HAL open science

Likelihood Ratio Test process for Quantitative Trait Loci detection.

Charles-Elie Rabier, Jean-Marc Azaïs, Céline Delmas

To cite this version:

Charles-Elie Rabier, Jean-Marc Azaïs, Céline Delmas. Likelihood Ratio Test process for Quantitative Trait Loci detection.. 2009. hal-00421215

HAL Id: hal-00421215

https://hal.science/hal-00421215v1

Preprint submitted on 1 Oct 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Likelihood Ratio Test process for Quantitative Trait Loci detection

Charles-Elie Rabier

Institut de Mathématiques de Toulouse, Toulouse, France.
INRA UR631, Auzeville, France.

Jean-Marc Azaïs

Institut de Mathématiques de Toulouse, Toulouse, France.

Céline Delmas

INRA UR631, Auzeville, France.

Summary. We consider the likelihood ratio test (LRT) process related to the test of the absence of QTL on the interval $[0, T]$ representing a chromosome (a QTL denotes a quantitative trait locus, i.e. a gene with quantitative effect on a trait). We give the asymptotic distribution of this LRT process under the null hypothesis that there is no QTL on $[0, T]$ and under the general alternative that there exist m QTL on $[0, T]$. We propose to estimate the number of QTL, their positions and their effects by penalized likelihood. Our results are extended to the case where individuals are structured into families.

Keywords: Gaussian process, Likelihood Ratio Test, Mixture models, Nuisance parameters present only under the alternative, QTL detection, χ^2 process.

1. Introduction

We study a population of progenies of a sire and 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 independent and identically distributed (iid). The mechanism of genetics, or more precisely of meiosis, implies that among the two chromosomes of each individual, one is inherited from the dame (whose effect will be neglected) and the other, inherited from the sire, consists of parts originated from chromosome 1 of the sire and parts originated from chromosome 2 of the sire, due to crossing-overs. Note that 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. Using the Haldane (1919) distance and modelling, each chromosome will be represented by a segment $[0, T]$. The distance on $[0, T]$ is called the genetic distance (which is measured in Morgans). The key point is that, if the true position of the QTL is $t = t^\star$, the response Y obeys to a mixture model with known weights :

$$p(t) f_{(\mu+q,\sigma)}(.) + \{1 - p(t)\} f_{(\mu-q,\sigma)}(.) \tag{1}$$

where $f_{(\mu,\sigma)}(.)$ denotes a Gaussian density with mean μ and variance σ^2. $(\mu,\ q,\ \sigma)$ are the unknown parameters. At every location $t \in [0, T]$, we perform a likelihood ratio test (LRT) of the hypothesis "$q = 0$" in formula (1) based on n observations $Y_1, ..., Y_n$. We

call $\Lambda_n(t)$ the obtained quantity. The dependence on t of the weights is precisely described in Section 3. We denote $p_j(t)$ the value of the weight $p(t)$ for the jth observation. The process $\{\Lambda_n(t),\ t \in [0,T]\}$ will be called "likelihood ratio test process" and taking as test statistic the maximum of this process comes down to perform a LRT in a model when the localisation of the QTL is an extra parameter.

In the special case where the weights are 0 or 1 depending on the individual, Lander and Botstein (1989) stated that the asymptotic distribution of the LRT process along $[0,T]$ is the square of an Ornstein-Uhlenbeck process. This result has been proved by Cierco (1998). Bounds for the distribution of the maximum of a regularization of an Ornstein-Uhlenbeck process were proposed by Azaïs and Cierco-Ayrolles (2002), Azaïs and Wschebor (2009). Some results about the asymptotic distribution of the LRT process under the null hypothesis are given in Rebaï et al. (1994) for a special modelling of the weights. Their results are inferred from the bounds given by Davies (1977), Davies (1987) for the maximum of sufficiently regular Gaussian and chi-square processes.

In this paper we consider the modelling of the weight used by geneticists to detect QTL, so called Interval Mapping. First we give the asymptotic distribution of the LRT process along the interval $[0,T]$ under the null hypothesis that there is no QTL on $[0,T]$ ($q=0$) and under the alternative that there is one QTL at $t^\star$ on $[0,T]$ which means that the quantitative trait for each individual is distributed as the mixture in formula (1) with $t=t^\star$. Then we compute the asymptotic distribution of the LRT process under the general alternative that there exist m QTL on $[0,T]$ at $t_1^\star,\cdots,t_m^\star$ with additive effects $q^1,\cdots,q^m$. The response is now a mixture of $M=2^m$ components of the form :

$$\sum_{\alpha=1}^{M} p_\alpha f_{(m_\alpha,\sigma)}(.)$$

where the p_αs and the m_αs are known functions of the unknown parameters μ, m, $t_1^\star$, ..., $t_m^\star$, q^1, ..., q^m. Under this general alternative, the LRT process is shown to converge towards a Gaussian process with mean function depending on these unknown parameters. We propose to estimate the unknown parameters by penalized likelihood.
Besides, we show that the LRT process is asymptotically the square of a "non linear interpoled process" (which means that the LRT statistics at each point can easily be deduced from the Wald or score statistics calculated at the positions where the auxiliary information is available). Note that in some remarks sections we also prove that the LRT process obtained by Rebaï et al. (1994), Rebaï et al. (1995) is asymptotically the square of a "linear interpoled process" and we generalize their results to the alternative hypothesis. Finally, our results are extended to the case where individuals are structured into families of sires. Recently, the law of the LRT process under the null hypothesis has also been obtained by Chang et al. (2009). Our work has been done independently. Technical differences are presented in appendix 8.5.

The originality of our paper is twofold. First we consider the true model used by geneticists to detect QTL whereas the model considered by Rebaï et al. is only an approximation. Then we obtain results not only under the null hypothesis, but also under the general alternative. This last result leads us to propose a new method, based on penalized likelihood, for estimating the number of QTL, their positions and their effects. We refer to the book of Van der Vaart (1998) for element of asymptotic statistics used in proofs. In a future paper, we will present the applications of the theoretical results presented here.

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 can be written mathematically : let $N(t)$ be a standard Poisson process, the law of $X(t)$ is $\frac{1}{2}(\delta_1 + \delta_{-1})$ and $X(t) = (-1)^{N(t)} X(t_1)$. The Haldane (1919)'s function $r : [0,\ T]^2 \longmapsto \left[0,\ \frac{1}{2}\right]$ is such as :

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

$\bar{r}(t, t')$ will be the function equal to $1 - r(t, t')$.
We are interested in a quantitative trait Y which depends on the value of $X(t)$ at $t^\star \in [t_1, t_K]$ which is the location of the QTL. The quantitative trait verifies :

$$Y_j = \mu\ +\ X(t^\star)\ q\ +\ \sigma\varepsilon$$

where ε is a Gaussian white noise and q the effect of the QTL.
Besides, the "genome information" is available only at locations of genetic markers, that is to say at $t_1, t_2, ..., t_K$. We denote by $X_j(t)$ the value of the variable $X(t)$ for the jth observation. So, in fact, our observation on each individual is $(Y_j,\ X_j(t_1),\ ...,\ X_j(t_K))$. These observations are supposed to be iid. The goal of this study is to test if q is equal to zero. The challenge is that $t^\star$ is unknown.

3. Only 2 genetic markers

To begin, we suppose that there are only two markers ($K = 2$) located at 0 and T : $0 = t_1 < t_2 = T$. As explained previously, we are looking for a QTL lying at a position $t^\star \in [t_1, t_2]$. Let $t \in [t_1, t_2]$. It is clear that the weight $p(t)$ satisfies $p(t) = \mathbb{P}\left\{X(t) = 1 \middle| X(t_1), X(t_2)\right\}$. Consider for example the case $X(t_1) = X(t_2) = 1$, then by the Bayes rule :

$$\mathbb{P}\left\{X(t) = 1 \middle| X(t_1) = 1, X(t_2) = 1\right\} = \frac{(1/2)\ \mathbb{P}\left\{N(t) - N(t_1)\ \text{even}\right\} \mathbb{P}\left\{N(t_2) - N(t)\ \text{even}\right\}}{(1/2)\ \mathbb{P}\left\{N(t_2) - N(t_1)\ \text{even}\right\}}$$

So that, in general $\forall t \in]t_1, t_2[$:

$$\begin{aligned} p(t) = {} & Q_t^{1,1}\ 1_{X(t_1)=1} 1_{X(t_2)=1}\ +\ Q_t^{1,-1}\ 1_{X(t_1)=1} 1_{X(t_2)=-1} \\ & + Q_t^{-1,1}\ 1_{X(t_1)=-1} 1_{X(t_2)=1}\ +\ Q_t^{-1,-1}\ 1_{X(t_1)=-1} 1_{X(t_2)=-1} \end{aligned} \tag{2}$$

where :

$$Q_t^{1,1} = \frac{\bar{r}(t_1, t)\ \bar{r}(t, t_2)}{\bar{r}(t_1, t_2)} \quad , \quad Q_t^{1,-1} = \frac{\bar{r}(t_1, t)\ r(t, t_2)}{r(t_1, t_2)}$$
$$Q_t^{-1,1} = \frac{r(t_1, t)\ \bar{r}(t, t_2)}{r(t_1, t_2)} \quad , \quad Q_t^{-1,-1} = \frac{r(t_1, t)\ r(t, t_2)}{\bar{r}(t_1, t_2)}$$

We can remark that we have :

$$Q_t^{-1,-1} = 1 - Q_t^{1,1} \quad \text{and} \quad Q_t^{-1,1} = 1 - Q_t^{1,-1}$$

Besides, $p(t_1) = 1_{X(t_1)=1}$ and $p(t_2) = 1_{X(t_2)=1}$. So, the weights $p(t)$ are continuous at t_1 and t_2.

Let $\theta = (q,\ \mu,\ \sigma)$ be the parameter of the model at t fixed and $\theta_0 = (0,\ \mu,\ \sigma)$ the true value of the parameter under H_0. 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 county measure on $\mathbb{N}$, is $\forall t \in [t_1, t_2]$:

$$L(\theta, t) = \left[p(t) f_{(\mu+q,\sigma)}(y) + \{1 - p(t)\}\, f_{(\mu-q,\sigma)}(y)\right]\ g(t) \tag{3}$$

where

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

The likelihood $L_n(\theta, t)$ for n observations is obtained by the product of n terms as above. $\hat{\theta} = (\hat{q},\ \hat{\mu},\ \hat{\sigma})$ will be the maximum likelihood estimator (MLE) of θ.

Under H_0, there is no QTL lying on the interval $[t_1, t_2]$. Besides, under H_1, it is supposed that there is only one location where the QTL lies. The location of the QTL, $t^\star$ ($t^\star \in [t_1, t_2]$), will be added in the definition of H_1. So, the alternative hypothesis can be written :

$H_{at^\star}$: "the QTL is located at the position $t^\star$ with effect $q = a/\sqrt{n}$ where $a \in \mathbb{R}^\star$ "

The QTL effect q is such as $q = a/\sqrt{n}$ in order to deal with Le Cam (1986)'s theory.

3.1. Results

Theorem 1 *With the previous defined notations,*

$$S_n(.) \Rightarrow Z(.) \quad , \quad \Lambda_n(.) \overset{F.d.}{\rightarrow} \{Z(.)\}^2$$

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

- *$S_n(.)$ is the score process for n observations*
- *$\Rightarrow$ is the weak convergence and $\overset{F.d.}{\rightarrow}$ is the convergence of finite-dimensional distributions*
- *$Z(.)$ is the Gaussian process with covariance function $\forall (t, t') \in [t_1, t_2]^2$:*

$$\Gamma(t, t') = \frac{4\mathbb{E}\left\{p(t)p(t')\right\} - 1}{\sqrt{\mathbb{E}\left[\{2p(t) - 1\}^2\right]}\sqrt{\mathbb{E}\left[\{2p(t') - 1\}^2\right]}}$$

  and expectation $\forall (t, t^\star) \in [t_1, t_2]^2$:

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

$$m_{t^\star}(t) = \frac{a\ \mathbb{E}\left[X(t^\star)\left\{2p(t) - 1\right\}\right]}{\sigma\ \sqrt{\mathbb{E}\left[\{2p(t) - 1\}^2\right]}}$$

Another way of characterizing $Z(.)$ is that $Z(.)$ is the non linear interpoled process such as $\forall t \in [t_1, t_2]$:

$$Z(t) = \{\ \alpha(t)\ Z(t_1)\ +\ \beta(t)\ Z(t_2)\ \} / \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}$$

where $\forall t \in]t_1, t_2[$, $\alpha(t) = Q_t^{1,1} + Q_t^{1,-1} - 1$, $\beta(t) = Q_t^{1,1} - Q_t^{1,-1}$ and $\alpha(t_1) = 1$, $\beta(t_1) = 0$, $\alpha(t_2) = 0$, $\beta(t_2) = 1$, $Cov\{Z(t_1), Z(t_2)\} = e^{-2t_2}$.
In the same way, $\forall (t, t^\star) \in [t_1, t_2]^2$:

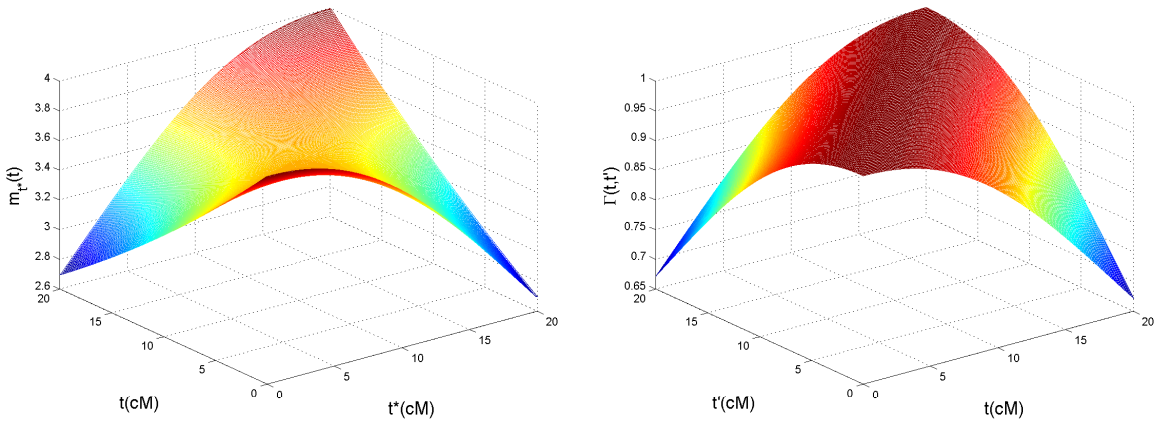
$$m_{t^\star}(t) = \{\ \alpha(t)\ m_{t^\star}(t_1)\ +\ \beta(t)\ m_{t^\star}(t_2)\ \} / \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}$$

The quantity $\mathbb{E}\left[\{2p(t)-1\}^2\right]$ is given in formula (10) of the proof of the theorem in Section 7.1. $\mathbb{E}\{p(t)p(t')\}$ is given in appendix 8.1. $\mathbb{E}[X(t^\star)\{2p(t)-1\}]$ is given in formula (15) of the proof in Section 7.1.

We limit our attention to finite dimensional convergence since for the applications, the interval studied is always discretized, Wu et al. (2007).

Figures 1 represent the covariance function $\Gamma(t, t')$ and also the mean function $m_{t^\star}(t)$. T is equal to 0.2M. We can remark that the covariance function is regular.

Contrary to Azaïs et al. (2006) and Azaïs et al. (2009), the shift at position t is not $\Gamma(t, t^\star)$. The model considered here is more complicated due to the fact that an observation includes the quantitative trait Y and the "genome information", $X(t_1)$ and $X(t_2)$. As it is well



Mean function Covariance function

Fig. 1. Mean function and Covariance function ($a = 4$, $\sigma = 1$, $T = 0.2$M)

known, for regular model, LRT is equivalent to Wald test, and score test in the sense that $\forall t \in [t_1, t_2]$:

$$\Lambda_n(t) = \{W_n(t)\}^2 + o_{P_{\theta_0}}(1) = \{S_n(t)\}^2 + o_{P_{\theta_0}}(1)$$

where $W_n(t)$ and $S_n(t)$ are respectively the Wald and the score test statistic for n observations. We remind that, as in the proof of the theorem in Section 7.1, the notation $o_{P_{\theta_0}}(1)$ is short for a sequence of random vectors that converges to zero in probability under H_0 (i.e. no QTL on the whole interval studied).
According to formula (9), given in Section 7.1 :

$$W_n(t) = \sqrt{n}\ \hat{q}\ \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}/\sigma \ \ , \ \ S_n(t) = \sum_{j=1}^{n} \frac{(y_j-\mu)\ (2p_j(t)-1)}{\sqrt{n}\ \sigma\ \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}} \tag{4}$$

Note that the Wald test can be obtained, replacing σ by $\hat{\sigma}$, according to Slutsky's lemma. The score test can be obtained, replacing μ by $\hat{\mu}$, according to Prohorov, and replacing σ by $\hat{\sigma}$, according to Slutsky's lemma. Nevertheless, in order to make the reading easier, the Wald and the score test statistic are defined as in formula (4). The score process considered in theorem 1 is based on this formula. Howewer, we have the same result as in theorem 1 for the other score process because the tightness of this process is obvious according to the proof of theorem 1.

After some calculations, we can remark that :

$$S_n(t) = \{\ \alpha(t)\ S_n(t_1)\ +\ \beta(t)\ S_n(t_2)\ \}\ /\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]} \tag{5}$$

with $\mathrm{Cov}\left\{S_n^0(t_1), S_n^0(t_2)\right\} = e^{-2t_2}$ where $S_n^0(.)$ is the score process under H_0.
It comes :

$$\begin{aligned}\Lambda_n(t) &= \{\ \alpha(t)\ S_n(t_1)\ +\ \beta(t)\ S_n(t_2)\ \}^2\ /\ \mathbb{E}\left[\{2p(t)-1\}^2\right]\ +\ o_{P_{\theta_0}}(1)\\ &= \{\ \alpha(t)\ W_n(t_1)\ +\ \beta(t)\ W_n(t_2)\ \}^2\ /\ \mathbb{E}\left[\{2p(t)-1\}^2\right]\ +\ o_{P_{\theta_0}}(1)\end{aligned}$$

Besides, by contiguity (cf. proof of theorem 1 in Section 7.1), the quantity $o_{P_{\theta_0}}(1)$ converges also to zero under $H_{at^\star}$. That is to say, the LRT statistic at a position t between the two genetic markers is asymptotically equal to the square of a non linear interpolation between the Wald or score test statistics on the markers.
Note that computing the square of this non linear interpoled process will be be quicker than calculating the observed process $\Lambda_n(.)$ on real data, which is time consuming because an EM algorithm is required to calculate the MLE's when the position tested is not on genetic markers.

3.2. Remarks

To construct an approximation of $S_n(.)$ (and $\Lambda_n(.)$), we introduce a new process $V_n(.)$ which is obtained from $S_n(.)$ by :

- linear (or polygonal) interpolation
- renormalization

More precisely :

$$V_n(t) = \left\{ \frac{t_2 - t}{t_2} \, S_n(t_1) \; + \; \frac{t}{t_2} \, S_n(t_2) \right\} / \sqrt{\tau(t)} \tag{6}$$

where

$$\tau(t) = \mathbb{V} \left\{ \frac{t_2 - t}{t_2} \, S_n^0(t_1) \; + \; \frac{t}{t_2} \, S_n^0(t_2) \right\} = \left(\frac{t_2 - t}{t_2} \right)^2 \; + \; 2 \, \frac{t(t_2 - t)}{(t_2)^2} e^{-2t_2} \; + \; \left(\frac{t}{t_2} \right)^2$$

It can be seen easily that $\tau(t) \neq 0$, $\forall t \in [t_1, t_2]$. $V_n(.)$ remains asymptotically a Gaussian process, centered under H_0, with unit variance and $\mathrm{Cov}\left\{ S_n^0(t_1), S_n^0(t_2) \right\} = e^{-2t_2}$.

Some comments about the linear interpoled process $V_n(.)$:

(a) According to formula (11) in Section 7.1 and after some calculations, we can establish that asymptotically, the process $V_n^2(.)$ corresponds to likelihood ratio tests for a mixture model whose weights verify :

$$p(t) = 1_{X(t_1)=1} 1_{X(t_2)=1} \; + \; \frac{t_2 - t}{t_2} \, 1_{X(t_1)=1} 1_{X(t_2)=-1} \; + \; \frac{t}{t_2} \, 1_{X(t_1)=-1} 1_{X(t_2)=1} \tag{7}$$

We can remark that these weights are an approximation at the first order of the weights considered previously in formula (2). So, the linear interpoled process will be a good approximation if and only if the genetic markers are close from each other.

(b) $V_n^2(.)$ is a generalization of the process studied, under H_0, by Rebaï et al. (1995) : the number of individuals in each class is not equal to the expectations (respectively $n\bar{r}(t_1, t_2)/2$, $nr(t_1, t_2)/2$, $nr(t_1, t_2)/2$, $n\bar{r}(t_1, t_2)/2$) but is still random (respectively $\sum_{j=1}^n 1_{X_j(t_1)=1} 1_{X_j(t_2)=1}$, $\sum_{j=1}^n 1_{X_j(t_1)=1} 1_{X_j(t_2)=-1}$, $\sum_{j=1}^n 1_{X_j(t_1)=-1} 1_{X_j(t_2)=1}$ and $\sum_{j=1}^n 1_{X_j(t_1)=-1} 1_{X_j(t_2)=-1}$).

(c) By contiguity (cf. proof of theorem 1 in Section 7.1), under $H_{at^\star}$, $V_n(.)$ is asymptotically the same process as under H_0 on which the mean function $\tilde{m}_{t^\star}(t)$ has been added. $\tilde{m}_{t^\star}(t)$ is such as :

$$\tilde{m}_{t^\star}(t) = \left\{ \frac{t_2 - t}{t_2} \, m_{t^\star}(t_1) \; + \; \frac{t}{t_2} \, m_{t^\star}(t_2) \right\} / \sqrt{\tau(t)}$$

(d) $V_n(.)$ is defined here with $\mathrm{Cov}\left\{ S_n^0(t_1), S_n^0(t_2) \right\} = e^{-2t_2}$. In order to consider other covariances between $S_n^0(t_1)$ and $S_n^0(t_2)$, $\tau(.)$ has to be adapted. It can easily be seen that the new process $V_n^2(.)$ is still a generalization of the process studied by Rebaï et al. (1995) for any covariance between $S_n^0(t_1)$ and $S_n^0(t_2)$ as soon as $\mathbb{E}\left[\{2p(t) - 1\}^2 \right] \neq 0$ ($p(t)$ verifies formula (7)).

4. Several markers : the "Interval Mapping" of Lander and Botstein (1989)

In that case suppose that there are K markers $0 = t_1 < t_2 < ... < t_K = T$. We consider values t, 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 :

$$t^\ell = sup \left\{ t_k \in \mathbb{T}_k : t_k < t \right\} \quad , \quad t^r = inf \left\{ t_k \in \mathbb{T}_k : t < t_k \right\}$$

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

Theorem 2 *We have the same result as in theorem 1 except that the following expressions are more complicated :*

$$\mathbb{E}\left[\{2p(t)-1\}^2\right] \quad , \quad \mathbb{E}\left\{p(t)p(t')\right\} \quad , \quad \mathbb{E}\left[X(t^\star)\left\{2p(t)-1\right\}\right] \quad , \quad \alpha(t) \quad , \quad \beta(t)$$

Besides, $Z(.)$ is now the non linear interpoled process such as :

$$Z(t) = \left\{\ \alpha(t)\ Z(t^\ell)\ +\ \beta(t)\ Z(t^r)\ \right\} / \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}$$

with $\forall k\ \forall k'$, $Cov\left\{Z(t_k), Z(t_{k'})\right\} = e^{-2|t_k - t_{k'}|}$.
In the same way, the mean function $m_{t^\star}(t)$ is now such as :

$$m_{t^\star}(t) = \left\{\ \alpha(t)\ m_{t^\star}(t^\ell)\ +\ \beta(t)\ m_{t^\star}(t^r)\ \right\} / \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}$$

All these expressions including a proof are given in appendix 8.2.

Note that $\forall k\ \ \forall k'$, $\Gamma(t_k, t_{k'}) = e^{-2|t_k - t_{k'}|}$. It is relative to an Ornstein-Uhlenbeck process, as studied in Lander and Botstein (1989), and Cierco (1998).
Besides, in the same way as what has been done in Section 3.1, we have :

$$\forall k \quad W_n(t_k) = \sqrt{n}\ \hat{q}/\sigma \quad , \quad S_n(t_k) = \sum_{j=1}^{n} \frac{(y_j - \mu)\ (2\ 1_{X_j(t_k)=1} - 1)}{\sigma\ \sqrt{n}}$$

$$\begin{aligned} \Lambda_n(t) &= \left\{\ \alpha(t)\ S_n(t^\ell)\ +\ \beta(t)\ S_n(t^r)\ \right\}^2 / \mathbb{E}\left[\{2p(t)-1\}^2\right]\ +\ o_{P_{\theta_0}}(1) \\ &= \left\{\ \alpha(t)\ W_n(t^\ell)\ +\ \beta(t)\ W_n(t^r)\ \right\}^2 / \mathbb{E}\left[\{2p(t)-1\}^2\right]\ +\ o_{P_{\theta_0}}(1) \end{aligned} \tag{8}$$

Note that $\forall k\ \forall k'$, $\mathrm{Cov}\left\{S_n^0(t_k), S_n^0(t_{k'})\right\} = e^{-2|t_k - t_{k'}|}$.
Besides, by contiguity (cf. appendix 8.2), the quantity $o_{P_{\theta_0}}(1)$ converges also to zero under $H_{at^\star}$.

4.1. Remarks

The linear interpoled process $V_n(.)$ presented in Section 3.2 can easily be generalized to the case of several markers. This process is a generalization of the process studied, under H_0, by Rebaï et al. (1994). The details are given in appendix 8.3.

On the other hand, the problem considered in this article can be viewed as a missing data problem. The auxiliary information $X(t)$ is available only at the location of genetic markers, otherwise the information is missing. Since in absence of missing data, the process is relative to an Ornstein-Uhlenbeck process, the missing observations can be obtained by a kriging method. The process refering to the kriging method will be called $M_n(.)$. After some easy

calculations, we obtain :

$$\begin{aligned} M_n(t) &= \left\{ e^{-2(t-t^\ell)} - \gamma(t)\, e^{-2(t^r-t^\ell)} \right\} S_n(t^\ell) \;\; + \;\; \gamma(t)\, S_n(t^r) \\ &= \left\{ e^{-2(t-t^\ell)} - \gamma(t)\, e^{-2(t^r-t^\ell)} \right\} W_n(t^\ell) \;\; + \;\; \gamma(t)\, W_n(t^r) \;+\; o_{P_{\theta_0}}(1) \end{aligned}$$

where $\gamma(t) = \frac{e^{-2(t^r-t)} - e^{-2(t-2t^\ell+t^r)}}{1-e^{-4(t^r-t^\ell)}}$. Note that this process is asymptotically a Gaussian process, centered under H_0 but with unit variance only at location of genetic markers.

Under $H_{at^\star}$, by contiguity (in the same way of what has been done in the proof of theorem 1), this is asymptotically the same process but the mean function, $\bar{m}_{t^\star}(t)$ is added to the process:

$$\bar{m}_{t^\star}(t) = \left\{ e^{-2(t-t^\ell)} - \gamma(t)\, e^{-2(t^r-t^\ell)} \right\} m_{t^\star}(t^\ell) \;\; + \;\; \gamma(t)\, m_{t^\star}(t^r)$$

If now a rescaling is done in order to obtain a process with unit variance, then by Taylor expansions at the first order, it can be seen that the process obtained by the kriging method is exactly the interpoled process $V_n(.)$ (asymptotic is not required).

4.2. Illustration

Figure 2 represents under H_0, the cumulative distribution function of the square of the sup of four different processes on $[0, 0.6]$ with 4 markers equally spaced every 0.2M : the asymptotic process $Z(.)$, the limiting process of the linear interpoled process $V_n(.)$, the limiting process of the kriging process $M_n(.)$ and the asymptotic process for which the tests are only done on the markers. 100000 sample paths of the different processes have been simulated. Each test is done every cM. The interest is on the quantile of order 95%. It is not surprising that when tests are only on markers, the cumulative density function is above the three other curves. Howewer, the curves of the linear interpoled process and the process $Z(.)$ are pretty close : the estimation of the quantile of order 95% is 6.06 for the linear interpoled and 6.08 for the process $Z(.)$. (5.86 if tests are done only on markers). It is not surprising because markers are close from each others. On the other hand, the kriging process is another way of analyzing data, that's why the corresponding cumulative distribution is so different (the estimation of the quantile is 6.77).

5. Generalization

In the previous sections, we were looking for a QTL lying on the interval $[0, T]$ using the concept of Interval Mapping. One population of progenies of a sire has been studied that is to say one family. In order to increase the power of the method, geneticists look for the QTL not in one family but simultanously in several families, each defined by a different sire. It increases the chances to study families whose sires are heterozygote at the QTL. In that case, the Interval Mapping method is also used : LRT are peformed at each position $t \in [0, T]$ and the supremum of these statistics is used as a unique test statistic. Naturally, the putative QTL is supposed to be lying at the same location in each family otherwise the concept of Interval Mapping has no sense.

Let's extend the model presented in Section 2 to the case of several families. Let $I \in \mathbb{N}^\star$ be the number of families. Let C be a discrete random variable refering to the family :

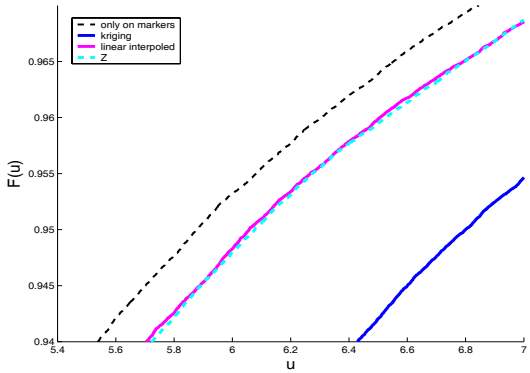


Fig. 2. Cumulative distribution function of the square of four different processes ($T = 0.6$M, 4 markers equaly spaced every 0.2M)

$\pi_i = P(C = i)$. In other words, the individual belongs to family i with probability π_i. When we deal with different families, the location and the number of the genetic markers usually differ in each family. However, in order to make reading easier, we will supposed here, that the location and the number of genetics markers do not differ with the family. The general results are present in Rabier (2009).

In our case, the process $X(.)$ is unchanged. Howewer, the quantitative trait verifies now :

$$\big(Y \big| C = i\big) \;=\; \mu_i \;+\; X(t^\star)\, q_i \;+\; \sigma\, \varepsilon$$

where μ_i and q_i are respectively a polygenic effect and the QTL effect inside family i. ε is a Gaussian white noise.

We denote by C_j the value of the variable C for the jth observation. In fact, our observation on each individual is $(Y_j,\ X_j(t_1),\ ...,\ X_j(t_K),\ C_j)$. These observations are supposed to be iid. The goal of this study is to test if all the q_i are equal to zero. As previously, the challenge is that $t^\star$ is unknown.

The same notations as in Section 4 will be used and as previously, we consider only values t, t' or $t^\star$ of the parameters that are distinct of the marker positions. The result will be prolonged by continuity at the markers positions.
Let $\theta = (q_1,\ ...,\ q_I,\ \mu_1,\ ...,\ \mu_I,\ \sigma)$ be the parameter of the model at t fixed and
$\theta_0 = (0,\ ...,\ 0,\ \mu_1,\ ...,\ \mu_I,\ \sigma)$ the true value of the parameter under H_0. The likelihood of the triplet $\big(Y,\ X(t^\ell),\ X(t^r),\ C\big)$ with respect to the measure $\lambda \otimes N \otimes N \otimes N$, λ being the Lebesgue measure, N the county measure on $\mathbb{N}$, is at a position t :

$$L(\theta, t) = \sum_{i=1}^{I} \left[p(t) f_{(\mu_i + q_i, \sigma)}(y) + \{1 - p(t)\}\, f_{(\mu_i - q_i, \sigma)}(y)\right]\ 1_{C=i}\ \frac{\pi_i}{2}\ g(t)$$

where $g(t)$ is the same function as in Section 3 adapted to the Marker interval (t^ℓ, t^r).
The likelihood $L_n(\theta, t)$ for n observations is obtained by the product of n terms as above.
$\hat{\theta} = (\hat{q}_1,\ ...,\ \hat{q}_I,\ \hat{\mu}_1, ...,\ \hat{\mu}_I,\ \hat{\sigma})$ will be the MLE of θ.

The alternative hypothesis can be written :

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

5.1. Results

Theorem 3 *With the previous defined notations,*

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

as n tends to infinity, under H_0 and $H_{at^\star}$ where the $Z^i(.)$ are independent Gaussian processes, with covariance function $\Gamma(t,t')$ and with expectation :

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

$$m^i_{t^\star}(t) = \frac{a_i \sqrt{\pi_i}\, \mathbb{E}\left[X(t^\star)\left\{2p(t)-1\right\}\right]}{\sigma \sqrt{\mathbb{E}\left[\left\{2p(t)-1\right\}^2\right]}}$$

Another way of characterizing $Z^i(.)$ is that $Z^i(.)$ is the non linear process such as :

$$Z^i(t) = \left\{ \alpha(t)\, Z^i(t^\ell) \;+\; \beta(t)\, Z^i(t^r) \right\} / \sqrt{\mathbb{E}\left[\left\{2p(t)-1\right\}^2\right]}$$

with $\forall k\ \forall k'$, $Cov\left\{Z^i(t_k), Z^i(t_{k'})\right\} = e^{-2|t_k - t_{k'}|}$.
In the same way :

$$m^i_{t^\star}(t) = \left\{ \alpha(t)\, m^i_{t^\star}(t^\ell) \;+\; \beta(t)\, m^i_{t^\star}(t^r) \right\} / \sqrt{\mathbb{E}\left[\left\{2p(t)-1\right\}^2\right]}$$

The quantities $\Gamma(t,t')$, $\mathbb{E}\left[X(t^\star)\left\{2p(t)-1\right\}\right]$, $\mathbb{E}\left[\left\{2p(t)-1\right\}^2\right]$, $\alpha(t)$ and $\beta(t)$ are the same as in theorem 2.

The proof of this theorem is given in Section 7.2. Note that this theorem is also suitable when the π_i's are unknown.

In the same way as what has been done in the previous Sections, let $W_n(t_k,i)$ and $S_n(t_k,i)$ be respectively the Wald statistic and the score statistic, which corresponds to testing the presence of a QTL in family i at a marker location t_k. According to the proof of theorem 3 in Section 7.2 :

$$W_n(t_k,i) = \sqrt{n\,\pi_i}\,\hat{q}_i/\sigma \quad , \quad S_n(t_k,i) = \sum_{j=1}^{n} \frac{(y_j-\mu_i)\,(2\,1_{X_j(t_k)=1}-1)}{\sigma\sqrt{n\,\pi_i}} 1_{C_j=i}$$

$$\begin{aligned}\Lambda_n(t) &= \sum_{i=1}^{I} \left\{ \alpha(t)\, S_n(t^\ell,i) \;+\; \beta(t)\, S_n(t^r,i) \right\}^2 / \mathbb{E}\left[\left\{2p(t)-1\right\}^2\right] \;+\; o_{P_{\theta_0}}(1) \\ &= \sum_{i=1}^{I} \left\{ \alpha(t)\, W_n(t^\ell,i) \;+\; \beta(t)\, W_n(t^r,i) \right\}^2 / \mathbb{E}\left[\left\{2p(t)-1\right\}^2\right] \;+\; o_{P_{\theta_0}}(1)\end{aligned}$$

Note that $\mathrm{Cov}\left\{S_n^0(t_k,i), S_n^0(t_{k'},i)\right\} = e^{-2|t_k - t_{k'}|}$.
Besides, by contiguity (cf. Section 7.2), the quantity $o_{P_{\theta_0}}(1)$ converges also to zero under $H_{at^\star}$.

5.2. Remarks

The linear interpoled process and the process obtained by kriging can be generalized to the case of several families. The details are presented in appendix 8.4.

6. Extension to several QTL

Until now, it has been supposed that there was only one QTL lying on the interval $[0,T]$. So, the test statistic used was a natural statistic, that is to say the supremum of the process. The interest is now on studying the same processes as previously, $\Lambda_n(.)$ and $S_n(.)$, but under the presence of several QTL on the interval $[0,T]$. In this case, the goal is not to perform a test anymore, but to be able to run a model selection in order to estimate the number of QTL and their locations.
In order to make the reading easier, we will deal with only one family. m will refer to the number of QTL and q^s to the QTL effect of the sth QTL. Its position will be called $t_s^\star$. We impose $t_1^\star < ... < t_m^\star$ and we will suppose that the QTL effects are additives and there is no interaction between them. So, the quantitative trait Y verifies :

$$Y = \mu \; + \; \sum_{s=1}^{m} X(t_s^\star)\; q^s \; + \; \sigma\varepsilon$$

where ε is a Gaussian white noise.

Let denote $\vec{t}^\star$ the quantity refering to the locations of the QTL. $H_{a\vec{t}^\star}$ will be the following assumption :

$H_{a\vec{t}^\star}$: " there are m QTL located respectively at $t_1^\star$, ..., $t_m^\star$ and with effect $q^1 = \frac{a^1}{\sqrt{n}}$, ..., $q^m = \frac{a^m}{\sqrt{n}}$ where $(a^1, ..., a^m) \in \mathbb{R}^{m\star}$ "

We will consider values t, t', $t_1^\star$, ..., $t_m^\star$ of the parameters that are distinct of the markers positions, and the result will be prolonged by continuity at the markers positions.

6.1. Results

Theorem 4 *With the previous defined notations,*

$$S_n(.) \Rightarrow Z^\star(.) \quad , \quad \Lambda_n(.) \overset{F.d.}{\rightarrow} \{Z^\star(.)\}^2$$

as n tends to infinity, under H_0 and $H_{a\vec{t}^\star}$ where the $Z^\star(.)$ is a Gaussian process, with covariance function $\Gamma(t,t')$ and with expectation :

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

$$m_{\vec{t}^\star}(t) = \sum_{s=1}^{m} \frac{a^s\; \mathbb{E}\left[X(t_s^\star)\;\{2p(t)-1\}\right]}{\sigma\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}}$$

Another way of characterizing $Z^\star(.)$ is that $Z^\star(.)$ is the non linear process such as :

$$Z^\star(t) = \left\{ \alpha(t)\ Z^\star(t^\ell)\ +\ \beta(t)\ Z^\star(t^r)\ \right\} / \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}$$

with $\forall k\ \forall k'$, $Cov\{Z^\star(t_k), Z^\star(t_{k'})\} = e^{-2|t_k - t_{k'}|}$.
In the same way :

$$m_{\vec{t}^\star}(t) = \sum_{s=1}^{m} \frac{a^s}{\sigma} \left\{ \alpha(t)\ \mathbb{E}\left[X(t_s^\star)\ \{2p(t^\ell)-1\}\right]\ +\ \beta(t)\ \mathbb{E}\left[X(t_s^\star)\ \{2p(t^r)-1\}\right]\ \right\} / \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}$$

The quantities $\Gamma(t,t')$, $\mathbb{E}\left[\{2p(t)-1\}^2\right]$, $\alpha(t)$, $\beta(t)$ are the same as in theorem 2.
$\mathbb{E}\left[X(t_s^\star)\ \{2p(t)-1\}\right]$ is given in formula (21) of the proof of the theorem in Section 7.3.

As we focus on the same LRT process as previously, formula (8) of Section 4 is still suitable. Besides, by contiguity (cf. Section 7.3), the quantity $o_{P_{\theta_0}}(1)$ converges also to zero under $H_{a\vec{t}^\star}$.

All the results presented in this Section 6 can easily be generalized to the case of several families and also to interactions between the QTL.

6.2. Estimation of the parameters

As for the application, the interval $[0,T]$ is always discretized, let consider only these points of discretization : $0 = s_1 < s_2 < ... < s_d = T$. Without loss of generality, it can be supposed that the QTL are located on these points of discretization.
We estimate the unknown parameters m, a^1,..., a^m and consequently $t_1^\star$,..., $t_m^\star$ by a penalized likelihood method (lasso Tibshirani (1996), elastic net Zou and Hastie (2005), dantzig selector Candes and Tao (2005)) applied to the model :

$$S_n(s_e) = \sum_{i=1}^{d} \frac{a^{s_i}\ \mathbb{E}\left[X(t_{s_i}^\star)\ \{2p(s_e)-1\}\right]}{\sigma\sqrt{\mathbb{E}\left[\{2p(s_e)-1\}^2\right]}}\ +\ \varepsilon_{s_e} \qquad e = 1, ..., d$$

where ε_{s_e} is a Gaussian white noise and $\mathrm{Cov}(\varepsilon_{s_e}, \varepsilon_{s_{e'}}) = \Gamma(s_e, s_{e'})$.
We remind that $S_n(s_e)$ is the score test for n observations performed at the position s_e.
This method will be investigated in a forthcoming paper.

7. Proofs

Notations : I_θ will be the Fisher information matrix taken at the point θ . $I_{ij}(\theta)$ refers to the element ij of I_θ. $I_{ij}^{-1}(\theta)$ refers to the element ij of I_θ^{-1}, the inverse of I_θ.

7.1. Proof of theorem 1

We first compute the score functions and the Fisher information matrix. Let $t \in [t_1, t_2]$.

$$\frac{\partial \log L}{\partial q}\ |_{\theta_0} = \frac{y-\mu}{\sigma^2}\ \{2p(t)-1\}$$

$$\frac{\partial \log L}{\partial \mu}\,|_{\theta_0} = \frac{y-\mu}{\sigma^2} \quad , \quad \frac{\partial \log L}{\partial \sigma}\,|_{\theta_0} = -\frac{1}{\sigma} + \frac{(y-\mu)^2}{\sigma^3}$$

$$I_{11}(\theta_0) = \frac{\mathbb{E}\left[\{2p(t)-1\}^2\right]}{\sigma^2} \quad , \quad I_{22}(\theta_0) = \frac{1}{\sigma^2}$$

As the fourth-order moment of a standard normal distribution is equal to three,

$$I_{33}(\theta_0) = \frac{2}{\sigma^2}$$

After some calculations, we find : $I_{12}(\theta_0) = I_{13}(\theta_0) = I_{23}(\theta_0) = 0$. So,

$$I_{\theta_0} = Diag\left[\frac{\mathbb{E}\left[\{2p(t)-1\}^2\right]}{\sigma^2}\ ,\ \frac{1}{\sigma^2}\ ,\ \frac{2}{\sigma^2}\right] \tag{9}$$

where $\mathbb{E}\left[\{2p(t_1)-1\}^2\right] = \mathbb{E}\left[\{2p(t_2)-1\}^2\right] = 1$ and $\forall t \in]t_1, t_2[$:

$$\mathbb{E}\left[\{2p(t)-1\}^2\right] = \bar{r}(t_1,t_2)\left(2Q_t^{1,1}-1\right)^2 + r(t_1,t_2)\left(2Q_t^{1,-1}-1\right)^2 \tag{10}$$

Indeed, $\forall t \in]t_1, t_2[$:

$$\begin{aligned}\mathbb{E}\left[\{2p(t)-1\}^2\right] &= 2\left\{\left(Q_t^{1,1}\right)^2\ \bar{r}(t_1,t_2) + \left(Q_t^{1,-1}\right)^2\ r(t_1,t_2)\right\} \\ &\quad + 2\left\{\left(Q_t^{-1,1}\right)^2\ r(t_1,t_2) + \left(Q_t^{-1,-1}\right)^2\ \bar{r}(t_1,t_2)\right\} - 1\end{aligned}$$

As $Q_t^{-1,1} = 1 - Q_t^{1,1}$, $Q_t^{-1,1} = 1 - Q_t^{1,-1}$ and $\bar{r}(t_1,t_2) + r(t_1,t_2) = 1$, we obtain formula (10).
$\mathbb{E}\left[\{2p(t)-1\}^2\right]$ is always different from zero since the parameter t is bounded. It comes $\forall t \in [t_1, t_2]$:

$$\Lambda_n(t) = \left[\sum_{j=1}^{n} \frac{(y_j-\mu)\ \{2p_j(t)-1\}}{\sigma\ \sqrt{n}\ \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}}\right]^2 + o_{P_{\theta_0}}(1) \tag{11}$$

By convention, the notation $o_{P_{\theta_0}}(1)$ is short for a sequence of random vectors that converges to zero in probability under H_0 (i.e. no QTL on the whole interval studied).

Study under H_0 :
Without loss of generality, we assume that $n = 1$ for the moment and we consider the score function :

$$S(t) = \frac{(y-\mu)\ \{2p(t)-1\}}{\sigma\ \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}} = \frac{y-\mu}{\sigma}\ h(t)$$

where the fact $h(.)$ is a random process independent of y.
It is easy to see that :

$$\mathbb{E}\left\{S(t)\right\} = 0 \ \ , \ \ \mathbb{V}\left\{S(t)\right\} = \mathbb{E}\left[\left\{h(t)\right\}^2\right] = 1$$

$\forall (t,t') \in [t_1,t_2]^2$:

$$\begin{aligned}\Gamma(t,t') := \mathrm{Cov}\left\{S(t),\ S(t')\right\} = \mathbb{E}\left\{h(t)h(t')\right\} &= \frac{\mathbb{E}\left[\{2p(t)-1\}\ \{2p(t')-1\}\right]}{\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}\sqrt{\mathbb{E}\left[\{2p(t')-1\}^2\right]}} \\ &= \frac{4\mathbb{E}\left\{p(t)p(t')\right\}-1}{\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}\sqrt{\mathbb{E}\left[\{2p(t')-1\}^2\right]}}\end{aligned} \tag{12}$$

The formula for $\mathbb{E}\left\{p(t)p(t')\right\}$ is given in appendix 8.1. As $|p(t)p(t')| \leqslant 1$, by dominated convergence theorem, $\mathbb{E}\left\{p(t)p(t')\right\}$ is continuous at (t_1,t'), (t_2,t') and (t_1,t_2). Then the covariance function is continuous at this points (because the denominator is also continuous). So, the covariance function is a continuous function on $[t_1,t_2]^2$.
Let $S_n(.)$ be the score process for n observations :

$$S_n(t) = \sum_{j=1}^{n} \frac{(y_j-\mu)\ (2p_j(t)-1)}{\sigma\ \sqrt{n}\ \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}} \tag{13}$$

When n tends to infinity, an application of the Multivariate Central Limit Theorem shows that for $0 \leqslant s_1 < s_2 < ... < s_d \leqslant T$:

$$(S_n(s_1),...,S_n(s_d))' \xrightarrow{\mathcal{L}} N(\underset{\sim}{0},\Sigma)$$

were Σ is the variance covariance matrix, with unit variance and covariance given by formula (12). $\underset{\sim}{0}$ is a column vector of zeros. As $\Lambda_n(t) = S_n^2(t) + o_{P_{\theta_0}}(1)$:

$$(\Lambda_n(s_1),...,\Lambda_n(s_d))' \xrightarrow{\mathcal{L}} \left\{N(\underset{\sim}{0},\Sigma)\right\}^2$$

Study under $H_{at^\star}$:
In this part, we set

$$Y_j = \mu\ +\ \frac{a}{\sqrt{n}}\ X_j(t^\star)\ +\ \sigma\ \varepsilon_j \tag{14}$$

where ε_j is a Gaussian white noise. According to formula (11), $\forall t \in [t_1,t_2]$:

$$\Lambda_n(t) = \left\{S_n(t)\right\}^2\ +\ o_{P_{\theta_0}}(1)$$

We remind that $o_{P_{\theta_0}}(1)$ is short for a sequence of random vectors that converges to zero in probability under H_0 (i.e. no QTL on the whole interval studied). Let $o_{P_{\theta_{0,t^\star}}}(1)$ be a

sequence of random vectors that converges to zeros if there is no QTL at position $t^\star$. Then, it is clear that :

$$\Lambda_n(t) = \{S_n(t)\}^2 \ + \ o_{P_{\theta_{0,t^\star}}}(1)$$

Let $\theta_{a,t^\star}$ be the parameter refering that we are under $H_{at^\star}$. Under $H_{at^\star}$, as the QTL is located at position $t^\star$, the density of $Y\big|X(t_0), X(t_1)$ verifies :

$$p(t^\star) f_{(\mu+q,\sigma)}(y) + \{1 - p(t^\star)\} f_{(\mu-q,\sigma)}(y)$$

Let Q_n and P_n two sequences of probability measures defined on the same space $(\Omega_n, \ \mathcal{A}_n)$. Q_n (respectively P_n) is the law corresponding to the density $L_n(\theta_{a,t^\star}, t^\star)$ (resp $L_n(\theta_0, t^\star)$). We will call the log likelihood ratio $\log \frac{dQ_n}{dP_n}$. It verifies : $\log \frac{dQ_n}{dP_n} = \log \left\{ \frac{L_n(\theta_{a,t^\star}, \ t^\star)}{L_n(\theta_0, \ t^\star)} \right\}$.

Notations : $Q_n \triangleleft P_n$ will mean the sequence Q_n is contiguous with the respect to the sequence P_n.

Let $b = (a, \ 0, \ 0)'$. As the model is differentiable in quadratic mean at $\theta_{a,t^\star}$:

$$\log\left(\frac{dQ_n}{dP_n}\right) = \frac{b'}{\sqrt{n}} \nabla \log L_n(\theta_0, t^\star) \ - \frac{1}{2} b' I_{\theta_0} b \ + \ o_{P_{\theta_0}}(1)$$

Then, by the central limit theorem :

$$\log\left(\frac{dQ_n}{dP_n}\right) \xrightarrow[H_0]{\mathcal{L}} N(-\frac{1}{2}\nu^2, \ \nu^2) \ \text{ with } \ \nu^2 = \frac{a^2}{\sigma^2} \, \mathbb{E}\left[\{2p(t^\star) - 1\}^2\right]$$

So, if we switch P_n and Q_n, by the ii) of Le Cam's first lemma, we have $P_n \triangleright Q_n$.

Up to now $\forall t \in [t_1, t_2]$:

$$\Lambda_n(t) = \{S_n(t)\}^2 \ + \ o_{P_{\theta_{0,t^\star}}}(1)$$

As $Q_n \triangleleft P_n$, according to iv) of Le Cam's first lemma :

$$\Lambda_n(t) = \{S_n(t)\}^2 \ + \ o_{P_{\theta_{a,t^\star}}}(1)$$

So, calculations can be done with the score process. According to formula (13) and (14), we have :

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \varepsilon_j h_j(t) \ + \ \sum_{j=1}^n \frac{a}{\sigma n} X_j(t^\star) h_j(t) = S_n^0(t) \ + \ \sum_{j=1}^n \frac{a}{\sigma n} X_j(t^\star) h_j(t)$$

where $h_j(.)$ is the equivalent of the process $h(.)$ defined above but for the individual j. $S_n^0(.)$ is the process obtained under H_0.
By the law of large number :

$$\frac{1}{n} \sum_{j=1}^n X_j(t^\star) h_j(t) \to \mathbb{E}\left\{X(t^\star) h(t)\right\}$$

Let suppose $K = 2$ for the moment and, for example $(t, t^\star) \in]t_1, t_2[^2$. Let us compute $\mathbb{E}\left[X(t^\star)\ \{2p(t)-1\}\right]$. We condition on $X(t_1)$ and $X(t_2)$. Consider, for example, the case $X(t_1) = X(t_2) = 1$. In this case, $p(t) = Q_t^{1,1}$ and we have :

$$\begin{aligned}
&\mathbb{E}\left[\ X(t^\star)\ \{2p(t)-1\}\ \middle|\ X(t_1) = X(t_2) = 1\right] = \mathbb{E}\left[\ X(t^\star)\ \left\{2Q_t^{1,1}-1\right\}\ \middle|\ X(t_1) = X(t_2) = 1\right] \\
&= \left\{2Q_t^{1,1}-1\right\}\ \mathbb{E}\left[\ X(t^\star)\ \middle|\ X(t_1) = X(t_2) = 1\right] \\
&= \left\{2Q_t^{1,1}-1\right\}\ \left\{\frac{\bar{r}(t_1,t^\star)\ \bar{r}(t^\star,t_2)}{\bar{r}(t_1,t_2)}\ -\ \frac{r(t_1,t^\star)r(t^\star,t_2)}{\bar{r}(t_1,t_2)}\right\} \\
&= \left\{2Q_t^{1,1}-1\right\}\ \left\{Q_{t^\star}^{1,1} - Q_{t^\star}^{-1,-1}\right\} = \left\{2Q_t^{1,1}-1\right\}\ \left\{2Q_{t^\star}^{1,1}-1\right\}
\end{aligned}$$

Considering the four cases :

$$\begin{aligned}
&\mathbb{E}\left[\ X(t^\star)\ \{2p(t)-1\}\right] \\
&= \left\{2Q_t^{1,1}-1\right\}\ \left\{2Q_{t^\star}^{1,1}-1\right\}\ \frac{1}{2}\ \bar{r}(t_1,t_2)\ +\ \left\{2Q_t^{1,-1}-1\right\}\ \left\{2Q_{t^\star}^{1,-1}-1\right\}\ \frac{1}{2}\ r(t_1,t_2) \\
&+ \left\{2Q_t^{-1,1}-1\right\}\ \left\{2Q_{t^\star}^{-1,1}-1\right\}\ \frac{1}{2}\ r(t_1,t_2)\ +\ \left\{2Q_t^{-1,-1}-1\right\}\ \left\{2Q_{t^\star}^{-1,-1}-1\right\}\ \frac{1}{2}\ \bar{r}(t_1,t_2) \\
&= \bar{r}(t_1,t_2)\ \left\{2Q_{t^\star}^{1,1}-1\right\}\ \left\{2Q_t^{1,1}-1\right\} \\
&+ r(t_1,t_2)\ \left\{2Q_{t^\star}^{1,-1}-1\right\}\ \left\{2Q_t^{1,-1}-1\right\}
\end{aligned} \tag{15}$$

According to dominated convergence theorem, $\mathbb{E}\left[\ X(t^\star)\ \{2p(t)-1\}\right]$ is continuous on $[t_1, t_2]^2$. As a conclusion, $\forall (t, t^\star) \in [t_1, t_2]^2$:

$$m_{t^\star}(t) = \frac{a\ \mathbb{E}\left[X(t^\star)\ \{2p(t)-1\}\right]}{\sigma\ \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}}$$

A non linear interpolation

After some easy calculations, we can remark that :

$$S_n(t) = \{\ \alpha(t)\ S_n(t_1)\ +\ \beta(t)\ S_n(t_2)\ \} / \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}$$

where $\mathrm{Cov}\left\{S_n^0(t_1), S_n^0(t_2)\right\} = e^{-2t_2}$, $\alpha(t_1) = 1$, $\beta(t_1) = 0$, $\alpha(t_2) = 0$, $\beta(t_2) = 1$ and $\forall t \in]t_1, t_2[$:

$$\alpha(t) = Q_t^{1,1} + Q_t^{1,-1} - 1 \ \text{ and } \ \beta(t) = Q_t^{1,1} - Q_t^{1,-1}$$

And it comes :

$$m_{t^\star}(t) = \{\ \alpha(t)\ m_{t^\star}(t_1)\ +\ \beta(t)\ m_{t^\star}(t_2)\ \} / \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}$$

Weak convergence of the score process

As $p(t)$ and $\mathbb{E}\left[\{2p(t)-1\}^2\right]$ are continuous functions, each trajectory of the process $S_n(.)$ is

a continuous function on $[0,T]$. Let define the modulus of continuity of a continous function x on $[0,T]$:

$$w_x(\delta) = sup_{|t'-t|<\delta} |x(t') - x(t)| \quad \text{where} \quad 0 < \delta \leqslant T$$

According to theorem 8.2 of Billingsley (1999), the score process is tight if and only if the two following conditions hold :

(a) the sequence $S_n(0)$ is tight.
(b) For each positive ϵ and η, there exist a δ, with $0 < \delta < T$, and an integer n_0 such that $\mathbb{P}\{w_{S_n}(\delta) \geqslant \eta\} \leqslant \epsilon \quad \forall n \geqslant n_0$.

According to Prohorov, the sequence $S_n(0)$ is tight. So, a) is verified.
Let define the functions $\tilde{\alpha}(t)$ and $\tilde{\beta}(t)$ such as :

$$\tilde{\alpha}(t) = \alpha(t)/\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]} \ , \ \tilde{\beta}(t) = \beta(t)/\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}$$

First, we can remark that $\forall\delta$ such as $0 < \delta \leqslant T$:

$$\begin{aligned} w_{S_n}(\delta) &= sup_{|t'-t|<\delta} |S_n(t') - S_n(t)| \\ &= sup_{|t'-t|<\delta} \left| \{\tilde{\alpha}(t') - \tilde{\alpha}(t)\} S_n(t_1) \ + \ \left\{\tilde{\beta}(t') - \tilde{\beta}(t)\right\} S_n(t_2) \right| \\ &\leqslant max\{|S_n(t_1)|, |S_n(t_2)|\} \ \left\{w_{\tilde{\alpha}}(\delta) + w_{\tilde{\beta}}(\delta)\right\} \end{aligned}$$

Let $\epsilon > 0$ and $\eta > 0$, as the sequence $max\{|S_n(t_1)|, |S_n(t_2)|\}$ is uniformly tight, $\exists M$ such as $\forall n \geqslant 1 \quad \mathbb{P}[\, max\{|S_n(t_1)|, |S_n(t_2)|\} \geqslant M \,] \leqslant \epsilon$.
It comes, $\mathbb{P}\left[\, max\{|S_n(t_1)|, |S_n(t_2)|\} \left\{w_{\tilde{\alpha}}(\delta) + w_{\tilde{\beta}}(\delta)\right\} \geqslant M \left\{w_{\tilde{\alpha}}(\delta) + w_{\tilde{\beta}}(\delta)\right\} \,\right] \leqslant \epsilon$
As $\tilde{\alpha}(t)$ and $\tilde{\beta}(t)$ are continuous on the compact $[0,T]$, according to Heine's theorem, these functions are uniformly continuous. So, let $\upsilon > 0$, $\exists\delta_1$ with $0 < \delta_1 < T$, such as $w_{\tilde{\alpha}}(\delta_1) < \upsilon/2$ and $\exists\delta_2$ with $0 < \delta_2 < T$ such as $w_{\tilde{\beta}}(\delta_2) < \upsilon/2$. Let $\delta = min(\delta_1, \delta_2)$ then $w_{\tilde{\alpha}}(\delta) + w_{\tilde{\beta}}(\delta) < \upsilon$. If we impose $\upsilon = \eta/M$, then $\forall n \geqslant 1$, $\mathbb{P}\{w_{S_n}(\delta) \geqslant \eta\} \leqslant \epsilon$ which means b) of theorem 8.2 of Billingsley (1999) is fulfilled. So, the tightness of the score process is proved.

To conclude, the tightness and the convergence of finite-dimensional imply the weak convergence of the score process.

7.2. Proof of theorem 3

We first compute the score functions and the Fisher Information matrix. Let $t \in [t_1, t_K] \backslash \mathbb{T}_k$:

$$\frac{\partial logL}{\partial q_i}|_{\theta_0} = \frac{y-\mu_i}{\sigma^2} \ \{2p(t)-1\} \ 1_{C=i} \quad , \quad \frac{\partial logL}{\partial \mu_i}|_{\theta_0} = \frac{y-\mu_i}{\sigma^2} \ 1_{C=i}$$

$$\frac{\partial logL}{\partial \sigma}|_{\theta_0} = -\frac{1}{\sigma} \ + \ \sum_{i=1}^{I} \frac{(y-\mu_i)^2}{\sigma^3} \ 1_{C=i}$$

$$I_{\theta_0} = Diag\left[\frac{\pi_1}{\sigma^2}\ \mathbb{E}\left[\{2p(t)-1\}^2\right],\ ...,\ \frac{\pi_I}{\sigma^2}\ \mathbb{E}\left[\{2p(t)-1\}^2\right],\ \frac{\pi_1}{\sigma^2}\ ,\ ...\ ,\ \frac{\pi_I}{\sigma^2}\ ,\ \frac{2}{\sigma^2}\right]$$

Remarks : If the π_i's were unknown, after some easy calculations, we find that the Fisher information matrix would still be diagonal, it would be the same as above, and the diagonal terms concerning π_i would be equal to $\frac{1}{\pi_i}$. So, the results established further are also true for all the π_i's unknown.

It comes $\forall t \in [t_1, t_K]\backslash\mathbb{T}_k$:

$$\Lambda_n(t) = \sum_{i=1}^{I}\left[\sum_{j=1}^{n}\frac{(y_j-\mu_i)\ \{2p_j(t)-1\}}{\sqrt{n\ \pi_i}\ \sigma\ \sqrt{\mathbb{E}\left\{\{2p(t)-1\}^2\right\}}}\ 1_{C_j=i}\right]^2 \ +\ o_{P_{\theta_0}}(1) \tag{16}$$

Study under H_0 :
Without loss of generality, we assume $n = 1$ for the moment and we consider the score test statistic refering to the test of the presence of a QTL in family i :

$$S(t,i) = \frac{(y-\mu_i)\ \{2p(t)-1\}}{\sqrt{\pi_i}\ \sigma\ \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}}\ 1_{C=i} = \frac{y-\mu_i}{\sigma\ \sqrt{\pi_i}}\ 1_{C=i}\ \ h(t)$$

where $h(.)$ is a random process (the same as in the proof of theorem 1), independent of y and C. It is easy to see that :

$$\mathbb{E}\left\{S(t,i)\right\} = 0\ \ ,\ \ \mathbb{V}\left\{S(t,i)\right\} = \mathbb{E}\left[\{h(t)\}^2\right] = 1$$

$\forall (t,t') \in [t_1, t_K]\backslash\mathbb{T}_k \times [t_1, t_K]\backslash\mathbb{T}_k$:

$$\mathrm{Cov}\left\{S(t,i),\ S(t',i)\right\} = \mathbb{E}\left\{h(t)h(t')\right\} = \Gamma(t,t')$$

This function $\Gamma(t,t')$ is the same as in theorem 2.
Let $S_n(.,i)$ be the score process for n observations, related to testing the presence of a QTL in family i on $[0,T]$:

$$S_n(t,i) = \sum_{j=1}^{n}\frac{(y_j-\mu_i)\ \{2p_j(t)-1\}}{\sqrt{n\ \pi_i}\ \sigma\ \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}}\ 1_{C_j=i} \tag{17}$$

When n tends to infinity, an application of the Multivariate Central Limit Theorem shows that for $0 \leqslant s_1 < s_2 < ... < s_d \leqslant T$:

$$(S_n(s_1,i), ..., S_n(s_d,i))' \xrightarrow{\mathcal{L}} N(\underset{\sim}{0}, \Sigma)$$

were Σ is the variance covariance matrix, with unit variance and covariance given by the function $\Gamma(t,t')$. $\underset{\sim}{0}$ is a column vector of zeros.

According to formula (16), we have $\Lambda_n(t) = \sum_{i=1}^{I}\{S_n(t,i)\}^2 + o_{P_{\theta_0}}(1)$. It comes :

$$(\Lambda_n(s_1), ..., \Lambda_n(s_d))' \xrightarrow{\mathcal{L}} \sum_{i=1}^{I}\left\{N(\underset{\sim}{0}, \Sigma)\right\}^2$$

Study under $H_{at^\star}$:
In this part, we set

$$Y_j = \mu_i \ + \ \frac{a_i}{\sqrt{n}} \, X_j(t^\star) \ + \ \sigma \, \varepsilon_j \tag{18}$$

where ε_j is a Gaussian white noise.
It is clear that we have also :

$$\Lambda_n(t) = \sum_{i=1}^{I} \{S_n(t,i)\}^2 \ + \ o_{P_{\theta_{0,t^\star}}}(1) \tag{19}$$

Under $H_{at^\star}$, as the QTL is located at position $t^\star$, the density of $Y \big| X(t^\ell), X(t^r), C = i$ verifies :

$$p(t^\star) f_{(\mu_i + q_i, \sigma)}(y) + \{1 - p(t^\star)\} \, f_{(\mu_i - q_i, \sigma)}(y)$$

By the central limit theorem :

$$log\left(\frac{dQ_n}{dP_n}\right) \xrightarrow[H_0]{\mathcal{L}} N(-\frac{1}{2}\nu^2, \ \nu^2) \ \text{ with } \ \nu^2 = \sum_{i=1}^{I} \frac{a_i^2 \, \pi_i}{\sigma^2} \, \mathbb{E}\left[\{2p(t^\star) - 1\}^2\right]$$

If we switch P_n and Q_n, by the ii) of Le Cam's first lemma, we have $P_n \triangleright Q_n$. According to iv) of Le Cam's first lemma and formula (19) :

$$\Lambda_n(t) = \sum_{i=1}^{I} \{S_n(t,i)\}^2 \ + \ o_{P_{\theta_{a,t^\star}}}(1)$$

So, calculations can be done with the score process. According to formula (17) and (18), we have :

$$\begin{aligned} S_n(t,i) &= \frac{1}{\sqrt{n \, \pi_i}} \sum_{j=1}^{n} \varepsilon_j \ 1_{C_j = i} \ h_j(t) \ + \ \sum_{j=1}^{n} \frac{a_i}{n \, \sigma \, \sqrt{\pi_i}} \ 1_{C_j = i} \ X_j(t^\star) \ h_j(t) \\ &= S_n^0(t,i) \ + \ \sum_{j=1}^{n} \frac{a_i}{n \, \sigma \, \sqrt{\pi_i}} \ 1_{C_j = i} \ X_j(t^\star) \ h_j(t) \end{aligned}$$

where $S_n^0(.,i)$ is the process obtained under H_0. We remind that $h_j(.)$ is the equivalent of the process $h(.)$ for the individual j. By the law of large number :

$$\frac{1}{n} \sum_{j=1}^{n} X_j(t^\star) \ h_j(t) \ 1_{C_j = i} \to \pi_i \, \mathbb{E}\left\{X(t^\star) h(t)\right\}$$

It comes : $\forall (t, t^\star) \in [t_1, t_K] \backslash \mathbb{T}_k \times [t_1, t_K] \backslash \mathbb{T}_k$:

$$m_{t^\star}^i(t) = \frac{a_i \ \sqrt{\pi_i} \ \mathbb{E}\left[X(t^\star)\{2p(t) - 1\}\right]}{\sigma \ \sqrt{\mathbb{E}\left[\{2p(t) - 1\}^2\right]}}$$

A non linear interpolation :
We can easily remark that $\forall t \in [t_1, t_K] \backslash \mathbb{T}_k$:

$$S_n(t,i) = \alpha(t)\ S_n(t^\ell, i)\ +\ \beta(t)\ S_n(t^r, i) / \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}$$

where $\alpha(t) = Q_t^{1,1} + Q_t^{1,-1} - 1$, $\beta(t) = Q_t^{1,1} - Q_t^{1,-1}$ and $\forall k\ \forall k'$, Cov $\left\{S_n^0(t_k, i), S_n^0(t_{k'}, i)\right\} = e^{-2|t_k - t_{k'}|}$.
It comes $\forall (t, t^\star) \in [t_1, t_K] \backslash \mathbb{T}_k \times [t_1, t_K] \backslash \mathbb{T}_k$:

$$m_{t^\star}^i(t) = \left\{\ \alpha(t)\ m_{t^\star}^i(t^\ell)\ +\ \beta(t)\ m_{t^\star}^i(t^r)\ \right\} / \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}$$

7.3. Proof of theorem 4

Study under H_0 :
$\Lambda_n(.)$ is the same process as in theorem 2.

Study under $H_{a\vec{t}^\star}$:
In this part, we set

$$Y_j = \mu\ +\ \sum_{s=1}^{m} X_j(t_s^\star)\ q^s\ +\ \sigma\varepsilon_j \tag{20}$$

where ε_j is a Gaussian white noise.
Let's introduce some notations :

- ξ : number of "Marker intervals" which contain the QTL.
  $\gamma = 1, ..., \xi$ will refer to the different intervals.

- m_γ : number of QTL in the interval γ.
  $\tau = 1, ..., m_\gamma$ refers to the τth QTL in the interval γ.

- the sth QTL on $[0, T]$, can be rewritten, $s = (\tau, \gamma) = \left\{\sum_{i=1}^{\gamma-1} m_i\right\}\ +\ \tau$

Let $\theta_{a,\vec{t}^\star} = \left(q^1,\ ...,\ q^m,\ \mu,\ \sigma\right)$ and $\theta_{0,\vec{t}^\star} = (0,\ ...,\ 0,\ \mu,\ \sigma)$.
After some calculations, the likelihood of $\left(Y,\ X\left\{t_{(1,1)}^{\star\ell}\right\},\ X\left\{t_{(1,1)}^{\star r}\right\},\ ...,\ X\left\{t_{(1,\xi)}^{\star\ell}\right\},\ X\left\{t_{(1,\xi)}^{\star r}\right\}\right)$ with respect to the measure $\lambda \otimes N \otimes ... \otimes N$, λ being the Lebesgue measure, N the county measure on $\mathbb{N}$, verifies :

$$L^\star(\theta_{a,\vec{t}^\star}) = \sum_{(u_1,...,u_m)\in\{-1,1\}^m} f_{(\mu+u_1q^1+...+u_mq^m,\sigma)}(y)$$
$$\times \left\{\left(\prod_{\gamma=1}^{\xi} A\left\{t_{(\tau,\gamma)}^{\star\ell}\ ,\ t_{(\tau,\gamma)}^\star\right\}\ \left[\prod_{\tau=1}^{m_\gamma-1} R\left\{t_{(\tau,\gamma)}^\star\ ,\ t_{(\tau+1,\gamma)}^\star\right\}\right]\ A\left\{t_{(m_\gamma,\gamma)}^{\star r}\ ,\ t_{(m_\gamma,\gamma)}^\star\right\}\ \right) g^\star(\vec{t}^\star)\right\}$$

where

$$
\begin{aligned}
u_s &= u_{(\tau,\gamma)} \\
A\left\{t\ ,\ t^\star_{(\tau,\gamma)}\right\} &= r\left\{t\ ,\ t^\star_{(\tau,\gamma)}\right\}\ 1_{X(t)u(\tau,\gamma)=-1}\ +\ \bar{r}\left\{t\ ,\ t^\star_{(\tau,\gamma)}\right\}\ 1_{X(t)u(\tau,\gamma)=1} \\
R\left\{t^\star_{(\tau,\gamma)}\ ,\ t^\star_{(\tau+1,\gamma)}\right\} &= \bar{r}\left\{t^\star_{(\tau,\gamma)}\ ,\ t^\star_{(\tau+1,\gamma)}\right\}\ 1_{u(\tau,\gamma)u(\tau+1,\gamma)=1} \\
&\quad + r\left\{t^\star_{(\tau,\gamma)}\ ,\ t^\star_{(\tau+1,\gamma)}\right\}\ 1_{u(\tau,\gamma)u(\tau+1,\gamma)=-1} \\
g^\star(\vec{t}^\star) &= \frac{1}{2}\ \prod_{\gamma=1}^{\xi-1}\ D\left\{t^{\star r}_{(m_\gamma,\gamma)}, t^{\star \ell}_{(1,\gamma+1)}\right\} \\
D(t,t') &= \bar{r}(t,t')\ 1_{X(t)X(t')=1}\ +\ r(t,t')\ 1_{X(t)X(t')=-1}
\end{aligned}
$$

The likelihood $L^\star_n(\theta_{a,\vec{t}^\star})$ for n observations is obtained by the product of n terms as above. Let Q_n and P_n two sequences of probability measures defined on the same space $(\Omega_n,\ \mathcal{A}_n)$. Q_n (respectively P_n) is the law corresponding to the density $L^\star_n(\theta_{a,\vec{t}^\star})$ (resp $L^\star_n(\theta_{0,\vec{t}^\star})$). We will call the log likelihood ratio $\log \frac{dQ_n}{dP_n}$. It verifies : $\log \frac{dQ_n}{dP_n} = \log\left\{\frac{L^\star_n(\theta_{a,\vec{t}^\star})}{L^\star_n(\theta_{0,\vec{t}^\star})}\right\}$.
As the model is differentiable in quadratic mean at $\theta_{a,\vec{t}^\star}$ and according to the central limit theorem :

$$log\left(\frac{dQ_n}{dP_n}\right) \overset{H_0}{\to} N(-\frac{1}{2}\vartheta^2,\ \vartheta^2)\ \text{ with } \vartheta^2 \in \mathbb{R}^{+\star}$$

If we switch P_n and Q_n, by the ii) of Le Cam's first lemma, we have $P_n \triangleright Q_n$.
We remind the notation $o_{P_{\theta_0}}(1)$ used in the proof of theorem 1 : $o_{P_{\theta_0}}(1)$ is short for a sequence of random vectors that converges to zeros in probability under H_0 (i.e. no QTL on the whole interval studied).
Besides, we remind the score test statistic on genetic markers for n observations. $\forall t \in [t_1, t_K]\backslash\mathbb{T}_k$, we have :

$$S_n(t) = \sum_{j=1}^{n} \frac{(y_j-\mu)\ (2\ p_j(t)-1)}{\sigma\ \sqrt{n}\ \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}}$$

According to the proof of theorem 1 :

$$\Lambda_n(t) = \{S_n(t)\}^2\ +\ o_{P_{\theta_0}}(1)$$

Let $o_{P_{\theta_{0,\vec{t}^\star}}}(1)$ be a sequence of random vectors that converges to zeros if there is no QTL at $t^\star_1$, ..., $t^\star_m$. Then, it is clear that :

$$\Lambda_n(t) = \{S_n(t)\}^2\ +\ o_{P_{\theta_{0,\vec{t}^\star}}}(1)$$

As $Q_n \triangleleft P_n$, according to iv) of Le Cam's first lemma :

$$\Lambda_n(t) = \{S_n(t)\}^2\ +\ o_{P_{\theta_{a,\vec{t}^\star}}}(1)$$

So, calculations can be done with the score process.
According to formula (20) :

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \varepsilon_j h_j(t) \; + \; \sum_{j=1}^{n} \left\{ \sum_{s=1}^{m} X_j(t_s^\star) \; a^s \right\} \frac{h_j(t)}{\sigma n}$$
$$= S_n^0(t) \; + \; \sum_{j=1}^{n} \left\{ \sum_{s=1}^{m} X_j(t_s^\star) \; a^s \right\} \frac{h_j(t)}{\sigma n}$$

where $h_j(t)$ is the same function as in the proof of theorem 3.
By the law of large number :

$$\frac{1}{n} \sum_{j=1}^{n} \left\{ \sum_{s=1}^{m} X_j(t_s^\star) \; a^s \right\} \; h_j(t) \to \mathbb{E} \left[\left\{ \sum_{s=1}^{m} X(t_s^\star) \; a^s \right\} h(t) \right]$$

We have :

$$\mathbb{E} \left[\left\{ \sum_{s=1}^{m} X(t_s^\star) \; a^s \right\} h(t) \right] = \sum_{s=1}^{m} \frac{a^s \; \mathbb{E} \left[X(t_s^\star) \; \{2p(t) - 1\} \right]}{\sqrt{\mathbb{E} \left[\{2p(t) - 1\}^2 \right]}}$$

According to formula (15) in Section 7.1 and formula (22) in appendix 8.2 :

$$\begin{aligned} &\mathbb{E} \left[X(t_s^\star) \; \{2p(t) - 1\} \right] \\ &= \left\{ \bar{r}(t^\ell, t^r) \left(2Q_{t_s^\star}^{1,1} - 1 \right) \left(2Q_t^{1,1} - 1 \right) \; + \; r(t^\ell, t^r) \left(2Q_{t_s^\star}^{1,-1} - 1 \right) \left(2Q_t^{1,-1} - 1 \right) \right\} \; 1_{t_s^\star \in]t^\ell, t^r[} \\ &+ e^{-2|t - t_s^\star|} \; 1_{t_s^\star \notin]t^\ell, t^r[} \end{aligned} \tag{21}$$

It comes $\forall (t, t_1^\star, ..., t_m^\star) \in \{[t_1, t_K] \backslash \mathbb{T}_k\}^{m+1}$:

$$m_{\vec{t}^\star}(t) = \sum_{s=1}^{m} \frac{a^s \; \mathbb{E} \left[X(t_s^\star) \; \{2p(t) - 1\} \right]}{\sigma \sqrt{\mathbb{E} \left[\{2p(t) - 1\}^2 \right]}}$$

A non linear interpolation :
The process $S_n(.)$ is a non linear interpolation according to formula (23) of appendix 8.2.
It comes $\forall (t, t_1^\star, ..., t_m^\star) \in \{[t_1, t_K] \backslash \mathbb{T}_k\}^{m+1}$:

$$m_{\vec{t}^\star}(t) = \sum_{s=1}^{m} \frac{a^s}{\sigma} \left\{ \; \alpha(t) \; \mathbb{E} \left[X(t_s^\star) \; \left\{ 2p(t^\ell) - 1 \right\} \right] \; + \; \beta(t) \; \mathbb{E} \left[X(t_s^\star) \; \{2p(t^r) - 1\} \right] \; \right\} \Big/ \sqrt{\mathbb{E} \left[\{2p(t) - 1\}^2 \right]}$$

Weak convergence of the score process :
Concerning the weak convergence of the score process, the proof is the same as in the proof of theorem 2 in appendix 8.2.

8. Appendix

8.1. Formula for $\mathbb{E}\{p(t)p(t')\}$

$\forall (t,t') \in]t_1,t_2[^2$:

$$\mathbb{E}\{p(t)p(t'\} = \frac{1}{2}\left\{Q_t^{1,1}Q_{t'}^{1,1}\ \bar{r}(t_1,t_2) + Q_t^{1,-1}Q_{t'}^{1,-1}\ r(t_1,t_2)\right\} + \frac{1}{2}\left\{Q_t^{-1,1}Q_{t'}^{-1,1}\ r(t_1,t_2) + Q_t^{-1,-1}Q_{t'}^{-1,-1}\ \bar{r}(t_1,t_2)\right\}$$

This quantity is continuous at t_1 and t_2 (cf. proof of theorem 1 in Section 7.1)

8.2. Sketch of the proof of theorem 2

Let $t \in [t_1,t_K]\backslash\mathbb{T}_k$. As t belongs to the "Marker interval" (t^ℓ,t^r), some adjustments with Section 3 have to be done : t_1 becomes t^ℓ and t_2 becomes t^r. So, $p(t)$ is now the quantity equal to $\mathbb{P}\left\{X(t)=1\middle|X(t^\ell),X(t^r)\right\}$. In the same way, $p(t)$, $Q_t^{1,1}$, $Q_t^{1,-1}$, $Q_t^{-1,1}$ and $Q_t^{-1,-1}$ described in formula (2) have to be adapted to the "Marker interval". The likelihood presented in formula (3), is unchanged except that the focus is on the triplet $\big(Y,\ X(t^\ell),\ X(t^r)\big)$ and the function $g(t)$ has to be adapted to the "Marker interval". Formula (11) of Section 7.1 is also suitable $t \in [t_1,t_K]\backslash\mathbb{T}_k$ because t is bounded. It comes, $\forall (t,t') \in [t_1,t_K]\backslash\mathbb{T}_k \times [t_1,t_K]\backslash\mathbb{T}_k$:

$$\Gamma(t,t') = \frac{4\mathbb{E}\{p(t)p(t')\} - 1}{\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}\sqrt{\mathbb{E}\left[\{2p(t')-1\}^2\right]}}$$

$\mathbb{E}\left[\{2p(t)-1\}^2\right]$ described in formula (10) of Section 7.1 has to be adapted to the "Marker interval".

$\forall (t,t') \in]t^\ell,t^r[^2$, the expression of $\mathbb{E}\{p(t)p(t')\}$ can be deduced from appendix 8.1 by adapting to the "Marker interval".

Besides, if $(t,t') \in]t^\ell,t^r[\times [t^r,t_K]\backslash\mathbb{T}_k$:

$$\begin{aligned}
&\mathbb{E}\{p(t)p(t')\}\\
&= \frac{1}{2}\bar{r}(t^\ell,t^r)\left[Q_{t'}^{1,1}\bar{r}\left\{(t')^\ell,(t')^r\right\} + Q_{t'}^{1,-1}r\left\{(t')^\ell,(t')^r\right\}\right]\left[Q_t^{1,1}\bar{r}\left\{t^r,(t')^\ell\right\} + Q_t^{-1,-1}r\left\{t^r,(t')^\ell\right\}\right]\\
&+ \frac{1}{2}\bar{r}(t^\ell,t^r)\left[Q_{t'}^{-1,1}r\left\{(t')^\ell,(t')^r\right\} + Q_{t'}^{-1,-1}\bar{r}\left\{(t')^\ell,(t')^r\right\}\right]\left[Q_t^{1,1}r\left\{t^r,(t')^\ell\right\} + Q_t^{-1,-1}\bar{r}\left\{t^r,(t')^\ell\right\}\right]\\
&+ \frac{1}{2}r(t^\ell,t^r)\left[Q_{t'}^{1,1}\bar{r}\left\{(t')^\ell,(t')^r\right\} + Q_{t'}^{1,-1}\ r\left\{(t')^\ell,(t')^r\right\}\right]\left[Q_t^{1,-1}r\left\{t^r,(t')^\ell\right\} + Q_t^{-1,1}\bar{r}\left\{t^r,(t')^\ell\right\}\right]\\
&+ \frac{1}{2}r(t^\ell,t^r)\left[Q_{t'}^{-1,1}r\left\{(t')^\ell,(t')^r\right\} + Q_{t'}^{-1,-1}\bar{r}\left\{(t')^\ell,(t')^r\right\}\right]\left[Q_t^{1,-1}\bar{r}\left\{t^r,(t')^\ell\right\} + Q_t^{-1,1}r\left\{t^r,(t')^\ell\right\}\right]
\end{aligned}$$

In the same way as what has been done in the proof of theorem 1 (cf. Section 7.1), $\forall (t,t^\star) \in [t_1,t_K]\backslash\mathbb{T}_k \times [t_1,t_K]\backslash\mathbb{T}_k$:

$$m_{t^\star}(t) = \frac{a\ \mathbb{E}\left[X(t^\star)\{2p(t)-1\}\right]}{\sigma\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}}$$

If $(t,t^\star) \in]t^\ell, t^r[^2$, then $\mathbb{E}\left[X(t^\star)\left\{2p(t)-1\right\}\right]$ has the same expression as in formula (15) of Section 7.1 provided that we adapt to the "Marker interval".
Besides, if $(t,t^\star) \in]t^\ell, t^r[\times [t^r, t_K]\backslash\mathbb{T}_k$:

$$\begin{aligned}
&\mathbb{E}\left[X(t^\star)\left\{2p(t)-1\right\}\right] \\
&= 2\, Q_t^{1,1}\, \mathbb{E}\left\{X(t^\star)1_{X(t^\ell)=1}1_{X(t^r)=1}\right\} \;+\; 2\, Q_t^{1,-1}\, \mathbb{E}\left\{X(t^\star)1_{X(t^\ell)=1}1_{X(t^r)=-1}\right\} \\
&+ 2\, Q_t^{-1,1}\, \mathbb{E}\left\{X(t^\star)1_{X(t^\ell)=-1}1_{X(t^r)=1}\right\} \;+\; 2\, Q_t^{-1,-1}\, \mathbb{E}\left\{X(t^\star)1_{X(t^\ell)=-1}1_{X(t^r)=-1}\right\} \\
&= \bar{r}(t^\ell,t)\,\bar{r}(t,t^r)\;\left\{1-2r(t^r,t^\star)\right\} \;+\; \bar{r}(t^\ell,t)\,r(t,t^r)\;\left\{2r(t^r,t^\star)-1\right\} \\
&+ r(t^\ell,t)\,\bar{r}(t,t^r)\;\left\{1-2r(t^r,t^\star)\right\} \;+\; r(t^\ell,t)\,r(t,t^r)\;\left\{2r(t^r,t^\star)-1\right\} \\
&= \left\{1-2r(t,t^r)\right\}\;\left\{1-2r(t^r,t^\star)\right\} = e^{-2(t^\star-t)}
\end{aligned}$$

As we deal with Poisson processes, it is reversible. So, If $(t,t^\star) \in\; [t^{\star r}, t_K]\backslash\mathbb{T}_k \times\,]t^{\star\ell}, t^{\star r}[$:

$$\mathbb{E}\left[X(t^\star)\left\{2p(t)-1\right\}\right] = \left\{1-2r(t^\star,t^\ell)\right\}\;\left\{1-2r(t^\ell,t)\right\} = e^{-2(t-t^\star)}$$

So, if t and $t^\star$ do not belong to the same "Marker interval" :

$$\mathbb{E}\left[X(t^\star)\left\{2p(t)-1\right\}\right] = e^{-2|t-t^\star|} \tag{22}$$

A non linear interpolation
Concerning the non linear interpolation, we have to adapt formula (5) of Section 3.1 to the "Marker interval". $\forall t \in [t_1, t_K]\backslash\mathbb{T}_k$ we have :

$$S_n(t) = \left\{\; \alpha(t)\; S_n(t^\ell) \;+\; \beta(t)\; S_n(t^r)\;\right\} / \sqrt{\mathbb{E}\left[\left\{2p(t)-1\right\}^2\right]} \tag{23}$$

where $\alpha(t) = Q_t^{1,1} + Q_t^{1,-1} - 1$, $\beta(t) = Q_t^{1,1} - Q_t^{1,-1}$ and $\forall k\; \forall k'$, Cov $\left\{S_n^0(t_k), S_n^0(t_{k'})\right\} = e^{-2|t_k - t_{k'}|}$.
It comes $\forall (t,t^\star) \in [t_1,t_K]\backslash\mathbb{T}_k \times [t_1,t_K]\backslash\mathbb{T}_k$:

$$m_{t^\star}(t) = \left\{\; \alpha(t)\; m_{t^\star}(t^\ell) \;+\; \beta(t)\; m_{t^\star}(t^r)\;\right\} / \sqrt{\mathbb{E}\left[\left\{2p(t)-1\right\}^2\right]}$$

Weak convergence of the score process
Each trajectory of the process $S_n(.)$ is a continuous function on $[0,T]$. In the same way as in the proof of theorem 1 in Section 7.1, in order to prove the tightness of the score process, we have to verify that conditions a) and b) of theorem 8.2 of Billingsley (1999) are fulfilled. According to Prohorov, $S_n(0)$ is tight, so a) is fulfilled.
We remind the modulus of continuity of $S_n(t)$:

$$w_{S_n}(\delta) = sup_{|t'-t|<\delta}\,|S_n(t') - S_n(t)| \quad \text{where} \quad 0 < \delta \leqslant T$$

Let define $w_{S_n}^k(\delta)$, the modulus of continuity of $S_n(t)$ only between the markers k and $k+1$:

$$w_{S_n}^k(\delta) = sup_{|t'-t|<\delta}\,|S_n(t'+t_k) - S_n(t+t_k)| \quad \text{where} \quad 0 < \delta \leqslant t_{k+1} - t_k$$

As the score process is tight when there are only two markers (cf. proof of theorem 1), according to b) of theorem 8.2 of Billingsley (1999), we have for a given k:

$$\forall \epsilon > 0\ \forall \eta > 0\ \exists \delta_k \text{ with } 0 < \delta_k < t_{k+1} - t_k \text{ such that } \mathbb{P}\left\{w^k_{S_n}(\delta_k) \geqslant \eta\right\} \leqslant \epsilon$$

So, let $\epsilon > 0$, $\epsilon' = \epsilon/(K-1)$, $\eta > 0$ and we impose $\delta = \min_{k\in\{1,...,K-1\}}(\delta_k)$
then $\forall k \in \{1, ..., K-1\}$ $\mathbb{P}\left\{w^k_{S_n}(\delta) \geqslant \eta\right\} \leqslant \epsilon'$.
As $w_{S_n}(\delta) \geqslant w^1_{S_n}(\delta) + ... + w^{K-1}_{S_n}(\delta)$, then $\mathbb{P}\left\{w_{S_n}(\delta) \geqslant \eta\right\} \leqslant \sum_{k=1}^{K-1} \mathbb{P}\left\{w^k_{S_n}(\delta) \geqslant \eta\right\} \leqslant \epsilon$ which means b) of theorem 8.2 of Billingsley (1999) is fulfilled. So, the tightness of the score process is proved.

To conclude, the tightness and the convergence of finite-dimensional imply the weak convergence of the score process.

8.3. Linear interpoled process under Interval Mapping

In presence of several markers, the linear interpoled process $V_n(.)$ is such as $\forall t \in [t_1, t_K]\backslash\mathbb{T}_k$:

$$\begin{aligned} V_n(t) &= \left\{\frac{t^r - t}{t^r - t^\ell}\, S_n(t^\ell) \;+\; \frac{t - t^\ell}{t^r - t^\ell}\, S_n(t^r)\right\} / \sqrt{\tau(t)} \\ &= \left\{\frac{t^r - t}{t^r - t^\ell}\, W_n(t^\ell) \;+\; \frac{t - t^\ell}{t^r - t^\ell}\, W_n(t^r)\right\} / \sqrt{\tau(t)} \;+\; o_{P_{\theta_0}}(1) \end{aligned}$$

where

$$\tau(t) = \left(\frac{t^r - t}{t^r - t^\ell}\right)^2 \;+\; 2\,\frac{(t^r - t)(t - t^\ell)}{(t^r - t^\ell)^2}\, e^{-2(t^r - t^\ell)} \;+\; \left(\frac{t - t^\ell}{t^r - t^\ell}\right)^2$$

It can be seen easily that $\tau(t) \neq 0$, $\forall t \in [t_1, t_K]\backslash\mathbb{T}_k$.
$V_n(.)$ remains asymptotically a Gaussian process with mean equal to 0 under H_0, unit variance, and $\forall k\ \forall k'$, $\mathrm{Cov}\left\{S^0_n(t_k), S^0_n(t_{k'})\right\} = e^{-2|t_k - t_{k'}|}$. In the same way as what has been done in Section 3.2, the weights of the model of mixture corresponding to this process verify :

$$p(t) = 1_{X(t^\ell)=1} 1_{X(t^r)=1} \;+\; \frac{t^r - t}{t^r - t^\ell}\, 1_{X(t^\ell)=1} 1_{X(t^r)=-1} \;+\; \frac{t - t^\ell}{t^r - t^\ell}\, 1_{X(t^\ell)=-1} 1_{X(t^r)=1}$$

This weights are an approximation at the first order of the original weights. So, the linear interpoled process will be a good approximation if and only if the genetic markers are close from each other. This process $V_n(.)$ is a generalization of the process studied, under H_0, by Rebaï et al. (1994). By contiguity (in the same way of what has been done in Section 7.1), under $H_{at^\star}$, $V_n(.)$ is asymptotically the same process as under H_0 on which the mean function, $\tilde{m}_{t^\star}(t)$, has been added :

$$\tilde{m}_{t^\star}(t) = \left\{\frac{t^r - t}{t^r - t^\ell}\, m_{t^\star}(t^\ell) \;+\; \frac{t - t^\ell}{t^r - t^\ell}\, m_{t^\star}(t^r)\right\} / \sqrt{\tau(t)}$$

8.4. Linear interpoled process and kriging process in presence of several families

Let $V_n(.,i)$ be the linear interpoled process for family i. $\forall t \in [t_1, t_K] \backslash \mathbb{T}_k$:

$$V_n(t,i) = \left(\frac{t^r - t}{t^r - t^\ell} \, S_n(t^\ell, i) \; + \; \frac{t - t^\ell}{t^r - t^\ell} \, S_n(t^r, i) \right) / \sqrt{\tau(t)}$$
$$= \left(\frac{t^r - t}{t^r - t^\ell} \, W_n(t^\ell, i) \; + \; \frac{t - t^\ell}{t^r - t^\ell} \, W_n(t^r, i) \right) / \sqrt{\tau(t)} \; + \; o_{P_{\theta_0}}(1)$$

where $\tau(t)$ has the same expression as in appendix 8.3 and $\forall k \; \forall k'$, $\mathrm{Cov}\left\{S_n^0(t_k, i), S_n^0(t_{k'}, i)\right\} = e^{-2|t_k - t_{k'}|}$.
$\sum_{i=1}^{I} \{V_n(.,i)\}^2$ is an approximation of the process $\Lambda_n(.)$ provided that the genetic markers are close from each other.
By contiguity (cf. Section 7.2), under $H_{at^\star}$, $V_n(.,i)$ is the same process as under H_0 on which the mean function, $\tilde{m}^i_{t^\star}(t)$, has been added :

$$\tilde{m}^i_{t^\star}(t) = \left\{ \frac{t^r - t}{t^r - t^\ell} \, m^i_{t^\star}(t^\ell) \; + \; \frac{t - t^\ell}{t^r - t^\ell} \, m^i_{t^\star}(t^r) \right\} / \sqrt{\tau(t)}$$

Let's generalize now the process obtained by kriging. $M_n(.,i)$ will be the kriging process for family i. $\forall t \in [t_1, t_K] \backslash \mathbb{T}_k$:

$$M_n(t,i) = \left\{ e^{-2(t - t^\ell)} - \gamma(t) \, e^{-2(t^r - t^\ell)} \right\} S_n(t^\ell, i) \;\; + \;\; \gamma(t) \, S_n(t^r, i)$$

where $\gamma(t)$ is given in Section 4.1.
$\sum_{i=1}^{I} \{M_n(.,i)\}^2$ will be the kriging process. By contiguity, under $H_{at^\star}$, $M_n(.,i)$ is the same process as under H_0 on which the mean function, $\bar{m}^i_{t^\star}(t)$ has been added :

$$\bar{m}^i_{t^\star}(t_k) = \left\{ e^{-2(t - t^\ell)} - \gamma(t) \, e^{-2(t^r - t^\ell)} \right\} m^i_{t^\star}(t^\ell) \;\; + \;\; \gamma(t) \, m^i_{t^\star}(t^r)$$

8.5. Comparison with Chang et al. (2009)

The law of the LRT process has also been obtained by Chang et al. (2009) under the null hypothesis. We propose here to present technical differences between our work and the work of Chang et al. (2009). As at a location t, the LRT is asymptotically the square of the score test, we will focus only on the score process as in Chang et al. (2009).

The main difference between the two approaches is that we consider the number of individuals in each class as a random variable whereas in Chang et al. (2009), the number of individuals in each class is supposed equal to the expectations (same remark as (b) of Section 3.2).

Our approach allows us to compute the score function $\frac{\partial \log L}{\partial q} \, |_{\theta_0}$ for only one observation and to calculate the Fisher information matrix without approximation.
Anyway, we obtain exactly the same Fisher information matrix as in Chang et al. (2009). However, there are some differences concerning other quantities.

8.5.1. Only two markers :

Let consider that there is only two markers as described in Section 3. Let $t \in]t_1, t_2[$. The result will be prolonged by continuity at the markers positions. According to formula (4)

of Section 3.2 and using the fact that $Q_t^{1,1} = 1 - Q_t^{-1,-1}$ and $Q_t^{1,-1} = 1 - Q_t^{-1,1}$, the score test statistic is :

$$S_n(t) = (1 - 2Q_t^{-1,-1}) \sum_{j=1}^{n} \frac{(y_j - \mu)\left\{1_{X_j(t_1)=1} 1_{X_j(t_2)=1} - 1_{X_j(t_1)=-1} 1_{X_j(t_2)=-1}\right\}}{\sigma\sqrt{n}\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}}$$
$$+ (1 - 2Q_t^{-1,1}) \sum_{j=1}^{n} \frac{(y_j - \mu)\left\{1_{X_j(t_1)=1} 1_{X_j(t_2)=-1} - 1_{X_j(t_1)=-1} 1_{X_j(t_2)=1}\right\}}{\sigma\sqrt{n}\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}}$$

With our notations, the test statistic used in formula (8) of Chang et al. (2009) is :

$$U^\star(t) = \frac{\sqrt{n}}{2}\,(1 - 2Q_t^{-1,-1})\,\frac{\bar{r}(t_1,t_2)\,(\bar{y}_{11} - \bar{y}_{-1-1})}{\hat{\sigma}\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}} \;+\; \frac{\sqrt{n}}{2}\,(1 - 2Q_t^{-1,1})\,\frac{r(t_1,t_2)\,(\bar{y}_{1-1} - \bar{y}_{-11})}{\hat{\sigma}\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}}$$

where $\bar{y}_{11} = \frac{2}{n\bar{r}(t_1,t_2)} \sum_{j=1}^{n} 1_{X_j(t_1)=1} 1_{X_j(t_2)=1}$, $\bar{y}_{-11} = \frac{2}{nr(t_1,t_2)} \sum_{j=1}^{n} 1_{X_j(t_1)=-1} 1_{X_j(t_2)=1}$ $\bar{y}_{1-1} = \frac{2}{nr(t_1,t_2)} \sum_{j=1}^{n} 1_{X_j(t_1)=1} 1_{X_j(t_2)=-1}$ and $\bar{y}_{-1-1} = \frac{2}{n\bar{r}(t_1,t_2)} \sum_{j=1}^{n} 1_{X_j(t_1)=-1} 1_{X_j(t_2)=-1}$.

We can remark $S_n(t) \neq U^\star(t) + o_{P_{\theta_0}}(1)$. It is due to the approximations done by Chang et al. (2009).
Let $G_n^1(t)$ and $G_n^2(t)$ be the quantities such as :

$$G_n^1(t) = \sum_{j=1}^{n} \frac{(y_j - \mu)\left\{1_{X_j(t_1)=1} 1_{X_j(t_2)=1} - 1_{X_j(t_1)=-1} 1_{X_j(t_2)=-1}\right\}}{\sigma\sqrt{n\,\bar{r}(t_1,t_2)}}$$
$$G_n^2(t) = \sum_{j=1}^{n} \frac{(y_j - \mu)\left\{1_{X_j(t_1)=1} 1_{X_j(t_2)=-1} - 1_{X_j(t_1)=-1} 1_{X_j(t_2)=1}\right\}}{\sigma\sqrt{n\,r(t_1,t_2)}}$$

$G_n^1(t)$ and $G_n^2(t)$ are asymptotically standard normal variables under H_0. Besides, $G_n^1(t)$ and $G_n^2(t)$ are independent. Note that $G_n^1(t)$ and $G_n^2(t)$ do not depend on t but we keep t as a parameter in order to adapt these test statistics to the case of several markers in the next Section.
Contrary to formula (9) of Chang et al. (2009) :

$$G_n^1(t) \neq \frac{1}{2}\sqrt{\bar{r}(t_1,t_2)n}\;\frac{\bar{y}_{11} - \bar{y}_{-1-1}}{\hat{\sigma}} + o_{P_{\theta_0}}(1)$$
$$G_n^2(t) \neq \frac{1}{2}\sqrt{r(t_1,t_2)n}\;\frac{\bar{y}_{1-1} - \bar{y}_{-11}}{\hat{\sigma}} + o_{P_{\theta_0}}(1)$$

We have :

$$S_n(t) = \left\{\sqrt{\bar{r}(t_1,t_2)}\,(1 - 2Q_t^{-1,-1})\,G_n^1(t) \;+\; \sqrt{r(t_1,t_2)}\,(1 - 2Q_t^{-1,1})\,G_n^2(t)\right\} / \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]} \tag{24}$$

This formula is the corrected version of formula (10) of Chang et al. (2009) without approximations here. According to formula (24), the score at a position t between two markers,

is an interpolation not linear between the test statistic $G_n^1(t)$ and $G_n^2(t)$. Naturally, when t tends to t_1 (resp. t_2), $S_n(t)$ tends to $S_n(t_1)$ (resp. $S_n(t_2)$). It becomes a linear interpolation between $S_n(t_1)$ and $S_n(t_2)$ if a Taylor linearization is done concerning the weights of the model of mixture (cf. Section 3.2).
Finally, we agree with formula (11) of Chang et al. (2009) concerning the covariance of the process, it is exactly the same function as $\Gamma(t,t')$ of theorem 1 of this paper.

Note that the non linear interpolation presented above, in formula (24), is not the same interpolation as presented in formula (5) of Section 3.1 of this paper. Our interpolation is more intuitive, because it is an interpolation between the test statistic on markers. Besides, we will see in the next Section that there is an advantage using our interpolation in terms of simulations.

8.5.2. Several markers : the "Interval Mapping" of Lander and Botstein (1989)
Let consider that there are several markers as described in Section 4. We consider values t, t' of the parameters that are distinct of markers positions. Let $t \in [t_1, t_K] \backslash \mathbb{T}_k$. We have :

$$G_n^1(t) = \sum_{j=1}^{n} \frac{(y_j - \mu)\left\{1_{X_j(t^\ell)=1} 1_{X_j(t^r)=1} - 1_{X_j(t^\ell)=-1} 1_{X_j(t^r)=-1}\right\}}{\sigma\ \sqrt{n\ \bar{r}(t^\ell, t^r)}}$$

$$G_n^2(t) = \sum_{j=1}^{n} \frac{(y_j - \mu)\left\{1_{X_j(t^\ell)=1} 1_{X_j(t^r)=-1} - 1_{X_j(t^\ell)=-1} 1_{X_j(t^r)=1}\right\}}{\sigma\ \sqrt{n\ r(t^\ell, t^r)}}$$

$$S_n(t) = \left\{\sqrt{\bar{r}(t^\ell, t^r)}\ (2Q_t^{1,1} - 1)\ G_n^1(t)\ +\ \sqrt{r(t^\ell, t^r)}\ (2Q_t^{1,-1} - 1)\ G_n^2(t)\right\} / \sqrt{\mathbb{E}\left[\{2p(t) - 1\}^2\right]}$$

This last formula is the corrected version of formula (14) of Chang et al. (2009).
Let $(t,t') \in]t^\ell, t^r[\ \times [t^r, t_K] \backslash \mathbb{T}_k$. The different covariances under H_0 are :

$$\mathrm{Cov}\left\{G_n^1(t), G_n^1(t')\right\} = \sqrt{\bar{r}(t^\ell, t^r)\ \bar{r}\left\{(t')^\ell, (t')^r\right\}}\ \ e^{-2\left\{(t')^\ell - t^r\right\}}$$

$$\mathrm{Cov}\left\{G_n^1(t), G_n^2(t')\right\} = \sqrt{\bar{r}(t^\ell, t^r)\ r\left\{(t')^\ell, (t')^r\right\}}\ \ e^{-2\left\{(t')^\ell - t^r\right\}}$$

$$\mathrm{Cov}\left\{G_n^2(t), G_n^1(t')\right\} = -\sqrt{r(t^\ell, t^r)\ \bar{r}\left\{(t')^\ell, (t')^r\right\}}\ \ e^{-2\left\{(t')^\ell - t^r\right\}}$$

$$\mathrm{Cov}\left\{G_n^2(t), G_n^2(t')\right\} = -\sqrt{r(t^\ell, t^r)\ r\left\{(t')^\ell, (t')^r\right\}}\ \ e^{-2\left\{(t')^\ell - t^r\right\}}$$

This is exactly the same covariances as in formula (19) of Chang et al. (2009). Besides, we agree with formula (20) of Chang et al. (2009) which establish a relationship between the test statistic G when t and t' belong to 2 consecutive marker interval (as above we suppose $t < t'$):

$$G_n^2(t') = \frac{1}{\sqrt{r(t^r, (t')^r)}} \left\{\sqrt{\bar{r}(t^\ell, t^r)}\ G_n^1(t) - \sqrt{r(t^\ell, t^r)}\ G_n^2(t) - \sqrt{\bar{r}(t^r, (t')^r)}\ G_n^1(t')\right\}$$

To conclude, the non linear interpolation proposed by Chang et al. (2009) is an approximation. We present here their interpolation without approximations. However, their

approximations don't affect the final results concerning the process. Their approach is very interesting because it characterizes the process by interpolation between the test statistic G, and in terms of covariance between the test statistic G.
In this paper, we have calculated the whole covariance function $\Gamma(t, t')$ (cf. theorem 2) in order to see if under the alternative, the shift at a position t was $\Gamma(t, t^\star)$ as in Azaïs et al. (2006) and Azaïs et al. (2009). We have also characterized the process by a non linear interpolation between the test statistic on markers. When tests are performed only on markers, the score process is a Discrete Ornstein Uhlenbeck (DOU) process (cf. Section 4). As it is well known that the DOU process is an AR(1) process, it will be easy to simulate the discrete process on markers, and to obtain the values between markers by interpolation.

9. Acknowledgements

The authors thank Jean-Michel Elsen for having proposed this subject of research and 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.

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

Azaïs, J. M., Gassiat, E., Mercadier, C. (2009) The likelihood ratio test for general mixture models with possibly structural parameter. *ESAIM*, To appear.

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

Billingsley, P., (1999) *Convergence of probability measures.* Wiley, New-York.

Candes, E. J. and Tao, T., (2005) The Dantzig selector : statistical estimation when p is much larger than n. *Annals of Statistics*, **35**, 2313-2351.

Chang, M. N., Wu, R., Wu, S. S., 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, R.B., (1977) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **64**, 247-254.

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

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.

Le Cam, L. (1986) *Asymptotic Methods in Statistical Decision Theory*, Springer.

Rabier, C-E. (2009) *PhD thesis*, Université Toulouse 3, Paul Sabatier.

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.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society - B*, **58**, **1**, 267-288.

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

Zou, H., Hastie, T. (2005) Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society - B*, **67**, **2**, 301-320.