

CONF-9508165-3

UCRL-JC-121819
PREPRINT

Using the Bootstrap in a Multivariate Data Problem:
An Example

RECEIVED
OCT 13 1995
OSTI

This paper was prepared for submittal to the
American Statistical Association Annual Meeting
Orlando, Florida
August 13-17, 1995

August 1, 1995

This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint is made available with the understanding that it will not be cited or reproduced without the permission of the author.

 Lawrence
Livermore
National
Laboratory

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED *JK*

DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

USING THE BOOTSTRAP IN A MULTIVARIATE DATA PROBLEM: AN EXAMPLE

Jeffrey G. Glosup, Michael C. Axelrod
Jeffrey G. Glosup, Lawrence Livermore National Laboratory, PO Box 808
Mail Stop L-195, Livermore, CA 94551

KEY WORDS: Resampling, Hypothesis Testing, Paired Comparison

ABSTRACT: The use of the bootstrap in the multivariate version of the paired *t*-test is considered and demonstrated through an example. The problem of interest involves comparing two different techniques for measuring the chemical constituents of an sample item. The bootstrap is used to form an empirical significance level for Hotelling's one-sample *T*-squared statistic. The bootstrap was selected to determine empirical significance levels because the implicit assumption of multivariate normality in the classic Hotelling's one-sample test might not hold. The results of both the classic and bootstrap test are presented and contrasted.

1. Introduction

In several recent problems we have encountered, there has been a need to compare measurements of the chemical isotopic composition of a group of items resulting from two different measurement techniques. In each case, the random variables representing the measurements are an example of multivariate compositional data since there are two or more isotopes of interest for each item and the isotopic fractions must sum to one. Aitchison (1986) discusses the analysis of multivariate compositional data and suggests that some appropriate transformation of the data can be considered to have a multivariate normal distribution and that the usual array of multivariate data analysis techniques can be used on the transformed data. However, little information is available on which transformation of the data is appropriate and there is no single best transformation for all situations.

Let some unspecified transformation of the measurement of the isotopic composition by a particular technique be represented by the random variable

$$\mathbf{X} = \mathbf{v} + \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1)$$

where \mathbf{X} is a p -dimensional vector representing the measured value, \mathbf{v} is a p -dimensional vector representing the true value of the variable, $\boldsymbol{\beta}$ is a p -dimensional vector representing the bias of the measurement technique, and $\boldsymbol{\varepsilon}$ is a p -dimensional vector of the random measurement errors. In general, it is assumed that the random error $\boldsymbol{\varepsilon}$ has zero mean and

some unknown $p \times p$ covariance matrix Σ . A particular measurement technique (labeled 1 or 2) will be designated by a superscript contained in parenthesis; that is $\mathbf{X}^{(2)}$ represents a measurement made using technique 2.

A variety of comparisons of the two measurement techniques are possible, but the one of interest here is to determine whether the two techniques are relatively unbiased. That is, we assume that the same item is measured by each of two techniques so that the true values \mathbf{v} for the two techniques are identical. This problem naturally leads to the use of the paired differences in inferences about the relative bias between the two measurement techniques. The difference vector \mathbf{Y} is defined so that

$$\mathbf{Y} = \mathbf{X}^{(2)} - \mathbf{X}^{(1)} = (\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}) + (\boldsymbol{\varepsilon}^{(2)} - \boldsymbol{\varepsilon}^{(1)}), \quad (2)$$

and the question of interest is whether all components of the relative bias vector $(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})$ are identically equal to zero. This problem can be treated by classical methods of multivariate analysis, and the relevant technique is reviewed in Section 2. Bootstrap techniques for testing this hypothesis can also be developed, and two such nonparametric bootstrap approaches are developed in Section 3 below. Both the classical and bootstrap techniques are demonstrated through an example in Section 4.

2. Review of the Classical Multivariate Approach to the Problem

The hypothesis test of interest is that the expected value of the difference vector is the zero vector for the case where the variance matrix for the difference vector \mathbf{Y} is unknown. Under the assumption that the difference vector has a p -dimensional multivariate normal distribution, the likelihood ratio test of this hypothesis leads to Hotelling's one-sample T^2 test, see Mardia (1979). For the most general case of this test, let \mathbf{Y} be a p -dimensional random variable having a multivariate normal distribution with unknown p -dimensional mean vector $\boldsymbol{\mu}$ and unknown $p \times p$ dimensional variance-covariance matrix Σ . Assume that a collection of n independent realizations of the random vector \mathbf{Y} have been observed and collected into an $n \times p$ observed value array Ψ and that the sample mean vector for the n realizations is the p -dimensional vector $\bar{\mathbf{Y}}$. Hotelling's one-sample T^2 test for the null hypothesis that the true mean is $\boldsymbol{\mu}_0$ is based on the test statistic

$$T^2 = \frac{n-p}{p} (\bar{\mathbf{Y}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\mathbf{Y}} - \boldsymbol{\mu}_0) \quad (3)$$

where S is the estimated variance-covariance matrix for \mathbf{Y} based on the n realizations of \mathbf{Y} , again see Mardia (1979). Note that for this definition of the T^2 test statistic, S is defined using n , not $(n-1)$ in the denominator. Under the null hypothesis (which includes the assumption of multivariate normality of \mathbf{Y}), this test statistic has Fisher's F distribution with p degrees of freedom in the numerator and $(n-p)$ degrees of freedom in the denominator. This distribution can be used to determine a significance level for an observed value of T^2 .

The assumption that the vector \mathbf{Y} has a multivariate normal distribution plays a key role in the demonstration that the T^2 statistic has an F distribution. Everitt (1979) performed a study of the robustness of the Hotelling's T^2 test to some departures from normality. However, the alternative distributions used in this study all relied on independence of the p components. It is uncertain what effects departures from normality have on the significance level derived from the F distribution for a more general case. In light of the uncertainty about the robustness of the F distribution, we explored bootstrap methods for determining the achieved significance level for Hotelling's T^2 statistic.

3. Bootstrap Approaches to Determining Significance Levels

The bootstrap provides a method for determining the significance level of a test statistic like Hotelling's T^2 statistic without relying on possibly unfounded assumptions about the statistical distribution of the underlying random variable. The bootstrap is based on estimating the distribution function of the test statistic by resampling from the empirical distribution function of the observed data; that is, by repeatedly drawing samples with replacement from the observed data itself. Details on the background and use of the bootstrap are contained in Efron (1982) and in Efron (1993) while a discussion of the use of the bootstrap for direct significance tests is contained in Hinkley (1988).

Two nonparametric versions of the bootstrap have been investigated here as alternative means of determining a significance level for Hotelling's one sample T^2 test. For both cases, the most general notation for Hotelling's multivariate one-sample test as defined in the previous section has been used. The first version is a naive bootstrap which relies on simply resampling with replacement from the rows of the observation array Ψ for the random variable \mathbf{Y} . Note that this implies that a single row can appear more than one time in the resampled array. The steps of the naive bootstrap algorithm are listed in Algorithm 1 below.

The second version is a bootstrap with scaling which is used to compensate for the variance-covariance structure of a particular data set since the bootstrap is known to be sensitive to the measurement scale used; see Chapter 12 of Efron (1993) for examples. In the univariate bootstrap, this type of scaling is usually accomplished by dividing the observed values of the random variable by the square root of the variance of the random variable (or an estimate of this quantity) before resampling and then re-scaling the resampled values. A similar technique can be used with multivariate data since the variance-covariance matrix is positive-definite and thus has a unique inverse and a unique square root of that inverse, see Mardia (1979). The square root of the inverse is defined by the Cholesky decomposition of the inverse which provides the factorization

$$\Sigma^{-1} = \mathbf{U}^T \mathbf{U} \quad (4)$$

where \mathbf{U} is an upper triangular matrix; see Bellman (1960) or Thisted (1988) for details. The steps required to perform the scaled bootstrap are provided in Algorithm 2 below. In this version, the resampling is performed not on the rows of the array Ψ but on the rows of the transformed array $\Psi(\Sigma^{-1/2})^T$.

Algorithm 1
The Naive Bootstrap Algorithm for the
Significance of Hotelling's One-Sample Test

1. Using the sample mean vector $\bar{\mathbf{Y}}$ for the n realizations of the random vector \mathbf{Y} and the hypothesized mean μ_0 , calculate T^2 using equation (3).
2. Select B independent bootstrap sample arrays (of dimension $n \times p$ each) $\Psi_1^*, \Psi_2^*, \dots, \Psi_B^*$ by sampling with replacement from the rows of the data realization array Ψ .
3. For each bootstrap sample, calculate the bootstrap statistic T_i^{*2} for $i = 1, \dots, B$ by using $\bar{\mathbf{Y}}_i^*$ in place of $\bar{\mathbf{Y}}$, and $\bar{\mathbf{Y}}$ in place of μ_0 in equation (3)
4. The bootstrap significance level p is¹ $\#\left[T_i^{*2} > T^2\right]/B$.

¹The function $\#[.]$ represents the number of times the argument is true for $i = 1, \dots, n$.

Algorithm 2
The Scaled Bootstrap Algorithm for the
Significance of Hotelling's One-Sample Test

1. Using the sample mean vector \bar{Y} for the n realizations of the random vector Y and the hypothesized mean μ_0 , calculate T^2 using equation (3).
2. Using the sample covariance matrix S for the n realizations of the random vector Y , calculate its square root $S^{1/2}$ and the square root of the inverse $S^{-1/2}$.
3. Form the scaled realization array $K = \Psi(S^{-1/2})^T$.
4. Select B independent scaled bootstrap sample arrays (of dimension $n \times p$ each) $K_1^*, K_2^*, \dots, K_B^*$ by sampling with replacement from the rows of the data realization array K .
5. Form the re-scaled bootstrap samples as $\Psi_i^* = K_i^*(S^{1/2})^T$ for $i = 1, \dots, B$
6. For each re-scaled bootstrap sample, calculate the bootstrap statistic T_i^{*2} for $i = 1, \dots, B$ by using \bar{Y}_i^* in place of \bar{Y} , and \bar{Y} in place of μ_0 in equation (3)
7. The bootstrap significance level p is $\#\{T_i^{*2} > T^2\}/B$.

4. Example

The use of the two bootstrap algorithms presented in the previous section are demonstrated here with an example. Due to a variety of circumstances, the chemical isotopic data that originally spawned this research cannot be used here. Therefore, Data Set 11 from Aitchison (1986) is used to compare the results from a classical Hotelling's one-sample test and the two bootstrap techniques for testing the claim of no relative bias. Data Set 11 consist of determinations of the composition of white blood cells by two different techniques (microscopic inspection and image analysis) for each of 30 blood specimens. The composition is expressed as a fraction for each of 3 categories (granulocytes, lymphocytes, and monocytes) such that the sums of these three fractions should add to 1 for each specimen.

Let the composition be denoted by the 3 dimensional vector $U = (u_1, u_2, u_3)$. Following a procedure described in Aitchison (1986), the compositional data is transformed out of a 2 dimensional simplex in \mathbb{R}^3 and into \mathbb{R}^2 by defining the vector $W = (u_1/u_3, u_2/u_3)$. This calculation is performed for all observations for both techniques. The two techniques are hereafter distinguished by a superscript (1) for microscopic inspection or (2) for image analysis. Two methods for defining the difference between the two techniques were

investigated and these correspond to assumptions about the distribution of the variables $W^{(1)}$ and $W^{(2)}$. The first method defines the difference by the simple difference formula $W^{(2)} - W^{(1)}$, which corresponds to an assumption of multivariate normality for both $W^{(1)}$ and $W^{(2)}$. The second method defines the difference as $\log(W^{(2)}) - \log(W^{(1)})$, which corresponds to an assumption of a logistic normal distribution for both $W^{(1)}$ and $W^{(2)}$; see Aitchison (1986) for details about the multivariate logistic normal distribution.

Preliminary analysis indicates that the differences $W^{(2)} - W^{(1)}$ do not behave as a multivariate normal distribution but that the difference defined as $\log(W^{(2)}) - \log(W^{(1)})$ at least appears to have normally distributed marginal distributions. This implies that the results of Hotelling's one-sample test are questionable in inferences involving the simple differences. The significance level (*p*-value) for Hotelling's one-sample test, as well as for the two bootstrap approaches are listed in Table 1 for the both the simple differences and the difference of logs. Note that for the simple differences, there is an order of magnitude difference between Hotelling's test *p*-value and the bootstrap *p*-value. While in this particular case, these differences would not lead to a different inference, it is conceivable that in other situations, differences in the *p*-values this large might impact the inference. For the log differences, the *p*-values from Hotelling's test and the bootstrap tests differ by a factor of 4 to 6. Again in this case, the difference does not have an impact on the inference, although it is conceivable that in another problem it might affect the inference.

The bootstrap techniques developed and demonstrated here provide a method of confirming inferences made using Hotelling's multivariate one-sample test. Additionally, these methods can be used to explore the robustness of Hotelling's test to departures from multivariate normality, and they provide an alternative technique for determining significance levels for this test in cases where the *F* distribution of the T^2 statistic does not hold.

5. Acknowledgments

This work was performed under the auspices of the U. S. Department of Energy by the Lawrence Livermore National Laboratory under contract number W-7405-ENG-78.

Table 1
Results for Hotelling's One Sample Test and Bootstrap
Test of Significant Difference of Population Mean

Difference Type	Value of Hotelling's T^2 statistic	p -value by Hotelling's One-Sample Test	p -value by Naive Bootstrap	p -value by scaled bootstrap
simple	7.8687	0.0019	0.010	0.011
log	13.5625	7.609×10^{-5}	0.0004	0.0002

6. References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. London, Chapman and Hall.

Bellman, R. (1960). *Introduction to Matrix Analysis*. New York, McGraw-Hill.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Volume 30 of CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM.

Efron, B., and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. London, Chapman and Hall.

Everitt, B. S. (1979). "A Monte Carlo Investigation of the Robustness of Hotelling's One- and Two-Sample T^2 test." *Journal of the American Statistical Association* 74: 48-51.

Hinkley, D. V. (1988). "Bootstrap Methods." *Journal of the Royal Statistical Society, B* 50(3): 321-337.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis*. New York, Academic Press.

Thisted, R. A. (1988). *Elements of Statistical Computing: Numerical Computation*. London, Chapman and Hall.