

Puissance et Robustesse des méthodes SNP par SNP dans des populations apparentées

Simon Teyssède¹ J-M. Elsen¹ A. Ricard²

¹INRA, UMR 1313, 78352 Jouy-en-Josas, France

²INRA, UR 631, 31326 Castanet-Tolosan, France

QTL-MAS Workshop 2010, Poznan, Poland
simon.teyssedre@toulouse.inra.fr



Outline

- 1 Introduction
 - Context
 - Objectives
 - Methods tested
- 2 Materials & Methods
 - Model and test
 - Model on SNP
- 3 Results
 - Regression
 - GRAMMAR
 - QTDT
- 4 Discussion & On going



Outline

- 1 Introduction
 - Context
 - Objectives
 - Methods tested
- 2 Materials & Methods
 - Model and test
 - Model on SNP
- 3 Results
 - Regression
 - GRAMMAR
 - QTDT
- 4 Discussion & On going

Context

- What we know :
 - Animal populations are often populations with related individuals
 - Conventional methods (LD) make the assumption that there is no relatedness between individuals \Rightarrow Population structure will affect the robustness of the methods
 - People try to control the population structure in their models



Context

- What we know :
 - Animal populations are often populations with related individuals
 - Conventional methods (LD) make the assumption that there is no relatedness between individuals \Rightarrow Population structure will affect the robustness of the methods
 - People try to control the population structure in their models
- But we often make the first analysis from simple methods (ex: simple regression) that do not correct for population structure (fast, easy to implement)



Context

- What we know :
 - Animal populations are often populations with related individuals
 - Conventional methods (LD) make the assumption that there is no relatedness between individuals \Rightarrow Population structure will affect the robustness of the methods
 - People try to control the population structure in their models
- But we often make the first analysis from simple methods (ex: simple regression) that do not correct for population structure (fast, easy to implement)
- Objective : Provide tools to show that some of these first analysis may give erroneous results in populations with related individuals

Objectives

- **Objective**: Evaluate algebraically the robustness and power of some methods from known parental structure (h^2 , pedigree, Nb individuals ...)
- **Hypothesis**: We assume that the true model is:

$$y = 1\mu + x\beta + Zu + e$$

with Y the vector of records, μ the overall mean, β the SNP effect, u the random polygenic effect with $u \sim N(0, A\sigma_u^2)$, A the relationship matrix and e the residuals with $e \sim N(0, I\sigma_e^2)$

- **Idea**: The user, who doesn't know the reality and therefore the true model, is wrong and uses a different model (for example, he uses the same model without the polygenic effect (Zu))
- **Question**: how many false positives should be expected if we use this simple regression? Will there be sufficient power to detect a QTL?



Methods tested

True model :

$$y = 1\mu + \mathbf{x}\beta + \mathbf{Z}u + \mathbf{e}$$

We tested 3 different methods (all are single SNP analysis) :

① Simple regression (LD):

$$y = 1\alpha + \mathbf{x}\beta + \epsilon$$

② GRAMMAR (LD): Aulchenko & al, Genetics, 2007

$$\begin{cases} y = 1\mu_1 + \mathbf{Z}u + \epsilon_1 \\ \hat{\epsilon}_1 = 1\mu_2 + \mathbf{x}\beta + \epsilon_2 \end{cases}$$

③ QTDT (LDLA): Abecasis & al, Am. J. Hum. Genet., 2000

$$y = 1\alpha + \frac{\mathbf{x}_p + \mathbf{x}_m}{2}\beta_b + \left(\mathbf{x} - \frac{\mathbf{x}_p + \mathbf{x}_m}{2}\right)\beta_w + \epsilon$$



Outline

- 1 Introduction
 - Context
 - Objectives
 - Methods tested
- 2 **Materials & Methods**
 - Model and test
 - Model on SNP
- 3 Results
 - Regression
 - GRAMMAR
 - QTDT
- 4 Discussion & On going



Regression model and student test (iid)

Regression model:

$$y = \mathbf{1}\alpha + \mathbf{x}\beta + \epsilon$$

Student test:

$$\frac{\hat{\beta} - E[\hat{\beta}]}{\sqrt{V(\hat{\beta})}} \sqrt{\frac{E[\hat{\epsilon}'\hat{\epsilon}]}{\hat{\epsilon}'\hat{\epsilon}}} \sim t_{N-2}$$

And in the case of i.i.d we have :

$$\begin{cases} E[\hat{\beta}] = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'E[\mathbf{y}] = \beta \\ V(\hat{\beta}) = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'V(\mathbf{y})\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1} = (\mathbf{x}'\mathbf{x})^{-1}\sigma_{\epsilon}^2 \\ E[\hat{\epsilon}'\hat{\epsilon}] = (N-2)\sigma_{\epsilon}^2 \end{cases}$$

⇒ For testing $\beta = 0$ against $\beta \neq 0$, we then use :

$$\frac{\hat{\beta} - \beta}{\sqrt{\hat{\epsilon}'\hat{\epsilon}}} \sqrt{(N-2)(\mathbf{x}'\mathbf{x})} \sim t_{N-2} \xrightarrow{N \rightarrow \infty} \begin{cases} N(0,1)_{H_0} \\ N\left(\frac{\beta}{\sqrt{V(\hat{\beta})}}, 1\right)_{H_1} \end{cases}$$



Regression model and student test (no iid)

Student test :

$$\frac{\hat{\beta} - E[\hat{\beta}]}{\sqrt{V(\hat{\beta})}} \sqrt{\frac{E[\hat{\epsilon}'\hat{\epsilon}]}{\hat{\epsilon}'\hat{\epsilon}}} \sim t_{N-2}$$

And now we have :

$$\begin{cases} E[\hat{\beta}] = (x'x)^{-1}x'E[y] = \beta \\ V(\hat{\beta}) = (x'x)^{-1}x'V(y)x(x'x)^{-1} = (x'x)^{-1}\sigma_{\epsilon}^2 + (x'x)^{-1}x'Ax(x'x)^{-1}\sigma_u^2 \\ E[\hat{\epsilon}'\hat{\epsilon}] = (N-2)\sigma_{\epsilon}^2 + (\text{tr}(A) - (x'x)^{-1}x'Ax - \frac{1}{N}1'A1)\sigma_u^2 \end{cases}$$

⇒ For testing $\beta = 0$ against $\beta \neq 0$, we then use :

$$\frac{\hat{\beta} - \beta}{\sqrt{\hat{\epsilon}'\hat{\epsilon}}} \sqrt{(N-2)(x'x) \frac{(1-h^2) + \frac{h^2}{N-2}(\text{tr}(A) - (x'x)^{-1}x'Ax - \frac{1}{N}1'A1)}{(1-h^2) + x'Ax(x'x)^{-1}h^2}}$$

$$t_{\text{user}} \sqrt{\lambda} \xrightarrow{N \rightarrow \infty} \begin{cases} N(0,1)_{H_0} \\ N(\frac{\beta}{\sqrt{V(\hat{\beta})}}, 1)_{H_1} \end{cases}$$



Regression model and student test (no iid)

So the user use the test :

$$t_{user} \xrightarrow{N \rightarrow \infty} \begin{cases} N(0, 1/\lambda)_{H_0} \\ N\left(\frac{\beta}{\sqrt{V(\hat{\beta})\lambda}}, 1/\lambda\right)_{H_1} \end{cases}$$

with :

$$\lambda = \frac{(1 - h^2) + \frac{h^2}{N-2} (tr(A) - (x'x)^{-1}x'Ax - \frac{1}{N}1'A1)}{(1 - h^2) + x'Ax(x'x)^{-1} - h^2}$$

Now, we want to : Explain λ as a function of known parameters.

But : We don't know x

So : We need to evaluate $E_x[\lambda]$



Regression model and student test (no iid)

We want to obtain :

$$E_x[\lambda] = \frac{(1 - h^2) + \frac{h^2}{N-2}(tr(A) - E_x[(x'x)^{-}]E_x[x'Ax]) - \frac{1}{N}1'A1)}{(1 - h^2) + E_x[x'Ax]E_x[(x'x)^{-}]h^2}$$

- Let $w = \{-2p, (1 - 2p), 2q\} / \sqrt{2pq}$ for genotype $\{11, 12, 22\}$
- By writing x as $x = w - \bar{w}$, and as $E(w_i w_j) = a_{ij}$ the relationship coefficient between individuals i and j , we found that :

$$E_x[x'x] = tr(A) - \frac{1}{n}(1'A1) = (n - 1)(\mu_D - \mu_O)$$

$$E_x[x'Ax] = tr(A'A) - \frac{2}{n}1'A'A1 + \frac{1}{n^2}1'A11'A1 \approx n^2 V_A$$

With :

- D is the diagonal of the relationship matrix A
- O is the out diagonal of the relationship matrix A



Conclusion : Regression

So the user use the test :

$$t_{user} \xrightarrow{N \rightarrow \infty} \begin{cases} N(0, 1/\lambda)_{H_0} \\ N\left(\frac{\beta}{\sqrt{V(\hat{\beta})\lambda}}, 1/\lambda\right)_{H_1} \end{cases}$$

with :

$$E_x[\lambda] = \frac{(1 - h^2) + C_2 h^2}{(1 - h^2) + C_1 h^2}$$

$$E_x\left[\frac{\beta}{\sqrt{V(\hat{\beta})\lambda}}\right] = \frac{\sqrt{(n-1)(\mu_D - \mu_O)}\beta/\sigma_y^2}{\sqrt{(1 - h^2) + C_2 h^2}}$$

and :

$$\begin{cases} C_1 = (\mu_D - \mu_O) + \frac{N[V_D + (N-1)V_O]}{(N-1)(\mu_D - \mu_O)} \\ \text{and} \\ C_2 = (\mu_D - \mu_O) - \frac{N[V_D + (N-1)V_O]}{(N-2)(N-1)(\mu_D - \mu_O)} \end{cases}$$

D the diagonal of the relationship matrix A

O the out diagonal of the relationship matrix A



Outline

- 1 Introduction
 - Context
 - Objectives
 - Methods tested
- 2 Materials & Methods
 - Model and test
 - Model on SNP
- 3 Results
 - Regression
 - GRAMMAR
 - QTDT
- 4 Discussion & On going



Résultats

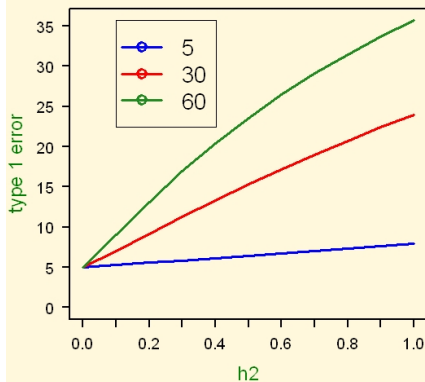
Population testée :

- 600 individus
- Familles avec 5, 30 et 60 descendants
- Fondateurs non apparentés
- Héritabilité comprise entre 0 et 1



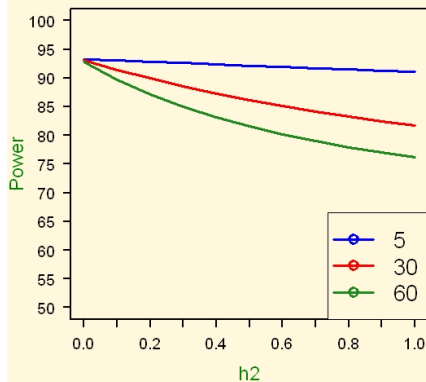
Regression

Robustness Regression



⇒ ↗ structure, ↗ FDR

Power Regression $\beta=0.14$

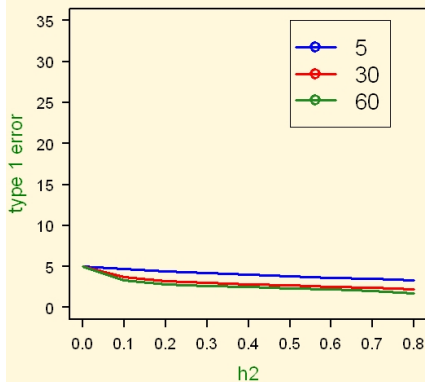


⇒ ↗ structure, ↘ Power



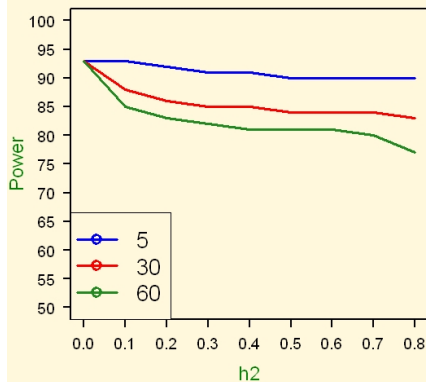
GRAMMAR

Robustness GRAMMAR



⇒ ↗ structure, ↘ FDR

Power GRAMMAR $\beta=0.14$

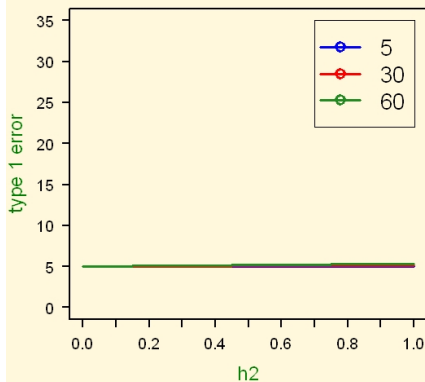


⇒ ↗ structure, ↘ Power

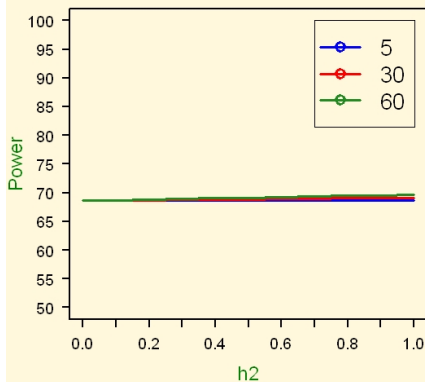


QTD

Robustness QTD



⇒ ↗ structure, = FDR

Power QTD $\beta=0.14$ 

⇒ ↗ structure, = Power



Outline

- 1 Introduction
 - Context
 - Objectives
 - Methods tested
- 2 Materials & Methods
 - Model and test
 - Model on SNP
- 3 Results
 - Regression
 - GRAMMAR
 - QTDT
- 4 Discussion & On going



Conclusion

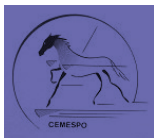
- **Objective : Provide tools to show that some of these first analysis may give erroneous results in populations with related individuals**
- Regression give a lot of problems when there is a parental structure between individuals
- GRAMMAR is more stable but have some problems with the power
- QTDT is stable and power is unaffected, but power isn't high

Acknowledgements

GENEQUIN funders :



GENEQUIN partners :



Thank you for your attention !

