

Chapter 7

Segregation Analysis

Quantitative traits, by definition, are controlled by the segregation of multiple genes. However, the continuous distribution of a quantitative trait does not require the segregation of too many genes. Segregation of just a few genes or even a single gene may be sufficient to generate a continuously distributed phenotype, provided that the environmental variant contributes substantial amount of the trait variation. It is often postulated that a quantitative trait may be controlled by one or a few “major genes” plus multiple modifier genes (genes with very small effects). Such a model is called oligogenic model, which is in contrast to the so called polygenic model where multiple genes with small and equal effects are assumed.

In this chapter, we will discuss a method to test the hypothesis that a quantitative trait is controlled by a single major gene even without observing the genotypes of the major gene. The method is called segregation analysis of quantitative traits. Although segregation analysis belongs to major gene detection, we discuss this topic separately from the previous topic to emphasize a slight difference between segregation analysis and the major gene detection discussed earlier. Here, we define major gene detection as an association study between a single-locus genotype with a quantitative trait where genotypes of the major gene are observed for all individuals. Segregation analysis, however, refers to a single-locus association study where genotypes of the major gene are not observed at all. Another reason for separating major gene detection from segregation analysis is that the statistical method and hypothesis test for segregation analysis can be quite different from those of the major gene detection.

7.1 Gaussian Mixture Distribution

We will use an F_2 population as an example to discuss the segregation analysis. Consider the three genotypes in the following order: A_1A_1 , A_1A_2 , and A_2A_2 . Let $k = 1, 2, 3$ indicate the three ordered genotypes. The means of individuals bearing the three ordered genotypes are denoted by μ_1 , μ_2 , and μ_3 , respectively.

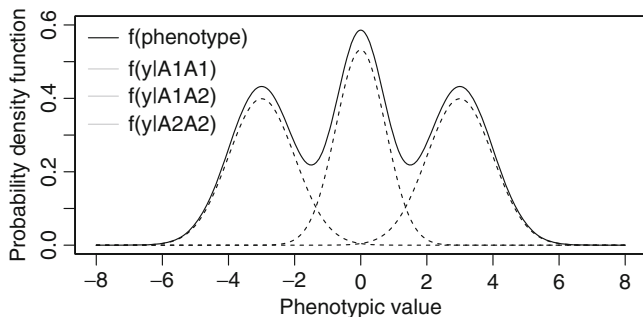


Fig. 7.1 Gaussian mixture with three components. The *solid line* represents the mixture distribution, while the three *dashed lines* represent the three components

Let y_j be the phenotypic value of individual j for $j = 1, \dots, n$, where n is the sample size. Given that individual j has the k th genotype, the linear model for y_j is

$$y_j = \mu_k + \epsilon_j, \quad (7.1)$$

where $\epsilon_j \sim N(0, \sigma^2)$ and σ^2 is the residual error variance. The probability density of y_j conditional on the k th genotype is

$$f_k(y_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(y_j - \mu_k)^2\right]. \quad (7.2)$$

In reality, the genotype of an individual is not observable, and thus, a mixture distribution is needed to describe the probability density of y_j . Let $\pi_k, \forall k = 1, 2, 3$, be the proportion of genotype k (also called the mixing proportion). Without any prior knowledge, π_k may be described by the Mendelian segregation ratio, i.e., $\pi_1 = \pi_3 = \frac{1}{2}\pi_2 = \frac{1}{4}$. Therefore, under the assumption of Mendelian segregation, the π_k 's are constants, not parameters. The distribution of y_j is a mixture of three normal distributions, each is weighted by the Mendelian mixing proportion. The mixture distribution is demonstrated by Fig. 7.1.

The probability density of y_j is

$$f(y_j) = \sum_{k=1}^3 \pi_k f_k(y_j). \quad (7.3)$$

The overall observed log likelihood function for parameters $\theta = \{\mu_1, \mu_2, \mu_3, \sigma^2\}$ is

$$L(\theta) = \sum_{j=1}^n \ln f(y_j) = \sum_{j=1}^n \ln \left[\sum_{k=1}^3 \pi_k f_k(y_j) \right]. \quad (7.4)$$

Any numerical algorithms may be used to estimate the parameters. However, the EM algorithm (Dempster et al. 1977) appears to be the most convenient method for such a mixture model problem and thus will be introduced in this chapter.

7.2 EM Algorithm

The expectation-maximization (EM) algorithm was developed by Dempster et al. (1977) as a special numerical algorithm for finding the maximum likelihood estimates (MLE) of parameters. In contrast to the Newton–Raphson algorithm, the EM algorithm is not a general algorithm for MLE; rather, it can only be applied to some special problems. If the following two conditions hold, then we should consider using the EM algorithm. The first condition is that the maximum likelihood problem can be formulated as a missing value problem. The second condition is that if the missing values were not missing, the MLE would have a closed form solution or, at least, a mathematically attractive form of the solution. We now evaluate the mixture model problem to see whether the two conditions apply.

7.2.1 Closed Form Solution

We introduce a label η_j to indicate the genotype of individual j . The definition of η_j is

$$\eta_j = \begin{cases} 1 & \text{for } A_1A_1 \\ 2 & \text{for } A_1A_2 \\ 3 & \text{for } A_2A_2 \end{cases} \quad (7.5)$$

Since the genotype of an individual is not observable, the label η_j is missing. Therefore, we can formulate the problem as a missing value problem. The missing values are the genotypes of the major gene and denoted by variable η_j for $j = 1, \dots, n$. Therefore, the first condition for using the EM algorithm is met. If η_j is not missing, do we have a closed form solution for the parameters? Let us now define three more variables as functions of η_j . These three variables are called $\delta(\eta_j, 1)$, $\delta(\eta_j, 2)$, and $\delta(\eta_j, 3)$, and their values are defined as

$$\delta(\eta_j, k) = \begin{cases} 1 & \text{if } \eta_j = k \\ 0 & \text{if } \eta_j \neq k \end{cases} \quad (7.6)$$

for $k = 1, 2, 3$. We now use $\delta(\eta_j, k)$ to represent the missing values. If $\delta(\eta_j, k)$ were not missing, the linear model would be described by

$$y_j = \delta(\eta_j, 1)\mu_1 + \delta(\eta_j, 2)\mu_2 + \delta(\eta_j, 3)\mu_3 + \epsilon_j. \quad (7.7)$$

Let us define $\delta_j = [\delta(\eta_j, 1) \delta(\eta_j, 2) \delta(\eta_j, 3)]$ as a 1×3 vector and $\beta = [\mu_1 \mu_2 \mu_3]^T$ as a 3×1 vector. The linear model can be rewritten as

$$y_j = \delta_j \beta + \epsilon_j. \quad (7.8)$$

When $\epsilon_j \sim N(0, \sigma^2)$ is assumed, the maximum likelihood estimates of parameters are

$$\hat{\beta} = \left[\sum_{j=1}^n \delta_j^T \delta_j \right]^{-1} \left[\sum_{j=1}^n \delta_j^T y_j \right] \quad (7.9)$$

for the means and

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n (y_j - \delta_j \hat{\beta})^2 \quad (7.10)$$

for the residual variance. We see that if the missing variables were not missing, the MLE of the parameters do have an attractive closed form solution. Since both requirements of the EM algorithm are met, we can adopt the EM algorithm to search for the MLE of parameters.

7.2.2 EM Steps

Before we derive the EM algorithm, let us show the expectation and maximization steps of the EM algorithm. The E-step involves calculating the expectations of all items containing the missing variables δ_j . The M-step is simply to estimate β and σ^2 using the closed form solutions given above with the items containing the missing variables replaced by the expectations obtained in the E-step, as shown below:

$$\beta = \left[\sum_{j=1}^n E(\delta_j^T \delta_j) \right]^{-1} \left[\sum_{j=1}^n E(\delta_j^T) y_j \right] \quad (7.11)$$

and

$$\sigma^2 = \frac{1}{n} \sum_{j=1}^n E[(y_j - \delta_j \beta)^2]. \quad (7.12)$$

We can see that the EM algorithm is better described by introducing the M-step first and then describing the E-step (in a reverse direction). The detail of the E-step is now given below:

$$E(\delta_j^T \delta_j) = \begin{bmatrix} E[\delta(\eta_j, 1)] & 0 & 0 \\ 0 & E[\delta(\eta_j, 2)] & 0 \\ 0 & 0 & E[\delta(\eta_j, 3)] \end{bmatrix}, \quad (7.13)$$

$$E(\delta_j^T)y_j = \begin{bmatrix} E[\delta(\eta_j, 1)]y_j \\ E[\delta(\eta_j, 2)]y_j \\ E[\delta(\eta_j, 3)]y_j \end{bmatrix} \quad (7.14)$$

and

$$E[(y_j - \delta_j\beta)^2] = \sum_{k=1}^3 E[\delta(\eta_j, k)](y_j - \mu_k)^2. \quad (7.15)$$

Here, we only need to calculate $E[\delta(\eta_j, k)]$, which is the conditional expectation of $\delta(\eta_j, k)$ given the parameter values and the phenotypic value. The full expression of the conditional expectation should be $E[\delta(\eta_j, k)|y_j, \beta, \sigma^2]$, but we use $E[\delta(\eta_j, k)]$ as a short notation.

$$E[\delta(\eta_j, k)] = \frac{\pi_k f_k(y_j|\theta)}{\sum_{k'=1}^3 \pi_{k'} f_{k'}(y_j|\theta)}. \quad (7.16)$$

where $\pi_1 = \pi_3 = \frac{1}{2}\pi_2 = \frac{1}{4}$ is the Mendelian segregation ratio and $f_k(y_j|\theta) = N(y_j|\mu_k, \sigma^2)$ is the normal density. In summary, the EM algorithm is described by

- Initialization: set $t = 0$ and let $\theta = \theta^{(t)}$.
- E-step: calculate $E[\delta(\eta_j, k)|y_j, \theta^{(t)}]$.
- M-step: update $\beta^{(t+1)}$ and $\sigma^{2(t+1)}$.
- Iteration: set $t = t + 1$ and iterate between the E-step and the M-step.

The convergence criterion is

$$\|\theta^{(t+1)} - \theta^{(t)}\| = \sqrt{(\theta^{(t+1)} - \theta^{(t)})'(\theta^{(t+1)} - \theta^{(t)})/\text{dim}(\theta)} \leq \epsilon, \quad (7.17)$$

where $\text{dim}(\theta) = 4$ is the dimension of the parameter vector and ϵ is an arbitrarily small positive number, say 10^{-8} .

Once the three genotypic values are estimated, the additive and dominance effects are estimated using linear contrasts of the genotypic values, e.g.,

$$\begin{cases} \hat{a} = \hat{\beta}_1 - \frac{1}{2}(\hat{\beta}_1 + \hat{\beta}_3) \\ \hat{d} = \hat{\beta}_2 - \frac{1}{2}(\hat{\beta}_1 + \hat{\beta}_3) \end{cases}. \quad (7.18)$$

7.2.3 Derivation of the EM Algorithm

The observed log likelihood function is given in (7.4). The MLE of θ is the (vector) value that maximizes this log likelihood function. The EM algorithm, however, does not directly maximize this likelihood function; instead, it maximizes the expectation

of the complete-data log likelihood function with the expectation taken with respect to the missing variable $\delta(\eta_j, k)$. The complete-data log likelihood function is

$$L_c(\theta) = - \sum_{j=1}^n \left[\frac{1}{2\sigma^2} \sum_{k=1}^3 \delta(\eta_j, k)(y_j - \mu_k)^2 + \sum_{k=1}^3 \delta(\eta_j, k) \ln(\pi_k) \right] - \frac{n}{2} \ln(\sigma^2) \quad (7.19)$$

The expectation of the complete-data log likelihood is $E_{\theta^{(t)}}[L_c(\theta)|y, \theta^{(t)}]$, which is denoted in short by $L(\theta|\theta^{(t)})$ and is defined as

$$L(\theta|\theta^{(t)}) = -\frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^n \sum_{k=1}^3 E[\delta(\eta_j, k)](y_j - \mu_k)^2 + \sum_{j=1}^n \sum_{k=1}^3 E[\delta(\eta_j, k)] \ln(\pi_k) \quad (7.20)$$

With the EM algorithm, the target likelihood function for maximization is neither the complete-data log likelihood function (7.19) nor the observed log likelihood function (7.4); rather, it is the expected complete-data log likelihood function (7.20). An alternative expression of the above equation is

$$L(\theta|\theta^{(t)}) = -\frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^n E[(y_j - \delta_j \beta)^2] + \sum_{j=1}^n \sum_{k=1}^3 E[\delta(\eta_j, k)] \ln(\pi_k). \quad (7.21)$$

The partial derivatives of $L(\theta|\theta^{(t)})$ with respect to β and σ^2 are

$$\frac{\partial}{\partial \beta} L(\theta|\theta^{(t)}) = \frac{1}{\sigma^2} E(\delta_j^T) y_j - \frac{1}{\sigma^2} \sum_{j=1}^n E(\delta_j^T \delta_j) \beta \quad (7.22)$$

and

$$\frac{\partial}{\partial \sigma^2} L(\theta|\theta^{(t)}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{j=1}^n E[(y_j - \delta_j^T \beta)^2], \quad (7.23)$$

respectively. Setting $\frac{\partial}{\partial \beta} L(\theta|\theta^{(t)}) = \frac{\partial}{\partial \sigma^2} L(\theta|\theta^{(t)}) = 0$, we get

$$\beta = \left[\sum_{j=1}^n E(\delta_j^T \delta_j) \right]^{-1} \left[\sum_{j=1}^n E(\delta_j^T) y_j \right] \quad (7.24)$$

and

$$\sigma^2 = \frac{1}{n} \sum_{j=1}^n E[(y_j - \delta_j \beta)^2]. \quad (7.25)$$

This concludes the derivation of the EM algorithm.

7.2.4 Proof of the EM Algorithm

The target likelihood function for maximization in the EM algorithm is the expectation of the complete-data log likelihood function. However, the actual MLE of θ is obtained by maximization of the observed log likelihood function. To prove that the EM solution of the parameters is indeed the MLE, we only need to show that the partial derivative of the expected complete-data likelihood is identical to the partial derivative of the observed log likelihood, i.e., $\frac{\partial}{\partial \theta} L(\theta | \theta^{(t)}) = \frac{\partial}{\partial \theta} L(\theta)$. If the two partial derivatives are the same, then the solutions must be the same because they both solve the same equation system, i.e., $\frac{\partial}{\partial \theta} L(\theta) = 0$.

Recall that the partial derivative of the expected complete-data log likelihood function with respect to β is

$$\frac{\partial}{\partial \beta} L(\theta | \theta^{(t)}) = \frac{1}{\sigma^2} E(\delta_j^T) y_j - \frac{1}{\sigma^2} E(\delta_j^T \delta_j) \beta, \quad (7.26)$$

which is a 3×1 vector as shown below:

$$\frac{\partial}{\partial \beta} L(\theta | \theta^{(t)}) = \left[\frac{\partial}{\partial \mu_1} L(\theta | \theta^{(t)}) \quad \frac{\partial}{\partial \mu_2} L(\theta | \theta^{(t)}) \quad \frac{\partial}{\partial \mu_3} L(\theta | \theta^{(t)}) \right]^T.$$

The k th component of this vector is

$$\begin{aligned} \frac{\partial}{\partial \mu_k} L(\theta | \theta^{(t)}) &= \frac{1}{\sigma^2} E[\delta(\eta_j, k)] y_j - \frac{1}{\sigma^2} E[\delta^2(\eta_j, k)] \mu_k \\ &= \frac{1}{\sigma^2} E[\delta(\eta_j, k)] y_j - \frac{1}{\sigma^2} E[\delta(\eta_j, k)] \mu_k \\ &= \frac{1}{\sigma^2} E[\delta(\eta_j, k)] (y_j - \mu_k) \end{aligned} \quad (7.27)$$

The equation holds because $E[\delta(\eta_j, k)] = E[\delta^2(\eta_j, k)]$, a property for the Bernoulli distribution. We now evaluate the partial derivative of the expected complete-data log likelihood with respect to σ^2 ,

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} L(\theta | \theta^{(t)}) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2} \sum_{j=1}^n E[(y_j - \delta_j^T \beta)^2] \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2} \sum_{j=1}^n E[\delta(\eta_j, k)] (y_j - \mu_k)^2 \end{aligned} \quad (7.28)$$

We now look at the partial derivatives of $L(\theta)$ with respect to the parameters. The observed log likelihood function is

$$L(\theta) = \sum_{j=1}^n \ln \sum_{k=1}^3 \pi_k f_k(y_j) \quad (7.29)$$

where

$$f_k(y_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(y_j - \mu_k)^2\right]. \quad (7.30)$$

The partial derivatives of $L(\theta)$ with respect to $\beta = [\mu_1 \mu_2 \mu_3]^T$ are

$$\frac{\partial}{\partial \mu_k} L(\theta) = \sum_{j=1}^n \frac{\pi_k}{\sum_{k'=1}^3 \pi_{k'} f_{k'}(y_j)} \frac{\partial}{\partial \mu_k} f_k(y_j), \quad (7.31)$$

where

$$\frac{\partial}{\partial \mu_k} f_k(y_j) = f_k(y_j) \left[\frac{1}{\sigma^2}(y_j - \mu_k) \right]. \quad (7.32)$$

Hence,

$$\frac{\partial}{\partial \mu_k} L(\theta) = \frac{1}{\sigma^2} \sum_{j=1}^n \frac{\pi_k f_k(y_j)}{\sum_{k'=1}^3 \pi_{k'} f_{k'}(y_j)} (y_j - \mu_k). \quad (7.33)$$

Recall that

$$E[\delta(\eta_j, k)] = \frac{\pi_k f_k(y_j)}{\sum_{k'=1}^3 \pi_{k'} f_{k'}(y_j)}. \quad (7.34)$$

Therefore,

$$\frac{\partial}{\partial \mu_k} L(\theta) = \frac{1}{\sigma^2} \sum_{j=1}^n E[\delta(\eta_j, k)](y_j - \mu_k), \quad (7.35)$$

which is exactly the same as $\frac{\partial}{\partial \mu_k} L(\theta|\theta^{(t)})$ given in (7.27). Now, let us look at the partial derivative of $L(\theta)$ with respect to σ^2 .

$$\frac{\partial}{\partial \sigma^2} L(\theta) = \sum_{j=1}^n \sum_{k=1}^3 \frac{\pi_k}{\sum_{k'=1}^3 \pi_{k'} f_{k'}(y_j)} \frac{\partial}{\partial \sigma^2} f_k(y_j), \quad (7.36)$$

where

$$\frac{\partial}{\partial \sigma^2} f_k(y_j) = -\frac{1}{2\sigma^2} f_k(y_j) + \left[\frac{1}{2\sigma^4}(y_j - \mu_k)^2 \right] f_k(y_j). \quad (7.37)$$

Hence,

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} L(\theta) = & -\frac{1}{2\sigma^2} \sum_{j=1}^n \sum_{k=1}^3 \frac{\pi_k f_k(y_j)}{\sum_{k'=1}^3 \pi_{k'} f_{k'}(y_j)} \\ & + \frac{1}{2\sigma^4} \sum_{j=1}^n \left[\sum_{k=1}^3 \frac{\pi_k f_k(y_j)}{\sum_{k'=1}^3 \pi_{k'} f_{k'}(y_j)} (y_j - \mu_k)^2 \right] \end{aligned} \quad (7.38)$$

Note that

$$\sum_{k=1}^3 \frac{\pi_k f_k(y_j)}{\sum_{k'=1}^3 \pi_{k'} f_{k'}(y_j)} = \sum_{k=1}^3 E[\delta(\eta_j, k)] = 1. \quad (7.39)$$

Therefore,

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} L(\theta) = & -\frac{1}{2\sigma^2} \sum_{j=1}^n \sum_{k=1}^3 E[\delta(\eta_j, k)] + \frac{1}{2\sigma^4} \sum_{j=1}^n \sum_{k=1}^3 E[\delta(\eta_j, k)] (y_j - \mu_k)^2 \\ = & -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{j=1}^n \sum_{k=1}^3 E[\delta(\eta_j, k)] (y_j - \mu_k)^2 \end{aligned} \quad (7.40)$$

which is exactly the same as $\frac{\partial}{\partial \sigma^2} L(\theta|\theta^{(t)})$ given in (7.28). We now have confirmed that

$$\frac{\partial}{\partial \sigma^2} L(\theta|\theta^{(t)}) = \frac{\partial}{\partial \sigma^2} L(\theta) \quad (7.41)$$

and

$$\frac{\partial}{\partial \mu_k} L(\theta|\theta^{(t)}) = \frac{\partial}{\partial \mu_k} L(\theta), \quad \forall k = 1, 2, 3. \quad (7.42)$$

This concludes the proof that the EM algorithm does lead to the MLE of the parameters.

7.3 Hypothesis Tests

The overall null hypothesis is “no major gene is segregating” denoted by

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu. \quad (7.43)$$

The alternative hypothesis is “at least one of the means is different from others,” denoted by

$$H_1 : \mu_1 \neq \mu_3 \text{ or } \mu_2 \neq \frac{1}{2}(\mu_1 + \mu_3). \quad (7.44)$$

The likelihood ratio test statistic is

$$\lambda = -2[L_0(\hat{\theta}) - L_1(\hat{\theta})], \quad (7.45)$$

where $L_1(\hat{\theta})$ is the observed log likelihood function evaluated at the MLE of θ for the full model, and

$$L_0(\hat{\theta}) = -\frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \sum_{j=1}^n (y_j - \hat{\mu})^2 \quad (7.46)$$

is the log likelihood values evaluated at the null model where

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^n y_j \quad (7.47)$$

and

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{\mu})^2. \quad (7.48)$$

Under the null hypothesis, λ will follow approximately a chi-square distribution with two degrees of freedom. Therefore, H_0 will be rejected if $\lambda > \chi_{2,1-\alpha}^2$, where $\alpha = 0.05$ may be chosen as the type I error.

7.4 Variances of Estimated Parameters

Unlike other iterative methods of parameter estimation, e.g., Newton–Raphson method, that variance–covariance matrix of the estimated parameters are provided automatically as a by-product of the iteration process, the EM algorithm does not facilitate an easy way for calculating the variance–covariance matrix of the estimated parameters. We now introduce a special method for calculating the variance–covariance matrix. The method was developed by Louis (1982) particularly for calculating the variance–covariance matrix of parameters that are estimated via the EM algorithm. The method requires the first and second partial derivatives of the complete-data log likelihood function (not the observed log likelihood function). The complete-data log likelihood function is

$$L(\theta, \delta) = \sum_{j=1}^n L_j(\theta, \delta), \quad (7.49)$$

where

$$L_j(\theta, \delta) = -\frac{1}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (y - \delta_j \beta)^2. \quad (7.50)$$

The first partial derivative of this log likelihood with respect to the parameter is called the score function, which is

$$S(\theta, \delta) = \frac{\partial}{\partial \theta} L(\theta, \delta) = \sum_{j=1}^n \frac{\partial}{\partial \theta} L_j(\theta, \delta) = \sum_{j=1}^n S_j(\theta, \delta), \quad (7.51)$$

where

$$S_j(\theta, \delta) = \frac{\partial}{\partial \theta} L_j(\theta, \delta) = \begin{bmatrix} \frac{\partial}{\partial \beta} L_j(\theta, \delta) \\ \frac{\partial}{\partial \sigma^2} L_j(\theta, \delta) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} \delta_j^T (y_j - \delta_j \beta) \\ -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} (y_j - \delta_j \beta)^2 \end{bmatrix}. \quad (7.52)$$

The second partial derivative is called the Hessian matrix $H(\theta, \delta)$. The negative value of the Hessian matrix is denoted by $B(\theta, \delta) = -H(\theta, \delta)$,

$$B(\theta, \delta) = -\frac{\partial^2 L(\theta, \delta)}{\partial \theta \partial \theta^T} = -\sum_{j=1}^n \frac{\partial^2 L_j(\theta, \delta)}{\partial \theta \partial \theta^T} = \sum_{j=1}^n B_j(\theta, \delta), \quad (7.53)$$

where

$$B_j(\theta, \delta) = -\frac{\partial^2 L_j(\theta, \delta)}{\partial \theta \partial \theta^T} = \begin{bmatrix} -\frac{\partial^2 L_j(\theta, \delta)}{\partial \beta \partial \beta^T} & -\frac{\partial^2 L_j(\theta, \delta)}{\partial \beta \partial \sigma^2} \\ -\frac{\partial^2 L_j(\theta, \delta)}{\partial \sigma^2 \partial \beta^T} & -\frac{\partial^2 L_j(\theta, \delta)}{\partial \sigma^2 \partial \sigma^2} \end{bmatrix} \quad (7.54)$$

Detailed expression of $B_j(\theta, \delta)$ is given below:

$$B_j(\theta, \delta) = \begin{bmatrix} \frac{1}{\sigma^2} \delta_j^T \delta_j & \frac{1}{\sigma^4} \delta_j^T (y_j - \delta_j \beta) \\ \frac{1}{\sigma^4} (y_j - \delta_j \beta)^T \delta_j & \frac{1}{\sigma^6} (y_j - \delta_j \beta)^2 - \frac{1}{2\sigma^2} \end{bmatrix}. \quad (7.55)$$

Louis (1982) gave the following information matrix:

$$\begin{aligned} I(\theta) &= \mathbb{E}[B(\theta, \delta)] - \text{var}[S(\theta, \delta)] \\ &= \sum_{j=1}^n \mathbb{E}[B_j(\theta, \delta)] - \sum_{j=1}^n \text{var}[S_j(\theta, \delta)], \end{aligned} \quad (7.56)$$

where the expectation and variance are taken with respect to the missing variable δ_j using the posterior probability of δ_j . Detailed expressions of $\mathbb{E}[B_j(\theta, \delta)]$ and $\text{var}[S_j(\theta, \delta)]$ are given in the end of this section. Readers may also refer to Han and Xu (2008) and Xu and Hu (2010) for the derivation and the results. Replacing θ by $\hat{\theta}$ and taking the inverse of the information matrix, we get the variance–covariance matrix of the estimated parameters,

$$\text{var}(\hat{\theta}) = I^{-1}(\hat{\theta}) = \{\mathbb{E}[B(\hat{\theta}, \delta)] - \text{var}[S(\hat{\theta}, \delta)]\}^{-1}. \quad (7.57)$$

This is a 4×4 variance–covariance matrix, as shown below:

$$\text{var}(\hat{\theta}) = \begin{bmatrix} \text{var}(\hat{\beta}) & \text{cov}(\hat{\beta}, \hat{\sigma}^2) \\ \text{cov}(\hat{\sigma}^2, \hat{\beta}^T) & \text{var}(\hat{\sigma}^2) \end{bmatrix}, \quad (7.58)$$

where $\text{var}(\hat{\beta})$ is a 3×3 variance matrix for the estimated genotypic values.

The additive and dominance effects can be expressed as linear functions of β , as demonstrated below:

$$\begin{bmatrix} a \\ d \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = L^T \beta, \quad (7.59)$$

where

$$L = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \end{bmatrix}^T. \quad (7.60)$$

The variance–covariance matrix for the estimated major gene effects is

$$\text{var} \begin{bmatrix} \hat{a} \\ \hat{d} \end{bmatrix} = L^T \text{var}(\hat{\beta}) L = \begin{bmatrix} \text{var}(\hat{a}) & \text{cov}(\hat{a}, \hat{d}) \\ \text{cov}(\hat{a}, \hat{d}) & \text{var}(\hat{d}) \end{bmatrix}. \quad (7.61)$$

The variance–covariance matrix of the estimated major gene effects also facilitates an alternative method for testing the hypothesis of $H_0 : a = d = 0$. This test is called the Wald-test statistic (Wald 1943),

$$W = \beta^T L [L^T \text{var}(\hat{\beta}) L]^{-1} L^T \beta = \begin{bmatrix} \hat{a} & \hat{d} \end{bmatrix} \begin{bmatrix} \text{var}(\hat{a}) & \text{cov}(\hat{a}, \hat{d}) \\ \text{cov}(\hat{a}, \hat{d}) & \text{var}(\hat{d}) \end{bmatrix}^{-1} \begin{bmatrix} \hat{a} \\ \hat{d} \end{bmatrix}. \quad (7.62)$$

The Wald-test statistic is much like the likelihood ratio test statistic. Under the null model, W follows approximately a χ^2 distribution with 2 degrees of freedom. However, Wald test is usually considered inferior compared to the likelihood ratio test statistic, especially when the sample size is small.

Before exiting this section, we now provide the derivation of $\text{E}[B_j(\theta, \delta)]$ and $\text{var}[S_j(\theta, \delta)]$. Recall that δ_j is a 1×3 multinomial variable with sample size 1 and defined as

$$\delta_j = [\delta(\eta_j, 1) \delta(\eta_j, 2) \delta(\eta_j, 3)] \quad (7.63)$$

This variable has the following properties:

$$\delta^2(\eta_j, k) = \delta(\eta_j, k) \quad (7.64)$$

and

$$\delta(\eta_j, k)\delta(\eta_j, k') = \begin{cases} \delta(\eta_j, k) & \text{for } k = k' \\ 0 & \text{for } k \neq k' \end{cases} \quad (7.65)$$

Therefore, the expectation of δ_j is

$$E(\delta_j) = [E[\delta(\eta_j, 1)] \ E[\delta(\eta_j, 2)] \ E[\delta(\eta_j, 3)]] \quad (7.66)$$

The expectation of its quadratic form is

$$E(\delta_j^T \delta_j) = \text{diag}[E(\delta_j)] = \begin{bmatrix} E[\delta(\eta_j, 1)] & 0 & 0 \\ 0 & E[\delta(\eta_j, 3)] & 0 \\ 0 & 0 & E[\delta(\eta_j, 3)] \end{bmatrix} \quad (7.67)$$

The variance–covariance matrix of δ_j is

$$\text{var}(\delta_j) = E(\delta_j^T \delta_j) - E(\delta_j)E^T(\delta_j) \quad (7.68)$$

To derive the observed information matrix, we need the first and second partial derivatives of the complete-data log likelihood with respect to the parameter vector $\theta = [\beta^T \ \sigma^2]^T$. The score vector is rewritten as

$$S_j(\theta, \delta) = \begin{bmatrix} \frac{1}{\sigma^2} \delta_j^T (y_j - \delta_j \beta) \\ \frac{1}{2\sigma^4} (y_j - \delta_j \beta)^2 \end{bmatrix} + \begin{bmatrix} 0_{3 \times 1} \\ -\frac{1}{2\sigma^2} \end{bmatrix} \quad (7.69)$$

where $0_{3 \times 1}$ is a 3×1 vector of zeros, and thus, the score is a 4×1 vector. The negative of the second partial derivative is

$$B_j(\theta, \delta) = \begin{bmatrix} \frac{1}{\sigma^2} \delta_j^T \delta_j & \frac{1}{\sigma^4} \delta_j^T (y_j - \delta_j \beta) \\ \frac{1}{\sigma^4} (y_j - \delta_j \beta)^T \delta_j & \frac{1}{\sigma^6} (y_j - \delta_j \beta)^2 \end{bmatrix} + \begin{bmatrix} 0_{3 \times 3} & 0_{3 \times 1} \\ 0_{1 \times 3} & -\frac{1}{2\sigma^2} \end{bmatrix} \quad (7.70)$$

where $0_{3 \times 3}$ is a 3×3 matrix of zeros, and thus, $B_j(\theta, \delta)$ is a 4×4 matrix. The expectation of $B_j(\theta, \delta)$ is easy to derive, but derivation of the variance–covariance matrix of the score vector is very difficult. Xu and Xu (2003) used a Monte Carlo approach to approximating the expectation and the variance–covariance matrix. They simulated multiple (e.g., 5,000) samples of δ_j from the posterior distribution and then took the sample mean of $B_j(\theta, \delta)$ and the sample variance–covariance matrix of $S_j(\theta, \delta)$ as the approximations of the corresponding terms. Here, we took a theoretical approach for the derivation and provide explicit expressions for the expectation and variance–covariance matrix. We can express the score vector as a linear function of δ_j and the $B_j(\theta, \delta)$ matrix as a quadratic function of δ_j . By trial and error, we found that

$$\begin{aligned}
S_j(\theta, \delta) &= \begin{bmatrix} \frac{1}{\sigma^2}(y_j - \beta_1) & 0 & 0 \\ 0 & \frac{1}{\sigma^2}(y_j - \beta_1) & 0 \\ 0 & 0 & \frac{1}{\sigma^2}(y_j - \beta_3) \\ \frac{1}{2\sigma^4}(y_j - \beta_1)^2 & \frac{1}{2\sigma^4}(y_j - \beta_2)^2 & \frac{1}{2\sigma^4}(y_j - \beta_3)^2 \end{bmatrix} \begin{bmatrix} \delta(\eta_j, 1) \\ \delta(\eta_j, 2) \\ \delta(\eta_j, 3) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ -\frac{1}{2\sigma^2} \end{bmatrix} \\
&= A_j^T \delta_j^T + C \tag{7.71}
\end{aligned}$$

where A_j^T is the 4×3 coefficient matrix and C is the 4×1 vector of constants. Let us define a 4×1 matrix H_j^T as

$$H_j^T = T_j^T \delta_j^T = \begin{bmatrix} \frac{1}{\sigma^2} & 0 & 0 \\ 0 & \frac{1}{\sigma^2} & 0 \\ 0 & 0 & \frac{1}{\sigma^2} \\ \frac{1}{\sigma^3}(y_j - \beta_1) & \frac{1}{\sigma^3}(y_j - \beta_2) & \frac{1}{\sigma^3}(y_j - \beta_3) \end{bmatrix} \begin{bmatrix} \delta_j(1) \\ \delta_j(2) \\ \delta_j(3) \end{bmatrix} \tag{7.72}$$

where T_j^T is the 4×3 coefficient matrix. We can now express matrix $B_j(\theta, \delta)$ as

$$B_j(\theta, \delta) = H_j^T H_j + D = T_j^T \delta_j^T \delta_j T_j + D \tag{7.73}$$

where

$$D = \text{diag}(C) = \begin{bmatrix} 0_{3 \times 3} & 0_{3 \times 1} \\ 0_{1 \times 3} & -\frac{1}{2\sigma^2} \end{bmatrix} \tag{7.74}$$

is a 4×4 constant matrix. The expectation of $B_j(\theta, \delta)$ is

$$E[B_j(\theta, \delta)] = T_j^T E(\delta_j^T \delta_j) T_j + D \tag{7.75}$$

The expectation vector and the variance-covariance matrix of $S_j(\theta, \delta)$ are

$$E[S_j(\theta, \delta)] = A_j^T E(\delta_j^T) + C \tag{7.76}$$

and

$$\text{var}[S_j(\theta, \delta)] = A_j^T \text{var}(\delta_j) A_j = A_j^T [E(\delta_j^T \delta_j) - E(\delta_j^T) E(\delta_j)] A_j \tag{7.77}$$

respectively. Expressing $S_j(\theta, \delta)$ and $B_j(\theta, \delta)$ as linear and quadratic functions of the missing vector δ_j has significantly simplified the derivation of the information matrix.

7.5 Estimation of the Mixing Proportions

We used an F_2 population as an example for segregation analysis. Extension of the segregation analysis to other populations is straightforward and will not be discussed here. For the F_2 population, we assumed that the major gene follows the Mendelian segregation ratio, i.e., $\pi_1 = \pi_3 = \frac{1}{2}\pi_2 = \frac{1}{4}$. Therefore, π_k is a constant, not a parameter for estimation. The method can be extended to a situation where the major gene does not follow the Mendelian segregation ratio. In this case, the values of π_k are also parameters for estimation. This section will introduce a method to estimate the π_k 's. These π_k 's are called the mixing proportions.

We simply add one more step in the EM algorithm to estimate π_k , $\forall k = 1, 2, 3$. Again, we maximize the expected complete-data log likelihood function. To enforce the restriction that $\sum_{k=1}^3 \pi_k = 1$, we introduce a Lagrange multiplier λ . Therefore, the actual function to be maximized is

$$L(\theta|\theta^{(t)}) = -\frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^n \sum_{k=1}^3 E[\delta(\eta_j, k)](y_j - \mu_k)^2 + \sum_{j=1}^n \sum_{k=1}^3 E[\delta(\eta_j, k)] \ln(\pi_k) + \lambda \left(1 - \sum_{k=1}^3 \pi_k \right). \quad (7.78)$$

The partial derivatives of $L(\theta|\theta^{(t)})$ with respect to π_k and λ are

$$\frac{\partial}{\partial \pi_k} L(\theta|\theta^{(t)}) = \frac{1}{\pi_k} \sum_{j=1}^n E[\delta(\eta_j, k)] - \lambda, \quad \forall k = 1, 2, 3 \quad (7.79)$$

and

$$\frac{\partial}{\partial \lambda} L(\theta|\theta^{(t)}) = 1 - \sum_{k=1}^3 \pi_k, \quad (7.80)$$

respectively. Let $\frac{\partial}{\partial \pi_k} L(\theta|\theta^{(t)}) = \frac{\partial}{\partial \lambda} L(\theta|\theta^{(t)}) = 0$, and solve for π_k 's and λ . The solution for π_k is

$$\pi_k = \frac{1}{\lambda} \sum_{j=1}^n E[\delta(\eta_j, k)], \quad \forall k = 1, 2, 3. \quad (7.81)$$

The solution for λ is obtained by

$$\sum_{k=1}^3 \pi_k = \frac{1}{\lambda} \sum_{k=1}^3 \sum_{j=1}^n E[\delta(\eta_j, k)] = \frac{1}{\lambda} \sum_{j=1}^n \sum_{k=1}^3 E[\delta(\eta_j, k)] = \frac{n}{\lambda} = 1. \quad (7.82)$$

This is because $\sum_{k=1}^3 E[\delta(\eta_j, k)] = 1$ and $\sum_{j=1}^n = n$. As a result, $\lambda = n$ and thus

$$\pi_k^{(t+1)} = \frac{1}{n} \sum_{j=1}^n E[\delta(\eta_j, k)], \quad \forall k = 1, 2, 3. \quad (7.83)$$