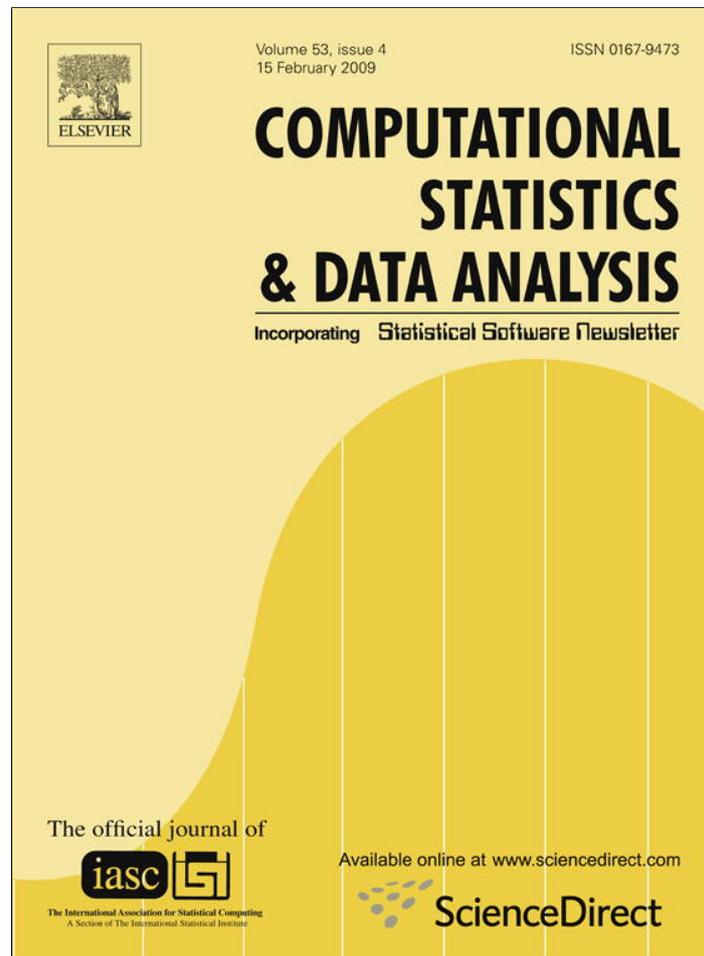


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/cstda

Obtaining similar null distributions in the normal linear model using computational methods

Irene Klugkist*, Herbert Hoijtink

University of Utrecht, The Netherlands

ARTICLE INFO

Article history:

Received 2 April 2007

Received in revised form 10 October 2008

Accepted 11 October 2008

Available online 1 November 2008

ABSTRACT

Hypothesis testing when the null hypothesis belongs to the univariate or multivariate normal linear model is discussed. More specifically it is shown how data can be replicated from the null distribution conditional on the sufficient statistics for the parameters of the null hypothesis at hand. This distribution will be called the similar null distribution of the data. It is shown how the similar null distribution of the data can be used to obtain level alpha tests for any test statistic that is of interest. The p -value that is obtained using the distribution of the test statistic conditional on the sufficient statistics is called the similar p -value.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Null hypothesis testing can be complicated for composite null hypotheses, i.e., in the presence of nuisance parameters. To compute a significance or p -value, two ingredients are needed in addition to the null hypothesis and observed data \mathbf{y} : A test statistic T and its null distribution $g(T|H_0)$. The derivation of $g(T|H_0)$ is not always straightforward and therefore null distributions are often approximated, leading to asymptotic p -values (cf. Bayarri and Berger (2000) and Robins et al. (2000)).

In this paper, composite null hypotheses that have the form of a normal linear model will be considered. In these models, sufficient statistics for the nuisance parameters typically are available. As a consequence, similar tests could be formulated leading to so-called similar p -values (p_{sim}). In similar tests the nuisance parameters are eliminated by conditioning on the observed sufficient statistics for the unknown parameters under the null model. Denoting the sufficient statistics by $S_r(\mathbf{y})$ ($r = 1, \dots, R$), where R denotes the total number of sufficient statistics for the null hypothesis at hand, a similar test evaluates a test statistic T using the distribution $g(T|S_1(\mathbf{y}), \dots, S_R(\mathbf{y}))$ (Lehmann, 1997, p. 140). Note that a similar test is a level alpha test, that is

$$P(p_{sim} < \alpha | H_0) = \alpha.$$

In general, the computation of p_{sim} can, however, be complicated; the distribution $g(T|S_1(\mathbf{y}), \dots, S_R(\mathbf{y}))$ is often difficult to obtain. As a consequence, one is either limited in the choice of T (restricting to those test statistics for which the distribution is available) or subjected to other procedures that may not have the desirable properties of the similar test. In the context of composite null hypotheses, Bayarri and Berger (2000), for instance, proposed two alternative formulations of p -values based on conditional predictive distributions: the conditional predictive p -value and the partial posterior predictive p -value. Bayarri et al. (2006) present sampling methods that can be applied to obtain these p -values for various test statistics in a univariate setting.

In this paper, we will present a method to obtain the *similar null distribution of the data* for both the univariate and multivariate normal linear null model. Note that the expression 'similar null distribution of the data' is not standard

* Corresponding address: Department of Methodology and Statistics, University of Utrecht, PO Box 80140, 3508 TC Utrecht, The Netherlands. Tel.: +31 30 253 5473; fax: +31 30 253 5797.

E-mail address: i.klugkist@uu.nl (I. Klugkist).

terminology. What we mean is that the distribution of the data (and not that of a specific test statistic) under the null is obtained, and that this distribution has the property that the expected proportion of rejections of the null hypothesis computed with respect to this distribution is level alpha for any test (Lehmann, 1997, p. 140). With the similar null distribution of the data, a similar p -value can be computed for any test statistic T .

Compared to the work presented by Bayarri and Berger (2000) and Bayarri et al. (2006) our approach is: (i) different, since we present a sampling procedure providing a similar p -value, that is, p_{sim} , (ii) more flexible, in the sense that one sampling approach delivers the null distribution of any test statistic of interest, (iii) more limited, since we consider only normal linear null hypotheses (although this does contain a wide range of applications and commonly used tests), and (iv) an extension, since we will present a sampling procedure for both univariate and multivariate null hypotheses.

The method proposed in this paper is based on repeatedly sampling data vectors from the similar null distribution of the data. Stated differently, the approach renders $g(\mathbf{z}|S_1(\mathbf{z}) = S_1(\mathbf{y}), \dots, S_R(\mathbf{z}) = S_R(\mathbf{y}))$, where \mathbf{z} denotes a data-matrix from the null population and \mathbf{y} the observed data vector. The distribution of the data \mathbf{z} conditional on the sufficient statistics is uniform, that is

$$g(\mathbf{z}|S_1(\mathbf{z}) = S_1(\mathbf{y}), \dots, S_R(\mathbf{z}) = S_R(\mathbf{y})) = c. \tag{1}$$

For notational convenience, in the sequel $\{S_1(\mathbf{z}) = S_1(\mathbf{y}), \dots, S_R(\mathbf{z}) = S_R(\mathbf{y})\}$ will be abbreviated by \mathbf{S} . An advantage of rendering the similar distribution of the data is that a similar test for any function of the data (i.e. any test statistic T) is obtained. The aim of the paper is therefore not to present a new test or specific test statistic, but to show that the similar null distribution of the data can be obtained and renders similar tests for any test statistic. Computation of $T(\mathbf{z})$ for each \mathbf{z} renders the similar distribution of the test statistic. This distribution is denoted $g(T(\mathbf{z})|\mathbf{S})$ and renders a similar p -value, that is

$$p_{sim} = P(T(\mathbf{z}) \geq T(\mathbf{y})|\mathbf{S}). \tag{2}$$

As was stated before, both univariate and multivariate normal null hypotheses will be considered. The general univariate normal null hypothesis for $i = 1, \dots, N$ respondents is:

$$y_i = \sum_{k=1}^K \beta_k d_{ki} + \sum_{p=1}^P \alpha_p x_{pi} + \varepsilon_i, \quad \text{with } \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2), \tag{3}$$

where y_i denotes the score of the i -th respondent on the outcome variable, x_{pi} the score of the i -th respondent on the p -th continuous predictor ($p = 1, \dots, P$), and, $d_{ki} = 1$ indicates that respondent i is in group k ($k = 1, \dots, K$), while $d_{ki} = 0$ indicates the opposite. Furthermore, β_k denotes the (adjusted) mean in group k , α_p is the regression coefficient relating the p -th continuous predictor to the dependent variable, and σ^2 is the residual variance. Specific null hypotheses are obtained by choosing appropriate values for K and P . For instance, the null hypothesis of a four group analysis of variance (ANOVA): $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta$, is (3) specified with $K = 4$, $d_{ki} = 1$ for $i = 1, \dots, N$ and $P = 0$. The standard ANOVA F -test tests against H_1 : not all β_j equal (where $j = 1, \dots, J$ denotes the number of groups under the alternative). In this case, the distribution $g(T|H_0)$ is known and exact p -values are obtained (provided the general ANOVA model assumptions are met). Testing against for instance H_2 : $\beta_1 \leq \beta_2 \leq \beta_3 \leq \beta_4$ is less straightforward and requires level probabilities (see Barlow et al. (1972), Robertson et al. (1988) and Silvapulle and Sen (2005)). Level probabilities are still easy to compute for the simple hypothesis just presented, but becomes more complicated for larger K , the addition of P covariates and variations in the type and number of order constraints. As another example of an alternative hypothesis that may be of interest, consider testing against H_1 using a test statistic that is robust against 'long-tailed error distributions which may not be symmetric' (see Silvapulle (1992a,b)). Silvapulle presents a test statistic that is useful in this context, but for this test only the asymptotic distribution is known. It could also be of interest to use a robust test statistic for testing against H_2 but, as far as is known by the authors, no applications of this have appeared yet in the literature. The reason may be that although the specification of a useful test statistic T is not that difficult, deriving the distribution $g(T|H_0)$ often is. With the approach presented in this paper, level alpha tests are obtained for any test statistic of interest as long as the null hypothesis is a specification of (3).

Furthermore, multivariate normal models will be considered. The general multivariate normal null hypothesis is:

$$\mathbf{y}_i = \sum_{k=1}^K \boldsymbol{\beta}_k d_{ki} + \sum_{p=1}^P \boldsymbol{\alpha}_p x_{pi} + \boldsymbol{\varepsilon}_i, \quad \text{with } \boldsymbol{\varepsilon}_i \stackrel{iid}{\sim} N(0, \boldsymbol{\Sigma}) \tag{4}$$

where \mathbf{y}_i , $\boldsymbol{\alpha}_p$, $\boldsymbol{\beta}_k$, and $\boldsymbol{\varepsilon}_i$ are vectors of length Q , and $\boldsymbol{\Sigma}$ is a $Q \times Q$ covariance matrix. A description will be given as to how the multivariate generalization of \mathbf{z} , that is, the $N \times Q$ matrix \mathbf{Z} can be sampled from the multivariate counterpart of (1). A Gibbs sampler will be used, that is based on repeated application of the univariate sampling procedure to each of the Q dependent variables. Specific multivariate null hypotheses are obtained by choosing appropriate values for K , P and Q . As an example, consider (4) with $K = 3$, $d_{ki} = 1$ for $i = 1, \dots, N$, $P = 0$ and $Q = 4$. This is, for instance, the null hypothesis for a three groups multivariate analysis of variance (MANOVA) with H_0 : $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_3 = \boldsymbol{\beta}$ against the alternative H_1 : not all $\boldsymbol{\beta}_j$ equal ($j = 1, \dots, J$ denote the groups under the alternative hypothesis). This is a classic multivariate testing problem for which a number of test-statistics have been developed: Wilk's lambda, Roy's largest root, the Hotelling-Lawley trace and Pillai-Bartlett trace (see, for example, Tatsuoka (1988), pp. 88–95 and 285–288, Stevens (1996), pp. 225–226). Many references to F -approximations of the null distributions of these statistics can be found in the literature. However, the finite sample null distribution is not known for each of these statistics (see, for example, Muller (1998)). As in the univariate

case, it may also be interesting and more informative to test H_0 against an ordered alternative, e.g. $H_2 : \beta_1 \leq \beta_2 \leq \beta_3$. We found only one reference with respect to the evaluation of this type of hypothesis in the literature. Sasabuchi et al. (2003) present a likelihood ratio test and a method to compute the supremum of its upper tail probability under the null hypothesis. In general, finding the (exact) distribution $g(T|H_0)$ for a test statistic that is useful in the context of multivariate ordered hypotheses is very difficult. With the approach presented in this paper, level alpha tests are also obtained for any test statistic of interest for null hypotheses that are a specification of (4).

The paper is organized as follows. In Section 2, an introduction is given on how to sample from the null distribution of the data based on a simple univariate null hypothesis (Section 2.1). Subsequently the approach will be presented for the general univariate null (Section 2.2) and a specification of the general univariate null model (Section 2.3). In Section 3, multivariate distributions are sampled using an extension of the approach that is based on a Gibbs sampling procedure (Section 3.1). Illustrations will be provided in the context of one-sided multivariate testing (Section 3.2) and hypotheses imposing order restrictions on multivariate means (Section 3.3). The paper will be concluded with a discussion in Section 4.

2. The univariate null hypothesis

2.1. Replication of data for a simple null hypothesis

Model (3) with $K = 1$ and $P = 0$ leads to the simple null hypothesis $y_i = \beta_1 + \varepsilon_i$, with $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$. The sufficient statistics for the nuisance parameters (β_1 and σ^2) of this null hypothesis are $S_1(\mathbf{y}) = \sum_{i=1}^N y_i$ and $S_2(\mathbf{y}) = \sum_{i=1}^N y_i^2$. The null distribution of the data conditional on the observed values of the sufficient statistics is not available in closed form, but data $\mathbf{z} = [z_1, \dots, z_N]$ can be replicated from

$$g\left(\mathbf{z} \mid \sum_{i=1}^N z_i = \sum_{i=1}^N y_i, \sum_{i=1}^N z_i^2 = \sum_{i=1}^N y_i^2\right) = c, \tag{5}$$

where the constant c implies that, conditional on the sufficient statistics, each data vector \mathbf{z} has the same density. In this section replication of a vector \mathbf{z} from (5) will be explained. In Section 2.1.1, a small numerical illustration for $N = 3$ is provided.

The replicated vectors \mathbf{z} give a discrete representation of the desired distribution (5). The precision of this representation increases with the number of replications and can therefore be controlled by the user. As will be shown, for numbers of replications that are easily obtained with modern personal computers, the obtained degree of precision is very high.

To replicate a vector \mathbf{z} from (5) amounts to randomly choosing any vector for which $\sum_{i=1}^N z_i = \sum_{i=1}^N y_i$ and $\sum_{i=1}^N z_i^2 = \sum_{i=1}^N y_i^2$. By solving $\sum_{i=1}^N z_i = \sum_{i=1}^N y_i$ for z_N and substituting $z_N = \sum_{i=1}^N y_i - \sum_{i=1}^{N-1} z_i$ in $\sum_{i=1}^N z_i^2 = \sum_{i=1}^N y_i^2$ the following condition is obtained:

$$\sum_{i=1}^{N-1} z_i^2 + \sum_{i=1}^{N-2} \sum_{l=i+1}^{N-1} 2z_i z_l - \sum_{i=1}^N y_i \sum_{i=1}^{N-1} z_i = \frac{1}{2} \sum_{i=1}^N y_i^2 - \frac{1}{2} \left(\sum_{i=1}^N y_i\right)^2, \tag{6}$$

which is the equation of an ellipsoid in \mathbb{R}^{N-1} . Note that (6) is a function of $\mathbf{z}^- = [z_1, \dots, z_{N-1}]$. To replicate a data vector \mathbf{z}^- , a point on this ellipsoid is needed, where each point has an equal probability of being sampled. Note also, that when a vector \mathbf{z}^- is obtained, the value for z_N can subsequently be computed using $z_N = \sum_{i=1}^N y_i - \sum_{i=1}^{N-1} z_i$.

It will be shown that (6) is equal to the formula for isodensity ellipsoids of a multivariate normal distribution of dimension $N - 1$ and this property will be used in the sampling scheme. To obtain an arbitrary point \mathbf{z}^- on the ellipsoid, the multivariate normal distribution corresponding to (6) must be defined. The general formula for isodensity ellipsoids of $N_{N-1}(\boldsymbol{\mu}^-, \boldsymbol{\Sigma}^-)$ is

$$(\mathbf{z}^- - \boldsymbol{\mu}^-)'(\boldsymbol{\Sigma}^-)^{-1}(\mathbf{z}^- - \boldsymbol{\mu}^-) = d, \tag{7}$$

where each different value of the constant d specifies a different ellipsoid (Tatsuoka, 1988, p. 70). The notation $\boldsymbol{\mu}^-$ and $\boldsymbol{\Sigma}^-$ is used to emphasize that the dimension of the multivariate normal equals the dimension of \mathbf{z}^- .

Denote $(\boldsymbol{\Sigma}^-)^{-1} = \boldsymbol{\Omega}^-$, with elements ω_{il} . Dropping the matrix notation, (7) can be written as

$$\sum_{i=1}^{N-1} \omega_{ii} z_i^2 + \sum_{i=1}^{N-2} \sum_{l=i+1}^{N-1} 2\omega_{il} z_i z_l - \sum_{i=1}^{N-1} \sum_{l=1}^{N-1} 2\omega_{il} \mu_l z_i = d - \sum_{i=1}^{N-1} \omega_{ii} \mu_i^2 - \sum_{i=1}^{N-2} \sum_{l=i+1}^{N-1} 2\omega_{il} \mu_l \mu_i. \tag{8}$$

Note that (8) is also a function of \mathbf{z}^- .

Comparing (8) with (6) it can be seen that for the simple null hypothesis

$$\boldsymbol{\mu}^- = \left[\frac{1}{N-1} \sum_{i=1}^N y_i, \dots, \frac{1}{N-1} \sum_{i=1}^N y_i \right], \tag{9}$$

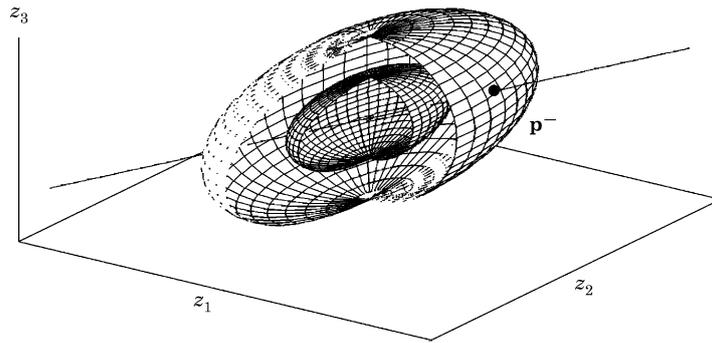


Fig. 1. Isodensity ellipsoids of a three dimensional normal distribution.

$$\Omega^- = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{2} \\ \frac{1}{2} & 1 & & \\ \vdots & & \ddots & \vdots \\ \frac{1}{2} & \cdots & \frac{1}{2} & 1 \end{bmatrix}, \tag{10}$$

and

$$d = \frac{1}{2} \sum_{i=1}^N y_i^2 + \frac{1}{2N-2} \left(\sum_{i=1}^N y_i \right)^2. \tag{11}$$

As will be explained hereafter, to replicate a vector \mathbf{z}^- on the ellipsoid that is specified by (6), the value of d , that is (11), is not required. Therefore, in the sequel, the derivation of d will be omitted. The procedure to obtain a point from exactly the ellipsoid that is specified by (6), consists of the following steps:

1. A point $\mathbf{p}^- = [p_1, \dots, p_{N-1}]$ is sampled from $N_{N-1}(\boldsymbol{\mu}^-, \boldsymbol{\Sigma}^-)$ with $\boldsymbol{\mu}^-$ as specified in (9), and $\boldsymbol{\Sigma}^-$ via Ω^- as specified in (10). This point \mathbf{p}^- belongs to one of the many isodensity ellipsoids, but typically not the required one, as specified in (6). This is illustrated in Fig. 1, for $N = 4$. The inner ellipsoid in this figure represents (6) but the sampled point \mathbf{p}^- belongs to the outer ellipsoid.
2. A line through the sampled point \mathbf{p}^- and the centroid of the ellipsoids $\boldsymbol{\mu}^-$ is drawn (see Fig. 1). This line has two intersections with the required ellipsoid, i.e. the inner ellipsoid in Fig. 1. One of the intersections (each with a probability .5) renders a vector \mathbf{z}^- . Since \mathbf{z}^- must be on the line, we have:

$$\mathbf{z}^- = \boldsymbol{\mu}^- + \lambda(\mathbf{p}^- - \boldsymbol{\mu}^-). \tag{12}$$

Note that (12) consists of $(N - 1)$ equations and N unknowns (\mathbf{z}^- and λ) and that (6) adds the N -th equation. Solving (6) and (12) for λ renders two solutions, from which one random choice is inserted in (12) to compute \mathbf{z}^- .

Obviously, not every vector \mathbf{p}^- has an equal probability of being sampled. However, all vectors that belong to one specific ellipsoid do have an equal probability and this holds for all isodensity ellipses. Since each vector (on any ellipsoid) has a corresponding position on the required ellipsoid that is obtained using the centroid of the ellipsoids, it holds that any value on the required ellipsoid has an equal probability of being sampled.

3. The last element is computed using $z_N = \sum_{i=1}^N y_i - \sum_{i=1}^{N-1} z_i$.

Going through steps 1 to 3 renders one replicated data vector \mathbf{z} from (5).

2.1.1. A small numerical illustration

In this section the derivation of the sampling procedure is illustrated using a small example of three observations, $\mathbf{y} = [1, 2, 3]$, with sufficient statistics $\sum_{i=1}^N y_i = 6$ and $\sum_{i=1}^N y_i^2 = 14$. Therefore we wish to sample replicated data $\mathbf{z} = [z_1, z_2, z_3]$ with:

$$\begin{cases} z_1 + z_2 + z_3 = 6 \\ z_1^2 + z_2^2 + z_3^2 = 14, \end{cases} \tag{13}$$

and, after rewriting (13) into the form (6):

$$z_1^2 + z_2^2 + z_1 z_2 - 6(z_1 + z_2) = -11. \tag{14}$$

Subsequently $z_3 = 6 - (z_1 + z_2)$.

To obtain a replicated vector $\mathbf{z}^- = [z_1, z_2]$, a random draw is obtained from a bivariate normal distribution with $\boldsymbol{\mu}^- = [3, 3]$ and a 2×2 covariance matrix $\boldsymbol{\Sigma}^-$. The inverse of $\boldsymbol{\Sigma}^-$ is $\boldsymbol{\Omega}^-$ and the latter has diagonal elements equal to 1 and off-diagonal elements equal to $\frac{1}{2}$. The obtained vector is denoted $\mathbf{p}^- = [p_1, p_2]$. The set of equations to be solved then consist of:

$$\begin{cases} z_1 = 3 + \lambda(p_1 - 3) \\ z_2 = 3 + \lambda(p_2 - 3) \\ z_1^2 + z_2^2 + z_1z_2 - 6(z_1 + z_2) = -11. \end{cases} \quad (15)$$

For each sampled $\mathbf{p}^- = [p_1, p_2]$, we have 3 equations and 3 unknowns, i.e. z_1, z_2 and λ . Solving (15) for λ provides 2 solutions for $[z_1, z_2]$, from which one is randomly chosen. Subsequently, z_3 can easily be computed and one replicated vector conditional on the sufficient statistics is obtained. For instance, if the sampled values are $p_1 = 0$ and $p_2 = 1$, solving (15) gives $\lambda_1 = 0.17$ and $\lambda_2 = 0.62$, leading to $\mathbf{z}_{\lambda_1} = \{2.49, 2.66, 0.85\}$ and $\mathbf{z}_{\lambda_2} = \{1.14, 1.76, 3.10\}$ respectively.

2.2. The general univariate null hypothesis

In this section the method of obtaining the null distribution for the general univariate null hypothesis (3) will be explained. For the nuisance parameters of (3), the number of sufficient statistics is $R = K + P + 1$. The density of the data conditional on the sufficient statistics $S_1(\mathbf{y}), \dots, S_R(\mathbf{y})$ is the similar null distribution of the data. To obtain this distribution, data $\mathbf{z} = [z_1, \dots, z_N]$ are replicated from $g(\mathbf{z}|\mathbf{S}) = c$ by execution of the following steps:

1. For the general univariate null hypothesis $g(\mathbf{z}|\mathbf{S}) = c$ is specified, using:

(a) For $r = 1, \dots, K$

$$S_r(\mathbf{z}) = \sum_{i=1}^N z_i d_{ri} = \sum_{i=1}^N y_i d_{ri}. \quad (16)$$

(b) For $r = K + 1, \dots, K + P$

$$S_r(\mathbf{z}) = \sum_{i=1}^N z_i x_{(r-K),i} = \sum_{i=1}^N y_i x_{(r-K),i}. \quad (17)$$

(c) For $r = K + P + 1$

$$S_r(\mathbf{z}) = \sum_{i=1}^N z_i^2 = \sum_{i=1}^N y_i^2. \quad (18)$$

Eqs. (16)–(18) together consist of $K + P + 1$ functions of \mathbf{z} , where \mathbf{z} consists of N elements (z_i), i.e. N unknowns. Note, that the $K + P$ equations of (16) and (17) are all linear functions of \mathbf{z} . Each of these $K + P$ linear equations is solved for one specific z_i as will be explained hereafter. The observations are assumed to be ordered by group. Thus, if group $k = 1$ consists of N_1 observations, they have indices $i = 1, \dots, N_1$ and the first observation of group $k = 2$ has index $i = N_1 + 1$. To denote the number of observations in group k , the notation N_k is used. To denote the last observation of group k , the notation $z_{N_{k+}}$ is used. In other words, $N_{k+} = \sum_{k'=1}^k N_{k'}$ is the cumulative sample size of the first k groups.

2. Using the notation introduced in step 1:

(a) (16) implies that, for $k = 1, \dots, K$

$$z_{N_{k+}} = \sum_{i=1}^N y_i d_{ki} - \sum_{i \in \mathcal{A}_k} z_i d_{ki}, \quad (19)$$

with $\mathcal{A}_k = \{1, \dots, N_{k+} - 1, N_{k+} + 1, \dots, N\}$ for $k = 1, \dots, K - 1$ and $\mathcal{A}_k = \{1, \dots, N - 1\}$ for $k = K$.

(b) (17) implies that, for $p = 1, \dots, P$

$$z_{N-p} = \left(\sum_{i=1}^N y_i x_{pi} - \sum_{i \in \mathcal{B}_p} z_i x_{pi} \right) / x_{p,(N-p)}, \quad (20)$$

with $\mathcal{B}_p = \{1, \dots, N - p - 1, N - p + 1, \dots, N\}$ for $p = 2, \dots, P$ and $\mathcal{B}_p = \{1, \dots, N - 2, N\}$ for $p = 1$. Note that the observations must be ordered such that $x_{p,(N-p)} \neq 0$ to avoid division by zero.

3. Using a stepwise elimination by substitution procedure, the $K + P$ linear equations from the previous step, render solutions for $K + P$ elements of \mathbf{z} . Note, that in principle it is arbitrary for which elements z_i the equations are solved. However to avoid divisions by zero in the substitution process, the ordering of the observations in the last group, where P equations are solved for the elements z_{N-p} ($p = 1, \dots, P$), must be such that $x_{p,(N-p)} \neq x_{p,N}, \forall p$.

Substitution of the linear solutions from (19) and (20) in the quadratic form (18) always renders an equation with only quadratic terms z_i^2 , products $z_i z_l$ ($i \neq l$), single terms z_i and terms that are not a function of \mathbf{z} . The resulting equation is a function of \mathbf{z}^- , where \mathbf{z}^- has dimension $N - K - P$, since $K + P$ elements are eliminated.

It can be written as

$$\sum_{i \in \mathcal{C}} c_{ii} z_i^2 + \sum_{i \in \mathcal{C}} \sum_{l > i \in \mathcal{C}} c_{il} z_i z_l - \sum_{i \in \mathcal{C}} b_i z_i = a, \quad (21)$$

with $\mathcal{C} = \{\mathcal{A}_1 \cap \dots \cap \mathcal{A}_K \cap \mathcal{B}_1 \cap \dots \cap \mathcal{B}_P\}$. Note that (21) is the general form of (6). The constants c_{ii} , c_{il} , b_i and a are based on the observed values of the sufficient statistics, and on the observed values of x_{pi} . Furthermore, (21) is an ellipsoid in \mathbb{R}^{N^-} , with dimension $N^- = N - K - P$.

4. The equation of an isodensity ellipsoid of a multivariate normal distribution $N_{N-K-P}(\boldsymbol{\mu}^-, \boldsymbol{\Sigma}^-)$, with $(\boldsymbol{\Sigma}^-)^{-1} = \boldsymbol{\Omega}^-$ can, analogous to (8), be written as

$$\sum_{i \in \mathcal{C}} \omega_{ii} z_i^2 + \sum_{i \in \mathcal{C}} \sum_{l > i \in \mathcal{C}} 2\omega_{il} z_i z_l - \sum_{i \in \mathcal{C}} \sum_{l \in \mathcal{C}} 2\omega_{il} \mu_l z_i = d - \sum_{i \in \mathcal{C}} \omega_{ii} \mu_i^2 - \sum_{i \in \mathcal{C}} \sum_{l > i \in \mathcal{C}} 2\omega_{il} \mu_i \mu_l. \tag{22}$$

Comparing (21) with (22) it can be seen that for the general null hypothesis $\omega_{ii} = c_{ii}$, $\omega_{il} = \frac{1}{2}c_{il}$ and the values for μ_l ($l \in \mathcal{C}$) can be derived from the set of equations

$$b_i = \sum_{l \in \mathcal{C}} 2\omega_{il} \mu_l \quad \text{for } i \in \mathcal{C}. \tag{23}$$

Note that there are $N - K - P$ equations in (23), and an equal number of parameters to solve, i.e. μ_l for $l \in \mathcal{C}$.

5. Sampling $\mathbf{p}^- = [p_1, \dots, p_{N^-}]$ from $N_{N-1}(\boldsymbol{\mu}^-, \boldsymbol{\Sigma}^-)$, with $\boldsymbol{\mu}^-$ and $\boldsymbol{\Sigma}^-$ as specified in step 4, and computation of the intersections of the line through \mathbf{p}^- and $\boldsymbol{\mu}^-$ with the required ellipsoid using (12) and (21), renders a vector $\mathbf{z}^- = [z_1, \dots, z_{N^-}]$. The remaining $K + P$ observations can be computed using the $K + P$ equations contained in (19) and (20).
 6. The previous step renders one vector \mathbf{z} from $g(\mathbf{z}|\mathbf{S}) = c$. Repeating it many times renders a discrete representation of $g(\mathbf{z}|\mathbf{S}) = c$.

2.3. A specification of the general univariate null hypothesis

This illustration deals with the general null hypothesis (3), with $K = 1, P = 1$ and $d_{1i} = 1$ for $i = 1, \dots, N$, that is:

$$y_i = \beta_1 + \alpha_1 x_{1i} + \varepsilon_i, \quad \text{with } \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2). \tag{24}$$

This could, for instance, be the null hypothesis of an analysis of covariance, where, under the null, several groups are assumed to have one common covariate adjusted mean. In the sequel, in x_{1i} the subscript 1 is omitted for notational convenience.

For the $K + P + 1 = 3$ nuisance parameters β_1, α_1 and σ^2 , three sufficient statistics exist. Vectors \mathbf{z} are replicated from $g(\mathbf{z}|\mathbf{S}) = c$, using:

$$S_1(\mathbf{z}) = \sum_{i=1}^N z_i = \sum_{i=1}^N y_i, \tag{25}$$

$$S_2(\mathbf{z}) = \sum_{i=1}^N z_i x_i = \sum_{i=1}^N y_i x_i, \tag{26}$$

and

$$S_3(\mathbf{az}) = \sum_{i=1}^N z_i^2 = \sum_{i=1}^N y_i^2. \tag{27}$$

The implication of (25) is

$$z_N = \sum_{i=1}^N y_i - \sum_{i \in \mathcal{A}_1} z_i, \tag{28}$$

with $\mathcal{A}_1 = \{1, \dots, N - 1\}$.

The implication of (26) is

$$z_{N-1} = \left(\sum_{i=1}^N y_i x_i - \sum_{i \in \mathcal{B}_1} z_i x_i \right) / x_{N-1}, \tag{29}$$

with $\mathcal{B}_1 = \{1, \dots, N - 2, N\}$. Note again, that the observations must be ordered such that $x_{N-1} \neq 0$, to avoid division by zero. Solving (28) and (29) for z_N and z_{N-1} , and substituting the solution in (27), renders an ellipsoid in \mathbb{R}^{N-2} :

$$\begin{aligned} & \sum_{i \in \mathcal{C}} (1 + h_i^2 - h_i) z_i^2 + \sum_{i \in \mathcal{C}} \sum_{l > i \in \mathcal{C}} (1 + 2h_i h_l - h_i - h_l) z_i z_l \\ & + \sum_{i \in \mathcal{C}} \left(\sum_{i'=1}^N y_{i'} - m_1 + m_2 + \left(2m_1 - 2m_2 - \sum_{i'=1}^N y_{i'} \right) h_i \right) z_i \\ & = \frac{1}{2} \left(\sum_{i=1}^N y_i^2 - \left(\sum_{i=1}^N y_i \right)^2 \right) - (m_1 - m_2)^2 + (m_1 - m_2) \sum_{i=1}^N y_i, \end{aligned} \tag{30}$$

with $h_i = \frac{x_i - x_N}{x_{N-1} - x_N}$, $h_l = \frac{x_l - x_N}{x_{N-1} - x_N}$, $m_1 = \frac{\sum_{i=1}^N x_i y_i}{x_{N-1} - x_N}$, $m_2 = \frac{x_N \sum_{i=1}^N y_i}{x_{N-1} - x_N}$ and $\mathcal{C} = \{\mathcal{A}_1 \cap \mathcal{B}_1\}$. Note, that (30) is a specification of (21) with the constants c_{ii} , c_{il} , b_i and a explicitly defined. For this example, it can be seen that the subjects must also be ordered such that $x_{N-1} \neq x_N$, because otherwise the coefficients h_i , h_l , m_1 and m_2 in (30) contain divisions through zero (see also step 3 in Section 2.2).

The ellipsoid described by (30) is equal to one of the isodensity ellipsoids from the multivariate normal $N_{N-2}(\boldsymbol{\mu}^-, \boldsymbol{\Sigma}^-)$ with $(\boldsymbol{\Sigma}^-)^{-1} = \boldsymbol{\Omega}^-$, with elements $\omega_{ii} = 1 + h_i^2 - h_i$, for $i \in \mathcal{C}$ and $\omega_{il} = \frac{1}{2}(1 + 2h_i h_l - h_i - h_l)$, for $i \neq l$; $i, l \in \mathcal{C}$. The values for μ_l ($l \in \mathcal{C}$) can be derived from the set of equations

$$\sum_{i'=1}^N y_{i'} - m_1 + m_2 + \left(2m_1 - 2m_2 - \sum_{i'=1}^N y_{i'}\right) h_i = \sum_{l \in \mathcal{C}} 2\omega_{il} \mu_l, \quad \text{for } i \in \mathcal{C}. \quad (31)$$

Similarly to (23), the number of equations in (31) is equal to the number of parameters to solve, i.e. μ_l for $l \in \mathcal{C}$. Again, sampling from this multivariate normal, deriving \mathbf{z}^- on the required ellipsoid, and computation of z_N and z_{N-1} using (28) and (29), renders a vector \mathbf{z} . Repetition of the procedure renders the similar distribution of the data.

To test (24) against any alternative an appropriate test statistic must be defined. Traditionally, defining a (new) test statistic implies the (often burdensome) task of deriving its null distribution. However, with the approach presented here, the similar null distribution of the data is rendered and therefore the similar distribution of any test statistic is made available. This enables, for instance, testing against an ordered alternative hypothesis, or, testing with robust test-statistics in the context of analysis of covariance.

3. The multivariate null hypothesis

3.1. Replication of data under the general multivariate null hypothesis

To obtain the null distribution of the data for the general multivariate null hypothesis (4), data $\mathbf{Z}' = [\mathbf{z}_1 \dots, \mathbf{z}_Q]$ is replicated from

$$g(\mathbf{Z} | S_1(\mathbf{Z}) = S_1(\mathbf{Y}), \dots, S_R(\mathbf{Z}) = S_R(\mathbf{Y})) = g(\mathbf{Z} | \mathbf{S}) = c, \quad (32)$$

where $\mathbf{Y}' = [\mathbf{y}_1 \dots, \mathbf{y}_Q]$ denotes the observed data, and R the total number of sufficient statistics. The sufficient statistics for (32) are:

1. For $q = 1, \dots, Q$ and $k = 1, \dots, K$ (i.e. Q times K sufficient statistics)

$$\sum_{i=1}^N z_{qi} d_{ki} = \sum_{i=1}^N y_{qi} d_{ki}. \quad (33)$$

2. For $q = 1, \dots, Q$ and $p = 1, \dots, P$ (i.e. Q times P sufficient statistics)

$$\sum_{i=1}^N z_{qi} x_{pi} = \sum_{i=1}^N y_{qi} x_{pi}. \quad (34)$$

3. For $q = 1, \dots, Q$ (i.e. Q sufficient statistics)

$$\sum_{i=1}^N z_{qi}^2 = \sum_{i=1}^N y_{qi}^2. \quad (35)$$

4. For $q = 1, \dots, Q - 1$ and $q' = q + 1, \dots, Q$ (i.e. $\frac{1}{2}(Q^2 - Q)$ sufficient statistics)

$$\sum_{i=1}^N z_{qi} z_{q'i} = \sum_{i=1}^N y_{qi} y_{q'i}. \quad (36)$$

Note that, $R = QK + QP + Q + \frac{1}{2}(Q^2 - Q)$, which is equal to Q times the number of sufficient statistics for the general univariate null hypothesis ($K + P + 1$) plus the $\frac{1}{2}(Q^2 - Q)$ sufficient statistics in (36), that concern the relations between the Q dependent variables. The presence of the equations in (36) and the fact that (35) consists of Q quadratic equations makes the sampling of replicated data \mathbf{Z} different from the univariate case presented in the previous section.

The sampling can, however, be partitioned in Q parts. This is done through the application of a Gibbs sampling approach. In the Gibbs sampler, sampling from multivariate distributions is simplified by iteratively sampling each parameter from its univariate distribution, conditional on the other parameters (Tierney, 1994; Smith and Roberts, 1993). Likewise, to sample data \mathbf{Z} from (32) an iterative procedure consisting of the following Q steps is applied:

1. Sample \mathbf{z}_1 from $g(\mathbf{z}_1 | \mathbf{z}_2, \dots, \mathbf{z}_Q, \mathbf{S}^{(1)}) = c$,

2. Sample \mathbf{z}_2 from $g(\mathbf{z}_2 | \mathbf{z}_1, \mathbf{z}_3, \dots, \mathbf{z}_Q, \mathbf{S}^{(2)}) = c$,

⋮

Q. Sample \mathbf{z}_Q from $g(\mathbf{z}_Q | \mathbf{z}_1, \dots, \mathbf{z}_{Q-1}, \mathbf{S}^{(Q)}) = c$,

where $\mathbf{S}^{(q)}$ ($q = 1, \dots, Q$) refers to those sufficient statistics that contain \mathbf{z}_q . Examining, for instance, $\mathbf{S}^{(1)}$ shows that K sufficient statistics of the form (33), P of the form (34), one of the form (35) and finally $Q - 1$ of a form that is similar to (34) remain. This holds for each $\mathbf{S}^{(q)}$ ($q = 1, \dots, Q$). Therefore, within each step the ‘active’ sufficient statistics in the conditioning, all fit in the general univariate framework. The sampling of a vector \mathbf{z}_q in each step is therefore equal to the sampling procedure described in Section 2.2.

The Gibbs sampler requires starting values for $\{\mathbf{z}_2, \dots, \mathbf{z}_Q\}$ in order to sample \mathbf{z}_1 in the first iteration. A simple choice is to use the observed scores $\{\mathbf{y}_2, \dots, \mathbf{y}_Q\}$. Note that any set of starting values that satisfies $\{\mathbf{S}^{(2)}, \dots, \mathbf{S}^{(Q)}\}$ is equally probable because the target distribution (32) is uniform. If the total number of iterations is large enough, the resulting p -value will not be affected by the initial values. Careful monitoring of the so-called convergence is however important (cf. Cowles and Carlin (1996)). In each of the two illustrations that follow in Sections 3.2 and 3.3, the Gibbs sampler is applied twice with different starting values. It will be seen that, with reasonable numbers of iterations, no effect of the initial values is seen, and stable p_{sim} -values are obtained.

3.2. Multivariate one-sided testing

The data ($N = 9$) used for this illustration are part of a larger research project on the efficacy of cognitive behavioral therapy for traumatic grief (Boelen et al., 2007). Before and after six cognitive therapy sessions the four criterion variables *grief* (Texas Revised Inventory of Grief), *traumatic grief* (Inventory of Traumatic Grief), *anxiety* (subscale of SCL-90) and *depression* (subscale of SCL-90) are measured. For each criterion variable, the decrease in symptoms is measured by the difference between the scores before and after the treatment. The research question investigated in this illustration is whether there is a decrease in the grief symptoms.

The null hypothesis is (4) with $Q = 4, K = 0, P = 0$, i.e

$$H_0 : \boldsymbol{\beta} = 0, \quad \boldsymbol{\varepsilon}_i \stackrel{iid}{\sim} N(0, \boldsymbol{\Sigma}).$$

It is tested against the alternative $\mathbf{y}_i = \boldsymbol{\beta}_1^* + \boldsymbol{\varepsilon}_i$ with the restriction that the mean vector must be positive, i.e.

$$H_1 : \boldsymbol{\beta}_1^* \geq 0.$$

The $*$ is used to explicitly denote that a parameter under the alternative hypothesis is considered.

Although we found some references in the literature with respect to multivariate one-sided testing, for most of the test statistics discussed, no finite sample distribution is available (Silvapulle, 1995b; Tang, 1994; Perlman and Wu, 2002). An exception is Follmann’s test (Follmann, 1996), although according to Chongcharoen et al. (2002) this test is not very powerful. There are other references that do present tests with their finite sample distribution, but for these tests it is assumed that the population covariance matrix is known, or known up to a multiplicative constant (Chongcharoen et al., 2002; Silvapulle, 1995a). Our method provides the similar distribution for any statistic that can be applied to test $\boldsymbol{\beta} = 0$ against $\boldsymbol{\beta}_1^* \geq 0$, when the covariance matrix is unknown.

The nuisance parameters under the null hypothesis are the 10 upper diagonal elements of the 4×4 covariance matrix $\boldsymbol{\Sigma}$. The sufficient statistics for these nuisance parameters are:

1. For $q = 1, \dots, 4$

$$\sum_{i=1}^N z_{qi}^2 = \sum_{i=1}^N y_{qi}^2.$$

2. For $q = 1, \dots, 3$ and $q' = q + 1, \dots, 4$

$$\sum_{i=1}^N z_{qi} z_{q'i} = \sum_{i=1}^N y_{qi} y_{q'i}.$$

The Gibbs sampler applied to sample from (32) consists of four steps. In each step, one vector \mathbf{z}_q is sampled, conditional on the current values of the other vectors and $\mathbf{S}^{(q)}$, i.e. the sufficient statistics that contain \mathbf{z}_q . Note that in each step, $\mathbf{S}^{(q)}$ consists of three linear and one quadratic function of \mathbf{z}_q , and this implies that in each step the general univariate approach as introduced in Section 2.2 can be applied, that is, $g(\mathbf{z} | \mathbf{S})$ with $K = 0$ and $P = 3$ (the number of dependent variables that are conditioned upon in each step, i.e., $Q - 1$).

To monitor convergence, several checks are performed. First of all, for each person and each outcome variable, a plot of all sampled z_{qi} values is visually inspected. As an illustration, in Fig. 2 the results of 1000 iterations are plotted for the first and ninth respondent, for the first (the two plots on top) and fourth (the two plots at bottom) outcome variables. Since the data are assumed to be i.i.d., the plots of z_{qi} values for a specific q but different respondents i are expected to be similar. Based on visual inspection of the plots, this is indeed the case. Also examination of summary measures per z_{qi} (e.g. the mean, standard deviation, minimum and maximum sampled value) shows virtually equal results for the nine respondents.

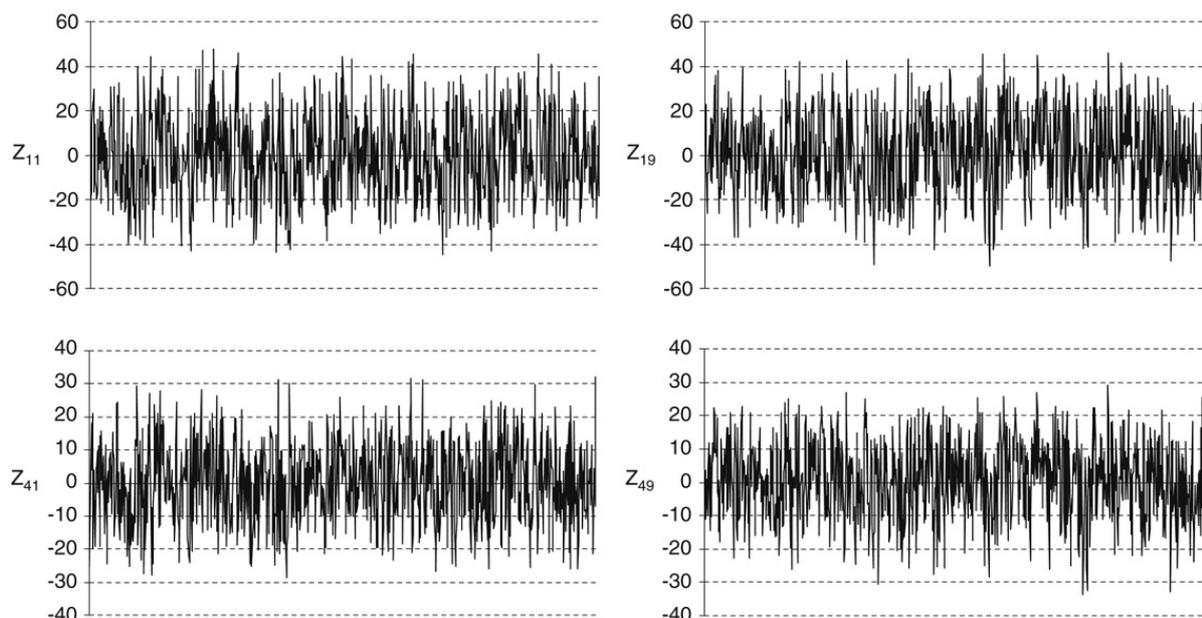


Fig. 2. Replicated data for the first and the ninth person on the first (top) and the fourth (bottom) dependent variable.

Another convergence check is the computation of the so-called \hat{R} measure (Gelman and Rubin, 1992). Loosely stated, with this measure, the variability within a chain of sampled values (or a part of a chain) is compared with the variability between chains (or different parts of a chain). A value of \hat{R} that is close to one is indicative of convergence and a rule of thumb states that values below 1.1 can be considered satisfactory (Gelman and Rubin, 1992). For the example data at hand, for each z_{qi} a chain of 50,000 iterations is partitioned in 5 parts of 10,000 and the value for \hat{R} is computed. The results confirm good convergence because, with a precision of two decimals, the resulting \hat{R} values are all equal to 1.00.

Now that we have replicated data \mathbf{Z} , any test statistic can be chosen to test the hypothesis of interest. In this illustration the test statistic used is:

$$T(\mathbf{Y}) = \sum_{q=1}^Q \frac{\bar{y}_q}{sd(y_q)}, \tag{37}$$

where \bar{y}_q denotes the sample mean and $sd(y_q)$ the sample standard deviation of the q -th dependent variable. The larger the value of $T(\mathbf{Y})$ the more the sample means are located in the positive orthant and, as such, the larger the evidence in favor of the alternative hypothesis. For each replicated data matrix \mathbf{Z} , $T(\mathbf{Z}) = \sum_{q=1}^Q \frac{\bar{z}_q}{sd(z_q)}$ is computed. The p_{sim} -value is obtained by (2).

Note that there are superior tests for the hypotheses at hand (e.g. Follmann (1996)). However, the aim of this illustration is not to find the best test statistic, but to demonstrate that our method obtains the similar distribution of any test statistic. The simplicity of (37) makes it attractive for this illustrative purpose, but our procedure can be applied to any statistic of interest.

Applying our approach to obtain the similar null distribution and the test statistic in (37) to the traumatic grief data, results in a significance $p = .0016$. It can be concluded that after six cognitive therapy sessions, the amount of grief has decreased.

To perform one more convergence check, two chains of iterations with different starting values are sampled. In Fig. 3, the p_{sim} -value is plotted against the iteration number. Note that these are cumulative plots. Therefore, at iteration 10,000 p_{sim} is computed based on the first 10,000 iterations, whereas p_{sim} at iteration 50,000 is computed based on the first 50,000 iterations. It can be seen that both chains converge to the same p_{sim} -value. Stable results up to the fourth decimal are obtained after approximately 250,000 iterations. Both chains were extended until a total length of 5 million iterations (not plotted) and the resulting p_{sim} -values both remained .0016.

3.3. Multivariate ordered means

The data used for this illustration comes from a large study concerning the effect of Sesame Street on six cognitive ability measures for children in the age range between three to five (Stevens, 1996, pp. 578–585). The cognitive abilities (e.g. knowledge of numbers and letters) were measured before Sesame Street was broadcast (pretests), and after the first year of the program (post-tests). Gain scores are obtained by taking the difference between the post-test and pretest scores. Furthermore, two groups are distinguished, based on whether children watched the show frequently (group 1), or, rarely

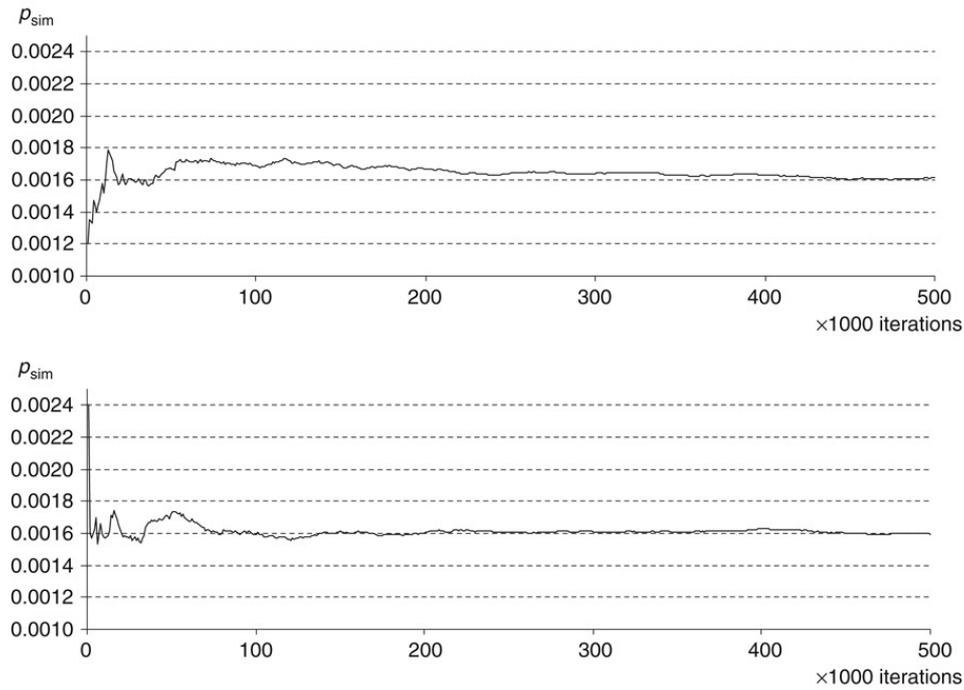


Fig. 3. Similar p -value after up to 500,000 iterations using two different starting values.

or never (group 2). The studied sample contains children from different social and environmental backgrounds, but in this illustration only one group is included: the disadvantaged rural children ($N = 43$). The research question of interest is whether watching the show regularly, compared to watching rarely or never, leads to a higher gain in the six measures of cognitive ability.

This leads to testing the null hypothesis

$$\mathbf{y}_i = \beta_1 d_{1i} + \boldsymbol{\varepsilon}_i, \quad \text{with } \boldsymbol{\varepsilon}_i \stackrel{iid}{\sim} N(0, \boldsymbol{\Sigma}), \tag{38}$$

i.e. (4) with $K = 1$, $d_{1i} = 1$ for $i = 1, \dots, N$, $P = 0$, and $Q = 6$, against the alternative

$$\mathbf{y}_i = \beta_1^* d_{1i} + \beta_2^* d_{2i} + \boldsymbol{\varepsilon}_i, \quad \text{with } \beta_1^* \geq \beta_2^*, \tag{39}$$

where the latter is interpreted coordinate-wise.

We found only one reference with respect to testing this kind of hypotheses (Sasabuchi et al., 2003). They present a method to compute the supremum of the upper tail probability under the null hypothesis for a likelihood ratio test. With our approach, the similar distribution of any test statistic and therefore level alpha tests are available.

Under the null hypothesis there are 27 nuisance parameters, i.e. six elements of β_1 and 21 elements of $\boldsymbol{\Sigma}$. The sufficient statistics for the nuisance parameters, are:

1. For $q = 1, \dots, 6$

$$\sum_{i=1}^N z_{qi} = \sum_{i=1}^N y_{qi}.$$

2. For $q = 1, \dots, 6$

$$\sum_{i=1}^N z_{qi}^2 = \sum_{i=1}^N y_{qi}^2.$$

3. For $q = 1, \dots, 5$ and $q' = q + 1, \dots, 6$

$$\sum_{i=1}^N z_{qi} z_{q'i} = \sum_{i=1}^N y_{qi} y_{q'i}.$$

The Gibbs sampler that is applied to sample from (32) consists of six steps. Again, each step of the Gibbs sampler involves the application of the general univariate approach introduced in Section 2.2, i.e., $g(\mathbf{z}|\mathbf{S})$ with $K = 1$ and $P = 5$.

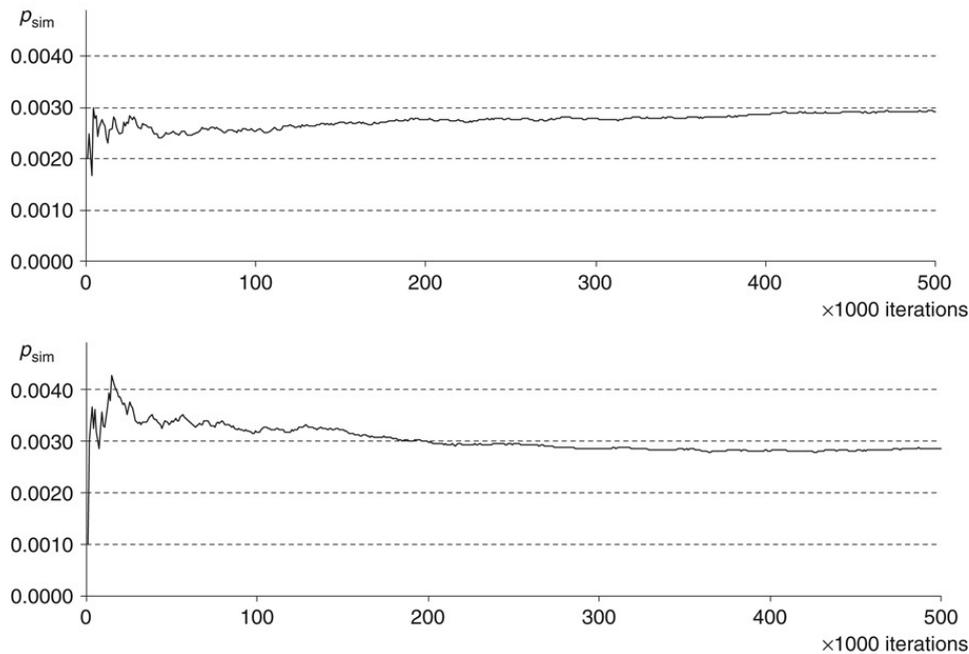


Fig. 4. Similar p -value after up to 500,000 iterations using two different starting values.

The test statistic used in this illustration is

$$T(\mathbf{Y}) = \sum_{q=1}^Q \frac{\bar{y}_{q1} - \bar{y}_{q2}}{sd_p(y_q)}, \tag{40}$$

where \bar{y}_{q1} (respectively \bar{y}_{q2}) denotes the sample mean of the q -th dependent variable of group 1 (respectively group 2), and $sd_p(y_q)$ denotes the pooled sample standard deviation of the two groups of the q -th dependent variable. The larger the value of this test statistic, the more the pairwise differences between the sample means are located in the positive orthant and, as such, the larger the evidence in favor of the alternative hypothesis. For each replicated data matrix \mathbf{Z} , $T(\mathbf{Z}) = \sum_{q=1}^Q \frac{\bar{z}_{q1} - \bar{z}_{q2}}{sd_p(z_q)}$ is computed. The p_{sim} -value is obtained by (2).

Testing (38) against (39) using (40) renders $p = .003$. It can be concluded that the disadvantaged rural children that watched Sesame Street regularly have learned more than the disadvantaged rural children that did not watch Sesame Street.

Convergence is monitored in a similar way as in the previous illustration and was again confirmed. Here we present just the plotted p_{sim} -values for two chains of 500,000 iterations with different starting values. In Fig. 4, it can be seen that both chains converge to virtually the same p_{sim} -value. Stable results up to the third decimal are obtained after approximately 200,000 iterations. Both chains were extended until a total length of 5 million iterations (not plotted) and the resulting p_{sim} -values both remained .003 (respectively .0028 and .0027).

Also, in this last example, a similar test is obtained for a multivariate, inequality constrained hypothesis. We did not find any references in the literature to such a test. Note that we used a relatively ad hoc test statistic, but the focus of this paper is not on the test statistic but on obtaining the similar null distribution of the data, which enables similar (level alpha) tests for any test statistic.

4. Discussion

This paper showed how to obtain similar null distributions of data, if the null hypothesis is a member of the univariate or multivariate normal model. We demonstrated that, with this distribution, similar p -values can be obtained for any test statistic of interest and, as such, level alpha tests are made available for a wide variety of hypotheses in the context of normal linear models.

The approach is based on repeatedly sampling replicated data from the null population, conditional on the observed values for the sufficient statistics. Computationally, the procedure may seem rather demanding but, for the examples presented, computation time was more than reasonable; we obtained stable p -values up to two decimals within minutes on an average modern pc for all examples and much faster for the smaller examples (i.e., univariate analyses and multivariate cases with a small sample size like the illustration in Section 3.2). The largest analysis was presented in Section 3.3, with a sample size $N = 43$ and number of outcome variables $Q = 6$. Again, results were obtained within a reasonable amount of time and investigation of convergence and stability of the resulting p_{sim} -value show satisfactory results. For still larger designs, such as larger values for Q , investigation would need to be done to find out how well the sampling procedure performs, but theoretically there are no limitations.

The examples given in this paper concerned null hypotheses with order restricted alternatives. Testing order constrained hypotheses has received considerable attention in the last few decades (cf. Barlow et al. (1972), Robertson et al. (1988) and Silvapulle and Sen (2005)) and is still ongoing (see, for instance, Singh and Liu (2006) and Tian et al. (2008)). For specifically *multivariate* applications, however, until now there have been hardly any exact or level alpha tests for constrained alternatives. Researchers are limited in the choice of a (powerful) test statistic since it is difficult to derive the null distribution of most test statistics. With the approach presented in this paper, the null distribution for the (multivariate) data is obtained and therefore leads to similar p -values for any test statistic of interest.

However, order constrained inference is not the only domain in which our method can be useful. A few more possibilities are mentioned. Consider the following example that introduces a test for potential outliers. In this example, the fit of the null model is checked using order statistics (cf. Mood et al. (1974), pp. 251–265). Let the data come from a normal distribution with mean β and variance σ^2 , i.e. the univariate null hypothesis (3) with $P = 0$ and $K = 1$. A test statistic T can be formulated that, for instance, determines whether the largest outlier is significant. For example, T can be the largest squared deviation between each observation and the sample average. The proportion of T for the replicated data matrices that have larger values than T in the observed data provides a p -value for the null hypothesis that the observed T is not an outlier.

The approach can also be used for model checking. As an example, consider the model assumption of equal variances in an analysis of variance with J groups. Under the null hypothesis, the J groups have one common variance σ^2 , i.e., the null hypothesis is (3) with $P = 0$ and $K = J$. Under the alternative, the groups have a group specific variance term. It is well-known that the classical tests for homogeneity of variances, e.g. F -test, Bartlett's test (Bartlett and Kendall, 1946), are highly sensitive to the assumption that population distributions are normal. Therefore, more robust test statistics have been developed, e.g. the Levene/Brown–Forsythe test statistic (Brown and Forsythe, 1974), which is approximately F -distributed. Recently Shoemaker (2003) proposed two other statistics for testing the homogeneity for two or more variances, that further improve the robustness properties. However, for these test statistics, as for the Levene/Brown–Forsythe test, only the asymptotic distribution is known. Since our approach renders the null distribution of the data, the similar distribution for any of these test statistics is obtained.

References

- Barlow, R.E., Bartholomew, D.J., Bremner, J.M., Brunk, H.D., 1972. *Statistical Inference Under Order Restrictions*. John Wiley & Sons, New York.
- Bartlett, M.S., Kendall, D.G., 1946. The statistical analysis of variance-heterogeneity and the logarithmic transformation. *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* 1 (suppl), 128–138.
- Bayarri, M.J., Berger, J.O., 2000. P values for composite null models. *J. Amer. Statist. Assoc.* 95, 1127–1142.
- Bayarri, M.J., Castellanos, M.E., Morales, J., 2006. MCMC methods to approximate conditional predictive distributions. *Comput. Statist. Data Anal.* 51, 621–640.
- Boelen, P.A., de Keijser, J., van den Hout, M.A., van den Bout, J., 2007. Treatment of complicated grief: A comparison between cognitive-behavioral therapy and supportive counseling. *J. Consulting and Clinical Psychology* 75, 277–284.
- Brown, M.B., Forsythe, A.B., 1974. Robust tests for the equality of variances. *J. Amer. Statist. Assoc.* 69, 364–367.
- Chongcharoen, S., Singh, B., Wright, F.T., 2002. Powers of some one-sided multivariate tests with the population covariance matrix known up to a multiplicative constant. *J. Statist. Plann. Inference* 107, 103–121.
- Cowles, M.K., Carlin, B.P., 1996. Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Amer. Statist. Assoc.* 91, 883–904.
- Follmann, D., 1996. A simple multivariate test for one-sided alternatives. *J. Amer. Statist. Assoc.* 91, 854–861.
- Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statist. Sci.* 7, 457–511.
- Lehmann, E.L., 1997. *Testing Statistical Hypotheses*, second edition. Springer-Verlag, New York.
- Mood, A.M., Graybill, F.A., Boes, D.C., 1974. *Introduction to the Theory of Statistics*, third edition. McGraw-Hill, Singapore.
- Muller, K.E., 1998. A new F -approximation for the Pillai–Bartlett trace under H_0 . *J. Comput. Graph. Statist.* 7, 131–138.
- Perlman, M.D., Wu, L., 2002. A class of conditional tests for a multivariate one-sided alternative. *J. Statist. Plann. Inference* 107, 155–171.
- Robertson, T., Wright, F.T., Dykstra, R.L., 1988. *Order Restricted Statistical Inference*. John Wiley & Sons, New York.
- Robins, J.M., van der Vaart, A., Ventura, V., 2000. The asymptotic distribution of p values in composite null models. *J. Amer. Statist. Assoc.* 95, 1143–1156.
- Sasabuchi, S., Tanaka, K., Tsukamoto, T., 2003. Testing homogeneity of multivariate normal mean vectors under an order restriction when the covariance matrices are common but unknown. *Ann. Statist.* 31, 1517–1536.
- Shoemaker, L.H., 2003. Fixing the F test for equal variances. *Amer. Statist.* 57, 105–114.
- Silvapulle, M.J., 1992a. Robust tests of inequality constraints and one-sided hypotheses in the linear model. *Biometrika* 79, 621–630.
- Silvapulle, M.J., 1992b. Robust Wald-type tests of one-sided hypotheses in the linear model. *J. Amer. Statist. Assoc.* 87, 156–161.
- Silvapulle, M.J., 1995a. On an F -type statistic for testing one-sided hypotheses and computation of chi-bar-squared weights. *Statist. Probab. Lett.* 28, 137–141.
- Silvapulle, M.J., 1995b. A Hotelling's T^2 -type statistic for testing against one-sided hypotheses. *J. Multivariate Statist.* 55, 312–319.
- Silvapulle, M.J., Sen, P.K., 2005. *Constrained Statistical Inference. Inequality, Order and Shape Restrictions*. John Wiley & Sons, New Jersey.
- Singh, P., Liu, W., 2006. A test against an umbrella ordered alternative. *Comput. Statist. Data Anal.* 3, 1957–1964.
- Smith, A.F.M., Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* 55, 3–23.
- Stevens, J., 1996. *Applied Multivariate Statistics for the Social Sciences*, third edition. Lawrence Erlbaum, Mahwah, New Jersey.
- Tang, D.I., 1994. Uniformly more powerful tests in a one-sided multivariate problem. *J. Amer. Statist. Assoc.* 89, 1006–1011.
- Tatsuoka, M.M., 1988. *Multivariate Analysis. Techniques for educational and psychological research*. Macmillan Publishing Company, New York.
- Tian, G.L., Wang Ng, K., Tan, M., 2008. EM-type algorithms for computing restricted MLEs in multivariate normal distributions and multivariate t -distributions. *Comput. Statist. Data Anal.* 52, 4768–4778.
- Tierney, L., 1994. Markov chains for exploring posterior distributions. *Ann. Statist.* 22, 1701–1762.