On quantitative trait locus mapping with an interference phenomenon

Charles-Elie Rabier

Test manuscript No.
(will be inserted by the editor)

On Quantitative Trait Locus mapping with an interference phenomenon

Charles-Elie Rabier

Received: date / Accepted: date

Abstract We consider the likelihood ratio test (LRT) process related to the test of the absence of QTL (a QTL denotes a gene with quantitative effect on a trait) on the interval $[0, T]$ representing a chromosome. The observation is the trait and the composition of the genome at some locations called "markers." We focus on the interference phenomenon, i.e. a recombination event inhibits the formation of another recombination event nearby. We give the asymptotic distribution of the LRT process under the null hypothesis that there is no QTL on $[0, T]$ and under local alternatives with a QTL at $t^\star$ on $[0, T]$. We show that the LRT process is asymptotically the square of a "linear interpolated and normalized process." We prove that under the null hypothesis, the distribution of the maximum of the LRT process is the same for a model with or without interference. However, the powers of detection are totally different between the two models.

Keywords Quantitative Trait Locus detection · Likelihood Ratio Test · Gaussian process · Chi-Square process.

PACS 62F03 · 62F05 · 62F12 · 62P10

1 Introduction

We study a backcross population: $A\times(A\times B)$, where A and B are purely homozygous lines. We address 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. The trait is observed on n individuals (progenies) and we denote by $Y_j,\ j = 1, ..., n$, the observations, which we will assume to be Gaussian, independent and identically distributed (i.i.d.). The mechanism of genetics, or more precisely of meiosis, implies that among the two chromosomes of each individual, one is purely inherited from A. The other (the "recombined"

Charles-Elie Rabier
Université de Toulouse, Institut de Mathématiques de Toulouse, U.P.S., Toulouse, France
INRA UR631, Station d'Amélioration Génétique des Animaux, Auzeville, France
University of Wisconsin-Madison, Statistic Department, Medical Science Center, 1300 University Avenue, Madison, WI 53706-1532, USA
Tel.: +(608)-265-9876, Fax: +(608)-262-0032
E-mail: rabier@stat.wisc.edu

one), due to crossing-overs, consists of parts originated from A and parts originated from B.

The chromosome will be represented by the segment $[0,T]$. The distance on $[0,T]$ is called the genetic distance and is measured in Morgans (see for instance Wu et al. [2007] or Siegmund and Yakir [2007]). K genetic markers are located at fixed locations $t_1 = 0 < t_2 < ... < t_K = T$. These markers will help us to find the QTL. $X(t_k)$ refers to the genetic information at marker k. For one individual, $X(t_k)$ takes the value $+1$ if, for example, the "recombined chromosome" is originated from A at location t_k and takes the value -1 if it is originated from B. We use the Haldane [1919] modeling for the genetic information at marker locations. It can be represented as follows: $X(0)$ is a random sign and $X(t_k) = X(0)(-1)^{N(t_k)}$ where $N(.)$ is a standard Poisson process on $[0,T]$. Due to the independence of increments of Poisson process, this model allows double recombinations between markers. For instance, if we consider 3 markers (i.e. $K = 3$), we can have the scenario $X(t_1) = 1$, $X(t_2) = -1$ and $X(t_3) = 1$, which means that there has been a recombination between markers 1 and 2, and also a recombination between markers 2 and 3. Obviously, in the same way, we can have the scenario $X(t_1) = -1$, $X(t_2) = 1$ and $X(t_3) = -1$.

A QTL is lying at an unknown position $t^\star$ between two genetic markers. $U(t^\star)$ is the genetic information at the QTL location. In the same way as for the genetic information at marker locations, $U(t^\star)$ takes value $+1$ if the "recombined chromosome" is originated from A at $t^\star$, and -1 if it is originated from B. The originality of this paper is in the fact that inside the marker interval which contains the QTL, we do not consider the classical Haldane model (contrary to Chang et al. [2009] and Azaïs et al. [2012]), but we focus on the model introduced by Rebaï et al. [1995] (see in particular their Section 2) in which double recombination between the QTL and its flanking markers is not allowed. As a consequence, under the model considered by Rebaï et al. [1995], if the QTL is lying for instance between the first two markers (i.e. $t^\star \in]t_1, t_2[$), we can not have the scenario $X(t_1) = 1$, $U(t^\star) = -1$ and $X(t_2) = 1$. Indeed, this would have supposed that there had been a recombination between the first marker and the QTL, and also a recombination between the second marker and the QTL. In particular, the model considers that if we have a recombination between the QTL and one of its flanking marker, we could not have a recombination between the QTL and the other flanking marker. In other words, if $X(t_1) = 1$ and $U(t^\star) = -1$, then we have automatically $X(t_2) = -1$. In the same way, if $X(t_2) = 1$ and $U(t^\star) = -1$, then we have automatically $X(t_1) = -1$. Using a particular choice for the recombination probabilities between the QTL and the markers, we shall prove that the law of $U(t^\star)$ given its flanking markers (still assuming that they are located at t_1 and t_2) is the following (see Section 2 for the details)

$$\mathbb{P}\left\{U(t^\star) = 1 \middle| X(t_1), X(t_2)\right\} = \begin{cases} 1 & \text{if } X(t_1) = 1 \text{ and } X(t_2) = 1 \\ \frac{t_2 - t^\star}{t_2 - t_1} & \text{if } X(t_1) = 1 \text{ and } X(t_2) = -1 \\ \frac{t^\star - t_1}{t_2 - t_1} & \text{if } X(t_1) = -1 \text{ and } X(t_2) = 1 \\ 0 & \text{if } X(t_1) = -1 \text{ and } X(t_2) = -1 \ . \end{cases} \tag{1}$$

Note that when the distance between $t^\star$ and t_1 (resp. t_2) increases, it is more likely to have one recombination between the QTL and the first (resp. second) marker.

This way, inside the marker interval which contains the QTL, we model the interference phenomenon: a recombination event inhibits the formation of another recombination event nearby (see for instance McPeek and Speed [1995]). This phenomenon was noticed a long time ago by geneticists working on the Drosophila (Sturtevant [1915], Muller [1916]). A key

experiment in this field is the one of Sturtevant which was recalled recently in the tutorial paper of Lobo and Shaw [2008]. Sturtevant investigated the frequencies of recombinations between the genes respectively responsible for body color (B), eye color (CO) and rudimentary wings (R), and placed in this order in the genome. Focusing only on the haplotypes for which a crossover between B and CO was observed, a recombination event between CO and R was obtained 9 times whereas no recombination between CO and R was observed 60 times. Due to the distance between CO and R on his original map, Sturtevant expected to observe a higher rate of haplotypes with a recombination between these two genes. As a result, this experiment reveals the presence of an interference phenomenon. Recent studies address also this point such as Martini et al. [2006] who focused on interference in yeast. They found on tetrad data, that when they could not detect a crossover in the interval lys5-met13, the adjacent met13-cyh2 interval was more likely to contain a recombination event. In the same way, when there was a recombination event in the interval lys5-met13, the adjacent interval met13-cyh2 was less likely to contain a crossover (cf. p.290 of Martini et al. [2006]).

Having explained the relevance of the interference model inside the marker interval, we will show that the overall model which is based on

- Haldane modeling for the genetic information at markers location
- the interference model inside the marker intervals,

is biologically motivated. Indeed, such an overall model is supported by experimental results reported in the following experimental results. Hillers and Villeneuve [2003] investigated interference in the nematode Caenorhabitis elegans by fusing multiple chromosomes together. They found that in most of meiosis, the chromosome meT7 which resulted of the fusion of 3 chromosomes, enjoyed only one crossover whereas three crossovers were expected according to the physical distance. They did not find double crossovers in adjacent interval (cf. intervals 1 and 2 in Figure 2C of Hillers and Villeneuve [2003], 0 crossover among 251 meiotic products assayed), but found double crossovers between non adjacent intervals (cf. intervals 1 and 3 in Figure 2C, 14 double crossovers among 251 meiotic products assayed). As a result, they claimed that "the nonindependent behavior of the different intervals implies that when two exchanges occur, they are still governed by an interference mechanism that acts along the length of meT7 to discourage nearby double exchanges, resulting in a wide spacing between crossing overs." In their study, Martini et al. [2006] also mention that crossovers tend to be widely spaced. In Figure 3A of Youds and Boulton [2011], it is noticeable that crossover sites "tend to be spaced for appart" on mouse chromosomes.

On the other hand, the same kind of arguments has been used by the statistician community. In example 11.3 p.248 of the statistical genetics textbook of Wu et al. [2007], the authors revisited Huang et al. [1997]'s experiment based on 12 rice chromosomes. Note that this population is equivalent to a backcross population. Wu et al. [2007] limited their study to chromosome 1. They considered 18 markers (i.e. 17 marker intervals), and performed their QTL analysis using a model which ignores double recombination inside each marker interval (model described in their Section 11.2.1. p.238). Then, each marker interval is analyzed independently to other marker intervals, that is to say the authors ignore double recombination inside a marker interval and another model is used for the genetic information on markers (Haldane for instance, cf. Figure 5.2 p.251).

In the present study, the overall model considered is the same as the one studied in Rebaï et al. [1994], where the authors extend their previous model (Rebaï et al. [1995]) to a whole chromosome using Haldane modeling on markers (see their Table 1). Note that in order to model the interference phenomenon, we could have focused on other models present in the

literature (e.g. Karlin and Liberman [1979], Stam [1979], King and Mortimer [1990], Foss et al. [1993]). However, we will see that the model considered in Rebaï et al. [1994] leads to interesting mathematical results.

We assume an "analysis of variance model" for the quantitative trait:

$$Y = \mu \ + \ U(t^\star) \ q \ + \ \sigma\varepsilon \tag{2}$$

where ε is a standard Gaussian random variable.

Indeed, it is well known that there is a finite number of loci underlying the variation in quantitative traits (e.g. in aquaculture and livestock, see Hayes [2007]). In this study, we will focus only on one locus (so-called QTL) and on only one quantitative trait. We will study the concept of QTL mapping: we will look for associations between allele variation at the QTL and variation in the quantitative trait of interest.

Since only the quantitative trait and the genetic information at marker locations are available, one observation will be

$$(Y, \ X(t_1), \ ..., \ X(t_K)) \, .$$

Conditionally on $X(t_1), \ldots, X(t_K)$, Y obeys to a mixture model with known weights:

$$p(t^*) f_{(\mu+q,\sigma)}(.) + \left\{1 - p(t^*)\right\} f_{(\mu-q,\sigma)}(.), \tag{3}$$

where $f_{(m,\sigma)}$ is the Gaussian density with parameters (m, σ) and where the function $p(t^\star)$ is the conditional probability that $U(t^\star) = 1$ conditionally on the flanking markers (cf. formula (1) if the flanking markers are located at t_1 and t_2).

We consider that we have n observations $(Y_j, \ X_j(t_1), \ ..., \ X_j(t_K))$, $j = 1, ..., n$ which are i.i.d., with the same distribution as described previously, and we want to test the presence of a QTL. Since its true location is unknown, we have to consider the location $t^\star$ as an unknown parameter t, and the likelihood process will also depend on the parameter t. The absence of a QTL is given by the null hypothesis H_0:"q=0," and the likelihood ratio test (LRT) of H_0 against its general alternative, has test statistic $\sup_t \Lambda_n(t)$, where $\Lambda_n(t)$ is the LRT statistic at location t. This paper gives the exact asymptotic distribution of this LRT statistic under the null hypothesis and under contiguous alternatives. These distributions have been given using some approximations under the null hypothesis, by Rebaï et al. [1995] and Rebaï et al. [1994]. In Cierco [1998], Azaïs and Cierco-Ayrolles [2002], Azaïs and Wschebor [2009], Chang et al. [2009] and Azaïs et al. [2012], the authors focus on other recombination models which do not model the interference phenomenon.

The main result of the paper (Theorems 1 and 2) is that the distribution of the LRT statistic is asymptotically that of the maximum of the square of a "linear normalized interpolated process." It is a generalization of the results obtained by Rebaï et al. [1995], Rebaï et al. [1994], where the authors focused only on the null hypothesis and characterized the process only by its covariance function. The computation of such a maximum is easy due to the interpolation. Note that recently, for a model without interference, Azaïs et al. [2012] have proved that the LRT statistic is asymptotically that of the maximum of the square of a "non linear normalized interpolated process." The second important result (Lemma 1) is that, under the null hypothesis, the maximum of the square of the "linear normalized interpolated process" is the same as the maximum of the square of the "non linear normalized interpolated process" obtained by Azaïs et al. [2012]. As a consequence, the Monte-Carlo Quasi Monte-Carlo method proposed by Azaïs et al. [2012] to compute thresholds is also suitable for our interference model. So, for our interference model, we have now a method to compute thresholds suitable whatever the genetic map is. This is not the case of the method

proposed in Rebaï et al. [1994] based on Davies [1977]. Using simulated data, we will see that, as expected, our method outperforms Rebaï's method in terms of false positives. Finally, we will compare the theoretical power of QTL detection, for a model without interference (Azaïs et al. [2012]) and a model with interference (this paper). We will show that it is largely more powerful to detect a QTL under interference than without interference. To sum up, we prove that we have exactly the same threshold with or without interference, but we have a totally different power.

We refer to the book of Van der Vaart [1998] for elements of asymptotic statistics used in proofs.

2 Main results: two genetic markers

To begin with, we suppose that there are only two markers ($K = 2$) located at 0 and T: $0 = t_1 < t_2 = T$. Furthermore, a QTL is lying between these two markers at $t^\star \in]t_1, t_2[$. Note that for the sake of clarity, we consider that the QTL is not located on markers. However, the main result of this section (Theorem 1) can be prolonged by continuity at marker locations.

Let $r(t_1, t_2)$ be the probability that there is a recombination between the two markers. Calculations on the Poisson distribution show that:

$$r(t_1, t_2) = \mathbb{P}(X(t_1)X(t_2) = -1) = \frac{1}{2}\,(1 - e^{-2|t_1 - t_2|}).$$

We set in addition $\bar{r}(t_1, t_2) = 1 - r(t_1, t_2)$. We will call $r_{t_1}(t^\star)$ (resp. $r_{t_2}(t^\star)$) the probability of recombination between the first (resp. second) marker and the QTL. So,

$$r_{t_1}(t^\star) = \mathbb{P}(X(t_1)U(t^\star) = -1)\ ,\ r_{t_2}(t^\star) = \mathbb{P}(X(t_2)U(t^\star) = -1).$$

As explained in Section 1, only one recombination is allowed between the QTL and the two markers. We have:

$$\{X(t_1)X(t_2) = -1\} \Leftrightarrow \left\{X(t_1)U(t^\star) = -1\right\} \cup \left\{X(t_2)U(t^\star) = -1\right\}.$$

Indeed, $X(t_1)U(t^\star) = -1$ means that there has been a recombination between the first marker and the QTL, so the second marker is not allowed to recombine with the QTL. As a consequence, $X(t_2) = U(t^\star)$ and we have $X(t_1)X(t_2) = -1$. Same remark for $X(t_2)U(t^\star) = -1$ but this time, it is the first marker which is not allowed to recombine with the QTL.

Note that since $\{X(t_1)U(t^\star) = -1\} \cap \{X(t_2)U(t^\star) = -1\} = \oslash$, we have

$$r(t_1, t_2) = r_{t_1}(t^\star) + r_{t_2}(t^\star). \tag{4}$$

In the same way as in Rebaï et al. [1995], we choose:

$$r_{t_1}(t^\star) = \frac{t^\star - t_1}{t_2 - t_1}\, r(t_1, t_2)\ ,\ r_{t_2}(t^\star) = \frac{t_2 - t^\star}{t_2 - t_1}\, r(t_1, t_2).$$

Then, the probability of recombination of the marker and the QTL is proportional to the probability of recombination of the two markers, and also proportional to the distance between between the QTL and the marker. Note that formula (4) stands with these expressions of $r_{t_1}(t^\star)$ and $r_{t_2}(t^\star)$.

Let's define now

$$p(t^\star) = \mathbb{P}\left\{U(t^\star) = 1 \middle| X(t_1), X(t_2)\right\}.$$

Obviously, since $U(t^\star)$ takes value $+1$ or -1, we have

$$1 - p(t^\star) = \mathbb{P}\left\{U(t^\star) = -1 \middle| X(t_1), X(t_2)\right\}.$$

Since only one recombination is allowed between the QTL and its flanking markers, we have

$$\mathbb{P}\left\{U(t^\star) = 1 \middle| X(t_1) = 1, X(t_2) = 1\right\} = 1\ , \quad \mathbb{P}\left\{U(t^\star) = 1 \middle| X(t_1) = -1, X(t_2) = -1\right\} = 0.$$

Besides, according to Bayes rules

$$\begin{aligned}
&\mathbb{P}\left\{U(t^\star) = 1 \middle| X(t_1) = 1, X(t_2) = -1\right\} \\
&= \frac{\mathbb{P}\left\{X(t_1) = 1 \middle| U(t^\star) = 1, X(t_2) = -1\right\} \mathbb{P}\left\{U(t^\star) = 1, X(t_2) = -1\right\}}{\mathbb{P}\left\{X(t_1) = 1, X(t_2) = -1\right\}} \\
&= \frac{r_{t_2}(t^\star)/2}{r(t_1,t_2)/2} = \frac{r_{t_2}(t^\star)}{r(t_1,t_2)} = \frac{t_2 - t^\star}{t_2 - t_1}.
\end{aligned}$$

In the same way,

$$\mathbb{P}\left\{U(t^\star) = 1 \middle| X(t_1) = -1, X(t_2) = 1\right\} = \frac{r_{t_1}(t^\star)}{r(t_1,t_2)} = \frac{t^\star - t_1}{t_2 - t_1}.$$

As a consequence,

$$p(t^\star) = 1_{X(t_1)=1} 1_{X(t_2)=1} \ + \ \frac{t_2 - t^\star}{t_2 - t_1}\ 1_{X(t_1)=1} 1_{X(t_2)=-1} \ + \ \frac{t^\star - t_1}{t_2 - t_1}\ 1_{X(t_1)=-1} 1_{X(t_2)=1}\ , \tag{5}$$

and there is agreement with formula (1) of Section 1. As explained in Section 1, conditionally on $X(t_1)$ and $X(t_2)$, Y obeys to the mixture model of formula (3). Note that, using the formula above for $p(t^\star)$, and using properties of conditional expectation, it is easy to check that $\mathbb{P}\{U(t^\star) = 1\} = 1/2$, so that $U(t^\star)$ takes values $+1$ and -1 with equal probability. As explained previously, since the location $t^\star$ of the QTL is unknown, we will have to perform tests at each position t between the two genetic markers. We will consider only positions t distinct of the marker locations and the result can be prolonged by continuity on markers.

Let $\theta = (q,\ \mu,\ \sigma)$ be the parameter of the model at t fixed. The likelihood of the triplet $(Y,\ X(t_1),\ X(t_2))$ with respect to the measure $\lambda \otimes N \otimes N$, λ being the Lebesgue measure, N the counting measure on $\mathbb{N}$, is:

$$L_t(\theta) = \left[p(t) f_{(\mu+q,\sigma)}(Y) + \{1 - p(t)\}\ f_{(\mu-q,\sigma)}(Y)\right]\ g(t) \tag{6}$$

where the function

$$\begin{aligned}
g(t) = &\frac{1}{2}\left\{\bar{r}(t_1,t_2)\ 1_{X(t_1)=1} 1_{X(t_2)=1} \ + \ r(t_1,t_2)\ 1_{X(t_1)=1} 1_{X(t_2)=-1}\right\} \\
&+ \frac{1}{2}\left\{r(t_1,t_2)\ 1_{X(t_1)=-1} 1_{X(t_2)=1} \ + \ \bar{r}(t_1,t_2)\ 1_{X(t_1)=-1} 1_{X(t_2)=-1}\right\}
\end{aligned}$$

can be removed because it does not depend on the parameters. By a small abuse of notation we still denote $L_t(\theta)$ for the likelihood without this function. Thus we set

$$L_t(\theta) = \left[p(t) f_{(\mu+q,\sigma)}(Y) + \{1 - p(t)\}\ f_{(\mu-q,\sigma)}(Y)\right]$$

and $l_t(\theta)$ will be the loglikelihood. Besides, we define the following quantity (with $p(t)$ given in formula (5)):

$$u(t) = 2p(t) - 1 \ .$$

We first compute the Fisher information at a point $\theta_0 = (0,\ \mu,\ \sigma)$ under which H_0 holds.

$$\frac{\partial l_t}{\partial q} \mid_{\theta_0} = \frac{Y-\mu}{\sigma^2}\ u(t) \quad , \tag{7}$$

$$\frac{\partial l_t}{\partial \mu} \mid_{\theta_0} = \frac{Y-\mu}{\sigma^2} \quad , \qquad \frac{\partial l_t}{\partial \sigma} \mid_{\theta_0} = -\frac{1}{\sigma} \ + \ \frac{(Y-\mu)^2}{\sigma^3} \quad .$$

After some calculations, we find

$$I_{\theta_0} = Diag\left[\frac{\mathbb{E}\left\{u^2(t)\right\}}{\sigma^2}\ ,\ \frac{1}{\sigma^2}\ ,\ \frac{2}{\sigma^2}\right] \ . \tag{8}$$

Before introducing our main theorem, let us define the score statistic and the LRT statistic at t. Since the Fisher Information matrix is diagonal, the score statistic of the hypothesis "$q = 0$" at t, for n independent observations, will be defined as

$$S_n(t) = \frac{\frac{\partial l_t^n}{\partial q} \mid_{\theta_0}}{\sqrt{\text{Var}\left(\frac{\partial l_t^n}{\partial q} \mid_{\theta_0}\right)}} \ ,$$

where $l_t^n(\theta)$ denotes the log likelihood at t, associated to n observations.

The LRT at t, for n independent observations, will be defined as

$$\Lambda_n(t) = 2\left\{l_t^n(\widehat{\theta}) - l_t^n(\widehat{\theta}_{|H_0})\right\} \quad ,$$

where $\widehat{\theta}$ is the maximum likelihood estimator (MLE), and $\widehat{\theta}_{|H_0}$ the MLE under H_0.

Our main result is the following:

Theorem 1 *Suppose that the parameters (q, μ, σ^2) vary in a compact and that σ^2 is bounded away from zero. Let H_0 be the null hypothesis $q = 0$ and define the following local alternative*

$H_{at^\star}$: *"the QTL is located at the position $t^\star$ with effect $q = a/\sqrt{n}$ where $a \neq 0$ ".*

With the previous defined notations,

$$S_n(.) \Rightarrow W(.) \quad , \quad \Lambda_n(.) \xrightarrow{F.d.} W^2(.) \quad , \quad \sup \Lambda_n(.) \Rightarrow \sup W^2(.)$$

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

- *$\Rightarrow$ is the weak convergence, $\xrightarrow{F.d.}$ is the convergence of finite-dimensional distributions*

- *$W(.)$ is the Gaussian process with unit variance such as:*

$$W(t) = \frac{\alpha(t)W(t_1) + \beta(t)W(t_2)}{\sqrt{Var\left\{\alpha(t)W(t_1) + \beta(t)W(t_2)\right\}}}$$

where

$$Cov\left\{W(t_1), W(t_2)\right\} = \rho(t_1, t_2) = \exp(-2|t_1 - t_2|) \ ,$$
$$\alpha(t) = \frac{t_2 - t}{t_2 - t_1} \quad , \quad \beta(t) = \frac{t - t_1}{t_2 - t_1}$$

and with expectation:

- *under H_0, $m(t) = 0$*
- *under $H_{at^\star}$*

$$m_{t^\star}(t) = \frac{\alpha(t)\ m_{t^\star}(t_1) + \beta(t)\ m_{t^\star}(t_2)}{\sqrt{Var\left\{\alpha(t)W(t_1) + \beta(t)W(t_2)\right\}}}$$

where

$$m_{t^\star}(t_1) = \frac{a}{\sigma}\left\{\alpha(t^\star) + \beta(t^\star)\rho(t_1, t_2)\right\} \ , \ m_{t^\star}(t_2) = \frac{a}{\sigma}\left\{\alpha(t^\star)\rho(t_1, t_2) + \beta(t^\star)\right\} \ .$$

Note that the functions $\alpha(t)$ and $\beta(t)$ are different from the ones present in Theorem 2.1 of Azaïs et al. [2012]. Our interpolation is linear whereas the interpolation described in Azaïs et al. [2012] is non linear. Therefore, our process $W(.)$ will be called a "linear normalized interpolated process." The proof of Theorem 1 is given in Appendix A.1.

In Azaïs et al. [2012], the authors present a lemma called Lemma 2.2, which is very useful to compute the supremum of the square of an interpolated process. Let us recall this lemma and the comments following this lemma.

Lemma 2.2 (Azaïs et al. [2012]) *Let $\gamma_1(t)$ and $\gamma_2(t)$ be two functions such that $\frac{\gamma_i(t)}{\gamma_1(t)+\gamma_2(t)}$ takes every value in $[0, 1]$, $i = 1, 2$. Let C_1 and C_2 be two real numbers and $0 < \tilde{\rho} < 1$ then*

$$\max_{t\in[t_1,t_2]} \frac{\left\{\gamma_1(t)C_1 + \gamma_2(t)C_2\right\}^2}{\gamma_1^2(t) + \gamma_2^2(t) + 2\tilde{\rho}\gamma_1(t)\gamma_2(t)} = \max\left(C_1^2 \ , \ C_2^2 \ , \ \frac{C_1^2 + C_2^2 - 2\tilde{\rho}C_1C_2}{1 - \tilde{\rho}^2} 1_{\frac{C_2}{C_1} \ \in \ ]\ \tilde{\rho}\ ,\ \frac{1}{\tilde{\rho}}\ [}\right).$$

In particular, if C_1 and C_2 are two random variables defined on the same probability space with $\text{Var}(C_i) = 1$, $i = 1, 2$, $\text{Cov}(C_1, C_2) = \tilde{\rho}$ with $0 < \tilde{\rho} < 1$ and if $\gamma_1(t)$ and $\gamma_2(t)$ are two functions as above, the lemma gives the distribution of the maximum on $[t_1, t_2]$ of the square of the following normalized interpolated process $D(.)$:

$$\forall t \in [t_1, t_2], \quad D(t) = \frac{\gamma_1(t)C_1 + \gamma_2(t)C_2}{\sqrt{\gamma_1^2(t) + \gamma_2^2(t) + 2\tilde{\rho}\gamma_1(t)\gamma_2(t)}}.$$

So, the lemma can be applied to our process $W(.)$ by taking $\gamma_1(t) = \alpha(t)$, $\gamma_2(t) = \beta(t)$, $\tilde{\rho} = \rho(t_1, t_2)$, $C_1 = W(t_1)$, $C_2 = W(t_2)$, since $\frac{\beta(t)}{\alpha(t)+\beta(t)}$ takes every value in $[0, 1]$. As a consequence, the computation of the maximum of our process $W^2(.)$ can be obtained easily using Lemma 2.2 of Azaïs et al. [2012].

On the other hand, we have this interesting result:

Lemma 1 *With the previous defined notations, under H_0,*

$$\max_{t\in[t_1,t_2]} W^2(t) = \max_{t\in[t_1,t_2]} Z^2(t) \ ,$$

where $Z(.)$ is the "non linear normalized interpolated process" obtained by Azaïs et al. [2012].

In other words, under the null hypothesis, according to Lemma 1, the maximum of the square of the "non linear normalized interpolated process" is the same as the maximum of the square of the "linear normalized interpolated process." Note that Lemma 1 stands only under the null hypothesis and not under the alternative.

Proof of Lemma 1: In order to prove this lemma, we just have to notice that under H_0 at marker locations, we have $Z(t_1) = W(t_1)$ and $Z(t_2) = W(t_2)$. Indeed, under H_0, the processes overlap at marker locations since there is no QTL affecting the processes and also because the recombination model (i.e. Haldane) is the same at marker locations. Then, using Lemma 2.2 of Azaïs et al. [2012], since only the process at points t_1 and t_2 is involved in the supremum, the computation of the maximum of $Z^2(.)$ and $W^2(.)$ is the same. Note that it can be proved that the arg max of $Z^2(.)$ and the arg max of $W^2(.)$ are different under some conditions. □

3 Several markers: the "genome scan"

We suppose now that there are K markers $0 = t_1 < t_2 < ... < t_K = T$. A QTL is lying at a position $t^\star$. So, in order to find the QTL, we will perform tests at every positions t on the chromosome. Note that we use the terminology "genome scan" instead of "interval mapping," since the "interval mapping" of Lander and Botstein [1989] is usually computed by geneticists with a model without interference (Haldane [1919]). So, in our case, since we consider an interference model, it will only be a "genome scan." We consider values t or $t^\star$ of the parameters that are distinct of the markers positions, and the result will be prolonged by continuity at the markers positions. For $t \in [t_1, t_K]\backslash\mathbb{T}_K$ where $\mathbb{T}_K = \{t_1, ..., t_K\}$, we define t^ℓ and t^r as:

$$t^\ell = \sup\{t_k \in \mathbb{T}_K : t_k < t\} \quad , \quad t^r = \inf\{t_k \in \mathbb{T}_K : t < t_k\}\,.$$

In other words, t belongs to the "Marker interval" (t^ℓ, t^r).

As explained in Section 1, in order to infer the value of $U(t^\star)$, we just need to keep the flanking markers. This means that the information brought by the other markers is useless. So, we have

$$\mathbb{P}\left\{U(t^\star) = 1 \middle| X(t_1), ..., X(t_K)\right\} = \mathbb{P}\left\{U(t^\star) = 1 \middle| X(t^{\star\ell}), X(t^{\star r})\right\} \ .$$

As a consequence, our problem becomes the same as the one with two genetic markers (see Section 2). In order to perform our tests at every positions t, we simply have to consider all the different marker intervals.

Theorem 2 *We have the same results as in Theorem 1 except that the following functions must be redefined:*

- t_1 *becomes* t^ℓ *and* t_2 *becomes* t^r *in all the expressions, except in the expressions* $\alpha(t^\star)$ *and* $\beta(t^\star)$*, where* t_1 *becomes* $t^{\star\ell}$ *and* t_2 *becomes* $t^{\star r}$
- $m_{t^\star}(t^\ell) = a\ \rho(t^\ell, t^{\star\ell}) \left\{\alpha(t^\star) + \beta(t^\star)\rho(t^{\star\ell}, t^{\star r})\right\}/\sigma$ *if* $t^\star > t^\ell$
- $m_{t^\star}(t^\ell) = a\ \rho(t^\ell, t^{\star r}) \left\{\alpha(t^\star)\rho(t^{\star r}, t^{\star\ell}) + \beta(t^\star)\right\}/\sigma$ *if* $t^\star < t^\ell$
- $m_{t^\star}(t^r) = a\ \rho(t^r, t^{\star\ell}) \left\{\alpha(t^\star) + \beta(t^\star)\rho(t^{\star\ell}, t^{\star r})\right\}/\sigma$ *if* $t^\star > t^r$
- $m_{t^\star}(t^r) = a\ \rho(t^r, t^{\star r}) \left\{\alpha(t^\star)\rho(t^{\star r}, t^{\star\ell}) + \beta(t^\star)\right\}/\sigma$ *if* $t^\star < t^r$.

The proof is given in Appendix A.2.

4 Application

In this Section, we present some applications of our theoretical study. We first focus on the null hypothesis and then we will move on to the alternative hypothesis.

4.1 Application to the computation of thresholds

In QTL detection, in order to conclude on the presence of a QTL or not, it is always important to use an appropriate threshold for the statistical test. The aim is to show from our theoretical study that we are now able to propose a threshold which gives better performances than the classical threshold proposed by Rebaï et al. [1995] and Rebaï et al. [1994] for the interference model.

Recall that $W(.)$ is our "linear normalized interpolated process" whereas $Z(.)$ is the "non linear normalized interpolated process" of Azaïs et al. [2012]. According to Lemma 1, when we consider only two genetic markers, the maximum of $W^2(.)$ is the same as the maximum of $Z^2(.)$ under the null hypothesis. Since when dealing with several markers, we just have to consider the different marker intervals, it is easy to check that Lemma 1 still holds when several markers are present. This way, the threshold will be the same for a model with interference (this paper) and for a model without interference (Azaïs et al. [2012]). In order to compute the threshold, Azaïs et al. [2012] propose a Monte-Carlo Quasi Monte-Carlo (MCQMC) method, based on Genz [1992]. This method is very fast, and the advantage of MCQMC is that it is more accurate than a simple Monte-Carlo method. We refer to Azaïs et al. [2012] and Genz [1992] for more details.

Let's explain now the method to compute thresholds, proposed by Rebaï et al. [1995] and Rebaï et al. [1994]. In Rebaï et al. [1995], the authors consider only two markers. They propose to use results of Davies [1977] and Davies [1987]. Indeed, in Davies, we can find an upper bound for a threshold corresponding to the supremum of a stochastic process (Gaussian process or Chi square process) which depends on a nuisance parameter only present under the alternative. In QTL detection, the nuisance parameter is the position of the QTL. Note that in Rebaï et al. [1995], the authors use as a scale the recombination units whereas in this paper, we use the genetic distance. In other words, if we call $W'(.)$ the process studied in Rebaï et al. [1995] with only two markers, we have the relationship $\forall t \in [t_1, t_2]$:

$$W(t) = W'\left\{r(t_1, t_2)\frac{t - t_1}{t_2 - t_1}\right\} \quad .$$

In their paper, they show that

$$\frac{\partial^2 \text{Cov}\left\{W'(t), W'(t')\right\}}{\partial t'^2} \mid_{t'=t} = -\frac{4\, r(t_1,t_2)\left\{1-r(t_1,t_2)\right\}}{\left[r(t_1,t_2) - 4r^2(t_1,t_2)\frac{t-t_1}{t_2-t_1} + 4\left\{r(t_1,t_2)\frac{t-t_1}{t_2-t_1}\right\}^2\right]^2} \ .$$

Then, since

$$\int_0^{r(t_1,t_2)} \sqrt{-\frac{\partial^2 \text{Cov}\left\{W'(t), W'(t')\right\}}{\partial t'^2} \mid_{t'=t}} \ dt = 2\arctan\left(\sqrt{\frac{r(t_1,t_2)}{1-r(t_1,t_2)}}\right)$$

and using Davies formula, they find that

$$\mathbb{P}\left\{\sup_{[0,r(t_1,t_2)]} W'(t) > c\right\} \leqslant \Phi(-c) \ + \ \frac{e^{-c^2/2}}{\pi} \ \arctan\left(\sqrt{\frac{r(t_1,t_2)}{1-r(t_1,t_2)}}\right) \ ,$$

where Φ is the cumulative distributive function of a standardized normal distribution. Note that since

$$\mathbb{P}\left\{\sup_{[t_1,t_2]} W(t) > c\right\} = \mathbb{P}\left\{\sup_{[0,r(t_1,t_2)]} W'(t) > c\right\} \ ,$$

it gives also the threshold for our process $W(.)$. In Rebaï et al. [1994], the authors generalize their approach to several markers. Their formula adapated to our process $W(.)$ becomes:

$$\mathbb{P}\left\{\sup_{[t_1,t_K]} W(t) > c\right\} \leqslant \Phi(-c) \ + \ \frac{e^{-c^2/2}}{\pi} \ \sum_{k=1}^{K-1} \arctan\left(\sqrt{\frac{r(t_k,t_{k+1})}{1-r(t_k,t_{k+1})}}\right) \ . \quad (9)$$

In order to obtain the threshold, we just have to find for which value of c, the right-side of formula (9) is equal to $\alpha/2$, and we will obtain the threshold c^2 for the supremum of our process $W^2(.)$. Note that this threshold c^2 will only correspond to a level lower or equal than α, due to the upper bound of formula (9).

Table 1 compares numerically the two approaches to compute thresholds for the interference model: Azaïs et al. [2012] and Rebaï et al. [1994]. For the genetic map, we consider the same configurations as in Table 1 of Rebaï et al. [1994], that is to say a chromosome of length $T = 1$M, different numbers of markers, and a level α equal to 5%. According to Table 1, we can see that the two approaches give different thresholds. It was expected since Rebaï's threshold corresponds only to a level lower or equal to 5%. Besides, the more markers there are, the more different the thresholds are. It is due to the fact that the derivative of the process $W(.)$ has a jump at each markers location, and Davies [1977] formula is suitable when the derivative of the process has a finite number of jumps. In other words, the more markers there are, the less appropriate Rebaï's threshold will be.

To conclude, since the two approaches are based on asymptotic results, we propose to check the asymptotic validity on simulated data. We simulated under the null hypothesis and under the interference model, 10000 samples of $n = 200$ individuals. We analyzed data using Lemma 2.2 of Azaïs et al. [2012] (still suitable here, cf. our Section 2), that is to say performing LRT on markers and performing only one test in each marker interval if the ratio of the score statistics on markers fulfilled a given condition. According to Table 1, Azaïs' method always gives a percentage of false positives close to 5% , whereas Rebai's method is

too conservative. So, for our interference model, we have now a method to compute thresholds which is suitable whatever the genetic map is, and which does not require the number of indivuals n to be too large.

Note that in Azaïs et al. [2012], using samples generated under a model without interference, the authors already highlighted that Rebai's method was too conservative. However, since they studied a model without interference and Rebai's method is for an interference model, the authors could not conclude if the method was too conservative because the derivative of the process had too many jumps or because of the two different models (with and without interference). Here, from our analysis of an interference model, we can now conclude that Rebai's method is too conservative because the derivative of the process has too many jumps.

4.2 About the power

We focus now on the alternative hypothesis. In our paper, double recombination between the QTL and its flanking markers is not allowed. This way, we model the interference phenomenon. In Azaïs et al. [2012], since the authors don't model interference, double recombination between the QTL and its flanking markers is allowed. The main difference is that, for an interference model, the LRT process is asymptotically the square of a linear interpolated and normalized process (i.e. $W(.)$), whereas for a model without interference, the LRT process is asymptotically the square of a non linear interpolated and normalized process (i.e. $Z(.)$). Table 2 compares the asymptotic power of the two approaches, using these asymptotic processes. We consider $a = 4$ (i.e. the constant for the QTL effect) and 100000 paths of each process. The different genetic maps studied are detailed in Table 3. First, we consider some sparse maps: map 1, map 2 and map 3. For map 1, we consider a chromosome of length $T = 5$M. 11 markers are located on the chromosome and the distance between two consecutive markers is either 40cM or 60cM. We can see that when the QTL is located at $t^\star = 80$cM and $t^\star = 370$cM, there are huge differences of power between the model with interference and the model without interference. For instance, we have 73.40% chances of detecting a QTL located at 80cM with interference, whereas we only have 58.13% chances of detecting the same QTL without interference. This is due to the fact that the mean functions are totally differents between the two asymptotic processes. We obtain the same kind of conclusions for map 2 and map 3. Map 4 is a more dense map: a chromosome of length $T = 1$M and 6 markers equally spaced every 20cM. We can see that there is now only a slight difference of power. To conclude, in the same way as what has been done in the previous section, we propose to check the asymptotic validity of our asymptotic results. In Table 4, we consider map 3: a chromosome of length $T = 4$M and 9 markers equally spaced every 50cM. We simulated 10000 samples of $n = 50$, $n = 100$, $n = 200$, $n = 1000$ individuals, according to the interference model. We can see that for $n = 200$, we are already close to the asymptotic results. For $n = 1000$, there is a very good agreement between the empirical power and the theoretical power. This validates our asymptotic study.

5 Conclusion

In this paper, we have presented some theoretical results related to QTL mapping under an interference phenomenon. We studied, in particular, the asymptotic properties of the test of the absence of QTL on an interval $[0, T]$. First, we have shown that the distribution of the

LRT statistic is asymptotically that of the maximum of the square of a "linear normalized interpolated process." On the other hand, we have proved that under the null hypothesis (i.e. no QTL on $[0,T]$), the maximum of the square of the "linear normalized interpolated process" is the same as the maximum of the square of the "non linear normalized interpolated process" obtained by Azaïs et al. [2012] for a model without interference. As a result, in order to compute thresholds, the Monte-Carlo Quasi Monte-Carlo method proposed by Azaïs et al. [2012] is still suitable under interference. Another important result is the following: although both LRT statistics (i.e. with or without interference) have the same distribution under the null hypothesis, they do not present the same distribution under the alternative hypothesis of one QTL located on $[0,T]$. Finally, we illustrated our theoretical results using a simulation study. Under the null hypothesis, we have shown that the Monte-Carlo Quasi Monte-Carlo method of Azaïs et al. [2012] is more efficient than Rebaï et al. [1994]'s method based on Davies [1977, 1987]. Under the alternative hypothesis, we pointed out that there is more statistical power to detect a QTL under interference than without interference.

Acknowledgements I thank Jean-Marc Azaïs, Céline Delmas, Jean-Michel Elsen and Brigitte Mangin for fruitful discussions. This work has been supported by the Animal Genetic Department of the French National Institute for Agricultural Research, SABRE, and the National Center for Scientific Research.

A Appendix: Proofs of theoretical results

A.1 Proof of Theorem 1

As mentioned before, we consider values of t and $t^\star$, distinct of marker locations and the result can be prolonged by continuity on markers.

A.1.1 Study of the score process under the null hypothesis

The study is based on the key lemma:

Lemma 2

$$u(t) = \alpha(t)X(t_1) + \beta(t)X(t_2)$$

with $\alpha(t) = \frac{t_2-t}{t_2-t_1}$ *and* $\beta(t) = \frac{t-t_1}{t_2-t_1}$.

To prove this lemma, use formula (5) and check that both coincide whatever the value of $X(t_1)$, $X(t_2)$ is. Now using formula (7), we have

$$\frac{\partial l_t^n}{\partial q} \mid_{\theta_0} = \sum_{j=1}^{n} \frac{Y_j - \mu}{\sigma^2} u_j(t) = 1/\sigma \sum_{j=1}^{n} \varepsilon_j u_j(t) = \frac{\alpha(t)}{\sigma} \sum_{j=1}^{n} \varepsilon_j X_j(t_1) + \frac{\beta(t)}{\sigma} \sum_{j=1}^{n} \varepsilon_j X_j(t_2) \quad (10)$$

this proves the interpolation. On the other hand

$$S_n(t_k) = \sum_{j=1}^{n} \frac{\varepsilon_j X_j(t_k)}{\sqrt{n}} \quad k = 1,2$$

and a direct application of central limit theorem implies that these two variables have a limit distribution which is Gaussian centered distribution with variance

$$\begin{pmatrix} 1 & \exp(-2|t_2 - t_1|) \\ \exp(-2|t_2 - t_1|) & 1 \end{pmatrix}.$$

This proves the expression of the covariance. The weak convergence of the score process, $S_n(.)$, is then a direct consequence of (10), the convergence of $(S_n(t_1), S_n(t_2))$ and the Continuous Mapping Theorem.

A.1.2 Study of the score process under the local alternative

Under the alternative

$$S_n(t) = \frac{a}{n\sigma} \sum_{j=1}^{n} \frac{U_j(t^*)u_j(t)}{\sqrt{\text{Var}\{u(t)\}}} + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \varepsilon_j \frac{u_j(t)}{\sqrt{\text{Var}\{u(t)\}}} \ .$$

The second term has the same distribution as under the null hypothesis and the first one gives the expectation. We have

$$\mathbb{E}\{S_n(t)\} = \frac{a\ \mathbb{E}\{U(t^*)u(t)\}}{\sigma\ \sqrt{\text{Var}\{u(t)\}}}.$$

According to Lemma 2, we have:

$$\mathbb{E}\{U(t^*)u(t)\} = \alpha(t)\ \mathbb{E}\{X(t_1)U(t^*)\}\ +\ \beta(t)\ \mathbb{E}\{U(t^*)X(t_2)\}\,.$$

So, we need now to calculate $\mathbb{E}\{X(t_1)U(t^*)\}$ and $\mathbb{E}\{U(t^*)X(t_2)\}$. We have

$$\begin{aligned}
\mathbb{P}\{X(t_1)U(t^\star) = -1\} &= \mathbb{P}\{U(t^\star) = 1 \mid X(t_1) = -1, X(t_2) = 1\}\,\mathbb{P}\{X(t_1) = -1, X(t_2) = 1\} \\
&+ \mathbb{P}\{U(t^\star) = 1 \mid X(t_1) = -1, X(t_2) = -1\}\,\mathbb{P}\{X(t_1) = -1, X(t_2) = -1\} \\
&+ \mathbb{P}\{U(t^\star) = -1 \mid X(t_1) = 1, X(t_2) = 1\}\,\mathbb{P}\{X(t_1) = 1, X(t_2) = 1\} \\
&+ \mathbb{P}\{U(t^\star) = -1 \mid X(t_1) = 1, X(t_2) = -1\}\,\mathbb{P}\{X(t_1) = 1, X(t_2) = -1\} \\
&= \frac{\beta(t^\star)r(t_1,t_2)}{2} + 0 + 0 + \frac{\beta(t^\star)r(t_1,t_2)}{2} = \beta(t^\star)r(t_1,t_2)\ .
\end{aligned}$$

As a consequence,

$$\mathbb{P}\{X(t_1)U(t^\star) = 1\} = 1 - \beta(t^\star)r(t_1,t_2)\ .$$

As a result,

$$\mathbb{E}\{X(t_1)U(t^\star)\} = 1 - 2\beta(t^\star)r(t_1,t_2) = \alpha(t^\star) + \beta(t^\star)\rho(t_1,t_2) \text{ with } \rho(t_1,t_2) = e^{-2|t_1-t_2|}\ .$$

In the same way, we obtain

$$\mathbb{E}\{U(t^\star)X(t_2)\} = \alpha(t^\star)\rho(t_1,t_2) + \beta(t^\star)\ .$$

This gives the result.

A.1.3 About the LRT process

Since the model with t fixed is regular, it is easy to prove that for fixed t

$$\Lambda_n(t) = S_n^2(t) + o_P(1) \tag{11}$$

under the null hypothesis.

Let us consider a local alternative defined by t^* and $q = a/\sqrt{n}$. The model with t^* fixed is differentiable in quadratic mean, this implies that the alternative defines a contiguous sequence of alternatives. By Le Cam's first Lemma, relation (11) remains true under the alternative. This gives the result for the convergence of finite-dimensional distribution. Concerning the study of the supremum of the LRT process, the proof is exactly the same as in Azaïs et al. [2012] which is based on results of Azaïs et al. [2006], Azaïs et al. [2009] and Gassiat [2002]. □

A.2 Proof of Theorem 2

We recall that we consider values t or $t^\star$ of the parameters that are distinct of the markers positions, and the result will be prolonged by continuity at the markers positions.

The proof of the theorem is the same as the proof of Theorem 1 as soon as we can limit our attention to the interval (t^ℓ, t^r) when considering a unique instant t. So, under H_0, the result is straightforward. However, under the local alternative, the proof is more complicated than the proof of Theorem 1. Indeed, the location $t^\star$ of the QTL and the location t, can belong to a different marker interval.

According to the proof of Theorem 1, under the alternative

$$S_n(t) = \frac{a}{n\sigma} \sum_{j=1}^{n} \frac{U_j(t^*)u_j(t)}{\sqrt{\text{Var}\left\{u(t)\right\}}} + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \varepsilon_j \frac{u_j(t)}{\sqrt{\text{Var}\left\{u(t)\right\}}} \ .$$

As previously, the second term has the same distribution as under the null hypothesis and the first one gives the expectation. We have

$$\mathbb{E}\left\{S_n(t)\right\} = \frac{a\ \mathbb{E}\left\{U(t^*)u(t)\right\}}{\sigma\ \sqrt{\text{Var}\left\{u(t)\right\}}}.$$

We notice that we have $u(t^\star) = \mathbb{E}\left\{U(t^\star) \mid X(t^{\star\ell})X(t^{\star r})\right\}$. Besides, $u(t)$ is a function of $X(t^\ell)$ and $X(t^r)$. As a consequence, by the properties of conditional expectancy, we have

$$\mathbb{E}\left\{U(t^*)u(t)\right\} = \mathbb{E}\left\{u(t^*)u(t)\right\} \ .$$

According to Lemma 2,

$$\begin{aligned}
\mathbb{E}\left\{u(t^*)u(t)\right\} &= \alpha(t^\star)\ \alpha(t)\ \mathbb{E}\left\{X(t^{\star\ell})X(t^\ell)\right\} \ + \ \beta(t^\star)\ \alpha(t)\ \mathbb{E}\left\{X(t^{\star r})X(t^\ell)\right\} \\
&+ \alpha(t^\star)\ \beta(t)\ \mathbb{E}\left\{X(t^{\star\ell})X(t^r)\right\} \ + \ \beta(t^\star)\ \beta(t)\ \mathbb{E}\left\{X(t^{\star r})X(t^r)\right\} \\
&= \alpha(t^\star)\ \alpha(t)\ \rho(t^\ell, t^{\star\ell}) \ + \ \beta(t^\star)\ \alpha(t)\ \rho(t^\ell, t^{\star r}) \\
&+ \alpha(t^\star)\ \beta(t)\ \rho(t^{\star\ell}, t^r) \ + \ \beta(t^\star)\ \beta(t)\ \rho(t^r, t^{\star r}) \ .
\end{aligned}$$

In order to obtain $\mathbb{E}\left\{u(t^*)u(t^\ell)\right\}$, we just have to use the dominated convergence theorem. As a result

$$\mathbb{E}\left\{u(t^*)u(t^\ell)\right\} = \alpha(t^\star)\ \rho(t^\ell, t^{\star\ell}) \ + \ \beta(t^\star)\ \rho(t^\ell, t^{\star r}) \ .$$

To conclude the proof, we just have to notice that

$$\begin{aligned}
\mathbb{E}\left\{u(t^*)u(t^\ell)\right\} &= \rho(t^\ell, t^{\star\ell})\left\{\alpha(t^\star) + \beta(t^\star)\rho(t^{\star\ell}, t^{\star r})\right\} \quad \text{if} \quad t^\star > t^\ell \\
&= \rho(t^\ell, t^{\star r})\left\{\alpha(t^\star)\rho(t^{\star r}, t^{\star\ell}) + \beta(t^\star)\right\} \quad \text{if} \quad t^\star < t^\ell \quad .
\end{aligned}$$

In order to obtain $\mathbb{E}\left\{u(t^*)u(t^r)\right\}$, we just have to replace t^ℓ by t^r. This gives the result. □

References

Azaïs JM, Cierco-Ayrolles C (2002) An asymptotic test for quantitative gene detection. Ann. Inst. Henri Poincaré (B) 38(6):1087-1092.

Azaïs JM, Delmas C, Rabier CE (2012) Likelihood ratio test process for Quantitative Trait Locus detection. to appear in Statistics.

Azaïs JM, Gassiat E, Mercadier C (2006) Asymptotic distribution and local power of the likelihood ratio test for mixtures. Bernoulli 12(5):775-799.

Azaïs JM, Gassiat E, Mercadier C (2009) The likelihood ratio test for general mixture models with possibly structural parameter. ESAIM 13:301-327.

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

Chang MN, Wu R, Wu SS, Casella G (2009) Score statistics for mapping quantitative trait loci. Statistical Application in Genetics and Molecular Biology 8(1):16.

Cierco C (1998) Asymptotic distribution of the maximum likelihood ratio test for gene detection. Statistics 31:261-285.
Davies RB (1977) Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 64:247-254.
Davies RB (1987) Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74:33-43.
Foss E, Lande R, Stahl FW, Steinberg CM (1993) Chiasma interference as a function of genetic distance. Genetics 133:681-691.
Gassiat E (2002) Likelihood ratio inequalities with applications to various mixtures. Ann. Inst. Henri Poincaré (B) 6:897-906.
Genz A (1992) Numerical computation of multivariate normal probabilities. J. Comp. Graph. Stat. 1:141-149.
Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. Journal of Genetics 8:299-309.
Hayes B (2007) QTL Mapping, MAS, and Genomic Selection. Short course organized by Iowa State University.
Hillers KJ, Villeneuve AM (2003) Chromosome-wide control of meiotic crossing over in C.elegans. Current Biology 13:1641-1647.
Huang N, Parco A, Mew T, Magpantay G, McCouch S, Guiderdoni E, Xu J, Subudhi P, Angeles ER, Khush GS (1997) RFLP mapping of isozimes, RAPD and QTLs for grain shape, brown planthopper resistance in a doubled haploid rice population. Molecular Breeding 3:105-113.
Karlin S, Liberman U (1979) A natural class of multilocus recombination processes and related measures of crossover interference. Adv. Appl. Probab. 11:479-501.
King JS, Mortimer RK (1990) A polymerization model of chiasma interference and corresponding computer simulation. Genetics 126:1127-1138.
Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 138:235-240.
Lobo I, Shaw K (2008) Thomas Hunt Morgan, genetic recombination, and gene mapping. Nature Education 1:1.
Martini E, Diaz LD, Hunter N, Keeney S (2006) Crossover homeostasis in yeast meiosis. Cell 126:285-295.
McPeek MS, Speed TP (1995) Modeling interference in genetic recombination. Genetics 139:1031-1044.
Muller HJ (1916) The mechanism of crossing-over. Am. Nat. 50:193-221, 284-305, 350-366, 421-434.
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.
Siegmund D, Yakir B (2007) The statistics of gene mapping. Springer, New-York.
Stam (1979) Interference in genetic crossing over and chromosome mapping. Genetics 92:573-594.
Sturtevant AH (1915) The behavior of the chromosomes as studied through linkage. Z. Indukt. Abstammungs. Vererbungsl. 13:234-287.
Van der Vaart AW (1998) Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics.
Youlds J, Boulton S (2011) The choice in meiosis-defining the factors that influence crossover or non-crossover formation J. Cell Sci. 124:501-513.
Wu R, Ma CX, Casella G (2007) Statistical Genetics of Quantitative Traits. Springer, New-York.

number of markers	101	51	41	26	6
Rebai	9.74 2.69%	9.09 3.23%	8.88 3.77%	8.43 4.04%	6.92 4.83%
Azaïs et al.	8.41 5.03%	8.27 4.80%	8.16 5.32%	7.91 5.21%	6.76 5.19%

Table 1 Threshold and Percentage of False Positives (10000 samples of $n = 200$) as a function of the number of markers and the method considered.The chromosome is of length $T = 1$ Morgan and the markers are equally spaced.

Genetic Map	$t^\star$	without interference	interference
map 1	0.10	73.35%	79.23%
	0.80	58.13%	73.40%
	1.30	74.83%	80.26%
	3.70	53.00%	70.35%
map 2	0.20	47.02%	66.39%
	1.90	71.93%	75.93%
	3.35	38.92%	62.19%
	5.75	78.12%	80.26%
map 3	0.25	61.77%	74.20%
	0.60	75.34%	82.01%
	2.80	64.14%	75.77%
	3.10	75.31%	81.87%
map 4	0.18	92.59%	93.52%
	0.44	91.34%	92.03%
	0.70	89.18%	90.45%

Table 2 Asymptotic power of the Interval Mapping as a function of the genetic map, the model considered and the location of the QTL $t^\star$ in Morgan ($a = 4$, $\sigma = 1$, 100000 paths).

	T	K	marker locations
map 1	5	11	$\forall k = 1, ..., 5 \;\; t_{2k} = k - 0.60$ and $\forall k = 0, ..., 5 \;\; t_{2k+1} = k$
map 2	7	15	$\forall k = 1, ..., 7 \;\; t_{2k} = k - 0.30$ and $\forall k = 0, ..., 7 \;\; t_{2k+1} = k$
map 3	4	9	$\forall k = 1, ..., 9 \;\; t_k = 0.50(k-1)$
map 4	1	6	$\forall k = 1, ..., 6 \;\; t_k = 0.20(k-1)$

Table 3 The different genetic maps considered (K is the number of markers, T is the length of the chromosome in Morgan, t_k is the location of marker k in Morgan).

$t^\star$	0.25	0.60	2.80	3.10
EP for $n = 50$	63.08%	72.86%	64.77%	73.27%
EP for $n = 100$	68.88%	77.71%	70.09%	77.17%
EP for $n = 200$	71.79%	79.87%	73.33%	80.02%
EP for $n = 1000$	74.24%	81.55%	74.96%	81.64%
Theoretical Power	74.20%	82.01%	75.77%	81.87%

Table 4 Theoretical Power and Empirical Power (EP) under the interference model and as a function of the location of the QTL, $t^\star$, in Morgan (map 3, $a = 4$, $\sigma = 1$, 100000 paths for the Theoretical Power, 10000 samples for EP).