1 The genomic model

The genomic model was implemented in a likelihood setting with Newton-Raphson and with the EM algorithm. Here the model is fitted using a Bayesian McMC approach.

To recapitulate, in its original parametrisation, the initial hierarchy of the genomic model is defined by the expressions

$$\begin{aligned} y|\mu, b, \sigma_e^2 &\sim N\left(1\mu + Wb, I\sigma_e^2\right), \\ b|\sigma_b^2 &\sim N\left(0, I\sigma_b^2\right). \end{aligned}$$

where the matrix of marker genotype codes W of dimension $n \times m$ is centred and scaled, and the vector of marker effects b has m elements. The genomic values are defined as $g = Wb$. The reparametrised model uses the decomposition

$$\begin{aligned} G &= \frac{1}{m}WW' \\ &= \frac{1}{m}U\Delta U' \\ &= U\Lambda U', \quad \Lambda = \frac{1}{m}\Delta. \end{aligned}$$

On defining $\alpha \sim N\left(0, \Lambda\sigma_g^2\right)$, where $\sigma_g^2 = m\sigma_b^2$, the alternative parametrisation as defined on page **??** is

$$\begin{aligned} y|\mu, \alpha, \sigma_e^2 &\sim N\left(1\mu + U\alpha, I\sigma_e^2\right), \\ \alpha|U, \sigma_g^2 &\sim SN\left(0, \Lambda\sigma_g^2\right), \\ G &= U\Lambda U', \\ y|\mu, \sigma_g^2, \sigma_e^2 &\sim N\left(1\mu, U\Lambda U'\sigma_g^2 + I\sigma_e^2\right). \end{aligned}$$

The data y has n elements and in contrast with (**??**), here a scalar mean μ is included. The diagonal matrix Λ contains the eigenvalues λ_i of U, $i = 1, \ldots, n$. The last eigenvalue is $\lambda_n = 0$. The vector

$$\begin{aligned} \alpha' &= (\alpha_1, \ldots, \alpha_{n-1}, \alpha_n)' \\ &= \left(\alpha'_{-n}\alpha_n\right) \end{aligned}$$

of dimension n has its first $n-1$ elements different from zero (the vector α_{-n} above) and the last element $\alpha_n = 0$. As mentioned before, the marginal density of α_n is a point mass at zero, because $\lambda_n = 0$ and therefore $\left[\alpha|U, \sigma_g^2\right]$ has the form of a singular normal distribution. In the reparametrised model the column vector of genomic values is $g = U\alpha$.

1.1 An alternative definition of genomic variance

The Appendix of the Overview chapter discusses the distinction between the parameters defined at the level of the quantitative genetic model, regarded as the data generating

mechanism, and the parameters defined at the level of the operational model, used to analyse the data. The genomic model is such an operational model. It was noted that the genomic variance of the genomic model (**??**), that is the same parameter σ_g^2 in the present example, has a tenuous connection with the parameter (**??**), that is argued, is the real focus of inference.

An alternative estimator of genomic variance (**??**) was defined on page **??**, that is better aligned with (**??**). This subsection shows that estimator (**??**) can be obtained when the Bayesian model is implemented in an McMC computing environment, with minimum extra code.

The $n \times 1$ vector of genomic values is $g = Wb = U\alpha$, where W is the $n \times p$ scaled matrix of marker genotypes and b is the $p \times 1$ vector of genetic marker effects. The genomic variance or variance of genomic values was defined in (**??**), and in terms of the reparametrised model, can be written as

$$\begin{aligned}
\sigma_G^2 &= \frac{1}{n} g'g - \left(\frac{1}{n} 1'g\right)^2 \\
&= \frac{1}{n} \alpha' U' U \alpha - \left(\frac{1}{n} 1' U \alpha\right)^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \alpha_i^2, \qquad (2)
\end{aligned}$$

because when matrix W is centred, the second term in the right hand side vanishes (see **NOTE** below for the proof). The vector α is unknown, and is therefore inferred from its marginal posterior distribution. When the Markov chain converges to its stationary distribution, extractions from this distribution are given by (**??**). For example, if $alpha^{[t]}$ is the vector of draws from (**??**) at round t of the McMC sampler, an extraction of σ_G^2 from its marginal posterior distribution is $\sigma_G^{2[t]} = \frac{1}{n} \alpha^{[t]\prime} \alpha^{[t]}$. The `R`-code is `mean(alpha*alpha)`.

1.1.1 NOTE

To show that the scalar $1'U\alpha = 0$ proceed as follows. The following assumptions are needed:

$$\begin{aligned}
1'W &= 0, \\
G &= WW' = U\Lambda U', \\
U'U &= UU' = I.
\end{aligned}$$

Then

$$\begin{aligned}
1'WW' &= 0, \\
1'U\Lambda U' &= 0, \\
1'U\Lambda &= 0.
\end{aligned}$$

Write

$$\begin{aligned} 1'U\Lambda &= 1'\left[U_1 U_2 \ldots U_n\right]\Lambda \\ &= \left[m_1, m_2, \ldots, m_n\right]\Lambda \\ &= \left[m_1\lambda_1, m_2\lambda_2, \ldots, m_n\lambda_n\right] \\ &= 0, \end{aligned} \tag{3}$$

where $m_i\lambda_i$ are scalars. The implication of (3) is that $m_1 = m_2 = \ldots = m_{n-1} = 0$, because Λ is a diagonal matrix of positive eigenvalues, except the last one, $\lambda_n = 0$, which is equal to zero. The vector α has elements $\alpha_i \sim N\left(0, \lambda_i\sigma_b^2\right)$, $i = 1, 2, \ldots, n-1$, with $\alpha_n \sim N\left(0, \lambda_n\sigma_b^2\right)$, $\lambda_n = 0$. Therefore $\alpha_n = 0$ a posteriori (with probability 1), and it follows that

$$\begin{aligned} 1'U\alpha &= \left[m_1, m_2, \ldots, m_n\right]\left[\alpha_1, \alpha_2, \ldots, \alpha_n\right]' \\ &= \left[0, 0, \ldots, m_n\right]\left[\alpha_1, \alpha_2, \ldots, 0\right]' = 0. \end{aligned}$$