# Constructing Confidence Intervals for QTL Location

**B. Mangin, B. Goffinet and A. Rebaï**

*Institut National de la Recherche Agronomique, Station de Biométrie et d'Intelligence Artificielle, 31326 Castanet-Tolosan Cedex, France*

Manuscript received December 6, 1993
Accepted for publication August 3, 1994

ABSTRACT

We describe a method for constructing the confidence interval of the QTL location parameter. This method is developed in the local asymptotic framework, leading to a linear model at each position of the putative QTL. The idea is to construct a likelihood ratio test, using statistics whose asymptotic distribution does not depend on the nuisance parameters and in particular on the effect of the QTL. We show theoretical properties of the confidence interval built with this test, and compare it with the classical confidence interval using simulations. We show in particular, that our confidence interval has the correct probability of containing the true map location of the QTL, for almost all QTLs, whereas the classical confidence interval can be very biased for QTLs having small effect.

FOLLOWING SAX (1923), many methods have been developed in the literature for detecting quantitative trait loci (QTL) using marker information. Estimating QTL map location is possible using "interval mapping" procedures based on maximum likelihood estimation (LANDER and BOTSTEIN 1989) or on linearized version of maximum likelihood (KNAPP *et al.* 1990; HALEY and KNOTT 1992). As pointed out by DARVASI *et al.* (1993), a very important quantity is the confidence interval of the QTL position on the chromosome.

CONNEALLY *et al.* (1985), in the field of linkage analysis, and LANDER and BOTSTEIN (1989) proposed the use of a confidence interval based on limiting $\chi^2$ distribution of the likelihood ratio test between two positions. This idea leads to a very simple construction of the confidence interval: take the maximum value of the LOD score, substract $c$, and the locations corresponding to this value of the LOD score are the extremes of the confidence interval (at 96.8% when using $c = 1$).

As we will see, simulations show that the $\chi^2$ approximation is not correct for QTLs having small effect, and therefore that the corresponding confidence intervals are biased, *e.g.*, the probability of the QTL being in an interval is less than the nominal level. The reason for this is that the regularity conditions for the convergence of the likelihood ratio test toward a $\chi^2$ distribution are not fulfilled for QTLs having small effect.

In this report, we will study the asymptotic properties of the likelihood ratio test for QTLs having small effect and construct a new test and therefore a new unbiased confidence interval.

## MODEL AND CLASSICAL CONFIDENCE INTERVAL

We consider a backcross sample of size $n$. Consider a QTL present at the position $d$ on a chromosome of length $L$. The trait value has a normal distribution with means $\mu_A$ and $\mu_B$ for the two QTL genotypes present in the backcross population and the same variance $\sigma^2$, for both genotypes. We will use $a = \mu_A - \mu_B$ and $\mu = (\mu_A + \mu_B)/2$. For each individual $k = 1, \ldots, n$, we score the phenotypic value of the trait $y_k$ and a set of $j = 1, \ldots, J$ markers with the information $M_{j,k}$ that takes the values $A$ or $B$ depending on the allele of the marker. The vector of phenotypic observation will be denoted by $Y$, and the vector of all marker information by $\mathcal{M}$.

The likelihood of $Y$, conditional on the marker information is $L_{\mathcal{M}}(Y, a, \mu, \sigma^2, d)$. Its complete expression is given by LANDER and BOTSTEIN (1989)

$$L_{\mathcal{M}}(Y, a, \mu, \sigma^2, d) = \prod_k [G_{\mathcal{M}}(k, d) l_k(1) + (1 - G_{\mathcal{M}}(k, d)] l_k(0))$$

where $G_{\mathcal{M}}(k, d)$ is the probability for individual $k$ to have genotype $A$ at position $d$, conditional on the marker information $\mathcal{M}$; $l_k(x) = \phi(y_k - \mu - xa, \sigma^2)$, for $x$ equal 0 or 1, is the probability density function for the normal distribution with mean 0 and variance $\sigma^2$. In the following, we assume no interference in recombination events and therefore use Haldane's map function. This function associates distance $d$ with recombination probability $r(d) = 0.5(1 - \exp(-2d))$.

A confidence interval can be built as follows (CONNEALLY *et al.* 1985): at each point $d_0$ along the chromosome, perform the likelihood ratio test

$$R(d_0) = 2 \ln \frac{\sup_{a,\mu,\sigma^2} L_{\mathcal{M}}(Y, a, \mu, \sigma^2, d_0)}{\sup_{\mu,\sigma^2} L_{\mathcal{M}}(Y, a = 0, \mu, \sigma^2, d_0)}.$$

Note that the LOD score test is essentially the same test as the likelihood ratio test $R(d_0)$, where the $\log_{10}$ is used instead of the Naperian logarithm and the ratio is not multiplied by 2. Therefore, it is asymptotically distributed as $\log_{10} e\chi^2/2 = 0.27\chi^2$.

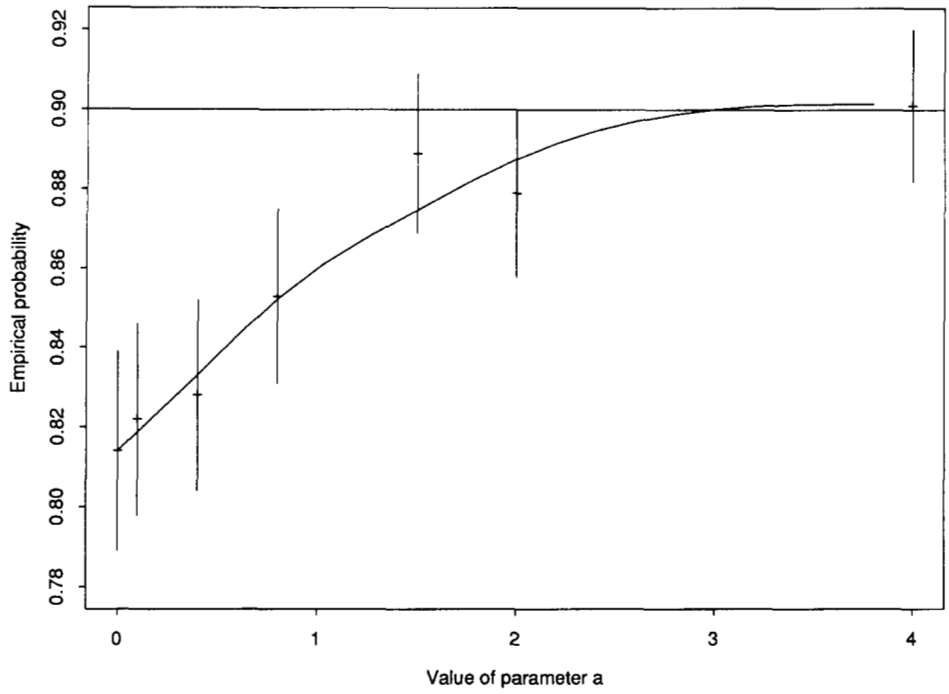


FIGURE 1.—Empirical probabilities that the confidence interval based on $T(d_0)$ contains the actual position of the QTL over 1,000 replications. Data for 200 backcross progeny were simulated with a 100 cM chromosome with markers each 20 cM. The QTL is located in the middle of the chromosome ($\sigma^2 = 1$). Simulations were performed with actual value a = {0, 0.1, 0.8, 1.5, 2, 4}. Vertical segments are for the 95% confidence interval of the empirical probabilities.

We can now calculate the test $T(d_0)$

$$T(d_0) = \sup_d [R(d)] - R(d_0)$$

$$= 2 \ln \frac{\sup_{a,\mu,\sigma^2,d} L_{\mathcal{M}}(Y, a, \mu, \sigma^2, d)}{\sup_{a,\mu,\sigma^2} L_{\mathcal{M}}(Y, a, \mu, \sigma^2, d_0)}.$$

Considering that $T(d_0)$ follows a $\chi^2$ distribution with one degree of freedom under the null hypothesis, that is $d_0$ is the correct position, the $(1 - \alpha)$ confidence interval is: $[d_{\text{inf}}, d_{\text{sup}}]$, where $d_{\text{inf}}$ ($d_{\text{sup}}$) is the smallest (the greatest) value of $d_0$ such that $T(d_0)$ is smaller than $\chi^2_{1,\alpha}$; $\chi^2_{1,\alpha}$ is the $\alpha$ quantile of a $\chi^2$ with 1 d.f.

The theory underlying this confidence interval is correct for any non-null finite value of $a$ and an infinite number $n$ of individuals. To investigate the quality of this confidence interval, we perform simulations for some $a$ values with $n = 200$, $\sigma^2 = 1$ and a chromosome of length 100 centiMorgan (cM) having markers each 20 cM (Figure 1) and 5 cM (Figure 2). The QTL is located in the middle of the chromosome in the first case, and at $d = 47.5$ cM in the second case. It appears that the confidence interval is unbiased for large values of $a$ but can be very biased for small values of $a$, particularly in the case of a dense map.

The reason for this is that for small value of $a$, the likelihood ratio test $T(d)$ does not follow a $\chi^2$ distribution, when the QTL is located at $d$. Table 1 shows that the quantiles of the distribution of $T(d)$ are different from those of a $\chi^2$. The difference depends on the $a$ values and is quite large when $a$ is small.

The following section gives a theoretical framework to deal with these small values of $a$.

## CONSTRUCTING A SIMILAR CONFIDENCE INTERVAL

A usual way to obtain a confidence interval based on the theory of tests is to defined a $1 - \alpha$ confidence interval as the set of values $d_0$ not "rejected" at level $\alpha$ using some function of $Y$, denoted $U(d_0)$, *i.e.*

$$\{d_0;\ U(d_0) \leq c_\alpha(d_0)\}$$

with

$$P[U(d_0 > c_\alpha(d_0)] = \alpha.$$

In models where there are nuisance parameters, the central requirement for this procedure is to be similar for all the nuisance parameters, *i.e.*, the probability of $U(d_0)$ being greater than $c(d_0)$ equals $\alpha$ for all the nuisance parameters. That is not the case for the classical procedure.

**Basic ideas:** To obtain a similar procedure for all the nuisance parameters $(\mu, \sigma^2, a)$, we propose to use a similar test, as described by COX and HINKLEY (1974). The basic idea for similarity is to find statistics whose distribution does not depend on the parameter $a$ under the null hypothesis: the QTL is located at $d_0$. Suppose that, under the null hypothesis, $\hat{a}_{d_0}$ is a sufficient statistic for the parameter $a$ then a good way to obtain similar procedure is to work with the conditional distribution of $Y$ given $\hat{a}_{d_0}$.

The second idea is to work in the local asymptotic framework. This framework is used in asymptotic theory to obtain the power of maximum likelihood ratio test, whose power is not trivially equal to 1. It is the correct framework to deal with QTLs that can be detected with powers ranging from 20 to 90% (REBAÏ *et al.* 1994). Note that when the classical interval is correct, the power of

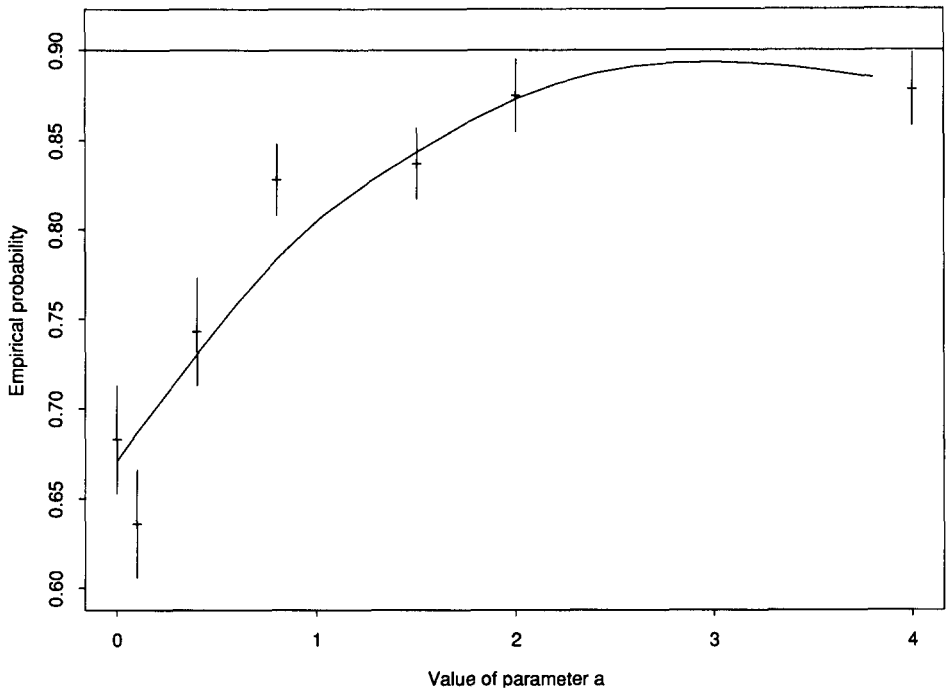


FIGURE 2.—Empirical probabilities that the confidence interval based on $T(d_0)$ contains the actual position of the QTL over 1,000 replications. Data for 200 backcross progeny were simulated with a 100 cM chromosome with markers each 5 cM. The QTL is located at a distance of 47.5 cM from one end of the chromosome ($\sigma^2 = 1$). Simulations were performed with actual value $a = \{0,0.1,0.8,1.5,2,4\}$. Vertical segments are for the 95% confidence interval of the empirical probabilities.

**TABLE 1**

**Empirical quantiles and their 95% confidence interval for the distribution of $T(d_0)$ over 1,000 replications**

| | Quantile | |
|---|---|---|
| | 10% | 5% |
| $\chi^2_1$ | 2.71 | 3.84 |
| $a = 0.1$ | $^{4.70}$ 5.07 $_{5.42}$ | $^{5.96}$ 6.36 $_{7.42}$ |
| $a = 0.4$ | $^{4.21}$ 4.51 $_{5.18}$ | $^{5.18}$ 5.96 $_{6.42}$ |
| $a = 0.8$ | $^{3.34}$ 3.75 $_{4.09}$ | $^{4.49}$ 4.89 $_{5.69}$ |

In each cell, the number in the middle is the empirical quantile and the italic numbers in the corners are the lower and upper bounds of a 95% confidence interval for the quantile.

Data for 200 backcross progeny where simulated with a 100-cM chromosome with markers each 5 cM. The QTL is located at a distance of 47.5 cM from one end of the chromosome ($\sigma^2 = 1$).

**TABLE 2**

**Empirical power (in %) of interval mapping for QTL detection over 1,000 replications**

| | $n$ | | | | | |
|---|---|---|---|---|---|---|
| | 200 | | | | 800 | |
| | 5% quantile at density (cM): | | 1% quantile at density (cM): | | 5% quantile at density (cM): | 1% quantile at density (cM): |
| $a$ | 20 | 5 | 20 | 5 | 20 | 20 |
| 0.1 | 8.5 | 9.0 | 1.0 | 1.2 | 18.2 | 4.2 |
| 0.4 | 59.8 | 64.0 | 28.4 | 30.0 | 99.3 | 95.2 |
| 0.8 | 99.3 | 99.9 | 95.2 | 98.5 | 100.0 | 100.0 |

The QTL is located in the middle of the chromosome for a marker density of 20 cM and at a distance of 47.5 cM from one end of the chromosome for a marker density of 5 cM ($\sigma^2 = 1$).

the maximum likelihood ratio test for QTL detection converges to 1 (Table 2).

Formally, in the local asymptotic framework, as $n \to \infty$, $a$ is assumed to tend to 0 in such a way that $a\sqrt{n}$ converges to a finite constant $\delta$ (COX and HINKLEY 1974, p.317). FEINGOLD *et al.* (1993) used the same asymptotic framework.

In this framework, $T(d_0)$ is not asymptotically distributed as a $\chi^2$ under the null hypothesis. This is because the information matrix is not positive-definite for $a = 0$, and therefore the classical Taylor expansions cannot be made in the neighborhood of $a = 0$. In particular, the parameter $d$ cannot be estimated consistently for $a = \delta/\sqrt{n}$, *i.e.*, the maximum likelihood estimator $\hat{d}$ of $d$ does not converge toward $d$ when the number of observations tends to infinity. This can be easily seen in a simple situation with only two markers that is treated in detail in the following section.

**The new test:** Working in the local asymptotic framework, asymptotically sufficient statistics can be found for the parameters $a$, $\mu$, $\sigma^2$, $d$. These are $\hat{\mu}$, the global mean, $\hat{\sigma}^2$, the classical estimator of the variance, and the mean class difference at each marker $S_j$; $j = 1, \ldots, J$

$$S_j = \frac{\sqrt{n}}{2}\left(\frac{\sum_k Y_k \mathbf{1}_{[M_{j,k}=A]}}{\sum_k \mathbf{1}_{[M_{j,k}=A]}} - \frac{\sum_k Y_k \mathbf{1}_{[M_{j,k}=B]}}{\sum_k Y_k \mathbf{1}_{[M_{j,k}=B]}}\right)$$

where $\mathbf{1}_{[M_{j,k}=\cdot]}$ is the indicator of the event $[M_{j,k} = \cdot]$.

Proof of asymptotic sufficiency of these statistics is straightforward using the work of REBAÏ *et al.* (1994). They showed that the maximum likelihood estimators in the complete model of LANDER and BOTSTEIN (1989) (known to be asymptotically sufficient statistics) and the

regression estimators in the linearised model of KNAPP *et al.* (1990) and HALEY and KNOTT (1992) are asymptotically equivalent (*e. g.*, convergent in probability).

Consider now $\hat{a}_{d_0}$ the maximum likelihood estimator of $a$ if the QTL is located at $d_0$ and $Z(d_0)$ the vector of components $Z_j(d_0)$; $j = 1, \ldots, J-1$ defined by:

$$Z_j(d_0) = \frac{1}{\sqrt{\hat{\sigma}^2}}\left(\frac{S_j}{1-2r_{j,d_0}} - \frac{S_{j+1}}{1-2r_{j+1,d_0}}\right)$$

where $r_j$, $d_0$ denotes the probability of recombination between the marker $j$ and a QTL located at $d_0$.

Proposition 1 (proven in APPENDIX [A1]) shows that $Z(d_0)$ is asymptotically a similar statistic for all the nuisance parameter when the QTL is supposed to be located at $d_0$.

**Proposition 1.** *Under the null hypothesis–the QTL is located at $d_0$–we get:*

$$\begin{pmatrix}\sqrt{n}\hat{a}_{d_0}\\ Z(d_0)\end{pmatrix} \xrightarrow{\mathcal{L}} N\left(\begin{pmatrix}\delta\\ 0\end{pmatrix}, \begin{pmatrix}\sigma^2 v_{d_0} & 0\\ 0 & V\end{pmatrix}\right)$$

*where the matrix V depends only on the length of the chromosome and the position of the makers.*

Proposition 2 (proven in APPENDIX [A2]) gives the asymptotic distribution of $Z(d_0)$ when the QTL is supposed to be located at $d$.

**Proposition 2.** *Under the alternative hypothesis–the QTL is located at d–, we get:*

$$Z(d_0) \xrightarrow{\mathcal{L}} N\left[X(d, d_0)\frac{\delta}{\sigma}, V\right]$$

*where the vector $X(d, d_0)$ depends only on the length of the chromosome, the position of the makers, d and $d_0$.*

Using the asymptotic distribution of $Z(d_0)$ we can build a maximum likelihood ratio test $T_Z(d_0)$,

$$T_Z(d_0) = 2\ln\frac{\sup_{\tilde{\delta}d} L^a(Z(d_0); \tilde{\delta}, d)}{L^a(Z(d_0); d_0)}$$

where $L^a(\cdot)$ means that the likelihood is calculated with the asymptotic distribution of $Z(d_0)$ and $\tilde{\delta} = \delta/\sigma$.

This gives:

$$T_Z(d_0) = \sup_d \frac{W^2(d, d_0)}{\mathrm{Var}_a(W(d, d_0))} \qquad (1)$$

where:

$$W(d, d_0) = X(d, d_0)' V^{-1} Z(d_0)$$

and $\mathrm{Var}_a(\cdot)$ denotes the variance for the asymptotic distribution of $Z(d_0)$.

The algebric expression of $T_Z(d_0)$ is given in the APPENDIX [A3].

## THRESHOLD CALCULATIONS

To be able to use the confidence interval built with $T_Z(d_0)$, we need the asymptotic distribution of this statistic under the hypothesis that the QTL is located at $d_0$. This distribution does not depend on the parameters $\delta$, $\sigma^2$ and $\mu$, but may depend on the length of the chromosome, the number and the position of the markers and the position $d_0$.

**The case with only two markers:** In this situation, the statistic $W(d, d_0)$ does not depend on $d$. So, the asymptotic distribution of $T_Z(d_0)$ under the null hypothesis is a $\chi^2_1$. Looking at the algebric expression of $W(d, d_0)$, given in the APPENDIX [A3], we see that:

$$T_Z(d_0) = \frac{Z_1^2(d_0)}{\left(\frac{1}{(1-2r(d_0))^2} + \frac{(1-2r(d_0))^2}{(1-2p)^2} - 2\right)}$$

where $p$ is the recombination probability between the two markers and $r(d_0)$ the recombination probability between the QTL and the first marker.

We get as a $1-\alpha$ confidence interval, the set of points:

$$\left\{ d_0; \left(\frac{S_1}{1-2r(d_0)} - \frac{S_2(1-2r(d_0))}{1-2p}\right)^2 < \chi^2_{1,\alpha}\left(\frac{1}{(1-2r(d_0))^2} + \frac{(1-2r(d_0))^2}{(1-2p)^2} - 2\right)\right\}.$$

The end points of the confidence region for the parameter $(1-2r(d_0))^2$ appear to be the solution of a quadratic function. In particular, the confidence region is not symmetric around the maximum likelihood estimator of $d$. Note that we observe here the same type of result as FIELLER (1954) with the confidence interval of the ratio of two random variables.

Another feature of the case with only two markers is the fact that the classical confidence interval and the new confidence interval could be the same. APPENDIX [A4] explains this particularity and the non-consistency of the likelihood estimator of $d$ in the local asymptotic framework.

**The case with more than 2 markers:** In this case, the asymptotic distribution of $T_Z(d_0)$ is the distribution of the supremum of a $\chi^2_1$ process with a covariance function depending on $d_0$. As it is difficult to obtain this distribution using analytical arguments, we propose to use simulations. These simulations are made using the asymptotic distributions of the $S_j$ for $j = 1, \ldots, J$.

Figures 3 and 4 show the $c_\alpha(d_0)$ threshold functions of this distribution for the 5 and 10% levels and for different numbers of markers equally spaced along the chromosome. In these cases, the threshold function $c_\alpha(d_0)$ is symmetric around the middle of the chromosome so we give this function for half of the chromosome.

## RESULTS AND DISCUSSION

Because the new confidence interval is constructed using asymptotic arguments in the local asymptotic framework, it is important to determine their qualities in real situations. This has been done using simulations.

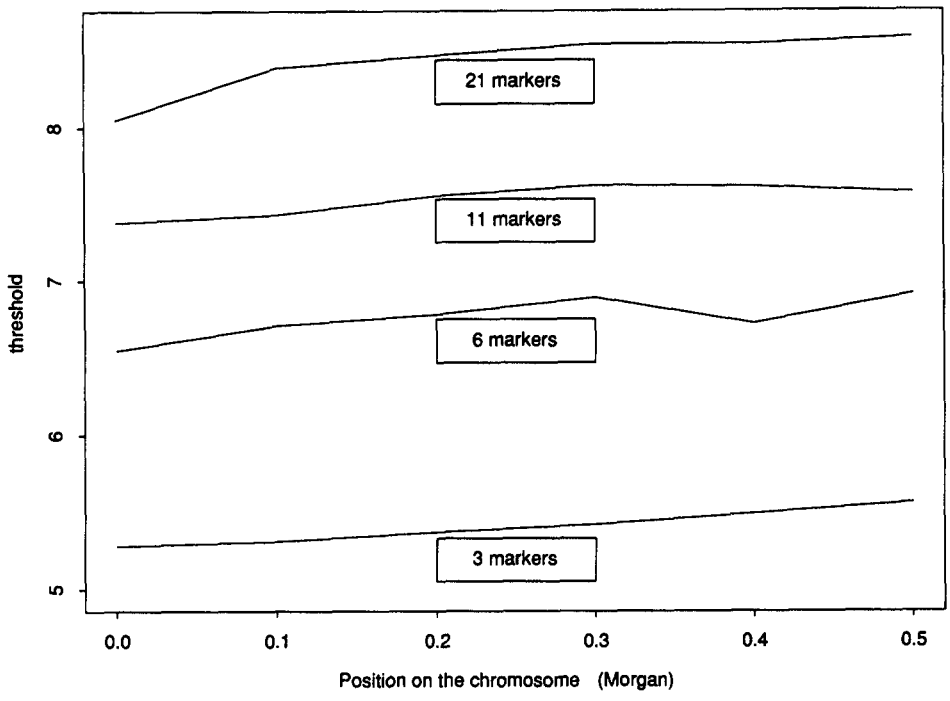


FIGURE 3.—Empirical threshold of the $T_Z(d_0)$ distribution for the 5% level over 50,000 replications.

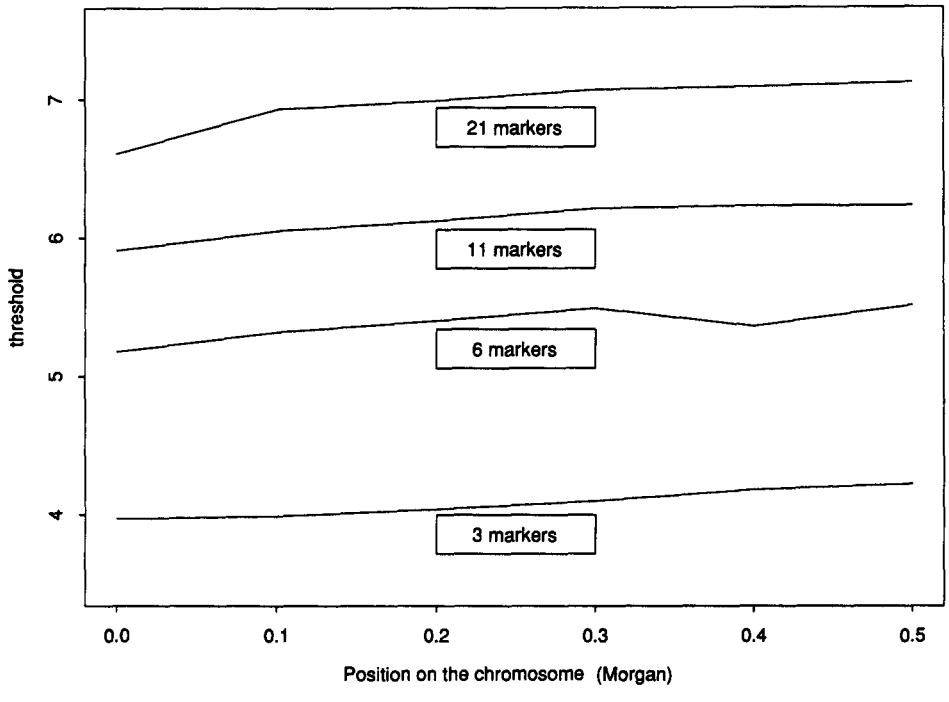


FIGURE 4.—Empirical threshold of the $T_Z(d_0)$ distribution for the 10% level over 50,000 replications.

Table 3 gives the empirical probability for the interval to contain the actual position of the QTL. Simulations are made with a chromosome of 1 Morgan, with markers at each 20 or 5 cM with $n = 200$ or $n = 800$. It appears that the confidence interval is unbiased for all values of a that have been used.

We show in detail a simulated example in Figure 5. This simulation has been performed with $n = 200$, $a = 0.4$, $\sigma^2 = 1$ and a marker at each 5 cM. The dashed line represents $T(d)$ and the classical confidence interval is the set of the points behind the threshold 2.71. The full line represents $T_Z(d)$ and the threshold is that shown in Figure 4. In this case, the actual position of the QTL, shown with an arrow on Figure 5, is not in the classical confidence interval but is in the new one. We obtain this type of result in about 20% of the replications.

The practical use of the new confidence interval needs the computation of $T_Z(d_0)$ using the formula (1) and its algebric expression given in APPENDIX [A3]. Then the correct threshold of the test statistic $T_Z(d_0)$ for each

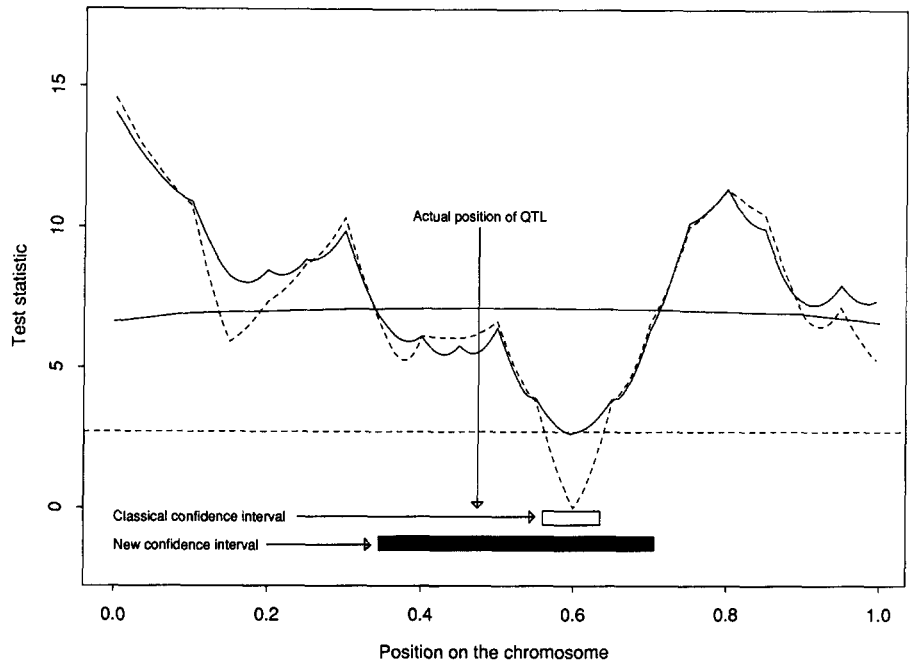


FIGURE 5.—Confidence intervals for the position of the QTL built respectively on $T(d_0)$ and $T_Z(d_0)$. Data for 200 backcross progeny were simulated with a 100 cM chromosome with markers each 5 cM ($\sigma^2 = 1$). The actual position of the QTL is pointed with the vertical arrow and its actual value is $a = 0.4$. The full line is $T_Z(d_0)$ and the threshold is that shown in Figure 4. The dash line is $T(d_0)$ and the confidence interval is the set of points below the threshold 2.71.

**TABLE 3**

**Empirical probabilities (in %) that the confidence interval based on $T_Z(d_0)$ contains the actual position of the QTL over 10,000 replications**

| | *n* | | |
|---|---|---|---|
| | 200 at density (cM): | | 800 at density (cM): |
| *a* | 20 | 5 | 20 |
| 0.1 | 90.0 | 90.3 | 90.6 |
| 0.5 | 89.5 | 89.6 | 89.9 |
| 1 | 89.6 | 89.8 | 89.9 |
| 2 | 90.4 | 89.8 | 89.4 |
| 4 | 89.7 | 89.6 | 89.8 |

The QTL is located in the middle of the chromosome for a marker density of 20 cM and at a distance of 47.5 cM from one end of the chromosome for a marker density of 5 cM ($\sigma^2 = 1$). All the empirical probabilities are in the 99% confidence region of 0.9.

position $d_0$ in a specific situation must be found. In Figures 3 and 4, we give the threshold function for some situations. For other situations, far from the equally spaced number of markers studied, specific simulations should be performed. Even if the use of $T_Z(d_0)$ seems to be more complicated than the use of $T(d_0)$, it must be preferred because it guarantees an unbiased confidence interval. Moreover, a correct threshold for $T(d_0)$ cannot be obtained because it depends on the value of the unknown parameter $a$.

In the general case, we have no information on the power of the test $T_Z(d_0)$ and therefore on the length of the new interval. However, we worked with asymptotic sufficient statistics and can argue for completion of $\{\hat{\mu}, \hat{\sigma}^2, \hat{a}_{d_0}\}$ in the local asymptotic framework under the null hypothesis. Then in this framework, $Z(d_0)$ contains all the informations coming from $Y$, concerning $d$ and not depending on the nuisance parameters under the hypothesis that the QTL is at position $d_0$. COX and HINKLEY (1974, p. 135) used the argument of completion to ensure that the region constructed with the likelihood ratio test of the distribution of the data conditioned on a complete sufficient statistic is the uniformly most powerful similar region. The main difference is that, in our work, we got the properties of sufficiency and completeness only asymptotically and locally.

Another problem is to calculate the probability that the confidence interval contains the actual position of the QTL, conditional on to the fact that the LOD score test is greater than its threshold.

FEINGOLD *et al.* (1993), in the framework of identity by descent mapping, gave approximate confidence regions for gene position based on sophisticated theoretical developments about point processes in the case of an ideally dense map. Their approach could be adapted in our situation and would give interesting results for this kind of map.

An interesting use of these confidence intervals is when one wants to test the consistency of a QTL over different environments or crosses. For instance, suppose a QTL was detected in some environment with a specific effect and position on the chromosome. A QTL in the same region of the chromosome was also detected in another environment but with different genetic effect (*e.g.*, PATTERSON *et al.* 1991). Some developments around our method could permit a test of whether the locations of the QTLs in both environments are the same.

Having a correct confidence interval for the QTL position is also of major interest in gene introgression experiments, by repeated backcrosses (*e.g.*, oligogenic disease resistance, MELCHINGER 1990). It would increase the efficiency and the reliability of such marker-facilitated selection programs.

Thanks are due to the referees whose comments clarified many issues and led to a better presentation.

## LITERATURE CITED


CONNEALLY, P. M., J. H. EDWARDS, K. K. KIDD, J. M. LALOUEL, N. E. MORTON *et al.*, 1985 Reports of the committee methods of linkage analysis and reporting. Cytogent. Cell Genet. **40:** 356–359.

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 gene effect and map location using a saturated genetic map. Genetics **134:** 943–951.

FEINGOLD, E., P. O. BROWN AND D. SIEGMUND, 1993 Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. Am. J. Hum. Genet. **53:** 234–251.

FIELLER, E. C., 1954 Some problems in interval estimation. J. R. Stat. Soc. B **16:** 175–185.

HALEY, C. S., AND S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci by using molecular markers. Heredity **69:** 315–324.

KNAPP, S. J., W. C. BRIDGES AND D. BIRKES, 1990 Mapping quantitative trait loci using molecular marker linkage maps. Theor. Appl. Genet. **79:** 583–592.

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

MELCHINGER, A. E., 1990 Use of molecular markers in breeding for oligogenic disease resistance. Plant Breed. **104:** 1–19.

PATERSON, A. H., S. DAMON, J. D. HEWITT, D. ZAMIR, H. D. RABINIWITCH *et al.*, 1991 Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. Genetics **127:** 181–197.

REBAÏ, A., B. GOFFINET AND B. MANGIN, 1994 Comparing power of different methods for QTL detection. Biometrics (in press).

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


Communicating editor: B. S. WEIR

## APPENDIX

**[A1]** In the local asymptotic framework, the maximum likelihood estimators are asymptotically equivalent to the regression estimators using the linearized model:

$$y_k = \mu + aG_M(k, d) + \epsilon_k$$

where the $\epsilon_k$ are independent and identically distributed as normal with mean 0 and variance $\sigma^2$ (REBAI *et al.* 1994).

Using the linearized model, it is simple to see that $\hat{\mu}$, $\hat{\sigma}^2$ and the $S_j$ for $j = 1, \ldots, J$ are asymptotic sufficient statistics for $\mu$, $\sigma^2$, $\delta$ and $d$.

In the following, we will use that $\sum_{k=1}^{n} \mathbf{1}_{[M_{j,k}=A]}/n$ and $\sum_{k=1}^{n} \mathbf{1}_{[M_{j,k}=B]}/n$ both converge in probability to ½.

For the null hypothesis–the QTL is located at $d_0$–we get:

$$\widehat{a_{d_0}} \simeq 2\,\frac{\left(x^2_{i_{d_0},d_0} - x^2_{i_{d_0},i_{d_0}+1}\right)\dfrac{S_{i_{d_0}}}{x_{i_{d_0},d_0}} + \left(\dfrac{x^2_{i_{d_0},i_{d_0}+1}}{x^2_{i_{d_0},d_0}} - x^2_{i_{d_0},i_{d_0}+1}\right)\dfrac{x_{i_{d_0},d_0}S_{i_{d_0}+1}}{x_{i_{d_0},i_{d_0}+1}}}{\sqrt{n}\left(\left(x^2_{i_{d_0},d_0} - x^2_{i_{d_0},i_{d_0}+1}\right) + \left(\dfrac{x^2_{i_{d_0},i_{d_0}+1}}{x^2_{i_{d_0},d_0}} - x^2_{i_{d_0},i_{d_0}+1}\right)\right)}$$

where $i_{d_0}$ is the left marker of the interval where the QTL is located, $r_{j,d_0}$ ($r_{i,j}$) is the recombination rate between the QTL and the marker $j$ (between the markers $i$ and $j$) and $x_{a,b} = (1 - 2r_{a,b})$.

As $\hat{\sigma}^2$ converge in probability to $\sigma^2$, it is sufficient to study the distribution of the statistics $S_j$ for $j = 1, \ldots, J$ in the linearized model, to get the asymptotic distribution of $(\sqrt{n}\widehat{a_{d_0}}\ Z(d_0))$.

In the linearised model, $S$, the vector of components $S_j$, is multinormal. This implies that $\widehat{a_{d_0}}$ and $Z(d_0)$ are asymptotically and locally multinormal.

Under the null hypothesis:

$$E(S_j) \simeq [(1 - 2r_{j,d_0})/2]\delta.$$

So we get:

$$\begin{aligned} E(Z_j(d_0)) &\simeq 0 \\ \mathrm{Cov}(S_j, S_{j+l}) &\simeq \sigma^2(\Pr(M_j = A, M_{j+l} = A) \\ &\quad + \Pr(M_j = B, M_{j+l} = B) \\ &\quad - \Pr(M_j = A, M_{j+l} = B) \\ &\quad - \Pr(M_j = B, M_{j+l} = A)) \\ &\simeq \sigma^2(1 - 2r_{j,j+l}) \end{aligned} \qquad (2)$$

So, the matrix $V$ depends only on the length of the chromosome and the location of the markers.

Besides, given three points on the genome, denoted $a$, $b$, $c$, located in term of probability of recombination by $r_{a,b}$, $r_{b,c}$ and $r_{a,c}$, if $b$ is located between $a$ and $c$ we get, assuming no interference in recombination events:

$$x_{a,c} = x_{a,b}x_{b,c} \qquad (3)$$

Using (3), we get:

$$\mathrm{Cov}(\widehat{a_{d_0}}, Z_j(d_0)) \simeq 0$$

**[A2]** To obtain the asymptotic distribution of $Z(d_0)$, we study the distribution of $S$ in the linearised model under the alternative hypothesis–the QTL is located at $d$–.

$$E(S_j) \simeq [(1 - 2r_{j,d})/2]\delta$$

We then get:

$$E(Z_{d_0}) \simeq X(d, d_0)(\delta/\sigma)$$

where the $j$th component of the vector $X(d, d_0)$ is:

$$X(d, d_0)_j = (x_{j,d}/x_{j,d_0} - x_{j+1,d}/x_{j+1,d_0}) \qquad (4)$$

The same arguments than in APPENDIX [A1] provides for $Z(d_0)$ an asymptotic normal distribution with variance V.

**[A3]** The expression of $W(d,d_0)$ is:

$$\begin{cases} W(d,d_0) = \alpha_{i_{d_0}} Z_{i_{d_0}}(d_0) + \alpha_{i_d} Z_{i_d}(d_0) + \sum_{j=i_d+1}^{i_{d_0}-1} Z_j(d_0) & \text{for } i_d < i_{d_0} \\ W(d, d_0) = Z_{i_{d_0}}(d_0) & \text{for } i_d = i_{d_0} \\ W(d, d_0) = \tilde{\alpha}_{i_{d_0}} Z_{i_{d_0}}(d_0) + \tilde{\alpha}_{i_d} Z_{i_d}(d_0) + \sum_{j=i_{d_0}+1}^{i_d-1} Z_j(d_0) & \text{for } i_d > i_{d_0} \end{cases}$$

with

$$\alpha_{i_{d_0}} = \frac{x^2_{i_{d_0}+1,d_0}(1 - x^2_{i_{d_0},d_0})}{x^2_{i_{d_0},d_0}(1 - x^2_{i_{d_0}+1,d_0}) + x^2_{i_{d_0}+1,d_0}(1 - x^2_{i_{d_0},d_0})}$$

$$\alpha_{i_d} = \frac{x^2_{i_d,d}(1 - x^2_{i_d+1,d})}{1 - x^2_{i_d,i_d+1}} \quad \bar{\alpha}_{i_{d_0}} = 1 - \alpha_{i_{d_0}} \quad \bar{\alpha}_{i_d} = \frac{x^2_{i_d+1,d}(1 - x^2_{i_d,d})}{1 - x^2_{i_d,i_d+1}}.$$

Using that $V$ is a diagonal matrix with diagonal element equal to:

$$V_{jj} = \frac{1}{x^2_{j,d_0}} - 2\frac{x_{j,j+1}}{x_{j,d_0}x_{j+1,d_0}} + \frac{1}{x^2_{j+1,d_0}}$$

the algebric expression of $T_Z(d_0)$ is easily found.

**Details of calculations:** Using (2) we get:

$$V_{j,j+l} = \frac{x_{j,j+l}}{x_{j,d_0}x_{j+l,d_0}} - \frac{x_{j,j+l+1}}{x_{j,d_0}x_{j+l+1,d_0}} - \frac{x_{j+1,j+l}}{x_{j+1,d_0}x_{j+l,d_0}} + \frac{x_{j+1,j+l+1}}{x_{j+1,d_0}x_{j+l+1,d_0}}.$$

Now, using (3) and the QLT position under the null hypothesis, give $V_{j,j+1} = 0$ for $l > 0$.

Suppose that $i_d < i_{d_0}$, we obtain using (2), (3) and (4):

$$\frac{X(d, d_0)_j}{V_{j,j}} = \begin{cases} 0 & j < i_d \\ x_{d,d_0}\left(\dfrac{x^2_{i_d,d}(1 - x^2_{i_d+1,d})}{1 - x^2_{i_d,i_d+1}}\right) & j = i_d \\ x_{d,d_0} & i_d < j < i_{d_0} \\ x_{d,d_0}\left(\dfrac{x^2_{i_{d_0}+1,d_0} - x^2_{i_{d_0}+1,d_0}x^2_{i_{d_0},d_0}}{x^2_{i_{d_0}+1,d_0} + x^2_{i_0,d_0} - 2x^2_{i_{d_0}+1,d_0}x^2_{i_{d_0},d_0}}\right) & j = i_{d_0} \\ 0 & i_{d_0} < j. \end{cases}$$

Because the test is invariant by scale, the constant $x_{d,d_0}$ can be left and we find the expression given for $W(d, d_0)$ in the case $i_d < i_{d_0}$.

For $i_d = i_{d_0}$, we get:

$$\frac{X(d, d_0)_j}{V_{j,j}} = \begin{cases} 0 & j < i_{d_0} \\ x_{d,d_0} & j = i_{d_0} \\ 0 & i_{d_0} < j. \end{cases}$$

And for $i_d > i_{d_0}$:

$$\frac{X(d, d_0)_j}{V_{j,j}} = \begin{cases} 0 & j < i_d \\ x_{d,d_0}\left(\dfrac{x^2_{i_d+1,d}(1 - x^2_{i_d,d})}{1 - x^2_{i_d,i_d+1}}\right) & j = i_d \\ x_{d,d_0} & i_d < j < i_{d_0} \\ x_{d,d_0}\left(\dfrac{x^2_{i_{d_0},d_0} - x^2_{i_{d_0}+1,d_0}x^2_{i_{d_0},d_0}}{x^2_{i_{d_0}+1,d_0} + x^2_{i_0,d_0} - 2x^2_{i_{d_0}+1,d_0}x^2_{i_{d_0},d_0}}\right) & j = i_{d_0} \\ 0 & i_{d_0} < j. \end{cases}$$

**[A4]** In the local asymptotic framework, consider the asymptotic distribution, under the null hypothesis, of the vector $S^r$, which elements are $S_1$ and $S_2$ divided by $\sqrt{\hat{\sigma}^2}$:

$$S^r \xrightarrow{\mathcal{L}} N\left(U(d_0)\frac{\delta}{\sigma}, W\right)$$

with

$$U(d_0) = \begin{pmatrix} \dfrac{x_{1,d_0}}{2} \\ \dfrac{x_{1,2}}{2x_{1,d_0}} \end{pmatrix} \qquad W = \begin{pmatrix} 1 & x_{1,2} \\ x_{1,2} & 1 \end{pmatrix}$$

where $x_{i,j}$ is defined in APPENDIX [A1].

**Maximum likelihood estimator for the parameter $d$:** Using the linearised model and the vector $U(d)$, we found that an asymptotic equivalent statistic of the maximum likelihood estimator for $x^2_{i,d} = (1 - 2r(d))^2$ is:

$$x_{1,2}\, S_1/S_2.$$

This estimator does not converge towards $(1 - 2r(d))^2$. Therefore, the maximum likelihood estimator for the parameter $d$ does not converge toward $d$.

**Asymptotic equivalence between $T_Z(d_0)$ and $T_Z(d_0)$:** Denote $P^{W^{-1}}_{U(d_0)}$ the projector onto the linear space generated by $U(d_0)$ for the $W^{-1}$ norm, $P^{W^{-1}}_{U^\perp(d_0)}$ the projector onto the orthogonal of the space for the $W^{-1}$ norm and $\|\ldots\|^2_{W^{-1}}$ the square of the $W^{-1}$ norm.

When only 2 markers are present on the chromosome, it can be proved that $T(d_0)$ and $T_Z(d_0)$ are equiva-lent when

$$\sup_d \|P^{W^{-1}}_{U(d)}(S^r)\|_{W^{-1}} = \|S^r\|_{W^{-1}}.$$

The probability of this event is not null, because the region covered by $\{d; U(d)\}$ and the vector $S^r$ are both in a two dimensional space.

This result follows from the equality

$$\|S^r\|_{W^{-1}} = \|P^{W^{-1}}_{U(d_0)}(S^r)\|_{W^{-1}} + \|P^{W^{-1}}_{U^\perp(d_0)}(S^r)\|_{W^{-1}},$$

the asymptotic equivalences

$$R(d_0) \simeq \|P^{W^{-1}}_{U(d_0)}(S^r)\|_{W^{-1}}$$

$$\sup_d R(d) \simeq \sup_d \|P^{W^{-1}}_{U(d)}(S^r)\|_{W^{-1}}$$

$$T_Z(d_0) \simeq \|P^{W^{-1}}_{U^\perp(d_0)}(S^r)\|_{W^{-1}}$$

and the definition of $T(d_0) = \sup_d R(d) - R(d_0)$.