

File S1

Proof of Equation (4) of the Main Text

This section provides a proof for equation (4) in the main text and shows the conditions under which the equation holds. This equation is reintroduced below,

$$y_j = X_j \beta + \sum_{k=1}^m Z_{jk} \gamma_k + \varepsilon_j \quad (1)$$

where

$$Z_{jk} = \bar{Z}_j(\lambda_k) = \Delta_k^{-1} \int_{\lambda_k - \frac{1}{2}\Delta_k}^{\lambda_k + \frac{1}{2}\Delta_k} Z_j(\lambda) d\lambda \quad (2)$$

is the average Z_j for all (infinite number of) markers within bin k and Δ_k is the size of the bin, defined as

$$\Delta_k = \int_{\lambda_k - \frac{1}{2}\Delta_k}^{\lambda_k + \frac{1}{2}\Delta_k} d\lambda \quad (3)$$

The location of the bin is denoted by λ_k , which is the middle point position of the bin in the genome. The genetic effect γ_k for bin k is defined as

$$\gamma_k = \bar{\gamma}(\lambda_k) \Delta_k = \int_{\lambda_k - \frac{1}{2}\Delta_k}^{\lambda_k + \frac{1}{2}\Delta_k} \gamma(\lambda) d\lambda \quad (4)$$

which is the sum of effects for all (infinite number of) markers in bin k . Equation (1) is an approximation. The exact version of the equation should be

$$y_j = X_j \beta + \sum_{k=1}^m \alpha_{jk} + \varepsilon_j \quad (5)$$

where

$$\alpha_{jk} = \int_{\lambda_k - \frac{1}{2}\Delta_k}^{\lambda_k + \frac{1}{2}\Delta_k} Z_j(\lambda) \gamma(\lambda) d\lambda \quad (6)$$

Let us rewrite equation (6) using

$$\alpha_{jk} = \int_{\lambda_k - \frac{1}{2}\Delta_k}^{\lambda_k + \frac{1}{2}\Delta_k} \left[Z_j(\lambda) - \bar{Z}_j(\lambda_k) + Z_{jk} \right] \left[\gamma(\lambda) - \bar{\gamma}(\lambda_k) + \Delta_k^{-1} \gamma_k \right] d\lambda \quad (7)$$

This is because $Z_{jk} = \bar{Z}_j(\lambda_k)$ and $\gamma_k = \bar{\gamma}(\lambda_k) \Delta_k$ as defined in equations (2) and (4), respectively. Expanding the product of equation (7) yields

$$\begin{aligned}
\alpha_{jk} &= \int_{\lambda_k - \frac{1}{2}\Delta_k}^{\lambda_k + \frac{1}{2}\Delta_k} Z_{jk} \gamma_k \Delta_k^{-1} d\lambda + \int_{\lambda_k - \frac{1}{2}\Delta_k}^{\lambda_k + \frac{1}{2}\Delta_k} [Z_j(\lambda) - \bar{Z}_j(\lambda_k)] [\gamma(\lambda) - \bar{\gamma}(\lambda_k)] d\lambda \\
&= Z_{jk} \gamma_k \Delta_k^{-1} \int_{\lambda_k - \frac{1}{2}\Delta_k}^{\lambda_k + \frac{1}{2}\Delta_k} d\lambda + \int_{\lambda_k - \frac{1}{2}\Delta_k}^{\lambda_k + \frac{1}{2}\Delta_k} [Z_j(\lambda) - \bar{Z}_j(\lambda_k)] [\gamma(\lambda) - \bar{\gamma}(\lambda_k)] d\lambda \quad (8) \\
&= Z_{jk} \gamma_k + \int_{\lambda_k - \frac{1}{2}\Delta_k}^{\lambda_k + \frac{1}{2}\Delta_k} [Z_j(\lambda) - \bar{Z}_j(\lambda_k)] [\gamma(\lambda) - \bar{\gamma}(\lambda_k)] d\lambda
\end{aligned}$$

When we derived the above equation, we used the following equivalents to simplify the derivation,

$$\int_{\lambda_k - \frac{1}{2}\Delta_k}^{\lambda_k + \frac{1}{2}\Delta_k} [Z_j(\lambda) - \bar{Z}_j(\lambda_k)] d\lambda = \int_{\lambda_k - \frac{1}{2}\Delta_k}^{\lambda_k + \frac{1}{2}\Delta_k} [\gamma(\lambda) - \bar{\gamma}(\lambda_k)] d\lambda = 0 \quad (9)$$

The second term of equation (8) will disappear if either $Z_j(\lambda) = \bar{Z}_j(\lambda_k)$ or $\gamma(\lambda) = \bar{\gamma}(\lambda_k)$ for $\lambda \in \Delta_k$. For the natural bins, all markers within a bin have an identical genotype and thus the first condition applies. The second condition means that all loci within the same bin have the same genetic effect. This condition is out of our control. Further examination of equation (8), we realized that this integration of the product can be interpreted as the covariance between Z_j and γ ,

$$\text{cov}(Z_j, \gamma) = \int_{\lambda_k - \frac{1}{2}\Delta_k}^{\lambda_k + \frac{1}{2}\Delta_k} [Z_j(\lambda) - \bar{Z}_j(\lambda_k)] [\gamma(\lambda) - \bar{\gamma}(\lambda_k)] d\lambda \quad (10)$$

There is no reason to believe that the genetic effect profile (a population parameter) is correlated to the segregation pattern of markers of an individual within the population. Therefore, we may safely ignore this covariance and replace α_{jk} by $Z_{jk} \gamma_k$.