Asymptotic distribution of the likelihood ratio test in QTL detection

Charles-Elie Rabier

Station d'Amélioration Génétique des Animaux, INRA Toulouse
Laboratoire statistiques et Probabilités, University Paul Sabatier (Toulouse 3)

Céline Delmas, Jean-Michel Elsen

INRA Toulouse

EAAP, Vilnius, August 2008

Outline

1. The dense map
   - Genome scan approach
   - Result of Lander & Botstein (1989) / Cierco (1996)
   - Population with family structure
   - Is it useful to consider the dense map ?
2. The sparse map
   - Test performed only on markers
3. Conclusion

Genome scan approach

H_0 : q=0 (no QTL) vs H_1 : q$\neq$0

- Likelihood ratio test at a position x :

$$T_x = 2ln\frac{L_x(\hat{\mu},\hat{q},\hat{\sigma}^2)}{L(\tilde{\mu},0,\tilde{\sigma}^2)}$$

- $T_{x_1}, T_{x_2}, ..., T_{x_K}$ define a process T(.)

Result of Lander & Botstein (1989) / Cierco (1996)

Hypothesis :

- all markers informative
- dense map
- N $\rightarrow \infty$ (N : number of individuals)

$\Rightarrow$ T(.) converges to the square of an Ornstein Uhlenbeck process under H_0

Definition (Ornstein Uhlenbeck process)

An O.U process is a gaussian stationary process, with mean equals to 0, variance equals to 1, covariance fonction equals to $r(t) = exp(-2 \mid t \mid)$

Population with family structure

- I sires each with J progenies
- s_i : polygenic effect of the sire i
- q_i : qtl effects of the allele present on the first chromosome of the sire i

$$H_0 : q_1 = ... = q_I = 0 \quad \text{vs} \quad H_1 : \exists\, q_i \neq 0$$

- Likelihood ratio test at a position x :

$$T_x = 2ln \frac{L_x(\hat{s_1}, ..., \hat{s_I}, \hat{q_1}, ..., \hat{q_I}, \hat{\sigma}^2)}{L(\tilde{s_1}, ..., \tilde{s_I}, 0, ..., 0, \tilde{\sigma}^2)}$$

Population with family structure

Hypothesis :

- all markers informative
- dense map
- J $\rightarrow \infty$ (J : number of progenies by sire)

$\Rightarrow$ T(.) converges to an Ornstein Uhlenbeck Chi Square process under H_0

Definition (Ornstein Uhlenbeck Chi Square process)

Let $Z^1(.), ..., Z^I(.)$ I independent O.U. process.
$Y(t) = \sum_{i=1}^{I} (Z^i(t))^2$ is named an O.U.C.S process

Ornstein Uhlenbeck

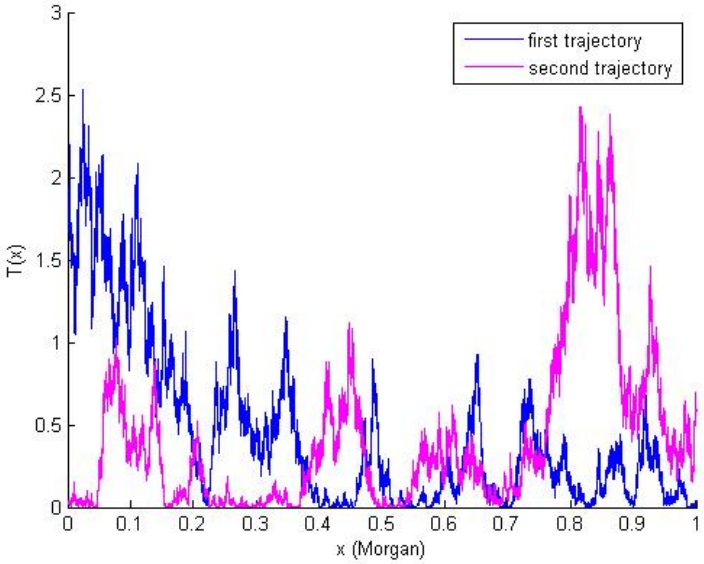

FIG.: 2 trajectories of the square of an O.U process

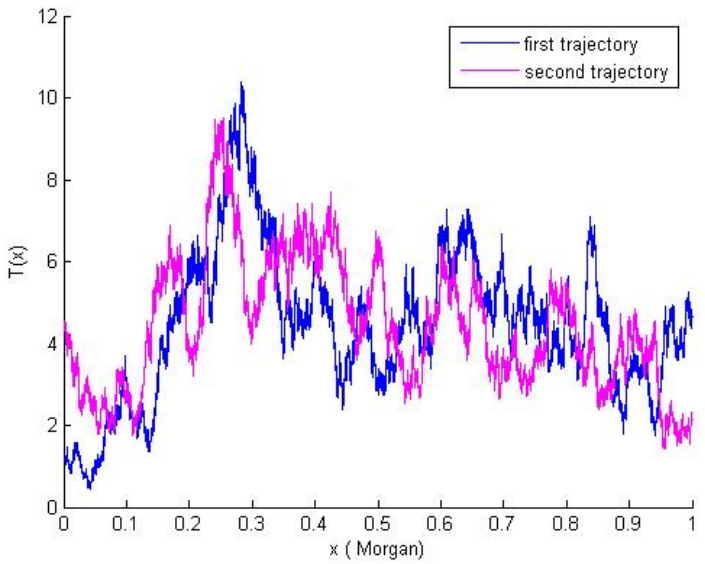

FIG.: 2 trajectories of an O.U.C.S process, I=5

Is it useful to consider the dense map ?

- Genome scan statistic

$$T_{GS} = sup_x T_x$$

Thresholds at the level 5% for T_{GS}
d : distance between markers (in Morgan)
100 000 trajectories, L=1M

d	O.U.C.S(1)	O.U.C.S(5)
10^{-5}	9.1696	18.8718
10^{-4}	9.0830	18.7576
10^{-3}	8.9230	18.5648

$\Rightarrow$ the dense map hypothesis is too strong

Test performed only on markers

- K markers (equally spaced for the illustration)
- d : distance between 2 adjacent markers
- $x_1, ..., x_K$: markers position
- only 1 family
- sparse map $\longrightarrow$ square of a **Discrete O.U. process**

$$\begin{bmatrix} T_{x_1} \\ \vdots \\ \vdots \\ T_{x_K} \end{bmatrix} \xrightarrow[J \to \infty]{H_0} W^2$$

with

$$W \sim N\left(\begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}, \begin{pmatrix} 1 & \rho & \dots & \rho^{K-1} \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho^{K-1} & \dots & \rho & 1 \end{pmatrix}\right) \text{ and } \rho = e^{-2d}$$

Test performed only on markers

I families $\Rightarrow$ Discrete O.U.C.S(I) process

- Application for I=5
  - d = 0.1 cM
  - L = 0.50 M

$\rightarrow$ 100 000 trajectories of the Discrete O.U.C.S(5) simulated

threshold=16.8391

$\rightarrow$ Thresholds obtained with classical methods
(40 000 populations simulated under H_0)

J	classical methods	real level
30	17.36	5.91%
100	16.94	5.15%
200	16.841	5.13%

$\rightarrow$ the Discrete O.U.C.S(5) goes 1680 times faster ! ! !

4 minutes vs almost 4 days

Test performed only on markers

- Application for I=5
  - d = 12 cM
  - L = 96 cM

$\longrightarrow$ 1 000 000 trajectories of the Discrete O.U.C.S(5) simulated

threshold=15.7129

$\longrightarrow$ Thresholds obtained with classical methods
(40 000 populations simulated under H_0)

J	classical methods
200	15.8246 (5.29%)
300	15.6817 (4.96%)
400	15.6612

Conclusion

To sum up :

- modeling the distribution of the test statistic as an O.U.C.S. process is very efficient in terms of computation time for finding rejection thresholds

Next work :

- Interval Mapping situation
- not fully informative markers

References

- « Mapping Mendelian Factors Underlying Quantitative Traits Using Linkage Maps »
  E.Lander, D.Botstein (Genetics 89)
- « Asymptotic distribution of the maximum likelihood ratio test for gene detection »
  C.Cierco (Statistics 98)