Rank-Based Statistical Methodologies for Quantitative Trait Locus Mapping

Fei Zou,*[,1] Brian S. Yandell[†] and Jason P. Fine[†]

*Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27599 and
[†]Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706

Manuscript received April 13, 2003
Accepted for publication July 25, 2003

ABSTRACT

This article addresses the identification of genetic loci (QTL and elsewhere) that influence nonnormal quantitative traits with focus on experimental crosses. QTL mapping is typically based on the assumption that the traits follow normal distributions, which may not be true in practice. Model-free tests have been proposed. However, nonparametric estimation of genetic effects has not been studied. We propose an estimation procedure based on the linear rank test statistics. The properties of the new procedure are compared with those of traditional likelihood-based interval mapping and regression interval mapping via simulations and a real data example. The results indicate that the nonparametric method is a competitive alternative to the existing parametric methodologies.

QUANTITATIVE genetics has developed rapidly, especially with progress in DNA-based genetic linkage maps. Various statistical approaches have been proposed to identify QTL by using molecular markers, such as SAX's (1923) single-marker *t*-test, LANDER and BOTSTEIN's (1989) maximum-likelihood-based interval mapping, HALEY and KNOTT's (1992) regression interval mapping, and ZENG's (1993, 1994) and JANSEN and STAM's (1994) composite interval mapping.

All the methods mentioned above are based on the normality assumption (or other parametric models) for the component distributions. The normal mixture model is the default analysis and is implemented in the widely used packages Mapmaker/QTL (LINCOLN *et al.* 1993) and QTL Cartographer (BASTEN *et al.* 1997). Many traits, however, are not normally distributed. An example is tumor counts, which arise in cancer studies and often appear to follow a negative binomial (DRINKWATER and KLOTZ 1981). Naively assuming normality of the underlying distributions greatly simplifies the form of the likelihood function. A problem is that if this assumption is violated, then false detection of a major locus may occur (MORTON 1984).

When the underlying distributions are suspected to be nonnormal, one strategy is to use a likelihood approach after transforming the data using, for example, the Box-Cox transformation (DRAPER and SMITH 1998). However, an appropriate transformation may not exist or may be difficult to find. Also this approach can raise serious issues of interpretation and the transformation involves an extra parametric assumption.

An alternative approach is to consider nonparametric methods. KRUGLYAK and LANDER (1995) apply the linear rank statistics to interval mapping, which is implemented in the latest version of Mapmaker/QTL (LINCOLN *et al.* 1993) and Qlink (DRINKWATER 1997). However, the method tests only for the presence of a QTL and does not provide an estimate of the phenotypic effect of the QTL. In this article, we extend the rank-based test statistic to the estimation of the quantitative trait effects.

Rank-based methods play an important role in nonparametric statistics. The linear rank statistic has been widely used in practice and its theoretical properties have been thoroughly studied (HAJEK and SIDAK 1967; HAJEK 1968). For linear regression, estimates of the regression coefficients based on linear rank statistics are available and have efficiency and robustness properties that are similar to those of the linear rank statistics. In this article, we adapt the existing methodology to construct rank-based estimates for genetic effects under the assumption that the underlying QTL component distributions have the same form and differ only by a shift. This appears to be the first attempt to apply linear rank-based estimates directly to the interval mapping and thus complements existing parametric methods. Simulations are conducted to compare the relative efficiencies of the nonparametric and parametric methods under a variety of distributions.

The article is arranged as follows. In the next section, we briefly introduce linear rank statistics and related estimation procedures for regression analysis. In NONPARAMETRIC INTERVAL MAPPING, the estimates of QTL effects are proposed in the context of interval mapping. In NUMERICAL STUDIES, the relative efficiencies of the proposed estimates are compared with the parametric estimates in simulation studies and the methods are illustrated in backcross data where the phenotype has a highly skewed distribution. In CONCLUSION AND RE-

[1] *Corresponding author:* Department of Biostatistics, University of North Carolina, 3107D McGavran-Greenberg Hall, CB 7420, Chapel Hill, NC 27599. E-mail: fzou@bios.unc.edu

Marks, we discuss the practical utility of the proposed methods.

RANK-BASED METHODS

First consider a simple regression model: $P(Y_i < y|X_i) = F(y - X_i\beta)$, where F is an unknown distribution, X_i are regressors, and Y_i are responses, $i = 1, \ldots, n$, and we are interested in testing H_0: $\beta = 0$. Define the shifted responses $Y_i(b) = Y_i - (X_i - \bar{X})\, b$ and their ranks $R_i(b) = \mathrm{rank}(Y_i(b))$. The ranks are 1 for the smallest observation, 2 for the next, and so on, preserving the order of the data but not the value. Under the null, the distribution of $R_i(0)$ is independent of the distribution F and uniformly distributed on $\{1, 2, 3, \ldots, n\}$. The Wilcoxon score statistic $L(b) = 1/(n+1)\Sigma_{i=1}^{n}(X_i - \bar{X})\, R_i(b)$ is a simple linear rank statistic (see Puri and Sen 1985 for some alternatives) and is widely used to test H_0. Statistical inquiry based on ranks can have dramatically smaller variances when data are not normal, leading to more efficient tests and estimators. Note that if we knew the true shift β, then the shifted values $Y_i(\beta)$ would all have the same distribution F and $E_\beta\{L(\beta)\} = 0$. All rank-based inference and estimation procedures are built on this premise.

The statistic $L(b)$ plays a fundamental role in nonparametric inference. Under the null hypothesis H_0: $\beta = 0$, $L(0)$ has the following asymptotic property:

$$Z^2 \overset{\text{def}}{=} \lim_{n\to\infty} \frac{L(0)^2}{\mathrm{Var}(L(0))} = \lim_{n\to\infty} \frac{(n-1)\{\sum_{i=1}^{n}(X_i - \bar{X})\, R_i(0)/(n+1)\}^2}{\sum_{i=1}^{n}(i/(n+1) - 1/2)^2 \sum_{i=1}^{n}(X_i - \bar{X})^2} \to \chi_1^2. \tag{1}$$

To estimate β, find the value b that shifts values of Y_i to $Y_i(b)$ such that the shifted values $Y_i(b)$ are not associated with X_i's anymore. A commonly used estimator is the Hodges-Lehmann estimator $\hat\beta$, which is the solution of the estimating equation $L(b) = 0$. However, the linear rank statistic $L(b)$ may not reach zero, so in practice $\hat\beta$ is taken to be the average of the closest values on either side of 0. In other words, $\hat\beta = \frac{1}{2}(\hat\beta_U + \hat\beta_L)$ with

$$\hat\beta_U = \inf\{b{:}L(b) < 0\} \quad \text{and} \quad \hat\beta_L = \sup\{b{:}L(b) > 0\}. \tag{2}$$

The asymptotic properties of the linear rank-based inferences and estimators and their relative efficiencies are discussed in detail in Puri and Sen (1985). The efficiency of the Wilcoxon rank sum test (Hodges-Lehmann estimate) relative to the t-test [maximum-likelihood estimate (MLE)] is ~95% if the distribution is normal and is never <86% for symmetric distributions. Thus the loss of efficiency in the normal case is slight and is offset by the robustness of the nonparametric method. For heavy tailed distributions, the gain in efficiency may be great. Later our simulations show that even for nonsymmetric error distributions, such as exponential, the rank-based method may be more efficient.

For multiple regression, all the above arguments can be extended in a straightforward manner. Suppose $P(Y_i < y|X_i) = F(y - X_i'\beta)$, where β, $X_i \in \mathcal{R}^p$. Again, F is totally unspecified. Similar to the simple regression, we define

$$Y_i(b) = Y_i - (X_i - \bar{X})'b$$

and

$$L(b) = \sum_{i=1}^{n}(X_i - \bar{X})R_i(b)/(n+1) = \{L^1(b), \ldots, L^p(b)\}',$$

where $b = (b_1, \ldots, b_p)' \in \mathcal{R}^p$.

Under some regularity conditions (Puri and Sen 1985, Chap. 5),

$$D_n^{-1}\{L'(0)\, C_n^{-}\, L(0)\} \xrightarrow{n\to\infty} \chi_p^2,$$

where $D_n = (n-1)^{-1}\Sigma_{i=1}^{n}(i/(n+1) - 1/2)^2$ and $C_n = \Sigma_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})'$. This result can be used to test H_0: $\beta = 0$.

To estimate β, define $\|L(b)\|^2 = \Sigma_{j=1}^{p} L^j(b)^2$, and let $\Delta_n = \{\arg\min_b \|L(b)\|^2\}$. Note that the set Δ_n may not be a single point. To obtain a unique estimator, we can let $\hat\beta$ be the center of mass of Δ_n. The computation of $\hat\beta$ usually requires some iterative procedures.

NONPARAMETRIC INTERVAL MAPPING

Backcross: In this section, we consider a backcross population $[(QQ \times qq) \times QQ]$. For a single-QTL model, we assume $P(Y_i < y|X_i) = F(y - \beta X_i)$, where $X_i = I(Q_i)$ is the indicator function that takes 1 if the QTL genotype $Q_i = QQ$, and 0 otherwise. We are interested in testing H_0: $\beta = 0$ *vs.* H_1: $\beta \neq 0$ and in estimating β, the genetic shift in distribution at the QTL between QQ and qQ genotypes.

If the QTL genotype Q_i's are known, we could apply the Wilcoxon rank sum tests and Hodges-Lehmann estimators directly in QTL analysis. However, in intervals between known loci, the QTL genotypes are not observed and the quantitative traits follow discrete mixture models and thus Q_i's are generally not available. A natural choice would be to use Haley-Knott regression (Haley and Knott 1992). That is, first, the mixing weights are calculated as the conditional probabilities of the QTL genotypes in intervals between marker loci using the genetic map and the genotypes of the flanking markers. Then, X_i is substituted with its conditional expectation $E(X_i|\text{flanking markers})$.

Since individuals with the same flanking markers have the same mixing weights and thus the same mixture distribution, for convenience, we can group the data into K groups by their flanking-marker genotypes. Suppose within each group the data have common distribution M_k, $k = 1, 2, \ldots, K$. Under the null hypothesis H_0, $M_1 = M_2 = \ldots = M_K$. After substituting X_i in (1)

TABLE 1

Comparison of parametric and nonparametric methods (20 cM)

$\beta = 2$	Power	cM	ARE	$\hat{\beta}$	ARE
Normal(0, 1)					
ML	1	10.13(1.100)	—	2.00(0.0043)	—
REG	1	10.06(1.148)	0.958	2.00(0.0051)	0.843
Rank	1	10.09(1.17)	0.94	2.00(0.005)	0.86
Exponential(0.5)					
ML	1	10.15(2.997)	—	2.00(0.0155)	—
REG	1	10.11(3.008)	0.996	1.977(0.0148)	1.047
Rank	1	10.22(2.375)	1.262	2.00(0.0068)	2.279
t(3)					
ML	1	10.35(2.836)	—	2.036(0.0109)	—
REG	1	10.25(3.199)	0.887	1.995(0.0130)	0.840
Rank	1	10.27(1.75)	1.62	2.00(0.007)	1.557
Logistic(0, 1)					
ML	1	9.8(2.9293)	—	2.015(0.0198)	—
REG	1	9.81(2.984)	0.982	2.009(0.0195)	1.020
Rank	1	9.78(2.557)	1.146	2.01(0.0175)	1.236
Cauchy(0, 1)					
ML	0.01	9.54(73.503)	—	15.448(44256)	—
REG	0.01	10.37(44.478)	1.653	18.714(66482)	0.666
Rank	1	10.23(2.886)	25.521	2.00(0.018)	∞

Tables 1–4 show the mean estimate of the QTL location (cM) and its gene effect (β), with the empirical variances of the estimates over replicates in parentheses. ARE is the estimated asymptotic relative efficiency of regression analysis or rank-based method *vs.* the ML method and is defined as var(ML)/var(REG) or var(ML)/var(Rank).

with $E(X_i|\text{flanking markers})$, we obtain the rank test statistic equivalent to the one in Kruglyak and Lander (1995). Note that instead of testing H_0 directly, here we instead test $M_1 = M_2 = \ldots = M_K$. Usually, K is much greater than the number of underlying distributions. For example, in the backcross population, we are interested in testing the difference between the two component distributions in the mixture model. In essence, we test for differences among the four mixtures, M_k, $k = 1, \ldots, 4$. Theoretically, it is unclear whether the relative efficiency of the rank sum test *vs.* the *t*-test [or, equivalently the likelihood-ratio test (LRT)] in linear regression still holds in this setting. However, we expect that the rank sum test performs better under most circumstances when data are nonnormal, which we investigate by simulations.

The estimation of β is more problematic than that for simple linear regression. In traditional linear regression, $E_\beta\{L(\beta)\} = 0$. Thus the estimator $\hat{\beta}$ is consistent. However, due to the mixture structure of QTL data, we can show that $E_\beta\{L(\beta)\}$ does not generally equal 0 when X_i is substituted by its conditional expectation. A theoretical formula of $E_\beta\{L(\beta)\}$ indicates that the magnitude of the deviation from 0 depends on the underlying distributions, the flanking marker distances, and the magnitude of β. It can be shown that, for a given distribution, the deviation goes to 0 as β goes to 0 or as the flanking marker distance goes to 0. Thus we expect the estimator $\hat{\beta}$ to work well in QTL analysis if either there is a relatively dense map (*e.g.*, < 20 cM, a common scenario of current genetic studies) or the QTL effect is relatively small. Efficiency is of less concern when the QTL effect is large than when it is small. In QTL mapping of complex traits, an individual QTL usually has small effect. For these reasons, we believe and our simulations as well show that the rank sum-based estimators are practically useful alternatives to the least-squares estimators from Haley and Knott's regression interval mapping.

The following are some properties of $\hat{\beta}$. To emphasize that $\hat{\beta}$ depends on $Y = \{Y_i\}$, we rewrite $\hat{\beta}$ as $\hat{\beta}(Y)$. From the definition of $\hat{\beta}$, it is not difficult to show that, for any $b \in R$,

i. $\hat{\beta}(Y) = \hat{\beta}(Y + b) \overset{\text{def}}{=} \hat{\beta}(Y_1 + b, \ldots, Y_n + b)$, and
ii. $\hat{\beta}(Y) = -\hat{\beta}(-Y)$.

In words, i indicates that adding a constant to the data has no effect on the estimator of QTL effect. Property ii says that if the data are multiplied by −1, the estimator has an opposite sign.

TABLE 2

Comparison of parametric and nonparametric methods (20 cM)

$\beta = 1$	Power	cM	ARE	$\hat{\beta}$	ARE
Normal(0, 1)					
ML	1	10.18(2.412)	—	1.00(0.0046)	—
REG	1	10.24(2.689)	0.897	0.999(0.0051)	0.902
Rank	1	10.20(2.444)	0.987	1.00(0.0054)	0.852
Exponential(0.5)					
ML	1	10.06(12.036)	—	1.00(0.0172)	—
REG	1	9.94(12.622)	0.954	0.999(0.018)	0.956
Rank	1	9.96(6.60)	1.824	1.00(0.0072)	2.389
$t(3)$					
ML	1	10.43(8.51)	—	1.00(0.0115)	—
REG	1	10.530(6.817)	1.248	1.00(0.014)	0.821
Rank	1	10.32(4.038)	2.107	1.00(0.0075)	1.533
Logistic(0, 1)					
ML	1	9.99(8.172)	—	0.997(0.0167)	—
REG	1	10.03(7.848)	1.041	0.999(0.0163)	1.020
Rank	1	9.86(7.031)	1.162	0.999(0.0153)	1.091
Cauchy(0, 1)					
ML	0	9.66(78.53)	—	5.011(1187)	—
REG	0	9.09(46.265)	1.697	5.842(1698)	0.699
Rank	1	10.89(8.766)	8.958	1.00(0.017)	∞

Extensions: Next we extend the methods to any other cross derived from two inbred lines, such as F_2. In general, the model can be expressed as $P(Y_i < y|X_i) = F(y - X_i'\beta)$, where $\beta = (a, d)'$ and $X_i = (X_{1,i}, X_{2,i})'$. The covariates

$X_{1,i} = -1$, 0, or 1 if individual i has QTL genotype qq, qQ, or QQ, and
$X_{2,i} = 1$ (or 0) if individual i has QTL genotype qQ (or else)

correspond to the additive and dominance genetic effects, a and d, respectively. In regression mapping, if the unknown $X_{j,i}$'s are replaced by their conditional expectations $E(X_{j,i}|\text{flanking markers})$, then the estimator $\hat{\beta}$ can be derived as described in RANK-BASED METHODS for multiple regression without any modifications. The methods may also be adapted to map multiple QTL (Kao *et al.* 1999) or to more complicated designs involving more than two inbred lines (Liu and Zeng 2000) by changing the dimension of β. Of course, the efficiency may be low if the dimension of β is large. This requires further investigation.

NUMERICAL STUDIES

Simulations were conducted to study the behavior of Z and $\hat{\beta}$ in a backcross population. For simplicity, only one chromosomal segment flanked by two markers is simulated. The two markers are either located at 0 and 10 cM with simulated QTL at 5 cM or located at 0 and 20 cM with simulated QTL at 10 cM, respectively. The setups are similar to those in Xu (1995). The QTL effect β is either 1 or 2. Standard normal, exponential(0.5), $t(3)$, standard logistic, and standard Cauchy are used as error distributions. One hundred simulations were conducted for each combination with sample size $n = 1000$. The average values and corresponding standard errors of estimated QTL position, QTL effect from parametric interval mapping (ML), and nonparametric Wilcoxon rank sum interval mapping (Rank) are given in Tables 1–4. As a comparison, we also run the regression analysis (REG) of Haley and Knott (1992) and the results are given in Table 1.

The estimates of QTL position and effect from the REG and the ML methods are very similar not only for normal data, which is consistent with Haley and Knott (1992) and with Xu (1995), but also for nonnormal data. Note that the nonparametric test and estimate generally are much more efficient than the parametric versions when data are not normally distributed. There is a modest loss of efficiency with normal data, which agrees with theory for simple linear regression. The marker distances and the magnitude of the QTL effect do not seem to have a large impact on the relative efficiencies of the estimators.

To estimate the power, the rank test statistic Z is first transformed to $\text{LOD}_R = \{2 \log(10)\}^{-1}Z^2$ and the test statistic from REG is also transformed to an equivalent

TABLE 3

Comparison of parametric and nonparametric methods (10 cM)

β = 2	Power	cM	ARE	$\hat{\beta}$	ARE
Normal(0, 1)					
ML	1	5.01(0.535)	—	1.99(0.0047)	—
REG	1	4.96(0.483)	1.108	1.989(0.0052)	0.905
Rank	1	4.99(0.454)	1.179	1.99(0.0054)	0.884
Exponential(0.5)					
ML	1	4.94(1.491)	—	2.03(0.0217)	—
REG	1	5.02(1.636)	0.911	2.017(0.0230)	0.945
Rank	1	4.97(0.999)	1.492	2.00(0.0084)	2.592
t(3)					
ML	1	5.18(1.26)	—	2.04(0.0111)	—
REG	1	5.10(1.323)	0.952	1.988(0.0136)	0.818
Rank	1	5.16(0.984)	1.280	1.995(0.0079)	1.414
Logistic(0, 1)					
ML	1	4.98(1.192)	—	2.00(0.0131)	—
REG	1	4.95(1.159)	1.028	2.01(0.0134)	0.973
Rank	1	5.000(1.030)	1.156	2.01(0.0120)	1.090
Cauchy(0, 1)					
ML	0.01	5.23(14.987)	—	2.72(75.157)	—
REG	0.01	5.28(9.82)	1.526	2.87(88.92)	0.845
Rank	1	4.94(1.835)	8.130	1.978(0.0145)	∞

LOD score. We then take threshold 3 for the LOD scores, which is recommended in practical genome-wide QTL analysis (see also Kruglyak and Lander 1995 for analytic genome-wide threshold calculations). The power is calculated as the proportion of significant tests from 100 simulated data sets. For the extreme case where data are Cauchy distributed, there is no power to detect the QTL by ML or REG interval mapping while Rank interval mapping does have power.

To further demonstrate the method, we consider the data on the time to death following infection with *Listeria monocytogenes* of 116 F_2 mice from an intercross between the BALB/cByJ and C57BL/6ByJ strains (Boyartchuk *et al.* 2001). The histograms of the log time to death of the nonsurvivors are given in Figure 1. Roughly 30% of mice survive beyond 264 days. From the histogram it is hard to justify that the log time to death of the nonsurvivors is normally distributed. Broman (2003) applied four different methods, including both the standard interval mapping and nonparametric interval mapping, to this data set and showed that the locus on chromosome 1 appears to have effect only on the average time to death among the nonsurvivors. For this reason, our analysis is restricted on chromosome 1 for those nonsurvivors.

The LOD scores obtained by standard interval mapping and the nonparametric interval mapping with the log time to death are plotted in Figure 2. It is clear that the two methods result in the maximums at the same position although the LOD curves are slightly different, which will result in some slightly different confidence intervals of the putative QTL locus by the conventional 1-LOD drop method. The additive and dominance estimators are 0.262 and 0.059, respectively, from standard interval mapping and are 0.257 and 0.038, respectively, based on our method. To assess whether the differences between the two methods are significant or not, 1000 bootstraps are performed. We restrict our analysis within chromosome 1. From our method, the 95% confidence interval (CI) of the QTL locus is (50 cM, 84 cM). The mean of the additive effect is 0.247 with standard error 0.077 and the mean of the dominant effect is 0.055 with standard error 0.122. Similarly, from standard interval mapping, we get the 95% CI of the QTL locus as (51 cM, 92 cM). The mean of the additive effect is 0.268 with standard error 0.071 and the mean of the dominant effect is 0.0284 with standard error 0.122. In all, the nonparametric QTL locus estimator is relatively more efficient than the parametric estimator and our nonparametric analysis confirms the results of Broman (2003).

CONCLUSION AND REMARKS

In this article, traditional rank-based estimators for linear regression have been adapted to analyze quantitative traits. The new method has been shown to be very similar to Haley and Knott's regression interval mapping when data are normally distributed and more efficient

TABLE 4

Comparison of parametric and nonparametric methods (10 cM)

$\beta = 1$	Power	cM	ARE	$\hat{\beta}$	ARE
Normal(0, 1)					
ML	1	4.98(1.252)	—	0.996(0.0039)	—
REG	1	4.96(1.17)	0.986	0.996(0.0044)	0.885
Rank	1	4.96(1.19)	1.052	0.997(0.0046)	0.837
Exponential(0.5)					
ML	1	5.03(4.252)	—	1.00(0.018)	—
REG	1	5.09(3.962)	1.073	1.00(0.0184)	0.980
Rank	1	5.09(2.26)	1.881	1.00(0.0064)	2.830
t(3)					
ML	1	4.97(2.999)	—	1.00(0.016)	—
REG	1	4.92(2.80)	1.071	1.00(0.0161)	0.996
Rank	1	4.96(1.594)	1.881	0.993(0.0096)	1.667
Logistic(0, 1)					
ML	1	5.12(3.581)	—	0.993(0.0114)	—
REG	1	5.07(3.157)	1.135	0.998(0.0116)	0.979
Rank	1	5.14(2.889)	1.239	0.996(0.0104)	1.096
Cauchy(0, 1)					
ML	0	5.49(15.543)	—	0.442(61.51)	—
REG	0	5.34(10.69)	1.454	0.396(68.26)	0.901
Rank	1	4.88(3.238)	4.800	0.991(0.0146)	∞

for nonnormal data. Our simulations indicate that the normal likelihood-ratio-based interval mapping is usually unbiased, even when the data are nonnormal, but may have very low efficiency. All our simulations are based on one QTL model. We believe the nonparametric model is very likely to produce ghost QTL as the parametric method does when two QTL are close to each other and multiple nonparametric QTL mapping is needed.

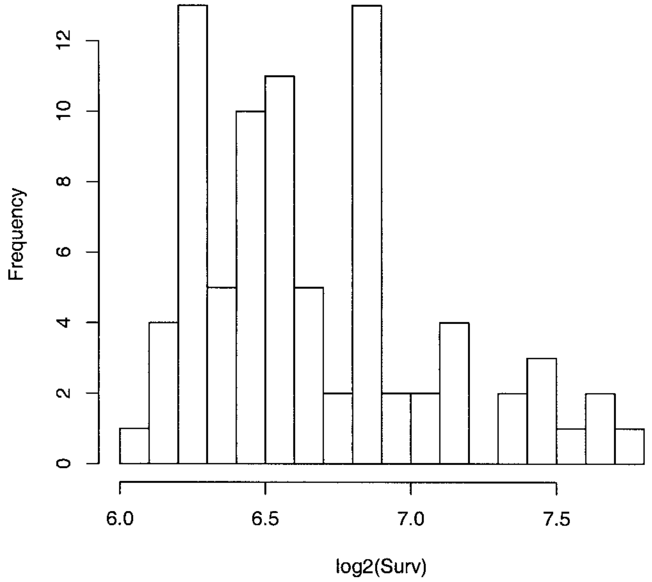

Figure 1.—Histogram of log 2(survival time), following infection with *Listeria monocytogenes* of 85 nonsurvival mice out of a total of 116 mice. The remaining 31 mice recovered from the infection and survived to the end of experiment, 264 hr [log 2(264) = 8].

In genetic studies of quantitative traits, adapting rank-

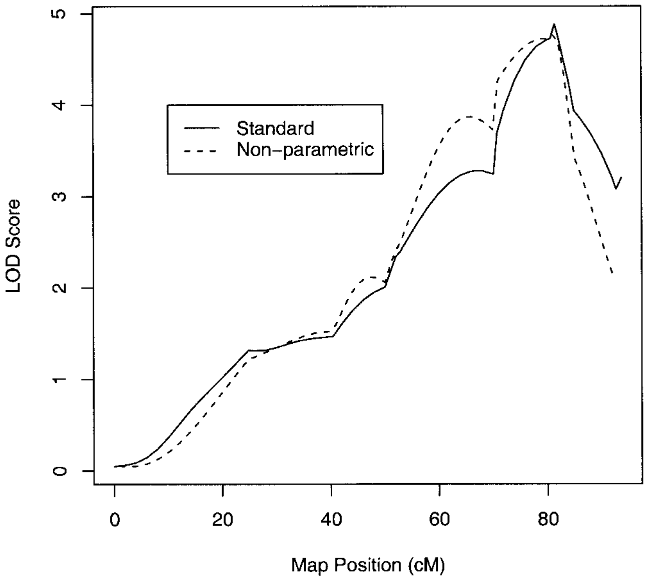

Figure 2.—LOD score curves from standard interval mapping (solid line) and nonparametric interval mapping (dashed line).

based methodologies is complicated because genetic markers are observed only at known loci and the QTL genotypes are usually unknown. Thus, the trait data arise from discrete mixtures of unknown distributions. The mixture structure of the data may distort certain properties of the underlying error distributions. For example, F may be unimodal even though the QTL data may not be. This means that the rank test in QTL mapping may have properties that differ from those for the rank test in linear regression.

As explained in NONPARAMETRIC INTERVAL MAPPING, the rank-based parameter estimate $\hat{\beta}$ is not generally unbiased with QTL data because the unknown regressors X_i are replaced by their expectations. On the basis of the theory of general estimating equations (LIANG and ZEGER 1986), one may show that the estimators of genetic effects from HALEY and KNOTT's (1992) regression method are unbiased, although the variances of the estimators may be larger than those from the Hodges-Lehmann estimators. While the rank-based estimators are theoretically biased, in simulations, the bias is negligible when compared with the regression and maximum-likelihood methods.

The computation of $\hat{\beta}$ usually is complicated if the dimension of β is >1 and requires some iterative procedures. KRAFT and VAN EEDEN (1972) proposed an easy one-step modification of the least-squares estimator of β to approximate $\hat{\beta}$. We may use this one-step modification if the calculation of $\hat{\beta}$ is too complicated,

$$\hat{\beta} \approx \tilde{\beta} + A_n^{-1}[C_n^{-1}L_n(\tilde{\beta})], \tag{3}$$

where $\tilde{\beta}$ is the least-squares estimator of β, and for any $d \in \mathcal{R}^p$,

$$A_n = \left(\frac{[L_n(\tilde{\beta}) - L_n(\tilde{\beta} - n^{-1/2}d)]'[L_n(\tilde{\beta}) - L_n(\tilde{\beta} - n^{-1/2}d)]}{(d'C_n^2d)/n^2}\right)^{1/2}.$$

LITERATURE CITED

BASTEN, C. J., B. S. WEIR and Z-B. ZENG, 1997 *QTL Cartographer: A Reference Manual and Tutorial for QTL Mapping*. Department of Statistics, North Carolina State University, Raleigh, NC.

BOYARTCHUK, V. L., K. W. BROMAN, R. E. MOSHER, S. E. F. DORAZIO, M. N. STARNBACH *et al.*, 2001 Multigenic control of Listeria monocytogenes susceptibility in mice. Nat. Genet. **27:** 259–260.

BROMAN, K. W., 2003 Mapping quantitative trait loci in the case of a spike in the phenotype distribution. Genetics **163:** 1169–1175.

DRAPER, N. R., and H. SMITH, 1998 *Applied Regression Analysis*, Ed. 3. John Wiley & Sons, New York.

DRINKWATER, N. R., 1997 *Qlink Documentation*. McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, WI.

DRINKWATER, N. R., and J. H. KLOTZ, 1981 Statistical methods for the analysis of tumor multiplicity data. Cancer Res. **41:** 113–119.

HAJEK, J., 1968 Asymptotic normality of simple linear rank statistics under alternatives. Ann. Math. Stat. **39:** 325–346.

HAJEK, J., and Z. SIDAK, 1967 *Theory of Rank Tests*. Academic Press, New York/London.

HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative traits in line crosses using flanking markers. Heredity **69:** 315–324.

JANSEN, R. C., and P. STAM, 1994 High resolution of quantitative traits into multiple loci via interval mapping. Genetics **136:** 1447–1455.

KAO, C. H., Z-B. ZENG and R. D. TEASDALE, 1999 Multiple interval mapping for quantitative trait loci. Genetics **152:** 1203–1216.

KRAFT, C. H., and C. VAN EEDEN, 1972 Linearized rank estimates and signed rank estimates for the general linear hypothesis. Ann. Math. Stat. **43:** 42–57.

KRUGLYAK, L., and E. S. LANDER, 1995 A nonparametric approach for mapping quantitative trait loci. Genetics **139:** 1421–1428.

LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

LIANG, K. Y., and S. L. ZEGER, 1986 Longitudinal data analysis using generalized linear models. Biometrika **73:** 13–22.

LINCOLN, S. E., M. J. DALY and E. S. LANDER, 1993 *A Tutorial and Reference Manual for MAPMAKER/QTL*. Whitehead Institute for Biometrical Research.

LIU, Y., and Z-B. ZENG, 2000 A general mixture model approach for mapping quantitative trait loci from diverse cross designs involving multiple inbred lines. Genet. Res. **75:** 345–355.

MORTON, N. E., 1984 Trials of segregation analysis by deterministic and macro simulation, pp. 83–107 in *Human Population Genetics: The Pittsburgh Symposium*, edited by A. CHAKRAVARTI. Van Nostrand Reinhold, New York.

PURI, M. L., and P. K. SEN, 1985 *Nonparametric Methods in General Linear Models*. John Wiley & Sons, New York.

SAX, K., 1923 The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics **8:** 552–560.

XU, S., 1995 A comment on the simple regression method for interval mapping. Genetics **141:** 1657–1659.

ZENG, Z-B., 1993 Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA **90:** 10972–10976.

ZENG, Z-B., 1994 Precision mapping of quantitative traits loci. Genetics **136:** 1457–1468.

Communicating editor: Z-B. ZENG