DERIVATIONS OF THE FULLY CONDITIONAL POSTERIOR DENSITIES

The joint posterior distribution of the unknown parameters and hidden variables, given the data, is proportional to the product of the joint prior and the likelihood, and the fully conditional posteriors of the parameters can be easily determined by selecting the terms including the parameter in question from the joint posterior. (For simplicity: $\star =$ ”the data, and the parameters except the one in question”)

The fully conditional posterior distribution of the population intercept $\beta_0$ is proportional to the likelihood, since the prior of $\beta_0$ is proportional to one, so

$$p(\beta_0|\star) \propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma_0^2}(y_i - \beta_0 - \sum_{j=1}^{p} \gamma_j \beta_j x_{ij} - u_i)^2\right).$$  \hspace{1cm} (S1)

This is a product of $n$ kernels of normal distributions, with a common variance $\sigma_0^2$, and means $(y_i - \sum_{j=1}^{p} \gamma_j \beta_j x_{ij} - u_i)$, $i = 1, \ldots, n$. The set of Gaussian functions is closed under multiplication, i.e. the product of normal densities is also a normal density, with the mean and the variance of a product density given by

$$\mu = \frac{\sum \mu_i \sigma_i^2}{\sum \frac{1}{\sigma_i^2}} \quad \text{and} \quad \sigma^2 = \frac{1}{\sum \frac{1}{\sigma_i^2}},$$

respectively. Hence, since in this case variance of the factors in the product is constant, the mean of product density reduces to the sum of the means of the individual distributions, divided by $n$, while the variance of the product density is given simply by the variance of the individual distributions divided by $n$.

The fully conditional posterior distribution of the regression coefficients $\beta_j$ is proportional to the product of the likelihood and the conditional prior $p(\beta_j|\sigma_j^2)$,

$$p(\beta_j|\star) \propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma_0^2}(y_i - \beta_0 - \sum_{l=1}^{p} \sum_{m=1}^{2} \gamma_{il} \beta_{lm} x_{ilm} - u_i)^2\right) \exp\left(-\frac{\beta_j^2}{2\sigma_j^2}\right),$$  \hspace{1cm} (S2)

that is a product of two types of kernels of normal distributions. One distribution comes from the prior, it has mean 0 and variance $\sigma_j^2$. The other part comes from the likelihood —
regarding \( \beta_j \) there are \( n \) kernels of normal distributions with means

\[
\frac{1}{\gamma_j x_{ij}} (y_i - \beta_0 - \sum_{(l,m) \neq (j,k)} \gamma_l \beta_m x_{ilm} - u_i), \quad i = 1, \ldots, n,
\]

and variances \( \sigma_0^2/(\gamma_j x_{ij})^2, \quad i = 1, \ldots, n \). Hence we get as a product a normal distribution with a mean

\[
\sum_{i=1}^{n} \left( \frac{\gamma_j x_{ij}}{\sigma_0^2} (y_i - \beta_0 - \sum_{(l,m) \neq (j,k)} \gamma_l \beta_m x_{ilm} - u_i) \right) / \left( \sum_{i=1}^{n} \frac{\gamma_j x_{ij}}{\sigma_0^2} + \frac{1}{\sigma_j^2} \right)
\]

and variance

\[
\left( \sum_{i=1}^{n} \frac{\gamma_j x_{ij}}{\sigma_0^2} + \frac{1}{\sigma_j^2} \right)^{-1},
\]

which equals the variance given in the Appendix when each term is multiplied by \( \sigma_0^2 \).

The conditional posterior distribution of the polygenic effect \( u \) is multivariate normal, as it is a product of the multivariate normal prior distribution, \( u|\sigma_u^2 \sim N_n(0, A\sigma_u^2) \) and a normal likelihood,

\[
p(u|\star) \propto p(u|\sigma_u^2)p(y|\star)
\]

\[
\propto \exp \left( -\frac{1}{2} u'(A\sigma_u^2)^{-1} u \right) \prod_{i=1}^{n} \exp \left( -\frac{1}{2 \sigma_0^2} (y_i - \beta_0 - \sum_{l=1}^{p} \sum_{m=1}^{2} \gamma_l \beta_m x_{ilm} - u_i)^2 \right).
\]

The kernel of the likelihood part of the posterior can also be interpreted as a kernel of a multivariate normal density regarding the polygenic effect

\[
u \sim N_n \left( y - \beta_0 - X\Gamma \beta, I_n \sigma_0^2 \right),
\]

so can simply use the same multiplication rule of normal densities as in previous cases, and we therefore get conditional posterior mean

\[
\left( \frac{1}{\sigma_0^2} I_n + \frac{1}{\sigma_u^2} A^{-1} \right)^{-1} \frac{1}{\sigma_0^2} \left( y - \beta_0 - X\Gamma \beta \right)
\]

(S3)

and covariance

\[
\left( \frac{1}{\sigma_0^2} I_n + \frac{1}{\sigma_u^2} A^{-1} \right)^{-1}.
\]
The variance parameters of the model have inverse $\chi^2$-priors which, due to conjugacy, leads to inverse $\chi^2$-posteriors. The prior of the residual variance $\sigma_0^2$ is proportional to $1/\sigma_0^2$, which leads to a posterior density
\[
p(\sigma_0^2|\star) \propto (\sigma_0^2)^{-(1+\frac{n}{2})} \exp \left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \gamma_j \beta_j x_{ij} - u_i)^2 \right).
\] (S4)

Regarding $\sigma_0^2$, this is an unnormalized probability density function of an inverse $\chi^2$-distribution, with $n$ degrees of freedom and scale parameter equal to
\[
\frac{1}{n} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \gamma_j \beta_j x_{ij} - u_i)^2.
\]
The inverse $\chi^2$-posteriors of the other variance parameters, $\sigma_j^2$ and $\sigma_u^2$, are derived with an identical logic.

In case of the Laplace($0$, $\lambda$) prior for the effect size, the effect variance $\sigma_j^2 \sim \text{Exp}(\lambda^2/2)$, and hence the fully conditional posterior of $\sigma_j^2$ is proportional to the product of the exponential prior and the normal, conditional prior of the effect size $\beta_j|\sigma_j^2$,
\[
p(\sigma_j^2|\star) \propto p(\beta_j|\sigma_j^2) p(\sigma_j^2|\lambda^2)
\propto (\sigma_j^2)^{-1/2} \exp \left(-\frac{\beta_j^2}{2\sigma_j^2} \right) \frac{\lambda^2}{2} \exp \left(-\frac{\lambda^2\sigma_j^2}{2} \right).
\]
Since exponential density is not conjugate to normal density, we need to consider the inverse of the variance
\[
p \left( \frac{1}{\sigma_j^2} | \star \right) \propto (\sigma_j^2)^{1/2} \exp \left(-\frac{\beta_j^2\sigma_j^2}{2} - \frac{\lambda^2}{2\sigma_j^2} \right) (\sigma_j^2)^{-2},
\]
where the last term is the Jacobian of the transformation $\sigma_j^2 \rightarrow \sigma_j^{-2}$. By rearranging the terms, we get
\[
p \left( \frac{1}{\sigma_j^2} | \star \right) \propto (\sigma_j^2)^{-3/2} \exp \left(-\lambda^2 \frac{\beta_j^2\sigma_j^2}{2\sigma_j^2} + 1 \right)
\]
and completing the numerator of the exponent into square
\[
p \left( \frac{1}{\sigma_j^2} | \star \right) \propto (\sigma_j^2)^{-3/2} \exp \left(-\frac{\lambda^2 \beta_j^2\sigma_j^2}{2\sigma_j^2\lambda^2} - 2\sigma_j^2 \frac{\beta_j^2}{\lambda^2} + 1 \right) - \frac{\lambda^2\sigma_j^2\sqrt{\beta_j^2}}{2\sigma_j^2}.
\]
\(\sigma_j^2\) is canceled out from the the last term of the exponent, hence the term being constant and left out, after which the exponent is expanded by \(\beta_j^2/\lambda^2\) and we get

\[
p\left(\frac{1}{\sigma_j^2}\right) \propto (\sigma_j^2)^{-3/2} \exp\left(-\frac{\lambda^2 \left( (\sigma_j^2)^2 - 2\sigma_j^2 \sqrt{\frac{\lambda^2}{\beta_j^2}} + \frac{\lambda^2}{\beta_j^2} \right)}{2\sigma_j^2 \beta_j^2}ight)
\]

\[= (\sigma_j^2)^{-3/2} \exp\left(-\frac{\lambda^2 \left( (\sigma_j^2)^2 - \sqrt{\frac{\lambda^2}{\beta_j^2}} \right)^2}{2\sigma_j^2 \beta_j^2}\right), \quad (S5)
\]

that is an inverse-Gaussian probability density function with mean \(\mu'\) and shape \(\lambda'\)

\[
\mu' = \sqrt{\frac{\lambda^2}{\beta_j^2}} \quad \text{and} \quad \lambda' = \lambda^2,
\]

the parametrization of the inverse-Gaussian density being

\[
f(x) \propto x^{-3/2} \exp\left(-\frac{\lambda'(x - \mu')^2}{2(\mu')^2 x}\right).
\]

The fully conditional posterior distribution of the indicators \(\gamma_j\) is Bernoulli. Directly from Bayes’ formula we get

\[
p(\gamma_j|\mathbf{y}, \star) = \frac{p(\gamma_j)p(\mathbf{y}|\gamma_j, \star)}{p(\mathbf{y}|\star)} = \frac{p(\gamma_j)p(\mathbf{y}|\gamma_j, \star)}{p(\gamma_j = 0)p(\mathbf{y}|\gamma_j = 0, \star) + p(\gamma_j = 1)p(\mathbf{y}|\gamma_j = 1, \star)}
\]

\[= \left( \frac{\pi \mathcal{R}_j}{1 - \pi} \right)^{\gamma_j} \left( \frac{1 - \pi}{(1 - \pi) + \pi \mathcal{R}_j} \right)^{1 - \gamma_j}
\]

where \(\pi = p(\gamma_j = 1) = 1 - p(\gamma_j = 0), \quad \text{and} \quad \mathcal{R}_j = \frac{p(\mathbf{y}|\gamma_j = 1, \star)}{p(\mathbf{y}|\gamma_j = 0, \star)} = \frac{\exp\left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{h \neq j} \gamma_h \beta_h x_{ih} - \beta_j x_{ij})^2\right)}{\exp\left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{h \neq j} \gamma_h \beta_h x_{ih})^2\right)}
\]

\[= \exp\left( \frac{1}{2\sigma_0^2} \sum_{i=1}^{n} \left(2 \left( \sum_{k=1}^{2} \beta_k x_{ij} \right) (y_i - \beta_0 - \sum_{h \neq j} \gamma_h \beta_h x_{ih}) - (\beta_j x_{ij})^2\right)\right).
\]
<table>
<thead>
<tr>
<th></th>
<th>Hierarchical Laplace</th>
<th>Non-hierarchical Laplace</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>with indicator</td>
<td>no indicator</td>
</tr>
<tr>
<td>$nqtl$</td>
<td>⋆ 30</td>
<td></td>
</tr>
<tr>
<td>$\xi$</td>
<td>1 1 1 3 0.5</td>
<td>10 100 100 100 100</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Original 75 75 74 43 105</td>
<td>164 44 37 39 44</td>
</tr>
<tr>
<td></td>
<td>Mean 75 75 75 43 105</td>
<td>150 44 39 39 45</td>
</tr>
<tr>
<td></td>
<td>Max 75 75 75 43 106</td>
<td>175 45 42 40 45</td>
</tr>
<tr>
<td></td>
<td>Min 74 75 74 43 105</td>
<td>106 42 37 38 44</td>
</tr>
<tr>
<td></td>
<td>Std 0.05 0.05 0.05 0.023 0.07</td>
<td>27 0.61 1.4 0.37 0.2</td>
</tr>
<tr>
<td>$\pi$</td>
<td>Original 0.96 0.01</td>
<td>0.15 0.69 0.02 0.03</td>
</tr>
<tr>
<td></td>
<td>Mean 0.96 0.01</td>
<td>0.17 0.69 0.03 0.03</td>
</tr>
<tr>
<td></td>
<td>Max 0.97 0.01</td>
<td>0.23 0.83 0.10 0.04</td>
</tr>
<tr>
<td></td>
<td>Min 0.94 0.01</td>
<td>0.15 0.58 0.02 0.03</td>
</tr>
<tr>
<td></td>
<td>Std 0.007 0.0005</td>
<td>0.038 0.047 0.015 0.0008</td>
</tr>
</tbody>
</table>

Table S1: Estimated hyperparameter values of $\lambda$ and $\pi$ in the original QTL-MAS data, in addition to the mean, maximum, minimum and standard deviance of the estimates in the analyses of the 100 replicated data sets. The $nqtl$ and $\xi$ denote alternative hyperprior parameter values for the prior of the indicator $\pi \sim \text{Beta}(nqtl, p−nqtl)$ and the rate parameter $\lambda^2 \sim \text{Gamma}(1, \xi)$ under the hierarchical Laplace model and $\lambda \sim \text{Gamma}(1, \xi)$ under the non-hierarchical Laplace model, with and without the indicator variable. ⋆ denotes the optional Beta(1,1) prior of the indicator.