Electronic Journal of Statistics
Vol. 0 (0000)
ISSN: 1935-7524
DOI: 10.1214/154957804100000000

Heritability estimation in high dimensional sparse linear mixed models

Anna Bonnet Elisabeth Gassiat and Céline Lévy-Leduc

Laboratoire de Mathématiques d'Orsay
Université Paris-Sud - 91405 Orsay Cedex, France
e-mail: elisabeth.gassiat@math.u-psud.fr

AgroParisTech - UMR INRA MIA 518
16, Rue Claude Bernard - 75005 Paris
e-mail: anna.bonnet@agroparistech.fr; celine.levy-leduc@agroparistech.fr

Abstract:
Motivated by applications in genetic fields, we propose to estimate the heritability in high-dimensional sparse linear mixed models. The heritability determines how the variance is shared between the different random components of a linear mixed model. The main novelty of our approach is to consider that the random effects can be sparse, that is may contain null components, but we do not know either their proportion or their positions. The estimator that we consider is strongly inspired by the one proposed by Pirinen, Donnelly and Spencer (2013), and is based on a maximum likelihood approach. We also study the theoretical properties of our estimator, namely we establish that our estimator of the heritability is $\sqrt{n}$-consistent when both the number of observations n and the number of random effects N tend to infinity under mild assumptions. We also prove that our estimator of the heritability satisfies a central limit theorem which gives as a byproduct a confidence interval for the heritability. Some Monte-Carlo experiments are also conducted in order to show the finite sample performances of our estimator.

Keywords and phrases: linear mixed model, heritability, sparsity, high dimension.

Contents

1. Introduction

Linear mixed models (LMMs) have been widely used in various fields such as agriculture, biology, medicine and genetics. In quantitative genetics, LMMs have been used for estimating the heritability of traits and breeding values as explained for instance by Lynch and Walsh (1998). In Genome Wide Association Studies (GWAS), which is the application field that inspired our work, Yang et al. (2011) suggested the use of linear mixed models to measure genotypes at a large number of single nucleotide polymorphisms (SNPs) in large samples of individuals in order to identify genetic variants that explain variations in phenotypes.

The model that we shall study in this paper is a LMM defined as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}\ , \tag{1}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ is the vector of observations, $\mathbf{X}$ is a $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector containing the unknown linear effects of the predictors, and $\mathbf{u}$ and $\mathbf{e}$ correspond to the random effects. Moreover, in (1), $\mathbf{Z}$ is a $n \times N$ random matrix which will be further described in Section 2.

We shall assume that the random effects can be sparse, that is only a proportion q of the components of $\mathbf{u}$ are non-zero:

$$u_i \overset{i.i.d.}{\sim} (1-q)\delta_0 + q\mathcal{N}(0, \sigma_u^{\star 2})\ \text{, for all } 1 \leqslant i \leqslant N \text{ and } \mathbf{e} \sim \mathcal{N}\left(0, \sigma_e^{\star 2}\mathrm{Id}_{\mathbb{R}^n}\right)\ , \tag{2}$$

where $\mathrm{Id}_{\mathbb{R}^n}$ denotes the $n \times n$ identity matrix, q is in $(0, 1]$, and δ_0 is the point mass at 0. Note that this corresponds to a more general situation than the usual assumption of (non-sparse) Gaussian random effects which is recovered when $q = 1$.

The use of linear mixed models to estimate heritability has been proposed by Yang et al. (2011) as an alternative to the regression models usually used in GWAS. The goal is to consider the joint effect of all SNPs on a phenotype, and the heritability corresponds to the proportion of phenotypic variance explained by all SNPs.

In the GWAS framework, $\mathbf{Z}$ is thus a matrix having a number of rows equal to the number of individuals in the experiment that is $n \approx 1000$ and a number of columns equal to the number of SNPs taken into account in the experiment, namely $N \approx 500,000$. This application motivated the framework that we chose where n and N tend to infinity.

The major difference between the framework of Yang et al. (2011) and ours is that they consider that the random effects are Gaussian while we consider a mixture model between a point mass at 0 and a Gaussian distribution. With this modeling, we assume

that all SNPs are not necessarily causal, that is that all SNPs do not explain a given phenotype.

Our main goal in this paper is to propose an estimator for the heritability in this possibly sparse framework and to establish its theoretical properties in the non standard theoretical context where n and N tend to infinity.

In this paper, we prove that using a strategy close to the one proposed by Pirinen, Donnelly and Spencer (2013), which has been devised in the case $q = 1$, provides consistent estimators even in the case where $q < 1$. Moreover, we prove that this estimator is $\sqrt{n}$-consistent in the following asymptotic framework: $n \to \infty$ and $N \to \infty$ such as $n/N \to a > 0$ and satisfies under mild assumptions a central limit theorem in both cases $q = 1$ and $q < 1$. It has to be noticed that the classical results that exist in linear mixed models are established only in the case where $q = 1$, n tends to infinity and N is constant.

The paper is organized as follows. Section 2 provides a detailed description of the model and the heritability estimator that we propose. Section 3 reviews existing methods for heritability estimation. Section 4 is dedicated to the theoretical properties of our estimator. The numerical results are presented in Section 5. They have been obtained thanks to the R package HiLMM that we have developed and which is available from the Comprehensive R Archive Network (CRAN). In Section 6, we provide some additional comments on our work as well as some prospects such as the estimation of the proportion q of non null components in the random effects. Finally, the proofs are given in Section 7.

2. Model and heritability estimator

2.1. Model

In the sequel, up to considering the projection of $\mathbf{Y}$ onto the orthogonal of the image of $\mathbf{X}$ and for notational simplicity, we shall focus on the following model

$$\mathbf{Y} = \mathbf{Z}\mathbf{u} + \mathbf{e}\ , \tag{3}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ is the vector of observations, $\mathbf{u}$ and $\mathbf{e}$ correspond to the random effects, which are defined in (2). Moreover, $\mathbf{Z}$ is a $n \times N$ random matrix such that the $Z_{i,j}$ are normalized random variables in the following sense: they are defined from a matrix $\mathbf{W} = (W_{i,j})_{1 \leqslant i \leqslant n,\, 1 \leqslant j \leqslant N}$ by

$$Z_{i,j} = \frac{W_{i,j} - \overline{W}_j}{s_j},\ i = 1, \ldots, n,\ j = 1, \ldots, N\ , \tag{4}$$

where

$$\overline{W}_j = \frac{1}{n}\sum_{i=1}^{n} W_{i,j},\ s_j^2 = \frac{1}{n}\sum_{i=1}^{n}(W_{i,j} - \overline{W}_j)^2,\ j = 1, \ldots, N\ . \tag{5}$$

In (4) and (5) the $W_{i,j}$'s are such that for each j in $\{1, \ldots, N\}$ the $(W_{i,j})_{1 \leqslant i \leqslant n}$ are independent and identically distributed random variables and such that the columns of

$\mathbf{W}$ are independent. With this definition the columns of $\mathbf{Z}$ are empirically centered and normalized.

In genetic applications, the matrix $\mathbf{W}$ contains all the genetic information about all the individuals in the study. More precisely, for each j, the $(W_{i,j})_{1\leqslant i\leqslant n}$ are i.i.d binomial random variables with parameters 2 and p_j. $W_{i,j} = 0$ (resp. 1, resp. 2) if the genotype of the ith individual at locus j is qq (resp. Qq, resp. QQ) where p_j is the frequency of Q allele at locus j.

In Model (1) with (4), (5), (2), one can observe that

$$\mathrm{Var}(\mathbf{Y}|\mathbf{Z}) = Nq\sigma_u^{\star 2}\mathbf{R} + \sigma_e^{\star 2}\mathrm{Id}_{\mathbb{R}^n}\ , \text{ where } \mathbf{R} = \frac{\mathbf{Z}\mathbf{Z}'}{N} \text{ and } q \text{ is defined in (2)}\ .$$

Inspired by Pirinen, Donnelly and Spencer (2013), Model (1) can be rewritten by using the following parameters:

$$\sigma^{\star 2} = Nq\sigma_u^{\star 2} + \sigma_e^{\star 2} \text{ and } \eta^\star = \frac{Nq\sigma_u^{\star 2}}{Nq\sigma_u^{\star 2} + \sigma_e^{\star 2}}\ . \tag{6}$$

Thus,

$$\mathrm{Var}(\mathbf{Y}|\mathbf{Z}) = \eta^\star\sigma^{\star 2}\mathbf{R} + (1-\eta^\star)\sigma^{\star 2}\mathrm{Id}_{\mathbb{R}^n}\ .$$

The parameter $\eta^\star$ which belongs to $[0,1]$ is commonly called the heritability in the case where $q = 1$, see for instance Yang et al. (2010), and determines how the variance is shared between $\mathbf{u}$ and $\mathbf{e}$ when all the components of $\mathbf{u}$ are non zero. We propose in (6) to extend this definition to the case where $\mathbf{u}$ may contain null components and q is in $(0,1]$. The parameter q actually corresponds to the proportion of non null components in $\mathbf{u}$ that is to the proportion of causal SNPs. Then, the heritability defined by $\eta^\star$ in (6) corresponds to the proportion of phenotypic variance explained by the causal variants.

2.2. Heritability estimator

In the case where $q = 1$, observe that

$$\mathbf{Y}|\mathbf{Z} \sim \mathcal{N}\left(0, \eta^\star\sigma^{\star 2}\mathbf{R} + (1-\eta^\star)\sigma^{\star 2}\mathrm{Id}_{\mathbb{R}^n}\right),$$

where $\eta^\star$ and $\sigma^\star$ are defined in (6).

Let $\mathbf{U}$ as the orthogonal matrix ($\mathbf{U}'\mathbf{U} = \mathbf{U}\mathbf{U}' = \mathrm{Id}_{\mathbb{R}^n}$) such that $\mathbf{U}\mathbf{R}\mathbf{U}' = \mathrm{diag}(\lambda_1,\ldots,\lambda_n)$ is a diagonal matrix having its diagonal entries equal to $\lambda_1,\ldots,\lambda_n$. Hence, in the case where $q = 1$ and conditionally to $\mathbf{Z}$, $\widetilde{\mathbf{Y}} = \mathbf{U}'\mathbf{Y}$ is a zero-mean Gaussian vector having a covariance matrix equal to $\mathrm{diag}(\eta^\star\sigma^{\star 2}\lambda_1 + (1-\eta^\star)\sigma^{\star 2},\ldots,\eta^\star\sigma^{\star 2}\lambda_n + (1-\eta^\star)\sigma^{\star 2})$, where the λ_i's are the eigenvalues of $\mathbf{R}$.

The method proposed by Pirinen, Donnelly and Spencer (2013) consists in computing the log-likelihood

$$L_n(\sigma^2,\eta) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^n \log(\eta(\lambda_i - 1) + 1) - \frac{1}{2\sigma^2}\sum_{i=1}^n \frac{\widetilde{Y}_i^{\,2}}{\eta(\lambda_i-1)+1} - \frac{n}{2}\log(2\pi)$$

and to maximize this function of two variables by iterative optimization techniques. Since in our case we are only interested in estimating $\eta^\star$, we plugged an estimator of $\sigma^{\star 2}$ that is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \frac{\widetilde{Y}_i^{\,2}}{\eta(\lambda_i - 1) + 1}$$

in L_n. Thus, in the case $q = 1$, the maximum likelihood strategy would lead to estimate $\eta^\star$, assumed to be in $[0, 1-\delta]$ with $\delta > 0$, by $\hat{\eta}$ defined as a maximizer of

$$L_n(\eta) = -\log\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\widetilde{Y}_i^2}{\eta(\lambda_i - 1) + 1}\right) - \frac{1}{n}\sum_{i=1}^{n} \log\left(\eta(\lambda_i - 1) + 1\right) \ , \qquad (7)$$

where the $\widetilde{Y}_i$'s are the components of the vector $\widetilde{\mathbf{Y}} = \mathbf{U}'\mathbf{Y}$.

We shall establish in Theorem 2, which is proved in Section **??**, that this strategy produces $\sqrt{n}$-consistent estimators of $\eta^\star$ in both cases: $q = 1$ and $q < 1$ and also that this estimator satisfies a central limit theorem which provides as a by-product confidence intervals for $\eta^\star$.

3. Existing methods for heritability estimation

Several approaches can be used for estimating the heritability in the case where $q = 1$ but to the best of our knowledge, no theoretical results concerning the estimation of the heritability or the estimation of $\sigma_u^\star$, $\sigma_e^\star$ have been established in the framework where both n and N tend to infinity. This is one of the contributions of our paper. Among these approaches, we can quote the REML (REstricted Maximum Likelihood) approach, originally proposed by Patterson and Thompson (1971) and described for instance in Searle, Casella and McCulloch (1992), which consists in estimating first $\sigma_u^\star$ and $\sigma_e^\star$ and then to estimate $\eta^\star$ as the ratio $\hat{\eta} = N\hat{\sigma_u}^2/(N\hat{\sigma_u}^2 + \hat{\sigma_e}^2)$. However, this type of approach is based on iterative procedures that require expensive matrix operations. Hence, several approximations have been proposed such as the AI algorithm (Gilmour, Thompson and Cullis (1995)) which is used for instance in the software GCTA (Genome-wide Complex Trait Analysis) described in Yang et al. (2011). Other approximations have also been proposed in the EMMA algorithm (Kang et al. (2008)). For further details on the different approximations that could be used we refer the reader to Pirinen, Donnelly and Spencer (2013). The latter paper proposes another methodology for estimating the heritability which consists in rewriting Model (1) with the parameters (6) and in using an eigenvalue decomposition of the matrix $\mathbf{R}$. Further details on their methodology are given hereafter. According to the numerical experiments conducted in Pirinen, Donnelly and Spencer (2013) their approach has the lowest computational burden among the available algorithms.

In the case of sparse high dimensional framework, most of the papers studied the case of linear models. Among them, we can quote: Meinshausen and Bühlmann (2010) and Beinrucker, Dogan and Blanchard (2014). The high dimensional linear mixed model where $\mathbf{u}$ is sparse, that is the case where $q < 1$, which is the most realistic case for the applications that we have in view, has received little attention. It has been

studied according two directions: detection and estimation. Concerning the detection field in this framework, we are only aware of the work of Arias-Castro, Candès and Plan (2011) in which a testing procedure for detecting a sparse vector in high dimensional linear sparse regression model is also proposed and compared with the one proposed by Ingster, Tsybakov and Verzelen (2010). As for the procedures dedicated to the heritability estimation, there exist, to the best of our knowledge, only three approaches: the approach of Yang et al. (2010) who propose to approximate the genetic correlation between every pairs of individuals across the set of causal SNPs by the genetic correlation across the set of all SNPs, the approach of Golan and Rosset (2011) who propose a methodology based on MCEM (Monte-Carlo expectation-maximization) developed by Wei and Tanner (1990) and the Bayesian approaches of Guan and Stephens (2011) and Zhou, Carbonetto and Stephens (2013). However, as far as the estimation issue in the high dimensional linear mixed model is concerned, the authors of these papers did not establish the theoretical properties of their estimators in the framework where both n and N tend to infinity.

4. Theoretical results

Observe that another way of writing Model (3) with the parameters defined in (6) consists in writing

$$\mathbf{Y} = \frac{1}{\sqrt{N}}\mathbf{Z}\mathbf{t} + \sigma^\star\sqrt{1-\eta^\star}\boldsymbol{\varepsilon}\ , \tag{8}$$

where $\boldsymbol{\varepsilon}$ is a $n \times 1$ Gaussian vector having a covariance matrix equal to identity and $\mathbf{t} = (t_1, \ldots, t_N)'$ is a random vector such that

$$t_i = \frac{\sigma^\star\sqrt{\eta^\star}}{\sqrt{q}} w_i \pi_i\ ,$$

where the w_i's and the π_i's are independent, $\mathbf{w} = (w_1, \ldots, w_N)'$ is a Gaussian vector with a covariance matrix equal to identity and the π_i's are i.i.d Bernoulli random variables such that $\mathbb{P}(\pi_1 = 1) = q$.

The estimator $\hat{\eta}$ is defined as a maximizer of $L_n(\eta)$ for $\eta \in [0, 1-\delta]$ for some small $\delta > 0$, L_n being given in (7). We shall study the asymptotic properties of $\hat{\eta}$ as n and N tend to infinity in a comparable way, that is when $n/N \to a > 0$. To understand the asymptotic behavior of $\hat{\eta}$, we shall first prove its consistency, then use a Taylor expansion of the derivative of L_n around $\hat{\eta}$ in the usual way. The computations as can be seen in (7) involve empirical means of functions of the eigenvalues λ_i of $\mathbf{R} = \frac{\mathbf{Z}\mathbf{Z}'}{N}$. Using Theorem 1.1 of Bai and Zhou (2008), we shall prove the almost sure convergence of such empirical quantities under a weak assumption denoted by Assumption 1 as follows.

(**A1**) Let $\mathbf{Z}$ and $\mathbf{W}$ be two matrices defined by (4) and (5). Recall that for each j in $\{1, \ldots, N\}$ the $(W_{i,j})_{1 \leqslant i \leqslant n}$ are independent and identically distributed random variables and such that the columns of $\mathbf{W}$ are independent (but not necessarily identically distributed). Assume that the entries $W_{i,j}$ of $\mathbf{W}$ are uniformly bounded, and have variance uniformly lower bounded, that is: there exist

$W_M < \infty$ and $\kappa > 0$ such that $0 \leqslant W_{i,j} \leqslant W_M$ and $\sigma_j^2 = \mathrm{Var}(W_{i,j}) \geqslant \kappa$, for all j.

The following lemma ensures that the result of Marchenko and Pastur (1968) which gives the empirical spectral distribution of sample covariance matrices $\mathbf{Z}\mathbf{Z}'/N$ holds even when the entries $Z_{i,j}$ of the matrix $\mathbf{Z}$ are not i.i.d. random variables but when $\mathbf{Z}$ is obtained by empirical standardization of a matrix $\mathbf{W}$ satisfying Assumption 1.

Lemma 1. *Under Assumption 1, as $n, N \to \infty$ such that $n/N \to a > 0$, the empirical spectral distribution of $R_N = \mathbf{Z}\mathbf{Z}'/N$: $F^{R_N}(x) = n^{-1}\sum_{k=1}^n \mathbb{1}_{\{\lambda_k \leqslant x\}}$ tends almost surely to the Marchenko-Pastur distribution defined as the distribution function of μ_a where, for any measurable set A,*

$$\mu_a(A) = \begin{cases} \left(1 - \frac{1}{a}\right)\mathbb{1}_{0\in A} + \nu_a(A) & \text{if } a > 1 \\ \nu_a(A) & \text{if } a \leqslant 1 \end{cases}$$

with

$$\mathrm{d}\nu_a(\lambda) = \frac{1}{2\pi}\frac{\sqrt{(a_+ - \lambda)(\lambda - a_-)}}{a\lambda}\mathbb{1}_{[a_-,a_+]}(x)\mathrm{d}x,\ a_\pm = (1 \pm \sqrt{a})^2\,. \tag{9}$$

In $F^{R_N}(x)$, the λ_k's denote the eigenvalues of R_N.

Our first main result is the $\sqrt{n}$-consistency of the estimator $\hat{\eta}$.

Theorem 1. *Let $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ satisfy Model (8) with $\eta^\star > 0$ and the entries $W_{i,j}$ of $\mathbf{W}$ satisfy Assumption 1. Then, for all q in $(0, 1]$, as $n, N \to \infty$ such that $n/N \to a \in (0, 1]$,*

$$\sqrt{n}(\hat{\eta} - \eta^\star) = O_P(1).$$

Such a result is a theoretical cornerstone to legitimate the use of an estimator. However, statistical inference has to be based on confidence sets. The next step is thus to find the asymptotic distribution of $\sqrt{n}(\hat{\eta} - \eta^\star)$. Define for any $\eta \in [0, 1]$ and $\lambda \geqslant 0$

$$g(\eta, \lambda) = \frac{\lambda - 1}{\eta(\lambda - 1) + 1}\,.$$

Define also

$$\gamma_n^2 = \left\{\frac{1}{n}\sum_{i=1}^n g(\hat{\eta}, \lambda_i)^2 - \left(\frac{1}{n}\sum_{i=1}^n g(\hat{\eta}, \lambda_i)\right)^2\right\}$$

and

$$\gamma^2(a, \eta^\star) = \left\{\int g(\eta, \lambda)^2 \mathrm{d}\mu_a(\lambda) - \left(\int g(\eta, \lambda)\mathrm{d}\mu_a(\lambda)\right)^2\right\}\,. \tag{10}$$

We are now ready to state our second main result about the asymptotic distribution of $\sqrt{n}(\hat{\eta} - \eta^\star)$. For general q, the result only holds when the entries of $\mathbf{Z}$, that is the random variables $Z_{i,j}$ are i.i.d. standard Gaussian. Indeed, as may be seen when computing the variances, we need to be able to find the asymptotic behavior of empirical means of functions of the eigenvalues together with the eigenvectors of the matrix $\mathbf{R} = \mathbf{Z}\mathbf{Z}'/N$.

Theorem 2. *Let $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ satisfy Model (8) with $\eta^\star > 0$ and assume that the random variables $Z_{i,j}$ are i.i.d. $\mathcal{N}(0,1)$. Then for any $q \in (0,1]$, as $n, N \to \infty$ such that $n/N \to a > 0$,*

$$\sqrt{n}(\hat{\eta} - \eta^\star)$$

converges in distribution to a centered Gaussian random variable with variance

$$\tau^2(a, \eta^\star, q) = \frac{2}{\gamma^2(a, \eta^\star)} + 3\frac{a^2 \eta^{\star 2}}{\gamma^4(a, \eta^\star)} \left(\frac{1}{q} - 1\right) S(a, \eta^\star)$$

where

$$S(a, \eta^\star) = \left[\int \frac{\lambda(\lambda-1)}{(\eta^\star(\lambda-1)+1)^2} \mathrm{d}\mu_a(\lambda) - \int \frac{\lambda}{(\eta^\star(\lambda-1)+1)} \mathrm{d}\mu_a(\lambda) \int \frac{\lambda-1}{(\eta^\star(\lambda-1)+1)} \mathrm{d}\mu_a(\lambda)\right]^2.$$

In the case where $q = 1$, the result holds in the general situation described in Assumption 1, and allows us to propose confidence sets with precise asymptotic confidence level.

Theorem 3. *Let $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ satisfy Model (8) with $q = 1$ and with $\eta^\star > 0$. Assume also that the entries $W_{i,j}$ of $\mathbf{W}$ satisfy Assumption 1 then, as $n, N \to \infty$ such that $n/N \to a > 0$,*

$$\gamma_n \sqrt{\frac{n}{2}} \, (\hat{\eta} - \eta^\star)$$

converges in distribution to $\mathcal{N}(0,1)$.

Let us now give some additional comments on the previous results. Firstly, it has to be noticed that none of the limiting variance depends on $\sigma^\star$. Secondly, Theorem 2 is proved here only in the case where the $Z_{i,j}$ are i.i.d. Gaussian. This is because we used several times that the matrix of eigenvectors of $\mathbf{Z}\mathbf{Z}'/N$ is independent of the eigenvalues, and uniformly distributed on the set of orthonormal matrices. We think that the result of Theorem 2 is also valid when the $Z_{i,j}$ are defined from the $W_{i,j}$ satisfying Assumption 1, as suggested by the numerical results obtained in Section 5. To prove it requires new results in an active research topic of the random matrix theory field. We can observe in the expression of $\tau^2(a, \eta^\star)$ given in Theorem 2 that the presence of q is counterbalanced by the presence of a^2. This will be confirmed by the results obtained in the numerical results given in Section 5. Finally, we can observe that $2/(n\gamma_n^2)$ corresponds to the usual inverse of the Fisher information associated to η. This result is classical in the case where N is fixed and n tends to infinity but did not exist in the framework where both n and N tend to infinity even if it was already used in biological applied papers for deriving standard errors and confidence intervals. Theorem 3 proves that this result still holds even in the case where both n and N tend to infinity.

To the best of our knowledge, the effect of the presence of null components in the random effects has never been taken into account for computing the asymptotic variance of an estimator of the heritability. This is the contribution of Theorem 2. This theorem shows that the asymptotic variance contains an additional term which increases its value in the case $q < 1$ with respect to the case $q = 1$. It is shown in Section 3.3 how the computation of the asymptotic variance can be altered if this additional term

is neglected. In practical situations, computing the standard error given by Theorem 2 requires the knowledge of q which is in general unknown. However, if an estimation of q is available for any practical reasons, the result of Theorem 2 can be used for computing confidence intervals and standard errors, see Section 6 for further details.

5. Numerical experiments

In this section, we first explain how to implement our method and then we illustrate the theoretical results of Section 4 on finite sample size observations for both cases: $q = 1$ and $q < 1$. We also compare the results obtained with our approach to those obtained by the GCTA software described in Yang et al. (2010) and Yang et al. (2011) which is a reference in quantitative genetics.

5.1. Implementation

In order to obtain $\hat{\eta}$, we used a Newton-Raphson approach which is based on the following recursion: starting from an initial value $\eta^{(0)}$,

$$\eta^{(k+1)} = \eta^{(k)} - \frac{L_n'(\eta^{(k)})}{L_n''(\eta^{(k)})} \ , \ k \geqslant 1 \ ,$$

where L_n' and L_n'' denote the first and second derivatives of L_n defined in (7), respectively. The closed form expression of these quantities are given in (13) and (25), respectively. In practice, this approach converges after at most 20 iterations and is not very sensitive to the initialization, namely to the value of $\eta^{(0)}$. However, in particular cases, the value of the initialization can have an influence on the estimation of $\eta^\star$. This is the case, for instance, when the real value $\eta^\star$ is close to 1. In these situations, our algorithm can provide an estimation bigger than 1 and we constrained our method to return a value equal to 0.99. Figure 1 shows the estimations obtained on 100 replications when $a = 0.1$ and $\eta^\star = 0.8$. From this figure, we can see that the estimation of $\eta^\star$ does not depend in general on the initialization, except in some cases. Moreover, the best choice for $\eta^{(0)}$ is not constant from one replication to another. In order to limit the effect of the initialization, our algorithm uses several values for $\eta^{(0)}$ and whenever the estimations differ, it keeps the estimation which is the farthest away from the boundaries.

5.2. Results in Model (2) when $q = 1$

We shall first consider the performance of the estimator $\hat{\eta}$ when $q = 1$ for $\eta^\star$ in $\{0.3, 0.5, 0.7\}$, $n = 1000$, $\sigma_u^\star = 0.1$ and for a in $\{0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1\}$, where $a = n/N$. We generated 500 data sets according to Model (1) using these parameters and $\mathbf{Z}$ as defined in (4) where the $W_{i,j}$ are binomial random variables with parameters 2 and p_j. In our experiments the p_j's are uniformly drawn in $[0.1, 0.5]$. The corresponding boxplots of $\hat{\eta}$ are displayed in Figure 2. We can see from this figure that

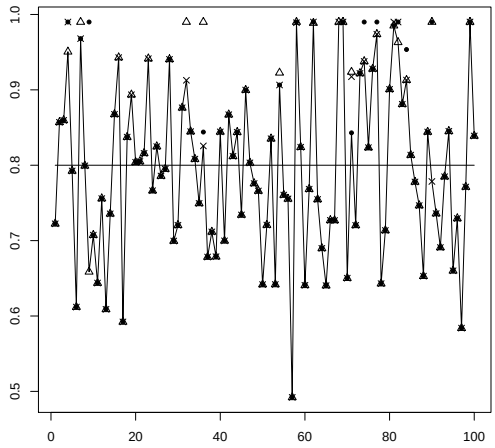

FIG 1. *Estimation of $\hat{\eta}$ obtained in the case $a = 0.1$ and $\eta^\star = 0.8$ for different values of initialization: $\eta^{(0)} = 0.1$ (dots), $\eta^{(0)} = 0.5$ (triangles) and $\eta^{(0)} = 0.9$ (crosses). The plain line displays the estimations obtained with our method to select the best initialization value and the x-axis is the replication number.*

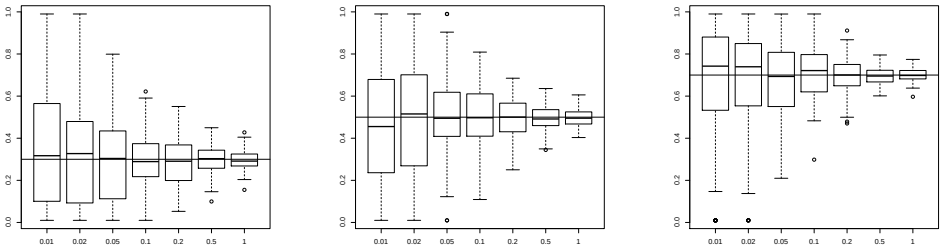

FIG 2. *Boxplots of $\hat{\eta}$ for different values of a, for $\eta^\star = 0.3$ (left), $\eta^\star = 0.5$ (middle) and $\eta^\star = 0.7$ (right). The horizontal line corresponds to the true value of $\eta^\star$. The whiskers of each boxplot correspond to the first and third quartiles.*

our approach provides unbiased estimators of $\eta^\star$ and that the smaller the a the larger the empirical variance.

In order to illustrate the central limit theorem given in Theorem 3, we first display in Figure 3 the histograms of $\gamma_n (n/2)^{1/2} (\hat{\eta} - \eta^\star)$ along with the p.d.f of a standard Gaussian random variable for $\eta^\star = 0.5$ and different values of a. We can see that the Gaussian p.d.f fits well the data in all the considered cases. We also display in Figure 4 the values of $n^{-1/2}\sqrt{2\gamma_n^{-2}}$ and the empirical standard deviation of $(\hat{\eta} - \eta^\star)$ averaged over all the experiments. As shown in Theorem 3, we also observe empirically that

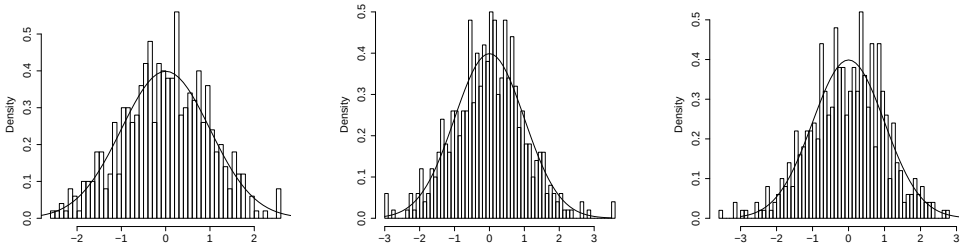

FIG 3. *Histograms of $\gamma_n (n/2)^{1/2} (\hat{\eta} - \eta^\star)$ for $\eta^\star = 0.5$ and $a = 0.05$ (left), $a = 0.1$ (middle), $a = 0.5$ (right) and the p.d.f of a standard Gaussian random variable in plain line.*

both quantities are very close.

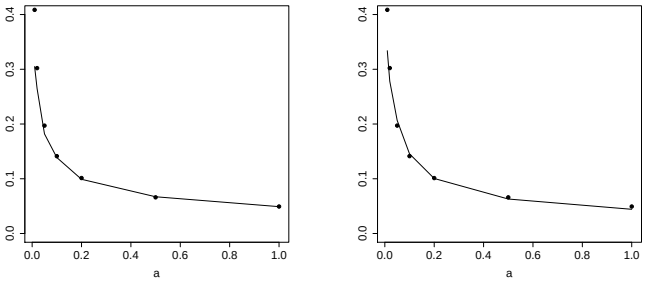

FIG 4. *Values of $n^{-1/2}\sqrt{2\gamma_n^{-2}}$ ("•") and the empirical standard deviation of $(\hat{\eta} - \eta^\star)$ (plain line) for several values of $\eta^\star$ (0.3 (left), 0.5 (right)).*

In practice, the value of $\gamma_n^{-1}(n/2)^{-1/2}$ can be used for deriving confidence intervals for $\eta^\star$. As we can see from Figure 4, our approach leads to very accurate confidence intervals for a larger than 0.1 even in finite sample size cases.

Let us now compare our results with those obtained with the software GCTA. As we can see from Figure 5 which displays the boxplots of $\hat{\eta}$ for different values of a when $\eta^\star = 0.7$, the results found by our approach and GCTA are very close. In both cases, we observe that when a is close to 1 the estimations of $\eta^\star$ are very accurate but when a is small the standard error becomes very high.

5.3. Results in model 2 when $q < 1$

This section is dedicated to the study of the performance of $\hat{\eta}$ when $q < 1$. We generated 500 data sets according to Model (1) for $\eta^\star = 0.7$, $a \in \{0.05, 0.1, 0.5, 1\}$, different

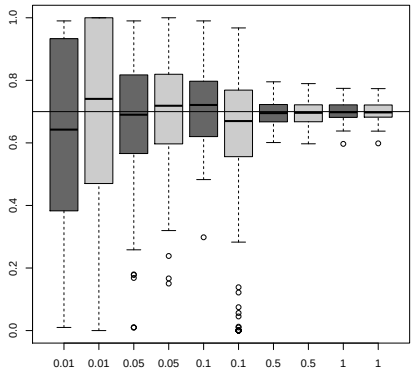

FIG 5. *Boxplots of $\hat{\eta}$ for different values of a, using our method (dark gray) and GCTA (light gray). The whiskers of each boxplot are the first and third quartiles.*

values of q and $\mathbf{Z}$ defined in (4) where the $W_{i,j}$ are binomial random variables with parameters 2 and p_j. In our experiments the p_j's are uniformly drawn in $[0.1, 0.5]$.

Figure 6 displays the boxplots of $\hat{\eta}$ for these parameters. We can see from this figure that for small values of a, the estimators of $\eta^\star$ have the same behavior for $q = 1$ and $q < 1$. However, when $a = 1$ or $a = 0.5$, we can see from this figure that the presence of null components strongly alter the performance of the estimator of $\eta^\star$. Since in typical GWAS experiments, $a = 0.01$ or even smaller, the results of Figure 6 could lead to conclude that considering the case $q < 1$ is not necessary for such values of the parameter a. However, as already noticed from Figure 2, the variance of $\hat{\eta}$ is very large for small values of a, hence considering the presence of null components and proposing a strategy for selecting only the non null components of $\mathbf{u}$ could be one way to increase a and thus to diminish the variance of $\hat{\eta}$.

In order to illustrate the central limit theorem given in Theorem 2, we first display in Figure 7 the histograms of $\tau_n^{-1} n^{1/2} (\hat{\eta} - \eta^\star)$ along with the p.d.f of a standard Gaussian random variable for $\eta^\star = 0.7$, two values of q: $q = 0.01$ and $q = 0.1$ and $a = 0.5$ (top) and two values of a: $a = 0.2$ and $a = 0.5$ with $q = 0.5$ (bottom). Here, τ_n is the empirical version of $\tau(a, \eta^\star, q)$ where γ is replaced by γ_n and $S(a, \eta^\star)$ is replaced by its empirical version with the eigenvalues of $\mathbf{R}$. When a is large ($a = 0.5$), we can see that the higher q the better the Gaussian p.d.f fits the histograms.

We also display in Figure 8 the values of $n^{-1/2}\tau_n$ and the empirical standard deviation of $(\hat{\eta} - \eta^\star)$ averaged over all the experiments for $\eta^\star = 0.7$ and $q = 0.5$. As shown in Theorem 2, we observe empirically that both quantities are very close. We also display in this figure the value of $n^{-1/2}\tau_n$ with $q = 1$ which boils down to consider the asymptotic standard deviation found in the non sparse model. We can see from this figure that neglecting the term depending on q leads to underestimate the asymptotic variance of $\hat{\eta}$ and that this difference is all the more striking that a is close to 1.

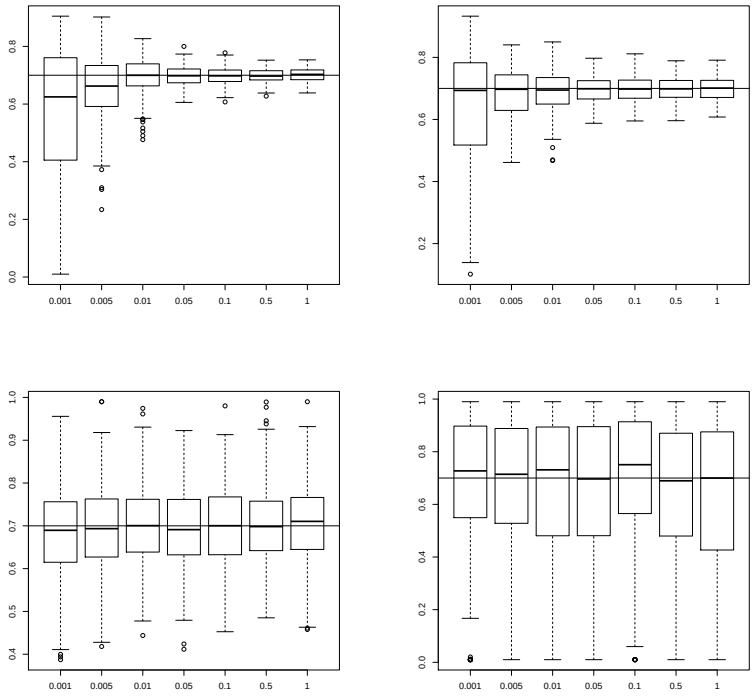

FIG 6. *Boxplots of $\hat{\eta}$ for different values of q, with $\eta^\star = 0.7$ and $a = 1$ (top left), $a = 0.5$ (top right), $a = 0.1$ (bottom left) and $a = 0.01$ (bottom right). The boxplots are based on 500 replications. The whiskers of each boxplot are the fist and third quartile.*

6. Discussion

In the course of this study, we have proposed a methodology for estimating the heritability in high dimensional linear mixed models. This methodology has two main features. Firstly, the theoretical performances of our estimator are established in a non standard theoretical framework where n and N tend to infinity and where the components of the random effect part can be equal to zero. Secondly, the computational burden of our approach is very low which makes its use possible on real data coming from GWAS experiments.

As a byproduct of the central limit theorem that we establish for $\eta^\star$ we can derive a confidence interval for the heritability. However, the confidence intervals depend on q which is the proportion of non null components in $\mathbf{u}$ and which is general unknown. For estimating q, several strategies can be considered. One could, for instance, use a GWAS approach to compute the p-values of the correlation tests of each SNP with the observations $\mathbf{Y}$ and then keep only the most significant ones. Such a practical approach can be used for providing a lower bound for q. A refinement of this approach has been

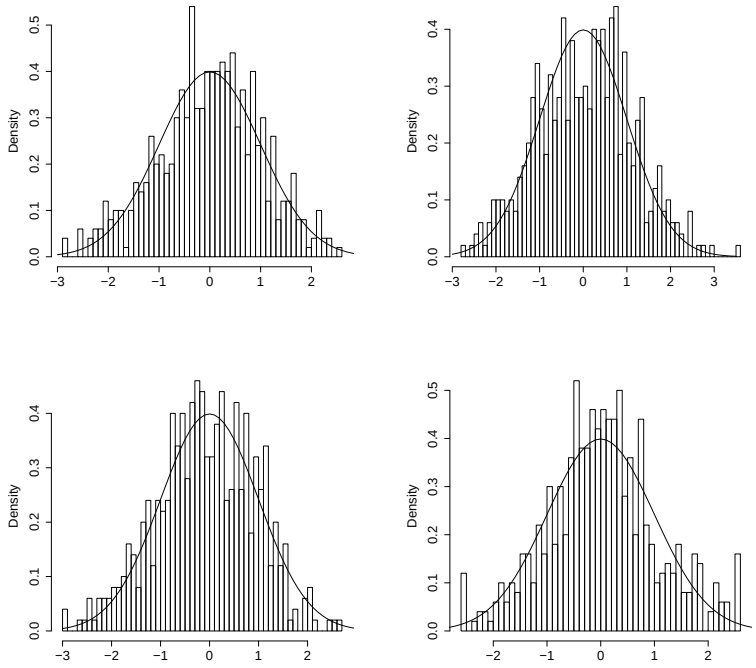

FIG 7. *Histograms of* $\tau_n^{-1} n^{1/2} (\hat{\eta} - \eta^\star)$ *for* $a = 0.5$ *and* $q = 0.5$ *(top left),* $a = 0.1$ *and* $q = 0.1$ *(top right), and for* $a = 0.1$ *and* $q = 0.01$ *(bottom left),* $a = 0.05$ *and* $q = 0.1$ *(bottom right).*

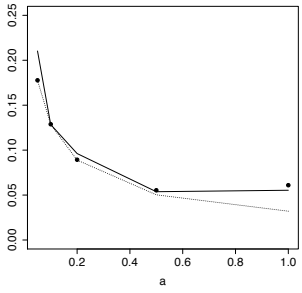

FIG 8. *Values of* $n^{-1/2}\tau_n$ *with the real value of* q *(*$q = 0.5$*) ("•"),* $q = 1$ *(dotted line) and the empirical standard deviation of* $(\hat{\eta} - \eta^\star)$ *(plain line) for* $\eta^\star = 0.7$.

proposed by Toro et al. (2014) who observed, through numerical studies, that for a fixed value of the heritability, the minimal p-value is all the more low that the number of causal SNPs is small. Hence, performing a GWAS approach on a given data set allows them to have an idea of the number of SNPs which are likely to be causal. One could also propose another practical method based on a variable selection technique. Such an approach has already been proposed by Fan and Li (2012) in the context of sparse linear mixed models. However, the framework in which their theoretical results are derived is different from the one that is considered in our paper. We are currently working on a paper Bonnet et al. (2014) which presents a variable selection method which is adapted to our framework and which could be used for estimating the proportion q of non null components in the random effects.

Moreover, we did not take into account the linkage disequilibrium issue which would require to extend our results to the case where the columns of the random matrix are correlated. This question will be the subject of a future work.

7. Proofs

Let us write the singular value decomposition (SVD) of the $n \times N$ matrix $\mathbf{Z}/\sqrt{N}$ as

$$\frac{1}{\sqrt{N}}\mathbf{Z} = \mathbf{U}\left(\sqrt{\mathbf{D}}\ \mathbf{0}\right)\mathbf{V}'$$

where $\mathbf{U}$ (already introduced in Section 1) is a $n \times n$ orthonormal matrix, $\mathbf{V}$ is a $N \times N$ orthonormal matrix and $\sqrt{\mathbf{D}}$ is a $n \times n$ diagonal matrix having its diagonal entries equal to $\sqrt{\lambda_i}$, the λ_i's being the eigenvalues of $\mathbf{R} = \mathbf{Z}\mathbf{Z}'/N$ previously defined. Thus, (8) rewrites as

$$\widetilde{\mathbf{Y}} = \mathbf{U}'\mathbf{Y} = \left(\sqrt{\mathbf{D}}\ \mathbf{0}\right)\mathbf{V}'\mathbf{t} + \sigma^\star\sqrt{1-\eta^\star}\,\widetilde{\boldsymbol{\varepsilon}}\,, \tag{11}$$

where $\widetilde{\boldsymbol{\varepsilon}} = \mathbf{U}'\boldsymbol{\varepsilon}$ is a $n \times 1$ centered Gaussian vector having a covariance matrix equal to identity.

We shall use repeatedly the following lemma which is proved in Section 7.4.

Lemma 2. *Let $\widetilde{\mathbf{Y}}$ be defined by (11) and $\mathbf{H}$ be a $n \times n$ diagonal matrix, then*

$$\mathrm{Var}\left(\widetilde{\mathbf{Y}}'\mathbf{H}\widetilde{\mathbf{Y}}|\mathbf{Z}\right) = 2\sigma^{\star 4}\,\mathrm{Tr}\left[\mathbf{H}^2\left\{(1-\eta^\star)Id_{\mathbb{R}^n} + \eta^\star\mathbf{D}\right\}^2\right] + 3\sigma^{\star 4}\eta^{\star 2}\left(\frac{1}{q}-1\right)\sum_{1\leqslant i\leqslant N} M_{ii}^2\,,$$

where

$$\mathbf{M} = \mathbf{V}\begin{pmatrix}\mathbf{D}\mathbf{H} & 0\\ 0 & 0\end{pmatrix}\mathbf{V}'\,,$$

and

$$\mathrm{Var}\left(\widetilde{\mathbf{Y}}'\mathbf{H}\widetilde{\mathbf{Y}}|\mathbf{Z}\right) \leqslant 2\sigma^{\star 4}\,\mathrm{Tr}\left[\mathbf{H}^2\left\{(1-\eta^\star)Id_{\mathbb{R}^n} + \eta^\star\mathbf{D}\right\}^2\right] + 3\sigma^{\star 4}\eta^{\star 2}\left(\frac{1}{q}-1\right)\mathrm{Tr}[\mathbf{D}^2\mathbf{H}^2].$$

Another useful lemma will be the following.

Lemma 3. *Under Assumption 1, let $h : \mathbb{R}^+ \to \mathbb{R}^+$ be such that there exist $\alpha > 0$ and C such that for all n,*

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} h(\lambda_i)^{1+\alpha}\right) \leqslant C.$$

Then

$$\frac{1}{n}\sum_{i=1}^{n} h(\lambda_i) = \int h(\lambda) d\nu_a(\lambda) + o_p(1).$$

The proof of this lemma follows from the application of Lemma 1 to the bounded function $h\mathbb{1}_{h\leqslant M}$, and the Markov inequality applied to the empirical mean of $h\mathbb{1}_{h>M}$.

Lemma 4. *Under Assumption 1 let $n, N \to \infty$ be such that $n/N \to a > 0$. Then there exists C such that for all n,*

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \lambda_i^2\right] \leqslant C.$$

To prove the lemma, notice that $\sum_{i=1}^{n} \lambda_i^2 = \mathrm{Tr}[\mathbf{Z}\mathbf{Z}'/N^2]$. But

$$\begin{aligned}\mathbb{E}\left(\mathrm{Tr}\left[(\mathbf{Z}\mathbf{Z}')^2\right]\right) &= \sum_{k\neq k'}\sum_{i,j}\mathbb{E}(Z_{i,k}Z_{j,k})\mathbb{E}(Z_{i,k'}Z_{j,k'}) + \sum_{k}\sum_{i}\mathbb{E}(Z_{i,k}^2) \\ &= nN(N-1) + N(N-1)n(n-1)\left(\frac{1}{n-1}\right)^2 + n^2N\end{aligned}$$

where the values of the involved expectations may be found in the proof of Lemma 1 in Section 7.4. We thus have

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \lambda_i^2\right] \leqslant 2 + \frac{n}{N}$$

which ends the proof.

7.1. Proof of Theorem 1

The first step is to prove the consistency of $\hat{\eta}$. We shall first prove that $L_n(\eta)$ converges uniformly for $\eta \in [0, 1-\delta]$ in probability to $L(\eta)$ given by

$$L(\eta) = -2\log\sigma^\star - \log\int\left[\frac{\eta^\star(\lambda-1)+1}{\eta(\lambda-1)+1}\right]d\mu_a(\lambda) - \int\log\left(\eta(\lambda-1)+1\right)d\mu_a(\lambda).$$

Using Lemma 2 with $\mathbf{H}$ with diagonal entries $1/(\eta(\lambda_i-1)+1)$, we get that

$$\begin{aligned}\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{Y}_i^2}{\eta(\lambda_i-1)+1}|\mathbf{Z}\right] &\leqslant \frac{\sigma^{\star 4}}{n^2}\sum_{i=1}^{n}\left[2\left(\frac{\eta^\star(\lambda_i-1)+1}{\eta(\lambda_i-1)+1}\right)^2\right. \\ &\qquad \left. + 3\left(\frac{1}{q}-1\right)\left(\frac{\eta^\star\lambda_i}{\eta(\lambda_i-1)+1}\right)^2\right] \\ &\leqslant \sigma^{\star 4}\left(2+3\left(\frac{1}{q}-1\right)\right)\frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{\lambda_i+1}{\delta}\right)^2\end{aligned}$$

since $\eta \in [0, 1-\delta]$. Now, using Lemma 4 we get that

$$\frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{\lambda_i+1}{\delta}\right)^2 = o_P(1)$$

which leads to

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{Y}_i^2}{\eta(\lambda_i-1)+1} &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{Y}_i^2}{\eta(\lambda_i-1)+1}|\mathbf{Z}\right]+o_p(1)\\ &= \sigma^{\star 2}\frac{1}{n}\sum_{i=1}^{n}\frac{\eta^\star(\lambda_i-1)+1}{\eta(\lambda_i-1)+1}+o_P(1).\end{aligned}$$

Now, using Lemma 3 we easily get that $\frac{1}{n}\sum_{i=1}^{n}\frac{\eta^\star(\lambda_i-1)+1}{\eta(\lambda_i-1)+1}$ converges in probability to $\int[\frac{\eta^\star(\lambda-1)+1}{\eta(\lambda-1)+1}]d\mu_a(\lambda)$ and $\frac{1}{n}\sum_{i=1}^{n}\log[(\eta(\lambda_i-1)+1)]$ converges in probability to $\int\log(\eta(\lambda-1)+1)d\mu_a(\lambda)$ so that $L_n(\eta) = L(\eta)+o_P(1)$.

In order to prove the uniform convergence of L_n to L in probability on $[0, 1-\delta]$, we shall use the following lemma which is proved in section 7.4.

Lemma 5. *Assume that for any $\eta \in [0, 1-\delta]$, $L_n(\eta)$ converges in probability to $L(\eta)$ and that*

$$\sup_{\eta\in[0,1-\delta]}\left|L_n'(\eta)\right| = O_P(1),\ \textit{as } n \textit{ tends to infinity}, \tag{12}$$

then

$$\sup_{\eta\in[0,1-\delta]}|L_n(\eta)-L(\eta)| = o_P(1),\ \textit{as } n \textit{ tends to infinity}.$$

Let us now prove that $\sup_{\eta\in[0,1-\delta]}|L_n'(\eta)| = O_P(1)$. Note that

$$L_n'(\eta) = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{Y}_i^2(\lambda_i-1)}{\{\eta(\lambda_i-1)+1\}^2}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{Y}_i^2}{\eta(\lambda_i-1)+1}\right)^{-1}-\frac{1}{n}\sum_{i=1}^{n}\frac{\lambda_i-1}{\eta(\lambda_i-1)+1}. \tag{13}$$

A study of $\eta \mapsto \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{Y}_i^2(\lambda_i-1)}{\{\eta(\lambda_i-1)+1\}^2}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{Y}_i^2}{\eta(\lambda_i-1)+1}\right)^{-1}$ shows that it is decreasing and that it takes negative values for $\eta \in [0, 1-\delta]$, so that its absolute value is maximum for $\eta = 1-\delta$. Thus

$$\begin{aligned}\sup_{\eta\in[0,1-\delta]}\left|L_n'(\eta)\right| &\leqslant \frac{1}{\delta}\left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^2|\lambda_i-1|\right)\left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^2\right)^{-1}+\frac{1}{n\delta}\sum_{i=1}^{n}|\lambda_i-1|\\ &\leqslant \frac{2}{\delta}+\frac{1}{\delta}\left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^2\lambda_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^2\right)^{-1}+\frac{1}{n\delta}\sum_{i=1}^{n}\lambda_i.\end{aligned}$$

By Lemma 2 with $\mathbf{H} = \mathrm{Id}$, we get

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^2 = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^2|\mathbf{Z}\right] + o_p(1) = \frac{\sigma^{2\star}}{n}\sum_{i=1}^{n}[\eta^\star(\lambda_i - 1) + 1)] + o_p(1)$$
$$= \sigma^{2\star}\int(\eta(\lambda-1)+1)d\mu_a(\lambda) + o_p(1),$$

where the last equality comes from Lemma 3. In the same way, we get by using Lemma 2 with $\mathbf{H}$ having its diagonal entries equal to λ_i and Lemma 3 that

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^2\lambda_i = \sigma^{2\star}\int\lambda(\eta(\lambda-1)+1)d\mu_a(\lambda) + o_p(1) = O_P(1).$$

Finally, we get from Lemma 3 that

$$\frac{1}{n}\sum_{i=1}^{n}\lambda_i = \int\lambda d\mu_a(\lambda) + o_p(1) = O_P(1)$$

which ends the proof of (12). By Lemma 5, we thus have proved that

$$\sup_{\eta\in[0,1-\delta]}|L_n(\eta) - L(\eta)| = o_P(1). \tag{14}$$

Now, using Jensen's inequality, we easily get that for all $\eta \in [0,1]$, $L(\eta) \leqslant L(\eta^\star)$, with equality if and only if $\eta = \eta^\star$. This together with (14) gives

$$\hat{\eta} = \eta^\star + o_P(1). \tag{15}$$

The next step is to prove that $\sqrt{n}(\hat{\eta} - \eta^\star) = O_P(1)$. Let us first note that $\hat{\eta}$ satisfies the following equation:

$$\sqrt{n}(\hat{\eta} - \eta^\star) = -\frac{\sqrt{n}L_n'(\eta^\star)}{L_n''(\widetilde{\eta})}, \quad \widetilde{\eta} \in (\hat{\eta}, \eta^\star)\ . \tag{16}$$

We first prove the asymptotic convergence of $L_n''(\widetilde{\eta})$.

Lemma 6. *Let $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ satisfy Model (8) with $\eta^\star > 0$ and the entries $W_{i,j}$ of $\mathbf{W}$ satisfy Assumption 1. Then, for all q in $(0,1]$, as $n, N \to \infty$ such that $n/N \to a \in (0,1]$, for any random variable $\widetilde{\eta}$ such that $\widetilde{\eta} \in (\hat{\eta}, \eta^\star)$,*

$$L_n''(\widetilde{\eta}) = -\sigma^{\star 2}\gamma^2(a, \eta^\star) + o_P(1).$$

Lemma 6 is proved in Section 7.4.
Let us now focus on the properties of $L_n'(\eta^\star)$. Using the following notation

$$U_i = \frac{\widetilde{Y}_i}{\sqrt{\eta^\star(\lambda_i - 1) + 1}}\ , \tag{17}$$

we see that $\sqrt{n}L_n'(\eta^\star)$ can be rewritten as follows:

$$\begin{aligned}
&\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\left({U_i}^2-\frac{1}{n}\sum_{j=1}^n U_j^2\right)g(\eta^\star,\lambda_i)\right\}\left(\frac{1}{n}\sum_{i=1}^n U_i^2\right)^{-1}\\
&=\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\left({U_i}^2-1\right)+\left(1-\frac{1}{n}\sum_{j=1}^n U_j^2\right)\right]g(\eta^\star,\lambda_i)\right\}\left(\frac{1}{n}\sum_{i=1}^n U_i^2\right)^{-1}\\
&=\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\left({U_i}^2-1\right)g(\eta^\star,\lambda_i)\right\}\left(\frac{1}{n}\sum_{i=1}^n U_i^2\right)^{-1}\\
&\qquad\qquad-\left\{\frac{1}{\sqrt{n}}\sum_{j=1}^n\left({U_j}^2-1\right)\right\}\left\{\frac{1}{n}\sum_{i=1}^n g(\eta^\star,\lambda_i)\right\}\left(\frac{1}{n}\sum_{i=1}^n U_i^2\right)^{-1},
\end{aligned}$$

where

$$g(\eta,\lambda)=\frac{\lambda-1}{\eta(\lambda-1)+1}\,.$$

But using Lemma 2 and Lemma 3 we get

$$\mathrm{Var}\left[n^{-1/2}\sum_{j=1}^n({U_j}^2-1)|\mathbf{Z}\right]=O_P(1)$$

Moreover, by Lemma 3, $n^{-1}\sum_{i=1}^n g(\eta^\star,\lambda_i)$ converges in probability to $\int g(\eta^\star,\lambda)\mathrm{d}\mu_a(\lambda)$. Thus,

$$\sqrt{n}L_n'(\eta^\star)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\left({U_i}^2-1\right)\left(g(\eta^\star,\lambda_i)-\int g(\eta^\star,\lambda)\mathrm{d}\mu_a(\lambda)\right)+o_P(1),\ \text{as } n\to\infty\,. \tag{18}$$

Using again Lemma 2 and Lemma 3 we obtain

$$\sqrt{n}L_n'(\eta^\star)=O_P(1).$$

This, together with Lemma 6 and (16) ends the proof of Theorem 1.

7.2. Proof of Theorem 2

Notice first that all previous results may be used, replacing Assumption 1 by the assumption that the $Z_{i,j}$ are i.i.d. standard Gaussian. Indeed, in this case, Lemma 1 reduces to the original result of Marchenko and Pastur (1968), Lemma 3 only involves Lemma 1 and truncation arguments, and the computations leading to Lemma 4 still hold. Thus, Theorem 1 and Lemma 6 also still hold.

Let us now prove that $\sqrt{n}L_n'(\eta^\star)$ converges in distribution to a centered Gaussian. Define $\mathbf{H}$ the diagonal $n\times n$ matrix with diagonal entries

$$H_i=\frac{1}{\eta^\star(\lambda_i-1)+1}\left[g(\eta^\star,\lambda_i)-\int g(\eta^\star,\lambda)\mathrm{d}\mu_a(\lambda)\right].$$

Define

$$\mathbf{L}_n = \frac{1}{\sqrt{n}}\widetilde{\mathbf{Y}}'\mathbf{H}\widetilde{\mathbf{Y}}.$$

Then using (18) and Lemma 3 we have

$$\sqrt{n}L_n'(\eta^\star) = \mathbf{L}_n - \mathbb{E}[\mathbf{L}_n|\mathbf{Z}] + o_P(1).$$

Now using Lemma 2 we get that setting $\gamma_n^2 = \mathrm{Var}\,[\mathbf{L}_n|\mathbf{Z}]$,

$$\begin{aligned}\gamma_n^2 &= 2\sigma^{\star 4}\frac{1}{n}\,\mathrm{Tr}\left[\mathbf{H}^2\left((1-\eta)^\star Id_{\mathbb{R}^n} + \eta^\star\mathbf{D}\right)^2\right] + 3\sigma^{\star 4}\eta^{\star 2}\left(\frac{1}{q}-1\right)\frac{1}{n}\sum_{i=1}^N M_{i,i}^2 \\ &= 2\sigma^{\star 4}\frac{1}{n}\sum_{i=1}^n\left(g(\eta^\star,\lambda_i) - \int g(\eta^\star,\lambda)\mathrm{d}\mu_a(\lambda)\right)^2 \\ &\qquad + 3\sigma^{\star 4}\eta^{\star 2}\left(\frac{1}{q}-1\right)\frac{1}{n}\sum_{i=1}^n\sum_{k,l=1}^n \lambda_k\lambda_l H_k H_l V_{i,k}^2 V_{i,l}^2.\end{aligned}$$

The first term in this sum converges as $n, N \to \infty$ to $2\sigma^{\star 4}\gamma^2(a,\eta^\star)$.
Under the assumption that the $Z_{i,j}$ are i.i.d. standard Gaussian, the matrix of eigenvectors $\mathbf{V}$ is Haar distributed on the orthonormal matrices, and is independent of $(\lambda_i)_{1\leqslant i\leqslant n}$, see Bai and Silverstein (2010) chapter 10. Conditionally to the eigenvalues $(\lambda_i)_{1\leqslant i\leqslant n}$, we thus get that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n\sum_{k,l=1}^n \lambda_k\lambda_l H_k H_l V_{i,k}^2 V_{i,l}^2|\mathbf{D}\right] = \left(\frac{1}{N}\sum_{k=1}^n \lambda_k H_k\right)^2(1+o(1))$$

converges to

$$a^2\left[\int\frac{\lambda(\lambda-1)}{(\eta^\star(\lambda-1)+1)^2}\mathrm{d}\mu_a(\lambda) - \int\frac{\lambda}{(\eta^\star(\lambda-1)+1)}\mathrm{d}\mu_a(\lambda)\int\frac{\lambda-1}{(\eta^\star(\lambda-1)+1)}\mathrm{d}\mu_a(\lambda)\right]^2$$

and

$$\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n\sum_{k,l=1}^n \lambda_k\lambda_l H_k H_l V_{i,k}^2 V_{i,l}^2|\mathbf{D}\right] = o_P(1)$$

so that

$$\gamma_n^2 = 2\sigma^{\star 4}\gamma^2(a,\eta^\star) + 3\sigma^{\star 4}\eta^{\star 2}\left(\frac{1}{q}-1\right)S(a,\eta^\star) + o_P(1).$$

Denote Δ the diagonal $N\times N$-matrix with diagonal entries $\Delta_i = \frac{\sigma^\star\sqrt{\eta^\star}}{\sqrt{q}}\pi_i$. Let us now write

$$\mathbf{L}_n - \mathbb{E}(\mathbf{L}_n|\mathbf{Z}) = \mathbf{L}_n - \mathbb{E}\left[\mathbf{L}_n|\Delta,\mathbf{Z}\right] + \mathbb{E}\left[\mathbf{L}_n|\Delta,\mathbf{Z}\right] - \mathbb{E}\left[\mathbf{L}_n|\mathbf{Z}\right].$$

We first have

$$\mathbb{E}\left[\mathbf{L}_n|\Delta,\mathbf{Z}\right] - E\left[\mathbf{L}_n|\mathbf{Z}\right] = \sigma^{\star 2}\eta^\star\frac{1}{\sqrt{n}}\sum_{i=1}^N\left(\frac{\pi_i^2}{q}-1\right)M_{i,i}$$

whose variance, conditionally to $\mathbf{Z}$ is

$$s_{n,1}^2 = \sigma^{\star 4}\eta^{\star 2}\left(\frac{1}{q}-1\right)\frac{1}{n}\sum_{i=1}^{N} M_{i,i}^2.$$

In the same way as for γ_n^2 we get that

$$s_{n,1}^2 = \sigma^{\star 4}\eta^{\star 2}\left(\frac{1}{q}-1\right)S(a,\eta^\star) + o_P(1).$$

Let

$$\xi_i = \left(\frac{\pi_i^2}{q}-1\right)M_{i,i} = \left(\frac{\pi_i^2}{q}-1\right)\sum_{k=1}^{n}\frac{\lambda_k(\lambda_k-1)}{(\eta^\star(\lambda_k-1)+1)^2}V_{i,k}^2.$$

Since $\eta^\star > 0$, the function $\lambda \mapsto \frac{\lambda(\lambda-1)}{(\eta^\star(\lambda-1)+1)^2}$ is bounded, and $\sum_{k=1}^n V_{i,k}^2 \leqslant \sum_{k=1}^N V_{i,k}^2 = 1$. Also, the variables $\left(\frac{\pi_i^2}{q}-1\right)$ are uniformly bounded by $1/q$. Thus

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\xi_i^2 \mathbb{1}_{|\xi_i|\geqslant cn}|\mathbf{Z}\right] = 0$$

for large enough n. Then, by Lindeberg's Theorem, conditionally to $\mathbf{Z}$,

$$\frac{1}{s_{n,1}}\left(\mathbb{E}\left[\mathbf{L}_n|\Delta,\mathbf{Z}\right] - \mathbb{E}\left[\mathbf{L}_n|\mathbf{Z}\right]\right)$$

converges in distribution to $\mathcal{N}(0,1)$.
Let us now set

$$s_{n,2}^2 = \gamma_n^2 - s_{n,1}^2$$

and notice that $s_{n,2}^2$ converges to

$$2\sigma^{\star 4}\gamma^2(a,\eta^\star) + 2\sigma^{\star 4}\eta^{\star 2}\left(\frac{1}{q}-1\right)S(a,\eta^\star).$$

We shall prove that, conditionally to $\mathbf{Z}$ and Δ, $(\mathbf{L}_n - \mathbb{E}(\mathbf{L}_n|\Delta,\mathbf{Z}))/s_{n,2}$ converges in distribution to $\mathcal{N}(0,1)$, and thus also unconditionally. Write

$$\mathbf{L}_n = \frac{1}{\sqrt{n}}\begin{pmatrix}\mathbf{w}' & \boldsymbol{\varepsilon}'\end{pmatrix} B \begin{pmatrix}\mathbf{w}\\ \boldsymbol{\varepsilon}\end{pmatrix}$$

where B is the $(N+n)\times(N+n)$-matrix

$$B = \begin{pmatrix}\Delta & 0\\ 0 & \sigma^\star(1-\eta^\star)^{\frac{1}{2}}Id_{\mathbb{R}^n}\end{pmatrix}\begin{pmatrix}\mathbf{V}\begin{pmatrix}\mathbf{DH} & 0\\ 0 & 0\end{pmatrix}\mathbf{V}' & \tilde{\mathbf{V}}\sqrt{\mathbf{D}}\mathbf{H}\\ \mathbf{H}\sqrt{\mathbf{D}}\tilde{\mathbf{V}}' & \mathbf{H}\end{pmatrix}\begin{pmatrix}\Delta & 0\\ 0 & \sigma^\star(1-\eta^\star)^{\frac{1}{2}}Id_{\mathbb{R}^n}\end{pmatrix}.$$

Here, $\tilde{\mathbf{V}}$ is the $N \times n$-matrix which consists of the first n columns of $\mathbf{V}$. Let ϕ be the characteristic function of $(\mathbf{L}_n - \mathbb{E}(\mathbf{L}_n|\Delta, \mathbf{Z}))/s_{n,2}$ conditionally to $\mathbf{Z}$ and Δ. Notice first that if b_j, $j = 1, \ldots, n+N$ are the eigenvalues of B, we may write

$$\mathbf{L}_n - \mathbb{E}\left[\mathbf{L}_n|\Delta, \mathbf{Z}\right] = \frac{1}{\sqrt{n}} \sum_{j=1}^{N+n} b_j(e_j^2 - 1).$$

for random variables e_j i.i.d. standard Gaussian. Thus

$$\phi\left(t\right) = \prod_{j=1}^{N+n} \left[\left(1 - 2i\frac{tb_j}{s_{n,2}\sqrt{n}}\right)^{-1/2} \exp\left(-i\frac{tb_j}{s_{n,2}\sqrt{n}}\right) \right]$$

and Taylor expansion leads to

$$\begin{aligned} \log\phi\left(t\right) &= \sum_{j=1}^{N+n} \left[-\frac{1}{2}\log\left(1 - 2i\frac{tb_j}{s_{n,2}\sqrt{n}}\right) - i\frac{tb_j}{s_{n,2}\sqrt{n}} \right] \\ &= -t^2 \frac{1}{n s_{n,2}^2} \sum_{j=1}^{N+n} b_j^2 + O\left[\frac{1}{n\sqrt{n} s_{n,2}^3} \sum_{j=1}^{N+n} b_j^3 \right]. \end{aligned}$$

We shall now prove that $\frac{1}{n s_{n,2}^2} \sum_{j=1}^{N+n} b_j^2$ converges to $1/2$. Tedious computations give

$$\begin{aligned} \sum_{j=1}^{N+n} b_j^2 &= \mathrm{Tr}(B^2) \\ &= \mathrm{Tr}(\Delta M \Delta^2 M \Delta) + \sigma^{\star 4}(1-\eta^{\star 2}\,\mathrm{Tr}(\mathbf{H}^2) + 2\sigma^{\star 2}(1-\eta^\star)\,\mathrm{Tr}[\Delta^2 \tilde{\mathbf{V}}\mathbf{D}\mathbf{H}^2\tilde{\mathbf{V}}']. \end{aligned}$$

Using the distribution of $\mathbf{V}$ and its independence on $\mathbf{D}$ we get

$$\begin{aligned} \mathbb{E}\left[\sum_{j=1}^{N+n} b_j^2 | \mathbf{D} \right] &= 2\sigma^{\star 4}\,\mathrm{Tr}\left[\mathbf{H}^2 \left((1-\eta)^\star Id_{\mathbb{R}^n} + \eta^\star \mathbf{D} \right)^2 \right] \\ &\quad + 2\sigma^{\star 4}\eta^{\star 2}\left(\frac{1}{q} - 1\right)\left(\frac{1}{N}\sum_{k=1}^{n} \lambda_k H_k\right)^2 (1 + o(1)) \end{aligned}$$

so that

$$\mathbb{E}\left[\frac{1}{n} \sum_{j=1}^{N+n} b_j^2 | \mathbf{D} \right] = 2\sigma^{\star 4}\gamma^2(a, \eta^\star) + 2\sigma^{\star 4}\eta^{\star 2}\left(\frac{1}{q} - 1\right) S(a, \eta^\star) + o_P(1).$$

Moreover, tedious computations again give

$$\mathrm{Var}\left[\frac{1}{n} \sum_{j=1}^{N+n} b_j^2 | \mathbf{D} \right] = o_P(1),$$

and we obtain that

$$\frac{1}{ns_{n,2}^2}\sum_{j=1}^{N+n} b_j^2 = \frac{1}{2} + o_P(1).$$

We shall now prove that $\frac{1}{n\sqrt{n}s_{n,2}^3}\sum_{j=1}^{N+n} b_j^3 = o_P(1)$. To do so, it is enough to prove that $\max_j |b_j| = o_P(\sqrt{n})$. Notice that for any normed vector $A = (A_1, A_2)$ in $\mathbb{R}^{N+n}$ where $A_1 \in \mathbb{R}^N$ and $A_2 \in \mathbb{R}^n$,

$$\max_j |b_j| \leqslant A'BA.$$

Now,

$$A'BA = A_1'(\Delta M \Delta)A_1 + 2\sigma^\star\sqrt{1-\eta^\star}A_1'(\Delta\tilde{\mathbf{V}}\sqrt{\mathbf{D}}\mathbf{H})A_2 + \sigma^{\star 2}(1-\eta^\star)A_2'\mathbf{H}A_2.$$

First, since $\eta^\star > 0$, all entries of $\mathbf{H}$ and $\mathbf{D}$ and $\mathbf{HD}$ are uniformly bounded and so are all entries of Δ. We thus get $A_2'\mathbf{H}A_2 = O(1)$ and $A_1'(\Delta\tilde{\mathbf{V}}\sqrt{\mathbf{D}}\mathbf{H})A_2 = O(1)$. Then, using the distribution of $\mathbf{V}$ and its independence on $\mathbf{D}$ we get

$$\mathbb{E}\left[A_1'(\Delta M\Delta)A_1|\mathbf{D}\right] = O\left(\frac{1}{N}\sum_{i=1}^n \lambda_i H_i\right)$$

and

$$\mathrm{Var}\left[A_1'(\Delta M\Delta)A_1|\mathbf{D}\right] = o_P(1),$$

so that $A'BA = O_P(1)$. We have thus proved that $\max_j |b_j| = O_P(1) = o_P(\sqrt{n})$.

Thus $\phi(t)$ converges in probability for all t to $\exp -\frac{t^2}{2}$ and the convergence may be strengthened by contradiction to an a.s. convergence, so that conditionally to $\mathbf{Z}$ and Δ, $(\mathbf{L}_n - \mathbb{E}(\mathbf{L}_n|\Delta,\mathbf{Z}))/s_{n,2}$ converges in distribution to $\mathcal{N}(0,1)$.

Now, conditionally to $\mathbf{Z}$ and Δ, $(\mathbf{L}_n - E(\mathbf{L}_n|\Delta,Z))/s_{n,2}$ converges in distribution to a Gaussian random variable independent of Δ. Thus conditionally to $\mathbf{Z}$, $\mathbf{L}_n - E\left[\mathbf{L}_n|\Delta,Z\right]$ and $E\left[\mathbf{L}_n|\Delta,Z\right] - E\left[\mathbf{L}_n|Z\right]$ converge in distribution to independent Gaussian variables, so that their sum converges in distribution to a centered Gaussian with variance the sum of the variances, namely the limit of γ_n^2, and Theorem 2 is proved.

7.3. Proof of Theorem 3

Using Lemma 6 and (16) , there remains to prove that $\sqrt{n}L_n'(\eta^\star)$ converges in distribution to $\mathcal{N}(0, 2\sigma^{\star 4}\gamma^2(a,\eta^\star))$ and that γ_n^2 converges in probability to $\gamma^2(a,\eta^\star)$.
Notice first that when $q = 1$, $(U_1,\ldots,U_n)|\mathbf{Z}$ is a centered Gaussian vector with a covariance matrix equal to $\sigma^{\star 2}$ times the identity matrix. We shall prove that conditionally to $\mathbf{Z}$, $\sqrt{n}L_n'(\eta^\star)$ converges in distribution to $\mathcal{N}(0, 2\sigma^{\star 4}\gamma^2(a,\eta^\star))$ so that the result still holds unconditionally. Using (18), it is only needed to prove it for

$\frac{1}{\sqrt{n}}\sum_{i=1}^n \left({U_i}^2-1\right)\left(g(\eta^\star,\lambda_i)-\int g(\eta^\star,\lambda)\mathrm{d}\mu_a(\lambda)\right)$. Now, conditionally to $\mathbf{Z}$, the variance of

$$\sum_{i=1}^n \left({U_i}^2-1\right)\left(g(\eta^\star,\lambda_i)-\int g(\eta^\star,\lambda)\mathrm{d}\mu_a(\lambda)\right)$$

is

$$\gamma_n^2=\frac{2{\sigma^\star}^4}{n}\sum_{i=1}^n\left(g(\eta^\star,\lambda_i)-\int g(\eta^\star,\lambda)\mathrm{d}\mu_a(\lambda)\right)^2.$$

Since $\eta^\star>0$, $g(\eta^\star,\lambda)$ is a bounded function of λ, and using Lemma 3,

$$\gamma_n^2=2{\sigma^\star}^4\gamma^2(a,\eta^\star))+o_P(1).$$

Also, setting $\xi_i=\left({U_i}^2-1\right)\left(g(\eta^\star,\lambda_i)-\int g(\eta^\star,\lambda)\mathrm{d}\mu_a(\lambda)\right)$ and C an upper bound of $|g(\eta^\star,\lambda)|$, we get that for any $c>0$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\mathbb{E}\left[\xi_i^2\mathbb{1}_{|\xi_i|\geqslant cn}|\mathbf{Z}\right] &\leqslant 4C^2{\sigma^\star}^4\mathbb{E}\left[\left({U_1}^2-1\right)^2\mathbb{1}_{2C|{U_1}^2-1|\geqslant cn}|Z\right]\\ &= 4C^2{\sigma^\star}^4\mathbb{E}\left[\left({U_1}^2-1\right)^2\mathbb{1}_{2C|{U_1}^2-1|\geqslant cn}\right]=o(1),\end{aligned}$$

where the first equality comes from the fact that the distribution of $(U_1,\dots,U_n)|\mathbf{Z}$ does not depend on $\mathbf{Z}$ and is thus also the distribution of $(U_1,\dots,U_n)$. Then, using Lindeberg's Theorem, conditionally to $\mathbf{Z}$, $\sqrt{n}L_n'(\eta^\star)$ converges in distribution to $\mathcal{N}(0,2{\sigma^\star}^4\gamma^2(a,\eta^\star))$ and thus also unconditionally.
The fact that γ_n^2 converges in probability to $\gamma^2(a,\eta^\star)$ is a straightforward consequence of Taylor expansion, the fact that $g(\eta^\star,\lambda)$ and its derivative with respect to η in the neighborhood of $\eta^\star$ are bounded functions of λ, and Slutzky's Lemma.

7.4. Proofs of technical lemmas

7.4.1. Proof of Lemma 1

As a byproduct of Theorem 1.1, Corollary 1.1 and Remark 1.1 of Bai and Zhou (2008), we use the following result to prove Lemma 1.

Theorem (Bai and Zhou (2008))**.** *Let $\mathbf{Z}$ be a matrix of size $n\times N$ which columns, denoted by $Z_1,\dots,Z_N$, are independent and let us denote $\bar{\mathbf{Z}}=\frac{1}{N}\sum_{k=1}^N Z_k$. Let us also recall that $\mathbf{R}=\mathbf{Z}\mathbf{Z}'/N$ and $F^{\mathbf{R}}$ is its empirical spectral distribution defined by $F^{\mathbf{R}}(x)=\frac{1}{n}\sum_{k=1}^n 1_{\{\lambda_k\geqslant x\}}$, where $\lambda_1,\dots,\lambda_n$ are the eigenvalues of $\mathbf{R}$. As $N\to\infty$, assume the following:*

1. *$T=(t_{i,j})$ is a matrix such that $\mathbb{E}(\bar{Z}_{i,j}Z_{m,j})=t_{m,i}$ for all j .*
2. *$\frac{1}{N}\max_{i\neq m}\mathbb{E}(\bar{Z}_{i,j}Z_{m,j})^2\to 0$ uniformly in $j\leqslant N$.*
3. *$\frac{1}{N^2}\sum_\Lambda\left(\mathbb{E}(\bar{Z}_{i,j}Z_{m,j}-t_{m,i})(Z_{i',j}\bar{Z}_{m',j}-t_{i',m'})\right)^2\to 0$ uniformly in $j\leqslant N$, with $\Lambda=\{(i,m,i',m'):1\leqslant i,m,i',m'\leqslant n\}\backslash\{(i,m,i',m'):i=i'\neq m=m'$ or $i=m'\neq i'=m\}$.*

4. $\frac{n}{N} \to a \in (0, +\infty)$.
5. *The norm of* T *is uniformly bounded and* F^T *tends to a degenerate distribution with mass at* $1/a$.

Then, with probability 1, F^R *converges to the Marchenko-Pastur distribution defined in (9).*

Observe that for all $j = 1, \ldots, N$,

$$\sum_{i=1}^{n} Z_{i,j} = 0 \tag{19}$$

and

$$\sum_{i=1}^{n} Z_{i,j}^2 = n. \tag{20}$$

Moreover, for each j, the random variables $(Z_{i,j})_{1 \leqslant i \leqslant n}$ are exchangeable. Thus, we deduce from (20) that for all $i = 1, \ldots, n$ and $j = 1, \ldots, N$, $\mathbb{E}(Z_{i,j}^2) = 1$. Hence, by (19), we get that

$$0 = \left(\sum_{i=1}^{n} Z_{i,j} \right)^2 = \sum_{i=1}^{n} Z_{i,j}^2 + \sum_{1 \leqslant i \neq m \leqslant n} Z_{i,j} Z_{m,j} \ ,$$

which, by (20), implies that for all $j = 1, \ldots, N$ and $i \neq m = 1, \ldots, n$,

$$\mathbb{E}(Z_{i,j} Z_{m,j}) = -\frac{n}{n(n-1)} = -\frac{1}{n-1} \ . \tag{21}$$

Thus, the matrix $\mathbf{T} = \mathbf{T}_n$ defined in Theorem (Bai and Zhou (2008)) is equal to $\mathbf{T} = n/(n-1)\mathrm{Id}_{\mathbb{R}^n} - \mathbf{J}_n/(n-1)$, where $\mathbf{J}_n$ is a $n \times n$ matrix having all its entries equal to 1. Hence the eigenvalues of $\mathbf{T}$ are 0 with multiplicity 1 and $n/(n-1)$ with multiplicity $(n-1)$, which gives Condition 5. of Theorem (Bai and Zhou (2008)).

Let us then check Condition 2. of Theorem (Bai and Zhou (2008)). Observe that, for $i \neq m$, $\mathbb{E}[(Z_{i,j} Z_{m,j} - t_{m,i})^2] = \mathbb{E}(Z_{i,j}^2 Z_{m,j}^2) - t_{m,i}^2$. By (20), for all $j = 1, \ldots, N$,

$$n^2 = \left(\sum_{i=1}^{n} Z_{i,j}^2 \right)^2 = \sum_{i=1}^{n} Z_{i,j}^4 + \sum_{1 \leqslant i \neq m \leqslant n} Z_{i,j}^2 Z_{m,j}^2 \ .$$

Since the $(Z_{i,j})_{1 \leqslant i \leqslant n}$ are exchangeable for each $j = 1, \ldots, N$, we get that for all $j = 1, \ldots, N$,

$$n = \mathbb{E}[Z_{1,j}^4] + (n-1)\mathbb{E}[Z_{1,j}^2 Z_{2,j}^2] \ .$$

Thus, for all $j = 1, \ldots, N$, $\mathbb{E}[Z_{1,j}^2 Z_{2,j}^2] \leqslant n/(n-1)$, which with the definition of the $t_{m,i}$'s gives the result.

Let us now check Condition 3. of Theorem (Bai and Zhou (2008)). Since the random variables $(Z_{i,j})_{1 \leqslant i \leqslant n}$ are exchangeable, it is enough to prove that, uniformly in k,

(i) $\mathbb{E}[Z_{1,k}^4] = o(\sqrt{n})$,

(ii) $\mathbb{E}[Z_{1,k}^2 Z_{2,k}^2] - 1 = o(1)$,
(iii) $\mathbb{E}[Z_{1,k}^3 Z_{2,k}] = o(1)$,
(iv) $\sqrt{n}\mathbb{E}[Z_{1,k}^2 Z_{2,k} Z_{3,k}] = o(1)$,
(v) $n\mathbb{E}[Z_{1,k} Z_{2,k} Z_{3,k} Z_{4,k}] = o(1)$, as $n \to \infty$.

Observe that (i) implies (ii). Using (19), by expanding $0 = \left(\sum_{i=1}^n Z_{i,k}\right)^2 \left(\sum_{i=1}^n Z_{i,k}^2\right)$ and taking the expectation, we get that (i) and (iii) imply (iv). By expanding $0 = \left(\sum_{i=1}^n Z_{i,k}\right)^4$, which comes from (19), and by taking the expectation, (i) and (iii) imply (v). Hence, it is enough to prove (i) and (iii) to conclude the proof of Lemma 1.

Let us first prove (i). By the definition of $Z_{1,k}$ given in (4), we get that for all k, $Z_{1,k}^2 \leqslant n$. Hence,

$$Z_{1,k}^2 \leqslant \frac{(W_{1,k} - \overline{W}_k)^2}{2\sigma_k^2} \mathbb{1}_{\{s_k^2 \geqslant 2\sigma_k^2\}} + n \mathbb{1}_{\{s_k^2 > 2\sigma_k^2\}} ,$$

and, by the assumptions on the $W_{i,k}$'s and on the σ_k's,

$$\mathbb{E}(Z_{1,k}^4) \leqslant \frac{W_M^2}{2\kappa^2} + 2n^2 \mathbb{P}(s_k^2 - \sigma_k^2 > \sigma_k^2) .$$

Theorem A of (Serfling, 1980, p. 201) implies that the second term of the previous inequality tends to zero as n tends to infinity uniformly in k, which concludes the proof of (i).

Let us now prove (iii). Using (19), we get $Z_{1,k}^3 \left(\sum_{i=1}^n Z_{i,k}\right) = 0$. By expanding this equation and taking the expectation, we obtain that $\mathbb{E}(Z_{1,k}^4) + \sum_{i=2}^n \mathbb{E}(Z_{1,k}^3 Z_{i,k}) = 0$. Since the $(Z_{i,k})_{1 \leqslant i \leqslant n}$ are exchangeable: $\mathbb{E}(Z_{1,k}^3 Z_{2,k}) = -\mathbb{E}(Z_{1,k}^4)/(n-1) = o(n^{-1/2})$, where the last equality comes from (i).

7.4.2. Proof of Lemma 2

Using (11) and the independence assumptions, we get

$$\begin{aligned}
&\mathrm{Var}(\widetilde{\mathbf{Y}}' \mathbf{H} \widetilde{\mathbf{Y}} | \mathbf{Z}) \\
&= \mathrm{Var}\left[\mathbf{v}'\mathbf{V} \begin{pmatrix} \mathbf{DH} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{V}'\mathbf{v} + 2\sigma^\star \sqrt{1-\eta^\star} \mathbf{v}'\mathbf{V} \begin{pmatrix} \sqrt{\mathbf{D}} \\ \mathbf{0} \end{pmatrix} \mathbf{H}\widetilde{\boldsymbol{\varepsilon}} + \sigma^{\star 2}(1-\eta^\star) \widetilde{\boldsymbol{\varepsilon}}' \mathbf{H} \widetilde{\boldsymbol{\varepsilon}} | \mathbf{Z} \right] \\
&= \mathrm{Var}\left[\mathbf{v}'\mathbf{M}\mathbf{v} | \mathbf{Z} \right] + 4\sigma^{\star 2}(1-\eta^\star) \mathrm{Var}\left[\mathbf{v}'\mathbf{V} \begin{pmatrix} \sqrt{\mathbf{D}} \\ \mathbf{0} \end{pmatrix} \mathbf{H}\widetilde{\boldsymbol{\varepsilon}} | \mathbf{Z} \right] + 2\sigma^{\star 4}(1-\eta^\star)^2 \mathrm{Tr}(\mathbf{H}^2) ,
\end{aligned} \tag{22}$$

where $\mathbf{M} = \mathbf{V} \begin{pmatrix} \mathbf{DH} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{V}'$. Using the independence assumptions, we get that

$$\begin{aligned}
4\sigma^{\star 2}(1-\eta^\star) \mathrm{Var}\left[\mathbf{v}'\mathbf{V} \begin{pmatrix} \sqrt{\mathbf{D}} \\ \mathbf{0} \end{pmatrix} \mathbf{H}\widetilde{\boldsymbol{\varepsilon}} | \mathbf{Z} \right] &= 4\sigma^{\star 4}\eta^\star(1-\eta^\star) \mathrm{Tr}(\mathbf{B}\mathbf{B}') \\
&= 4\sigma^{\star 4}\eta^\star(1-\eta^\star) \mathrm{Tr}(\mathbf{D}\mathbf{H}^2) ,
\end{aligned} \tag{23}$$

where $\mathbf{B} = \mathbf{V}\begin{pmatrix}\sqrt{\mathbf{D}}\\ \mathbf{0}\end{pmatrix}\mathbf{H}$. Moreover, $\mathbb{E}(\mathbf{v}'\mathbf{M}\mathbf{v}|\mathbf{Z}) = \sigma^{\star 2}\eta^\star \operatorname{Tr}(\mathbf{D}^2\mathbf{H}^2)$ and

$$\begin{aligned}\mathbb{E}\left[(\mathbf{v}'\mathbf{M}\mathbf{v})^2|\mathbf{Z}\right] &= \frac{\sigma^{\star 4}\eta^{\star 2}}{q^2}\left[2q^2\sum_{1\leqslant i\neq j\leqslant N}M_{ij}^2 + q^2\sum_{1\leqslant i\neq i'\leqslant N}M_{ii}M_{i'i'} + 3q\sum_{1\leqslant i\leqslant N}M_{ii}^2\right]\\ &= \sigma^{\star 4}\eta^{\star 2}\left[2\operatorname{Tr}(\mathbf{M}^2) - 2\sum_{1\leqslant i\leqslant N}M_{ii}^2 + \operatorname{Tr}(\mathbf{M})^2 - \sum_{1\leqslant i\leqslant N}M_{ii}^2 + \frac{3}{q}\sum_{1\leqslant i\leqslant N}M_{ii}^2\right]\\ &= \sigma^{\star 4}\eta^{\star 2}\left[2\operatorname{Tr}(\mathbf{D}^2\mathbf{H}^2) + \operatorname{Tr}(\mathbf{M})^2 + 3\left(\frac{1}{q}-1\right)\sum_{1\leqslant i\leqslant N}M_{ii}^2\right].\end{aligned}$$

Thus,

$$\operatorname{Var}\left[\mathbf{v}'\mathbf{M}\mathbf{v}\middle|\,\mathbf{Z}\right] = \sigma^{\star 4}\eta^{\star 2}\left[2\operatorname{Tr}(\mathbf{D}^2\mathbf{H}^2) + 3\left(\frac{1}{q}-1\right)\sum_{1\leqslant i\leqslant N}M_{ii}^2\right]. \tag{24}$$

The proof of the equality in Lemma 2 follows from (22), (23) and (24). The proof of the inequality in Lemma 2 follows now from

$$\sum_{1\leqslant i\leqslant N}M_{ii}^2 \leqslant \sum_{1\leqslant i,j\leqslant N}M_{ij}^2 = \operatorname{Tr}[\mathbf{D}^2\mathbf{H}^2].$$

7.4.3. Proof of Lemma 5

Let $\epsilon > 0$ and let $\{\eta_1 < \cdots < \eta_{K(\epsilon)}\}$ be a grid of $[0, 1-\delta]$ such that $|\eta_j - \eta_{j+1}| < \epsilon$ for all $j \in \{0, \ldots, K_\epsilon\}$ then

$$\begin{aligned}\sup_{\eta\in[0,1-\delta]}|L_n(\eta) - L(\eta)| &\leqslant \sup_{j\in\{1,\ldots,K_\epsilon\}}\left[\sup_{\eta'\in[\eta_j,\eta_{j+1}]}|L_n(\eta') - L_n(\eta_j)| + |L_n(\eta_j) - L(\eta_j)|\right.\\ &\qquad\left. + \sup_{\eta'\in[\eta_j,\eta_{j+1}]}|L(\eta_j) - L(\eta')|\right]\\ &\leqslant \epsilon\sup_{\eta\in[0,1-\delta]}|L_n'(\eta)| + \sup_{j\in\{1,\ldots,K_\epsilon\}}|L_n(\eta_j) - L(\eta_j)| + \omega(\epsilon),\end{aligned}$$

where $\omega(\epsilon)$ is the modulus of continuity of L, which is continuous on the compact $[0, 1-\delta]$ and thus uniformly continuous on this compact. Since $\sup_{\eta\in[0,1-\delta]}|L_n'(\eta)| = O_P(1)$ then, for every $\beta > 0$, there exists C such that for all n, $\mathbb{P}(\sup_{\eta\in[0,1-\delta]}|L_n'(\eta)| \geqslant C) \leqslant \beta$. Let $\alpha > 0$ and let us consider the ϵ-grid such that $\epsilon \leqslant \alpha/3C$ and $\omega(\epsilon) \leqslant \alpha/3$,

thus we get that

$$\begin{aligned}
&\mathbb{P}(\sup_{\eta\in[0,1-\delta]}|L_n(\eta)-L(\eta)|\geqslant\alpha)\\
&\leqslant\mathbb{P}(\sup_{\eta\in[0,1-\delta]}\left|L_n'(\eta)\right|\geqslant C)+\mathbb{P}(\sup_{j\in\{1,\dots,K_\epsilon\}}|L_n(\eta_j)-L(\eta_j)|\geqslant\alpha-C\epsilon-\omega(\epsilon))\\
&\leqslant\mathbb{P}(\sup_{\eta\in[0,1-\delta]}\left|L_n'(\eta)\right|\geqslant C)+\mathbb{P}(\sup_{j\in\{1,\dots,K_\epsilon\}}|L_n(\eta_j)-L(\eta_j)|\geqslant\frac{\alpha}{3})\\
&\leqslant\mathbb{P}(\sup_{\eta\in[0,1-\delta]}\left|L_n'(\eta)\right|\geqslant C)+\sum_{j=1}^{K_\epsilon}\mathbb{P}(|L_n(\eta_j)-L(\eta_j)|\geqslant\frac{\alpha}{3}),
\end{aligned}$$

which concludes the proof of Lemma 5 since each term tends to zero as n tends to infinity.

7.4.4. Proof of Lemma 6

The second derivative of L_n is given by

$$\begin{aligned}
L_n''(\eta)=&\left(-\frac{2}{n}\sum_{i=1}^n\frac{\widetilde{Y}_i^2(\lambda_i-1)^2}{\{\eta(\lambda_i-1)+1\}^3}\right)\left(\frac{1}{n}\sum_{i=1}^n\frac{\widetilde{Y}_i^2}{\{\eta(\lambda_i-1)+1\}}\right)^{-1}\\
&+\left(\frac{1}{n}\sum_{i=1}^n\frac{\widetilde{Y}_i^2(\lambda_i-1)}{\{\eta(\lambda_i-1)+1\}^2}\right)^2\left(\frac{1}{n}\sum_{i=1}^n\frac{\widetilde{Y}_i^2}{\{\eta(\lambda_i-1)+1\}}\right)^{-2}\\
&+\frac{1}{n}\sum_{i=1}^n\frac{(\lambda_i-1)^2}{\{\eta(\lambda_i-1)+1\}^2}\,.
\end{aligned}\tag{25}$$

In particular for $\eta=\eta^\star$, we have

$$\frac{1}{n}\sum_{i=1}^n\frac{\widetilde{Y}_i^2}{\{\eta^\star(\lambda_i-1)+1\}}=1+o_P(1),$$

and using as previously Lemma 2, Lemma 3 and the fact that all functions of λ involved in the empirical means are bounded since $\eta^\star>0$, we get

$$\begin{aligned}
\frac{2}{n}\sum_{i=1}^n\frac{\widetilde{Y}_i^2(\lambda_i-1)^2}{\{\eta(\lambda_i-1)+1\}^3}&=\frac{2\sigma^{\star2}}{n}\sum_{i=1}^n\frac{(\lambda_i-1)^2}{\{\eta(\lambda_i-1)+1\}^2}+o_P(1)\\
&=2\sigma^{\star2}\int\frac{(\lambda-1)^2}{\{\eta(\lambda-1)+1\}^2}d\mu_a(\lambda)+o_P(1)
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n\frac{\widetilde{Y}_i^2(\lambda_i-1)}{\{\eta(\lambda_i-1)+1\}^2}&=\frac{\sigma^{\star2}}{n}\sum_{i=1}^n\frac{(\lambda_i-1)}{\{\eta(\lambda_i-1)+1\}}+o_P(1)\\
&=\sigma^{\star2}\int\frac{(\lambda-1)}{\{\eta(\lambda-1)+1\}}d\mu_a(\lambda)+o_P(1)
\end{aligned}$$

leading to

$$L_n''(\eta) = -\sigma^{\star 2}\gamma^2(a, \eta^\star) + o_P(1).$$

Using Slutzky's Lemma and $\hat{\eta} = \eta^\star + o_P(1)$, there just remains to prove that for small enough $\alpha > 0$,

$$\sup_{|\eta - \eta^\star| \leqslant \alpha} |L_n''(\eta) - L_n''(\eta)| = O_p(\alpha).$$

But this comes easily from

$$\sup_{|\eta - \eta^\star| \leqslant \alpha} |L_n''(\eta) - L_n''(\eta)| \leqslant \alpha \sup_{|\eta - \eta^\star|} |L_n^{(3)}(\eta)|$$

where $L_n^{(3)}(\eta)$ is the third derivative of $L_n(\eta)$, and a similar handling of empirical means as before. Indeed, all functions of λ involved are bounded as soon as α is such that $\eta^\star \geqslant 2\alpha$.

Acknowledgments

The authors would like to thank Edouard Maurel-Segala and Maxime Février for stimulating discussions on random matrix theory and Thomas Bourgeron and Roberto Toro for having led us to study this very interesting subject and for the discussions that we had together on genetic topics.

References

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *The Annals of Statistics* **39** 2533–2556.

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral analysis of large dimensional random matrices*, second ed. *Springer Series in Statistics*. Springer, New York. MR2567175 (2011d:60014)

BAI, Z. and ZHOU, W. (2008). Large sample covariance matrices without independence structures in columns. *Statistica Sinica* **18** 425–442.

BEINRUCKER, A., DOGAN, U. and BLANCHARD, G. (2014). Extensions of Stability Selection using subsamples of observations and covariates. arXiv:1407.4916v1.

BONNET, A., TORO, R., LÉVY-LEDUC, C., GASSIAT, E. and BOURGERON, T. (2014). Improvement of heritability estimation through a variable selection approach in sparse high dimensional linear mixed models. Work in progress.

FAN, Y. and LI, R. (2012). Variable selection in mixed effects models. *Annals of Statistics* **40** 2043–2068.

GILMOUR, A., THOMPSON, R. and CULLIS, B. (1995). Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51** 1440–1450.

GOLAN, D. and ROSSET, S. (2011). Accurate estimation of heritability in genome wide studies using random effects models. *Bioinformatics [ISMB/ECCB]* **27** 317-323.

GUAN, Y. and STEPHENS, M. (2011). Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics* **5** 1780–1815.

INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electronic Journal of Statistics* **4** 1476–1526.

KANG, H. M., ZAITLEN, N. A., WADE, C. M., KIRBY, A., HECKERMAN, D., DALY, M. J. and ESKIN, E. (2008). Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics* **178** 1709–1723.

LYNCH, M. and WALSH, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, M.

MARCHENKO, V. A. and PASTUR, L. A. (1968). Distribution of eigenvalues for some sets of random matrices. *Math. USSR, Sb.* **1** 457–483.

MEINSHAUSEN, N. and BÜHLMANN, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72** 417–473.

PATTERSON, H. D. and THOMPSON, R. (1971). Recovery of Inter-block Information when Block Sizes are Unequal. *Biometrika* **58** 545-554.

PIRINEN, M., DONNELLY, P. and SPENCER, C. C. A. (2013). Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. *The Annals of Applied Statistics* **7** 369–390.

SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (1992). *Variance Components*. *Wiley Series in Probability and Statistics*. Wiley, New Jersey.

SERFLING, R. (1980). *Approximation theorems of mathematical statistics*. *Wiley series in probability and mathematical statistics*. Wiley, New York, NY.

TORO, R., POLINE, J.-B., HUGUET, G., LOTH, E., FROUIN, V., BANASCHEWSKI, T., BARKER, G. J., BOKDE, A., BÜCHEL, C., CARVALHO, F., CONROD, P., FAUTH-BÜHLER, M., FLOR, H., GALLINAT, J., GARAVAN, H., GOWLOAN, P., HEINZ, A., ITTERMANN, B., LAWRENCE, C., LEMAÎTRE, H., MANN, K., NEES, F., PAUS, T., PAUSOVA, Z., RIETSCHEL, M., ROBBINS, T., SMOLKA, M., STRÖHLE, A., SCHUMANN, G. and BOURGERON, T. (2014). Genomic architecture of human neuroanatomical diversity. *Molecular Psychiatry*.

WEI, G. C. G. and TANNER, M. A. (1990). A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms. *Journal of the American Statistical Association* **85** 699–704.

YANG, J., BENYAMIN, B., MCEVOY, B. P., GORDON, S., HENDERS, A. K., NYHOLT, D. R., MADDEN, P. A., HEATH, A. C., MARTIN, N. G., MONTGOMERY, G. W., GODDARD, M. E. and VISSCHER, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42** 565–569.

YANG, J., LEE, S. H., GODDARD, M. E. and VISSCHER, P. M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics* **88** 76 - 82.

ZHOU, X., CARBONETTO, P. and STEPHENS, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLoS genetics* **9** e1003264.