



HAL
open science

The supremum of Chi-Square processes

Charles-Elie Rabier, Alan Genz

► **To cite this version:**

Charles-Elie Rabier, Alan Genz. The supremum of Chi-Square processes. *Methodology and Computing in Applied Probability*, 2014, 16 (3), pp.715-729. 10.1007/s11009-013-9331-1 . hal-00658577

HAL Id: hal-00658577

<https://hal.science/hal-00658577>

Submitted on 10 Jan 2012

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.

The supremum of Chi-Square processes

Charles-Elie Rabier · Alan Genz

Received: date / Accepted: date

Abstract We describe a lower bound for the critical value of the supremum of a Chi-Square process. This bound can be approximated using a MCQMC simulation. We compare numerically this bound with the upper bound given by Davies, only suitable for a regular Chi-Square process. In a second part, we focus a non regular Chi-Square process : the Ornstein-Uhlenbeck Chi-Square process. Recently, Rabier et al. (2009) have shown that this process has an application in genetics : it is the limiting process of the likelihood ratio test process related to the test of a gene on an interval representing a chromosome. Using results from Delong (1981), we propose a theoretical formula for the supremum of such a process and we compare it in particular with our simulated lower bound.

Keywords Chi-Square process · Monte-Carlo Quasi Monte-Carlo · Ornstein-Uhlenbeck process · Quantitative Trait Locus detection.

PACS 60G99 · 6008 · 65C05 · 65C10 · 65D30 · 62M86 · 62P10

1 The Davies Upper bound

In the article of Davies (1987), the focus is on hypothesis testing when a nuisance parameter t^* is present only under an alternative. So, t^* is meaningless under the null hypothesis. If t^* were known, the natural way to perform the test is to consider

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

Tel.: +(608)-265-8765, Fax: +(608)-262-0032

E-mail: rabier@stat.wisc.edu

Present address: University of Wisconsin-Madison, Statistic Department, Medical Science Center, 1300 University Avenue, Madison, WI 53706-1532, USA

Alan Genz

Department of Mathematics, Washington State University, Pullman, WA 99164-3113, USA

$t = t^*$. However, as it is only known that t belongs to the interval $[\mathcal{L}, \mathcal{U}]$, Davies suggests the use of the test statistic :

$$\sup \{S(t) : \mathcal{L} \leq t \leq \mathcal{U}\} \quad (1)$$

where $S(t)$ denotes the test statistic at t . Davies also considers the case :

$$S(t) = V_1(t)^2 + \dots + V_d(t)^2 \quad (2)$$

where the $V_i(t)$ are independent for each t and distributed as a standardized normal under the null hypothesis. The process $S(\cdot)$ is called a Chi-Square process with d degrees of freedom. The main results of Davies (1987) is the following formula :

$$\text{pr} \left\{ \sup_{t \in [\mathcal{L}, \mathcal{U}]} S(t) > c \right\} \leq \text{pr} (\chi_d^2 > c) + \int_{\mathcal{L}}^{\mathcal{U}} \Psi(t) dt \quad (3)$$

where

$$\Psi(t) = \mathbb{E} (\|\eta(t)\|) c^{\frac{d-1}{2}} e^{-c/2} \pi^{-1/2} 2^{-d/2} / \Gamma(d/2 + 1/2)$$

Γ is the Gamma function and χ_d^2 is a random variable which follows a Chi-Square with d degrees of freedom. We refer to Davies (1987) to obtain the general expression of the quantity $\mathbb{E} (\|\eta(t)\|)$. The author specifies that formula (3) is suitable when the processes $V_i(\cdot)$ have a derivative with a finite number of jumps.

In what follows, we will call Davies upper bound the right side of formula (3). We will focus only on Chi-Square processes $S(\cdot)$ where the $V_i(\cdot)$ are independent (a particular case of formula 2).

2 Computation of the Discretized Lower Bound

2.1 Introduction

In this section, we present a lower bound for the critical value of the supremum of a Chi-Square process. This is a lower bound because we discretize the process. The probabilities needed for the lower bounds can then be explicitly written as multivariate normal integrals, which can be approximated with simulation methods.

The time interval $[\mathcal{L}, \mathcal{U}]$ for the process is discretized using $t_i = \mathcal{L} + i(\mathcal{U} - \mathcal{L})/m$ for $i = 1, 2, \dots, m$. h will be the stepsize of discretization : $h = (\mathcal{U} - \mathcal{L})/m$. Then we define A to be the $m \times m$ covariance matrix for the discretized process, with entries $a_{ij} = r(t_i - t_j)$. If we also define X to be an $d \times m$ matrix, with columns \mathbf{x}_j , for $j = 1, \dots, m$, the integrals that are needed for computation of the lower bound are multivariate Normal probability integrals over a product of m d -dimensional spheres, given by

$$P(u) = \int_{\|\mathbf{x}_1\|^2 < u^2} \int_{\|\mathbf{x}_2\|^2 < u^2} \dots \int_{\|\mathbf{x}_m\|^2 < u^2} \frac{e^{-\frac{1}{2} \sum_{i=1}^d [x_{i1}, \dots, x_{im}] A^{-1} [x_{i1}, \dots, x_{im}]'}}{\{(2\pi)^m |A|\}^{\frac{d}{2}}} \prod_{i=1}^d \prod_{j=1}^m dx_{ij},$$

for real $u \geq 0$. If we determine u so that $P(u) = 1 - \alpha$, then u will be a lower bound for the critical value of the supremum of the Chi-Square process with covariance function $r(t)$.

In order to describe some simulation methods for approximation of the $P(u)$ integrals, we start with a change of variables designed to simplify the multivariate Normal density. Let L be the $m \times m$ lower triangular Cholesky factor for A (so that $A = LL^T$). Now define the change variables to an $d \times m$ matrix of variables Y by $X = YL^T$, so that $dX = |A|^{\frac{d}{2}} dY$ and therefore

$$P(u) = \int_{\|\mathbf{x}_1(Y)\|^2 < u^2} \int_{\|\mathbf{x}_2(Y)\|^2 < u^2} \cdots \int_{\|\mathbf{x}_m(Y)\|^2 < u^2} \frac{e^{-\frac{1}{2} \sum_{i=1}^d \sum_{j=1}^m y_{ij}^2}}{(2\pi)^{\frac{md}{2}}} \prod_{i=1}^d \prod_{j=1}^m dy_{ij}, \quad (4)$$

where $x_{ij}(Y) = \sum_{k=1}^j l_{jk} y_{ik}$, and $\mathbf{x}_j(Y)$ is a function of $\mathbf{y}_1, \dots, \mathbf{y}_j$.

2.2 Direct Simulation

A direct simulation method for approximating the integrals $I(u)$ uses simulation from the univariate Normal distribution. Let $Y_{ij}^{(k)} \sim N(0, 1)$, and define

$$f\{Y^{(k)}\} = \max_{1 \leq j \leq m} [\|\mathbf{x}_j\{Y^{(k)}\}\|] \quad \text{and} \quad g\{Y^{(k)}\} = \begin{cases} 1 & \text{if } f\{Y^{(k)}\} \leq u \\ 0 & \text{otherwise} \end{cases}.$$

Then

$$P(u) \approx P_N = \frac{1}{N} \sum_{k=1}^N g\{Y^{(k)}\}$$

with standard error

$$E_N = \left(\frac{1}{N(N-1)} \sum_{k=1}^N [g\{Y^{(k)}\} - P_N]^2 \right)^{\frac{1}{2}}.$$

Because the $g\{Y^{(k)}\}$ is 0 or 1, $E_N \approx \left[\frac{P(u)\{1-P(u)\}}{N} \right]^{\frac{1}{2}} \approx \left\{ \frac{P_N(1-P_N)}{N} \right\}^{\frac{1}{2}}$ (see Fishman (1996)).

If u_p is the value of u where $P(u) = p$ for a given probability p , an approximate value for u_p can easily be determined from this simulation. Define \mathbf{F} to be the vector of sorted (ascending order) $f\{Y^{(k)}\}$ values and define $[pN]$ to be the value of pN rounded to the nearest integer. Then $u_p \approx F_{[pN]}$.

2.3 Conditional Simulation

The direct simulation method described in the previous section is an "acceptance-rejection" algorithm which can be inefficient for some combinations of u and A , so a potentially more efficient method is considered in this section. This method is a generalization of the method described in the paper by Genz (1992), where MVN

probabilities over hyper-rectangular regions were considered. The method in this paper begins with equation (4) for $P(u)$ written in more detail in the form

$$P(u) = \int_{\sum_{i=1}^d (l_{11}y_{i1})^2 < u^2} \frac{e^{-\frac{1}{2} \sum_{i=1}^d y_{i1}^2}}{(2\pi)^{\frac{d}{2}}} \int_{\sum_{i=1}^d (l_{21}y_{i1} + l_{22}y_{i2})^2 < u^2} \frac{e^{-\frac{1}{2} \sum_{i=1}^d y_{i2}^2}}{(2\pi)^{\frac{d}{2}}} \dots \int_{\sum_{i=1}^d (l_{m1}y_{i1} + \dots + l_{mm}y_{im})^2 < u^2} \frac{e^{-\frac{1}{2} \sum_{i=1}^d y_{im}^2}}{(2\pi)^{\frac{d}{2}}} \prod_{j=m}^1 \prod_{i=d}^1 dy_{ij}. \quad (5)$$

In order to simplify the notation, we assume that A is nonsingular and define a matrix C , determined by scaling the rows of L by the successive diagonal elements of L , so that $c_{ij} = l_{ij}/l_{ii}$, which makes $c_{ii} = 1$. If the scaled sphere radii are defined by $u_i = u/l_{ii}$, then

$$P(u) = \int_{\sum_{i=1}^d y_{i1}^2 < u_1^2} \frac{e^{-\frac{1}{2} \sum_{i=1}^d y_{i1}^2}}{(2\pi)^{\frac{d}{2}}} \int_{\sum_{i=1}^d (c_{21}y_{i1} + y_{i2})^2 < u_2^2} \frac{e^{-\frac{1}{2} \sum_{i=1}^d y_{i2}^2}}{(2\pi)^{\frac{d}{2}}} \dots \int_{\sum_{i=1}^d (c_{m1}y_{i1} + \dots + c_{m-1,1}y_{i,m-1} + y_{im})^2 < u_m^2} \frac{e^{-\frac{1}{2} \sum_{i=1}^d y_{im}^2}}{(2\pi)^{\frac{d}{2}}} \prod_{j=m}^1 \prod_{i=d}^1 dy_{ij}.$$

The structure of the integration constraints permit the conditional integrations to be completed with many possible orderings for the variables. We describe a method which uses a variable ordering by rows, starting with $y_{11}, y_{12}, \dots, y_{1m}$, followed by $y_{21}, y_{22}, \dots, y_{2m}$, and so on, finishing with $y_{d1}, y_{d2}, \dots, y_{dm}$, as is indicated by the $\prod_{j=m}^1 \prod_{i=d}^1 dy_{ij}$ measure in equation (5).

The ‘‘outermost’’ variable y_{11} , has constraint $-u_1 < y_{11} < u_1$. Given y_{11} , the next variable y_{12} , has constraint $-u_2 - c_{11}y_{11} < y_{12} < u_2 - c_{11}y_{11}$, and so on, with y_{1j} constrained by $-u_j < \sum_{k=1}^{j-1} c_{jk}y_{1k} + y_{1j} < u_j$. If values are given for y_{ik} , $i = 1, \dots, l-1$, $k = 1, \dots, m$, and y_{lk} , $k = 1, \dots, j-1$, then the integral for y_{lj} has the constraint

$$\sum_{i=1}^{l-1} \left(\sum_{k=1}^{j-1} c_{jk}y_{ik} \right)^2 + \left(\sum_{k=1}^{j-1} c_{jk}y_{lk} + y_{lj} \right)^2 < u_j^2.$$

Solving for y_{lj} , the constraint for y_{lj} becomes

$$- \left\{ u_j^2 - \sum_{i=1}^{l-1} \left(\sum_{k=1}^{j-1} c_{jk}y_{ik} \right)^2 \right\}^{1/2} - \sum_{k=1}^{j-1} c_{jk}y_{lk} \leq y_{lj} \leq \left\{ u_j^2 - \sum_{i=1}^{l-1} \left(\sum_{k=1}^{j-1} c_{jk}y_{ik} \right)^2 \right\}^{1/2} - \sum_{k=1}^{j-1} c_{jk}y_{lk}. \quad (6)$$

So the limits for y_{ij} , which depend on

$$\mathbf{y}_{ij} = (y_{11}, y_{12}, \dots, y_{1,j-1}, y_{21}, y_{22}, \dots, y_{2,j-1}, \dots, y_{i1}, \dots, y_{i,j-1}),$$

are

$$M_{ij}^{\pm}(\mathbf{y}_{ij}) = \pm \left\{ u_j^2 - \sum_{s=1}^{i-1} \left(\sum_{k=1}^{j-1} c_{jk} y_{sk} \right)^2 \right\}^{1/2} - \sum_{k=1}^{j-1} c_{jk} y_{ik}.$$

Using these limit expressions,

$$\begin{aligned} P(u) &= \int_{M_{11}^-(\mathbf{y}_{11})}^{M_{11}^+(\mathbf{y}_{11})} \phi(y_{11}) \int_{M_{12}^-(\mathbf{y}_{12})}^{M_{12}^+(\mathbf{y}_{21})} \phi(y_{12}) \int_{M_{1m}^-(\mathbf{y}_{1m})}^{M_{1m}^+(\mathbf{y}_{1m})} \phi(y_{1m}) \\ &\quad \int_{M_{21}^-(\mathbf{y}_{21})}^{M_{21}^+(\mathbf{y}_{21})} \phi(y_{21}) \int_{M_{22}^-(\mathbf{y}_{22})}^{M_{22}^+(\mathbf{y}_{22})} \phi(y_{22}) \int_{M_{2m}^-(\mathbf{y}_{2m})}^{M_{2m}^+(\mathbf{y}_{2m})} \phi(y_{2m}) \\ &\quad \cdots \int_{M_{d1}^-(\mathbf{y}_{d1})}^{M_{d1}^+(\mathbf{y}_{d1})} \phi(y_{d1}) \int_{M_{d2}^-(\mathbf{y}_{d2})}^{M_{d2}^+(\mathbf{y}_{d2})} \phi(y_{d2}) \int_{M_{dm}^-(\mathbf{y}_{dm})}^{M_{dm}^+(\mathbf{y}_{dm})} \phi(y_{dm}) \prod_{j=m}^1 \prod_{i=d}^1 dy_{ij}, \end{aligned}$$

where $\phi(y) = e^{-\frac{1}{2}y^2} / (2\pi)^{\frac{1}{2}}$, the standard univariate Normal pdf. If the changes of variables $z_{ij} = \Phi(y_{ij})$, where $\Phi(y)$ is the standard univariate Normal cdf, with $dz_{ij} = \phi(y_{ij}) dy_{ij}$, are completed, then

$$\begin{aligned} I(u) &= \int_{N_{11}^-(\mathbf{z}_{11})}^{N_{11}^+(\mathbf{z}_{11})} \int_{N_{12}^-(\mathbf{z}_{12})}^{N_{12}^+(\mathbf{z}_{12})} \cdots \int_{N_{1m}^-(\mathbf{z}_{1m})}^{N_{1m}^+(\mathbf{z}_{1m})} \int_{N_{21}^-(\mathbf{z}_{21})}^{N_{21}^+(\mathbf{z}_{21})} \int_{N_{22}^-(\mathbf{z}_{22})}^{N_{22}^+(\mathbf{z}_{22})} \cdots \int_{N_{2m}^-(\mathbf{z}_{2m})}^{N_{2m}^+(\mathbf{z}_{2m})} \\ &\quad \cdots \int_{N_{d1}^-(\mathbf{z}_{d1})}^{N_{d1}^+(\mathbf{z}_{d1})} \int_{N_{d2}^-(\mathbf{z}_{d2})}^{N_{d2}^+(\mathbf{z}_{d2})} \cdots \int_{N_{dm}^-(\mathbf{z}_{dm})}^{N_{dm}^+(\mathbf{z}_{dm})} \prod_{j=m}^1 \prod_{i=d}^1 dz_{ij}. \end{aligned}$$

with $N_{ij}^{\pm}(\mathbf{z}_{ij}) = \Phi \left[M_{ij}^{\pm} \left\{ \Phi^{-1}(\mathbf{z}_{ij}) \right\} \right]$ and $\mathbf{z}_{ij} = (z_{11}, z_{12}, \dots, z_{1m}, \dots, z_{i1}, \dots, z_{i,j-1})$. Now make the final changes to $(0, 1)$ variables using

$$z_{ij} = N_{ij}^-(\mathbf{z}_{ij}) + D_{ij}(\mathbf{z}_{ij}) w_{ij}, \quad D_{ij}(\mathbf{z}_{ij}) = N_{ij}^+(\mathbf{z}_{ij}) - N_{ij}^-(\mathbf{z}_{ij}),$$

and then

$$\begin{aligned} P(u) &= \int_0^1 D_{11} \{ \mathbf{z}_{11}(\mathbf{w}_{11}) \} \cdots \int_0^1 D_{1m} \{ \mathbf{z}_{n1}(\mathbf{w}_{1m}) \} \\ &\quad \int_0^1 D_{21} \{ \mathbf{z}_{21}(\mathbf{w}_{21}) \} \cdots \int_0^1 D_{2m} \{ \mathbf{z}_{2m}(\mathbf{w}_{1m}) \} \\ &\quad \cdots \int_0^1 D_{d1} \{ \mathbf{z}_{d1}(\mathbf{w}_{d1}) \} \cdots \int_0^1 D_{dm} \{ \mathbf{z}_{nm}(\mathbf{w}_{dm}) \} \prod_{j=m}^1 \prod_{i=d}^1 dw_{ij} \\ &\equiv \int_0^1 f(\mathbf{w}) \prod_{j=m}^1 \prod_{i=d}^1 dw_{ij}, \end{aligned}$$

with

$$f(\mathbf{w}) = \prod_{i=1}^d \prod_{j=1}^m [N_{ij}^+ \{ \mathbf{z}_{ij}(\mathbf{w}_{ij}) \} - N_{ij}^- \{ \mathbf{z}_{ij}(\mathbf{w}_{ij}) \}],$$

so that $P(u)$ can be approximated using any numerical integration method for the unit hypercube $H^{(dm)} = [0, 1]^{dm}$.

If u_p is the value of u where $P(u) = p$ for a given probability p , an approximate value for u_p can be determined by applying a numerical root-finding method (e.g. the bisection or secant method) to function $h(u) = P(u) - p$.

2.4 Numerical Integration

A simple Monte Carlo (MC) method for the approximate computation of $P(u)$, which uses $U(0, 1)$ random numbers, takes the form

$$P(u) \approx P_N = \frac{1}{N} \sum_{k=1}^N f(\mathbf{W}_k),$$

with standard error

$$E_N = \left[\frac{1}{N(N-1)} \sum_{k=1}^N \{f(\mathbf{W}_k) - P_N\}^2 \right]^{\frac{1}{2}},$$

with all components of $\mathbf{W}_k \sim U(0, 1)$.

MC methods using N points have errors that are typically $O(1/N^{\frac{1}{2}})$, so quasi-Monte Carlo (QMC) methods (see Fox (1999), Sloan et al. (1994)), with asymptotic errors which can be approximately $O(1/N)$ for N points, are often used to provide improved simulation approximations for these kinds of computations. Given a set of QMC points $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_M$ from $H^{(dm)}$, a typical M -point QMC method for $P(u)$ uses

$$P(u) \approx Q_M = \frac{1}{M} \sum_{s=1}^M f(\mathbf{Z}_s).$$

Error estimates for Q_M can be computed if the QMC method is randomized. A simple method for randomization uses random shifts of the QMC approximations in the form

$$Q_M(\mathbf{W}) = \frac{1}{M} \sum_{s=1}^M f(\{\mathbf{Z}_s + \mathbf{W}\}^*),$$

where \mathbf{W} has random $U(0, 1)$ components and $\{\mathbf{X}\}^*$ denotes the vector of fractional parts of the components in \mathbf{X} . Then an MCQMC approximation for $P(u)$ is given by

$$P(u) \approx Q_{N,M} = \frac{1}{N} \sum_{k=1}^N Q_M(\mathbf{W}_k),$$

with standard error

$$E_{N,M} = \left[\frac{1}{N(N-1)} \sum_{k=1}^N \{Q_M(\mathbf{W}_k) - Q_{N,M}\}^2 \right]^{\frac{1}{2}}.$$

For these approximations, N is usually chosen to be small (e.g. $N = 12$) relative to M .

3 Illustration of Davies upper bound and the discretized lower bound

In order to illustrate Davies upper bound and the MCQMC lower bound, we have to choose a process suitable for Davies assumptions : we will consider the Chi-Square process where each $V_i(\cdot)$ is a stationary process with covariance function $r(\Delta) = e^{-\Delta^2/2}$. As mentioned previously, we consider independent processes $V_i(\cdot)$. Let $V'_i(t) = \partial V_i(t)/\partial t$. As the processes $V_i(\cdot)$ are independent, we have $\forall j \neq i$:

$$\text{cov} \{V'_i(t), V'_j(t)\} = 0 \text{ and } \text{cov} \{V'_i(t), V_j(t)\} = 0$$

Since $V'_i(t) = \lim_{\Delta \rightarrow 0} \frac{V_i(t+\Delta) - V_i(t)}{\Delta}$, then $\mathbb{E} \{V'_i(t)\} = 0$. Also, $r''(0) = -1$, so $\text{var} \{V'_i(t)\} = 1$. On the other hand,

$$\mathbb{E} \{V_i(t)V'_i(t)\} = \lim_{\Delta \rightarrow 0} \frac{\mathbb{E} \{V_i(t)V_i(t+\Delta)\}}{\Delta} - \lim_{\Delta \rightarrow 0} \frac{\mathbb{E} \{V_i^2(t)\}}{\Delta} = \lim_{\Delta \rightarrow 0} \frac{r(\Delta) - 1}{\Delta} = 0$$

As a consequence, if we denote $\mathbf{V}(t) = \begin{pmatrix} V_1(t) \\ \vdots \\ V_d(t) \end{pmatrix}$, $\mathbf{V}'(t) = \begin{pmatrix} V'_1(t) \\ \vdots \\ V'_d(t) \end{pmatrix}$ and I_{2d} the identity matrix $2d \times 2d$, then

$$\text{var} \begin{pmatrix} \mathbf{V}(t) \\ \mathbf{V}'(t) \end{pmatrix} = I_{2d} \quad (7)$$

As the covariance matrix (7) is the identity, we can use the following result of Davies :

$$\mathbb{E}(\|\eta\|) = 2^{1/2} \Gamma(d/2 + 1/2)/\Gamma(d/2)$$

Then, according to formula (3) :

$$\text{pr} \left\{ \sup_{t \in [\mathcal{L}, \mathcal{U}]} S(t) > c \right\} \leq \text{pr}(\chi_d^2 > c) + (\mathcal{U} - \mathcal{L}) c^{(d-1)/2} e^{-c/2} \pi^{-1/2} 2^{(1-d)/2} / \Gamma(d/2)$$

In order to perform a test at the level α , the challenge is to obtain the critical value c such as $\text{pr} \left\{ \sup_{t \in [\mathcal{L}, \mathcal{U}]} S(t) > c \right\} = \alpha$: the MCQMC bound and Davies' bound allows us to obtain respectively a lower and an upper bound for the critical value c .

Tables 1, 2 and 3 give these bounds as a function of the level of the test α , the interval $[\mathcal{L}, \mathcal{U}]$ and the number of degrees of freedom d of the Chi-Square process. The discretization used for the MCQMC method results in these tables was $h = \frac{1}{4}$ (so $m = 4(\mathcal{U} - \mathcal{L})$) for all cases.

4 About the supremum of a non regular Chi-Square process : the Ornstein-Uhlenbeck process

4.1 Introduction

In this section, the interest is on the critical value for the supremum of a non regular Chi-Square process : the Ornstein-Uhlenbeck Chi-Square process. It corresponds to the process $S(\cdot)$ of formula (2) with the processes $V_i(\cdot)$ independent and $\text{cov}\{V_i(t), V_i(t')\} = e^{-2|t-t'|}$. Such a process has a direct application in genetics, in particular in Quantitative Trait Locus (QTL) detection.

4.2 The Ornstein-Uhlenbeck Chi-Square process in the QTL detection

A QTL denotes a gene with quantitative effect on a trait. The method used by most of geneticists in order to detect a QTL on a chromosome, is the Interval Mapping proposed by Lander and Botstein (1989) (see Wu et al. (2007)). Using the Haldane (1919) distance and modelling, each chromosome is represented by a segment $[0, \mathcal{U}]$. The distance on $[0, \mathcal{U}]$ is called the genetic distance (measured in Morgans). At each location $t \in [0, \mathcal{U}]$, using the “genome information” brought by genetic markers, a likelihood ratio test (LRT) is performed, testing the presence of a QTL at this position. So, multi-testing leads to a LRT process, and taking as test statistic the supremum of this process comes down to perform a LRT in a model when the localisation of the QTL is an extra parameter.

In Rabier et al. (“Likelihood Ratio Test process for Quantitative Trait Loci detection”, hal-00421215), the authors considered a population of progenies which are structured into families of sires. They prove that when the number of genetic markers and the number of progenies tends to infinity, the limiting process of the LRT process is an Ornstein-Uhlenbeck Chi-Square process (the number of degrees of freedom corresponds to the number of sires) under the null hypothesis of the absence of QTL on the interval $[0, \mathcal{U}]$. So, in order to take decision about the presence of a QTL on $[0, \mathcal{U}]$, the critical value for the supremum of an Ornstein-Uhlenbeck Chi-Square process has to be calculated.

4.3 Critical value calculation for the Ornstein-Uhlenbeck Chi-Square process

Let OU denote an Ornstein-Uhlenbeck process and OUCS(d) an Ornstein-Uhlenbeck Chi-Square process with d degrees of freedom. We propose here different ways to calculate the critical value of the supremum of an OUCS(d). Since an OU is an AR(1) process, an OUCS(d) is the sum of d independent AR(1) processes. As a consequence, the critical value can easily be obtained using a Monte-Carlo method. On the other hand, an upper bound can be obtained using the MCQMC lower bound introduced in Section 2. Finally, we propose a formula in order to calculate the critical value theoretically.

Let $\mathbf{W}(t) = \begin{pmatrix} W_1(t) \\ \vdots \\ W_d(t) \end{pmatrix}$ a brownian motion in dimension d and $\mathbf{X}(t) = \begin{pmatrix} X_1(t) \\ \vdots \\ X_d(t) \end{pmatrix}$

the process such as $\forall t \forall i, X_i(t) = \frac{W_i(e^{2t})}{e^t}$. We can remark that $\text{cov}\{X_i(t), X_i(t')\} = e^{-|t-t'|}$ and that $W_i(t) = \sqrt{t}X_i\left\{\frac{\log(t)}{2}\right\}$.

Besides :

$$\|\mathbf{X}(t)\|^2 = e^{-2t} \|\mathbf{W}(e^{2t})\|^2$$

We can remark that $\text{cov}\{X_i(2t), X_i(2t')\} = e^{-2|t-t'|}$ which corresponds to the covariance of an OU. So, we impose $V_i(t) = X_i(2t)$. Let $T \in \mathbb{R}^{+\star}$, it comes :

$$\sup_{t \in [0, \frac{\log(T)}{4}]} S(t) = \sup_{t \in [0, \frac{\log(T)}{4}]} \|\mathbf{V}(t)\|^2 = \sup_{t \in [0, \frac{\log(T)}{2}]} \|\mathbf{X}(t)\|^2 = \sup_{t \in [1, T]} \left\{ \frac{\|\mathbf{W}(t)\|}{\sqrt{t}} \right\}^2 \quad (8)$$

Note that here \mathcal{L} and \mathcal{U} of formula (1) are respectively equal to 0 and $\log(T)/4$. This is convenient to deal with this case in order to relate with Delong (1981)'s work. Indeed, in Delong (1981), there are some important results about :

$$\text{pr} \left\{ \sup_{t \in [1, T]} \frac{\|\mathbf{W}(t)\|}{\sqrt{t}} < c \right\}$$

In order to calculate this quantity, Delong uses very difficult methods. His results are presented in some exact tables. As a consequence, using formula (8), it is easy to calculate exact critical values for the supremum of the process $S(\cdot)$.

In his article, Delong also gives an approximative formula (cf. page 2205 of the article) suitable for c and T large :

$$\text{pr} \left\{ \sup_{t \in [1, T]} \frac{\|\mathbf{W}(t)\|}{\sqrt{t}} < c \right\} = \frac{(c^2/2)^{d/2} e^{-c^2/2}}{\Gamma(d/2)} \left\{ \log(T) \left(1 - \frac{d}{c^2}\right) + \frac{2}{c^2} + O\left(\frac{1}{c^4}\right) \right\} \quad (9)$$

According to formulas (8) and (9), we can deduce an approximative formula for the process $S(\cdot)$ suitable for c and T large :

$$\text{pr} \left\{ \sup_{t \in [0, \frac{\log(T)}{4}]} S(t) < c \right\} = \frac{(c/2)^{d/2} e^{-c/2}}{\Gamma(d/2)} \left\{ \log(T) \left(1 - \frac{d}{c}\right) + \frac{2}{c} + O\left(\frac{1}{c^2}\right) \right\} \quad (10)$$

With the help of a Newton's method, it becomes easy to obtain the critical value c corresponding to a test at the α level, that is to say a test such as $\text{pr} \left\{ \sup_{t \in [0, \frac{\log(T)}{4}]} S(t) > c \right\} = \alpha$.

A numerical study is presented in Tables 4, 5, 6, and 7. Critical values for the process $S(\cdot)$ are calculated according to the different methods :

- DE refers to Delong Exact table (based on formula (8))
- DF refers to Delong Approximative Formula (based on formula (10))
- MC refers to the Monte-Carlo method using AR(1) processes
- MCQMC refers to the MCQMC lower bound

Note that Davies bound has not been considered since it is only suitable when the processes $V_i(\cdot)$ have a derivative with a finite number of jumps, which is not the case here.

The processes $S(\cdot)$ studied in Tables 4 to 7 are respectively the OUCS(4), the OUCS(5), the OUCS(6) and the OUCS(7). Note that since the exact tables of Delong are available only for $d \leq 4$, DE has only been computed in Table 4. The MCQMC bounds were computed using the discretization stepsize $h = \log(T)/256$ (so $m = 64$). Note that for MC, we also use a discretization ($h = 10^{-5}$) but as h is very small, it is closer to the continuous process.

According to Table 4, DF, DE and MC give in most of cases close results. However, MCQMC give smaller critical values due to the discretization. In the same way, according to Tables 5 to 7, DF and MC give close results whereas MCQMC give smaller critical values. As a consequence, we can say that DF is a very good approximation for finding the critical value of an OUCS process, which is not the case of the MC-QMC method. The problem is that the discretization stepsize h we use for MCQMC is too large. Table 8 shows for the OUCS(4) how the CPU time and the critical value increase when h decreases. We can observe that as expected the critical value is affected by the discretization, and the CPU time increases fairly rapidly with increasing m .

4.4 Why is our MCQMC critical value interesting for QTL detection ?

Until now, we have computed thresholds for the OUCS(d) process. The problem is that in real life, the number of genetic markers never tends to infinity. As a consequence, we will never have to consider the OUCS(d) process to compute thresholds. In fact, we have to deal with dense map, that is to say a map with a large (but not infinite) number of markers equally spaced on the chromosome. Since we perform tests at each marker location to find QTL, the process is not continuous anymore : it is a discrete process. So, we have to compute thresholds for the Discrete OUCS process and not for the continuous one. The stepsize of discretization will be the distance between consecutive markers. In such a context, the MCQMC critical value can be computed and it corresponds to the true threshold (contrary to the previous Section). Table 9 presents a comparison between MC and MCQMC thresholds. We study a map of size 1 Morgan and we study two configurations where markers are located every 1cM or every 2cM which is usual in genetics. We can see that MC and MC-QMC give almost same results, but we should rather use MCQMC since it is more accurate (cf. Section 2.4).

Acknowledgements The authors thank Jean-Marc Azaïs for having proposed this subject of research and fruitful discussions. This work has been supported by the French National Center for Scientific Research, the Animal Genetic Department of the French National Institute for Agricultural Research and SABRE.

References

- Davies, R.B., (1987) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **74**, 33-43
- Delong, D. M., (1981) Crossing probabilities for a square root boundary by a Bessel process. *Commun. Statist.-Theor. Meth.*, **A10(21)**, 2197-2213
- Fishman, G. S., (1996) *Monte Carlo: Concepts, Algorithms, and Applications*, Springer-Verlag, 68
- Fox, B. L., (1999) *Strategies for Quasi-Monte Carlo* (International Series in Operations Research & Management Science, 22), Kluwer Academic Publishers
- 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
- Sloan, I. H. and Joe, S. (1994) *Lattice Methods for Multiple Integration*. Oxford University Press
- Wu, R., MA, C.X., Casella, G. (2007) *Statistical Genetics of Quantitative Traits*, Springer

α		10%		5%		1%	
$[\mathcal{L}, \mathcal{U}]$	Method	MCQMC	DAV	MCQMC	DAV	MCQMC	DAV
	(0, 2)		6.28	6.87	8.34	8.39	11.72
(0, 5)		8.32	8.44	9.90	9.97	13.34	13.45
(0, 10)		9.67	9.81	11.22	11.33	14.76	14.80

Table 1 Bounds for the critical value c of the Chi-Square process with 2 degrees of freedom, as a function of the level of the test α and the interval $[\mathcal{L}, \mathcal{U}]$ considered. The upper bound refers to Davies' method (DAV) whereas the lower bound to MCQMC

α		10%		5%		1%	
$[\mathcal{L}, \mathcal{U}]$	Method	MCQMC	DAV	MCQMC	DAV	MCQMC	DAV
	(0, 2)		10.54	10.61	12.33	12.40	16.27
(0, 5)		12.34	12.48	14.15	14.24	18.03	18.15
(0, 10)		13.92	14.07	15.67	15.79	19.55	19.65

Table 2 Bounds for the critical value c of the Chi-Square process with 4 degrees of freedom, as a function of the level of the test α and the interval $[\mathcal{L}, \mathcal{U}]$ considered. The upper bound refers to Davies' method (DAV) whereas the lower bound to MCQMC

α		10%		5%		1%	
$[\mathcal{L}, \mathcal{U}]$	Method	MCQMC	DAV	MCQMC	DAV	MCQMC	DAV
	(0, 2)		15.34	15.45	17.44	17.53	21.94
(0, 5)		17.47	17.64	19.52	19.65	23.92	24.04
(0, 10)		19.26	19.46	21.27	21.41	25.69	25.71

Table 3 Bounds for the critical value c of the Chi-Square process with 7 degrees of freedom, as a function of the level of the test α and the interval $[\mathcal{L}, \mathcal{U}]$ considered. The upper bound refers to Davies' method (DAV) whereas the lower bound to MCQMC

α	Method	10%			5%			1%					
		DE	DF	MC	MCQMC	DE	DF	MC	MCQMC	DE	DF	MC	MCQMC
$(T, \frac{\log(T)}{4})$		14.21	14.09	14.21	13.28	16.16	16.10	16.08	15.16	20.43	20.58	20.16	19.27
(20, 0.75)		14.52	14.44	14.52	13.49	16.48	16.43	16.56	15.41	20.70	20.89	20.34	19.46
(30, 0.85)		14.67	14.66	14.67	13.65	16.65	16.65	16.65	15.58	20.88	21.10	20.70	19.63
(40, 0.92)		14.82	14.83	14.75	13.76	16.81	16.81	16.65	15.64	20.98	21.23	20.88	19.75
(50, 0.98)		14.90	14.95	14.82	13.86	16.89	16.93	16.65	15.71	21.07	21.34	21.25	19.84
(60, 1.02)		15.05	15.05	15.98	13.91	16.97	17.02	16.81	15.79	21.16	21.44	20.79	19.91
(70, 1.06)		15.13	15.14	15.13	13.97	17.06	17.11	17.06	15.85	21.25	21.51	21.44	19.98
(80, 1.10)		15.21	15.28	15.13	14.05	17.14	17.23	17.22	15.94	21.34	21.64	21.07	20.05
(100, 1.15)													

Table 4 Critical values for the OUCS(4) as a function of the level α of the test and the interval considered. The critical values c correspond to $\text{pr} \left\{ \sup_{t \in [0, \frac{\log(T)}{4}]} S(t) > c \right\} = \alpha$ where the process $S(\cdot)$ is an OUCS(4). These values are calculated by different methods : DE (Delong Exact table), DF (Delong Approximate Formula), MC (Monte-Carlo method using AR(1) processes), MCQMC. For the simulation of the four Autoregressive processes, 10000 paths have been generated with a step of 10^{-5}

α	Method	10%			5%			1%		
		DF	MC	MCQMC	DF	MC	MCQMC	DF	MC	MCQMC
$(T, \frac{\log(T)}{4})$		16.02	16.00	15.13	18.13	18.06	17.13	22.78	22.75	21.44
(20, 0.75)		16.39	16.48	15.37	18.48	18.58	17.38	23.11	23.23	21.65
(30, 0.85)		16.63	16.65	15.55	18.71	18.84	17.50	23.31	23.14	21.74
(40, 0.92)		16.80	16.81	15.62	18.86	18.92	17.66	23.46	23.14	21.89
(50, 0.98)		16.93	16.89	15.67	18.99	19.01	17.69	23.58	23.33	21.95
(60, 1.02)		17.04	17.06	15.81	19.09	19.10	17.75	23.67	23.52	22.00
(70, 1.06)		17.13	16.97	15.85	19.18	19.10	17.85	23.75	23.62	22.07
(80, 1.10)		17.27	17.22	15.95	19.32	19.18	17.94	23.87	23.62	22.16
(100, 1.15)										

Table 5 Critical values for the OUCS(5) as a function of the level α of the test and the interval considered. The critical values c correspond to $\text{pr} \left\{ \sup_{t \in [0, \frac{\log(T)}{4}]} S(t) > c \right\} = \alpha$ where the process $S(\cdot)$ is an OUCS(5). These values are calculated by different methods : DE (Delong Exact table), DF (Delong Approximate Formula), AR (Monte-Carlo method using AR(1) processes), MCQMC. For the simulation of the five Autoregressive processes, 10000 paths have been generated with a step of 10^{-5}

α	Method	10%			5%			1%		
		DF	MC	MCQMC	DF	MC	MCQMC	DF	MC	MCQMC
$(T, \frac{\log(T)}{4})$		17.86	17.81	16.90	20.05	19.89	18.99	24.86	24.40	23.44
(20, 0.75)		18.24	18.23	17.17	20.41	20.34	19.22	25.20	25.00	23.65
(30, 0.85)		18.49	18.20	17.29	20.65	20.52	19.36	25.41	24.90	23.85
(40, 0.92)		18.67	18.66	17.41	20.81	20.79	19.50	25.56	25.10	23.92
(50, 0.98)		18.81	18.66	17.52	20.94	20.70	19.61	25.68	25.10	23.99
(60, 1.02)		18.92	18.92	17.60	21.05	20.98	19.64	25.78	25.20	24.06
(70, 1.06)		19.02	18.84	17.69	21.13	21.07	19.77	25.86	25.20	24.12
(80, 1.10)		19.17	19.18	17.76	21.28	21.44	19.80	25.99	25.91	24.20
(100, 1.15)										

Table 6 Critical values for the OUCS(6) as a function of the level α of the test and the interval considered. The critical values c correspond to $\text{pr} \left\{ \sup_{t \in [0, \frac{\log(T)}{4}]} S(t) > c \right\} = \alpha$ where the process $S(\cdot)$ is an OUCS(6). These values are calculated by different methods : DE (Delong Exact table), DF (Delong Approximate Formula), MC (Monte-Carlo method using AR(1) processes), MCQMC. For the simulation of the six Autoregressive processes, 10000 paths have been generated with a step of 10^{-5}

α	Method	10%			5%			1%		
		DF	MC	MCQMC	DF	MC	MCQMC	DF	MC	MCQMC
$(T, \frac{\log(T)}{4})$		19.62	19.71	18.61	21.88	21.81	20.75	26.85	26.73	25.34
(20, 0.75)		20.03	19.80	18.87	22.27	22.09	21.03	27.20	26.52	25.53
(30, 0.85)		20.28	19.80	19.03	22.51	22.28	21.19	27.42	27.03	25.70
(40, 0.92)		20.47	20.52	19.13	22.68	22.56	21.28	27.57	27.25	25.80
(50, 0.98)		20.61	20.52	19.26	22.81	22.66	21.37	27.70	27.46	25.87
(60, 1.02)		20.72	20.52	19.35	22.92	22.66	21.40	27.79	27.20	25.95
(70, 1.06)		20.82	20.70	19.40	23.01	23.04	21.51	27.88	27.88	25.99
(80, 1.10)		20.98	20.79	19.47	23.16	23.04	21.61	28.02	27.67	26.08
(100, 1.15)										

Table 7 Critical values for the OUCS(7) as a function of the level α of the test and the interval considered. The critical values c correspond to $\Pr \left\{ \sup_{t \in [0, \frac{\log(T)}{4}]} S(t) > c \right\} = \alpha$ where the process $S(\cdot)$ is an OUCS(7). These values are calculated by different methods : DE (Delong Exact table), DF (Delong Approximate Formula), MC (Monte-Carlo method using AR(1) processes), MCQMC. For the simulation of the seven Autoregressive processes, 10000 paths have been generated with a step of 10^{-5}

h	m	c	CPU
0.047	16	12.47	10.76
0.023	32	12.88	20.87
0.012	64	13.30	46.82
0.006	128	13.61	125.23

Table 8 Critical values c and CPU time for the OUCS(4) using the MCQMC method and as a function of the discretization stepsize h ($\frac{\log(T)}{4} = 0.75$)

α (d, h, m)	10%		5%		1%	
	MC	MCQMC	MC	MCQMC	MC	MCQMC
(4, 0.02, 50)	13.65	13.72	15.47	15.43	19.66	19.74
(4, 0.01, 100)	13.99	14.04	15.91	15.75	19.94	19.72
(5, 0.02, 50)	15.57	15.33	17.45	17.48	21.70	21.88
(5, 0.01, 100)	15.86	15.88	17.90	17.62	22.23	22.16
(6, 0.02, 50)	17.36	17.31	19.36	19.30	23.67	23.65
(6, 0.01, 100)	17.67	17.87	19.78	19.50	24.29	23.72
(7, 0.02, 50)	19.03	19.01	21.08	21.45	25.81	25.09
(7, 0.01, 100)	19.42	19.57	21.55	21.74	26.34	26.03

Table 9 Critical values c for the DOUCS as a function of the number of degrees of freedom d , the discretization m , the level α and the method ($\frac{\log(T)}{4} = 1$)