Chapter 9
Interval Mapping

Interval mapping is an extension of the individual marker analysis so that two markers are analyzed at a time. In the marker analysis (Chap. 8), we cannot estimate the exact position of a QTL. With interval mapping, we use two markers to determine an interval, within which a putative QTL position is proposed. The genotype of the putative QTL is not observable but can be inferred with a certain probability using the three-point or multipoint method introduced in Chap. 4. Once the genotype of the QTL is inferred, we can estimate and test the QTL effect at that particular position. We divide the interval into many putative positions of QTL with one or two cM apart and investigate every putative position within the interval. Once we have searched the current interval, we move on to the next interval and so on until all intervals have been searched. The putative QTL position (not necessarily at a marker) that has the maximum test statistical value is the estimated QTL position. Figure 9.1 demonstrates the process of genome scanning for markers only (panel a), for markers and virtual markers (panel b), and for every point of the chromosome (panel c).

Interval mapping was originally developed by Lander and Botstein (1989) and further modified by numerous authors. Interval mapping has revolutionized genetic mapping because we can really pinpoint the exact location of a QTL. In each of the four sections that follow, we will introduce one specific statistical method of interval mapping based on the F_2 design. Methods of interval mapping for a BC design are straightforward and thus will not be discussed in this chapter. Maximum likelihood (ML) method of interval mapping (Lander and Botstein 1989) is the optimal method for interval mapping. Least-squares (LS) method (Haley and Knott 1992) is a simplified approximation of Lander and Botstein method. The iteratively reweighted least-squares (IRLS) method (Xu 1998a,b) is a further improved method over the least-squares method. Recently Feenstra et al. (2006) developed an estimating equation (EE) method for QTL mapping, which is an extension of the IRLS with improved performance. Han and Xu (2008) developed a Fisher scoring algorithm (FISHER) for QTL mapping. Both the EE and FISHER algorithms maximize the same likelihood function, and thus, they generate identical result. In this chapter, we introduce the methods based on their simplicity rather than their chronological

S. Xu, *Principles of Statistical Genomics*, DOI 10.1007/978-0-387-70807-2_9,

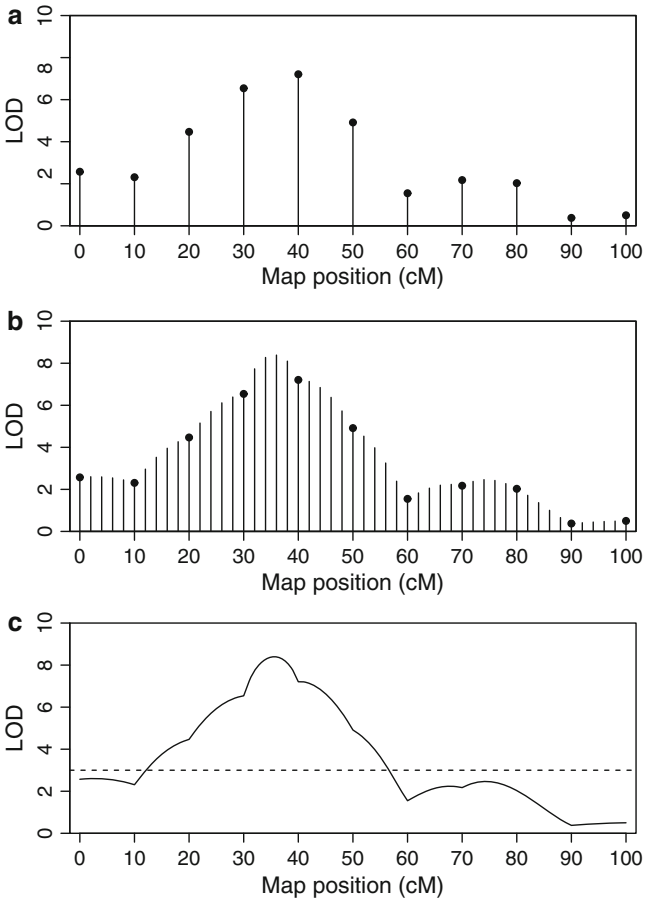

Fig. 9.1 The LOD test statistics for (**a**) marker effects (*top panel*), (**b**) virtual marker effects (panel in the *middle*), and (**c**) every point of a simulated chromosome (*bottom panel*)

orders of development. Therefore, the methods will be introduced in the following order: LS, IRLS, FISHER, and ML. Bayesian method will be discussed in a later chapter where multiple QTL mapping is addressed.

9.1 Least-Squares Method

The LS method was introduced by Haley and Knott (1992) aiming to improving the computational speed. The statistical model for the phenotypic value of the jth individual is

$$y_j = X_j\beta + Z_j\gamma + \varepsilon_j \tag{9.1}$$

where β is a $p \times 1$ vector for some model effects that are irrelevant to QTL effects, X_j is a $1 \times p$ known design vector, $\gamma = \{a, d\}$ is a 2×1 vector for QTL effects of a putative locus (a for additive effect and d for dominance effect), and Z_j is a 1×2 vector for the genotype indicator variable defined as

$$Z_j = \begin{cases} H_1 & \text{for } A_1A_1 \\ H_2 & \text{for } A_1A_2 \\ H_3 & \text{for } A_2A_2 \end{cases} \tag{9.2}$$

where H_k for $k = 1, 2, 3$ is the kth row of matrix

$$H = \begin{bmatrix} +1 & 0 \\ 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{9.3}$$

The residual error ε_j is assumed to be a $N(0, \sigma^2)$ variable. Although normal distribution for ε_j is not a required assumption for the LS method, it is required for the ML method. It is important to include non-QTL effects β in the model to control the residual error variance as small as possible. For example, location and year effects are common in replicated experiments. These effects are not related to QTL but will contribute to the residual error if not included in the model. If there is no such a non-QTL effect to consider in a nice designed experiment, β will be a single parameter (intercept) and X_j will be unity across all $j = 1, \ldots, n$.

With interval mapping, the QTL genotype is never known unless the putative QTL position overlaps with a fully informative marker. Therefore, Haley and Knott (1992) suggested to replace the unknown Z_j by the expectation of Z_j conditional on flanking marker genotype. Let $p_j(1)$, $p_j(0)$, and $p_j(-1)$ be the conditional probabilities for the three genotypes given flanking marker information (see Chap. 4 for the method of calculating conditional probability). The LS model of Haley and Knott (1992) is

$$y_j = X_j\beta + U_j\gamma + e_j \tag{9.4}$$

where

$$U_j = E(Z_j) = p_j(+1)H_1 + p_j(0)H_2 + p_j(-1)H_3 \tag{9.5}$$

is the conditional expectation of Z_j. The residual error e_j (different from ε_j) remains normal with mean zero and variance σ^2, although this assumption has been violated (see next section). The least-squares estimate of β and γ is

$$\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} X_j^T X_j & \sum_{j=1}^{n} X_j^T U_j \\ \sum_{j=1}^{n} U_j^T X_j & \sum_{j=1}^{n} U_j^T U_j \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{n} X_j^T y_j \\ \sum_{j=1}^{n} U_j^T y_j \end{bmatrix} \tag{9.6}$$

and the estimated residual error variance is

$$\hat{\sigma}^2 = \frac{1}{n-p-2}\sum_{j=1}^{n}(y_j - X_j\hat{\beta} - U_j\hat{\gamma})^2 \tag{9.7}$$

The variance–covariance matrix of the estimated parameters is

$$\text{var}\begin{bmatrix}\hat{\beta}\\ \hat{\gamma}\end{bmatrix} = \begin{bmatrix}\sum_{j=1}^{n} X_j^T X_j & \sum_{j=1}^{n} X_j^T U_j\\ \sum_{j=1}^{n} U_j^T X_j & \sum_{j=1}^{n} U_j^T U_j\end{bmatrix}^{-1}\hat{\sigma}^2 \tag{9.8}$$

which is a $(p+2)\times(p+2)$ matrix. Let

$$\text{var}(\hat{\gamma}) = V = \begin{bmatrix}\text{var}(\hat{a}) & \text{cov}(\hat{a},\hat{d})\\ \text{cov}(\hat{a},\hat{d}) & \text{var}(\hat{d})\end{bmatrix} \tag{9.9}$$

be the 2×2 lower diagonal bock of matrix (9.8). The standard errors of the estimated additive and dominance effects are the square roots of the diagonal elements of matrix (9.9).

We can use either the F-test or the W-test statistic to test the hypothesis of H_0 : $\gamma = 0$. The W-test statistic is

$$W = \hat{\gamma}^T V^{-1}\hat{\gamma} = \begin{bmatrix}\hat{a} & \hat{d}\end{bmatrix}\begin{bmatrix}\text{var}(\hat{a}) & \text{cov}(\hat{a},\hat{d})\\ \text{cov}(\hat{a},\hat{d}) & \text{var}(\hat{d})\end{bmatrix}^{-1}\begin{bmatrix}\hat{a}\\ \hat{d}\end{bmatrix} \tag{9.10}$$

The likelihood ratio test statistic can also be applied if we assume that $e_j \sim N(0,\sigma^2)$ for all $j = 1,\ldots,n$. The log likelihood function for the full model is

$$\begin{aligned} L_1 &= -\frac{n}{2}\ln(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2}\sum_{j=1}^{n}(y - X_j\hat{\beta} - U_j\hat{\gamma})^2\\ &\approx -\frac{n}{2}\left[\ln(\hat{\sigma}^2) + 1\right] \end{aligned} \tag{9.11}$$

The reduced model under $H_0 : \gamma = 0$ is

$$\begin{aligned} L_0 &= -\frac{n}{2}\ln(\hat{\hat{\sigma}}^2) - \frac{1}{2\hat{\hat{\sigma}}^2}\sum_{j=1}^{n}(y - X_j\hat{\hat{\beta}})^2\\ &\approx -\frac{n}{2}\left[\ln(\hat{\hat{\sigma}}^2) + 1\right] \end{aligned} \tag{9.12}$$

where

$$\hat{\hat{\beta}} = \left[\sum_{j=1}^{n} X_j^T X_j\right]^{-1} \left[\sum_{j=1}^{n} X_j^T y_j\right] \tag{9.13}$$

and

$$\hat{\hat{\sigma}}^2 = \frac{1}{n-p}\sum_{j=1}^{n}(y_j - X_j\hat{\hat{\beta}})^2 \tag{9.14}$$

The likelihood ratio test statistic is

$$\lambda = -2(L_0 - L_1) \tag{9.15}$$

9.2 Weighted Least Squares

Xu (1995) realized that the LS method is flawed because the residual variance is heterogeneous after replacing X_j by its conditional expectation U_j. The conditional variance of X_j given marker information varies from one individual to another, and it will contribute to the residual variance. Xu (1998a,b) modified the exact model

$$y_j = X_j\beta + Z_j\gamma + \varepsilon_j \tag{9.16}$$

by

$$y_j = X_j\beta + U_j\gamma + (Z_j - U_j)\gamma + \varepsilon_j \tag{9.17}$$

which differs from the Haley and Knott's (1992) model by $(Z_j - U_j)\gamma$. Since Z_j is not observable, this additional term is merged into the residual error if ignored. Let

$$e_j = (Z_j - U_j)\gamma + \varepsilon_j \tag{9.18}$$

be the new residual error. The Haley and Knott's (1992) model can be rewritten as

$$y_j = X_j\beta + U_j\gamma + e_j \tag{9.19}$$

Although we assume $\varepsilon_j \sim N(0, \sigma^2)$, this does not validate the normal assumption of e_j. The expectation for e_j is

$$E(e_j) = [E(Z_j) - U_j]\gamma + E(\varepsilon_j) = 0 \tag{9.20}$$

The variance of e_j is

$$\text{var}(e_j) = \sigma_j^2 = \gamma^T \text{var}(Z_j)\gamma + \sigma^2 = \left(\frac{1}{\sigma^2}\gamma^T \Sigma_j \gamma + 1\right)\sigma^2 \tag{9.21}$$

where $\Sigma_j = \text{var}(Z_j)$, which is defined as a conditional variance–covariance matrix given flanking marker information. The explicit forms of Σ_j are

$$\Sigma_j = E(Z_j^T Z_j) - E(Z_j^T)E(Z_j), \tag{9.22}$$

where

$$E(Z_j^T Z_j) = p_j(1)H_1^T H_1 + p_j(0)H_2^T H_2 + p_j(-1)H_3^T H_3 \tag{9.23}$$

and

$$E(Z_j) = U_j = p_j(1)H_1 + p_j(0)H_2 + p_j(-1)H_3. \tag{9.24}$$

Let

$$\sigma_j^2 = \left(\frac{1}{\sigma^2}\gamma^T \Sigma_j \gamma + 1\right)\sigma^2 = \frac{1}{W_j}\sigma^2 \tag{9.25}$$

where

$$W_j = \left(\frac{1}{\sigma^2}\gamma^T \Sigma_j \gamma + 1\right)^{-1} \tag{9.26}$$

is the weight variable for the jth individual. The weighted least-squares estimate of the parameters is

$$\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} X_j^T W_j X_j & \sum_{j=1}^{n} X_j^T W_j U_j \\ \sum_{j=1}^{n} U_j^T W_j X_j & \sum_{j=1}^{n} U_j^T W_j U_j \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{n} X_j^T W_j y_j \\ \sum_{j=1}^{n} U_j^T W_j y_j \end{bmatrix} \tag{9.27}$$

and

$$\hat{\sigma}^2 = \frac{1}{n-p-2}\sum_{j=1}^{n} W_j(y_j - X_j\hat{\beta} - U_j\hat{\gamma})^2 \tag{9.28}$$

Since W_j is a function of σ^2, iterations are required. The iteration process is demonstrated as below:

1. Initialize γ and σ^2.
2. Update β and γ using 9.27.
3. Update σ^2 using 9.28.
4. Repeat Step 2 to Step 3 until a certain criterion of convergence is satisfied.

The iteration process is very fast, usually taking less than 5 iterations to converge. Since the weight is not a constant (it is a function of the parameters), repeatedly updating the weight is required. Therefore, the weighted least-squares method is also called iteratively reweighted least squares (IRLS). The few cycles of iterations make the results of IRLS very close to that of the maximum likelihood method (to be introduced later). A nice property of the IRLS is that the variance–covariance matrix

of the estimated parameters is automatically given as a by-product of the iteration process. This matrix is

$$\text{var}\begin{bmatrix}\hat{\beta}\\ \hat{\gamma}\end{bmatrix} = \begin{bmatrix}\sum_{j=1}^{n} X_j^T W_j X_j & \sum_{j=1}^{n} X_j^T W_j U_j\\ \sum_{j=1}^{n} U_j^T W_j X_j & \sum_{j=1}^{n} U_j^T W_j U_j\end{bmatrix}^{-1} \hat{\sigma}^2 \tag{9.29}$$

As a result, the F- or W-test statistic can be used for significance test. Like the least-squares method, a likelihood ratio test statistic can also be established for significance test. The L_0 under the null model is the same as that described in the section of least-squares method. The L_1 under the alternative model is

$$\begin{aligned} L_1 &= -\frac{n}{2}\ln(\hat{\sigma}^2) + \frac{1}{2}\sum_{j=1}^{n}\ln(W_j) - \frac{1}{2\hat{\sigma}^2}\sum_{j=1}^{n} W_j(y - X_j\hat{\beta} - U_j\hat{\gamma})^2 \\ &\approx -\frac{n}{2}\left[\ln(\hat{\sigma}^2) + 1\right] + \frac{1}{2}\sum_{j=1}^{n}\ln(W_j) \end{aligned} \tag{9.30}$$

9.3 Fisher Scoring

The weighted least-squares solution described in the previous section does not maximize the log likelihood function (9.30). We can prove that it actually maximizes (9.30) if W_j is treated as a constant. The fact that W_j is a function of parameters makes the above weighted least-squares estimates suboptimal. The optimal solution should be obtained by maximizing (9.30) fully without assuming W_j being a constant.

Recall that the linear model for y_j is

$$y_j = X_j\beta + U_j\gamma + e_j \tag{9.31}$$

where the residual error $e_j = (Z_j - U_j)\gamma + \varepsilon_j$ has a zero mean and variance

$$\sigma_j^2 = \left(\frac{1}{\sigma^2}\gamma^T\Sigma_j\gamma + 1\right)\sigma^2 = \frac{1}{W_j}\sigma^2 \tag{9.32}$$

If we assume that $e_j \sim N(0, \sigma_j^2)$, we can construct the following log likelihood function:

$$L(\theta) = -\frac{n}{2}\ln(\sigma^2) + \frac{1}{2}\sum_{j=1}^{n}\ln(W_j) - \frac{1}{2\sigma^2}\sum_{j=1}^{n} W_j(y - X_j\beta - U_j\gamma)^2 \tag{9.33}$$

where $\theta = \{\beta, \gamma, \sigma^2\}$ is the vector of parameters. The maximum likelihood solution for the above likelihood function is hard to obtain because W_j is not a constant but a function of the parameters. The Newton–Raphson algorithm may be adopted, but it requires the second partial derivative of the log likelihood function with respect to the parameter, which is very complicated. In addition, the Newton–Raphson algorithm often misbehaves when the dimensionality of θ is high. We now introduce the Fisher scoring algorithm for finding the MLE of θ. The method requires the first partial derivative of $L(\theta)$ with respect to the parameters, called the score vector and denoted by $S(\theta)$, and the information matrix, denoted by $I(\theta)$. The score vector has the following form:

$$S(\theta) = \begin{bmatrix} \frac{1}{\sigma^2} \sum_{j=1}^{n} X_j^T W_j (y_j - \mu_j) \\ \frac{1}{\sigma^2} \sum_{j=1}^{n} U_j^T W_j (y_j - \mu_j) - \frac{1}{\sigma^2} \sum_{j=1}^{n} W_j \Sigma_j \gamma + \frac{1}{\sigma^4} \sum_{j=1}^{n} (y_j - \mu_j)^2 W_j^2 \Sigma_j \gamma \\ \frac{1}{2\sigma^4} \sum_{j=1}^{n} W_j^2 (y_j - \mu_j)^2 - \frac{1}{2\sigma^2} \sum_{j=1}^{n} W_j \end{bmatrix} \tag{9.34}$$

where

$$\mu_j = X_j \beta + U_j \gamma \tag{9.35}$$

The information matrix is given below

$$I(\theta) = \begin{bmatrix} \frac{1}{\sigma^2} \sum_{j=1}^{n} X_j^T W_j X_j & \frac{1}{\sigma^2} \sum_{j=1}^{n} X_j^T W_j U_j & 0 \\ \frac{1}{\sigma^2} \sum_{j=1}^{n} U_j W_j X_j, & \frac{1}{\sigma^2} \sum_{j=1}^{n} U_j^T W_j U_j + \frac{2}{\sigma^4} \sum_{j=1}^{n} W_j^2 \Sigma_j \gamma \gamma^T \Sigma_j, & \frac{1}{\sigma^4} \sum_{j=1}^{n} W_j^2 \Sigma_j \gamma \\ 0 & \frac{1}{\sigma^4} \sum_{j=1}^{n} W_j^2 \gamma^T \Sigma_j & \frac{1}{2\sigma^4} \sum_{j=1}^{n} W_j^2 \end{bmatrix} \tag{9.36}$$

The Fisher scoring algorithm is implemented using the following iteration equation:

$$\theta^{(t+1)} = \theta^{(t)} + I^{-1}(\theta^{(t)}) S(\theta^{(t)}) \tag{9.37}$$

where $\theta^{(t)}$ is the parameter value at iteration t and $\theta^{(t+1)}$ is the updated value. Once the iteration process converges, the variance–covariance matrix of the estimated parameters is automatically given, which is

$$\text{var}(\hat{\theta}) = I^{-1}(\hat{\theta}) \tag{9.38}$$

The detailed expression of this matrix is

$$\text{var}\begin{bmatrix}\hat{\beta}\\ \hat{\gamma}\\ \hat{\sigma}^2\end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} X_j^T W_j X_j & \sum_{j=1}^{n} X_j^T W_j U_j & 0 \\ \sum_{j=1}^{n} U_j W_j X_j, & \sum_{j=1}^{n} U_j^T W_j U_j + \frac{2}{\hat{\sigma}^2}\sum_{j=1}^{n} W_j^2 \Sigma_j \hat{\gamma}\hat{\gamma}^T \Sigma_j, & \frac{1}{\sigma^2}\sum_{j=1}^{n} W_j^2 \Sigma_j \hat{\gamma} \\ 0 & \frac{1}{\hat{\sigma}^2}\sum_{j=1}^{n} W_j^2 \hat{\gamma}^T \Sigma_j & \frac{1}{2\hat{\sigma}^2}\sum_{j=1}^{n} W_j^2 \end{bmatrix}^{-1} \hat{\sigma}^2 \tag{9.39}$$

which can be compared with the variance–covariance matrix of the iteratively reweighted least-squares estimate given in the previous section (9.29).

We now give the derivation of the score vector and the information matrix. We can write the log likelihood function as

$$L(\theta) = \sum_{j=1}^{n} L_j(\theta) \tag{9.40}$$

where

$$L_j(\theta) = -\frac{1}{2}\ln(\sigma^2) + \frac{1}{2}\ln W_j - \frac{1}{2\sigma^2} W_j (y_j - \mu_j)^2 \tag{9.41}$$

and

$$\mu_j = X_j \beta + U_j \gamma \tag{9.42}$$

The score vector is a vector of the first partial derivatives, as shown below:

$$S(\theta) = \sum_{j=1}^{n} S_j(\theta) \tag{9.43}$$

where

$$S_j(\theta) = \begin{bmatrix} \frac{\partial}{\partial \beta} L_j(\theta) \\ \frac{\partial}{\partial \gamma} L_j(\theta) \\ \frac{\partial}{\partial \sigma^2} L_j(\theta) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} X_j^T W_j (y_j - \mu_j) \\ \frac{1}{\sigma^2} U_j^T W_j (y_j - \mu_j) - \frac{1}{\sigma^2} W_j \Sigma_j \gamma + \frac{1}{\sigma^4}(y_j - \mu_j)^2 W_j^2 \Sigma_j \gamma \\ \frac{1}{2\sigma^4} W_j^2 (y_j - \mu_j)^2 - \frac{1}{2\sigma^2} W_j \end{bmatrix} \tag{9.44}$$

Therefore, we only need to take the sum of the first partial derivatives across individuals to get the score vector. Note that when deriving $S_j(\theta)$, we need the following derivatives:

$$\frac{\partial W_j}{\partial \theta} = \begin{bmatrix} \frac{\partial W_j}{\partial \beta} \\ \frac{\partial W_j}{\partial \gamma} \\ \frac{\partial W_j}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{2}{\sigma^2} W_j^2 \Sigma_j \gamma \\ \frac{1}{\sigma^2} W_j (1 - W_j) \end{bmatrix} \tag{9.45}$$

and

$$\frac{\partial \mu_j}{\partial \theta} = \begin{bmatrix} \frac{\partial \mu_j}{\partial \beta} \\ \frac{\partial \mu_j}{\partial \gamma} \\ \frac{\partial \mu_j}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} X_j^T \\ U_j^T \\ 0 \end{bmatrix} \tag{9.46}$$

The information matrix is

$$I(\theta) = \sum_{j=1}^{n} I_j(\theta) = \sum_{j=1}^{n} -E[H_j(\theta)] \tag{9.47}$$

where

$$H_j(\theta) = \frac{\partial^2 L_j(\theta)}{\partial \theta \partial \theta^T} = \begin{bmatrix} \frac{\partial^2 L_j(\theta)}{\partial \beta \partial \beta^T} & \frac{\partial^2 L_j(\theta)}{\partial \beta \partial \gamma^T} & \frac{\partial^2 L_j(\theta)}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 L_j(\theta)}{\partial \gamma \partial \beta^T} & \frac{\partial^2 L_j(\theta)}{\partial \gamma \partial \gamma^T} & \frac{\partial^2 L_j(\theta)}{\partial \gamma \partial \sigma^2} \\ \frac{\partial^2 L_j(\theta)}{\partial \sigma^2 \partial \beta^T} & \frac{\partial^2 L_j(\theta)}{\partial \sigma^2 \partial \gamma^T} & \frac{\partial^2 L_j(\theta)}{\partial \sigma^2 \partial \sigma^2} \end{bmatrix} \tag{9.48}$$

is the second partial derivative of $L_j(\theta)$ with respect to the parameters and called the Hessian matrix. Derivation of this matrix is very tedious, but the negative expectation of the Hessian matrix is identical to the expectation of the product of the score vector (Wedderburn 1974),

$$-E[H_j(\theta)] = E[S_j(\theta) S_j^T(\theta)] \tag{9.49}$$

Using this identity, we can avoid the Hessian matrix. Therefore, the information matrix is

$$I(\theta) = \sum_{j=1}^{n} I_j(\theta) = \sum_{j=1}^{n} E[S_j(\theta) S_j^T(\theta)] \tag{9.50}$$

where

$$E[S_j(\theta) S_j^T(\theta)] = \begin{bmatrix} E\left(\frac{\partial L_j(\theta)}{\partial \beta} \frac{\partial L_j(\theta)}{\partial \beta^T}\right) & E\left(\frac{\partial L_j(\theta)}{\partial \beta} \frac{\partial L_j(\theta)}{\partial \gamma^T}\right) & E\left(\frac{\partial L_j(\theta)}{\partial \beta} \frac{\partial L_j(\theta)}{\partial \sigma^2}\right) \\ E\left(\frac{\partial L_j(\theta)}{\partial \gamma} \frac{\partial L_j(\theta)}{\partial \beta^T}\right) & E\left(\frac{\partial L_j(\theta)}{\partial \gamma} \frac{\partial L_j(\theta)}{\partial \gamma^T}\right) & E\left(\frac{\partial L_j(\theta)}{\partial \gamma} \frac{\partial L_j(\theta)}{\partial \sigma^2}\right) \\ E\left(\frac{\partial L_j(\theta)}{\partial \sigma^2} \frac{\partial L_j(\theta)}{\partial \beta^T}\right) & E\left(\frac{\partial L_j(\theta)}{\partial \sigma^2} \frac{\partial L_j(\theta)}{\partial \gamma^T}\right) & E\left(\frac{\partial L_j(\theta)}{\partial \sigma^2} \frac{\partial L_j(\theta)}{\partial \sigma^2}\right) \end{bmatrix} \tag{9.51}$$

Note that the expectation is taken with respect to the phenotypic value y_j. In other words, after taking the expectation, variable y_j will disappear from the expressions. There are six different blocks in the above matrix. We will only provide the derivation for one block as an example. The derivations of the remaining five blocks are left to students for practice. The result can be found in Han and Xu (2008). We now show the derivation of the first block of the matrix. The product (before taking the expectation) is

$$\begin{aligned}\frac{\partial L_j(\theta)}{\partial \beta}\frac{\partial L_j(\theta)}{\partial \beta^T} &= \left[\frac{1}{\sigma^2}X_j^T W_j(y_j-\mu_j)\right]\left[\frac{1}{\sigma^2}X_j^T W_j(y_j-\mu_j)\right]^T \\ &= \frac{1}{\sigma^4}X_j^T W_j^2 X_j^T (y_j-\mu_j)^2 \end{aligned} \tag{9.52}$$

The expectation of it is

$$\begin{aligned}E\left(\frac{\partial L_j(\theta)}{\partial \beta}\frac{\partial L_j(\theta)}{\partial \beta^T}\right) &= \frac{1}{\sigma^4}X_j^T W_j^2 X_j^T E\left[(y_j-\mu_j)^2\right] \\ &= \frac{1}{\sigma^2}X_j^T W_j X_j^T \end{aligned} \tag{9.53}$$

The second line of the above equation requires the following identity:

$$E\left[(y_j-\mu_j)^2\right] = \frac{1}{W_j}\sigma^2 \tag{9.54}$$

Taking the sum of (9.53) across individuals, we get

$$I_{11}(\theta) = \frac{1}{\sigma^2}\sum_{j=1}^{n} X_j^T W_j X_j \tag{9.55}$$

which is the first block of the information matrix. When deriving the expectations for the remaining five blocks, we need the following expectations:

$$E[(y_j-\mu_j)^k] = \begin{cases} 0 & \text{for odd } k \\ W_j^{-1}\sigma^2 & \text{for } k=2 \\ 3W_j^{-2}\sigma^4 & \text{for } k=4 \end{cases} \tag{9.56}$$

The above expectations requires the assumption of $y_j \sim N(\mu_j, \sigma_j^2)$ where $\sigma_j^2 = W_j^{-1}\sigma^2$.

9.4 Maximum Likelihood Method

The maximum likelihood method (Lander and Botstein 1989) is the optimal one compared to all other methods described in this chapter. The linear model for the phenotypic value of y_j is

$$y_j = X_j\beta + Z_j\gamma + \varepsilon_j \tag{9.57}$$

where $\varepsilon_j \sim N(0, \sigma^2)$ is assumed. The genotype indicator variable Z_j is a missing value because we cannot observe the genotype of a putative QTL. Rather than replacing Z_j by U_j as done in the least-squares and the weighted least-squares methods, the maximum likelihood method takes into consideration the mixture distribution of y_j. We have learned the mixture distribution in Chap. 7 when we deal with segregation analysis of quantitative traits. We now extend the mixture model to interval mapping. When the genotype of the putative QTL is observed, the probability density of y_j is

$$\begin{aligned} f_k(y_j) &= \Pr(y_j | Z_j = H_k) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(y_j - X_j\beta + H_k\gamma)^2\right] \end{aligned} \tag{9.58}$$

When flanking marker information is used, the conditional probability that $Z_j = H_k$ is

$$p_j(k) = \Pr(Z_j = H_k), \forall k = 1, 2, 3 \tag{9.59}$$

for the three genotypes, A_1A_1, A_1A_2, and A_2A_2. These probabilities are different from the Mendelian segregation ratio $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ as described in the segregation analysis. They are the conditional probabilities given marker information and thus vary from one individual to another because different individuals may have different marker genotypes. Using the conditional probabilities as weights, we get the mixture distribution

$$f(y_j) = \sum_{k=1}^{3} p_j(2-k) f_k(y_j) \tag{9.60}$$

where

$$p_j(2-k) = \begin{cases} p_j(-1) & \text{for } k = 1 \\ p_j(0) & \text{for } k = 2 \\ p_j(+1) & \text{for } k = 3 \end{cases} \tag{9.61}$$

is a special notation for the conditional probability and should not be interpreted as p_j times $(2-k)$. The log likelihood function is

$$L(\theta) = \sum_{j=1}^{n} L_j(\theta) \tag{9.62}$$

where $L_j(\theta) = \ln f(y_j)$.

9.4.1 EM Algorithm

The MLE of θ can be obtained using any numerical algorithms but the EM algorithm is generally more preferable than others because we can take advantage of the mixture distribution. Derivation of the EM algorithm has been given in Chap. 7 when segregation analysis was introduced. Here we simply give the result of the EM algorithm. Assuming that the genotypes of all individuals are observed, the maximum likelihood estimates of parameters would be

$$\begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} X_j^T X_j & \sum_{j=1}^{n} X_j^T Z_j \\ \sum_{j=1}^{n} Z_j^T X_j & \sum_{j=1}^{n} Z_j^T Z_j \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{n} X_j^T y_j \\ \sum_{j=1}^{n} Z_j^T y_j \end{bmatrix} \tag{9.63}$$

and

$$\sigma^2 = \frac{1}{n} \sum_{j=1}^{n} (y_j - X_j\beta - Z_j\gamma)^2 \tag{9.64}$$

The EM algorithm takes advantage of the above explicit solutions of the parameters by substituting all entities containing the missing value Z_j by their posterior expectations, i.e.,

$$\begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} X_j^T X_j & \sum_{j=1}^{n} X_j^T E(Z_j) \\ \sum_{j=1}^{n} E(Z_j^T) X_j & \sum_{j=1}^{n} E(Z_j^T Z_j) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{n} X_j^T y_j \\ \sum_{j=1}^{n} E(Z_j^T) y_j \end{bmatrix} \tag{9.65}$$

and

$$\sigma^2 = \frac{1}{n} \sum_{j=1}^{n} E\left[(y_j - X_j\beta - Z_j\gamma)^2\right] \tag{9.66}$$

where the expectations are taken using the posterior probabilities of QTL genotypes, which is defined as

$$p_j^*(2-k) = \frac{p_j(2-k) f_k(y_j)}{\sum_{k'=1}^{3} p_j(2-k') f_{k'}(y_j)}, \forall k = 1, 2, 3 \tag{9.67}$$

The posterior expectations are

$$E(Z_j) = \sum_{k=1}^{3} p_j^*(2-k)H_k$$

$$E(Z_j^T Z_j) = \sum_{k=1}^{3} p_j^*(2-k)H_k^T H_k$$

$$E\left[(y_j - X_j\beta - Z_j\gamma)^2\right] = \sum_{k=1}^{3} p_j^*(2-k)(y_j - X_j\beta - H_k\gamma)^2 \tag{9.68}$$

Since $f_k(y_j)$ is a function of parameters, thus $p_j^*(2-k)$ is also a function of the parameters. However, the parameters are unknown, and they are the very quantities we want to find out. Therefore, iterations are required. Here is the iteration process:

1. Initialize $\theta = \theta^{(t)}$ for $t = 0$.
2. Calculate the posterior expectations using (9.67) and (9.68).
3. Update parameters using (9.65) and (9.66).
4. Increment t by 1, and repeat Step 2 to Step 3 until a certain criterion of convergence is satisfied.

Once the iteration converges, the MLE of the parameters is $\hat{\theta} = \theta^{(t)}$, where t is the number of iterations required for convergence.

9.4.2 Variance–Covariance Matrix of $\hat{\theta}$

Unlike the weighted least-squares and the Fisher scoring algorithms where the variance–covariance matrix of the estimated parameters is automatically given as a by-product of the iteration process, the EM algorithm requires an additional step to calculate this matrix. The method was developed by Louis (1982), and it requires the score vectors and the Hessian matrix for the complete-data log likelihood function rather than the actual observed log likelihood function. The complete-data log likelihood function is the log likelihood function as if Z_j were observed, which is

$$L(\theta, Z) = \sum_{j=1}^{n} L_j(\theta, Z) \tag{9.69}$$

wheres

$$L_j(\theta, Z) = -\frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y_j - X_j\beta - Z_j\gamma)^2 \tag{9.70}$$

The score vector is

$$S(\theta, Z) = \sum_{j=1}^{n} S_j(\theta, Z) \tag{9.71}$$

where

$$S_j(\theta, Z) = \begin{bmatrix} \frac{\partial}{\partial \beta} L_j(\theta, Z) \\ \frac{\partial}{\partial \gamma} L_j(\theta, Z) \\ \frac{\partial}{\partial \sigma^2} L_j(\theta, Z) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} X_j^T (y_j - X_j \beta - Z_j \gamma) \\ \frac{1}{\sigma^2} Z_j^T (y_j - X_j \beta - Z_j \gamma) \\ -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} (y_j - X_j \beta - Z_j \gamma)^2 \end{bmatrix} \tag{9.72}$$

The second partial derivative (Hessian matrix) is

$$H(\theta, Z) = \sum_{j=1}^{n} H_j(\theta, Z) \tag{9.73}$$

where

$$H_j(\theta, Z) = \begin{bmatrix} \frac{\partial^2 L_j(\theta,Z)}{\partial \beta \partial \beta^T} & \frac{\partial^2 L_j(\theta,Z)}{\partial \beta \partial \gamma^T} & \frac{\partial^2 L_j(\theta,Z)}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 L_j(\theta,Z)}{\partial \gamma \partial \beta^T} & \frac{\partial^2 L_j(\theta,Z)}{\partial \gamma \partial \gamma^T} & \frac{\partial^2 L_j(\theta,Z)}{\partial \gamma \partial \sigma^2} \\ \frac{L_j(\theta,Z)}{\partial \sigma^2 \partial \beta^T} & \frac{L_j(\theta,Z)}{\partial \sigma^2 \partial \gamma^T} & \frac{L_j(\theta,Z)}{\partial \sigma^2 \partial \sigma^2} \end{bmatrix} \tag{9.74}$$

The six different blocks of the above matrix are

$$\begin{aligned} \frac{\partial^2 L_j(\theta)}{\partial \beta \partial \beta^T} &= -\frac{1}{\sigma^2} X_j^T X_j \\ \frac{\partial^2 L_j(\theta)}{\partial \beta \partial \gamma^T} &= -\frac{1}{\sigma^2} X_j^T Z_j \\ \frac{\partial^2 L_j(\theta)}{\partial \beta \partial \sigma^2} &= -\frac{1}{\sigma^4} X_j^T (y_j - X_j \beta - Z_j \gamma) \\ \frac{\partial^2 L_j(\theta)}{\partial \gamma \partial \gamma^T} &= -\frac{1}{\sigma^2} Z_j^T Z_j \\ \frac{\partial^2 L_j(\theta)}{\partial \gamma \partial \sigma^2} &= -\frac{1}{\sigma^4} Z_j^T (y_j - X_j \beta - Z_j \gamma) \\ \frac{\partial^2 L_j(\theta)}{\partial \sigma^2 \partial \sigma^2} &= \frac{1}{2\sigma^4} - \frac{1}{\sigma^6} (y_j - X_j \beta - Z_j \gamma)^2 \end{aligned} \tag{9.75}$$

We now have the score vector and the Hessian matrix available for the complete-data log likelihood function. The Louis information matrix is

$$I(\theta) = -E[H(\theta, Z)] - E[S(\theta, Z)S^T(\theta, Z)] \tag{9.76}$$

where the expectations are taken with respect to the missing value (Z_j) using the posterior probabilities of QTL genotypes. At the MLE of parameters, $E[S(\hat{\theta}, Z)] = 0$. Therefore,

$$\begin{aligned} E[S(\theta, Z)S^T(\theta, Z)] &= \text{var}[S(\theta, Z)] + E[S(\theta, Z)]E[S^T(\theta, Z)] \\ &= \text{var}[S(\theta, Z)] \end{aligned} \tag{9.77}$$

As a result, an alternative expression of the Louis information matrix is

$$I(\theta) = -E[H(\theta, Z)] - \text{var}[S(\theta, Z)] \tag{9.78}$$

$$= -\sum_{j=1}^{n} E[H_j(\theta, Z)] - \sum_{j=1}^{n} \text{var}[S_j(\theta, Z)] \tag{9.79}$$

The expectations are

$$E[H_j(\theta, Z)] = \begin{bmatrix} E\left(\frac{\partial^2 L_j(\theta,Z)}{\partial\beta\partial\beta^T}\right) & E\left(\frac{\partial^2 L_j(\theta,Z)}{\partial\beta\partial\gamma^T}\right) & E\left(\frac{\partial^2 L_j(\theta,Z)}{\partial\beta\partial\sigma^2}\right) \\ E\left(\frac{\partial^2 L_j(\theta,Z)}{\partial\gamma\partial\beta^T}\right) & E\left(\frac{\partial^2 L_j(\theta,Z)}{\partial\gamma\partial\gamma^T}\right) & E\left(\frac{\partial^2 L_j(\theta,Z)}{\partial\gamma\partial\sigma^2}\right) \\ E\left(\frac{L_j(\theta,Z)}{\partial\sigma^2\partial\beta^T}\right) & E\left(\frac{L_j(\theta,Z)}{\partial\sigma^2\partial\gamma^T}\right) & E\left(\frac{L_j(\theta,Z)}{\partial\sigma^2\partial\sigma^2}\right) \end{bmatrix} \tag{9.80}$$

The six different blocks of the above matrix are

$$\begin{aligned} E\left(\frac{\partial^2 L_j(\theta)}{\partial\beta\partial\beta^T}\right) &= -\frac{1}{\sigma^2} X_j^T X_j \\ E\left(\frac{\partial^2 L_j(\theta)}{\partial\beta\partial\gamma^T}\right) &= -\frac{1}{\sigma^2} X_j^T E(Z_j) \\ E\left(\frac{\partial^2 L_j(\theta)}{\partial\beta\partial\sigma^2}\right) &= -\frac{1}{\sigma^4} X_j^T \left[y_j - X_j\beta - E(Z_j)\gamma\right] \\ E\left(\frac{\partial^2 L_j(\theta)}{\partial\gamma\partial\gamma^T}\right) &= -\frac{1}{\sigma^2} E(Z_j^T Z_j) \\ E\left(\frac{\partial^2 L_j(\theta)}{\partial\gamma\partial\sigma^2}\right) &= -\frac{1}{\sigma^4} E\left[Z_j^T (y_j - X_j\beta - Z_j\gamma)\right] \\ E\left(\frac{\partial^2 L_j(\theta)}{\partial\sigma^2\partial\sigma^2}\right) &= \frac{1}{2\sigma^4} - \frac{1}{\sigma^6} E\left[(y_j - X_j\beta - Z_j\gamma)^2\right] \end{aligned} \tag{9.81}$$

Again, all the expectations are taken with respect to the missing value Z_j, not the observed phenotype y_j. This is very different from the information matrix of the Fisher scoring algorithm. The variance–covariance matrix of the score vector is

$$\text{var}[S(\theta, Z)] = \sum_{j=1}^{n} \text{var}[S_j(\theta, Z)] \tag{9.82}$$

where $\text{var}[S_j(\theta, Z)]$ is a symmetric matrix as shown below:

$$\begin{bmatrix} \text{var}\left(\frac{\partial L_j(\theta,Z)}{\partial \beta}\right) & \text{cov}\left(\frac{\partial L_j(\theta,Z)}{\partial \beta}, \frac{\partial L_j(\theta,Z)}{\partial \gamma^T}\right) & \text{cov}\left(\frac{\partial L_j(\theta,Z)}{\partial \beta}, \frac{\partial L_j(\theta,Z)}{\partial \sigma^2}\right) \\ \text{cov}\left(\frac{\partial L_j(\theta,Z)}{\partial \gamma}, \frac{\partial L_j(\theta,Z)}{\partial \beta^T}\right) & \text{var}\left(\frac{\partial L_j(\theta,Z)}{\partial \gamma}\right) & \text{cov}\left(\frac{\partial L_j(\theta,Z)}{\partial \gamma}, \frac{\partial L_j(\theta,Z)}{\partial \sigma^2}\right) \\ \text{cov}\left(\frac{\partial L_j(\theta,Z)}{\partial \sigma^2}, \frac{\partial L_j(\theta,Z)}{\partial \beta^T}\right) & \text{cov}\left(\frac{\partial L_j(\theta,Z)}{\partial \sigma^2}, \frac{\partial L_j(\theta,Z)}{\partial \gamma^T}\right) & \text{var}\left(\frac{\partial L_j(\theta,Z)}{\partial \sigma^2}\right) \end{bmatrix} \tag{9.83}$$

The variances are calculated with respect to the missing value Z_j using the posterior probabilities of QTL genotypes. We only provide the detailed expression of one block of the above matrix. The remaining blocks are left to students for practice. The block that is used as an example is the (1,2) block.

$$\begin{aligned} \text{cov}\left(\frac{\partial L_j(\theta, Z)}{\partial \beta}, \frac{\partial L_j(\theta, Z)}{\partial \gamma^T}\right) &= E\left[\frac{\partial L_j(\theta, Z)}{\partial \beta}\frac{\partial L_j(\theta, Z)}{\partial \gamma^T}\right] \\ &\quad - E\left[\frac{\partial L_j(\theta, Z)}{\partial \beta}\right] E\left[\frac{\partial L_j(\theta, Z)}{\partial \gamma^T}\right] \end{aligned} \tag{9.84}$$

where

$$\begin{aligned} E\left[\frac{\partial L_j(\theta, Z)}{\partial \beta}\right] &= \frac{1}{\sigma^2} X_j^T [y_j - X_j\beta - E(Z_j)\gamma] \\ E\left[\frac{\partial L_j(\theta, Z)}{\partial \gamma^T}\right] &= \frac{1}{\sigma^2} E[(y_j - X_j\beta - Z_j\gamma)Z_j] \\ E\left[\frac{\partial L_j(\theta, Z)}{\partial \beta}\frac{\partial L_j(\theta, Z)}{\partial \gamma^T}\right] &= \frac{1}{\sigma^4} E\left[X_j^T (y_j - X_j\beta - Z_j\gamma)^2 Z_j\right] \end{aligned} \tag{9.85}$$

We already learned how to calculate $E(Z_j)$ using the posterior probability of QTL genotype. The other expectations are

$$\begin{aligned} E[(y_j - X_j\beta - Z_j\gamma)Z_j] &= \sum_{k=1}^{3} p_j^*(2-k)(y_j - X_j\beta - H_k\gamma)H_k \\ E\left[X_j^T (y_j - X_j\beta - Z_j\gamma)^2 Z_j^T\right] &= \sum_{k=1}^{3} p_j^*(2-k) X_j^T (y_j - X_j\beta - H_k\gamma)^2 H_k^T \end{aligned} \tag{9.86}$$

When calculating the information matrix, the parameter θ is substituted by $\hat{\theta}$, the MLE of θ. Therefore, the observed information matrix is

$$I(\hat{\theta}) = -E[H(\hat{\theta}, Z)] - \text{var}[S(\hat{\theta}, Z)] \tag{9.87}$$

and the variance–covariance matrix of the estimated parameters is $\text{var}(\hat{\theta}) = I^{-1}(\hat{\theta})$.

9.4.3 Hypothesis Test

The hypothesis that $H_0 : \gamma = 0$ can be tested using several different ways. If $\text{var}(\hat{\theta})$ is already calculated, we can use the F- or W-test statistic, which requires $\text{var}(\hat{\gamma})$, the variance–covariance matrix of the estimated QTL effects. It is a submatrix of $\text{var}(\hat{\theta})$. The W-test statistic is

$$W = \hat{\gamma}^T \text{var}^{-1}(\hat{\gamma})\hat{\gamma} \tag{9.88}$$

Alternatively, the likelihood ratio test statistic can be applied to test H_0. We have presented two log likelihood functions; one is the complete-data log likelihood function, denoted by $L(\theta, Z)$, and the other is the observed log likelihood function, denoted by $L(\theta)$. The log likelihood function used to construct the likelihood ratio test statistic is $L(\theta)$, not $L(\theta, Z)$. This complete-data log likelihood function, $L(\theta, Z)$, is only used to derive the EM algorithm and the observed information matrix. The likelihood ratio test statistic is

$$\lambda = -2(L_0 - L_1)$$

where $L_1 = L(\hat{\theta})$ is the observed log likelihood function evaluated at $\hat{\theta} = \{\hat{\beta}, \hat{\gamma}, \hat{\sigma}^2\}$ and L_0 is the log likelihood function evaluated at $\hat{\hat{\theta}} = \{\hat{\hat{\beta}}, 0, \hat{\hat{\sigma}}^2\}$ under the restricted model. The estimated parameter $\hat{\hat{\theta}}$ under the restricted model and L_0 are the same as those given in the section of the least-squares method.

9.5 Remarks on the Four Methods of Interval Mapping

The LS method (Haley and Knott 1992) is an approximation of the ML method, aiming to improve the computational speed. The method has been extended substantially to many other situations, e.g., multiple-trait QTL mapping (Knott and Haley 2000) and QTL mapping for binary traits (Visscher et al. 1996). When used for binary and other nonnormal traits, the method is no longer called LS. Because of the fast speed, the method remains a popular method, even though

the computer power has increased by many orders of magnitude since the LS was developed. In some literature (e.g., Feenstra et al. 2006), the LS method is also called the H–K method in honor of the authors, Haley and Knott (1992). Xu (1995) noticed that the LS method, although a good approximation to ML in terms of estimates of QTL effects and test statistic, may lead to a biased (inflated) estimate for the residual error variance. Based on this work, Xu (1998a,b) eventually developed the iteratively reweighted least-squares (IRLS) method. In these works (Xu 1998a,b), the iteratively reweighted least squares was abbreviated IRWLS. Xu (1998b) compared LS, IRLS, and ML in a variety of situations and concluded that IRLS is always better than LS and as efficient as ML. When the residual error does not have a normal distribution, which is required by the ML method, LS and IRLS can be better than ML. In other words, LS and IRLS are more robust than ML to the departure from normality. Kao (2000) and Feenstra et al. (2006) conducted more comprehensive investigation on LS, IRLS, and ML and found that when epistatic effects exist, LS can generate unsatisfactory results, but IRLS and ML usually map QTL better than LS. In addition, Feenstra et al. (2006) modified the weighted least-squares method by using the estimating equations (EE) algorithm. This algorithm further improved the efficiency of the weighted least squares by maximizing an approximate likelihood function. Most recently, Han and Xu (2008) developed a Fisher scoring (FISHER) algorithm to maximize the approximate likelihood function. Both the EE and Fisher algorithm maximize the same likelihood function, and thus, they produce identical results.

The LS method ignores the uncertainty of the QTL genotype. The IRLS, FISHER (or EE), and ML methods use different ways to extract information from the uncertainty of QTL genotype. If the putative location of QTL overlaps with a fully informative marker, all four methods produce identical result. Therefore, if the marker density is sufficiently high, there is virtually no difference for the four methods. For low marker density, when the putative position is far away from either flanking marker, the four methods will show some difference. This difference will be magnified by large QTL. Han and Xu (2008) compared the four methods in a simulation experiment and showed that when the putative QTL position is fixed in the middle of a 10-cM interval, the four methods generated almost identical results. However, when the interval expands to 20 cM, the differences among the four methods become noticeable.

Interval mapping with a 1-cM increment for the mouse 10th-week body weight data was conducted using all the four methods by Han and Xu (2008). The LOD test statistic profiles are shown in Fig. 9.2 for the four methods of interval mapping (LS, IRLS, FISHER, and ML). There is virtually no difference for the four methods. The difference in LOD profiles is noticeable when the marker density is low. Comparisons for the estimated QTL effects were also conducted for the mouse data. Figure 9.3 shows the estimated QTL effect profiles along the genome for the four methods. Again the difference is barely noticeable.

A final remark on interval mapping is the way to infer the QTL genotype using flanking markers. If only flanking markers are used to infer the genotype of a putative position bracketed by the two markers, the method is called interval

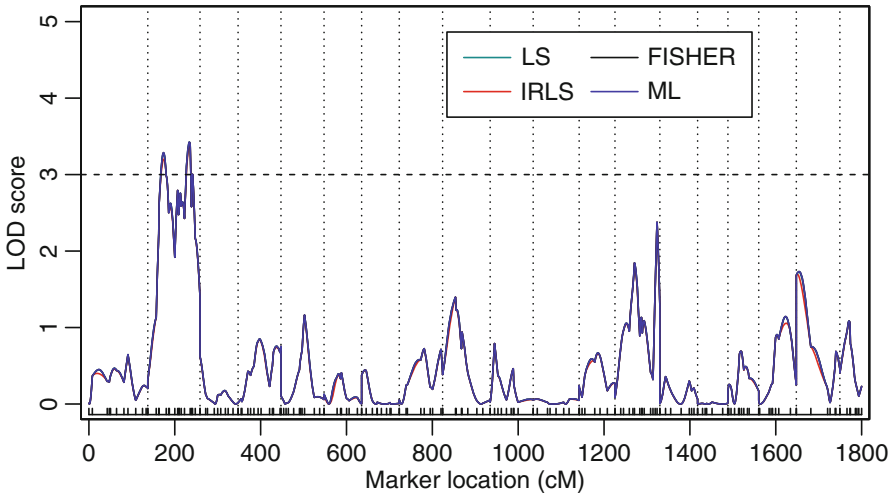

Fig. 9.2 The LOD test statistic profiles for four methods of interval mapping (LS, least square; IRLS, iteratively reweighted least square; FISHER, Fisher scoring; and ML, maximum likelihood). The mouse data were obtained from Lan et al. (2006). The trait investigated is the 10th-week body weight. The 19 chromosomes (excluding the sex chromosome) are separated by the *vertical dotted lines*. The unevenly distributed *black ticks* on the horizontal axis indicate the marker locations

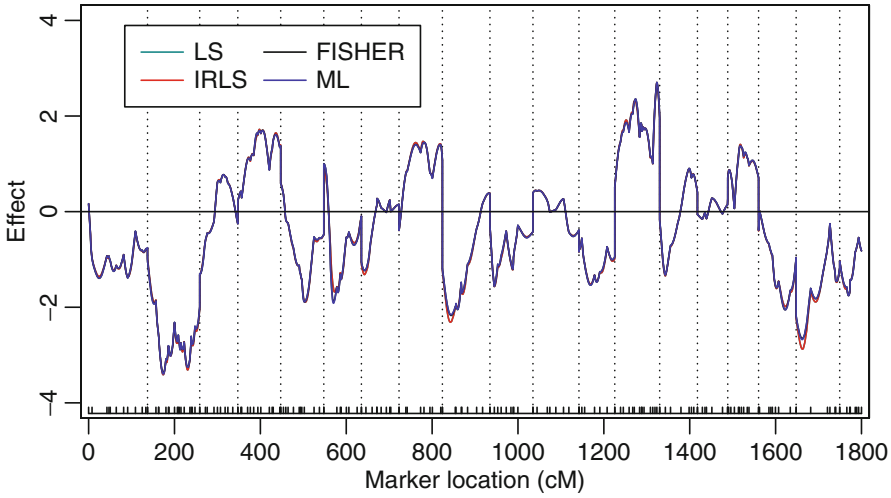

Fig. 9.3 The QTL effect profiles for four methods of interval mapping (LS, least square; IRLS, iteratively reweighted least square; FISHER, Fisher scoring; and ML, maximum likelihood). The mouse data were obtained from Lan et al. (2006). The trait investigated is the 10th-week body weight. The 19 chromosomes (excluding the sex chromosome) are separated by the *vertical dotted lines*. The unevenly distributed *black ticks* on the horizontal axis indicate the marker locations

mapping. Strictly speaking, interval mapping only applies to fully informative markers because we always use flanking markers to infer the QTL genotype. However, almost all datasets obtained from real-life experiments contain missing, uninformative, or partially informative markers. To extract maximum information from markers, people always use the multipoint method (Jiang and Zeng 1997) to infer a QTL genotype. The multipoint method uses more markers or even all markers of the entire chromosome (not just flanking markers) to infer the genotype of a putative position. With the multipoint analysis, we no longer have the notion of interval, and thus, interval mapping is no longer an appropriate phrase to describe QTL mapping. Unfortunately, a more appropriate phrase has not been proposed, and people are used to the phrase of interval mapping. Therefore, the so-called interval mapping in the current literature means QTL mapping under a single QTL model, regardless of whether the genotype of a putative QTL position is inferred from flanking markers or all markers.