



Permutation Testing in the Presence of Polygenic Variation

Mark Abney*

Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America

Received 5 September 2014; Revised 9 January 2015; accepted revised manuscript 26 January 2015.

Published online 10 March 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/gepi.21893

ABSTRACT: This article discusses problems with and solutions to performing valid permutation tests for quantitative trait loci in the presence of polygenic effects. Although permutation testing is a popular approach for determining statistical significance of a test statistic with an unknown distribution—for instance, the maximum of multiple correlated statistics or some omnibus test statistic for a gene, gene-set, or pathway—naïve application of permutations may result in an invalid test. The risk of performing an invalid permutation test is particularly acute in complex trait mapping where polygenicity may combine with a structured population resulting from the presence of families, cryptic relatedness, admixture, or population stratification. I give both analytical derivations and a conceptual understanding of why typical permutation procedures fail and suggest an alternative permutation-based algorithm, MVN_{permute}, that succeeds. In particular, I examine the case where a linear mixed model is used to analyze a quantitative trait and show that both phenotype and genotype permutations may result in an invalid permutation test. I provide a formula that predicts the amount of inflation of the type I error rate depending on the degree of misspecification of the covariance structure of the polygenic effect and the heritability of the trait. I validate this formula by doing simulations, showing that the permutation distribution matches the theoretical expectation, and that my suggested permutation-based test obtains the correct null distribution. Finally, I discuss situations where naïve permutations of the phenotype or genotype are valid and the applicability of the results to other test statistics.

Genet Epidemiol 39:249–258, 2015. © 2015 Wiley Periodicals, Inc.

KEY WORDS: permutation test; polygenic effect; family studies; population structure; type I error rate; QTL

Introduction

In the search for genetic determinants of complex traits, we may be faced with the difficulty of determining the statistical significance of a given test statistic that does not necessarily follow any known probability distribution. This arises when correcting for the multiple comparisons of many correlated tests—for example, to determine genome-wide significance [Abney et al., 2002; Cheng and Palmer, 2010]—or in methods where multiple variants (e.g., rare variants) are aggregated into an omnibus test. Methods that use weights that vary depending on the phenotype data, for instance, typically do not have a known asymptotic distribution and require resampling methods to estimate significance [Fang et al., 2012; Sha et al., 2012]. Even when an asymptotic distribution is known, realities of genetic data, such as population structure or linkage disequilibrium, may result in an inflated false-positive rate [Epstein et al., 2012; Liu et al., 2013; Tintle et al., 2011]. Family-based methods, though often robust to population stratification, can also have false-positive rates above the nominal level [Greco et al., 2014; Kazma and Bailey, 2011]. Permutation tests can be a solution in such cases [Basu and Pan, 2011; Lin and Tang, 2011], but they rely on the assumption that the subjects are independent. This assumption is violated, for instance, in the presence of

population stratification [Epstein et al., 2012; Liu et al., 2013] or familial relatedness [e.g., Abney et al., 2002; Bourgain and Genin, 2005; Kazma and Bailey, 2011], preventing the valid application of a permutation test.

At the heart of the invalidity of a permutation test in the presence of population stratification or relatedness is the presence of polygenic effects and its confounding with genotypes. As I discuss below, this may result in a lack of exchangeability between subjects, a fundamental requirement of a permutation test. It is worth noting that relatedness is not always a barrier to a valid permutation test. For instance, in some model organism breeding designs, exchangeability exists, allowing a valid permutation test [Churchill and Doerge, 1994], and more-complicated breeding designs can also, with careful thought, lead to valid permutation tests [Cheng et al., 2013; Cheng and Palmer, 2010; Churchill and Doerge, 2008; Peirce et al., 2008]. Similarly, given specific restrictions on the types of relatedness that is present among the subjects (e.g., only siblings), it may be possible to formulate a valid permutation test [e.g., Allison et al., 1999; Fang et al., 2012]. However, many forms of population structure, including familial relatedness, can cause confounding that can invalidate a permutation test. Although in a simple population stratification scenario—where a limited number of principal components can adjust for the background genetic confounding—it is possible to formulate a valid permutation test [Epstein et al., 2012], in more-complicated scenarios, as

*Correspondence to: Mark Abney, Department of Human Genetics, University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA. E-mail: abney@uchicago.edu

often exist in human studies—where close or distant relatedness, cryptic or otherwise, may possibly combine with other forms of population structure—a clear statistical framework to help researchers determine the applicability of a permutation test, or how precisely to do such a test, has been lacking.

Here, I consider possible permutation approaches for quantitative traits where arbitrary forms of population structure may exist in the sample. The presence of both population structure and polygenicity leads me to using the linear mixed model (LMM) for multivariate normal data as a foundation on which to build, as this is a standard model used in the genetic analysis of quantitative traits. Although the approaches used here may be applicable to nonnormal types of data, I do not consider this issue. Certainly, permutation tests in LMMs have been considered in the past [e.g., Anderson and Robinson, 2001; Anderson and Ter Braak, 2003], however these studies consider the case where either the “treatment” (e.g., genotypes) is assigned randomly by the researcher or the stochastic components of the model (i.e., the random effect plus the error terms) are independent. Neither of these situations generally holds true in the genetic analysis of a complex trait, where the researcher is not at liberty to assign genotypes at random and the polygenic effect generally results in nonindependence of the random effect. In addition to defining the LMM, I show how misspecification of the covariance matrix leads to an altered asymptotic distribution of the standard test statistic, and how different permutation approaches can be modeled through different forms of misspecification of the covariance matrix. I discuss this issue further below. Finally, I also discuss what precisely should be permuted—phenotypes, residuals, or genotypes—and provide simulation results supporting the analytical findings.

Statistical Model

Here I define the statistical model used in the remainder of this article and the resultant likelihood. Given this model, I propose a standard test statistic that, under the right set of conditions, asymptotically follows a central chi-squared distribution with 1 degree of freedom (χ_1^2). Although it is not really necessary to use a permutation test when the distribution of the statistic is known, its analytical tractability allows for insights that also apply to more general cases. Given the model, I then define exchangeability and the conditions that are needed to ensure that exchangeability holds.

Given n subjects with phenotype data $\mathbf{y} = (y_1, \dots, y_n)^t$, where the superscript $()^t$ indicates transpose; the $n \times p$ covariate data matrix \mathbf{X} , which includes the intercept term; and the predictor of interest (e.g., genotypes) $\mathbf{g} = (g_1, \dots, g_n)^t$, the LMM is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{g}\gamma + \mathbf{e}, \quad (1)$$

where $\boldsymbol{\beta}$ is the vector of parameters for the covariates, γ is the scalar parameter for the predictor of interest, and $\mathbf{e} \sim \text{MVN}(0, \Sigma\sigma^2)$ is an error term. The error term encompasses both a random effect and residual error \mathbf{e}^* , $\mathbf{e} = \mathbf{u} + \mathbf{e}^*$. The residual errors are distributed as independent normals

with variance σ_e^2 . In the genetic context the random effect \mathbf{u} will typically be the polygenic effect, and if we further assume that it is the sum of a large number of independent additive genetic effects in an outbred sample, the central limit theorem dictates that \mathbf{u} is multivariate normally distributed with correlation matrix \mathbf{K} [Lange, 1978] with the result extended to the case of inbreeding and dominance variance in Abney et al. [2000]. Although I do not assume a particular structure for Σ , a common parameterization in a genetic LMM is $\Sigma = \mathbf{K}h^2 + \mathbf{I}(1 - h^2)$, where \mathbf{K} is an additive genetic relationship matrix (GRM), \mathbf{I} is the identity matrix, and h^2 is the narrow-sense heritability. The matrix \mathbf{K} may be estimated from available genotype data or determined from a pedigree, in which case it is equal to 2Φ , where Φ is the matrix of kinship coefficients. The log likelihood of this model is

$$l = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log |\Sigma| - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{g}\gamma)^t \Sigma^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{g}\gamma). \quad (2)$$

The quantity of interest is the parameter γ , and under the null hypothesis $\gamma = 0$. To test against the alternative $\gamma \neq 0$, the statistic $T = \frac{\hat{\gamma}^2}{\widehat{\text{Var}}(\hat{\gamma})}$, where $\hat{\gamma}$ is the best linear unbiased estimator (BLUE, equivalently the maximum-likelihood estimator) of γ , has a χ_1^2 distribution under the null hypothesis when σ^2 is known. In practice, we use

$$\hat{T} = \frac{\hat{\gamma}}{\widehat{\text{Var}}(\hat{\gamma})}, \quad (3)$$

where the estimated variance $\widehat{\text{Var}}(\hat{\gamma})$ uses an estimator S^2 in place of the true variance σ^2 . In an LMM approach $S^2 = \frac{1}{n-p-1} (\mathbf{y} - \hat{\mathbf{y}})^t \Sigma^{-1} (\mathbf{y} - \hat{\mathbf{y}})$ and is an unbiased estimator of σ^2 . This results in \hat{T} being asymptotically χ_1^2 distributed. However, in a genetic analysis Σ is not always known, leading to the question of what the distribution of \hat{T} is when Σ is misspecified.

In genetic analyses of complex traits, using an LMM with a misspecified covariance matrix is likely a common occurrence. Perhaps the simplest example of this is when unrelated individuals are unknowingly sampled from two populations, with different allele frequencies at the tested marker, but are assumed to be from a single population. If the trait is associated with population membership, we see an inflated false-positive rate. This sort of confounding is easily corrected by including either a covariate with an indicator of population membership or a block structured correlation matrix with elements equal to 1 when a pair is from the same population or 0 when they are not. At the other end of the population structure scale, misspecification may occur in family studies with a known pedigree when the pedigree is wrong or incomplete. In fact, even if the pedigree is known without error, misspecification exists when the kinship matrix (as computed from the pedigree) is used as the additive GRM because under a polygenic model the kinship coefficients give only the expected identity by descent (IBD) sharing across the genome whereas the correlation in phenotype values will be the result of the realized IBD sharing. In spite of this last form of misfit,

the successful use of the kinship coefficient in the GRM over many decades of pedigree studies in both humans and animals suggests a degree of robustness to the use of the expected covariance in place of the realized covariance.

In order to quantify the effects of covariance matrix misspecification on hypothesis testing in the LMM, in the Appendix I derive the distribution of the test statistic \hat{T} in the case where the incorrect matrix Ψ is used instead of Σ . I find that \hat{T} is asymptotically distributed as a scaled chi-squared distribution, $\frac{\hat{T}}{\eta} \sim \chi^2_1$, where η is a constant. In the case of no covariates and assuming that both \mathbf{y} and \mathbf{g} have been centered by their mean values, η takes the form

$$\eta = \frac{n \mathbf{g}^t \Psi^{-1} \Sigma \Psi^{-1} \mathbf{g}}{\mathbf{g}^t \Psi^{-1} \mathbf{g} \text{Tr}(\Sigma \Psi^{-1})}.$$

The scalar η is, in essence, the genomic control [Bacanu et al., 2002; Devlin and Roeder, 1999] parameter, and its general analytical form in the asymptotic limit of very large sample sizes is in the Appendix. In any real dataset we do not know Σ and, hence, cannot know η , but having an analytical form will allow us to determine the degree of miscalibration in \hat{T} for hypothesized circumstances, as we will see below.

Confounding, Exchangeability, and Permutations

If we expect to perform a permutation test given purely observational data, we should also be concerned with the possibility of confounding. Consider the linear model $y = \mu + x_1\beta_1 + x_2\beta_2 + e$, where e is an independent error and we wish to test the null hypothesis $\beta_2 = 0$. In a designed experiment we can ensure that x_2 has no confounders by random assignment of its values to each subject, and we can safely permute the labels of x_2 to obtain a valid test. With purely observational data, however, x_2 may be confounded with x_1 due to unknown structure in the data, and permutations of the x_2 subject labels would result in an invalid test. Note that joint permutation of x_1 and x_2 would be valid, but this strategy fails when x_1 is not observed. This situation arises in genetic studies when there is population structure in the sample. In this case x_2 would be genotype and x_1 is a predictor that also depends on the population structure, for example, a polygenic effect or an indicator of population membership (there would be $P - 1$ such indicators for P populations in the sample) with each population having a distinct effect on the outcome y . If genotype x_2 is dependent on the population structure, we would not want to naively permute all the subject labels of x_2 , as this would give an incorrect type 1 error rate. In this case, even if population membership is not recorded, it can often be inferred if there is sufficient genetic data.

Less well appreciated is that, given structured data, confounding can occur even when the predictor x_2 is sampled independently from the unobserved structured-population predictor x_1 . Note that this independence is conditional given the covariance matrices (representing population structure) for x_1 and x_2 in the sense that they are vectors drawn independently from two distinct multivariate distributions each with a given covariance matrix. To understand this,

consider an example where the subjects are connected by some pedigree with x_2 being their genotypes at a marker that has no genetic effect and is not in linkage disequilibrium with any causal locus, and with x_1 representing the polygenic effect. The genotype x_2 and polygenic effect x_1 necessarily have the same correlation matrix \mathbf{K} induced by the pedigree. Thus, when we sample x_1 and x_2 independently from their distributions, having the same correlation matrix \mathbf{K} , it is actually equivalent to x_1 and x_2 being *conditionally* independent given the population or pedigree structure. Marginally (i.e., unconditional on the underlying structure) x_1 and x_2 are correlated with each other. That is, similar genotype values will tend to match up with similar polygenic (and, hence, trait) values simply because these vectors have a similar correlation structure. It is this correlation between x_1 and x_2 that leads to confounding when ignoring x_1 in the model. Note that this argument does not depend on whether \mathbf{K} is the result of population structure resulting from a pedigree or the block diagonal form, with constant off-diagonals in each block, that results from assuming a population-specific genetic effect. Every genetic trait will depend on genotypes with some population structure correlation \mathbf{K} , resulting in confounding when testing a genetic marker that also has correlation \mathbf{K} , thus altering the type 1 error away from the expected amount unless the confounding is corrected for in the test [Newman et al., 2001]. Conversely, if the elements of either x_2 or x_1 are *unconditionally* independent—meaning either one has the identity as the covariance matrix—there will be no confounding of x_2 with y . For instance, if the x_2 genotypes were independent binomials as would be the case in an unstructured population, there would neither be inflation of the test statistic nor any problems with permuting the values of x_2 . Unfortunately, with observational data verifying the absence of confounding, and the permissibility of a permutation test, may not be possible.

Developing a permutation test for observational data requires assessing whether the permuted quantities are exchangeable. Quantities are exchangeable if, upon permutation of the labels of those quantities, their distribution function is unchanged [Bernardo and Smith, 2000, Sec. 4.2]. In particular, because we want to know the distribution of the test statistic under the null hypothesis, we require exchangeability when $\gamma = 0$. In an LMM, the natural quantities to permute are the residuals $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$. In the Appendix I show that the residuals are exchangeable only in the special case where $\Sigma_{ii} = a$ and $\Sigma_{ij} = b$, $i \neq j$ for some scalar constants a and b , where Σ_{ij} is the i, j -th element of Σ . Note that because we do not in general know $\boldsymbol{\beta}$ or σ^2 but must instead estimate them, even when Σ has an exchangeable structure permuting the residuals technically provides only an approximate permutation test, though the approximation tends to be very accurate [Anderson and Robinson, 2001].

In general, when using an LMM to model polygenic variation the matrix Σ will not have an exchangeable structure. Nevertheless, we might undertake a permutation test where the residuals are permuted under the assumption that the phenotype has an exchangeable correlation matrix Ψ rather than true correlation matrix Σ . The fundamental question

is will these permutations give an unbiased estimate of the threshold for rejecting the null hypothesis at some specified false positive rate? To address this question exactly, we would need to understand the properties of the order statistics $T_{(k)}$, $k = 1, \dots, n!$ under permutations of the residuals. Instead, I address this in an approximate, but more intuitive, approach by treating the statistics $T_{(k)}$ as samples from a distribution with covariance matrix that has an exchangeable structure. In the simulation results below, we will see that the empiric distribution we get by doing permutations closely matches the distribution obtained from assuming $\Psi = \mathbf{I}$.

Simulations

The simulations are done in a sample of 1,415 Hutterite individuals with a known 13-generation pedigree [Abney et al., 2000]. Phenotypes for the sample are generated under the null model to have a mean of 3.0 and covariance matrix $\Sigma\sigma^2$ with $\Sigma = 2\Phi h^2 + \mathbf{I}(1 - h^2)$, with Φ the kinship coefficient matrix as computed from the pedigree and h^2 the narrow-sense heritability. Genotypes are simulated by randomly assigning the founders of the pedigree a genotype from a biallelic marker with minor allele frequency of 0.3 and using Mendelian segregation to randomly determine the genotypes of all the other pedigree members.

First, I address the question of what happens when “naive” permutations are done. That is, the residuals under the null model are permuted regardless of whether the correlation matrix is exchangeable and the new phenotype (i.e., the covariate effects plus the permuted residuals) is put through the same LMM analysis as the original data. More precisely, we assume the null model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ where $\mathbf{e} \sim \text{MVN}(0, \sigma^2 \Sigma)$. That is, analyses done under the null model use exactly the same model as that used to generate the data. Using Equation (2) with $\gamma = 0$, I first fit the null model and obtain maximum-likelihood estimates for the parameters, $\hat{\boldsymbol{\beta}}_0$, \hat{h}_0^2 , $\hat{\sigma}_0^2$. Using generalized least squares (GLS), I test the null hypothesis $\gamma = 0$ against the alternative $\gamma \neq 0$ using the test statistic \hat{T} (Eq. (3)) computed under the alternative model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{g}\gamma + \mathbf{e}, \text{ where } \mathbf{e} \sim \text{MVN}(0, \sigma^2 \hat{\Sigma}_0) \text{ and} \\ \hat{\Sigma}_0 = 2\Phi \hat{h}_0^2 + \mathbf{I}(1 - \hat{h}_0^2). \quad (4)$$

Note that \hat{T} is necessarily asymptotically distributed as a χ_1^2 because $\hat{\Sigma}_0$ asymptotically converges to the true correlation matrix Σ . Also note that σ^2 in Equation (4) is estimated by the sample variance when computing \hat{T} . I want to compare this asymptotic distribution with the empirical distribution one obtains by doing naive permutations. To do this, I first obtain the estimated residuals under the null, $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_0$. I permute these residuals to obtain $\hat{\mathbf{e}}_{\pi_1}$ and a new phenotype vector $\mathbf{y}_{\pi_1} = \mathbf{X}\hat{\boldsymbol{\beta}}_0 + \hat{\mathbf{e}}_{\pi_1}$. Under the alternative model of Equation (4) but with \mathbf{y}_{π_1} in place of \mathbf{y} , I obtain a test statistic $\hat{T}_{\pi_1}^v$, where the v superscript indicates the use of naive permutations. I repeat this process $L = 10^4$ times to obtain $\hat{T}_{\pi_1}^v, \dots, \hat{T}_{\pi_L}^v$. If doing naive permutations were to provide the correct empiric

distribution for our original test statistic \hat{T} , then the samples $\hat{T}_{\pi_1}^v, \dots, \hat{T}_{\pi_L}^v$ should follow a χ_1^2 distribution.

As shown in Figure 1 the empiric distribution clearly fails to follow the desired asymptotic distribution. The reason for this is that the permutations fail to maintain the correlation structure of the original phenotype data. As discussed above, this form of permutation would give an accurate distribution only when the true correlation matrix of the estimated residuals has an exchangeable structure. We can use the methods in the Appendix to quantify the inaccuracy of the empiric permutation distribution. We can model the statistics $\hat{T}_{\pi_1}^v, \dots, \hat{T}_{\pi_L}^v$ as coming from the distribution $\eta \chi_1^2$ that results from assuming the incorrect correlation matrix $\Psi = \mathbf{I}$ rather than the correlation matrix $\hat{\Sigma}_0$. Using the theoretically computed value of η , as given in the Appendix, and plotting $\hat{T}_{\pi_1}^v/\eta, \dots, \hat{T}_{\pi_M}^v/\eta$ against a χ_1^2 in Figure 1 B we see that the distributions match well. From this, we see that computing the significance of \hat{T} from $\{\hat{T}_{\pi_i}^v\}$ would lead to an anticonservative estimated level of significance. For instance, to get nominal levels of significance of 10^{-4} and 10^{-5} , the permuted distribution would select threshold levels of 12.7 and 16.3, respectively. Because the actual distribution of the test statistic is χ_1^2 , the observed type 1 error rates would be 3.7×10^{-4} and 5.4×10^{-5} , respectively, a substantial inflation.

Though it is not possible to do an exact permutation test when the residuals have a nonexchangeable correlation matrix, it is possible to do a valid permutation-based test. The approach (referred to here as MVNpermute) is described in Abney et al. [2002] and the Appendix and it relies on the fact that there exists a linear transformation of the residuals that results in a vector (i.e., the transformed residuals) whose covariance matrix is proportional to the identity matrix, and is therefore exchangeable. Because MVNpermute is based on permutations of an invertible transformation of the phenotype residuals, all structure in the genotype data (e.g., linkage disequilibrium, allele frequencies) is preserved. Inverting the transformation following permutations then results in new simulated datasets that maintain the structure in the entire original data (i.e., phenotype correlations and genotype structure).

I repeated the above simulations but with the permutations generated using MVNpermute. This gave statistics $\hat{T}_{\pi_1}^M, \dots, \hat{T}_{\pi_L}^M$, where the M superscript indicates the use of MVNpermute. As shown in Figure 2, this results in statistics that follow the expected distribution. That is, by first decorrelating the residuals—ensuring exchangeability for a normally distributed trait—permutations allow us to estimate the proper threshold for a given false-positive rate. In practice, MVNpermute is not necessary to determine the P -value at a single SNP, but obtaining L permutation-based datasets $\{\mathbf{y}_{\pi_i}\}$ allows us to do an empiric multiple testing correction to determine genome-wide significance, for instance [Abney et al., 2002].

It is not unusual to realize that permuting the phenotypes (or rather the residuals) does not result in a valid permutation test when individuals are related. A possible alternative is to permute the genotypes instead of the phenotypes. A

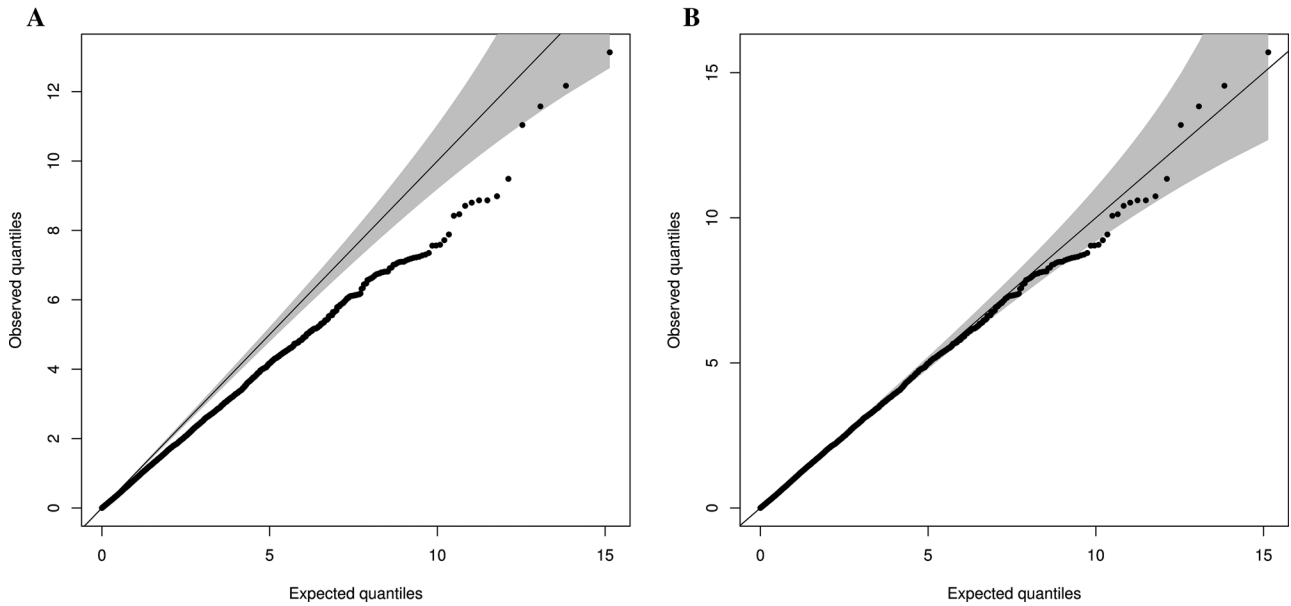


Figure 1. QQ plots under naive phenotype residual permutations. In both plots the expected quantiles are for a χ^2 distribution and the shaded area is the 95% confidence region. (A) The observed quantiles are the values of the test statistic under permutations of the trait values. (B) The observed quantiles are the values in (A) divided by the theoretical inflation factor.

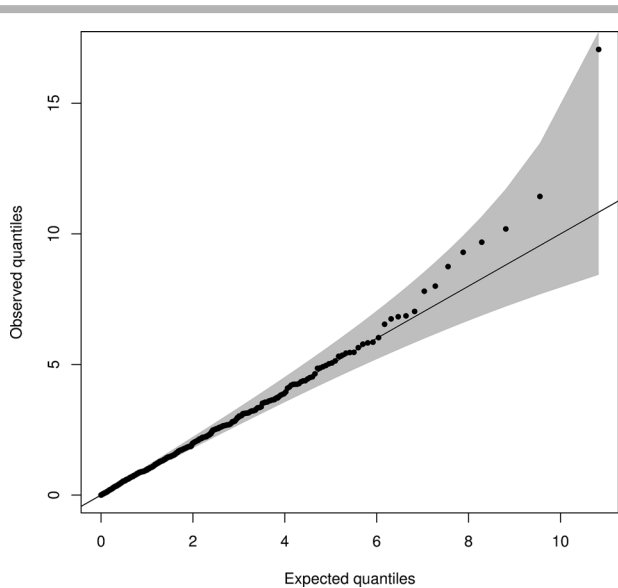


Figure 2. QQ plot of the MVNpermute method. The observed quantiles are the values of the test statistic from 10,000 MVNpermutations, whereas the theoretical quantiles are those from a χ^2_1 distribution. The shaded region is the 95% confidence bounds.

rationale is that the LMM inference is based on the conditional distribution of the phenotypes given the genotypes. As the phenotypes remain fixed while new genotypes get assigned to individuals via permutation, the correlation structure in the phenotypes is preserved, resulting in a valid permutation test. In fact, many statistics used in complex trait mapping assume a distribution that is conditional on the

genotype data, even if the form of the test statistic distribution is not known. Permuting the genotype data, then, to estimate this distribution seems a natural approach.

To understand the consequences of a genotype permutation procedure on an arbitrary test statistic, let us first consider the standard ordinary least squares (OLS) statistic,

$$\tilde{T}_r = \frac{\tilde{\gamma}}{\sqrt{\text{Var}(\tilde{\gamma})}} \quad (5)$$

$$\begin{pmatrix} \tilde{\beta} \\ \tilde{\gamma} \end{pmatrix} = (\mathbf{M}^t \mathbf{M})^{-1} \mathbf{M}^t \mathbf{y} \quad (6)$$

$$\mathbf{M} = (\mathbf{X} \ \mathbf{g}), \quad (7)$$

where $\tilde{\cdot}$ indicates estimation with a scaled covariance matrix Ψ rather than Σ (in this case $\Psi = \mathbf{I}$). That is, the statistic we use does not explicitly account for the polygenic effect. We may be aware that relatedness in our sample will result in \tilde{T}_r not being χ^2_1 distributed and, thus, perform genotype permutations to obtain the empiric distribution of \tilde{T}_r . To see if genotype permutations recover the correct distribution, I performed $L = 10^4$ permutations of the genotype data while keeping the phenotype data constant to obtain $\{\tilde{T}_{r,\pi_1}, \dots, \tilde{T}_{r,\pi_L}\}$ where π_1, \dots, π_L index the permutations. I then compare this to a sample from the true null distribution that I obtain by performing L gene dropping simulations, $\tilde{T}_{r,1}, \dots, \tilde{T}_{r,L}$. The results in Figure 3 show that the distribution obtained by genotype permutations is highly deflated relative to the distribution from gene dropping. Using this approach to obtain an empiric threshold of significance would result in a highly inflated false-positive rate.

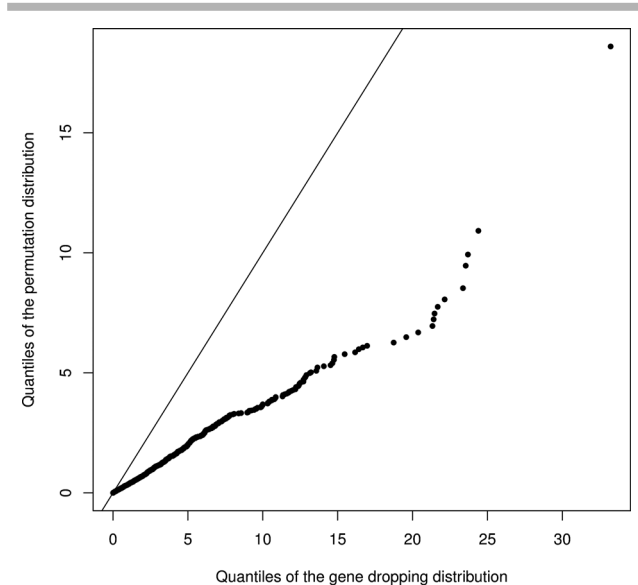


Figure 3. Q-Q plot of the empiric null distribution for the OLS statistic against the expected null distribution. The expected null distribution is a sample obtained by doing gene dropping. The solid line is the $y = x$ line.

The source of the problem with genotype permutations can be understood by returning to the notion of confounding. Because genotypes were not randomly assigned by the researcher to subjects, the absence of confounding is not guaranteed. In fact, the subject genotypes are correlated as a consequence of Mendelian segregation and all markers in the genome, whether causative or not, share the same pedigree for a given set of individuals. Hence, the covariance of the marker being tested is equal (up to a scalar constant) to the covariance of the polygenic effect. If genotypes are permuted, the similarity of the covariance structures of phenotype and genotype will not be preserved. Thus, I expect that the null distribution of an arbitrary test statistic, not just the OLS statistic, will not be correctly estimated by genotype permutations, in general.

Although the null distribution of an arbitrary test statistic cannot be inferred from genotype permutations, the null distribution of the GLS statistic can be. That is, if instead of using \hat{T}_r we use \hat{T} as defined in Equation (3) and do the genotype permutation procedure as described above, we find the permutation distribution of \hat{T} matches the gene dropping permutation distribution (data not shown). We can understand this result by looking at the definition of the GLS statistic \hat{T} . We can view the GLS statistic as the OLS statistic computed on the data following a decorrelation step. That is, if we define $\Sigma^{1/2}$ as the symmetric square root matrix of Σ and $\mathbf{z} = \Sigma^{-1/2}\mathbf{y}$, $\mathbf{W} = \Sigma^{-1/2}\mathbf{X}$, $\mathbf{f} = \Sigma^{-1/2}\mathbf{g}$, $\boldsymbol{\epsilon} = \Sigma^{-1/2}\mathbf{e}$, then we obtain the linear model $\mathbf{z} = \mathbf{W}\boldsymbol{\beta} + \mathbf{f}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon} \sim \text{MVN}(0, \mathbf{I}\sigma^2)$. The GLS statistic on the original data is equivalent to the OLS statistic on the decorrelated data \mathbf{z} . Because our new trait data \mathbf{z} are normally distributed and uncorrelated, they are independent and can no longer be confounded

with the genotypes under the null hypothesis. In the absence of confounding, then, permuting the genotypes recovers the true null distribution of the test statistic \hat{T} .

Software

The MVNpermute algorithm is implemented in the R programming language and is available for download from the Comprehensive R Archive Network (<http://cran.r-project.org>) as the “MVNpermute” package.

Discussion

The fundamental challenge with performing a permutation test is ensuring exchangeability in the permuted quantities. In a genetic association test, it is generally not possible to do an exact permutation test when the trait under study has a polygenic component. The reason is that confounding due to population structure exists between the genotype being tested and the unknown polygenic effect, both of which have similar covariance structures. Only when all individuals are equally related, as in an F2 cross [Churchill and Doerge, 1994], will a naive permutation approach obtain the correct type 1 error rate. Nevertheless, with an accurate estimate of the trait covariance structure, it may be possible to remove the confounding and perform a valid permutation test. I have described an approach we have previously proposed [Abney et al., 2002] for removing the correlation in the phenotype residuals and shown that it generates the correct null distribution. Strictly, the method is valid when the phenotype data are multivariate normally distributed, where removing the correlation is sufficient to ensure exchangeability. Another permutation approach was proposed by Aulchenko et al. [2007]. They estimate the polygenic effect and obtain estimates of the residual error term. Although, under multivariate normality, the residual errors in Equation (1) are exchangeable, the *estimated* residual errors, in general, will not be. Nevertheless, this may be a case of “close enough,” allowing for a reasonably accurate estimate of significance thresholds, though I have not investigated this question.

Other resampling strategies are possible, though they also have limitations. Gene dropping is one such approach. In this strategy one simulates the Mendelian segregation of the founder genotypes through all descendants. Because Mendelian segregation is random and independent of the phenotype, it provides a valid distribution of the test statistic under the null hypothesis. The primary difficulties with gene dropping is the need for a complete pedigree and knowledge of the founder genotypes. If the pedigree is known, but the founder genotypes are not, it may be possible to reconstruct, or simply guess, them from the available data. Doing so, however, runs the risk of introducing unknown biases as the observed genotypes may be confounded with the phenotypes. On the other hand, if the pedigree is not known, gene dropping is simply not feasible.

Instead of gene dropping we might try to permute genotypes, leaving the covariance structure of the phenotype

intact. As discussed above, for an arbitrary test statistic this does not necessarily result in a valid test as permutations of the genotypes will not preserve their covariance structure. In addition, applying a “decorrelating” transformation to the genotypes is not sufficient to ensure their exchangeability because, unlike the multivariate normal distribution of the phenotype, the joint distribution of the genotypes has higher-order dependencies. Nevertheless, in the case of a multivariate normal phenotype being analyzed with a linear mixed model, the standard test statistic \hat{T} naturally transforms the trait data to being independent. Once independent, any sets of dependent or independent genotypes, including permutations of the original ones, can be used to recover the correct null distribution of the test statistic. This approach has been used in mouse cross data to obtain proper genome-wide significance levels [Cheng et al., 2013; Cheng and Palmer, 2010], and more recently in humans [Zhang et al., 2014]. It seems likely that any test statistic that removes the correlation in the phenotype data and does not depend on Mendelian segregation under the null hypothesis would allow genotype permutations to be valid, though I have not investigated this further. An additional caveat arises, however, with genotype permutations when there are other covariates. In particular, if any of the covariates are associated with the genotype, genotype permutations may estimate the incorrect null distribution. That is, it will give the null distribution for when the covariate and genotype are not associated rather than for when they are. This may arise, for instance, when a covariate is itself a genetic trait, or when it is a principal component vector obtained from a population structure analysis. It may also arise when testing effects such as gene-by-environment interaction. In this situation the null model has a nonzero genetic main effect. In general, an approximate permutation test of interaction effects is done by computing and permuting outcome residuals [Anderson and Ter Braak, 2003] as done by MVNpermute. Additional work is needed to understand this effectiveness and validity of MVNpermute and genotype permutations in the presence of genetic interaction terms and of genotype permutations when other genetic predictors are in the null model.

Another resampling strategy is the parametric bootstrap. In this approach one assumes the phenotypes follow a particular parametric distribution with parameter values equal to those estimated from the observed data under the null hypothesis. Samples are then drawn from this distribution and a test statistic computed for each sample, thus obtaining an empiric null distribution. For instance, one might assume the phenotype follows a multivariate normal distribution with fixed effect parameters and variance components estimated by maximum likelihood under the null model. Drawing many phenotypes from this distribution and testing the genotype at an SNP against each randomly drawn phenotype provides a null distribution for the test statistic. This approach relies on the parametric distribution accurately representing the observed data. Insofar as the data deviate from the assumed distribution, biases in the estimated significance threshold may ensue. A true permutation test has the advantage of not needing to make such parametric assumptions.

The MVNpermute method also relies on certain distributional assumptions. Namely, that exchangeability under the null is determined by the structure of the correlation matrix. Intuitively, this assumption appears weaker than those used in a parametric bootstrap, suggesting that there may be greater robustness to the permutation-based approach, though I have not investigated this question.

The analyses I performed here were based on using a statistic known to asymptotically follow a χ_1^2 distribution. This allowed me to easily show the invalidity of particular permutation procedures. In practice, one would not need a permutation test for such a statistic, but the lessons extend to other statistics as well. For instance, we might want to determine statistical significance after correcting for multiple correlated tests, as when doing a genome-wide scan or a scan over a smaller region. In this case the statistic of interest would be the maximum over all statistics in the scan. Similarly, statistics that jointly combine information across SNPs or use phenotype dependent weights, for which there may not be a clear generative model, may not have a known distribution under the null hypothesis. Situations such as these would benefit from a permutation test, if one exists. The presence of polygenic variation may make a true permutation test difficult or impossible, but a permutation-based test may be achievable by carefully considering the sources of correlation, or nonexchangeability, in the data. Hopefully, the examples and discussion I provided here will help bring insight into the development of such tests.

Acknowledgments

I would like to thank Abraham Palmer, Riyan Cheng, Peter Carbonetto, and Lei Sun for repeatedly raising this topic with me and motivating me to write this manuscript. Peter Carbonetto also provided helpful comments on the MVNpermute R code. This work was supported by NIH grant HG002899. The author has no conflicts of interest to declare.

Appendix A

Distribution with a Misspecified Covariance Matrix

The LMM of the main text is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{g}\gamma + \mathbf{e}$, where $\mathbf{e} \sim N(0, \Sigma\sigma^2)$. The asymptotic distribution of the test statistic $\hat{T} = \hat{\gamma}^2 / \widehat{\text{Var}}(\hat{\gamma})$ is χ_1^2 under the null hypothesis, but only if we use the correct scaled covariance matrix Σ in our estimate of $\hat{\gamma}$ and its variance. If we misspecify this matrix, the distribution of \hat{T} becomes a scaled chi-squared distribution $\eta \cdot \chi_1^2$ with the scalar η depending on the amount of misspecification. To determine η we can derive \hat{T} and its distribution assuming that we have used the incorrect covariance matrix $\Psi\sigma^2$ in place of the correct $\Sigma\sigma^2$. First, we obtain the BLUE for γ . If we let $\Psi^{1/2}$ be the symmetric positive definite square root matrix of Ψ and define

$$\begin{aligned} \mathbf{z} &= \Psi^{-1/2}\mathbf{y} \\ \mathbf{W} &= \Psi^{-1/2}\mathbf{X} \\ \mathbf{f} &= \Psi^{-1/2}\mathbf{g} \\ \boldsymbol{\epsilon} &= \Psi^{-1/2}\mathbf{e} \end{aligned}$$

$$\mathbf{H}_W = \mathbf{W}(\mathbf{W}^t\mathbf{W})^{-1}\mathbf{W}^t \quad (\text{A.1})$$

then the BLUE for γ is

$$\tilde{\gamma} = [\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\mathbf{f}]^{-1}\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\mathbf{z},$$

where $\tilde{\cdot}$ indicates an estimate using Ψ rather than Σ . The variance of this estimator is

$$\begin{aligned} \text{Var}(\tilde{\gamma}) &= [\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\mathbf{f}]^{-1}\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\Theta(\mathbf{I} - \mathbf{H}_W) \\ &\quad \times \mathbf{f}[\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\mathbf{f}]^{-1}\sigma^2. \end{aligned} \quad (\text{A.2})$$

where $\text{Var}(\mathbf{z}) = \Theta\sigma^2 = \Psi^{-1/2}\Sigma\Psi^{-1/2}\sigma^2$. The statistic

$$T = \frac{\tilde{\gamma}^2}{\text{Var}(\tilde{\gamma})} = \frac{(\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\mathbf{z})^2}{\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\Theta(\mathbf{I} - \mathbf{H}_W)\mathbf{f}\sigma^2} \quad (\text{A.3})$$

is χ_1^2 because \mathbf{z} is multivariate normal. The statistic T , however, is not an adequate test statistic because it depends on the unknown matrix Σ and the unknown parameter σ^2 . In practice we use the test statistic

$$\tilde{T} = \frac{\tilde{\gamma}^2}{\widetilde{\text{Var}}(\tilde{\gamma})} = \frac{(\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\mathbf{z})^2}{\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\mathbf{f}S^2}, \quad (\text{A.4})$$

where in place of σ^2 we have the sample variance

$$S^2 = \frac{1}{n-p-1}(\mathbf{z} - \bar{\mathbf{z}})^t(\mathbf{z} - \bar{\mathbf{z}}),$$

where $\bar{\mathbf{z}} = \mathbf{H}_A\mathbf{z} = \mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t\mathbf{z}$ and $\mathbf{A} = (\mathbf{W} \ \mathbf{f}) = \Psi^{-1/2}\mathbf{M}$, $\mathbf{M} = (\mathbf{X} \ \mathbf{g})$. It is straightforward to show that if $\Psi = \Sigma$ then the expectation $E(S^2) = \sigma^2$. In general, however, we have,

$$\begin{aligned} E(S^2) &= \frac{1}{n-p-1} \text{Tr}((\mathbf{I} - \mathbf{H}_A)\Theta)\sigma^2 \\ &= \frac{\sigma^2}{n-p-1} \left[\text{Tr}(\Sigma\Psi^{-1}) - \text{Tr}([\mathbf{M}^t\Psi^{-1}\mathbf{M}]^{-1} \right. \\ &\quad \left. \times \mathbf{M}^t\Psi^{-1}\Sigma\Psi^{-1}\mathbf{M}) \right]. \end{aligned} \quad (\text{A.5})$$

The result of misspecifying the covariance matrix is given by the following lemma.

Lemma 1. Let $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{g}\gamma, \Sigma\sigma^2)$ with Σ nonnegative definite and Ψ be some symmetric nonnegative definite matrix. Define $\Theta = \Psi^{-1/2}\Sigma\Psi^{-1/2}$, $\mathbf{M} = (\mathbf{X} \ \mathbf{g})$, with \mathbf{z} , \mathbf{f} , \mathbf{H}_W , Θ as in Equation (A.1) and \tilde{T} as in Equation (A.4). If $\lambda_1(\Sigma\Psi^{-1}) = o(n^{1/2})$, where $\lambda_1(\Sigma\Psi^{-1})$ is the largest eigenvalue of matrix $\Sigma\Psi^{-1}$ (equivalently the largest eigenvalue of Θ), then as $n \rightarrow \infty$

$\frac{\tilde{T}}{\eta} \rightarrow \chi_1^2$ in distribution, where

$$\eta = \frac{(n-p-1)\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\Theta(\mathbf{I} - \mathbf{H}_W)\mathbf{f}}{\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\mathbf{f}[\text{Tr}(\Sigma\Psi^{-1}) - \text{Tr}([\mathbf{M}^t\Psi^{-1}\mathbf{M}]^{-1}\mathbf{M}^t\Psi^{-1}\Sigma\Psi^{-1}\mathbf{M})]}. \quad (\text{A.6})$$

Proof. We can write $\tilde{T} = (\tilde{\gamma}^2/\text{Var}(\tilde{\gamma})) \times (\text{Var}(\tilde{\gamma})/\widetilde{\text{Var}}(\tilde{\gamma}))$, which is the product of a χ_1^2 random variable, as given in Equation (A.3), and the ratio of the true to estimated variance of $\tilde{\gamma}$. The true variance is given by Equation (A.2), but the estimated variance is

$$\widetilde{\text{Var}}(\tilde{\gamma}) = [\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\mathbf{f}]^{-1}S^2, \text{ where}$$

$$S^2 = \frac{1}{n-p-1}\mathbf{z}^t(\mathbf{I} - \mathbf{H}_A)\mathbf{z}.$$

Thus, the test statistic is

$$\tilde{T} = \frac{\tilde{\gamma}^2}{\text{Var}(\tilde{\gamma})} \times \frac{\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\Theta(\mathbf{I} - \mathbf{H}_W)\mathbf{f}}{\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\mathbf{f}} \times \frac{\sigma^2}{S^2},$$

the product of a χ_1^2 random variable and the quantity

$$\eta = \frac{\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\Theta(\mathbf{I} - \mathbf{H}_W)\mathbf{f}}{\mathbf{f}^t(\mathbf{I} - \mathbf{H}_W)\mathbf{f}} \times \frac{\sigma^2}{S^2}.$$

If as $n \rightarrow \infty$ $\text{Var}(S^2) \rightarrow 0$, then $\frac{\sigma^2}{S^2} \rightarrow \frac{\sigma^2}{E(S^2)}$ in probability, where $E(S^2)$ is given by Equation (A.5). Thus, we obtain Equation (A.6) of Lemma 1 when $\text{Var}(S^2) \rightarrow 0$. We can obtain a sufficient condition for $\text{Var}(S^2) \rightarrow 0$ by considering

$$\begin{aligned} \text{Var}(S^2) &= \frac{1}{(n-p-1)^2} \text{Var}(\mathbf{z}^t(\mathbf{I} - \mathbf{H}_A)\mathbf{z}) \\ &= \frac{2\sigma^4}{(n-p-1)^2} \text{Tr}(\Theta(\mathbf{I} - \mathbf{H}_A)\Theta(\mathbf{I} - \mathbf{H}_A)). \end{aligned}$$

Thus, $\text{Var}(S^2) \rightarrow 0$ when $\text{Tr}(\Theta(\mathbf{I} - \mathbf{H}_A)\Theta(\mathbf{I} - \mathbf{H}_A)) = o(n^2)$.

Now, consider the following eigenvalue result [Zhang, 2011, Theorem 8.12, p. 274]. Let $\lambda_i(\mathbf{P})$ be the i th eigenvalue for some $n \times n$ matrix \mathbf{P} ordered such that $\lambda_1(\mathbf{P}) \geq \lambda_2(\mathbf{P}) \geq \dots \geq \lambda_n(\mathbf{P})$. Then, for any $n \times n$ nonnegative definite, Hermitian matrices \mathbf{P} , \mathbf{Q}

$$\lambda_i(\mathbf{P})\lambda_n(\mathbf{Q}) \leq \lambda_i(\mathbf{PQ}) \leq \lambda_i(\mathbf{P})\lambda_1(\mathbf{Q}).$$

It follows that $\lambda_i(\Theta[\mathbf{I} - \mathbf{H}_A]) \leq \lambda_i(\Theta)$ because Θ is symmetric, nonnegative definite and $\mathbf{I} - \mathbf{H}_A$ is symmetric and idempotent with all eigenvalues equal to 0 or 1. Furthermore, $\Theta[\mathbf{I} - \mathbf{H}_A]$ is nonnegative definite because $\lambda_i(\Theta[\mathbf{I} - \mathbf{H}_A]) \geq \lambda_i(\Theta)\lambda_n(\mathbf{I} - \mathbf{H}_A) \geq 0$, and thus $\lambda_i(\Theta[\mathbf{I} - \mathbf{H}_A])^2 \leq \lambda_i(\Theta)^2$. Recalling that the trace of a matrix is the sum of the eigenvalues we have $\text{Tr}(\Theta(\mathbf{I} - \mathbf{H}_A)\Theta(\mathbf{I} - \mathbf{H}_A)) \leq \text{Tr}(\Theta^2)$. If we define $\mathbf{B} = \Sigma\Psi^{-1}$, then $\text{Tr}(\Theta^2) = \text{Tr}(\mathbf{B}^2)$. Hence,

$$\begin{aligned} \text{Var}(S^2) &= \frac{2\sigma^4}{(n-p-1)^2} \text{Tr}(\Theta(\mathbf{I} - \mathbf{H}_A)\Theta(\mathbf{I} - \mathbf{H}_A)) \\ &\leq \frac{2\sigma^4}{(n-p-1)^2} \text{Tr}(\mathbf{B}^2) \\ &\leq \frac{2\sigma^4}{(n-p-1)^2} n[\lambda_1(\mathbf{B})]^2. \end{aligned}$$

Thus, a sufficient condition for $\text{Var}(S^2) \rightarrow 0$, and hence $\sigma^2/S^2 \rightarrow \sigma^2/E(S^2)$ in probability, is $\lambda_1(\Sigma\Psi^{-1}) = o(n^{1/2})$. \square

Note that a possibly tighter sufficient condition is $\text{Tr}(\mathbf{B}^2) = \sum_{i=1}^n \lambda_i(\mathbf{B})^2 = o(n^2)$.

Exchangeability of the Multivariate Normal Distribution

Given a random vector $\mathbf{y} = (y_1, \dots, y_n)$ distributed as a multivariate normal $f(\mathbf{y}) = N(\boldsymbol{\mu}, \Sigma)$, under what conditions are the elements of \mathbf{y} exchangeable? If we let \mathbf{P} be a permutation matrix so that $\mathbf{P}\mathbf{y}$ is a permutation of the elements of \mathbf{y} , then \mathbf{y} is exchangeable when $f(\mathbf{y}) = f(\mathbf{P}\mathbf{y})$ for every

permutation matrix \mathbf{P} [Bernardo and Smith, 2000, Sec. 4.2]. Taking the log of both sides, this reduces to,

$$(\mathbf{y} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) = (\mathbf{P}\mathbf{y} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{P}\mathbf{y} - \boldsymbol{\mu}),$$

which implies the condition,

$$\mathbf{y}^t (\mathbf{P}^t \boldsymbol{\Sigma}^{-1} \mathbf{P} - \boldsymbol{\Sigma}^{-1}) \mathbf{y} + 2\boldsymbol{\mu}^t \boldsymbol{\Sigma}^{-1} (\mathbf{P}\mathbf{y} - \boldsymbol{\mu}) = 0. \quad (\text{A.7})$$

In order for Equation (A.7) to hold for every \mathbf{y} and \mathbf{P} , each of the two terms must be zero. If the first term is to be zero for all \mathbf{y} and \mathbf{P} , then $\mathbf{P}^t \boldsymbol{\Sigma}^{-1} \mathbf{P} = \boldsymbol{\Sigma}^{-1}$. This condition is met if and only if all the diagonal elements $\Sigma_{ii} = \nu$, for some constant ν , and all the off-diagonals $\Sigma_{ij} = \nu\rho$, for some constant ρ . To set the second term to zero we first note that for some vector \mathbf{w} , the condition $\mathbf{w}^t (\mathbf{P}\mathbf{y} - \boldsymbol{\mu}) = 0$ implies $\mathbf{w} = \alpha \mathbf{1}$, for some constant α and where $\mathbf{1} = (1, \dots, 1)^t$. Hence, given the structure of $\boldsymbol{\Sigma}$ we already determined, the second term is zero whenever $\boldsymbol{\mu} = (\mu, \dots, \mu)^t$ for some constant μ .

In the text, the null model corresponds to $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$. Let us assume the $\boldsymbol{\Sigma}$ has an exchangeable structure. In general, however, the vector $\mathbf{X}\boldsymbol{\beta} \neq \mu \mathbf{1}$ for some fixed μ and \mathbf{y} is still not exchangeable. The vector $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, though, does satisfy the requirements for exchangeability, and a permutation test can be based on permutations of the residuals. In practice, the vector $\boldsymbol{\beta}$ is unknown and must be estimated, resulting in estimated residuals $\hat{\mathbf{e}}$. Permuting the estimated residuals, then, results in only an asymptotically exact permutation test when $\boldsymbol{\Sigma}$ has an exchangeable structure.

MVNpermute Algorithm

The permutation-based algorithm was originally presented in Abney et al. [2002, pp. 926–927] and I review it here for completeness. Assume the outcome \mathbf{y} follows the model as given in Equation (1) in the main text, and let the errors \mathbf{e} have known covariance matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}\sigma^2$. In practice, this matrix may not be known, in which case a consistent estimator will maintain the asymptotic properties of the permutation based procedure. For instance, a maximum-likelihood estimate under the null model ($\gamma = 0$) could be used. The residuals under the null model $\hat{\mathbf{e}}_0 = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_0$, where $\hat{\boldsymbol{\beta}}_0 = (\mathbf{X}^t \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \boldsymbol{\Omega}^{-1} \mathbf{y}$, have covariance matrix $\mathbf{V}^* = \boldsymbol{\Omega} - \mathbf{X}(\mathbf{X}^t \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^t$.

The goal, then, is to transform the residuals, which are not exchangeable, to a new vector whose elements are exchangeable. We can accomplish this by premultiplying Equation (1) by \mathbf{C}^{-t} where \mathbf{C} is given by the Cholesky decomposition $\boldsymbol{\Omega} = \mathbf{C}^t \mathbf{C}$. The resulting model under the null hypothesis $\gamma = 0$ is $\mathbf{z} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\mathbf{z} = \mathbf{C}^{-t} \mathbf{y}$, $\mathbf{W} = \mathbf{C}^{-t} \mathbf{X}$, and $\boldsymbol{\epsilon} = \mathbf{C}^{-t} \mathbf{e}$. The covariance matrix of the residuals $\hat{\boldsymbol{\epsilon}} = \mathbf{z} - \mathbf{W}\hat{\boldsymbol{\beta}}_0$ is $\mathbf{V} = \mathbf{I} - \mathbf{W}(\mathbf{W}^t \mathbf{W})^{-1} \mathbf{W}^t$. Note that \mathbf{V} is symmetric and idempotent (i.e., $\mathbf{V}^2 = \mathbf{V}$) and if \mathbf{X} is of rank p then \mathbf{V} has rank $n - p$. By the spectral theorem we can make the decomposition $\mathbf{V} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^t$, where $\boldsymbol{\Lambda}$ is a diagonal matrix with the first $n - p$ elements equal to the eigenvalue 1 and the last p elements equal to the eigenvalue 0, and \mathbf{U} is the matrix whose columns are eigenvectors. Let $\mathbf{U} = (\mathbf{U}_1 \mathbf{U}_0)$, where \mathbf{U}_1 is the matrix whose $n - p$ columns are the eigenvectors associated with eigenvalue 1. Then, we have $\mathbf{V} = \mathbf{U}_1 \mathbf{U}_1^t$ and $\mathbf{U}_1^t \mathbf{U}_1 = \mathbf{I}_{n-p}$. The vector $\boldsymbol{\xi} = \mathbf{U}_1^t \hat{\boldsymbol{\epsilon}}$ has covariance matrix $\mathbf{U}_1^t \mathbf{V} \mathbf{U}_1 = \mathbf{I}_{n-p}$ and its elements, under the assumption of multivariate normality of the residuals, are exchangeable. The elements of $\boldsymbol{\xi}$ are now permuted to obtain $\boldsymbol{\xi}^\pi = \mathbf{P}\boldsymbol{\xi}$ where \mathbf{P} is a permutation matrix, and then transformed by \mathbf{U}_1 to get $\hat{\boldsymbol{\epsilon}}_\pi = \mathbf{U}_1 \boldsymbol{\xi}^\pi$. Note

that I use the convention that “ π ” used as a superscript denotes the variable is permuted, whereas “ π ” used as a subscript denotes that the variable is derived from permuted and nonpermuted quantities. A new shuffled dataset obtained from the permutation is

$$\begin{aligned} \mathbf{y}_\pi &= \mathbf{X}\hat{\boldsymbol{\beta}}_0 + \mathbf{C}^t \hat{\boldsymbol{\epsilon}}_\pi \\ &= \mathbf{X}\hat{\boldsymbol{\beta}}_0 + \mathbf{C}^t \mathbf{U}_1 \mathbf{P} \mathbf{U}_1^t \mathbf{C}^{-t} \hat{\mathbf{e}}_0. \end{aligned}$$

The MVNpermute algorithm is coded as an R function that takes as input the outcome vector \mathbf{y} , matrix of covariates \mathbf{X} , assumed covariance matrix $\boldsymbol{\Omega}$, and the desired number of permutations. The output is a matrix with columns being the permutation-based outcome vectors. The MVNpermute function is available as a download from CRAN.

References

- Abney M, McPeck MS, Ober C. 2000. Estimation of variance components of quantitative traits in inbred populations. *Am J Hum Genet* 66:629–650.
- Abney M, Ober C, McPeck MS. 2002. Quantitative-trait homozygosity and association mapping and empirical genome-wide significance in large, complex pedigrees: fasting serum-insulin level in the Hutterites. *Am J Hum Genet* 70:920–934.
- Allison DB, Heo M, Kaplan N, Martin ER. 1999. Sibling-based tests of linkage and association for quantitative traits. *Am J Hum Genet* 64:1754–1763.
- Anderson M, Robinson J. 2001. Permutation tests for linear models. *Aust N Z J Stat* 43:75–88.
- Anderson MJ, Ter Braak CJF. 2003. Permutation tests for multi-factorial analysis of variance. *J Stat Comput Simul* 73:85–113.
- Aulchenko YS, de Koning DJ, Haley C. 2007. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* 177:577–585.
- Bacanu SA, Devlin B, Roeder K. 2002. Association studies for quantitative traits in structured populations association studies for quantitative traits in structured populations. *Genet Epidemiol* 22:78–93.
- Basu S, Pan W. 2011. Comparison of statistical tests for disease association with rare variants. *Genet Epidemiol* 35:606–619.
- Bernardo JM, Smith AFM. 2000. *Bayesian Theory*. New York: Wiley.
- Bourgain C, Genin E. 2005. Complex trait mapping in isolated populations: are specific statistical methods required? *Eur J Hum Genet* 13:698–706.
- Cheng R, Palmer AA. 2013. A simulation study of permutation, bootstrap and gene dropping for assessing statistical significance in the case of unequal relatedness. *Genetics* 193:1015–1018.
- Cheng R, Lim JE, Samocha KE, Sokoloff G, Abney M, Skol AD, Palmer AA. 2010. Genome-wide association studies and the problem of relatedness among advanced intercross lines and other highly recombinant populations. *Genetics* 185:1033–1044.
- Churchill GA, Doerge RW. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* 138:963–971.
- Churchill GA, Doerge RW. 2008. Naive application of permutation testing leads to inflated type I error rates. *Genetics* 178:609–610.
- Devlin B, Roeder K. 1999. Genomic control for association studies. *Biometrics* 55:997–1004.
- Epstein MP, Duncan R, Jiang Y, Conneely KN, Allen AS, Satten GA. 2012. A permutation procedure to correct for confounders in case-control studies, including tests of rare variation. *Am J Hum Genet* 91:215–223.
- Fang S, Sha Q, Zhang S. 2012. Two adaptive weighting methods to test for rare variant associations in family-based designs. *Genet Epidemiol* 36:499–507.
- Greco B, Luedtke A, Hainline A, Alvarez C, Beck A, Tintle NL. 2014. Application of family-based tests of association for rare variants to pathways. *BMC Proc* 8(Suppl 1):S105.
- Kazma R, Bailey JN. 2011. Population-based and family-based designs to analyze rare variants in complex diseases. *Genet Epidemiol* 35(Suppl 1):S41–S47.
- Lange K. 1978. Central limit theorems for pedigrees. *J Math Biol* 6:59–66.
- Lin DY, Tang ZZ. 2011. A general framework for detecting disease associations with rare variants in sequencing studies. *Am J Hum Genet* 89:354–367.
- Liu Q, Nicolae DL, Chen LS. 2013. Marbled inflation from population structure in gene-based association studies with rare variants. *Genet Epidemiol* 37:286–292.

- Newman DL, Abney M, McPeck MS, Ober C, Cox NJ. 2001. The importance of genealogy in determining genetic associations with complex traits. *Am J Hum Genet* 69:1146–1148.
- Peirce JL, Broman KW, Lu L, Chesler EJ, Zhou G, Airey DC, Birmingham AE, Williams RW. 2008. Genome reshuffling for advanced intercross permutation (GRAIP): simulation and permutation for advanced intercross population analysis. *PLoS One* 3:e1977.
- Sha Q, Wang X, Wang X, Zhang S. 2012. Detecting association of rare and common variants by testing an optimally weighted combination of variants. *Genet Epidemiol* 36:561–571.
- Tintle N, Aschard H, Hu I, Nock N, Wang H, Pugh E. 2011. Inflated type I error rates when using aggregation methods to analyze rare variants in the 1000 genomes project exon sequencing data in unrelated individuals: summary results from Group 7 at Genetic Analysis Workshop 17. *Genet Epidemiol* 35(Suppl 1): S56–S60.
- Zhang F. 2011. *Matrix Theory: Basic Results and Techniques (2nd ed.)*. New York: Springer.
- Zhang Q, Wang L, Koboldt D, Boreki IB, Province MA. 2014. Adjusting family relatedness in data-driven burden test of rare variants. *Genet Epidemiol* 38:722–727.