Threshold and power for Quantitative Trait Locus detection

Charles-Elie Rabier, Jean-Marc Azais, Jean-Michel Elsen, Céline Delmas

Threshold and power for Quantitative Trait Locus detection[☆]

C-E. Rabier[*,a,b], J-M. Azaïs[a], J-M. Elsen[b], C. Delmas[b]

[a] *Université de Toulouse, Institut de Mathématiques de Toulouse, U.P.S., F-31062 Toulouse Cedex 9, France.*
[b] *INRA UR631, Station d'Amélioration Génétique des Animaux, BP 52627-31326 Castanet-Tolosan Cedex, France.*

Abstract

We propose several new methods to calculate threshold and power for Quantitative Trait Locus (QTL) detection. They are based on asymptotic theoretical results presented in Rabier et al. (2009) . The asymptotic validity is checked by simulations. The methods proposed are fast and easy to implement. A comparison of power between a multiple testing procedure and a global test has been realized, showing far better performances of the global test for the detection of a QTL.

Key words: QTL detection, Likelihood Ratio Test, Chi-Square process, Multiple Testing, Threshold, Monte-Carlo methods

1. Introduction

We study the problem of detecting a Quantitative Trait Locus, so-called QTL (a gene influencing a quantitative trait which is able to be measured) on a given chromosome in a population of progenies which are structured into sire families. The back-cross population, $A \times (A \times B)$, where A and B are purely homozygous lines, is a particular case of such a population.
A method largely used in order to detect a QTL, is the Interval Mapping proposed by Lander and Botstein (1989). Using the Haldane (1919) distance

[*]Corresponding author. Tel.:+33 5 61 28 52 64; fax.:+33 5 61 28 53 53
Email addresses: charles-elie.rabier@toulouse.inra.fr (C-E. Rabier), azais@cict.fr (J-M. Azaïs), jean-michel.elsen@toulouse.inra.fr (J-M. Elsen), celine.delmas@toulouse.inra.fr (C. Delmas)

and modelling, each chromosome is represented by a segment $[0,T]$. The distance on $[0,T]$ is called the genetic distance (measured in Morgans). At each location $t \in [0,T]$, the presence of a QTL is tested with a Likelihood Ratio Test (LRT). So, multi-testing leads to a LRT process, and taking as test statistic the supremum of this process comes down to perform a LRT in a model when the localisation of the QTL is an extra parameter.
Some theoretical results are present in Rebaï et al. (1994, 1995), in Cierco (1998), and in Azaïs and Cierco-Ayrolles (2002). Howewer, these papers use some approximations. In Rabier et al. (2009), article submitted, the focus is on the exact model. The asymptotic distributions of the LRT process are given under the null hypothesis (no QTL on $[0,T]$), under the alternative that there is one QTL at $t^\star$ on $[0,T]$, and under the general alternative that there are m QTL on $[0,T]$. The focus here is on the null hypothesis and on the particular alternative that there is one QTL on $[0,T]$.

In this paper :

1. we propose methods, as a function of the genetic map, to calculate thresholds for the supremum of the LRT process under H_0.
2. as all these methods are based on asymptotic results, we check the validity of the asymptotic assumption by simulating samples of different sizes.
3. we study the asymptotic power of the Interval Mapping and we give advices on how to optimize the detecting process.

The methods studied are available in a Matlab package with graphical user interface : "imapping.zip".
It can be downloaded at www.math.univ-toulouse.fr/∼rabier .

2. Model

The chromosome is the segment $[0,T]$. K genetic markers are located on the chromosome, one at each extremity. $t_1 = 0 < t_2 < ... < t_K = T$ are the locations of the markers. The "genome information" at t will be denoted $X(t)$. The Haldane (1919) model is the following : the law of $X(t_1)$ is $\frac{1}{2}(\delta_1 + \delta_{-1})$ and $X(t) = (-1)^{N(t)}X(t_1)$ where $N(t)$ is a standard Poisson process. Indeed, the Haldane model assumes that crossovers occur at random and independently of each other. The Haldane (1919)'s function

$r : [0,\ T]^2 \longmapsto \left[0,\ \frac{1}{2}\right]$ is such as :

$$r(t,t') = \mathbb{P}(X(t)X(t') = -1) = \mathbb{P}(|N(t) - N(t')| \text{ odd}) = \frac{1}{2}\left(1 - e^{-2|t-t'|}\right)$$

This function links the recombination rate $r(t,t')$ between two loci located respectively at t and t', and the distance $|t-t'|$ between the two loci.

A sire family is defined by a set of progenies. I families will be considered. C is a discrete random variable refering to the family. The individual belongs to family i with probability $\pi_i = P(C = i)$.
The quantitative trait Y depends on the value of $X(t)$ at $t^\star \in [t_1, t_K]$ which is the location of the QTL. It also depends on the family it belongs to. The quantitative trait verifies :

$$\left(Y \middle| C = i\right) \ = \ \mu_i \ + \ X(t^\star)\, q_i \ + \ \sigma\, \varepsilon$$

where μ_i and q_i are respectively a polygenic effect and the QTL effect within family i. ε is a Gaussian white noise.
Besides, the "genome information" is available only at locations of genetic markers, that is to say at $t_1, t_2, ..., t_K$.
n is the number of observations j, (Y_j, $X_j(t_1)$, ..., $X_j(t_K)$, C_j). These observations are supposed to be independent and identically-distributed. We will call one population a sample of n observations.
The goal of this study is to test if there is a QTL on the chromosome with at least one of the sires heterozygous. The challenge is that $t^\star$ is unknown. So, the alternative hypothesis can be written :

$H_{\lambda t^\star}$: "there is at least one $q_i = \lambda_i/\sqrt{n}$, with $\lambda_i \in \mathbb{R}^\star$, at the position $t^\star$ "

In this context, we remind Theorem 3 of Rabier et al. (2009), which gives the asymptotic distribution of the LRT process, $\Lambda_n(.)$, under the null and the alternative hypothesis.

Theorem *With the previous defined notation,*

$$\Lambda_n(.) \overset{F.d.}{\rightarrow} \sum_{i=1}^{I} \left\{Z^i(.)\right\}^2 \tag{1}$$

as n tends to infinity, under H_0 and $H_{\lambda t^\star}$ where :

- $\overset{F.d.}{\rightarrow}$ *is the convergence of fini-dimensional distributions*
- *the* $Z^i(.)$ *are independent Gaussian processes. More precisely,* $Z^i(.)$ *is the continuous and the non linear process such as* $\forall t \in]t_k, t_{k+1}[$ *:*

$$Z^i(t) = \left\{ \alpha(t)\, Z^i(t_k) \;+\; \beta(t)\, Z^i(t_{k+1}) \right\} \tag{2}$$

where $Cov\{Z^i(t_k), Z^i(t_{k'})\} = e^{-2|t_k - t_{k'}|}$. *The mean function of* $Z^i(.)$ *verifies :*

  - *under* H_0, $m(t) = 0$
  - *under* $H_{\lambda t^\star}$, $m^i_{t^\star}(t)$ *is proportional to* $\lambda_i \sqrt{\pi_i}$.

We refer to Rabier et al. (2009) for the rather complicated expressions of the functions $m^i_{t^\star}(t)$, $\alpha(t)$, $\beta(t)$ *and the covariance* $\Gamma(t, t')$ *of* $Z^i(.)$.

Note that when the number of genetic markers is infinite, the process $Z^i(.)$ is an Ornstein-Uhlenbeck process. The paths of three processes are presented in Figure 1 (the length of the chromosome is $T = 1$ Morgan):

- the Ornstein-Uhlenbeck process.
- the process $Z^1(.)$ with only 2 markers, located at $t_1 = 0$ and $t_2 = 1$M.
- the process $Z^1(.)$ with markers located every 10cM.

The paths of the last two processes are smooth whereas the paths of the Ornstein-Uhlenbeck process are very jerky. It's not suprising because the Ornstein-Uhlenbeck process can be viewed as a stationnary version of the Brownian motion.

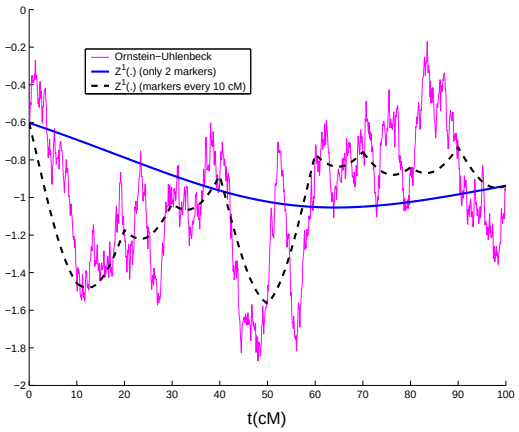

Figure 1: Paths of three different Gaussian processes

3. Different methods to obtain thresholds as a function of the map considered

3.1. Introducing the methods

We propose several new methods as a function of the map considered to calculate thresholds for the supremum of the LRT process under H_0. In particular, two kinds of maps are studied :

- a sparse map : a few markers covering the chromosome
- a dense map : a high density of markers pretty close to each other

We will suppose that when the map is dense, tests are performed only on markers, whereas when the map is sparse, test are also performed between markers.

Under a sparse map, thresholds can be obtained using the most appropriate methods as function of I :

- for $I = 1$, the problem comes down to computing the distribution of the maximum, i-absolute, value of a Gaussian vector. This can be done by a Discrete Monte-Carlo Quasi Monte-Carlo method (DMCQMC) : the method for numerical computation of a multivariate normal probability (Genz, 1992) can be considered. It uses a transformation that simplifies the problem and places it into a form that allows efficient calculation

using MCQMC methods. Note that a Newton's method can been used in order to obtain the threshold. This method is faster than using a simple Monte-Carlo method.

- for $I > 1$, a Discrete Monte-Carlo (DMC) method can be performed. According to the Theorem, when tests are performed only at the markers locations, the asymptotic process is a Discrete Ornstein-Uhlenbeck Chi-Square process with I degrees of freedoms (DOUCS(I)). The definition of such a process is given in the right-hand side of formula (1). When considered at the markers locations, the processes $Z^i(.)$ are simply AR(1) processes and we can interpolate by formula (2). In this situation, the threshold is easily obtained by a Discrete Monte-Carlo method based on a large number of sample paths ($nspaths$) of the asymptotic process.

Under a dense map, we propose theoretical methods to obtain the thresholds. Assuming that the number of genetic markers is infinite, the LRT process is asymptotically an Ornstein-Uhlenbeck Chi Square process with I degrees of freedom (OUCS(I)).
In a paper in progress, Rabier and Genz propose an approximative formula (named DF here) for the threshold of the supremum of such a OUCS(I) process. It is based on Delong (1981)'s work on Brownian motion. This formula is suitable when I and the threshold are large.
Besides, statistical tables given by Estrella (2003), for the threshold of the supremum of the OUCS(I), are also available. In order to obtain its exact tables, Estrella improved Delong's work on hypergeometrics functions. Estrella's method will be denoted ET.

Table 1 is a summary of all the methods proposed for the two kind of maps.

Map	Method
Dense (testing on markers)	ET (table available for $I \leq 20$) DF (formula available for I and threshold large)
Sparse (testing between markers)	$DMCQMC$ (available only for $I = 1$) DMC for $I > 1$

Table 1: Summary of all the methods studied as a function of the map considered and the way of performing tests (DMC for Discrete Monte-Carlo, $DMCQMC$ for Discrete Monte-Carlo Quasi Monte-Carlo, ET for Estrella exact table, DF for Delong approximative formula)

3.2. Applications under the null hypothesis

In this Section, the focus is on thresholds corresponding to the 95% quantile of the supremum of the LRT process under H_0. In order to illustrate the different methods, a sparse map and a dense map are considered. Since all the methods are based on asymptotic results (cf. Theorem), in order to check the validity of this results, some populations of different sizes have been simulated (*npop* denotes the number of populations whereas n denotes the size of a population).

Sparse map

The sparse map consists of a chromosome of length $T = 60$cM with 4 genetic markers equally spaced every 20cM. The presence of a QTL is tested every 5cM.

In Table 2, thresholds are presented as a function of I. In Table 3, the focus is on the number of false positives (NFP) as a function of the number of individuals n (thresholds taken from Table 2). Using Binomial distribution, a 95% confidence interval is calculated (into brackets in the tables) for the true percentage of the number of false positives.

According to Table 3, when there are in mean 200 individuals per family, that is to say $n = 200\,I$, NFP is not significantly different from 5%. When $n = 50\,I$, we can consider that NFP is still fair (even if it is significantly different from 5%) whereas when $n = 30\,I$, NFP is not so nominal.

Dense map

The dense map consists of a chromosome of length $T = 50$cM with 501 genetic markers equally spaced every 0.1cM.

The thresholds are compared in Table 4, and the NFP in Table 5. This aspect suggests fast convergence to asymptotic regime.

3.3. Remark

ET is not appropriate for sparse map for two reasons :

1. *ET* is based on Ornstein-Uhlenbeck (OU) process which is much more irregular than the process $Z^1(.)$ (see Figure 1). When $I = 1$, this can be formalized by the use of Slepian type inequalities, specially lemma (2.1) in Azaïs and Wschebor (2009) which comes from Plackett (1954). It can be proved that the covariances are smaller in the case of OU process than for the process $Z^1(.)$. It implies that the maximum of OU is stochastically greater than the $Z^1(.)$ one. Since $\mathbb{P}\left(\sup|Z^1(.)| > u\right) \approx$

$2\mathbb{P}\left(\sup Z^1(.) > u\right)$, this argument can be approximatively extended to the absolute value.

2. for the sparse map, the focus is not on continuous process but on dicrete process : the maximum of continuous process is always greater than the discrete one.

To sum up, ET will give too large thresholds.

Method	$DMCQMC\ (I=1)$	$DMC\ (I=3)$	$DMC\ (I=5)$
Threshold	6.06	10.76	14.47

Table 2: Thresholds obtained using the appropriate method as a function of the value of I considered ($nspaths = 1000000$). The map consists of 4 genetic markers equally spaced every 20cM (T=60cM). A test is done every 5cM.

Method / n	$DMCQMC\ (I=1)$	$DMC\ (I=3)$	$DMC\ (I=5)$
$n = 200\ I$	5.20% [4.98%; 5.42%]	5.03% [4.82%; 5.24%]	5.22% [5.00%; 5.44%]
$n = 50\ I$	5.78% [5.55%; 6.01%]	5.97% [5.74%; 6.20%]	6.11% [5.88%; 6.34%]
$n = 30\ I$	6.60% [6.36%; 6.84%]	6.77% [6.52%; 7.02%]	7.08% [6.83%; 7.33%]

Table 3: Number of False Positives (NFP) as a function of the number of individuals n and the method considered. The map consists of 4 genetic markers equally spaced every 20cM (T=60cM). A test is done every 5cM ($\sigma = 1$, $\mu_1 = -0.37$, $\mu_2 = 0.03$, $\mu_3 = 0.06$, $\mu_4 = -0.26$, $\mu_5 = 0.27$, $npop = 40000$).

	$I=1$			$I=3$		
Method	ET	DF	DMC	ET	DF	DMC
Threshold	7.84	7.61	7.68	13.09	12.91	12.86

Table 4: Thresholds obtained using theoretical methods ET, DF as function of the value of I considered. DMC for checking ($nspaths = 1000000$). The map consists of 501 genetic markers equally spaced every 0.1cM ($T = 0.5$M). A test is done on each marker.

Method / n	DF	DMC	ET
$n = 1000$	4.78% [4.57%; 4.99%]	5.13% [4.91%; 5.35%]	4.41% [4.21%; 4.61%]
$n = 500$	4.96% [4.75%; 5.17%]	5.15% [4.93%; 5.37%]	4.64% [4.43%; 4.85%]
$n = 150$	5.67% [5.44%; 5.90%]	5.91% [5.68%; 6.14%]	5.34% [5.12%; 5.56%]

Table 5: Number of False Positives (NFP) as a function of the number of individuals n and the method used. $I = 5$ here. The map consists of 501 genetic markers equally spaced every 0.1cM ($T = 0.5$M). A test is done on each marker ($\sigma = 1$, $\mu_1 = -0.37$, $\mu_2 = 0.03$, $\mu_3 = 0.06$, $\mu_4 = -0.26$, $\mu_5 = 0.27$, $npop = 40000$).

4. Study of the statistical power

Motivation

Some of the sires are heterozygotes at the QTL and others homozygous. QTL can only be detected in heterozygous sires families. Thus, two questions arise :

1. is it always profitable to include all the families in the analysis ?
2. do we have to analyze families all together or separately ?

We consider here, the sparse map of Section 3.2. As previously, tests are performed every 5 cM. The level considered is 5%.

About the QTL effects

When we deal with I families, since the total number of individuals is n, the expectation of the number of individuals in family i is only $n\pi_i$. So, in order to see the evolution of the power of the Interval Mapping with the number of families, we will consider $\lambda_i = \frac{\lambda}{\sqrt{\pi_i}}$ (note that if $I = 1$, $\lambda_1 = \lambda$ because $\pi_1 = 1$). As a consequence, the mean function, $m^i_{t^\star}(t)$, of the asymptotic process $Z^i(.)$, is proportional to λ and does not depend on i (cf. Theorem).

How to optimize the QTL detecting process

Only asymptotic results are studied here (cf. Theorem). Figures 2, 3, 4 illustrate question 1 whereas Figures 5, 6, 7 illustrate question 2.
In Figures 2, 3, 4, the power is plotted as a function of $t^\star$, I and the values of the λ_i's. In Figures 5, 6, 7, we compare the power of the approach which

consists of analyzing all families together (as previously), and the power of the approach which consists of analyzing families separately. For all of these Figures, $t^\star$ has been discretized with a step of 5cM.

Figures 2, 3, 4
For all of these three figures, two curves with the same colour on different Figures represent the same quantities.

As expected, the power increases with the number I of families (Figure 2 for $\lambda = 2$) and, for a given I, with the proportion of heterozygous sires (Figure 3 for $I = 5$, $\lambda = 2$ and various number number nz of non zeros λ_i's).

According to Figure 4, it is almost as powerful to consider $I = 1$ with $nz = 1$ (cf. grey curve) as $I = 5$ with $nz = 2$ (cf. green curve). So, it is much more powerful to consider $I = 1$ with $nz = 1$ (cf. grey curve) than $I = 5$ with $nz = 1$ (cf. brown curve). As a consequence, if they could be sorted in advance, it would be more powerful to concentrate the analysis on the families with a segregating QTL. Furthermore, once the families targeted, it would be more powerful to remove the families with very small QTL effects (not illustrated here). Indeed, it is like these families add noise to the model.

Figures 5, 6, 7
Practically, the segregating families are not known before the analysis and the true question is : do we have to analyze all the families together (global approach) or analyze families separately (Bonferroni approach) ? Indeed, since all the results are asymptotic, the variance is not better estimated when the global approach is considered.

Figures 5, 6,7 represent the two approaches. In these Figures, when the curve is :

- blue, $I = 5$
- cyan, $I = 7$
- orange, $I = 12$
- in solid line, it refers to the global approach
- in dashed line, it refers to the Bonferroni approach

We remind the null hypothesis,

$$H_0 : \text{“ } \forall i \in \{1, ..., I\}\,,\ \ q_i = 0 \text{ ”}$$

and the alternative hypothesis,

$$H_{\lambda t^\star} : \text{“ there is at least one } q_i = \lambda_i/\sqrt{n}\text{, with } \lambda_i \in \mathbb{R}^\star\text{, at the position } t^\star \text{ ”}$$

For the global approach, when H_0 is rejected, it only comes out that there is a QTL in at least one family, but this family is not known. For the Bonferroni approach, in order to answer the same question as for the global approach, we define the test statitistic U and the critical region CR, which results from a Bonferroni correction :

$$U = \left(\sup\left\{Z^1(.)\right\}^2,\ ...,\ \sup\left\{Z^I(.)\right\}^2\right)$$

$$CR = \left\{u = (u_1, ..., u_I) \in \mathbb{R}^I \text{ such as there is at least one } u_i \text{ verifying } u_i \geq c\right\}$$

where c is the threshold verifying : $\mathbb{P}\left(\sup\left\{Z_0^1(.)\right\}^2 \geq c\right) = \frac{0.05}{I}$.
$Z_0^1(.)$ is the Gaussian process centered and with covariance function $\Gamma(t, t')$. The Bonferroni correction allows to have $\mathbb{P}_{H_0}(U \in CR) \leq 0.05$. Obviously, the power of the Bonferroni approach is $\mathbb{P}_{H_{\lambda t^\star}}(U \in CR)$.
In Figure 5, the focus is on the particular case where there is only a QTL in family 1. $\lambda = 2$ has been taken. In the Figure, are represented the power of the two approaches as a function of $t^\star$ and I. It is noticeable that the Bonferroni approach is more powerful than the global approach. In Figure 6, the focus is on the particular case where there is a QTL in each family. In that case, the Bonferroni approach is outperformed by the global approach.
Figure 7 represents the mean power of the two approaches. Every alternative hypotheses have been considered (except the null hypothesis), i.e. for a given I, $nz = 1, ..., I$. Equiprobability concerning all these hypotheses has been supposed. According to the Figure, for a given I, there is a mean increase in term of power of at least 15% when the global approach is considered.

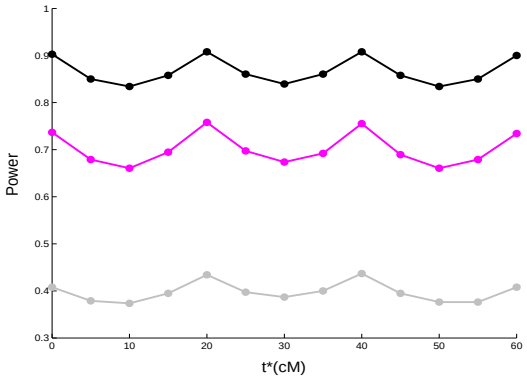

Figure 2: Power as a function of $t^\star$ and I. From top to bottom, $I = 5$, $I = 3$, $I = 1$ ($\lambda = 2$, $\sigma = 1$, $nspaths = 100000$).

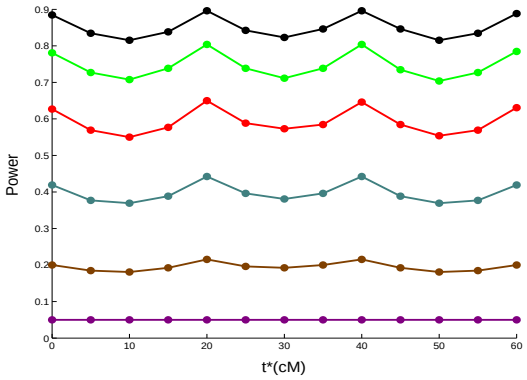

Figure 3: Power as a function of $t^\star$ and the number nz of non zero. From top to bottom, $nz = 5,\ 4,\ 3,\ 2,\ 1,\ 0$ ($I = 5$, $\lambda = 2$, $\sigma = 1$, $nspaths = 100000$).

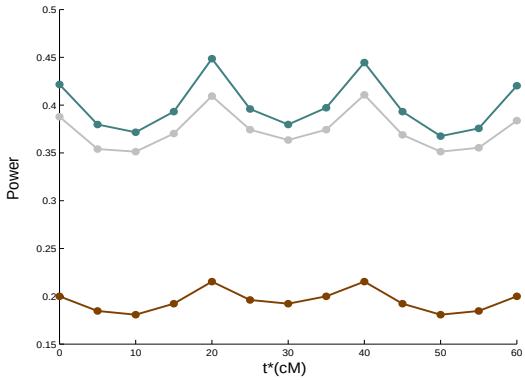

Figure 4: Power as a function of $t^\star$, I and the number nz of non zero. From top to bottom : $I = 5$ with $nz = 2$, $I = 1$ with $nz = 1$, $I = 5$ with $nz = 1$ ($\lambda = 2$, $\sigma = 1$, $nspaths = 100000$).

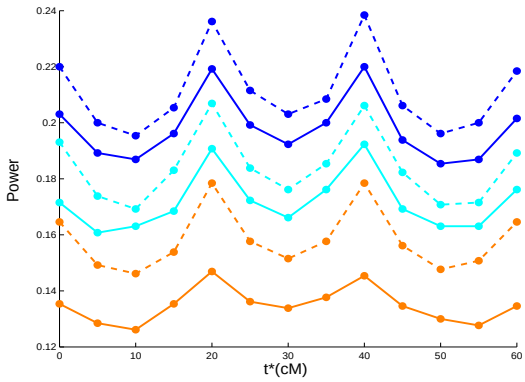

Figure 5: Power of the global approach (in solid line) and power of the Bonferroni approach (in dashed line), as a function of $t^\star$ and in the particular case of $nz = 1$. Orange refers to $I = 12$, cyan to $I = 7$ and blue to $I = 5$ ($\lambda = 2$, $\sigma = 1$, $nspaths = 100000$).

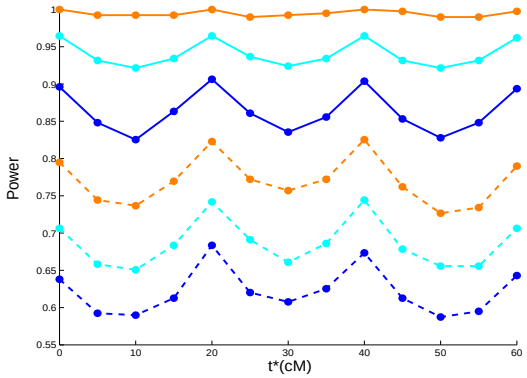

Figure 6: Power of the global approach (in solid line) and power of the Bonferroni approach (in dashed line), as a function of $t^\star$ and in the particular case of $nz = I$. Orange refers to $I = 12$, cyan to $I = 7$ and blue to $I = 5$ ($\lambda = 2$, $\sigma = 1$, $nspaths = 100000$).

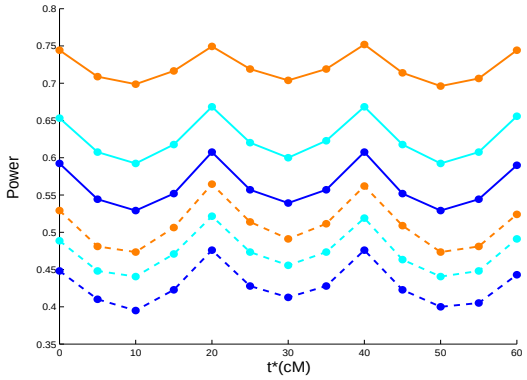

Figure 7: Mean power of the global approach (in solid line) and mean power of the Bonferroni approach (in dashed line), as a function of $t^\star$. Orange refers to $I = 12$, cyan to $I = 7$ and blue to $I = 5$ ($\lambda = 2$, $\sigma = 1$, $nspaths = 100000$).

4.1. Discussion

As all this study of the statistical power is based on asymptotic results, it is important to check the validity of the asymptotic assumption. A numerical evaluation has been performed for $\lambda = 2$ and a QTL located at $t^\star = 25$cM (cf. Table 6). The Theoretical Power, based on results of the Theorem, has been calculated using a DMC method. The Empirical Power (EP) has been computed assuming π_i equal to $1/I$ and a 95% confidence interval for the true value of the power is given into brackets. According to Table 6, the Theoretical Power is always located in the confidence interval whatever the value of n, demonstrating on this example, that the Theoretical Power should also be suitable for moderate values of n.

It can be seen that for all the figures shown here, the method is more powerful when the QTL is located on a marker. This is not surprising since on markers, the distribution from which belongs the quantitative trait Y is exactly known whereas between markers, since this distribution is unknown, a mixture model is used.

According to the Theorem, the LRT process, $\Lambda_n(.)$, is asymptotically the sum of the square of independent interpolated processes. So, it is easy to test every positions between genetic markers. In this article, as far as the sparse map is concerned, tests have been performed only every 5cM. In Figure 8, the focus is on an interval of two genetic markers spaced from 20cM ($I = 3$, $\lambda = 2$). We compare the power of the Interval Mapping when tests are performed every cM and every 5cM, as function of the location of the QTL, $t^\star$ ($I = 3$, $\lambda = 2$). It is noticeable that the two approaches give almost the same power : testing only every 5cM is convenient enough.

	$I = 1$	$I = 3$	$I = 5$
Theoretical Power	37.59%	68.57%	85.58%
EP for $n = 200$ I	38.08% [37.13%; 39.03%]	68.80% [67.89%; 69.71%]	85.00% [84.30%; 85.70%]
EP for $n = 50$ I	37.54% [36.59%; 38.49%]	68.37% [67.46%; 69.28%]	84.74% [84.04%; 85.44%]
EP for $n = 30$ I	37.83% [36.88%; 38.78%]	68.57% [67.66%; 69.48%]	85.15% [84.45%; 85.85%]

Table 6: Theoretical Power and Empirical Power (EP) as a function of I ($\lambda = 2$, $t^\star = 25$cM, $nspaths = 100000$ for the Theoretical Power and $npop = 10000$ for the Empirical Power, $\mu_1 = -0.37$, $\mu_2 = 0.03$, $\mu_3 = 0.06$, $\mu_4 = -0.26$, $\mu_5 = 0.27$, $\sigma = 1$)

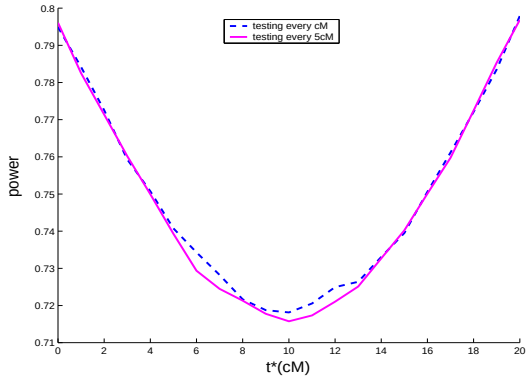

Figure 8: Power as a function of $t^\star$ and the way of testing ($\lambda = 2$, $\sigma = 1$, $nspaths = 100000$, $I = 3$). The map consists of 2 markers spaced from 20cM.

Conclusion

In order to optimize the QTL detecting process, we can advise :

1. to target, whenever possible, families with the biggest QTL effects and then, to analyze all these families together.
2. when it is not possible to target families, to analyze all the families together directly.

5. Acknowledgements

This work has been supported by the Animal Genetic Department of the French National Institute for Agricultural Research (INRA), SABRE, and the National Center for Scientific Research (CNRS).

References

Azaïs, J. M. and Cierco-Ayrolles, C., (2002) An asymptotic test for quantitative gene detection. *Ann. I. H. Poincaré*, **38**, **6**, 1087-1092.

Azaïs, J. M. and Wschebor, M., (2009) *Level sets and extrema of random processes and fields*. Wiley, New-York.

Cierco, C., (1998) Asymptotic distribution of the maximum likelihood ratio test for gene detection. *Statistics*, **31**, 261-285.

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

Delong, D. M., (1981) Crossing probabilities for a square root boundary by a Bessel process. *Commun. Statist.-Theor. Meth.*, **A10(21)**, 2197-2213.

Estrella, A., (2003) Critical values and p values of bessel process distributions : computation and application to structural break tests. *Econometric Theory*, **19(6)**, 1128-1143.

Genz, A., (1992) Numerical computation of multivariate normal probabilities. *J. Comp. Graph. Stat.*, 141-149.

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

Lander, E.S., Botstein, D., (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **138**, 235-240.

Plackett, R.I. (1954) A reduction formula for normal multivariate integrals. *Biometrika*, **41**, 351-360.

Rabier, C-E., Azaïs, J-M., Delmas, C. (2009) Likelihood Ratio Test for Quantitative Trait Loci detection. *hal-00421215*.

Rebaï, A., Goffinet, B., Mangin, B. (1994) Approximate thresholds of interval mapping tests for QTL detection. *Genetics*, **138**, 235-240.

Rebaï, A., Goffinet, B., Mangin, B. (1995) Comparing power of different methods for QTL detection. *Biometrics*, **51**, 87-99.

Van der Vaart, A.W. (1998) *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics.

Wu, R., MA, C.X., Casella, G. (2007) *Statistical Genetics of Quantitative Traits*, Springer