf{x}) = \mathcal{N}\left(\mathbf{x}|\mu, \mathbf{\Lambda}^{-1}\right) \tag{6}$$

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}\left(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}\right).$$
(7)

Translating that notation into our own notation, we get the following equivalences:

$$\mathbf{x} \equiv \operatorname{vec}\left(\omega_{T,C}\right) \qquad \mu \equiv \mathbf{0} \qquad \mathbf{\Lambda}^{-1} \equiv \sigma^2 \mathbf{V}_C \otimes \mathbf{I}_T \qquad (8)$$

$$\mathbf{y} \equiv \operatorname{vec}\left(\mathbf{Y}_{G,C}\right) \qquad \mathbf{A} \equiv \mathbf{M}_{T,G}^{\mathsf{T}} \qquad \mathbf{b} \equiv \mathbf{0} \qquad \mathbf{L}^{-1} \equiv \delta \mathbf{\Sigma}_C \otimes \mathbf{I}_G.$$
 (9)

In our own notation, and following Eq. 7, the distribution of $\mathbf{Y}_{G,C}$ given $\omega_{T,C}$ can be rewritten as:

$$\operatorname{vec}\left(\mathbf{Y}_{G,C}|\omega_{T,C}\right) \sim \mathcal{N}\left(\mathbf{0}, \sigma^{2}\mathbf{V}_{C} \otimes \mathbf{\Pi}_{G} + \delta \boldsymbol{\Sigma}_{C} \otimes \mathbf{I}_{G}\right),\tag{10}$$

where $\Pi_G = \mathbf{M}_{T,G}^{\mathsf{T}} \mathbf{M}_{T,G}$.

Going back to the general formula from ?, the marginal distribution of \mathbf{y} and conditional distribution of \mathbf{x} given \mathbf{y} are computed according to the following:

$$p(\mathbf{y}) = \mathcal{N}\left(\mathbf{y}|\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{T}\right)$$
(11)

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}\left(\mathbf{x}|\mathbf{\Sigma}\left\{\mathbf{A}^{T}\mathbf{L}\left(\mathbf{y}-\mathbf{b}\right)+\mathbf{\Lambda}\boldsymbol{\mu}\right\},\mathbf{\Sigma}\right),$$
(12)

with
$$\boldsymbol{\Sigma} = \left(\boldsymbol{\Lambda} + \boldsymbol{A}^T \boldsymbol{\Lambda} \boldsymbol{A} \right)^{-1}$$
. (13)

Hence, based on Eq. 11 and Eq. 12, the posterior distribution of $\omega_{T,C}$ given $\mathbf{Y}_{G,C}$ is then

$$\operatorname{vec}\left(\omega_{T,C}|\mathbf{Y}_{G,C}\right) \sim \mathcal{N}\left(\left(\mathbf{V}_{C} \otimes \mathbf{M}_{T,G}^{\mathsf{T}}\mathbf{I}_{T}\right) \mathbf{\Lambda}_{\mathbf{CG} \otimes \mathbf{CG}}^{-1} \operatorname{vec}\left(\mathbf{Y}_{G,C}\right), \mathbf{\Lambda}_{\mathbf{CG} \otimes \mathbf{CG}}\right)$$
(14)

with

$$\Lambda_{\mathbf{CG}\otimes\mathbf{CG}} = \sigma^2 \mathbf{V}_C \otimes \mathbf{\Pi}_G + \delta \mathbf{\Sigma}_C \otimes \mathbf{I}_G.$$
(15)

For the reformulation, we used the *Woodbury matrix identity* (Woodbury, 1950). For the computation of the posterior values of $\omega_{T,C}$, with notation $\hat{\omega}_{T,C}$, it suffices to compute the mean of Eq. 14:

$$\operatorname{vec}\left(\hat{\omega}_{T,C}|\mathbf{Y}_{G,C}\right) = \left(\mathbf{V}_{C} \otimes \mathbf{M}_{T,G}^{\mathsf{T}}\mathbf{I}_{T}\right) \left[\sigma^{2}\mathbf{V}_{C} \otimes \mathbf{\Pi}_{G} + \delta \boldsymbol{\Sigma}_{C} \otimes \mathbf{I}_{G}\right]^{-1} \operatorname{vec}\left(\mathbf{Y}_{G,C}\right)$$
(16)

As the covariance matrix is assumed to be the sum of Kronecker products, the runtime complexity is reduced to $O(G^3 + C^3)$ in a $O(G^2 + C^2)$ space instead of $O(G^3C^3)$ runtime and a memory requirement of $O(G^2C^2)$ (Lippert *et al.*, 2014).