

Statistical Applications in Genetics and Molecular Biology

Volume 8, Issue 1

2009

Article 16

Score Statistics for Mapping Quantitative Trait Loci

Myron N. Chang*

Rongling Wu†

Samuel S. Wu‡

George Casella**

*University of Florida, mchang@cog.ufl.edu

†University of Florida, rwu@ufl.edu

‡University of Florida, samwu@biostat.ufl.edu

**University of Florida, casella@ufl.edu

Score Statistics for Mapping Quantitative Trait Loci

Myron N. Chang, Rongling Wu, Samuel S. Wu, and George Casella

Abstract

We propose a method to detect the existence of quantitative trait loci (QTL) in a backcross population using a score test. Since the score test only uses the MLEs of parameters under the null hypothesis, it is computationally simpler than the likelihood ratio test (LRT). Moreover, because the location parameter of the QTL is unidentifiable under the null hypothesis, the distribution of the maximum of the LRT statistics, typically the statistic of choice for testing H_0 : no QTL, does not have the standard chi-square distribution asymptotically under the null hypothesis. From the simple structure of the score test statistics, the asymptotic null distribution can be derived for the maximum of the square of score test statistics. Numerical methods are proposed to compute the asymptotic null distribution and the critical thresholds can be obtained accordingly. We show that the maximum of the LR test statistics and the maximum of the square of score statistics are asymptotically equivalent. Therefore, the critical threshold for the score test can be used for the LR test also. A simple backcross design is used to demonstrate the application of the score test to QTL mapping.

KEYWORDS: score test, likelihood ratio test, asymptotic equivalence, backcross design

1 Introduction

Since the publication of the seminal mapping paper by Lander and Botstein (1989), there has been a vast wealth of literature concerning the development of statistical methods for mapping quantitative trait loci (QTL) that affect complex traits (reviewed in Doerge et al. 1997; Jansen 2001; Hoeschele 2001; Wu et al. 2007). In QTL mapping, one of the thorniest issues is the the characterization of critical thresholds to declare the statistical significance of a QTL (Lander and Schork 1994). Typically the profile of likelihood-ratio (LR) test statistics is constructed over the grid of possible QTL locations in a linkage group or an entire genome and the maximum of the LR (MLR) is used as a global test statistic. At a given position of the QTL, the LR test statistic is asymptotically χ^2 -distributed under the null hypothesis with degrees of freedom equal to the number of associated QTL effects. However, the QTL position is unidentified under the null hypothesis

$$(1) \quad H_0: \text{no QTL,}$$

therefore, the MLR test statistic does not have the standard χ^2 -distribution asymptotically.

To overcome the limitations due to the failure of the test statistic to follow a standard statistical distribution, a distribution-free simulation approach has been proposed to calculate critical values for different experimental settings with intermediate marker densities (Lander and Botstein 1989; van Ooijen 1992; Darvasi et al 1993). A more empirical approach for determining critical thresholds is based on a permutation test procedure (see Churchill and Doerge 1994 and Doerge and Churchill 1996). The simulation- or permutation-based approach requires a very high computational workload, and this makes its application impractical in many situations. Several researchers have derived formulae for computing approximate critical thresholds to control the type I error rate at a chromosome- or genome-wide level (Rebai et al. 1994; Dupuis and Siegmund 1999; Piepho 2001).

The purpose of this article is to propose a general statistical framework to detect the existence of a QTL using a score test. A score-statistic approach for mapping QTL was considered with sibships or a natural population in humans (Schaid et al. 2002; Wang and Huang 2002). Zou et al. (2004) utilized score statistics for genome-wide mapping of QTL in controlled crosses, in which the resampling method was proposed to assess the significance. Since the score test only uses maximum likelihood estimates (MLEs) of parameters under the null hypothesis (1), under which all phenotypes in the sample are independent and identically distributed (i.i.d.), this approach is much less demanding computationally than the LR test. We form score test statistics for given locations in the permissible range and then use

the maximum of the square of these score test statistics as a global test statistic. We show that under the H_0 , the maximum of the square of score test statistics asymptotically follows the distribution of $\sup Z^2(d)$, where $Z(d)$ is a Gaussian stochastic process with mean zero and well-formulated covariances, and the supremum is over the permissible range of d , the Haldane (1919) distance between one end of the genome and the location of the QTL. The distribution of $\sup Z^2(d)$ only depends on Haldane distances or equivalently on recombination fractions between markers, and the critical thresholds can be obtained by simulation. We show that the maximum of the square of score test statistics and the MLR are asymptotically equivalent, and hence the critical thresholds obtained for the maximum of the square of score test statistics can be used for MLR test also.

The article is organized as follows. We provide background information on QTL mapping in backcross population in Section 2. We derive the score test statistic and the asymptotic null distribution of the maximum of the square of score test statistics for interval mapping In Section 3. We further show that the maximum of the LR test statistics and the maximum of the square of score statistics are asymptotically equivalent under the null hypothesis. In Section 4, the score test and the null distribution are extended to composite interval mapping, where we show that, in the dense map case, the distribution of $\sup Z^2(d)$ is the same as that derived by Lander and Botstein (1989). Numerical methods to compute the null distribution are addressed in Section 5. In Section 6, examples are presented and the thresholds obtained by our method are compared with those from other procedures. Section 7 is a discussion section. We provide some technical details in the Appendix.

2 QTL Mapping in Backcross Populations

This section provides a short introduction to QTL mapping in backcross populations. A more thorough introduction can be found in Doerge *et al.* (1997).

2.1 Basics

The aim of QTL mapping is to associate genes with quantitative phenotypic traits. For example, we might be interested in QTL that affect the growth rate of pine trees, so we might be looking for regions on a chromosome that are associated with growth rate in pines.

In the simplest case we assume that there are two parents that have alleles QQ and qq at a certain (unknown) place, or locus, on the chromosome. If the parents mate, the offspring (called the F_1 generation) will have the allele Qq . In a *backcross*, the offspring is mated back to a parent, say QQ , producing a new generation with

alleles QQ, Qq . We want to investigate the association between the new generation (sometimes called BC_1) and the quantitative trait.

Now the problems arise. We typically cannot see if the locus is Q or q , but rather we see a "marker", with alleles M and m . If the marker is "close" to the gene, where closeness is measured by recombination distance, r , then we hope that seeing an M implies that a Q is there, and so on. If r is small, with $r = 0$ being the limiting case of no recombination, then M is closely linked to Q . But if r is large, with $r = 1/2$ being the limiting case, then there may be no linkage between M and Q .

A simple statistical model for all of this is the following. If Y is a random variable signifying the quantitative trait, assume that Y is normal with variance σ^2 and mean μ_{QQ} , μ_{qQ} or μ_{Qq} depending on the alleles at the particular locus under consideration. Clearly the parents are $N(\mu_{QQ}, \sigma^2)$ and $N(\mu_{qQ}, \sigma^2)$, respectively, and the F_1 is $N(\mu_{Qq}, \sigma^2)$. The distribution of the backcross to the QQ parent is a mixture of normals with means μ_{QQ} and μ_{Qq} .

If the Q alleles could be observed, the mixing fraction would be $1/2$, but we can only observe the marker genotypes M or m . If r denotes the recombination probability, then the distribution of the phenotype in the backcross population (Doerge et al. 1997) is

$$Y|MM \sim \begin{cases} N(\mu_{QQ}, \sigma^2) & \text{with probability } 1 - r \\ N(\mu_{Qq}, \sigma^2) & \text{with probability } r \end{cases},$$

and

$$Y|Mm \sim \begin{cases} N(\mu_{QQ}, \sigma^2) & \text{with probability } r \\ N(\mu_{Qq}, \sigma^2) & \text{with probability } 1 - r \end{cases}.$$

When categorized by the markers, the difference in means of the populations is

$$\mu_{MM} - \mu_{Mm} = (1 - 2r)(\mu_{QQ} - \mu_{Qq}).$$

Assuming that $\mu_{QQ} - \mu_{Qq} \neq 0$, a test of H_0 : no linkage between M and Q can be carried out by testing $H_0: \mu_{MM} - \mu_{Mm}$ (which is also equivalent to the null hypothesis $H_0: r = 1/2$). Using this mixture model in a likelihood analysis, we can not only test for linkage but also estimate r . (The backcross is the simplest design in which we get enough information to estimate r .)

2.2 Interval mapping

In practice, we typically use a somewhat more complex model, where we assume that the QTL is *bracketed* by two flanking markers $\mathcal{M}^{\ell-1}$ (with alleles $M^{\ell-1}$ and

$m^{\ell-1}$) and \mathcal{M}^ℓ (with alleles M^ℓ and m^ℓ) from a genetic linkage map. These two markers form four distinct genotype groups

$$(2) \quad \{M^{\ell-1}m^{\ell-1}M^\ell m^\ell\}, \{M^{\ell-1}m^{\ell-1}m^\ell m^\ell\}, \{m^{\ell-1}m^{\ell-1}M^\ell m^\ell\}, \{m^{\ell-1}m^{\ell-1}m^\ell m^\ell\}.$$

The recombination fraction between the two markers $\mathcal{M}^{\ell-1}$ and \mathcal{M}^ℓ is denoted by r , and that between $\mathcal{M}^{\ell-1}$ and the QTL by r_1 , and that between the QTL and \mathcal{M}^ℓ by r_2 . The corresponding genetic distances between these three loci are denoted by D , x and $D - x$, respectively. Genetic distances and recombination fractions are related through a *map function* (Wu et al. 2007). Here we use the Haldane (1919) map function, given by $r = (1 - e^{-2D})/2$, to convert the genetic distance to the corresponding recombination fraction.

To express the conditional probability of a QTL genotype, say Qq , given each of the four two-marker genotypes, we introduce two ratio parameters,

$$(3) \quad \begin{aligned} \theta_1 &= \frac{r_1 r_2}{1-r} = \frac{r_1 r_2}{(1-r_1)(1-r_2) + r_1 r_2} = \frac{1 + e^{-2D} - e^{-2x} - e^{-2(D-x)}}{2(1 + e^{-2D})}, \\ \text{and} \\ \theta_2 &= \frac{r_1(1-r_2)}{r} = \frac{r_1(1-r_2)}{r_1(1-r_2) + (1-r_1)r_2} = \frac{1 - e^{-2D} - e^{-2x} + e^{-2(D-x)}}{2(1 - e^{-2D})}. \end{aligned}$$

Furthermore, assume that the putative QTL has an additive effect on the trait, and define the expected genotypic values as $\mu + \Delta$ and $\mu - \Delta$ for the two QTL genotypes Qq and qq , respectively, where μ is the overall mean of the backcross population. The corresponding distributions of the phenotypic values for the four marker genotypes can be modelled, respectively, as

$$\begin{aligned} &(1 - \theta_1)N(\mu - \Delta, \sigma^2) + \theta_1 N(\mu + \Delta, \sigma^2), \\ &(1 - \theta_2)N(\mu - \Delta, \sigma^2) + \theta_2 N(\mu + \Delta, \sigma^2), \\ &\theta_2 N(\mu - \Delta, \sigma^2) + (1 - \theta_2)N(\mu + \Delta, \sigma^2), \\ &\theta_1 N(\mu - \Delta, \sigma^2) + (1 - \theta_1)N(\mu + \Delta, \sigma^2), \end{aligned}$$

where μ, Δ, σ^2 , and x , $0 \leq x \leq D$, are the four unknown parameters contained in the above mixture models.

Lastly, suppose that we take a sample of size N from the backcross populations, with respective sample sizes of n_1, n_2, n_3 and n_4 in the four marker groups of (2). We then have

$$(4) \quad \begin{aligned} \lim(n_1/N) &= \lim(n_4/N) = \frac{1}{2}(1 - r) \\ \lim(n_2/N) &= \lim(n_3/N) = \frac{1}{2}r, \end{aligned}$$

as $N \rightarrow \infty$.

3 Single Interval Mapping

Since the existence of the QTL in the interval is indicated by $\Delta \neq 0$, its statistical significance can be tested by the hypotheses:

$$H_0 : \Delta = 0 \text{ vs. } H_a : \Delta \neq 0,$$

where H_0 is equivalent to (1). Note that under the null hypothesis, all observations are i.i.d. from $N(\mu, \sigma^2)$. Thus, the QTL position parameter $\theta = (\theta_1, \theta_2) = \theta(x, D)$ is unidentifiable because there is no information on θ if H_0 is true. In this case, the MLR test statistic does not follow a standard χ^2 -distribution with two degrees of freedom.

We derive a score statistic to overcome this problem. Recall that n_1, n_2, n_3 and n_4 are the sample sizes of the four marker groups. If we define $N_0 = 0, N_1 = n_1, N_2 = N_1 + n_2, N_3 = N_2 + n_3$ and $N_4 = N = N_3 + n_4$, the log likelihood of the phenotype data conditional on the marker information can be written

$$\begin{aligned} \ell(\Delta, \mu, \sigma^2, x) &= -\frac{N}{2} \log 2\pi - \frac{N}{2} \log \sigma^2 \\ &+ \sum_{i=1}^{N_1} \log \left[(1 - \theta_1) e^{-\frac{1}{2\sigma^2}(y_i - \mu + \Delta)^2} + \theta_1 e^{-\frac{1}{2\sigma^2}(y_i - \mu - \Delta)^2} \right] \\ &+ \sum_{i=N_1+1}^{N_2} \log \left[(1 - \theta_2) e^{-\frac{1}{2\sigma^2}(y_i - \mu + \Delta)^2} + \theta_2 e^{-\frac{1}{2\sigma^2}(y_i - \mu - \Delta)^2} \right] \\ (5) \quad &+ \sum_{i=N_2+1}^{N_3} \log \left[\theta_2 e^{-\frac{1}{2\sigma^2}(y_i - \mu + \Delta)^2} + (1 - \theta_2) e^{-\frac{1}{2\sigma^2}(y_i - \mu - \Delta)^2} \right] \\ &+ \sum_{i=N_3+1}^N \log \left[\theta_1 e^{-\frac{1}{2\sigma^2}(y_i - \mu + \Delta)^2} + (1 - \theta_1) e^{-\frac{1}{2\sigma^2}(y_i - \mu - \Delta)^2} \right]. \end{aligned}$$

The score function is the derivative of (5) with respect to Δ evaluated at $\Delta = 0$:

$$\begin{aligned} u(\mu, \sigma^2, x) &= \left. \frac{\partial \ell}{\partial \Delta} \right|_{\Delta=0} = \frac{1}{\sigma^2} \left\{ -(1 - 2\theta_1) \sum_{i=1}^{N_1} (y_i - \mu) \right. \\ &- (1 - 2\theta_2) \sum_{i=N_1+1}^{N_2} (y_i - \mu) + (1 - 2\theta_2) \sum_{i=N_2+1}^{N_3} (y_i - \mu) \\ &\left. + (1 - 2\theta_1) \sum_{i=N_3+1}^N (y_i - \mu) \right\}. \end{aligned}$$

Under the null hypothesis, the MLEs of μ and σ^2 are

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N y_i \text{ and } \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{\mu})^2,$$

respectively. For a given $x \in [0, D]$, to form the score test statistic we need to replace μ and σ^2 in $u(\mu, \sigma^2, x)$ by $\hat{\mu}$ and $\hat{\sigma}^2$.

For ease of notation, denote

$$\begin{aligned} \bar{y}_t &= \frac{1}{n_t} \sum_{i=N_{t-1}+1}^{N_t} y_i, \quad t = 1, 2, 3, 4. \\ a(x, \mathbf{N}) &= (\theta_1 - \theta_2)n_2 - (1 - \theta_1 - \theta_2)n_3 - (1 - 2\theta_1)n_4, \\ b(x, \mathbf{N}) &= (\theta_2 - \theta_1)n_1 - (1 - 2\theta_2)n_3 - (1 - \theta_1 - \theta_2)n_4, \\ c(x, \mathbf{N}) &= (1 - \theta_1 - \theta_2)n_1 + (1 - 2\theta_2)n_2 + (\theta_1 - \theta_2)n_4 \\ d(x, \mathbf{N}) &= (1 - 2\theta_1)n_1 + (1 - \theta_1 - \theta_2)n_2 + (\theta_2 - \theta_1)n_3, \end{aligned}$$

where $\mathbf{N} = (n_1, n_2, n_3, n_4)$. We can then write

$$u(\hat{\mu}, \hat{\sigma}^2, x) = \frac{2}{N\hat{\sigma}^2} [a(x, \mathbf{N})n_1\bar{y}_1 + b(x, \mathbf{N})n_2\bar{y}_2 + c(x, \mathbf{N})n_3\bar{y}_3 + d(x, \mathbf{N})n_4\bar{y}_4],$$

where, under the null hypothesis, the variance of $u(\hat{\mu}, \hat{\sigma}^2, x)$ is

$$\text{Var}(u(\hat{\mu}, \hat{\sigma}^2, x)) = \left(\frac{2}{N\hat{\sigma}^2} \right)^2 [a^2(x, \mathbf{N})n_1 + b^2(x, \mathbf{N})n_2 + c^2(x, \mathbf{N})n_3 + d^2(x, \mathbf{N})n_4].$$

The score test statistic, $U(x)$, is $u(\hat{\mu}, \hat{\sigma}^2, x)$ divided by its standard deviation, that is,

$$(6) \quad U(x) = \frac{u(\hat{\mu}, \hat{\sigma}^2, x)}{\sqrt{\text{Var}(u(\hat{\mu}, \hat{\sigma}^2, x))}}$$

Calculation of the score test statistic $U(x)$ is simple (in contrast to the computation of the MLEs $\hat{\Delta}_x$, $\hat{\mu}_x$, and $\hat{\sigma}_x^2$, which can be complex). Moreover, the maximum score test statistic is asymptotically equivalent to the maximum likelihood ratio statistic, as we now show.

For a given $x \in [0, D]$, let $\hat{\Delta}_x$, $\hat{\mu}_x$, and $\hat{\sigma}_x^2$ be the MLEs of Δ , μ and σ^2 . By Cox and Hinkley (1974, pages 323-324), the LR test statistic and the square of the score test statistic are asymptotically equivalent at this x , that is,

$$(7) \quad 2[\ell(\hat{\Delta}_x, \hat{\mu}_x, \hat{\sigma}_x^2, x) - \ell(0, \hat{\mu}, \hat{\sigma}^2, x)] \approx U^2(x).$$

Over the range of $x \in [0, D]$, the maximum of the left-hand side of (7) (MLR) equals the maximum of the right-hand side of (7) asymptotically. The proof is deferred to Appendix 8.3.

Theorem 1

$$\lim_{N \rightarrow \infty} \left| \sup_{x \in [0, D]} 2\{\ell(\hat{\eta}(x), x) - \ell(\hat{\eta}_0)\} - \sup_{x \in [0, D]} U^2(x) \right| = 0 \text{ a.s.}$$

where $\hat{\eta}(x) = (\hat{\Delta}_x, \hat{\mu}_x, \hat{\sigma}_x)$, $\hat{\eta}_0 = (0, \hat{\mu}, \hat{\sigma})$, and the almost sure convergence is with respect to the distribution of the phenotypic trait under H_0 .

In Appendix 8.3, we show that the difference between the two sides in (7) converges to zero with probability one uniformly for $x \in [0, D]$ as $N \rightarrow \infty$.

To test (1) we need the null distribution of the maximum of $U^2(x)$ (which is equivalent to the MLR test statistic). To do this we derive a simplified and asymptotically equivalent form of the score test statistic. Because of (4), we can replace n_1, n_2, n_3 , and n_4 in $U(x)$ by $\frac{1}{2}(1-r)N, \frac{1}{2}rN, \frac{1}{2}rN$, and $\frac{1}{2}(1-r)N$, respectively. We then have

$$(8) \quad U(x) \approx U^*(x) = \left(\frac{\sqrt{N}}{2\hat{\sigma}} \right) \frac{(1-r)(1-2\theta_1)(\bar{y}_4 - \bar{y}_1) + r(1-2\theta_2)(\bar{y}_3 - \bar{y}_2)}{\sqrt{(1-r)(1-2\theta_1)^2 + r(1-2\theta_2)^2}}.$$

Under the null hypothesis,

$$(9) \quad \frac{1}{2}\sqrt{(1-r)N} \frac{(\bar{y}_4 - \bar{y}_1)}{\hat{\sigma}} \approx W_1 \text{ and } \frac{1}{2}\sqrt{rN} \frac{(\bar{y}_3 - \bar{y}_2)}{\hat{\sigma}} \approx W_2,$$

where W_1 and W_2 are two independent standard normal random variables. Thus, $U^*(x)$ is asymptotically equivalent to

$$(10) \quad Z(x) = \frac{(1-2\theta_1)\sqrt{1-r}W_1 + (1-2\theta_2)\sqrt{r}W_2}{\sqrt{(1-r)(1-2\theta_1)^2 + r(1-2\theta_2)^2}}.$$

Next, consider two QTL positions x' and $x'' \in [0, D]$, whose corresponding probabilities in (3) can be written as $(\theta'_1, \theta'_2) = \theta(x', D)$ and $(\theta''_1, \theta''_2) = \theta(x'', D)$. The covariance between $Z(x')$ and $Z(x'')$ is

$$(11) \quad \begin{aligned} & \text{cov}(Z(x'), Z(x'')) \\ &= \frac{(1-r)(1-2\theta'_1)(1-2\theta''_1) + r(1-2\theta'_2)(1-2\theta''_2)}{\sqrt{(1-r)(1-2\theta'_1)^2 + r(1-2\theta'_2)^2} \sqrt{(1-r)(1-2\theta''_1)^2 + r(1-2\theta''_2)^2}}. \end{aligned}$$

Therefore, the MLR test statistic or the maximum of the square of score test statistics asymptotically distributed as

$$(12) \quad \sup_{0 \leq x \leq D} Z^2(x),$$

where $Z(x)$ is a Gaussian stochastic process with mean 0 and covariances (11).

4 Composite Interval Mapping

In the preceding section, we discussed the asymptotic theory of the score test statistic for QTL mapping from a single marker interval. In this section, we use the full marker information in *composite interval mapping*. Statistically, this is a combination of interval mapping based on two given flanking markers and a partial regression analysis on all markers except for the two ones bracketing the QTL.

4.1 QTL mapping on a sparse map

Suppose there are $k + 1$ diallelic markers, $\mathcal{M}^0, \mathcal{M}^1, \dots, \mathcal{M}^k$, on a sparse linkage map. These markers encompass k marker intervals. In a backcross population, four marker genotypes at a given interval $[\mathcal{M}^{\ell-1}, \mathcal{M}^\ell]$ ($\ell = 1, 2, \dots, k$) are indicated by

$$G^{(\ell)} = \begin{cases} 1 & \text{if a marker genotype is } M^{\ell-1}m^{\ell-1}M^\ell m^\ell \\ 2 & \text{if a marker genotype is } M^{\ell-1}m^{\ell-1}m^\ell m^\ell \\ 3 & \text{if a marker genotype is } m^{\ell-1}m^{\ell-1}M^\ell m^\ell \\ 4 & \text{if a marker genotype is } m^{\ell-1}m^{\ell-1}m^\ell m^\ell. \end{cases}$$

The number of observations within each of these four marker genotypes at a given marker interval can be expressed as

$$n_t^{(\ell)} = \sum_{i=1}^N I(G_i^{(\ell)} = t), \quad t = 1, 2, 3, 4$$

where $I(A)$ is the indicator function of event A . We then use

$$\mathbf{N}^{(\ell)} = (n_1^{(\ell)}, n_2^{(\ell)}, n_3^{(\ell)}, n_4^{(\ell)})$$

to array the observations of all the four marker genotypes at the interval. We can then express the phenotypic mean of a quantitative trait (y) for a marker genotype t at the interval $[\mathcal{M}^{\ell-1}, \mathcal{M}^\ell]$ as

$$\bar{y}_t^{(\ell)} = \frac{1}{n_t^{(\ell)}} \sum_{i=1}^N y_i I(G_i^{(\ell)} = t).$$

Let D_ℓ and r_ℓ ($\ell = 1, 2, \dots, k$) denote the Haldane distance and recombination fraction between $\mathcal{M}^{\ell-1}$ and \mathcal{M}^ℓ , respectively, and denote the Haldane distance between the two extreme markers \mathcal{M}^0 and \mathcal{M}^k by $D = \sum_{\ell=1}^k D_\ell$. A putative QTL

must be located somewhere on the map, at unknown position $d \in [0, D]$. For each $d \in [0, D]$, define

$$(13) \quad \ell_d = \min\{\ell : \sum_{i=1}^{\ell-1} D_i \leq d < \sum_{i=1}^{\ell} D_i\} \text{ and } x_d = D - \sum_{i=1}^{\ell-1} D_i,$$

to identify the possible QTL within the interval $[\mathcal{M}^{\ell_d-1}, \mathcal{M}^{\ell_d}]$ that is associated with the distance d . As derived in Section 3, the score test statistic (6) for detecting the QTL at $d \in [0, D]$, the asymptotically equivalent forms (8) and (10), and the relationship (9) are all valid with \mathbf{N} , \bar{y}_t , and r replaced by $\mathbf{N}^{(\ell)}$, $\bar{y}_t^{(\ell)}$, and r_ℓ , respectively. In (6), (8), and (10), $(\theta_1, \theta_2) = \theta(x_d, D_{\ell_d})$ is the function of x_d and D_{ℓ_d} defined in (3).

For example, we express (10) as

$$(14) \quad Z(d) = \frac{(1 - 2\theta_1)\sqrt{1 - r_\ell}W_1^{(\ell)} + (1 - 2\theta_2)\sqrt{r_\ell}W_2^{(\ell)}}{\sqrt{(1 - r_\ell)(1 - 2\theta_1)^2 + r_\ell(1 - 2\theta_2)^2}},$$

where $W_1^{(\ell)}$ and $W_2^{(\ell)}$ are two independent standard normal random variables,

$$(15) \quad W_1^{(\ell)} \approx \frac{1}{2}\sqrt{(1 - r_\ell)N}\frac{(\bar{y}_4^{(\ell)} - \bar{y}_1^{(\ell)})}{\hat{\sigma}} \text{ and } W_2^{(\ell)} \approx \frac{1}{2}\sqrt{r_\ell N}\frac{(\bar{y}_3^{(\ell)} - \bar{y}_2^{(\ell)})}{\hat{\sigma}}.$$

In order to search for a QTL throughout the entire genome, we need to compute the covariance between $Z(d')$ and $Z(d'')$, $0 \leq d' \leq d'' \leq D$. Assume (ℓ', x') and (ℓ'', x'') are associated with d' and d'' through (13). Let $(\theta'_1, \theta'_2) = \theta(x', D_{\ell'})$ and $(\theta''_1, \theta''_2) = \theta(x'', D_{\ell''})$ as defined in (3). If $\ell' = \ell'' = \ell$, then the covariance between $Z(d')$ and $Z(d'')$ is given in (11) with r replaced by r_ℓ . If $\ell' \neq \ell''$, then the covariance between $Z(d')$ and $Z(d'')$ is derived in (A.6) in Appendix 8.1. It now follows that the MLR test statistic or the maximum of the square of score test statistics is asymptotically distributed as

$$(16) \quad \sup_{0 \leq d \leq D} Z^2(d),$$

where $Z(d)$ is a Gaussian stochastic process with mean 0 and covariance in (11) when $\ell' = \ell''$ or in (A.6) when $\ell' < \ell''$. Note that the null distribution of (16) only depends on Haldane distances or equivalently on recombination fractions between the markers. In Section 5, we will use a numerical approach to compute the distribution of (16).

4.2 QTL mapping on a dense map

For a dense genetic map, we can assume that markers are located at every point along the genome. This can be considered as the limit of the sparse map when k (the number of diallelic markers) converges to infinity and the genomic distance between any two consecutive markers converges to zero.

For a dense map, two markers $\mathcal{M}^{\ell-1}$ and \mathcal{M}^ℓ can be assumed to have the identical distance (d) to marker \mathcal{M}^0 . Furthermore, for any interval $[\mathcal{M}^{\ell-1}, \mathcal{M}^\ell]$, we have $n_2^{(\ell)} \approx n_3^{(\ell)} \approx 0$, $n_1^{(\ell)} \approx n_4^{(\ell)} \approx \frac{1}{2}N$, $\theta_1 \approx 0$, $a(x, \mathbf{N}^{(\ell)}) \approx -n_4^{(\ell)}$, and $d(x, \mathbf{N}^{(\ell)}) \approx n_1^{(\ell)}$. If the Haldane distance between \mathcal{M}^0 and the QTL, which is located within interval $[\mathcal{M}^{\ell-1}, \mathcal{M}^\ell]$, is $d \in [0, D]$, the score statistic (6) for detecting the QTL at d is reduced to

$$U(d) = \frac{\bar{y}_4^{(\ell)} - \bar{y}_1^{(\ell)}}{\hat{\sigma} \sqrt{\frac{1}{n_1^{(\ell)}} + \frac{1}{n_4^{(\ell)}}}}.$$

For two possible locations d' and d'' of the QTL, $d' < d''$, where d' and d'' belong to $[\mathcal{M}^{\ell'-1}, \mathcal{M}^{\ell'}]$ and $[\mathcal{M}^{\ell''-1}, \mathcal{M}^{\ell''}]$, the covariance (see Chang et al. 2003 for details) is reduced to

$$(17) \quad \text{cov}(U(d'), U(d'')) \approx 1 - 2r_{\ell', \ell''-1} \approx 1 - 2r_{\ell' \ell''}.$$

This suggests that the asymptotic distribution of MLR test statistic, or the maximum of the square of the score test statistics, is the same as the distribution of $\sup_{0 \leq d \leq D} Z^2(d)$, where $Z(\cdot)$ is a Gaussian stochastic process with mean 0 and covariance (17). This result is consistent with the finding of Lander and Botstein (1989).

5 Simulation of the Null Distribution

Although it is difficult to derive an explicit formula for the asymptotic null distributions derived in Sections 3 and 4.1, it is relatively straightforward to simulate them.

5.1 Interval mapping

For any given values of two standard normal random variables W_1 and W_2 , we can compute $\sup_{0 \leq x \leq D} Z^2(x)$ where $Z(x)$ is given in (10). Although x does not appear explicitly in $Z(x)$, recall from (3) that $(\theta_1, \theta_2) = \theta(x, D)$, so $Z^2(x)$ is a function of

x. Define

$$y_{1,2} = \frac{-2e^{2D}W_1W_2 \pm (W_1^2 + W_2^2)\sqrt{e^{4D} - 1}}{(W_1^2 - W_2^2)\sqrt{e^{4D} - 1} - 2W_1W_2},$$

where y_1 correspond to + of “ \pm ” and y_2 correspond to - of “ \pm ”. By straightforward computation, it can be shown that the solutions to $\frac{dZ^2(x)}{dx} = 0$ are

$$x_i = \frac{1}{4}(\log y_i + 2D), \quad i = 1, 2,$$

provided that $y_i > 0$. We can then simulate

$$(18) \quad \sup_{0 \leq x \leq D} Z^2(x) = \begin{cases} \max\{Z^2(0), Z^2(x_i), Z^2(D)\} & \text{if } y_i > 0 \text{ and } 0 \leq x_i \leq D \\ \max\{Z^2(0), Z^2(D)\} & \text{otherwise} \end{cases}.$$

by simulating W_1 and W_2 i.i.d. standard normal.

5.2 Composite interval mapping

The expression (14), which is asymptotically equivalent to the score test statistic, involves standard normal random variables $W_i^{(\ell)}$, $i = 1, 2$; $\ell = 1, 2, \dots, k$. It can be shown (see Appendix 8.1 for details) that $W_1^{(\ell)}, W_2^{(\ell)}$, $\ell = 1, 2, \dots, k$, follow a joint multivariate normal distribution with mean $\mathbf{0}$, variances equal to 1, $\text{cov}(W_1^{(1)}, W_2^{(1)}) = 0$, and remaining covariances

$$(19) \quad \begin{aligned} \text{cov}(W_1^{(\ell)}, W_1^{(\ell'')}) &\approx \sqrt{(1 - r_{\ell'}) (1 - r_{\ell''})} (1 - 2r_{\ell', \ell''-1}) \\ \text{cov}(W_1^{(\ell)}, W_2^{(\ell'')}) &\approx \sqrt{(1 - r_{\ell'}) r_{\ell''}} (1 - 2r_{\ell', \ell''-1}), \\ \text{cov}(W_2^{(\ell)}, W_1^{(\ell'')}) &\approx -\sqrt{r_{\ell'} (1 - r_{\ell''})} (1 - 2r_{\ell', \ell''-1}) \\ \text{cov}(W_2^{(\ell)}, W_2^{(\ell'')}) &\approx -\sqrt{r_{\ell'} r_{\ell''}} (1 - 2r_{\ell', \ell''-1}). \end{aligned}$$

where $\ell' < \ell''$ and $r_{\ell', \ell''}$ is the recombination fraction between markers $\mathcal{M}^{\ell'}$ and $\mathcal{M}^{\ell''}$. Thus the variance-covariance matrix Σ of $W_1^{(1)}, W_2^{(1)}$ and $W_1^{(\ell)}, \ell = 2, 3, \dots, k$ is well defined. It can also be shown (see Appendix 8.2) that $W_2^{(\ell+1)}$ may be recursively calculated through

$$(20) \quad W_2^{(\ell+1)} = \frac{1}{\sqrt{r_{\ell+1}}} \left(\sqrt{1 - r_{\ell}} W_1^{(\ell)} - \sqrt{r_{\ell}} W_2^{(\ell)} - \sqrt{1 - r_{\ell+1}} W_1^{(\ell+1)} \right).$$

To simulate the null distribution of $\sup_{0 \leq d \leq D} Z^2(d)$, where $Z(d)$ is in (14), we can do the following:

- (i) Generate $(W_1^{(1)}, W_2^{(1)}, W_1^{(\ell)}, \ell = 2, 3, \dots, k) \sim N(\mathbf{0}, \Sigma)$;
- (ii) Compute $W_2^{(\ell+1)}, \ell = 1, 2, \dots, k - 1$ by (20);
- (iii) Using (18), for each ℓ , find

$$(21) \quad \sup_{\sum_{i=1}^{\ell-1} D_i \leq d \leq \sum_{i=1}^{\ell} D_i} Z^2(d),$$

and take the maximum over $\ell = 1, 2, \dots, k$.

Repeat the above steps many times (for example, 10,000 times), to obtain the approximate distribution of $\sup_{0 \leq d \leq D} Z^2(d)$, the asymptotic null distribution of the MLR test or the maximum of the square of score test statistics. The critical thresholds can be obtained accordingly.

Note that the random variables $W_1^{(1)}, W_2^{(1)}$ and $W_1^{(\ell)}, \ell = 2, 3, \dots, k$, can be calculate as $\mathbf{A}'\mathbf{Z}$, where \mathbf{A} is the Cholesky factorization of Σ (that is, $\mathbf{A}'\mathbf{A} = \Sigma$) and \mathbf{Z} follows multivariate standard normal distribution. Note that we only need to factor the matrix once for the entire simulation.

6 Examples

To examine the statistical properties of our score test statistic proposed, we perform a simulation study to determine the asymptotic null distribution and the critical threshold for declaring the existence of a significant QTL. Our results here are compared to those obtained by Lander and Botstein (1989), Doerge and Churchill (1996) and Piepho (2001). The score test statistic is further validated using a case study from a forest tree genome project.

6.1 Simulation

We use the same simulation design as used by Lander and Botstein (1989) to simulate a backcross mapping population of size 250. Thus, our results can be directly compared with those of Lander and Botstein. In our simulation study, equal spacing of the six markers is assumed throughout 12 chromosomes of 100 cM each. Five QTL are hypothesized to be located at genetic positions 70, 49, 27, 8 and 3 cM from the left end on the first five chromosomes with effects of 1.5, 1.25, 1.0, 0.75 and 0.50, respectively. For each individual, genotypes at the markers were generated assuming no interference. The corresponding quantitative trait was simulated by summing the five QTL effects and random standard normal noise.

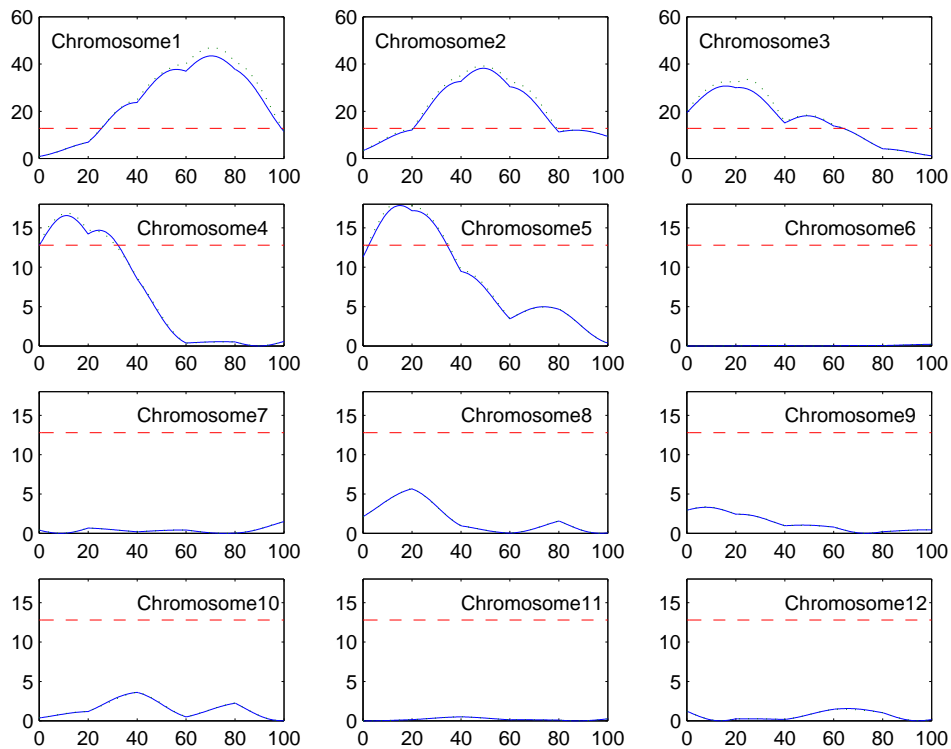


Figure 1: Score test for a simulated data set with 250 backcross progeny. As in Lander and Botstein (1989), we assumed that markers are spaced 20 cM throughout 12 chromosomes of 1 Morgan each. The QTLs, located at genetic positions 70, 49, 27, 8 and 3 cM from the left end on the first five chromosomes, were assumed to have effects 1.5, 1.25, 1.0, 0.75 and 0.50, respectively. For each individual, genotypes at markers were generated assuming no interference. The corresponding quantitative trait was simulated by summing the five QTL effects and random standard normal noise. The solid line is for score test statistics and the dotted line is for likelihood ratio test statistics ($-2\log(LR)$). The dashed line indicates the threshold of 12.83 at the 0.05 level based on the simulated asymptotic null distribution. All five QTLs attained this threshold.

Both the likelihood approach of Lander and Botstein (1989) and the score-statistic approach are used to analyze the simulated marker and phenotype data. The QTL likelihood profiles are drawn from the likelihood ratio test statistics and the score test statistics. As shown in Figure 1, these two statistics gave similar profiles across each of the simulated linkage groups. The 95% percentile of score test statistics based on 10,000 simulations under the null model is 12.83, which is used as the threshold for declaring a QTL at the significance level $\alpha = 0.05$.

Other approaches to estimate the threshold for the same data set, are summarized in the following table:

Reference	Method	.05 Threshold
This paper	Simulation	12.83
Lander and Botstein (1989)	Dense markers	14.51
Lander and Botstein (1989)	Bonferroni	11.16
Piepho (2001)	Approximation	12.96
Churchill and Doerge (1994)	Permutation Test	11.44

From the example, it can be seen that the threshold obtained by our method is close to Piepho's threshold and higher than that by Churchill and Doerge. Our large scale simulation study showed that thresholds by Bonferroni method and thresholds based on dense markers tend to under estimate and to overestimate the cutoff point, respectively.

6.2 A case study

The study we consider was derived from an interspecific hybridization of *Populus* (poplar). A *P. deltoides* clone (designated I-69) was used as a female parent to mate with a *P. euramericana* clone (designated I-45) as a male parent (Wu et al. 1992). Both *P. deltoides* I-69 and *P. euramericana* I-45 were selected at the Research Institute for Poplars in Italy in the 1950s and were introduced to China in 1972. A genetic linkage map has been constructed using 90 genotypes randomly selected from the 450 hybrids with random amplified polymorphic DNAs (RAPDs), amplified fraction length polymorphisms (AFLPs) and inter-simple sequence repeats (ISSRs) (Yin et al. 2002). This map comprises the 19 largest linkage groups for each parental map, which roughly represent 19 pairs of chromosomes. The 90 hybrid genotypes used for map construction were measured for wood density with wood samples collected from 11-year-old stems in a field trial in a completely randomized design.

The 19 linkage groups constructed for each parent were scanned for possible existence of QTL affecting wood density using both our score statistic and

the likelihood approach. We successfully detected a wood density QTL on linkage group, D17, composed of 16 normally segregated markers. The QTL profiles obtained from these two approaches, as shown in Figure 2, consistently exhibit a marked peak within a narrow marker interval AG/CGA-480 – AG/CGA-330. Both the score statistic value and log-likelihood ratio at the peak are greater than the threshold of 10.10 ($\alpha = 0.05$) obtained from the simulated asymptotic null distribution. Other approaches are also used to estimate the $\alpha = 0.05$ level threshold, which give 7.87 for permutation tests, 8.87 for Piepho's quick method, 8.73 for the Bonferroni method and 10.20 for Lander and Botstein's formula for dense markers. All these suggest that both approaches have consistently detected a significant QTL, located at the peak of the profiles, affecting wood density in hybrid poplars.

This example illustrated that the score statistic has the power to detect the hypothesized QTL as well as the likelihood approach. Also, the QTL profiles from our score statistic and the likelihood approach are similar. However, in the wood density example, the calculation of the score statistic threshold based on 10,000 simulations from the asymptotic null distribution only took 5 minutes, where the calculation of the threshold based on 10,000 permutation likelihood ratio tests took 4,100 minutes. (All calculations were performed using Matlab software on a Dell Inspiron 7000 with a 300 MHz CPU.)

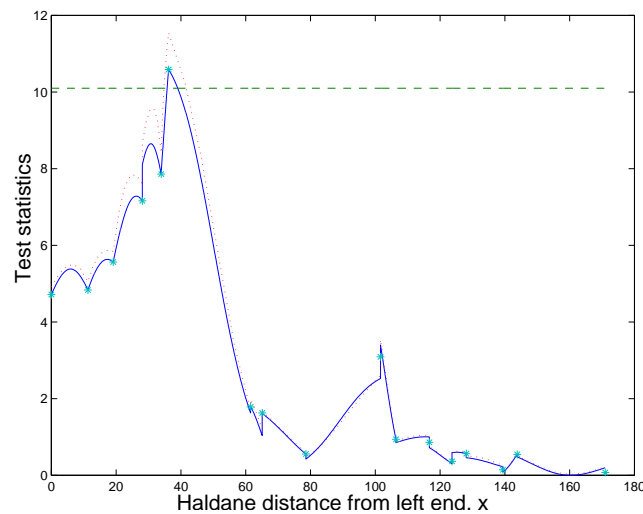


Figure 2: The value of the test statistics along chromosome 17, which has 16 markers. The solid line is for the score test, the dotted line is for the likelihood ratio test ($-2\log(LR)$), and the dashed line indicates the threshold of 10.10 at the 0.05 level based on the simulated asymptotic null distribution.

7 Discussion

In this article, we devise a score-statistic method for QTL mapping based on a genetic linkage map. The asymptotic null distribution of the score statistics is derived. The score test statistics $U(x)$ in equation (6) converges to its asymptotic equivalent $Z(x)$ quickly, because the convergence only depends on two things: (a) sample proportions in the four marker groups converge to the asymptotic proportions and (b) the standardized sample mean difference in equation (9) converges to standard normal distribution. In a variety of cases we studied, a sample size of 150 to 200 is usually sufficient to warrant the use of the threshold based on asymptotic null distribution.

Determining threshold of test statistics for QTL mapping is a first important step toward the identification of individual genes that contribute to genetic variation in a complex trait. However, this is not a trivial task because many factors, such as map density, marker distribution pattern, genome size, and the proportion of missing marker and phenotypic data, may affect the distribution of the test statistic. For an infinitely dense map, the test statistic may be approximated by an Ornstein-Uhlenbeck diffusion process for a backcross of large samples (Lander and Botstein 1989). This discovery was then used in more complicated genetic designs (Dupuis and Siegmund 1999; Zou et al. 2001). Rebai et al. (1994, 1995) derived a conservative threshold by considering the position of a QTL as a nuisance parameter present only under the alternative hypothesis in that the theoretical result of Davies (1977, 1987) can be at work. In this article, we devise a score-statistic method for QTL mapping based on a genetic linkage map. A score test statistic inherits the optimal property of the maximum likelihood method, yet it is much easier to compute. It is asymptotically equivalent to the likelihood ratio test statistic with asymptotic null distribution being the distribution of the maximum of the square of a well-defined Gaussian stochastic process. The critical thresholds obtained in Section 5 can be used for both the MLR test and the maximum of the square of score test statistics.

The computational method in Section 5 is feasible even for a large k , the number of molecular markers in a genetic linkage map. For a dense map, we can assume that a QTL is located at the same position as a marker. In this case, the numbers of recombinant groups $n_2^{(\ell)}$ and $n_3^{(\ell)}$ approaches zero for all intervals and our results on the asymptotic distribution of MLR test statistic in Section 4 reduce to that of Lander and Botstein (1989). Results from numerical analyses suggest that our score statistical model can be reasonably well applied in sparse maps, thus making the model more useful in general genetic studies. As an alternative to permutation tests or resampling strategies with no assumption about the distribution of the test statistic (Churchill and Doerge 1994; Zou et al. 2004), our model displays significant computing advantages.

Current genetic mapping has achieved a point at which almost every kind of biological traits and organisms can be studied by QTL mapping approaches. This requires that our model be of great usefulness and effectiveness in a broad spectrum of genetic settings. For outcrossing species, like forest trees, a full-sib family derived from two heterozygous parents provides an informative design for QTL mapping (Haley et al. 1994; Xu 1996; Wu et al. 2002; Lin et al. 2004). For humans and other species, QTL mapping may be based on a set of random samples drawn from a natural population in which associations between markers and QTL are studied in terms of linkage disequilibrium. For many complex traits, a one-QTL model is too simple to characterize their genetic architecture of high complexity. Traits may be likely to be controlled by a web of genes, each interacting with one another and environmental factor. With strong theoretical derivations, the proposed method can be readily extended to consider these more realistic situations.

Readers can find our Matlab code for computing the score statistic and the critical threshold at the webiste: <http://www.ehpr.ufl.edu/code/235>.

8 Appendix: Some Technical Details

8.1 The covariance of $Z(d')$ and $Z(d'')$

Assume (ℓ', x') and (ℓ'', x'') are associated with d' and d'' through (13), respectively. Since the covariance of $Z(d')$ and $Z(d'')$ is available in (11) for $\ell' = \ell''$, we derive the covariance for $\ell' < \ell''$.

Let $\mathcal{A}_{s,t;\ell',\ell''}$ denote the subset of individuals who belong to Group s in interval ℓ' and belong to Group t in interval ℓ'' , i.e.,

$$\mathcal{A}_{s,t;\ell',\ell''} = \{i : 1 \leq i \leq N, \quad G_i^{(\ell')} = s, \quad G_i^{(\ell'')} = t\}, 1 \leq s, t \leq 4, 1 \leq \ell' < \ell'' \leq k.$$

Denote the size of $\mathcal{A}_{s,t;\ell',\ell''}$ by $n_{s,t;\ell',\ell''}$, i.e.,

$$(A.1) \quad n_{s,t;\ell',\ell''} = \sum_{i=1}^N I(G_i^{(\ell')} = s)I(G_i^{(\ell'')} = t), 1 \leq s, t \leq 4, 1 \leq \ell' < \ell'' \leq k.$$

Let $\mathbf{N}^{(\ell',\ell'')} = (n_{s,t;\ell',\ell''})_{s,t=1,2,3,4}$ be the 4×4 matrix with s, t as the row index and column index, respectively. It is seen that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{N}^{(\ell',\ell'')} = \frac{1}{2} \begin{pmatrix} 1 - r_{\ell'} & 0 & 0 & 0 \\ 0 & r_{\ell'} & 0 & 0 \\ 0 & 0 & r_{\ell'} & 0 \\ 0 & 0 & 0 & 1 - r_{\ell'} \end{pmatrix}$$

$$(A.2) \quad \begin{aligned} & \times \begin{pmatrix} 1 - r_{\ell', \ell''-1} & 1 - r_{\ell', \ell''-1} & r_{\ell', \ell''-1} & r_{\ell', \ell''-1} \\ r_{\ell', \ell''-1} & r_{\ell', \ell''-1} & 1 - r_{\ell', \ell''-1} & 1 - r_{\ell', \ell''-1} \\ 1 - r_{\ell', \ell''-1} & 1 - r_{\ell', \ell''-1} & r_{\ell', \ell''-1} & r_{\ell', \ell''-1} \\ r_{\ell', \ell''-1} & r_{\ell', \ell''-1} & 1 - r_{\ell', \ell''-1} & 1 - r_{\ell', \ell''-1} \end{pmatrix} \\ & \times \begin{pmatrix} 1 - r_{\ell''} & 0 & 0 & 0 \\ 0 & r_{\ell''} & 0 & 0 \\ 0 & 0 & r_{\ell''} & 0 \\ 0 & 0 & 0 & 1 - r_{\ell''} \end{pmatrix}. \end{aligned}$$

In addition, we have results similar to (4):

$$(A.3) \quad \begin{aligned} \lim_{N \rightarrow \infty} \frac{n_1^{(\ell')}}{N} &= \lim_{N \rightarrow \infty} \frac{n_4^{(\ell')}}{N} = \frac{1}{2}(1 - r_{\ell'}), \\ \lim_{N \rightarrow \infty} \frac{n_2^{(\ell')}}{N} &= \lim_{N \rightarrow \infty} \frac{n_3^{(\ell')}}{N} = \frac{1}{2}r_{\ell'}, \\ \lim_{N \rightarrow \infty} \frac{n_1^{(\ell'')}}{N} &= \lim_{N \rightarrow \infty} \frac{n_4^{(\ell'')}}{N} = \frac{1}{2}(1 - r_{\ell''}), \\ \lim_{N \rightarrow \infty} \frac{n_2^{(\ell'')}}{N} &= \lim_{N \rightarrow \infty} \frac{n_3^{(\ell'')}}{N} = \frac{1}{2}r_{\ell''}. \end{aligned}$$

Using (15), (A.2), and (A.3), we have

$$(A.4) \quad \begin{aligned} \text{cov}(W_1^{(\ell')}, W_1^{(\ell'')}) &\approx \frac{N}{4\sigma^2} \sqrt{(1 - r_{\ell'})(1 - r_{\ell''})} \\ &\times \text{cov}(\bar{y}_4^{(\ell')} - \bar{y}_1^{(\ell')}, \bar{y}_4^{(\ell'')} - \bar{y}_1^{(\ell'')}) \\ &= \frac{N}{4} \sqrt{(1 - r_{\ell'})(1 - r_{\ell''})} \\ &\times \left(\frac{n_{4,4;\ell',\ell''}}{n_4^{(\ell')} n_4^{(\ell'')}} - \frac{n_{4,1;\ell',\ell''}}{n_4^{(\ell')} n_1^{(\ell'')}} - \frac{n_{1,4;\ell',\ell''}}{n_1^{(\ell')} n_4^{(\ell'')}} + \frac{n_{1,1;\ell',\ell''}}{n_1^{(\ell')} n_1^{(\ell'')}} \right) \\ &\approx \sqrt{(1 - r_{\ell'})(1 - r_{\ell''})(1 - 2r_{\ell', \ell''-1})} \end{aligned}$$

Similarly, we have

$$(A.5) \quad \begin{aligned} \text{cov}(W_1^{(\ell')}, W_2^{(\ell'')}) &\approx \sqrt{(1 - r_{\ell'})r_{\ell''}}(1 - 2r_{\ell', \ell''-1}), \\ \text{cov}(W_2^{(\ell')}, W_1^{(\ell'')}) &\approx -\sqrt{r_{\ell'}(1 - r_{\ell''})}(1 - 2r_{\ell', \ell''-1}) \\ \text{cov}(W_2^{(\ell')}, W_2^{(\ell'')}) &\approx -\sqrt{r_{\ell'}r_{\ell''}}(1 - 2r_{\ell', \ell''-1}). \end{aligned}$$

By (14), (A.4) and (A.5), for $\ell' < \ell''$ the covariance of $Z(d')$ and $Z(d'')$ is

$$\text{cov}(Z(d'), Z(d''))$$

$$\begin{aligned}
 &= (1 - 2r_{\ell', \ell''-1}) \\
 &\quad \times \left[(1 - 2\theta_1')(1 - 2\theta_1'')(1 - r_{\ell'})(1 - r_{\ell''}) \right. \\
 &\quad \left. + (1 - 2\theta_1')(1 - 2\theta_2'')(1 - r_{\ell'})r_{\ell''} - (1 - 2\theta_2')(1 - 2\theta_1'')r_{\ell'}(1 - r_{\ell''}) \right. \\
 &\quad \left. - (1 - 2\theta_2')(1 - 2\theta_2'')r_{\ell'}r_{\ell''} \right] / \\
 &\quad \sqrt{(1 - r_{\ell'})(1 - 2\theta_1')^2 + r_{\ell'}(1 - 2\theta_2')^2} / \\
 &\quad \sqrt{(1 - r_{\ell''})(1 - 2\theta_1'')^2 + r_{\ell''}(1 - 2\theta_2'')^2},
 \end{aligned}
 \tag{A.6}$$

where $r_{\ell' \ell''}$ denotes the recombination fraction between two markers $\mathcal{M}^{\ell'}$ and $\mathcal{M}^{\ell''}$ ($\ell' < \ell''$), where $r_{\ell-1, \ell} = r_{\ell}$ and $r_{\ell, \ell} = 0$.

8.2 The linear relationship between $W_1^{(\ell)}, W_2^{(\ell)}, W_1^{(\ell+1)}$, and $W_2^{(\ell+1)}$ in (20)

From (A.4) and (A.5), the variance-covariance matrix of $W_1^{(\ell)}, W_2^{(\ell)}, W_1^{(\ell+1)}$, and $W_2^{(\ell+1)}$ has the form

$$\begin{pmatrix} I_2 & \Gamma \\ \Gamma' & I_2 \end{pmatrix}, \text{ where } \Gamma = \begin{pmatrix} \sqrt{(1 - r_{\ell})(1 - r_{\ell+1})} & \sqrt{(1 - r_{\ell})r_{\ell+1}} \\ -\sqrt{r_{\ell}(1 - r_{\ell+1})} & -\sqrt{r_{\ell}r_{\ell+1}} \end{pmatrix}.$$

It can be seen that the above matrix is singular and hence (20) holds.

The relationship (20) has an intuitive interpretation. Denote by $\mathcal{B}_{s; \ell}$ the subset of individuals who belong to Group s in interval ℓ , i.e.,

$$\mathcal{B}_{s; \ell} = \{i : 1 \leq i \leq N, G_{i\ell} = s\}, \quad 1 \leq s \leq 4; \quad 1 \leq \ell \leq k.$$

Consider two consecutive intervals $[M_{\ell-1}, M_{\ell}]$ and $[M_{\ell}, M_{\ell+1}]$. For any individual in Group 1 or Group 3 (Group 2 or Group 4), the marker condition is (M^{ℓ}, m^{ℓ}) ((m^{ℓ}, m^{ℓ})) at marker M_{ℓ} . Thus this individual must belong to Group 1 or Group 2 (Group 3 or Group 4) in interval $[M_{\ell}, M_{\ell+1}]$. Therefore,

$$\mathcal{B}_{1; \ell} \cup \mathcal{B}_{3; \ell} = \mathcal{B}_{1; \ell+1} \cup \mathcal{B}_{2; \ell+1}$$

and similarly,

$$\mathcal{B}_{2; \ell} \cup \mathcal{B}_{4; \ell} = \mathcal{B}_{3; \ell+1} \cup \mathcal{B}_{4; \ell+1}.$$

Consequently,

$$n_1^{(\ell)} \bar{y}_1^{(\ell)} + n_3^{(\ell)} \bar{y}_3^{(\ell)} = n_1^{(\ell+1)} \bar{y}_1^{(\ell+1)} + n_2^{(\ell+1)} \bar{y}_2^{(\ell+1)}$$

and

$$(A.7) \quad n_2^{(\ell)} \bar{y}_2^{(\ell)} + n_4^{(\ell)} \bar{y}_4^{(\ell)} = n_3^{(\ell+1)} \bar{y}_3^{(\ell+1)} + n_4^{(\ell+1)} \bar{y}_4^{(\ell+1)}.$$

Combining (15), (A.3), and (A.7) we see that

$$\begin{aligned} & \sqrt{1-r_\ell} W_1^{(\ell)} - \sqrt{r_\ell} W_2^{(\ell)} - \sqrt{1-r_{\ell+1}} W_1^{(\ell+1)} - \sqrt{r_{\ell+1}} W_2^{(\ell+1)} \\ \approx & \frac{\sqrt{N}}{2} \left[\left(1-r_\ell\right) \left(\bar{y}_4^{(\ell)} - \bar{y}_1^{(\ell)}\right) - r_\ell \left(\bar{y}_3^{(\ell)} - \bar{y}_2^{(\ell)}\right) \right. \\ & \left. - \left(1-r_{\ell+1}\right) \left(\bar{y}_4^{(\ell+1)} - \bar{y}_1^{(\ell+1)}\right) - r_{\ell+1} \left(\bar{y}_3^{(\ell+1)} - \bar{y}_2^{(\ell+1)}\right) \right] \\ \approx & \frac{1}{\sqrt{N}} \left[\left(n_4^{(\ell)} y_4^{(\ell)} - n_1^{(\ell)} \bar{y}_1^{(\ell)}\right) - \left(n_3^{(\ell)} \bar{y}_3^{(\ell)} - n_2^{(\ell)} \bar{y}_2^{(\ell)}\right) \right. \\ & \left. - \left(n_4^{(\ell+1)} \bar{y}_4^{(\ell+1)} - n_1^{(\ell+1)} \bar{y}_1^{(\ell+1)}\right) - \left(n_3^{(\ell+1)} \bar{y}_3^{(\ell+1)} - n_2^{(\ell+1)} \bar{y}_2^{(\ell+1)}\right) \right] \\ = & 0, \end{aligned}$$

which provides a justification of (20).

8.3 Equivalence of the suprema of the MLR and score statistic

In this section we demonstrate the equivalence of the suprema of the MLR statistic and the maximum of the square of score test statistic over one marker interval $[M^0, M^1]$. The equivalence over multiple intervals follows immediately. Specifically, we prove Theorem 1 of Section 3:

$$\lim_{N \rightarrow \infty} \left| \sup_{x \in [0, D]} 2\{\ell_N(\hat{\eta}^{(N)}(x), x) - \ell_N(\hat{\eta}_0^{(N)})\} - \sup_{x \in [0, D]} U^2(x) \right| = 0 \text{ a.s.}$$

where the log likelihood ℓ and the score statistics $U(x)^2$ are described in Section 3, and the almost sure convergence is with respect to the distribution of the phenotypic trait under H_0 .

The proof of Theorem 1 is based on two lemmas. We will state the two lemmas and then, using them, will prove the theorem. In separate sections we will prove the two lemmas, and two other needed results. Before stating the lemma we review some notation.

The genotype variable G^ℓ is defined at the beginning of Section 4.1. The superscript “ ℓ ” will be omitted since we focus on only one interval $[M^0, M^1]$. The

probability function of G is defined by

$$p(g) = P(G = g) = \begin{cases} \frac{1}{2}(1-r), & g = 1 \\ \frac{1}{2}r, & g = 2 \\ \frac{1}{2}r, & g = 3 \\ \frac{1}{2}(1-r), & g = 4. \end{cases}$$

To express the likelihood function in a compact form, we define

$$\varphi(y; \lambda, \theta) = (1 - \theta)e^{-\frac{1}{2\sigma^2}(y-\mu_1)^2} + \theta e^{-\frac{1}{2\sigma^2}(y-\mu_2)^2}.$$

where $\lambda = (\mu_1, \mu_2, \sigma)^T$. We will also use the alternative parameterization $\mu_1 = \mu - \Delta$, $\mu_2 = \mu + \Delta$ and $\eta = (\Delta, \mu, \sigma)^T$

The density function of the phenotype variable Y conditional on $G = j$ is

$$f_{Y|G=j}(y; \lambda, x) = \frac{1}{\sqrt{2\pi\sigma}} \varphi(y; \lambda, \theta_j^*),$$

where $\theta_j^* = \theta_1, \theta_2, 1 - \theta_2, 1 - \theta_1$, for $j = 1, 2, 3, 4$, respectively, and θ_1, θ_2 are functions of x as defined in (3).

Assume the observations are $(Y_i, G_i) = (y_i, g_i)$, $i = 1, \dots, N$, and are arranged so that the first n_1 observations belong to group 1; the next n_2 belong to group 2; and so on. The likelihood functions conditional on the genotype variable are

$$f(y^{(N)}; \lambda, x) = \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^N \prod_{i=1}^{N_1} \varphi(y_i; \lambda, \theta_1) \prod_{i=N_1+1}^{N_2} \varphi(y_i; \lambda, \theta_2) \prod_{i=N_2+1}^{N_3} \varphi(y_i; \lambda, 1 - \theta_2) \prod_{i=N_3+1}^N \varphi(y_i; \lambda, 1 - \theta_1),$$

where $y^{(N)} = (y_1, \dots, y_N)$. The full likelihood function for observations $(Y_i, G_i) = (y_i, g_i)$, $i = 1, \dots, N$, is

$$(A.8) \quad L_N(\lambda; x) = f(y^{(N)}; \lambda, x) \prod_{j=1}^4 [p(j)]^{n_j}.$$

Since the second factor does not contain parameters of interest, we consider the log likelihood function for $y^{(N)}$ conditional on the genotype variable $\ell_N(\lambda; x) = \log f(y^{(N)}; \lambda, x)$, which is given in (5).

For a fixed $x \in [0, D]$, the mle's of $\mu_1, \mu_2, \sigma, \Delta, \mu, \lambda, \eta$ are denoted, respectively, by $\hat{\mu}_1^{(N)}(x), \hat{\mu}_2^{(N)}(x), \hat{\sigma}^{(N)}(x), \hat{\Delta}^{(N)}(x), \hat{\mu}^{(N)}(x),$

$$\begin{aligned}\hat{\lambda}^{(N)}(x) &= \left(\hat{\mu}_1^{(N)}(x), \hat{\mu}_2^{(N)}(x), \hat{\sigma}^{(N)}(x) \right)^T, \\ \hat{\eta}^{(N)}(x) &= \left(\hat{\Delta}^{(N)}(x), \hat{\mu}^{(N)}(x), \hat{\sigma}^{(N)}(x) \right)^T,\end{aligned}$$

with $\hat{\mu}_1^{(N)}(x) = \hat{\mu}^{(N)}(x) - \hat{\Delta}^{(N)}(x)$ and $\hat{\mu}_2^{(N)}(x) = \hat{\mu}^{(N)}(x) + \hat{\Delta}^{(N)}(x)$.

Let μ_0 be the true common value of μ_1 and μ_2 and σ_0 be the true value of σ under H_0 , and denote the normal density function with mean μ_0 , and variance σ_0^2 by $n(y; \mu_0, \sigma_0)$. Under $\lambda = \lambda_0 = (\mu_0, \mu_0, \sigma)$, the densities of (Y, G) and Y are

$$f_{Y,G}(y, g; \lambda_0, x) = n(y; \mu_0, \sigma_0)p(g) \text{ and } f_Y(y; \lambda_0, x) = n(y; \mu_0, \sigma_0),$$

where neither depend on x under H_0 .

From now on, all probabilities $P_0(\cdot)$, expectations $E_0(\cdot)$, variances $Var_0(\cdot)$, and all probability statements such as “converges in probability”, “converges almost surely (a.s.)”, and “with probability one” are under $\lambda = \lambda_0$.

Lemma 1 (Information Matrix) *The information matrix $I(x)$ is given by*

$$(A.9) \quad I(x) = -E_0 \frac{1}{N} \begin{pmatrix} \frac{\partial^2 \ell}{\partial \Delta^2} & \frac{\partial^2 \ell}{\partial \Delta \partial \mu} & \frac{\partial^2 \ell}{\partial \Delta \partial \sigma} \\ \frac{\partial^2 \ell}{\partial \mu \partial \Delta} & \frac{\partial^2 \ell}{\partial \mu^2} & \frac{\partial^2 \ell}{\partial \mu \partial \sigma} \\ \frac{\partial^2 \ell}{\partial \sigma \partial \Delta} & \frac{\partial^2 \ell}{\partial \sigma \partial \mu} & \frac{\partial^2 \ell}{\partial \sigma^2} \end{pmatrix}_{\eta = \eta_0} = \frac{1}{\sigma_0^2} \begin{pmatrix} \gamma(x) & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

where $\gamma(x) = (1 - 2\theta_1)^2(1 - r) + (1 - 2\theta_2)^2r$ (recall that θ_1 and θ_2 depend on x ; see (3)).

A sequence of random variables $X_N(\eta, x)$ is *almost surely (a.s.) uniformly bounded* with respect to $\eta \in \Gamma = \{(\Delta, \mu, \sigma) : |\mu - \Delta| \leq \xi, |\mu + \Delta| < \xi, 0 < \sigma_1 \leq \sigma \leq \sigma_2 < \infty\}$ and $x \in [0, D]$ if for any $\varepsilon > 0$, there exists a constant K such that

$$P_0 \left[\sup_{\substack{\eta \in \Gamma \\ x \in [0, D] \\ N=1, 2, \dots}} |X_N(\eta, x)| \leq K \right] > 1 - \varepsilon.$$

A sequence of random variables $X_N(x)$ converges to a random variable $Y(x)$ *almost surely and uniformly* for $x \in [0, D]$ if

$$\lim_{N \rightarrow \infty} \sup_{x \in [0, D]} |X_N(x) - Y(x)| = 0 \text{ a.s.}$$

Two sequences of random variables $X_N(s)$ and $Y_N(x)$ are called *asymptotically equivalent*, denoted by $X_N(x) \doteq Y_N(x)$, if $X_N(x) - Y_N(x)$ converges to zero a.s. and uniformly for $x \in [0, D]$.

Lemma 2 (Higher Order Terms)

(A) All elements of the $3 \times 3 \times 3$ matrix

$$(A.10) \quad \frac{1}{N} \frac{\partial^3}{\partial \eta^3} \ell_N(\eta, x)$$

are a.s. bounded uniformly for $\eta \in \Gamma$ and $x \in [0, D]$.

(B)

$$\sqrt{N} [I(x)]^{\frac{1}{2}} (\hat{\eta}^{(N)}(x) - \eta_0) \doteq \frac{1}{\sqrt{N}} [I(x)]^{-\frac{1}{2}} \frac{\partial}{\partial \eta} \ell_N(\eta, x) \Big|_{\eta=\eta_0}.$$

8.3.1 Proof of Theorem 1:

It is sufficient to show that

$$(A.11) \quad 2\{\ell_N(\hat{\eta}^{(N)}(x), x) - \ell_N(\hat{\eta}_0^{(N)})\} \doteq U^2(x).$$

Recall from Section 3 that, under H_0 , we can write $U(x)$ as

$$U(x) = \frac{u(\hat{\mu}_0^{(N)}, \hat{\sigma}_0^{(N)}, x)}{\sqrt{\text{Var}_0[u(\hat{\mu}_0^{(N)}, \hat{\sigma}_0^{(N)}, x)]}}.$$

It follows that,

$$\begin{aligned} & \frac{u(\mu_0, \sigma_0, x) - u(\hat{\mu}_0^{(N)}, \hat{\sigma}_0^{(N)}, x)}{\sqrt{\text{Var}_0[u(\hat{\mu}_0^{(N)}, \hat{\sigma}_0^{(N)}, x)]}} \\ &= \frac{1}{\sigma_0^2} \frac{(\hat{\mu}_0^{(N)} - \mu_0)[(1 - 2\theta_1)(n_4 - n_1) + (1 - 2\theta_2)(n_3 - n_2)]}{\sqrt{\text{Var}_0[u(\hat{\mu}_0^{(N)}, \hat{\sigma}_0^{(N)}, x)]}} \\ &+ \left(\frac{1}{\sigma_0^2} - \frac{1}{(\hat{\sigma}_0^{(N)})^2} \right) \frac{u(\hat{\mu}_0^{(N)}, \hat{\sigma}_0^{(N)}, x)}{\sqrt{\text{Var}_0[u(\hat{\mu}_0^{(N)}, \hat{\sigma}_0^{(N)}, x)]}} \\ &\doteq 0, \end{aligned}$$

and hence

$$U(x) \doteq \frac{u(\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0, x)}{\sqrt{\text{Var}_0[u(\hat{\boldsymbol{\mu}}_0^{(N)}, \hat{\boldsymbol{\sigma}}_0^{(N)}, x)]}}$$

Next, using the facts that

$$\frac{1}{N} \text{Var}_0[u(\hat{\boldsymbol{\mu}}_0^{(N)}, \hat{\boldsymbol{\sigma}}_0^{(N)}, x)] \doteq \frac{1}{\sigma_0^2} \gamma(x) \text{ and } \left. \frac{\partial}{\partial \Delta} \ell_N(\boldsymbol{\eta}, x) \right|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0} = u(\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0, x),$$

we can write

$$\begin{aligned} \text{(A.12)} \quad U(x) &\doteq \frac{u(\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0, x)}{\sqrt{\text{Var}_0[u(\hat{\boldsymbol{\mu}}_0^{(N)}, \hat{\boldsymbol{\sigma}}_0^{(N)}, x)]}} \doteq \frac{\left. \frac{\partial}{\partial \Delta} \ell_N(\boldsymbol{\eta}, x) \right|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0}}{\sqrt{\frac{N}{\sigma_0^2} \gamma(x)}} \\ &\doteq \frac{\sqrt{N}}{\sigma_0} \hat{\Delta}^{(N)}(x) \sqrt{\gamma(x)}, \end{aligned}$$

where the last equality follows from Lemma 2 (B).

Next we consider the likelihood ratio statistic. Using a Taylor expansion, together with Lemma 2 (A), we can write

$$\text{(A.13)} \quad 2[\ell_N(\hat{\boldsymbol{\eta}}^{(N)}(x), x) - \ell_N(\boldsymbol{\eta}_0)] \doteq N(\hat{\boldsymbol{\eta}}^{(N)}(x) - \boldsymbol{\eta}_0)^T I(x)(\hat{\boldsymbol{\eta}}^{(N)}(x) - \boldsymbol{\eta}_0),$$

and, under H_0 , since there is no dependence on x , we can use classical results to write

$$\text{(A.14)} \quad 2[\ell_N(\hat{\boldsymbol{\eta}}_0^{(N)}) - \ell_N(\boldsymbol{\eta}_0)] \doteq \frac{N}{\sigma_0^2} (\hat{\mathbf{v}}_0^{(N)} - \mathbf{v}_0)^T \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} (\hat{\mathbf{v}}_0^{(N)} - \mathbf{v}_0),$$

where $\mathbf{v} = (\boldsymbol{\mu}, \boldsymbol{\sigma})^T$, $\mathbf{v}_0 = (\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0)^T$, and $\hat{\mathbf{v}}_0^{(N)} = (\hat{\boldsymbol{\mu}}_0^{(N)}, \hat{\boldsymbol{\sigma}}_0^{(N)})^T$.

Using the derivatives of the log likelihood (see the proof of Lemma 1 we have

$$\sqrt{N}(\hat{\mathbf{v}}_0^{(N)} - \mathbf{v}_0) \doteq \frac{\sigma_0^2}{\sqrt{N}} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \left. \frac{\partial}{\partial \mathbf{v}} \ell_N(\boldsymbol{\eta}, x) \right|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0} \doteq \sqrt{N} \left[\begin{pmatrix} \hat{\boldsymbol{\mu}}^{(N)}(x) \\ \hat{\boldsymbol{\sigma}}^{(N)}(x) \end{pmatrix} - \mathbf{v}_0 \right],$$

where the last equality follows from Lemma 2 (B). The likelihood ratio statistic is obtained by subtracting (A.14) from (A.13), which results in

$$\begin{aligned} 2[\ell(y^{(N)}; \boldsymbol{\eta}^{(N)}(x), x) - \ell(y^{(N)}; \hat{\boldsymbol{\eta}}_0^{(N)})] &\doteq \frac{N}{\sigma_0^2} (\hat{\Delta}^{(N)}(x))^2 \gamma(x) \\ &\doteq U^2(x), \end{aligned}$$

where the last equality follows from (A.12). The proof of Theorem 1 is now complete.

8.3.2 Proof of Lemma 1

The proof of Lemma 1 is quite straightforward, and we only sketch it here. Evaluation of the derivatives of the log likelihood is long and tedious, and we omit the details. However, for completeness, we give the derivatives evaluated at the null parameter.

$$\begin{aligned} \left. \frac{\partial \ell_N}{\partial \Delta} \right|_{\eta=\eta_0} &= \frac{1}{\sigma_0^2} \left[-(1-2\theta_1) \sum_1^{N_1} (y_i - \mu_0) - (1-2\theta_2) \sum_{N_1+1}^{N_2} (y_i - \mu_0) \right. \\ &\quad \left. + (1-2\theta_2) \sum_{N_2+1}^{N_3} (y_i - \mu_0) + (1-2\theta_1) \sum_{N_3+1}^N (y_i - \mu_0) \right], \\ \left. \frac{\partial \ell_N}{\partial \mu} \right|_{\eta=\eta_0} &= \frac{1}{\sigma_0^2} \sum_{i=1}^N (y_i - \mu_0), \\ \left. \frac{\partial \ell_N}{\partial \sigma} \right|_{\eta=\eta_0} &= -\frac{N}{\sigma_0} + \frac{1}{\sigma_0^3} \sum_{i=1}^N (y_i - \mu_0)^2 \\ \left. \frac{\partial^2 \ell_N}{\partial \Delta^2} \right|_{\eta=\eta_0} &= -\frac{N}{\sigma_0^2} + \frac{1}{\sigma_0^4} \left[\sum_{i=1}^N (y_i - \mu_0)^2 - (1-2\theta_1)^2 \sum_{i=1}^{N_1} (y_i - \mu_0)^2 \right. \\ &\quad \left. - (1-2\theta_2)^2 \sum_{i=N_1+1}^{N_2} (y_i - \mu_0)^2 - (1-2\theta_2)^2 \sum_{i=N_2+1}^{N_3} (y_i - \mu_0)^2 \right. \\ &\quad \left. - (1-2\theta_1)^2 \sum_{i=N_3+1}^N (y_i - \mu_0)^2 \right], \\ \left. \frac{\partial^2 \ell_N}{\partial \mu^2} \right|_{\eta=\eta_0} &= -\frac{N}{\sigma_0^2}, \\ \left. \frac{\partial^2 \ell_N}{\partial \sigma^2} \right|_{\eta=\eta_0} &= \frac{N}{\sigma_0^2} - \frac{3}{\sigma_0^4} \sum_{i=1}^n (y_i - \mu_0)^2, \\ \left. \frac{\partial^2 \ell_N}{\partial \Delta \partial \mu} \right|_{\eta=\eta_0} &= \frac{1}{\sigma_0^2} \left[(1-2\theta_1)(n_1 - n_4) + (1-2\theta_2)(n_2 - n_3) \right], \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell_N}{\partial \Delta \partial \sigma} \Big|_{\eta=\eta_0} &= \frac{2}{\sigma_0^3} \left[(1-2\theta_1) \sum_1^{N_1} (y_i - \mu_0) + (1-2\theta_2) \sum_{N_1+1}^{N_2} (y_i - \mu_0) \right. \\ &\quad \left. - (1-2\theta_2) \sum_{N_2+1}^{N_3} (y_i - \mu_0) - (1-2\theta_1) \sum_{N_3+1}^N (y_i - \mu_0) \right], \\ \frac{\partial^2 \ell_N}{\partial \mu \partial \sigma} \Big|_{\eta=\eta_0} &= \frac{-2}{\sigma_0^3} \sum_{i=1}^N (y_i - \mu_0). \end{aligned}$$

8.3.3 Proof of Lemma 2

The proof of Lemma 2 relies on two other results, which we state here and prove in the next sections.

Lemma 3 (Boundedness of MLE)

- (A) For any $x \in [0, D]$, the mle $\hat{\lambda}^{(N)}(x)$ exists with probability one.
- (B) For any $\delta > 0$, there exist a measurable subset A of the sample space with $P(A) \geq 1 - \delta$, real numbers σ_1, σ_2 satisfying $0 < \sigma_1 < \sigma_0 < \sigma_2 < \infty$, a real number $\xi > 0$ and an integer $M > 0$, such that for any sample point in A , $x \in [0, D]$, and $N > M$, we have

$$\left| \hat{\mu}_j^{(N)}(x) \right| \leq \xi, \quad j = 1, 2, \text{ and } \sigma_1 \leq \left| \hat{\sigma}^{(N)}(x) \right| \leq \sigma_2,$$

where $\delta, A, \sigma_1, \sigma_2, \xi$, and M are independent of x .

Lemma 4 (Uniform convergence of the MLE) The MLEs $\hat{\mu}_1^{(N)}(x), \hat{\mu}_2^{(N)}(x)$, and $\hat{\sigma}^{(N)}(x)$ converge uniformly to μ_0, μ_0 , and σ_0 , respectively, a.s. and uniformly for $x \in [0, D]$.

Since the second order moments of the elements in (A.10) are bounded uniformly for $\eta \in \Gamma$ and $x \in [0, D]$, the conclusion in (A) follows from Chebyshev's inequality, which also implies that

$$(A.15) \quad \frac{1}{\sqrt{N}} \frac{\partial}{\partial \eta} \ell_N(\eta, x) \Big|_{\eta=\eta_0}$$

is a.s. bounded uniformly for $x \in [0, D]$, since the elements in (A.15) have mean zero and variances bounded uniformly for $x \in [0, D]$. Applying Chebyshev's inequality once more, we can show that

$$(A.16) \quad \frac{1}{N} \frac{\partial^2}{\partial \eta^2} \ell_N(\eta, x) \Big|_{\eta=\eta_0} \doteq I(x),$$

since the right-hand side of (A.16) is the mean of the left-hand side and the variances of the elements of the left-hand side converge to zero uniformly for $x \in [0, D]$.

From the Taylor expansion

$$\begin{aligned} & \frac{\partial}{\partial \eta} \ell_N(\hat{\eta}^{(N)}(x), x) - \frac{\partial}{\partial \eta} \ell_N(\eta, x) \Big|_{\eta=\eta_0} \\ &= \frac{\partial}{\partial \eta^2} \ell_N(\eta, x) \Big|_{\eta=\eta_0} (\hat{\eta}^{(N)}(x) - \eta_0) + \frac{1}{2} \frac{\partial^3}{\partial \eta^3} \ell_N(\eta^*(x), x) [\hat{\eta}^{(N)}(x) \\ & \quad - \eta_0, \hat{\eta}^{(N)}(x) - \eta_0], \end{aligned}$$

where $\eta^*(x) = \eta_0 + \zeta(\hat{\eta}^{(N)}(x) - \eta_0)$ with $0 < \zeta < 1$, we have

$$\begin{aligned} & \sqrt{N} [I(x)]^{\frac{1}{2}} (\hat{\eta}^{(N)}(x) - \eta_0) - \frac{1}{\sqrt{N}} [I(x)]^{-\frac{1}{2}} \frac{\partial}{\partial \eta} \ell_N(\eta, x) \Big|_{\eta=\eta_0} \\ &= \{I(x) \left[\frac{1}{N} \frac{\partial^2}{\partial \eta^2} \ell_N(\eta, x) \Big|_{\eta=\eta_0} + \frac{1}{2} \frac{1}{N} \frac{\partial^3}{\partial \eta^3} \ell_N(\eta^*(x), x) (\hat{\eta}^{(N)}(x) - \eta_0) \right]^{-1} - I\} \\ (A.17) \quad & \cdot \frac{1}{\sqrt{N}} I^{-\frac{1}{2}}(x) \frac{\partial}{\partial \eta} \ell_N(\eta, x) \Big|_{\eta=\eta_0}, \end{aligned}$$

where I is the 3×3 identity matrix. The argument in the first paragraph of this proof and Lemmas 3 and 4 imply that the right-hand side of (A.17) is asymptotically equivalent to 0.

8.3.4 Proof of Lemma 3

Under H_0 , the mle's of $\mu_1 = \mu_2 = \mu_0$ and $\sigma^2 = \sigma_0^2$ are the sample mean and sample variance, respectively, and let $\hat{\lambda}_0^{(N)} = (\hat{\mu}_0^{(N)}, \hat{\mu}_0^{(N)}, \hat{\sigma}_0^{(N)})$. Clearly,

$$(A.18) \quad \sup_{\lambda} f(y^{(N)}; \lambda, x) \geq f(y^{(N)}; \hat{\lambda}_0^{(N)}, x) = \left(\frac{1}{\sqrt{2\pi} \hat{\sigma}_0^{(N)}} \right)^N e^{-\frac{N}{2}}.$$

For any positive number b , the variance of Y conditional on $|Y| \leq b$ is denoted by

$$\sigma_b^2 = \text{Var}_0(Y | |Y| \leq b),$$

which converges to $\sigma_0^2 = \text{Var}_0(Y)$ as $b \rightarrow \infty$. For any $\varepsilon > 0$, select a positive number b such that

$$P_0(|Y| \leq b) > 1 - \varepsilon \text{ and } \sigma_b > (1 - \varepsilon)\sigma_0.$$

The sample mean and sample variance using y_i 's with $|y_i| \leq b$ are denoted by $\bar{y}_b^{(N)}$ and $(\hat{\sigma}_b^{(N)})^2$, respectively. We have

$$(A.19) \quad \lim_{N \rightarrow \infty} \hat{\sigma}_b^{(N)} = \sigma_b > (1 - \varepsilon)\sigma_0.$$

Now choose intervals $I_1 = [-2, -1]$, $I_2 = [0, 1]$, $I_3 = [2, 3]$, and $I_4 = [-b, b]$. The choice of these intervals is somewhat arbitrary, we are only using $I_1 - I_3$ to get a lower bound on σ^2 , and I_4 to get a bound on μ .

Let $W_j^{(N)}$ be the number of Y_1, \dots, Y_N in I_j , $j = 1, 2, 3, 4$. Denote by $n_j^{(b)}$ the number of Y_1, \dots, Y_N in $[-b, b]$ and in group j , $j = 1, 2, 3, 4$. Clearly,

$$W_4^{(N)} = \sum_{j=1}^4 n_j^{(b)}.$$

Following from the fact that

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{W_j^{(N)}}{N} &= P_0(Y \in I_j), \quad j = 1, 2, 3, 4, \\ \lim_{N \rightarrow \infty} \hat{\sigma}_0^{(N)} &= \sigma_0, \end{aligned}$$

and the limits in (4) and (A.19), there exist a measurable subset A of the sample space with $P_0(A) > 1 - \delta$, a positive number α , and an integer M such that for any sample point in A and $N > M$,

$$(A.20) \quad \begin{aligned} W_j^{(N)} &> \alpha N, \quad j = 1, 2, 3, \\ W_4^{(N)} &> (1 - \varepsilon)N, \\ \hat{\sigma}_0^{(N)} &< \frac{3}{2}\sigma_0, \\ n_j &> (1 - \varepsilon)\frac{1}{2}(1 - r)N, \quad j = 1, 4, \\ n_j &> (1 - \varepsilon)\frac{1}{2}rN, \quad j = 2, 3, \\ \hat{\sigma}_b^{(N)} &> (1 - \varepsilon)\sigma_0. \end{aligned}$$

Hereafter in the proof of Lemma 3, we assume that the sample point is in A and $N > M$ and then, (A.20) hold. We then have,

$$(A.21) \quad \sup_{\lambda} f(y^{(N)}; \lambda, x) \geq \left(\frac{1}{\sqrt{2\pi\hat{\sigma}_0^{(N)}}} \right)^N e^{-\frac{N}{2}} > \left(\frac{2e^{-\frac{1}{2}}}{3\sqrt{2\pi\sigma_0}} \right)^N.$$

Now, because of the gaps between $I_1 - I_3$, for any real values of μ_1 and μ_2 , there exists an interval, I^* , among I_1, I_2 and I_3 such that $\inf \left\{ |y - \mu_j| : j = 1, 2, y \in I^* \right\} \geq \frac{1}{2}$. This fact, together with the inequality that $f(y^{(N)}; \lambda, x)$ is less than or equal to

$$(A.22) \quad \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^N \prod_{y_i \in I_1 \cup I_2 \cup I_3} \exp \left\{ -\frac{1}{2\sigma^2} \min \left[(y_i - \mu_1)^2, (y_i - \mu_2)^2 \right] \right\}$$

and (A.20), imply that

$$(A.23) \quad f(y^{(N)}; \lambda, x) \leq \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^N \left[\exp \left(-\frac{1}{2\sigma^2} \cdot \frac{1}{4} \right) \right]^{\alpha N} = \left(\frac{e^{-\frac{\alpha}{8\sigma^2}}}{\sqrt{2\pi\sigma}} \right)^N.$$

As $\sigma \rightarrow 0$ or $\sigma \rightarrow \infty$, the quantity in parentheses converges to zero and then (A.23) < right-hand side of (A.21). As a result, the supremum of $f(\cdot)$ can be achieved in the area of $0 < \sigma_1 \leq \sigma \leq \sigma_2 < \infty$, where σ_1 and σ_2 are two constants independent of x . Furthermore, for any $y_i \in I_1 \cup I_2 \cup I_3$, the limit

$$\lim_{\substack{|\mu_1| \rightarrow \infty \\ |\mu_2| \rightarrow \infty}} \min \left[(y_i - \mu_1)^2, (y_i - \mu_2)^2 \right] = \infty,$$

implies that the right-hand side of (A.22) < right-hand side of (A.21) for a bounded σ as both $|\mu_1|$ and $|\mu_2|$ converge to ∞ . Therefore, we have shown that the supremum of $f(\cdot)$ can be achieved in the area where $\sigma_1 \leq \sigma \leq \sigma_2$ and at least one of μ_1 and μ_2 is bounded.

Now we show that for $|\mu_1| \leq b$ and as $|\mu_2| \rightarrow \infty$, the value of the likelihood function is smaller than the right-hand side of (A.21). Following from the inequalities (A.20), we have

$$n_j^{(b)} > n_j - \varepsilon N > (1 - \varepsilon) \frac{1}{2} (1 - r) N - \varepsilon N \geq (1 - k\varepsilon) \frac{1}{2} (1 - r) N, \quad j = 1, 4,$$

and

$$n_j^{(b)} > n_j - \varepsilon N > (1 - \varepsilon) \frac{1}{2} r N - \varepsilon N \geq (1 - k\varepsilon) \frac{1}{2} r N, \quad j = 2, 3,$$

where

$$k = \max \left(\frac{3-r}{1-r}, \frac{2+r}{r} \right).$$

Consequently,

$$\begin{aligned} & \overline{\lim}_{|\mu_2| \rightarrow \infty} f(y^{(N)}, \lambda, x) \\ & \leq \overline{\lim}_{|\mu_2| \rightarrow \infty} \left(\frac{1}{\sqrt{2\pi\sigma}} \right) \prod_{\substack{i=1 \\ y_i \in I_4}}^{N_1} \varphi(y_i; \lambda, \theta_1) \prod_{\substack{i=N_1+1 \\ y_i \in I_4}}^{N_2} \varphi(y_i; \lambda, \theta_2) \\ & \quad \times \prod_{\substack{i=N_2+1 \\ y_i \in I_4}}^{N_3} \varphi(y_i; \lambda, 1 - \theta_2) \prod_{\substack{i=N_3+1 \\ y_i \in I_4}}^{N_4} \varphi(y_i; \lambda, 1 - \theta_1) \\ & = \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^N \cdot (1 - \theta_1)^{n_1^{(b)}} (1 - \theta_2)^{n_2^{(b)}} \theta_2^{n_3^{(b)}} \cdot \theta_1^{n_4^{(b)}} \prod_{\substack{i=1 \\ y_i \in I_4}}^N \exp \left[-\frac{1}{2\sigma^2} (y_i - \mu_1)^2 \right] \\ & \leq \left(\frac{1}{\sqrt{2\pi} \frac{W_4^{(N)}}{N} \hat{\sigma}_b^{(N)}} \right)^N [(1 - \theta_1)\theta_1]^{(1-k\varepsilon)\frac{1}{2}(1-r)N} [(1 - \theta_2)\theta_2]^{(1-k\varepsilon)\frac{1}{2}rN} e^{-\frac{N}{2}} \\ & \leq \left(\frac{e^{-\frac{1}{2}}}{2^{1-k\varepsilon} \sqrt{2\pi} (1 - \varepsilon)^2 \sigma_0} \right)^N. \end{aligned}$$

Note that to obtain the second last inequality, the MLEs $\bar{y}_b^{(N)}$ and $\hat{\sigma}_b^{(N)}$ have been used. For small ε , clearly the last term is less than the right-hand side of (A.21). Symmetrically, for $|\mu_2| \leq b$ and as $|\mu_1| \rightarrow \infty$, the value of the likelihood function is also smaller than the right-hand side of (A.21). Therefore, we have shown that the supremum of $f(\cdot)$ is achieved in the area of μ_1, μ_2 , and σ bounded uniformly with respect to $x \in [0, D]$. The existence of mle's of μ_1, μ_2 , and σ for any $x \in [0, D]$ follows from the continuity of $f(\cdot)$ with respect to λ .

8.3.5 Proof of Lemma 4

Before we prove Lemma 4 we need a few preliminaries.

For any function $h(\lambda, x)$ of λ and x , any subset Λ of the parameter space of λ , and subset Φ of $[0, D]$, define

$$h(\Lambda, \Phi) = \sup_{\lambda \in \Lambda, x \in \Phi} h(\lambda, x).$$

For any $\varepsilon > 0$, $\lambda^* = (\mu_1^*, \mu_2^*, \sigma^*)$, and $x^* \in [0, D]$, define the ε -neighborhood of λ^* by

$$S(\lambda^*, \varepsilon) = \{\lambda = (\mu_1, \mu_2, \sigma) : |\mu_1 - \mu_1^*| < \varepsilon, |\mu_2 - \mu_2^*| < \varepsilon, |\sigma - \sigma^*| < \varepsilon\}$$

and the ε -neighborhood of x^* by

$$S(x^*, \varepsilon) = \{x \in [0, D] : |x - x^*| < \varepsilon\}.$$

As $\varepsilon \downarrow 0$, we have

$$f_{Y|G=i}(y; S(\lambda, \varepsilon), S(x, \varepsilon)) \downarrow f_{Y|G=i}(y; \lambda, x)$$

and

$$(A.24) \quad f_{Y,G}(y, g; S(\lambda, \varepsilon), S(x, \varepsilon)) \downarrow f_{Y,G}(y, g; \lambda, x).$$

From Rao (Section 6.6, page 59), or an application of Jensen's inequality, for $\lambda \neq \lambda_0$,

$$E_0 \left[\log \left(\frac{f_{Y,G}(Y, G; \lambda, x)}{f_{Y,G}(Y, G; \lambda_0)} \right) \right] < 0.$$

Now we prove Lemma 4.

Let $\delta, M, \sigma_1, \sigma_2, \xi$ and A be as in Lemma 3. It is sufficient to show that for any $\varepsilon > 0$, there exists a M_1 such that when $N > M_1$,

$$\hat{\lambda}^{(N)}(x) \in S(\lambda_0, \varepsilon), \quad x \in [0, D],$$

with probability $1 - 2\delta$. Define $\Gamma = \{\lambda : |\mu_1| \leq \xi, \sigma_1 \leq \sigma \leq \sigma_2\}$ and observe that

$$(A.25) \quad \left| \log \left(\frac{f_{Y,G}(y, g; \Gamma, [0, D])}{f_{Y,G}(y, g; \lambda_0)} \right) \right| \leq K|y| + Q,$$

where K and Q are positive constants depending on ξ, σ_1 , and σ_2 . Let $B = \Gamma - S(\lambda_0, \varepsilon)$ and note that for any $\lambda \in B$ and $x \in [0, D]$, (A.24), (A.25) and the Lebesgue Dominated Convergence Theorem imply that

$$\begin{aligned} \lim_{\varepsilon \downarrow 0} E_0 \left[\log \left(\frac{f_{Y,G}(Y, G; S(\lambda, \varepsilon), S(x, \varepsilon))}{f_{Y,G}(Y, G; \lambda_0)} \right) \right] &= E_0 \left[\log \left(\frac{f_{Y,G}(Y, G; \lambda, x)}{f_{Y,G}(Y, G; \lambda_0, x)} \right) \right] \\ &< C(\lambda, x) < 0. \end{aligned}$$

Therefore, there exists an $\varepsilon = \varepsilon(\lambda, x)$ such that

$$E_0 \left[\log \left(\frac{f_{Y,G}(Y, G; S(\lambda, \varepsilon), S(x, \varepsilon))}{f_{Y,G}(Y, G; \lambda_0)} \right) \right] < C(\lambda, x) < 0.$$

The preceding argument applies to each $(\lambda, x) \in B \times [0, D]$. Since $B \times [0, D]$ is compact, there exist points $(\lambda_1, x_1), \dots, (\lambda_k, x_k)$ and corresponding $\varepsilon_1, \dots, \varepsilon_k$ and C_1, \dots, C_k such that

$$B \times [0, D] \subset \bigcup_{n=1}^k S(\lambda_i, \varepsilon) \times S(x_i, \varepsilon)$$

and for $i = 1, \dots, k$,

$$(A.26) \quad E_0 \left[\log \left(\frac{f_{Y,G}(Y, G; S(\lambda_i, \varepsilon_i), S(x_i, \varepsilon_i))}{f_{Y,G}(Y, G; \lambda_0)} \right) \right] < C_i < 0.$$

Let $C = \max(C_1, \dots, C_k)$. Because

$$(A.27) \quad \frac{1}{N} \log \left(\frac{L_N(S(\lambda_i, \varepsilon_i), S(x_i, \varepsilon_i))}{L_N(\lambda_0)} \right)$$

converges a.s. to the left-hand side of (A.26), there exist a measurable subset $A_1 \subset A$ with $P_0(A_1) > 1 - 2\delta$ and an integer $M_1 > M$ such that when $N > M_1$,

$$(A.27) < C < 0, \quad \text{for } i = 1, \dots, k.$$

Now for any $\lambda \in B$ and any $x \in [0, D]$, there exists an i , $1 \leq i \leq k$, such that $(\lambda, x) \in S(\lambda_i, \varepsilon_i) \times S(x_i, \varepsilon_i)$ and for any sample point in A_1 and $N > M_1$,

$$\begin{aligned} \frac{1}{N} [\ell_N(\lambda, x) - \ell_N(\lambda_0)] &= \frac{1}{N} \log \left(\frac{L_N(\lambda, x)}{L_N(\lambda_0)} \right) \\ &< \frac{1}{N} \log \left(\frac{L_N(S(\lambda_i, \varepsilon_i), S(x_i, \varepsilon_i))}{L_N(\lambda_0)} \right) < C < 0, \end{aligned}$$

which implies that $\hat{\lambda}^{(N)}(x) \in B$, for all $x \in [0, D]$. By Lemma 3,

$$\hat{\lambda}^{(N)}(x) \in \Gamma, \quad \text{for all } x \in [0, D].$$

Thus, for any sample point in A_1 and $N > M_1$,

$$\hat{\lambda}^{(N)}(x) \in S(\lambda_0, \varepsilon), \quad \text{for all } x \in [0, D],$$

which concludes the proof.

References

Churchill, G. A., and R. W. Doerge. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963-971.

Cox, D. R. and D. V. Hinkley. (1974). *Theoretical Statistics*. Chapman & Hall: London.

Darvasi, A., A. Weinreb, V. Minke, J. I. Weller and M. Soller. (1993). Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134: 943-951.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64: 247-254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74: 33-43.

Doerge, R. W., and G. A. Churchill. (1996). Permutation tests for multiple loci affecting a quantitative character. *Genetics* 142: 285-294.

Doerge, R. W., Z-B. Zeng and B. S. Weir (1997). Statistical Issues in the Search for Genes Affecting Quantitative Traits in Experimental Populations. *Statistical Science* 12: 195-219.

Dupuis, J. and D. Siegmund. (1999). Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* 151: 373-386.

Haldane, J. B. S. (1919) The combination of linkage values and the calculation of distance between the loci of linkage factors. *J. Genet.* 8: 299-309.

Haley, C. S., S. A. Knott and J. M. Elsen. (1994). Genetic mapping of quantitative trait loci in cross between outbred lines using least squares. *Genetics* 136: 1195-1207.

Hoeschele, I. (2001). Mapping quantitative trait loci in outbred pedigrees. In: *Handbook of Statistical Genetics*, Edited by D. J. Balding, M. Bishop and C. Cannings. Wiley New York 599-644.

- Jansen, R. C. (2001). Quantitative trait loci in inbred lines. In: *Handbook of Statistical Genetics*, Edited by D. J. Balding, M. Bishop and C. Cannings. Wiley New York. 567-597.
- Lander, E. S., and D. Botstein. (1989). Mapping Mendelian factors underlying quantitative traits using RELP linkage maps. *Genetics* 121: 185-199.
- Lander, E. S., and N. J. Schorck. (1994). Genetic dissection of complex traits. *Science* 265: 2037-2048.
- Lin, M., X.-Y. Lou, M. Chang and R. L. Wu. (2004). A general statistical framework for mapping quantitative trait loci in non-model systems: Issue for characterizing linkage phases. *Genetics* 165: 901-913.
- Piepho, H-P. (2001). A quick method for computing approximate thresholds for quantitative trait loci detection. *Genetics* 157: 425-432.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd Edition. Wiley, New York.
- Rebai, A., B. Goffinet, and B. Mangin. (1994). Approximate thresholds of interval mapping tests for QTL detection. *Genetics* 138: 235-240.
- Rebai, A., B. Goffinet and B. Mangin. (1995) Comparing power of different methods for QTL detection. *Biometrics* 51: 87-99.
- Schaid, D. J., C. M. Rowland, D. E. Tines, R. M. Jacobson and G. A. Poland. (2002). Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am. J. Hum. Genet.* 70: 425-34.
- Van Ooijen, J.W. (1992). Accuracy of mapping quantitative trait loci in autogamous species. *Theor. Appl. Genet.* 84: 803-811.
- Wang, K., and J. Huang. (2002). A score-statistic approach for mapping quantitative trait loci with sibships of arbitrary size. *Am. J. Hum. Genet.* 70: 412-424.
- Wu, R. L., C.-X. Ma and G. Casella. (2007) *Statistical Genetics of Quantitative Traits: Linkage, Maps, and QTL*. Springer-Verlag, New York.

- Wu, R. L., C.-X. Ma, I. Painter and Z.-B.Zeng. (2002). Simultaneous maximum likelihood estimation of linkages and linkage phases over a heterogeneous genome. *Theor. Pop. Biol.* 61: 349-363.
- Wu, R. L., M. X. Wang and M. R. Huang. (1992). Quantitative genetics of yield breeding for Populus short rotation culture. I. Dynamics of genetic control and selection models of yield traits. *Can. J. Forest Res.* 22: 175-182.
- Xu, S. Z. (1996). Mapping quantitative trait loci using four-way crosses. *Genet. Res.* 68: 175-181.
- Yin, T. M., X. Y. Zhang, M. R. Huang, M. X. Wang, Q. Zhuge, S. M. Tu, L. H. Zhu and R. L. Wu. (2002). Molecular linkage maps of the Populus genome. *Genome* 45: 541-555.
- Zeng, Z.-B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136: 1457-1468.
- Zou, F., B. S. Yandell and J. P. Fine. (2001). Statistical issues in the analysis of quantitative traits in combined crosses. *Genetics* 158: 1339-1346.
- Zou, F, J. P. Jason, J. Hu, and D. Y. Lin (2004). An efficient resampling method for assessing genome-wide statistical significance in mapping quantitative trait loci. *Genetics* 168 (4): 2307-2316.