Likelihood Ratio Test process for Quantitative Trait Locus detection

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

Likelihood Ratio Test process for Quantitative Trait Locus detection

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.

Jean-Marc Azaïs

Université de Toulouse, Institut de Mathématiques de Toulouse, U.P.S, Toulouse, France.

Céline Delmas

INRA UR631, Station d'Amélioration Génétique des Animaux, 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 alternative that there exists a QTL at $t^\star$ on $[0, T]$. We show that the LRT is asymptotically the square of a non linear interpolated process. We propose a simple and original method to calculate the maximum and the argmax of the LRT process using only statistics on markers and their ratio. We finally propose a new method to calculate thresholds for QTL detection.

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 backcross population: $A \times (A \times B)$, where A and B are purely homozygous lines 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 purely inherited from A while the other (the "recombined" one), consists of parts originated from A and parts originated from B, due to crossing-overs. 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. 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 $t = t^\star$.

The main result of the paper (theorem 1 and theorem 3) is that the LRT process is asymptotically the square of a "non linear interpolated process". It describes the fact that, when we analyze data, the likelihood profile (ie. the path observed of the LRT process) is smooth between markers. Besides, we have a close formula (lemma 1 and lemma 2) to compute the maximum of the LRT process. This formula allows us to give advice on how to analyze data : we should first perform tests on markers and then calculate only one other statistic in each marker interval if the ratio between the score statistics on the flanking markers fulfill a given condition. Finally, we propose a new method suitable whatever the genetic map, using Monte-Carlo Quasi Monte-Carlo (Genz (1992)), to calculate thresholds for QTL detection. This method will be compared with Rebaï et al. (1994)'s method based on Davies (1977), and with Feingold and al. (1993)'s method based on Siegmund (1985).

Note that in this article, we also prove that the LRT process obtained by Rebaï et al. (1994), Rebaï et al. (1995) is asymptotically the square of a "linear interpolated process" and we generalize their results to the alternative hypothesis. Besides, we show that the law of the maximum of the square of the "non linear interpolated process" is the same as the law of the maximum of the square of the "linear interpolated process". We refer to the book of Van der Vaart (1998) for element of asymptotic statistics used in proofs.

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 :

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

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

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

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. A "non linear interpolated process"

Theorem 1 *With the previous defined notations, and defining respectively $\Lambda_n(.)$ and $S_n(.)$, the LRT process and the score process for n observations,*

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

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

- *$\Rightarrow$ 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 interpolated process such as $\forall t \in [t_1,t_2]$:

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

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* $\alpha(t_1) = 1$, $\beta(t_1) = 0$, $\alpha(t_2) = 0$, $\beta(t_2) = 1$, $\mathrm{Cov}\{\mathrm{Z}(\mathrm{t}_1), \mathrm{Z}(\mathrm{t}_2)\} = \mathrm{e}^{-2(\mathrm{t}_2 - \mathrm{t}_1)}$.
In the same way, $\forall (t, t^\star) \in [t_1, t_2]^2$ *:*

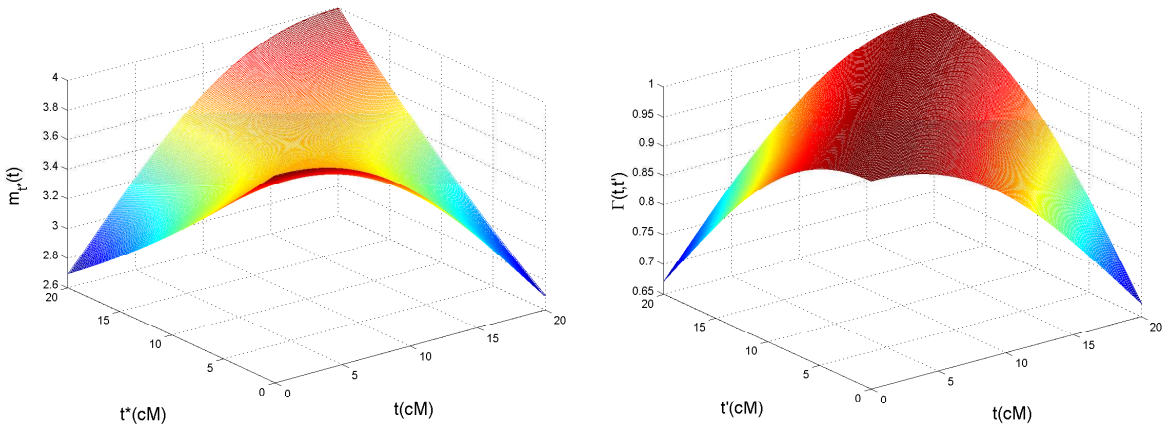
$$m_{t^\star}(t) = \{\ \alpha(t)\ m_{t^\star}(t_1)\ +\ \beta(t)\ m_{t^\star}(t_2)\ \} \Big/ \sqrt{\mathbb{E}\left[\{2p(t) - 1\}^2\right]}$$

The quantity $\mathbb{E}\left[\{2p(t) - 1\}^2\right]$ *is given in formula (9) of the proof of the theorem in Section 5.1.* $\mathbb{E}\{p(t)p(t')\}$ *is given in appendix 7.1.* $\mathbb{E}[X(t^\star)\{2p(t) - 1\}]$ *is given in formula (14) of the proof in Section 5.1.*

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 $X(t_2)$.



Mean function Covariance function

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

3.2. Remarks

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

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

We remind that, as in the proof of the theorem in Section 5.1, 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).
A little algebra shows (see Section 5.1) :

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

The score test can be obtained, replacing μ by $\hat{\mu} = \sum y_j/n$, according to Prohorov, and replacing σ by $\hat{\sigma} = \sum(y_j - \hat{\mu})^2/n$, according to Slutsky's lemma. Nevertheless, in this article, in order to make the reading easier, the score test statistic is 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 $\mathrm{Cov}_{\mathrm{H_0}}\ \{\mathrm{S_n(t_1), S_n(t_2)}\} = \mathrm{e}^{-2(\mathrm{t_2 - t_1})}$. $\alpha(t)$ and $\beta(t)$ are given quantities in theorem 1. It comes :

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

Besides, by contiguity (cf. proof of theorem 1 in Section 5.1), 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 score test statistics on the markers.

3.3. A "linear interpolated process"

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 :

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

where

$$\begin{aligned}\tau(t) &= \mathbb{V}_{\mathrm{H_0}}\left\{\frac{t_2 - t}{t_2 - t_1}\ S_n(t_1)\ +\ \frac{t - t_1}{t_2 - t_1}\ S_n(t_2)\right\}\\ &= \left(\frac{t_2 - t}{t_2 - t_1}\right)^2\ +\ 2\,\frac{(t - t_1)(t_2 - t)}{(t_2 - t_1)^2}\,e^{-2(t_2 - t_1)}\ +\ \left(\frac{t - t_1}{t_2 - t_1}\right)^2\end{aligned}$$

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 $\mathrm{Cov}_{\mathrm{H_0}}\ \{\mathrm{S_n(t_1), S_n(t_2)}\} = \mathrm{e}^{-2(\mathrm{t_2 - t_1})}$.

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

(a) According to formula (10) in Section 5.1 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 :

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

We can remark that these weights are an approximation at the first order of the weights considered previously in formula (2). So, $V_n(.)$ will be a good approximation if and only if the genetic markers are close to 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 5.1), 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 :

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

(d) $V_n(.)$ is defined here with $\mathrm{Cov}_{\mathrm{H_0}} \{\mathrm{S_n(t_1)}, \mathrm{S_n(t_2)}\} = \mathrm{e}^{-2(\mathrm{t_2}-\mathrm{t_1})}$. In order to consider other covariances between $S_n(t_1)$ and $S_n(t_2)$, $\tau(.)$ 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) provided that $\mathbb{E}\left[\{2p(t)-1\}^2\right] \neq 0$ ($p(t)$ verifies formula (7)).

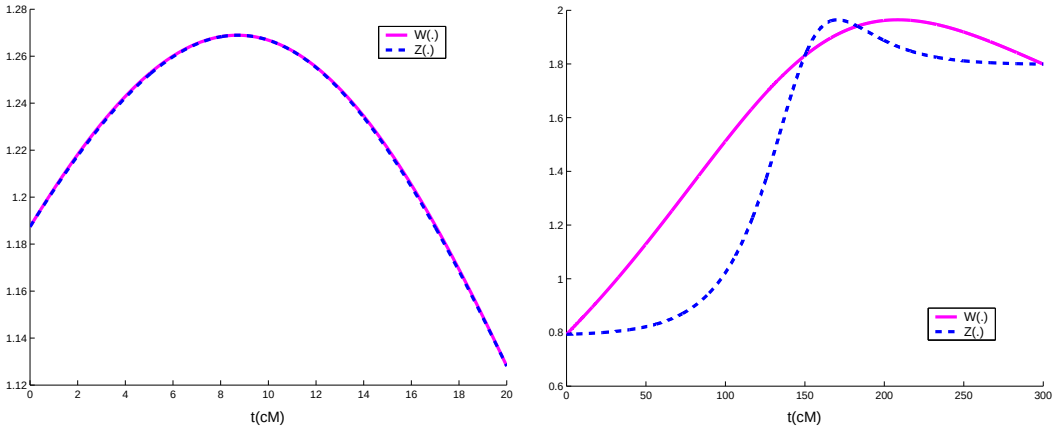
We will name $W(.)$ the limiting process of $V_n(.)$: $W(.)$ is the linear interpolated process. Figure 2 represents two paths of the processes $W(.)$ and $Z(.)$ under the null hypothesis. We remind that $Z(.)$ is the non linear interpolated process. We can observe that the paths of the two processes overlap when the distance between the two genetic markers is 20cM, that is to say 0.2M. As mentioned before, $W(.)$ is a good approximation of $Z(.)$ when the genetic markers are close to each other. As expected, we can remark that when the distance between the markers is 3M, the paths of the two processes don't overlap anymore. Note that same conclusions hold under contiguous alternatives (cf. Rabier (2010)).

3.4. Impact of the interpolations on data analysis

In this Section, we state why the results about the non linear interpolation and about the linear interpolation are important for data analysis. To begin, we present a theorem and a lemma.

Theorem 2 *Let C_1 and C_2 be two continuous random variables, and let $\tilde{\rho}$ such as $0 < \tilde{\rho} < 1$. Let consider $\gamma_1(t)$ and $\gamma_2(t)$, two continuous functions on $[t_1,t_2]$, and let define the process $D(.)$ on $[t_1,t_2]$ such as :*

$$D(t) = \frac{\gamma_1(t)\, C_1 \; + \; \gamma_2(t)\, C_2}{\sqrt{\{\gamma_1(t)\}^2 \; + \; \{\gamma_2(t)\}^2 \; + \; 2\, \tilde{\rho}\, \gamma_1(t)\, \gamma_2(t)}}$$



One path for the two processes ($T = 0.2$M) One path for the two processes ($T = 3$M)

Fig. 2. Comparison between the paths, under H_0, of the linear interpolated process $W_{(.)}$ and those of the non linear interpolated process $Z_{(.)}$

then if the function $\frac{\gamma_2(t)}{\gamma_1(t)+\gamma_2(t)}$ *is bounded by 0 and 1 on* $[t_1, t_2]$*, and if these bounds are reached, then*

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

Lemma 1 *With the previous defined notations and reminding that* $W(t_1) = Z(t_1)$ *and* $W(t_2) = Z(t_2)$*,*

$$\text{let} \quad \xi = \frac{(t_2-t_1)\ \left\{e^{-2(t_2-t_1)}\ W(t_1) - W(t_2)\right\}}{\left\{e^{-2(t_2-t_1)}-1\right\}\left\{W(t_1)+W(t_2)\right\}} + t_1 \ \text{, then under } H_0 \text{ and } H_{at^\star}$$

$$\{W(\xi)\}^2 = \frac{\{W(t_1)\}^2 + \{W(t_2)\}^2 - 2\,e^{-2(t_2-t_1)}\,W(t_1)\,W(t_2)}{\left\{1+e^{-2(t_2-t_1)}\right\}\left\{1-e^{-2(t_2-t_1)}\right\}} \qquad \text{and}$$

$$\sup_{t\in[t_1,t_2]} \{Z(t)\}^2 = \sup_{t\in[t_1,t_2]} \{W(t)\}^2$$
$$= \max\left[\{W(t_1)\}^2 \ , \ \{W(\xi)\}^2\ 1_{\frac{W(t_2)}{W(t_1)} \in \,]\,e^{-2(t_2-t_1)}\ ,\ e^{2(t_2-t_1)}\,[} \ , \ \{W(t_2)\}^2\right]$$

The proof of theorem 2 and lemma 1 are respectively given in Sections 5.2 and 5.3. According to lemma 1, even when the genetic markers are not close to each other, the law of the supremum of the square of the two interpolated processes is the same (see Figure 2). Howewer, when the supremum is obtained between markers, it is not obtained at the same

positions. These locations are ξ for $\{W(.)\}^2$ and ξ' for $\{Z(.)\}^2$, where ξ' is such as :

$$\frac{(t_2 - t_1)\ \beta(\xi')}{\alpha(\xi')\ +\ \beta(\xi')}\ +\ t_1 = \xi$$

On the other hand, lemma 1 can easily be adapted to the non asymptotic processes. Indeed, we can replace $Z(.)$ by $S_n(.)$, and $W(.)$ by $V_n(.)$ everywhere in this lemma, since the focus is on the same interpolations. As previously, when the supremum is obtained between markers, it is obtained at ξ for $V_n(.)$ and ξ' for $S_n(.)$. Furthermore, we can advise not to perform a large number of tests between the genetic markers anymore. First, we should perform score tests on markers. Then, only if the ratio observed between the score statistics on markers (ie. the ratio $S_n(t_2)/S_n(t_1)$) belongs to the interval $]\ e^{-2(t_2-t_1)}\ ,\ e^{2(t_2-t_1)}\ [$, we have to calculate another quantity :

$$\zeta = \frac{\{S_n(t_1)\}^2 + \{S_n(t_2)\}^2 - 2\ e^{-2(t_2-t_1)}\ S_n(t_1)\ S_n(t_2)}{\left\{1 + e^{-2(t_2-t_1)}\right\}\left\{1 - e^{-2(t_2-t_1)}\right\}}$$

To conclude, we should use as a test statistic :

$$\max\left[\{S_n(t_1)\}^2\ ,\ \zeta\ 1_{\frac{S_n(t_2)}{S_n(t_1)}\ \in\ ]\ e^{-2(t_2-t_1)}\ ,\ e^{2(t_2-t_1)}\ [}\ ,\ \{S_n(t_2)\}^2\right]$$

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 3 *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 interpolated 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 $\forall k\ \forall k'$, $\mathrm{Cov}\left\{\mathrm{Z(t_k), Z(t_{k'})}\right\} = \mathrm{e}^{-2|\mathrm{t_k} - \mathrm{t_{k'}}|}$.
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 7.2.

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).
The paths of three processes are presented in Figure 3 ($T = 1$M):

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

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

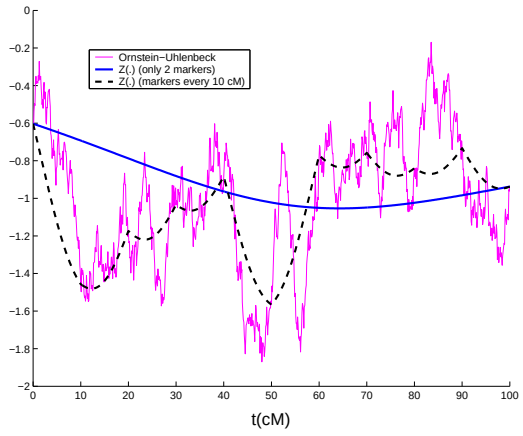


Fig. 3. Paths of three different Gaussian processes

Recently, the law of the LRT process under the null hypothesis has also been obtained by Chang et al. (2009). Technical differences are presented in appendix 7.4. The originality of our work is first, that we focus not only on the null hypothesis. Secondly, we show that the LRT process is asymptotically the square of a "non linear interpolated process". It describes the fact that, when we analyze data, the likelihood profile (ie. the path observed of the LRT process) is smooth between markers.

4.1. Remarks

The linear interpolated process $W(.)$ presented in Section 3.3 can easily be generalized to the case of several markers. This is a generalization of the process studied, under H_0, by Rebaï et al. (1994). The details are given in appendix 7.3.

In the same way, lemma 1 can be generalized :

Lemma 2 *With the previous defined notations*

and reminding that $\forall k \;\; W(t_k) = Z(t_k)$,

$$\text{let} \quad \xi(t^\ell, t^r) = \frac{(t^r - t^\ell)\,\left\{e^{-2(t^r-t^\ell)}\,W(t^\ell) - W(t^r)\right\}}{\left\{e^{-2(t^r-t^\ell)} - 1\right\}\left\{W(t^\ell) + W(t^r)\right\}} \;+\; t^\ell$$

, then under H_0 *and* $H_{at^\star}$

$$\left[W\left\{\xi(t^\ell, t^r)\right\}\right]^2 = \frac{\left\{W(t^\ell)\right\}^2 + \left\{W(t^r)\right\}^2 - 2\,e^{-2(t^r-t^\ell)}\,W(t^\ell)\,W(t^r)}{\left\{1 + e^{-2(t^r-t^\ell)}\right\}\left\{1 - e^{-2(t^r-t^\ell)}\right\}}$$

and

$$\sup_{t\in[0,T]} \left\{Z(t)\right\}^2 = \sup_{t\in[0,T]} \left\{W(t)\right\}^2$$

$$\sup_{t\in[0,T]} \left\{W(t)\right\}^2 = \max\left[\left\{W(t_1)\right\}^2, ..., \left\{W(t_K)\right\}^2, \left[W\left\{\xi(t_1,t_2)\right\}\right]^2 1_{\frac{W(t_2)}{W(t_1)} \in \,]\, e^{-2(t_2-t_1)}\,,\, e^{2(t_2-t_1)}\,[} \right.$$

$$\left., ..., \left[W\left\{\xi(t_{K-1},t_K)\right\}\right]^2 1_{\frac{W(t_K)}{W(t_{K-1})} \in \,]\, e^{-2(t_K-t_{K-1})}\,,\, e^{2(t_K-t_{K-1})}\,[}\right]$$

The proof of this lemma is largely inspired from the proof of lemma 1. Using lemma 2 and using the same arguments as in Section 3.4, we can advise to perform tests on markers and to calculate one other statistic in each marker interval when it is required.

4.2. Application to the calculation of thresholds

The theoretical results presented in this article allow us to propose a new method to obtain the $\alpha\%$ quantile of the supremum of the process $\left\{Z(.)\right\}^2$ under H_0. This method is a direct application of lemma 2. Besides, Monte-Carlo Quasi Monte-Carlo (MCQMC) methods of Genz (1992) which are very fast have been considered. As the numerical computation of a multivariate normal distribution is often a difficult problem, Genz described in his paper, a transformation that simplifies the problem and places into a form that allows efficient calculations using standard numerical multiple integration algorithms. He suggests to use in particular MCQMC algorithms. Indeed, a simple Monte-Carlo method (MC) using N points have errors that are typically $O(1/\sqrt{N})$ whereas Quasi Monte-Carlo methods (QMC) have errors $O(1/N)$. In order to be sure that the functions studied have nice properties for QMC, another Monte-Carlo step is required, this is MCQMC. We refer to Genz (1992) for more details. We use here function QSIMVNEF of Genz, which is a Matlab function with supporting functions, for the numerical computation of multivariate normal distribution expected values. This function has been adapted and a Newton method has been used in order to obtain the thresholds.

Our method is available in a Matlab package with graphical user interface : “imapping.zip". It can be downloaded at www.math.univ-toulouse.fr/∼rabier .

In this Section, we propose to compare the performances of our method with other methods usually used in QTL detection.
In Rebaï et al. (1994), we can find an upper bound for the threshold. This bound is the

quantity c^2 such as :

$$1-\alpha = 2\,\Phi(-c) \;+\; \frac{2\,e^{-c^2/2}}{\pi} \sum_{k=1}^{K-1} \arctan\left(\sqrt{\frac{1-e^{-2(t_{k+1}-t_k)}}{1+e^{-2(t_{k+1}-t_k)}}} \right)$$

where Φ is the cumulative distribution of the standardized normal distribution.
This method is based on Davies (1977). However, it is sensitive to the number of genetic markers. Indeed, the derivative of the process $W(.)$ has a jump at each markers location, and Davies (1977) upper bound is suitable when the derivative of the process has a finite number of jumps.

In Feingold and al. (1993), the authors propose a threshold based on the discrete process resulting from tests only on markers. Besides, they suppose constant the distance between genetic markers. The threshold c^2 is such as :

$$1-\alpha = 1 \;-\; \Phi(c) \;+\; 2\,T\,c\,\varphi(c)\,\nu(2c\sqrt{\Delta})$$

where φ is the density of a normal standardized, Δ is distance between two consecutive markers.
This method is inspired from Siegmund (1985) where the function ν is fully described.

In Figures 4 and 5, thresholds corresponding to different methods are computed. As expected, Rebaï's method is very sensitive to the number of genetic markers. We can observe that Feingold's method and our method give almost same results.

Howewer, our method give different results than Feingold when the number of genetic markers is very small (cf. Figure 6). Indeed, Feingold's method requires the number of genetic makers to be not too small (cf. Feingold and al. (1993)). The advantage of our method is that this method is appropriate whatever the map.

Method	*this paper*	*Rebaï*	*Feingold*
Threshold	6.76	6.92	6.78

Fig. 4. Thresholds as a function of the method considered. The map consists of 6 genetic markers equally spaced every 20cM ($T = 1$M, $\alpha = 95\%$).

Method	*this paper*	*Rebaï*	*Feingold*
Threshold	8.23	9.09	8.26

Fig. 5. Thresholds as a function of the method considered. The map consists of 51 genetic markers equally spaced every 2cM ($T = 1$M, $\alpha = 95\%$).

Method	*this paper*	*Feingold*
Threshold	5.40	5.78

Fig. 6. Thresholds as a function of the method considered. The map consists of 2 genetic markers ($T = 1$M, $\alpha = 95\%$).

5. 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.

5.1. Proof of theorem 1

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

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

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

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

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,

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

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

$$\mathbb{E}\left[\{2p(t)-1\}^2\right] = \bar{r}(t_1,t_2)\left(2Q_t^{1,1}-1\right)^2 \ + \ r(t_1,t_2)\left(2Q_t^{1,-1}-1\right)^2 \tag{9}$$

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 (9).
$\mathbb{E}\left[\{2p(t)-1\}^2\right]$ is always different from zero since the parameter t is bounded. It comes $\forall t \in [t_1,t_2]$:

$$\Lambda_n(t) = \left[\sum_{j=1}^{n} \frac{(y_j-\mu)\ \{2p_j(t)-1\}}{\sigma\ \sqrt{n}\ \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}}\right]^2 \ +\ o_{P_{\theta_0}}(1) \tag{10}$$

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 :

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

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

$$\begin{aligned}\Gamma(t,t') := \mathrm{Cov}\{\mathrm{S(t)},\ \mathrm{S(t')}\} = \mathbb{E}\{\mathrm{h(t)h(t')}\} &= \frac{\mathbb{E}\left[\{2p(t)-1\}\ \{2p(t')-1\}\right]}{\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}\sqrt{\mathbb{E}\left[\{2p(t')-1\}^2\right]}} \\ &= \frac{4\mathbb{E}\{p(t)p(t')\}-1}{\sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}\sqrt{\mathbb{E}\left[\{2p(t')-1\}^2\right]}}\end{aligned} \tag{11}$$

The formula for $\mathbb{E}\{p(t)p(t')\}$ is given in appendix 7.1. As $|p(t)p(t')| \leqslant 1$, by dominated convergence theorem, $\mathbb{E}\{p(t)p(t')\}$ 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 :

$$S_n(t) = \sum_{j=1}^{n} \frac{(y_j-\mu)\ (2p_j(t)-1)}{\sigma\ \sqrt{n}\ \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]}} \tag{12}$$

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

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

were Σ is the variance covariance matrix, with unit variance and covariance given by formula (11). $\underset{\sim}{0}$ 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

$$Y_j = \mu \ + \ \frac{a}{\sqrt{n}} \, X_j(t^\star) \ + \ \sigma \, \varepsilon_j \tag{13}$$

where ε_j is a Gaussian white noise. According to formula (10), $\forall t \in [t_1, t_2]$:

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

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, by the iii) of Le Cam's first lemma, we have $Q_n \triangleleft P_n$.

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 (12) and (13), 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 :

$$\begin{aligned}
&\mathbb{E}\left[\ X(t^\star) \ \{2p(t) - 1\}\right] \\
&= \left\{2Q_t^{1,1} - 1\right\} \ \left\{2Q_{t^\star}^{1,1} - 1\right\} \ \frac{1}{2} \ \bar{r}(t_1, t_2) \ + \ \left\{2Q_t^{1,-1} - 1\right\} \ \left\{2Q_{t^\star}^{1,-1} - 1\right\} \ \frac{1}{2} \ r(t_1, t_2) \\
&+ \left\{2Q_t^{-1,1} - 1\right\} \ \left\{2Q_{t^\star}^{-1,1} - 1\right\} \ \frac{1}{2} \ r(t_1, t_2) \ + \ \left\{2Q_t^{-1,-1} - 1\right\} \ \left\{2Q_{t^\star}^{-1,-1} - 1\right\} \ \frac{1}{2} \ \bar{r}(t_1, t_2) \\
&= \bar{r}(t_1, t_2) \ \left\{2Q_{t^\star}^{1,1} - 1\right\} \ \left\{2Q_t^{1,1} - 1\right\} \\
&+ r(t_1, t_2) \ \left\{2Q_{t^\star}^{1,-1} - 1\right\} \ \left\{2Q_t^{1,-1} - 1\right\} \qquad (14)
\end{aligned}$$

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 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 $\mathrm{Cov}_{\mathrm{H_0}}\{\mathrm{S_n(t_1), S_n(t_2)}\} = \mathrm{e}^{-2\mathrm{t_2}}$, $\alpha(t_1) = 1$, $\beta(t_1) = 0$, $\alpha(t_2) = 0$, $\beta(t_2) = 1$ 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
To begin, we remind that $t_1 = 0$ and $t_2 = T$. 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.

5.2. Proof of theorem 2

Introducing the process $\tilde{W}(.)$:
We consider the process $\tilde{W}(.)$ on $[0,1]$ such as :

$$\tilde{W}(t) = \frac{(1-t)\, C_1 \; + \; t\, C_2}{\sqrt{(1-t)^2 + t^2 + 2\, \tilde{\rho}\, t\, (1-t)}}$$

We can remark that $\tilde{W}(0) = C_1$ and $\tilde{W}(1) = C_2$.
The interest is on the supremum of the process $\left\{\tilde{W}(.)\right\}^2$:

$$\left\{\tilde{W}(t)\right\}^2 = \frac{(1-t)^2 \{C_1\}^2 \; + \; 2t(1-t)C_1C_2 \; + \; t^2 \{C_2\}^2}{(1-t)^2 \; + \; t^2 \; + \; 2\, \tilde{\rho}\, t\, (1-t)}$$

We will call respectively $N(t)$ and $D(t)$ the numerator and the denominator of the fraction above.

$$\frac{\partial N(t)}{\partial t} = 2\; \{(1-t)\, C_1 \; + \; t\, C_2\}\; \{C_2 - C_1\}$$

We can remark that :

$$(1-t)^2 + t^2 + 2\, t\, \tilde{\rho}\, (1-t) = 1 - 2(\, 1-\tilde{\rho})\, t\, (1-t) \tag{15}$$

It comes : $\frac{\partial D(t)}{\partial t} = -2\, (1-\tilde{\rho})\, (1-2t)$. So,

$$\begin{aligned}\frac{\partial \left\{\tilde{W}(t)\right\}^2}{\partial t} &= [\, 2\, \{(1-t)C_1 + tC_2\}\, \{C_2 - C_1\}\, \{1 - 2(1-\tilde{\rho})t(1-t)\} \\ &\quad + \; 2(1-\tilde{\rho})(1-2t)\, \{(1-t)C_1 + tC_2\}^2 \,\big]\, / \{D(t)\}^2\end{aligned}$$

We have :

$$\begin{aligned}&\frac{\partial \left\{\tilde{W}(t)\right\}^2}{\partial t} = 0 \\ &\Leftrightarrow \{(1-t)C_1 + tC_2\} \\ &\times \; [\{C_2 - C_1\}\, \{1 - 2(1-\tilde{\rho})t(1-t)\} + \; (1-\tilde{\rho})(1-2t)\, \{(1-t)C_1 + tC_2\}] = 0\end{aligned}$$

As $\{(1-t)C_1 + tC_2\}$ corresponds to a minimum, the focus is on the second term. This second term is equal to zero if :

$$\frac{C_2}{C_1} = \frac{1+\tilde{\rho}}{1 + (\tilde{\rho} - 1)t} - 1$$

Let define the function $\psi_{\tilde{\rho}}(t)$ such as :

$$\psi_{\tilde{\rho}}(t) = \frac{1+\tilde{\rho}}{1+(\tilde{\rho}-1)t} - 1$$

As $(\tilde{\rho}-1)t$ is a decreasing function on $[0,1]$, then $\psi_{\tilde{\rho}}(t)$ is an increasing function on $[0,1]$ with $\psi_{\tilde{\rho}}(0) = \tilde{\rho}$ and $\psi_{\tilde{\rho}}(1) = \frac{1}{\tilde{\rho}}$.
Let define $\psi_{\tilde{\rho}}^{-1}$ the inverse function of $\psi_{\tilde{\rho}}$. After straightforward calculations, we find :

$$\psi_{\tilde{\rho}}^{-1}(u) = \frac{\tilde{\rho}-u}{(\tilde{\rho}-1)(u+1)}$$

So, the extremum between 0 and 1 is obtained for :

$$\tilde{\xi} = \frac{\tilde{\rho}\, C_1 - C_2}{(\tilde{\rho}-1)\{C_2+C_1\}}$$

After some calculations, using formula (15), we find that :

$$(1-\tilde{\xi})^2 + \tilde{\xi}^2 + 2\,\tilde{\xi}\,\tilde{\rho}\,(1-\tilde{\xi}) = \frac{1+\tilde{\rho}}{1-\tilde{\rho}}\ \frac{\{C_1\}^2 + \{C_2\}^2 - 2\,\tilde{\rho}\,C_1\,C_2}{\{C_1+C_2\}^2}$$

It comes :

$$\left\{\tilde{W}(\tilde{\xi})\right\}^2 = \frac{\{C_1\}^2 + \{C_2\}^2 - 2\,\tilde{\rho}\,C_1\,C_2}{(1+\tilde{\rho})(1-\tilde{\rho})}$$

So :

$$\begin{aligned}\sup_{t\in[0,1]}\left\{\tilde{W}(t)\right\}^2 = \left\{\tilde{W}(\tilde{\xi})\right\}^2 & 1_{\frac{C_2}{C_1}\in\,]\,\tilde{\rho}\,,\,\frac{1}{\tilde{\rho}}\,[} \;+\; \{C_2\}^2\; 1_{\frac{C_2}{C_1}\in\,[\,\frac{1}{\tilde{\rho}}\,,\,+\infty\,[} \\ & + \{C_1\}^2\; 1_{\frac{C_2}{C_1}\in\,[\,0\,,\,\tilde{\rho}\,]} \;+\; \{C_1\}^2\; 1_{\frac{C_2}{C_1}\in\,]\,-\infty\,,\,0[\,\bigcap\,|C_1|>|C_2|} \\ & + \{C_2\}^2\; 1_{\frac{C_2}{C_1}\in\,]\,-\infty\,,\,0[\,\bigcap\,|C_2|>|C_1|} \end{aligned} \tag{16}$$

A concise version of this formula is that the supremum of $\left\{\tilde{W}(.)\right\}^2$ is the maximum of three random variables :

$$\sup_{t\in[0,1]}\left\{\tilde{W}(t)\right\}^2 = \max\left[\{C_1\}^2\ ,\ \left\{\tilde{W}(\tilde{\xi})\right\}^2\ 1_{\frac{C_2}{C_1}\in\,]\,\tilde{\rho}\,,\,\frac{1}{\tilde{\rho}}\,[}\ ,\ \{C_2\}^2\right]$$

Let $\gamma_1(t)$ and $\gamma_2(t)$ be two continuous functions on $[t_1,t_2]$. Besides, as in theorem 2, let suppose that $\frac{\gamma_2(t)}{\gamma_1(t)+\gamma_2(t)}$ is bounded by 0 and 1, and that these bounds are reached. We have $\forall t \in [t_1,t_2]$:

$$\tilde{W}\left(\frac{\gamma_2(t)}{\gamma_1(t)+\gamma_2(t)}\right) = D(t)$$

And,

$$\sup_{t\in[t_1,t_2]}\{D(t)\}^2 = \sup_{t\in[0,1]}\left\{\tilde{W}(t)\right\}^2$$

It concludes the proof.

5.3. Proof of lemma 1

Study of the supremum of the linear interpolated process $W(.)$:
Let consider the process $W(.)$ on $[t_1, t_2]$. It verifies $\forall t \in [t_1, t_2]$:

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

where

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

Using same notations as in theorem 2, let consider :

$$\gamma_1(t) = \frac{t_2 - t}{t_2 - t_1} \;\; , \;\; \gamma_2(t) = \frac{t - t_1}{t_2 - t_1} \;\; , \;\; C_1 = W(t_1) \;\; , \;\; C_2 = W(t_2) \;\; , \;\; \tilde{\rho} = e^{-2(t_2 - t_1)}$$

We will call $\gamma_3(t)$ the ratio $\frac{\gamma_2(t)}{\gamma_1(t) + \gamma_2(t)}$. We have $\gamma_3(t) = \frac{t - t_1}{t_2 - t_1}$. So, $\gamma_3(t_2) = 1$, $\gamma_3(t_1) = 0$, and $0 \leqslant \gamma_3(t) \leqslant 1$. As a consequence, according to theorem 2 :

$$\sup_{t \in [t_1, t_2]} \{W(t)\}^2 = \max \Big[\{W(t_1)\}^2 \;\; , \;\; \{W(t_2)\}^2 \;\; , \;\; \frac{\{W(t_1)\}^2 + \{W(t_2)\}^2 - 2 \, e^{-2(t_2 - t_1)} \, W(t_1) \, W(t_2)}{\left\{1 + e^{-2(t_2 - t_1)}\right\} \left\{1 - e^{-2(t_2 - t_1)}\right\}} \, 1_{\frac{W(t_2)}{W(t_1)} \, \in \,] \, e^{-2(t_2 - t_1)} \, , \, e^{2(t_2 - t_1)} \, [} \Big]$$

In the same way as in the proof of theorem 2 (cf. Section 5.2), we have :

$$\{W(\xi)\}^2 = \frac{\{W(t_1)\}^2 + \{W(t_2)\}^2 - 2 \, e^{-2(t_2 - t_1)} \, W(t_1) \, W(t_2)}{\left\{1 + e^{-2(t_2 - t_1)}\right\} \left\{1 - e^{-2(t_2 - t_1)}\right\}}$$

with :

$$\xi = \frac{(t_2 - t_1) \, \left\{ e^{-2(t_2 - t_1)} \, W(t_1) - W(t_2) \right\}}{\left\{ e^{-2(t_2 - t_1)} - 1 \right\} \{W(t_1) + W(t_2)\}} \; + \; t_1$$

Study of the supremum of the non linear interpolated process $Z(.)$:
Let consider the process $Z(.)$ on $[t_1, t_2]$. It verifies $\forall t \in [t_1, t_2]$:

$$Z(t) = \{\alpha(t) Z(t_1) \; + \; \beta(t) Z(t_2)\} \, / \sqrt{\{\alpha(t)\}^2 + \{\beta(t)\}^2 + 2 \, \alpha(t) \, \beta(t) \, e^{-2(t_2 - t_1)}}$$

Indeed, according to the proof of theorem 1 (cf. Section 5.1), $Z(.)$ has unit variance.
Using same notations as in theorem 2, let consider :

$$\gamma_1(t) = \alpha(t) \; , \; \gamma_2(t) = \beta(t) \; , \; C_1 = Z(t_1) \; , \; C_2 = Z(t_2) \; , \; \tilde{\rho} = e^{-2(t_2 - t_1)}$$

As previously, we will call $\gamma_3(t)$ the ratio $\frac{\gamma_2(t)}{\gamma_1(t) + \gamma_2(t)}$. We have $\gamma_3(t) = \frac{\beta(t)}{\alpha(t) + \beta(t)}$. To begin, we will admit that $\forall t \in [t_1, t_2]$, $0 \leqslant \gamma_3(t) \leqslant 1$. Besides, $\gamma_3(t_2) = 1$ and $\gamma_3(t_1) = 0$.

As a consequence, according to theorem 2 :

$$\sup_{t\in[t_1,t_2]} \{Z(t)\}^2 = \max \left[\{Z(t_1)\}^2 \ , \ \{Z(t_2)\}^2 \ , \right.$$
$$\left. \frac{\{Z(t_1)\}^2 + \{Z(t_2)\}^2 - 2\, e^{-2(t_2-t_1)}\ Z(t_1)\ Z(t_2)}{\left\{1+e^{-2(t_2-t_1)}\right\}\left\{1-e^{-2(t_2-t_1)}\right\}}\ 1_{\frac{Z(t_2)}{Z(t_1)}\ \in\ ]\ e^{-2(t_2-t_1)}\ ,\ e^{2(t_2-t_1)}\ [} \right]$$

To conclude, as $Z(t_1) = W(t_1)$ and $Z(t_2) = W(t_2)$:

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

The interest is now on the function $\gamma_3(t)$. $\forall t \in [t_1, t_2]$, we have :

$$\gamma_3(t) = 1 - \frac{Q_t^{1,1} + Q_t^{1,-1} - 1}{2Q_t^{1,1} - 1}$$

In order to proove that $\forall t \in]t_1, t_2[\ 0 < \gamma_3(t) < 1$, we will proove that $0 < \bar{\gamma}_3(t) < 1$ with $\bar{\gamma}_3(t) = (Q_t^{1,1} + Q_t^{1,-1} - 1)/(2Q_t^{1,1} - 1)$.
The calculation of the derivative of $Q_t^{1,1}$ shows that $Q_t^{1,1}$ is a decreasing function on $]t_1, (t_1+t_2)/2]$ and an increasing function on $[(t_1+t_2)/2, t_2[$. The minimum is $Q^{1,1}_{(t_1+t_2)/2}$. Since $Q^{1,1}_{(t_1+t_2)/2} = \left\{1 + e^{-2(t_1+t_2)} + 2e^{-t_1-t_2}\right\} / \left\{2 + 2e^{-2(t_2-t_1)}\right\}$, we have $Q^{1,1}_{(t_1+t_2)/2} > 1/2$. By continuity, $0 < 2Q_t^{1,1} - 1$. The focus is now on the numerator of $\bar{\gamma}_3(t)$. After calculations, we obtain :

$$\frac{\partial(Q_t^{1,1} + Q_t^{1,-1})}{\partial t} = \frac{-e^{-2(t-t_1)} - e^{-4t_2+2t+2t_1}}{2\ r(t_1,t_2)\ \bar{r}(t_1,t_2)}$$

As a consequence, $Q_t^{1,1} + Q_t^{1,-1} - 1$ is a decreasing function on $]t_1, t_2[$. By continuity, $0 < Q_t^{1,1} + Q_t^{1,-1} - 1$. So, $\bar{\gamma}_3(t) > 0$.
On the other hand, the study of the derivative of $Q_t^{1,1} - Q_t^{1,-1}$ shows that $Q_t^{1,1} > Q_t^{1,-1}$. It comes $\forall t \in]t_1, t_2[,\ \ 0 < \bar{\gamma}_3(t) < 1$ and $0 < \gamma_3(t) < 1$.

6. 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.

7. Appendix

7.1. 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 5.1)

7.2. Sketch of the proof of theorem 3

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 (10) of Section 5.1 is also suitable $t \in [t_1, t_K]\backslash\mathbb{T}_k$ because t is bounded. It comes, $\forall (t, t') \in [t_1, t_K]\backslash\mathbb{T}_k \times [t_1, t_K]\backslash\mathbb{T}_k$:

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

$\mathbb{E}\left[\left\{2p(t)-1\right\}^2\right]$ described in formula (9) of Section 5.1 has to be adapted to the "Marker interval".

$\forall (t, t') \in]t^\ell, t^r[^2$, the expression of $\mathbb{E}\left\{p(t)p(t')\right\}$ can be deduced from appendix 7.1 by adapting to the "Marker interval".

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

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

In the same way as what has been done in the proof of theorem 1 (cf. Section 5.1), $\forall (t, t^\star) \in [t_1, t_K]\backslash\mathbb{T}_k \times [t_1, t_K]\backslash\mathbb{T}_k$:

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

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 (14) of Section 5.1 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" :

$$\mathbb{E}\left[X(t^\star)\left\{2p(t)-1\right\}\right] = e^{-2|t-t^\star|} \tag{17}$$

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 :

$$S_n(t) = \left\{\ \alpha(t)\ S_n(t^\ell)\ +\ \beta(t)\ S_n(t^r)\ \right\} / \sqrt{\mathbb{E}\left[\left\{2p(t)-1\right\}^2\right]} \tag{18}$$

where $\alpha(t) = Q_t^{1,1} + Q_t^{1,-1} - 1$, $\beta(t) = Q_t^{1,1} - Q_t^{1,-1}$ and $\forall k\ \forall k'$, $\mathrm{Cov}_{\mathrm{H_0}}\left\{\mathrm{S_n(t_k)}, \mathrm{S_n(t_{k'})}\right\} = \mathrm{e}^{-2|\mathrm{t_k}-\mathrm{t_{k'}}|}$.

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 5.1, 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 $S_n(t)$:

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

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

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

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 $\delta = \min_{k\in\{1,\ldots,K-1\}}(\delta_k)$
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.

7.3. Linear interpolated process in presence of several markers

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

$$V_n(t) = \left\{ \frac{t^r - t}{t^r - t^\ell}\, S_n(t^\ell) \;+\; \frac{t - t^\ell}{t^r - t^\ell}\, S_n(t^r) \right\} / \sqrt{\tau(t)}$$

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 $\forall k\ \forall k'$, $\mathrm{Cov}_{\mathrm{H_0}}\left\{\mathrm{S_n(t_k)}, \mathrm{S_n(t_{k'})}\right\} = \mathrm{e}^{-2|\mathrm{t_k} - \mathrm{t_{k'}}|}$. In the same way as what has been done in Section 3.3, 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, $V_n(.)$ will be a good approximation if and only if the genetic markers are close to 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 5.1), 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 :

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

As previously, $W(.)$, the limiting process of $V_n(.)$, is named the linear interpolated process.

7.4. 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 3.3).

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.

7.4.1. 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 3.3 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 :

$$S_n(t) = \left\{\sqrt{\bar{r}(t_1,t_2)}\,(1 - 2Q_t^{-1,-1})\,G_n^1(t) + \sqrt{r(t_1,t_2)}\,(1 - 2Q_t^{-1,1})\,G_n^2(t)\right\} / \sqrt{\mathbb{E}\left[\{2p(t)-1\}^2\right]} \tag{19}$$

This formula is the corrected version of formula (10) of Chang et al. (2009) without approximations here. According to formula (19), 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 3.3).
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 (19), is not the same interpolation as presented in formula (5) of Section 3.2 of this paper. Our interpolation is more intuitive, because it is an interpolation between the test statistic on markers. Besides, it explains why the likelihood profiles (ie. the paths of the process $\Lambda_n(.)$) are smooth between markers.

7.4.2. 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 :

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

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

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

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 :

$$\mathrm{Cov}_{H_0}\left\{G_n^1(t), G_n^1(t')\right\} = \sqrt{\bar{r}(t^\ell,t^r)\;\bar{r}\,\{(t')^\ell,(t')^r\}}\;\; e^{-2\left\{(t')^\ell - t^r\right\}}$$

$$\mathrm{Cov}_{H_0}\left\{G_n^1(t), G_n^2(t')\right\} = \sqrt{\bar{r}(t^\ell,t^r)\;r\,\{(t')^\ell,(t')^r\}}\;\; e^{-2\left\{(t')^\ell - t^r\right\}}$$

$$\mathrm{Cov}_{H_0}\left\{G_n^2(t), G_n^1(t')\right\} = -\sqrt{r(t^\ell,t^r)\;\bar{r}\,\{(t')^\ell,(t')^r\}}\;\; e^{-2\left\{(t')^\ell - t^r\right\}}$$

$$\mathrm{Cov}_{H_0}\left\{G_n^2(t), G_n^2(t')\right\} = -\sqrt{r(t^\ell,t^r)\;r\,\{(t')^\ell,(t')^r\}}\;\; e^{-2\left\{(t')^\ell - t^r\right\}}$$

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. Howewer, their approximations don't affect the final results concerning the process.

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.

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.

Feingold, E., Brown, P.O., Siegmund, D. (1993). Gaussian models for genetic linkage analysis using complete high-resolution maps of idendity by descent. *Am. J. Human. Genet.*, **53**, 234-251.

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

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

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

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

Rabier, C-E. (2010). *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.

Siegmund, D. (1985). Sequential anlysis : tests and confidence intervals. *Springer, New York*.

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