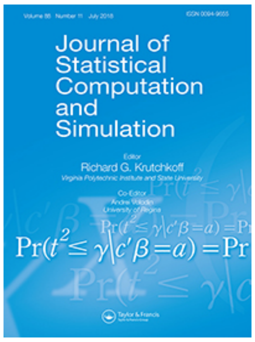

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: http://www.tandfonline.com/loi/gscs20

Bootstrap confidence intervals for the contributions of individual variables to a Mahalanobis distance

Zillur R. Shabuz & Paul H. Garthwaite

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION, 2018
VOL. 88, NO. 11, 2232–2258
https://doi.org/10.1080/00949655.2018.1453813

Bootstrap confidence intervals for the contributions of individual variables to a Mahalanobis distance

Zillur R. Shabuz and Paul H. Garthwaite

School of Mathematics and Statistics, The Open University, Milton Keynes, UK

ABSTRACT

Hotelling's T^2 and Mahalanobis distance are widely used in the statistical analysis of multivariate data. When either of these quantities is large, a natural question is: *How do individual variables contribute to its size?* The Garthwaite–Koch partition has been proposed as a means of assessing the contribution of each variable. This yields point estimates of each variable's contribution and here we consider bootstrap methods for forming interval estimates of these contributions. New bootstrap methods are proposed and compared with the percentile, bias-corrected percentile, non-studentized pivotal and studentized pivotal methods via a large simulation study. The new methods enable use of a broader range of pivotal quantities than with standard pivotal methods, including vector pivotal quantities. In the context considered here, this obviates the need for transformations and leads to intervals that have higher coverage, and yet are narrower, than intervals given by the standard pivotal methods. These results held both when the population distributions were multivariate normal and when they were skew with heavy tails. Both equal-tailed intervals and shortest intervals are constructed; the latter are particularly attractive when (as here) squared quantities are of interest.

ARTICLE HISTORY
Received 8 March 2017
Accepted 14 March 2018

KEYWORDS
Bootstrap; Garthwaite–Koch partition; Mahalanobis distance; percentile method; pivot; studentized pivotal; vector pivotal

1. Introduction

In multivariate analysis, Mahalanobis distance (MD) is the most commonly used distance measure between two vectors. It was proposed by Mahalanobis [1] as a measure of the distance between groups that takes account of multiple characteristics and the correlations between these characteristics. The initial motivation was to analyse and classify human skulls into groups, based on multiple characteristics and the MD continues to be widely used in classification problems. MD also underlies Hotelling's one-sample and two-sample T^2 tests: it forms the test statistic when multiplied by appropriate constants determined from sample size(s).

To give some specific applications of MD, in environmental and health science it has been used to identify and map suitable habitats for a species. For instance, Liu and Weng [2] calculated MD between a vector of environmental variables and the mean vector of

CONTACT Zillur R. Shabuz md.shabuz@open.ac.uk School of Mathematics and Statistics, The Open University, Milton Keynes MK7 6AA, UK

environmental factors at the closest locations to mosquito infections. Small MD values indicated a more favourable habitat for mosquitoes. In multivariate calibration, MD is used to determine multivariate outliers [3] and evaluate the representativity between two multivariate data sets [4]. In analytical chemistry, Shah and Gemperline [5] used MD in pattern recognition to classify new samples by comparing them to a set of measurements of predetermined classes. In process control, MD is used in Hotelling's T^2-test to build multivariate control charts using the original or latent variables [6]. In the field of wildlife biology, MD can be used to find the ideal landscape of some wildlife species. Clark et al. [7] developed a multivariate model based on MD in a Geographic Information System (*GIS*) to identify areas of high habitat of female black bears.

When the value of an MD or Hotelling's T^2 is large, then an obvious question is *Which variables cause it to be large?* One approach to answering this question is to form a partition of the MD, where each element of the partition is associated with one variable and an element's size measures the contribution of the variable. Garthwaite and Koch [8] recently proposed a partition of this form that has attractive properties. They partition the (squared) MD, $\Delta^2 = (\boldsymbol{X} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$, by a linear transformation from the random vector $\boldsymbol{X}$ to $\boldsymbol{W}$ such that $\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W} = \Delta^2$ and the components of $\boldsymbol{W}$ are uncorrelated, with the transformation chosen to maximize the sum of correlations between corresponding elements of $\boldsymbol{X}$ and $\boldsymbol{W}$. Rogers [9] developed global predictive risk maps for an important tropical disease, dengue, and used the partition to identify the most important predictors in determining the presence or absence of dengue in an area. Following Rogers, we refer to the partition as the Garthwaite–Koch partition.

The partition gives point estimates of the contribution of individual variables and scientists would often want interval estimates of these contributions. The task of forming confidence intervals for (un-partitioned) MDs has, of course, attracted attention. Madansky and Olkin [10] provide an approximate confidence interval based on the asymptotic distribution of the likelihood ratio statistic. More recently, Reiser [11] gave a method for constructing exact confidence intervals using the non-central F-distribution. A Bayesian approach is also possible [12]. In contrast, little work has been conducted on forming confidence intervals for the contribution to an MD given by the Garthwaite–Koch partition, although Garthwaite and Koch [8] illustrated that bootstrap confidence intervals are readily constructed. Here we consider common non-parametric bootstrap methods for forming confidence intervals and propose new methods. We then use simulation to compare their performance for confidence intervals of individual contributions of variables to MD.

If we let $\boldsymbol{W} = (W_1, \ldots, W_m)^{\mathrm{T}}$, then the contribution of the ith variable is either expressed as an absolute value, W_i^2, or as a proportion $W_i^2 / \sum_{j=1}^m W_j^2$. The standard bootstrap pivotal methods apply a one-to-one transformation to W_i^2 or $W_i^2 / \sum_{j=1}^m W_j^2$ and assume the transformed quantity is pivotal. Our new methods broaden the range of pivotal quantities that can be used. For W_i^2, a one-to-one function of W_i is treated as a pivotal quantity. (Standard bootstrap methods cannot use a one-to-one function of W_i as a pivotal quantity because the function would not have a one-to-one mapping to W_i^2.) For $W_i^2 / \sum_{j=1}^m W_j^2$, a multivariate function of $(W_1, \ldots, W_m)$ is taken as a pivotal quantity.

In the simulation study, both equal-tailed and shortest intervals are constructed. An attraction of the shortest interval for W_i^2 is that its lower limit will be 0 if the equal-tailed interval for W_i contains 0. This is intuitively desirable, as only an upper bound for W_i^2

seems of interest when it is unclear whether W_i is positive or negative. This is also the case for $W_i^2/\sum_{j=1}^m W_j^2$ if it is unclear whether $W_i/\{\sum_{j=1}^m W_j^2\}^{1/2}$ is positive or negative.

In Section 2, we briefly discuss the MD and the Garthwaite–Koch partition. In Section 3, we describe the methods used to construct bootstrap confidence intervals, including the new methods. An extensive simulation study is reported in Section 4 that examines the performance of methods when population distributions are multivariate normal. In Section 5, further simulations are reported where population distributions are skew. Concluding comments are given in Section 6.

2. MD and the Garthwaite–Koch partition

Suppose we have two distinct groups (populations) which we shall label as π_1 and π_2. For example, π_1 and π_2 might represent genuine bank notes and fake bank notes, or, in a medical diagnosis situation, those with an illness and those without it. Each individual in these groups has a number (say, m) of variables or characteristics. These characteristics may include, for example, physical measurements such as length or height, and medical features, such as body temperature or blood pressure. Let $\boldsymbol{X}$ denote a (random) vector that contains the values of these variables on an item, individual or experimental unit.

Suppose the two populations have means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and share a common covariance matrix $\boldsymbol{\Sigma}$. Then the MD between the two means is the non-negative square root of

$$\Delta_1^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2). \tag{1}$$

Of course, the population parameters are rarely known and it is usual for them to be estimated by the corresponding sample values. Suppose we have two independent random samples of sizes n_1 and n_2 $(n_1 + n_2 = n)$ from populations π_1 and π_2, yielding sample means $\bar{\boldsymbol{X}}_1$ and $\bar{\boldsymbol{X}}_2$ and sample covariance matrices $\mathbf{S}_1$ and $\mathbf{S}_2$. If the populations have the same covariance $\boldsymbol{\Sigma}$, the sample MD, D_1, can be similarly defined by

$$D_1^2 = (\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)^{\mathrm{T}}\mathbf{S}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2), \tag{2}$$

where $\mathbf{S} = \{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2\}/(n - 2)$ is an unbiased estimate of $\boldsymbol{\Sigma}$. Hotelling's two-sample T^2 statistic is $\{n_1 n_2/(n_1 + n_2)\}D_1^2$ and is used to test the hypothesis that $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are equal.

Other forms of MD are also commonly of interest. The MD between a vector $\boldsymbol{X} = (X_1, \dots, X_m)^{\mathrm{T}}$ and the mean $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)^{\mathrm{T}}$ of a population with covariance matrix $\boldsymbol{\Sigma}$ is the non-negative square root of

$$\Delta_2^2 = (\boldsymbol{X} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}), \tag{3}$$

while a useful dissimilarity measure between two random vectors $\boldsymbol{X}_{[1]}$ and $\boldsymbol{X}_{[2]}$ drawn from a distribution with the common covariance matrix $\boldsymbol{\Sigma}$ is given by

$$\Delta_3 = \{(\boldsymbol{X}_{[1]} - \boldsymbol{X}_{[2]})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{[1]} - \boldsymbol{X}_{[2]})\}^{1/2}. \tag{4}$$

Also, Hotelling's one-sample T^2 statistic is

$$T_1^2 = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_{\mathbf{0}})^{\mathrm{T}}\mathbf{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_{\mathbf{0}}) \tag{5}$$

when n observations are taken from a multivariate normal (MVN) distribution whose hypothesized mean is $\boldsymbol{\mu}_0$ and $\bar{\boldsymbol{X}}$ and $\mathbf{S}$ are the sample mean and covariance matrix.

The square of the MD is often referred to as the *Mahalanobis Index* (MI) and we shall do so here. Garthwaite and Koch [8] consider the MI given by Equation (3) for the case where $\boldsymbol{X}$ is an $m \times 1$ random vector whose covariance matrix is proportional to $\boldsymbol{\Sigma}$, and $\boldsymbol{\mu}$ is an $m \times 1$ vector that is not necessarily the mean of $\boldsymbol{X}$. They address the task of partitioning Δ_2^2 into the contribution of individual variables, thus giving an evaluation of each variable's contribution to the MI.

If $\boldsymbol{\Sigma}$ is the $m \times m$ identity matrix then clearly the contribution of each variable is the square of the corresponding component of $\boldsymbol{X} - \boldsymbol{\mu}$. This partitioning extends easily if $\boldsymbol{\Sigma}$ is a diagonal matrix. However, if the variables are correlated with each other (so $\boldsymbol{\Sigma}$ is not diagonal) then the MI cannot be partitioned in a straightforward way. Garthwaite and Kock consider transformations of the form

$$\boldsymbol{X} \rightarrow \boldsymbol{W} = \mathbf{A}(\boldsymbol{X} - \boldsymbol{\mu}) \tag{6}$$

such that $\boldsymbol{W}$ is an $m \times 1$ vector and

$$\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W} = (\boldsymbol{X} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \tag{7}$$

for any vector $\boldsymbol{X}$. Then

$$\Delta_2^2 = \sum_{i=1}^{m} W_i^2 \tag{8}$$

where $\boldsymbol{W} = (W_1, \ldots, W_m)^{\mathrm{T}}$ and so $\boldsymbol{W}$ yields a partition of Δ_2^2. They argue that the contribution of X_i to Δ_2^2 can sensibly be defined as W_i^2 if

(a) it is reasonable to identify W_i with X_i and
(b) the components of $\boldsymbol{W}$ are uncorrelated and have identical variances.

To this end, they choose $\mathbf{A}$ in (6) such that $\sum_{i=1}^{m} \mathrm{corr}(X_i, W_i)$ is maximized under the condition that Equation (7) holds for all $\boldsymbol{X}$. This yields

$$\mathbf{A} = (\mathbf{G}\boldsymbol{\Sigma}\mathbf{G})^{-1/2}\mathbf{G} \tag{9}$$

where $\mathbf{G}$ is a positive-definite diagonal matrix and $\mathbf{G}\boldsymbol{\Sigma}\mathbf{G}$ has diagonal elements of 1. That is, $\mathbf{G}$ has diagonal elements equal to the reciprocal of the square root of the corresponding diagonal element of $\boldsymbol{\Sigma}$. As $\sum_{i=1}^{m} \mathrm{corr}(X_i, W_i)$ is maximized, each component X_i is identified with the corresponding component W_i in a one-to-one relationship. The sample estimate of $\boldsymbol{W}$, denoted by $\widehat{\boldsymbol{W}} = (\hat{W}_1, \ldots, \hat{W}_m)^{\mathrm{T}}$, is obtained by replacing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with (unbiased) sample estimates.

In examples given in Garthwaite and Koch [8], the partition always gives a sensible evaluation of the contributions of individual X variables. Rogers [9] uses the partition for disease mapping and notes that 'Identifying the key model variables in predicting the changing spatial pattern of vector-borne diseases over time is now made possible by the Garthwaite–Koch technique'.

3. Bootstrap confidence intervals

Bootstrapping is a resampling technique for estimating the sampling distribution of estimators and making inference about the corresponding parameters when there are no theoretical results on which to base inferences. Bootstrap methods were introduced by Efron [13] and they have become the standard means of forming confidence intervals when analytically formed intervals are unavailable. It is assumed that sample observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are independent realizations of a random variable $\boldsymbol{X}$ whose probability density function (*p.d.f.*), f, and cumulative distribution function (*c.d.f.*), F, are unknown. The aim is to construct a confidence interval for some population characteristic, which we denote by θ. While $\boldsymbol{X}$ may be a vector, θ is a scalar.

Steps in the algorithm for the nonparametric bootstrap are as follows [14].

(a) Sample n observations randomly with equal probability and replacement from $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ to create a bootstrap set of the same size as the original study data.
(b) Calculate the bootstrap version (replication) of the statistic of interest ($\hat{\theta}$) in the same way as for the study data set.
(c) Repeat stages (a) and (b) N times, where N is large, to obtain N bootstrap replications which together form the bootstrap distribution of θ.

We let $\hat{\theta}$ denote the estimate of θ given by the original data and $\hat{\theta}_1^\star, \ldots, \hat{\theta}_N^\star$ denote the estimates given by the N bootstrap replications.

Various methods have been proposed for constructing a confidence for θ from $\hat{\theta}_1^\star, \ldots, \hat{\theta}_N^\star$. For good reviews of a number of methods, see [15–17]. We will consider four commonly used methods: the percentile method, the bias-corrected percentile method, the non-studentized pivotal method and the studentized pivotal method. We also propose new methods that deviate from the above algorithm for obtaining bootstrap replications of θ. The new methods introduce a parameter, γ say, that determines θ, but while θ must be a function of γ, the function need not be one-to-one. It is an estimate of γ that is determined from each bootstrap set and the estimates are manipulated to form a confidence interval for θ, using a method that has similarities to a bootstrap pivotal method (see Section 3.3).

In the present paper, the characteristic of interest, θ, reflects the contribution of the ith component of $\boldsymbol{X}$ to the MI. This contribution is defined to be W_i^2 $(i = 1, \ldots, m)$, where W_i is the ith component of $(\mathbf{G\Sigma G})^{-1/2}\mathbf{G}(\boldsymbol{X} - \boldsymbol{\mu})$. We examine bootstrap methods for forming confidence intervals for (i) W_i^2 and (ii) $W_i^2 / \sum_{j=1}^m W_j^2$. The latter quantity is the *proportion* of the MI that is attributable to the ith X-variable and is a readily interpretable measure of the ith variable's importance. As W_i^2 is non-negative, the distribution of its sample estimate will be markedly skew when the point estimate of W_i is near 0. Thus an equal-tailed interval will sometimes be markedly longer than the shortest interval that has the same level of confidence. Partly for this reason, we consider shortest confidence intervals as well as equal-tailed confidence intervals.

The other reason is that we believe there should be some coherence between a confidence interval for W_i^2 and a confidence interval for W_i (and similarly with $W_i^2 / \sum_{j=1}^m W_j^2$ and $W_i / \{\sum_{j=1}^m W_j^2\}^{1/2}$). With regard to this, consider the question:

What is a sensible confidence interval for W_i^2 if the confidence interval for W_i includes 0?

When W_i has the sampling distribution given in Figure 1(a), then W_i^2 has the sampling distribution in Figure 1(b). The equal-tailed confidence interval for W_i is indicated in Figure 1(a) and the shortest interval for W_i^2 is marked in Figure 1(b). The latter interval not only includes all the plausible values for W_i^2 but also the square of the most plausible values for W_i. This is not true of an equal-tailed confidence intervals for W_i^2, as the interval

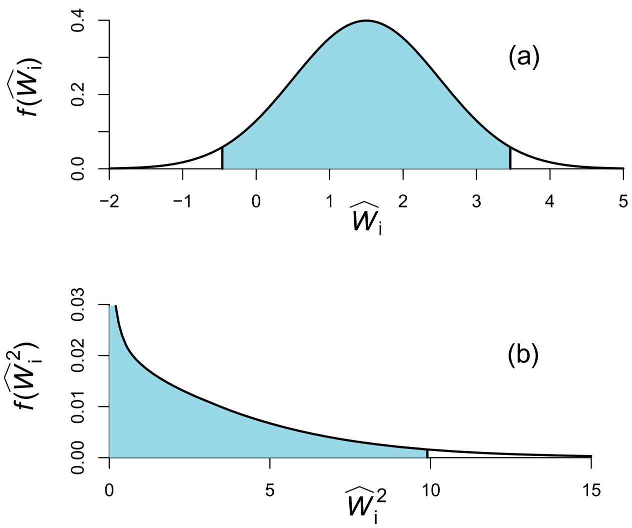

Figure 1. (a) Probability distribution for $\hat{W}_i$ and equal-tailed confidence interval for W_i, and (b) corresponding probability distribution for $\hat{W}_i^2$ and shortest confidence interval for W_i^2.

would not contain 0. (Obviously the shortest confidence intervals for W_i^2 will not have 0 as its lower endpoint when the sign of W_i is clear.)

As Figures 1(a) and 1(b) illustrate, a shortest confidence interval for W_i^2 should sometimes have 0 as its lower endpoint. However, an empirical bootstrap distribution is discrete and the smallest bootstrap estimate of W_i^2 is unlikely to equal 0 precisely. It follows that the lower endpoint of a confidence interval will not equal 0, as the lower endpoint cannot be less than the smallest bootstrap estimate. As a 'continuity correction', for a 95% confidence interval (for both W_i^2 and $W_i^2/\sum_{j=1}^m W_j^2$) we form intervals for the following combinations of (α_1, α_2): (0.000, 0.050), (0.005, 0.045), ..., (0.045, 0.005) and (0.050, 0.000), where the lower interval endpoint is taken as 0 when $\alpha_1 = 0$. We observe which of the intervals is shortest and take that as the 'shortest' confidence interval. This often gives 0 as the lower limit. A similar approach can be used for other levels of confidence.

We next describe the bootstrap methods of interest in this paper. We suppose we wish to form a confidence interval for which the lower and upper tail areas are α_1 and α_2, respectively. For an equal-tailed $100(1-2\alpha)\%$ confidence interval, $\alpha_1 = \alpha_2 = \alpha$. For a shortest interval, α_1 and α_2 are varied, subject to $\alpha_1 + \alpha_2 = 2\alpha$.

3.1. Percentile methods

3.1.1. Percentile method

The simplest method of forming a bootstrap confidence interval is the percentile method, which simply equates quantiles of the distribution of $\hat{\theta}$ to the equivalent quantiles of the bootstrap distribution of $\hat{\theta}^\star$. This gives $(\hat{\theta}^\star(\alpha_1), \hat{\theta}^\star(1-\alpha_2))$ as a $100(1-2\alpha)\%$ confidence interval for θ, where $\hat{\theta}^\star(q)$ denotes the qth quantile of the bootstrap distribution.

This method has simplicity, can be applied to any statistic, and no invalid parameter values will be included in the confidence interval, as the method is range-preserving. Also, the method is transformation respecting, implying that if (θ_L, θ_U) is a $100(1-2\alpha)\%$

confidence interval for θ and g is a monotonic increasing transformation of θ, then $(g(\theta_L), g(\theta_U))$ is a $100(1-2\alpha)\%$ confidence interval for $g(\theta)$. Largely for these reasons, the method is widely used. However, if the distribution of $\hat{\theta}$ is markedly skew, the coverage error of equal-tailed intervals is often substantial [15].

In applying both this method and the bias-corrected percentile method, θ will be equated to either W_i^2 or $W_i^2/\sum_{j=1}^m W_j^2$.

3.1.2. Bias-corrected percentile method

The bias-corrected percentile method (*BC* method) is a modification of the percentile method that aims to improve coverage for non-symmetric distributions. Its steps are as follows:

(1) Let $\hat{\theta}_k^\star$ denote the estimate of θ given by the kth bootstrap resample. Count the number of members of $\hat{\theta}_1^\star, \hat{\theta}_2^\star, \ldots, \hat{\theta}_N^\star$ that are less than $\hat{\theta}$ (calculated from the original data). Call this number p and set $p^\star = p/N$. Set $z_0 = \Phi^{-1}(p^\star)$, where Φ denotes the *c.d.f.* of the standard normal distribution.
(2) Define $\hat{\alpha}_1$ and $\hat{\alpha}_2$ as $\hat{\alpha}_1 = \Phi(2z_0 + z_1)$ and $\hat{\alpha}_2 = 1 - \Phi(2z_0 + z_2)$, where $z_1 = \Phi^{-1}(\alpha_1)$ and $z_2 = \Phi^{-1}(1-\alpha_2)$.
(3) Take $\hat{\theta}^\star(\hat{\alpha_1})$ and $\hat{\theta}^\star(1-\hat{\alpha_2})$ as the endpoints of the confidence interval.

This method is as easily implemented as the percentile method. If the distribution of $\hat{\theta}^\star$ is symmetric about $\hat{\theta}$, that is when $z_0 = 0$, the bias-corrected percentile interval and percentile interval are the same. Hence the method may be thought of as a 'fine-tuning' of the percentile method.

3.2. Pivotal methods

Pivotal methods form a function of θ and $\hat{\theta}$ that is treated as pivotal: it is assumed that the sampling distribution of the function does not depend upon any unknown quantities. The most commonly used functions for the non-studentized and studentized pivotal methods are $\hat{\theta} - \theta$ and $(\hat{\theta} - \theta)/\hat{\sigma}(\hat{\theta})$, respectively, where $\hat{\sigma}(\hat{\theta})$ is the standard error of $\hat{\theta}$.

The methods can yield confidence intervals that include invalid parameter values if the range of θ is bounded. Moreover, neither $\hat{\theta} - \theta$ nor $(\hat{\theta} - \theta)/\hat{\sigma}(\hat{\theta})$ could be a pivotal function if the range of θ is bounded. Transformations are the usual approach to counter this problem. If the parameter of real interest has a bounded range, then θ is equated to some monotonic increasing function of the parameter. A confidence interval for θ is constructed and its endpoints transformed back, giving a confidence interval for the true parameter of interest. However, pivotal methods are not transformation respecting, so the choice of transformation will affect the endpoints of confidence intervals. Here we put $\theta = \log W_i^2$ or $\theta = \mathrm{logit}[W_i^2/\sum_{j=1}^m W_j^2]$ when seeking a confidence interval for W_i^2 or $W_i^2/\sum_{j=1}^m W_j^2$, respectively. In both cases the resulting θ has a range of $(-\infty, \infty)$.

3.2.1. Non-studentized pivotal (basic) method

The non-studentized pivotal method makes the assumption that the distribution of $\psi = \hat{\theta} - \theta$ is similar to the distribution of $\hat{\psi}^\star = \hat{\theta}^\star - \hat{\theta}$. Quantiles of the bootstrap distribution of $\hat{\psi}^\star$ are used to form confidence intervals for θ. The steps are as follows:

(1) Set $\hat{\psi}_k^\star = \hat{\theta}_k^\star - \hat{\theta}$, for $k = 1, 2, \ldots, N$.
(2) Determine $\hat{\psi}^\star(\alpha_2)$ and $\hat{\psi}^\star(1-\alpha_1)$, where $\hat{\psi}^\star(q)$ denotes the qth quantile of the bootstrap distribution of $\hat{\psi}^\star$.
(3) Then a $100(1-2\alpha)\%$ confidence interval for θ is given by $(\hat{\theta} - \hat{\psi}^\star(1-\alpha_1),\ \hat{\theta} - \hat{\psi}^\star(\alpha_2))$, which can equivalently be written as $(2\hat{\theta} - \hat{\theta}^\star(1-\alpha_1),\ 2\hat{\theta} - \hat{\theta}^\star(\alpha_2))$.

This method is simple to use, but the coverage error can be substantial if the distributions of ψ and $\hat{\psi}^\star$ differ in a clearly noticeably manner.

3.2.2. Studentized pivotal method (bootstrap t method)

The bootstrap t method aims to improve on the basic bootstrap method by treating $(\hat{\theta} - \theta)/\hat{\sigma}(\hat{\theta})$ as a pivotal quantity, where $\hat{\sigma}(\hat{\theta})$ denotes the estimated standard error of $\hat{\theta}$. It derives its name from the fact that when $\hat{\theta} \sim N(\theta, \sigma^2)$, then $(\hat{\theta} - \theta)/\hat{\sigma}(\hat{\theta})$ is a pivotal quantity that has a t-distribution. The method assumes that $(\hat{\theta} - \theta)/\hat{\sigma}(\hat{\theta})$ and $(\hat{\theta}^\star - \hat{\theta})/\hat{\sigma}(\hat{\theta}^\star)$ have similar distributions. The following are its primary steps.

(1) Set $\hat{\xi}_k^\star = (\hat{\theta}_k^\star - \hat{\theta})/\hat{\sigma}_k(\hat{\theta}^\star)$, for $k = 1, 2, \ldots, N$, where $\hat{\sigma}_k(\hat{\theta}^\star)$ is an estimate of the standard error of $\hat{\theta}_k^\star$ (see below).
(2) Determine $\hat{\xi}^\star(\alpha_2)$ and $\hat{\xi}^\star(1-\alpha_1)$, where $\hat{\xi}^\star(q)$ denotes the qth quantile of the bootstrap distribution of $\hat{\xi}^\star$.
(3) Then a $100(1-2\alpha)\%$ confidence interval for θ is given as $(\hat{\theta} - \hat{\sigma}(\hat{\theta})\hat{\xi}^\star(1-\alpha_1),\ \hat{\theta} - \hat{\sigma}(\hat{\theta})\hat{\xi}^\star(\alpha_2))$.

The method requires estimates $\hat{\sigma}(\hat{\theta})$ and $\hat{\sigma}_k(\hat{\theta}^\star)$. The former is obtained from

$$\hat{\sigma}^2(\hat{\theta}) = \frac{1}{N-1} \sum_{k=1}^{N} (\hat{\theta}_k^\star - \bar{\theta}^\star)^2, \tag{10}$$

where $\bar{\theta}^\star$ is the mean of $\hat{\theta}_1^\star, \ldots, \hat{\theta}_N^\star$. The latter might be estimated using the jackknife. The alternative, that we use, is to carry out a computationally intensive, but routine, 'second-level bootstrap' to estimate $\hat{\sigma}_k(\hat{\theta}^\star)$, as follows.

Let $\boldsymbol{x}_{k1}^\star, \ldots, \boldsymbol{x}_{kn}^\star$ be the kth ($k = 1, 2, \ldots, N$) bootstrap sample and let $\hat{\theta}_k^\star$ denote the estimate of θ it gives. Obtain a second-level bootstrap sample $\boldsymbol{x}_{k1}^{\star\star}, \ldots, \boldsymbol{x}_{kn}^{\star\star}$ by sampling with replacement from $\boldsymbol{x}_{k1}^\star, \ldots, \boldsymbol{x}_{kn}^\star$ and evaluate the estimate of θ. Repeat this B times and let $\hat{\theta}_{k\ell}^{\star\star}$ denote the estimate given by the ℓth second-level sample ($\ell = 1, 2, \ldots, B$). Then the estimate of the variance of $\hat{\theta}_k^\star$ is

$$\hat{\sigma}_k^2(\hat{\theta}^\star) = \frac{1}{B-1} \sum_{\ell=1}^{B} (\hat{\theta}_{k\ell}^{\star\star} - \bar{\theta}_k^\star)^2, \tag{11}$$

where $\bar{\theta}_k^\star$ is the mean of $\hat{\theta}_{k1}^{\star\star}, \hat{\theta}_{k2}^{\star\star}, \ldots, \hat{\theta}_{kB}^{\star\star}$.

From each bootstrap resample, at least 25 second-level bootstrap samples should be taken [17]. The obvious drawback of the studentized pivotal method is that the process is computationally intensive – to generate a total of N values of $\hat{\theta}^\star$, a total of BN bootstrap samples are required. The method can perform very poorly if $\hat{\sigma}(\hat{\theta})$ is not independent of θ, but simulation results reported in the literature [17] suggest that the method often gives more accurate coverage than other bootstrap methods.

3.3. New methods

The new methods broaden the range of pivotal quantities that can be used to form bootstrap confidence intervals for θ. The methods are partly Bayesian, in that parameters are treated as variables that have probability distributions – but no prior distributions are specified.

Let $\theta = h(\gamma)$, where γ may be a vector and h is not necessarily a monotonic function, nor necessarily a one-to-one function. The sample data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ yield an estimate $\hat{\gamma}$ of γ. From a bootstrap resample, we determine an estimate $\hat{\gamma}^*$ of $\hat{\gamma}$ in the same way as $\hat{\gamma}$ was determined from the original sample. Let $\hat{\gamma}_1^*, \ldots, \hat{\gamma}_N^*$ denote the estimates given by the N resamples.

When seeking a confidence interval for W_i^2, we set $\theta = W_i^2$ and $\gamma = W_i$. For an interval for $W_i^2 / \sum_{j=1}^m W_j^2$, we put $\theta = W_i^2 / \sum_{j=1}^m W_j^2$ and $\gamma = (W_1, \ldots, W_m)^{\mathrm{T}}$.

3.3.1. Method A

The first of our new methods, Method A, treats $\hat{\gamma} - \gamma$ as a pivotal quantity and makes the following assumption.

Assumption A: *Given any γ, the statistics $\hat{\gamma} - \gamma$ and $\hat{\gamma}^* - \hat{\gamma}$ are from the same distribution.*

Let $\hat{P}_{\hat{\gamma}^*|\hat{\gamma}}$ denote bootstrap probabilities when $\hat{\gamma}^*$ is considered a random variable and $\hat{\gamma}$ is non-random. Similarly, let $\hat{P}_{\hat{\gamma}|\gamma}$ denote bootstrap probabilities when $\hat{\gamma}$ is random while γ is non-random. We have,

$$\hat{P}_{\hat{\gamma}^*\,|\,\hat{\gamma}}(\hat{\gamma}^* - \hat{\gamma} = \nu) = \begin{cases} 1/N & \nu = \hat{\gamma}_k^* - \hat{\gamma}; \quad k = 1, \ldots, N \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

so, from Assumption A,

$$\hat{P}_{\hat{\gamma}\,|\,\gamma}(\hat{\gamma} - \gamma = \nu) = \begin{cases} 1/N & \nu = \hat{\gamma}_k^* - \hat{\gamma}; \quad k = 1, \ldots, N \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

To add flexibility, we wish to allow γ to have a probability distribution, so we adopt a Bayesian approach and let $\hat{P}_{\gamma\,|\,\hat{\gamma}}$ denote bootstrap posterior probabilities, where γ is now random while $\hat{\gamma}$ is non-random. So that Bayesian credible intervals match frequentist confidence intervals, we assume that γ has a probability matching prior distribution. (See, for example, Datta and Mukerjee [18] for details of probability matching priors.) Then, from

Equation (13),

$$\hat{P}_{\gamma\,|\,\hat{\gamma}}(\hat{\gamma}-\gamma=\nu)=\begin{cases}1/N & \nu=\hat{\gamma}_k^*-\hat{\gamma}; \quad k=1,\dots,N\\ 0 & \text{otherwise,}\end{cases} \tag{14}$$

so

$$\hat{P}_{\gamma\,|\,\hat{\gamma}}(\gamma=\eta)=\begin{cases}1/N & \eta=2\hat{\gamma}-\hat{\gamma}_k^*; \quad k=1,\dots,N\\ 0 & \text{otherwise.}\end{cases} \tag{15}$$

Let $\hat{P}_{\theta\,|\,\hat{\gamma}}$ denote the bootstrap posterior probability distribution of θ. As $\theta=h(\gamma)$, Equation (15) gives

$$\hat{P}_{\theta\,|\,\hat{\gamma}}[\theta=h(\eta)]=\begin{cases}1/N & \eta=2\hat{\gamma}-\hat{\gamma}_k^*; \quad k=1,\dots,N\\ 0 & \text{otherwise.}\end{cases} \tag{16}$$

Quantiles from the distribution in Equation (16) yield a Bayesian credible interval for θ and, because a probability matching prior has been used, we take this as the confidence interval. That is, if $\theta_A^{\#}(q)$ denotes the qth quantile of $\hat{P}_{\theta\,|\,\hat{\gamma}}(\theta)$ given by (16), then $(\theta_A^{\#}(\alpha_1),\ \theta_A^{\#}(1-\alpha_2))$ is the $100(1-2\alpha)\%$ confidence interval for θ given by Method A.

The following summarizes the steps for Method A.

(1) Determine $\hat{\gamma}_1^*,\dots,\hat{\gamma}_N^*$ from the N bootstrap resamples.
(2) Put $\hat{\vartheta}_k^*=h(2\hat{\gamma}-\hat{\gamma}_k^*)$ for $k=1,\dots,N$.
(3) Then a $100(1-2\alpha)\%$ confidence interval for θ is given as $(\hat{\vartheta}^*(\alpha_1),\ \hat{\vartheta}^*(1-\alpha_2))$, where $\hat{\vartheta}^*(q)$ is the qth sample quantile of the sample $\hat{\vartheta}_1^*,\dots,\hat{\vartheta}_N^*$.

3.3.2. Method B

Suppose γ is an r-dimensional vector. Let $\hat{\sigma}(\hat{\gamma}_{(j)})$ and $\hat{\sigma}(\hat{\gamma}_{(j)}^*)$ denote the estimated standard errors of the jth components of $\hat{\gamma}$ and $\hat{\gamma}^*$, respectively. Also, let $\hat{\tau}(\hat{\gamma})$ and $\hat{\tau}(\hat{\gamma}^*)$ denote $r\times r$ diagonal matrices whose jth diagonal elements are $[\hat{\sigma}(\hat{\gamma}_{(j)})]^{-1}$ and $[\hat{\sigma}(\hat{\gamma}_{(j)}^*)]^{-1}$, respectively. Method B makes the following assumption.

Assumption B: *Given any γ, the statistics $\hat{\tau}(\hat{\gamma})(\hat{\gamma}-\gamma)$ and $\hat{\tau}(\hat{\gamma}^*)(\hat{\gamma}^*-\hat{\gamma})$ are from the same distribution.*

Consequently, the difference between Method A and Method B is similar to the difference between the non-studentized and studentized pivotal methods. Corresponding to Equations (12) and (14), the bootstrap probabilities are

$$\hat{P}_{\hat{\gamma}^*\,|\,\hat{\gamma}}\left[\hat{\tau}(\hat{\gamma}^*)\{\hat{\gamma}^*-\hat{\gamma}\}=\nu\right]=\begin{cases}1/N & \nu=\hat{\tau}_k(\hat{\gamma}^*)\{\hat{\gamma}_k^*-\hat{\gamma}\}; \quad k=1,\dots,N\\ 0 & \text{otherwise,}\end{cases} \tag{17}$$

and the resulting bootstrap posterior probabilities are:

$$\hat{P}_{\gamma\,|\,\hat{\gamma}}\left[\hat{\tau}(\hat{\gamma})\{\hat{\gamma}-\gamma\}=\nu\right]=\begin{cases}1/N & \nu=\hat{\tau}_k(\hat{\gamma}^*)\{\hat{\gamma}_k^*-\hat{\gamma}\}; \quad k=1,\dots,N\\ 0 & \text{otherwise.}\end{cases} \tag{18}$$

This yields the bootstrap posterior distribution of θ:

$$\hat{P}_{\theta\,|\,\hat{\gamma}}[\theta = h(\eta)] = \begin{cases} 1/N & \eta = \hat{\gamma} - \{\hat{\tau}(\hat{\gamma})\}^{-1}\hat{\tau}_k(\hat{\gamma}^*)\{\hat{\gamma}_k^* - \hat{\gamma}\}; \quad k = 1, \ldots, N \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

Let $\theta_B^{\#}(q)$ denote the qth quantile of $\hat{P}_{\theta\,|\,\hat{\gamma}}(\theta)$ given by (19). Then $(\theta_B^{\#}(\alpha_1),\ \theta_B^{\#}(1-\alpha_2))$ is the $100(1-2\alpha)\%$ confidence interval for θ given by Method B.

Method B requires estimates of the standard errors, $\hat{\sigma}(\hat{\gamma}_{(j)})$ and $\hat{\sigma}_k(\hat{\gamma}^*_{(j)})$ for $k = 1, \ldots, N; j = 1, \ldots, r$. These are obtained in a way analogous to the procedure for obtaining $\hat{\sigma}(\hat{\theta})$ and $\hat{\sigma}_k(\hat{\theta}^*)$ in the studentized pivotal method. The following summarizes the steps in Method B.

(1) Generate N bootstrap resamples to obtain estimates $\hat{\gamma}_1^*, \ldots, \hat{\gamma}_N^*$. The sample standard deviation of the jth components of the $\hat{\gamma}_k^*$ is taken as $\hat{\sigma}(\hat{\gamma}_{(j)})$. The diagonal elements of $\hat{\tau}(\hat{\gamma})$ are set equal to $\{\hat{\sigma}(\hat{\gamma}_{(1)})\}^{-1}, \ldots, \{\hat{\sigma}(\hat{\gamma}_{(r)})\}^{-1}$.
(2) From the kth bootstrap resample ($k = 1, \ldots, N$), generate B second-level bootstrap samples and estimate γ in each. Let $\hat{\gamma}_{kl}^{**}$ denote the estimate of γ given by the lth second-level sample ($l = 1, \ldots, B$). The sample standard deviation of the jth components of the $\hat{\gamma}_{kl}^{**}$ is taken as $\hat{\sigma}_k(\hat{\gamma}^*_{(j)})$. For $k = 1, \ldots, N$, the diagonal elements of $\hat{\tau}_k(\hat{\gamma}^*)$ are set equal to $\{\hat{\sigma}_k(\hat{\gamma}^*_{(1)})\}^{-1}, \ldots, \{\hat{\sigma}_k(\hat{\gamma}^*_{(r)})\}^{-1}$.
(3) Put $\hat{\lambda}_k^* = h[\hat{\gamma} - \{\hat{\tau}(\hat{\gamma})\}^{-1}\hat{\tau}_k(\hat{\gamma}^*)\{\hat{\gamma}_k^* - \hat{\gamma}\}]$ for $k = 1, \ldots, N$.
(4) Then a $100(1-2\alpha)\%$ confidence interval for θ is given as $(\hat{\lambda}^*(\alpha_1),\ \hat{\lambda}^*(1-\alpha_2))$, where $\hat{\lambda}^*(q)$ is the qth sample quantile of the sample $\hat{\lambda}_1^*, \ldots, \hat{\lambda}_N^*$.

When γ is a scalar and h is a monotonic transformation, standard bootstrap methods can be used to first form a bootstrap confidence interval for γ and then the endpoints of the interval can be back-transformed to obtain a bootstrap confidence interval for θ. If the non-studentized pivotal method is used to form the bootstrap confidence for γ, then the resulting confidence interval for θ is identical to the interval given by Method A. If the studentized pivotal method is used, the resulting confidence interval is identical to the interval given by Method B. The advantages of Methods A and B are that they can be used when γ is not a scalar and h is not a monotonic transformation.

4. Simulation study: MVN distributions

A large simulation study was conducted to evaluate the coverage probabilities of the six methods. In this section, we use an MVN distribution to describe each population because this is consistent with the assumptions underlying Hotelling's T^2 hypothesis test and the test of whether an MD is unusually large. In Sections 4.1–4.3, the mechanics of the simulations are described and results are presented in Section 4.4.

4.1. Population distributions

We require a number of known population distributions. To mimic reality, we set the mean and variance of each population distribution equal to the sample mean and variance of a real data set, using the following five data sets.

Table 1. Features of the data sets.

Data set	Sample sizes	No. of variables	Range of absolute correlations
Bank notes	100 and 100	6	0.000–0.664
Athletes	102 and 100	9	0.017–0.967
Skulls	17 and 15	5	0.011–0.718
Psychological tests	32 and 32	4	0.322–0.628
Flea beetles	19 and 20	4	0.074–0.727

(1) *Swiss bank notes*: The data set is given in Flury and Riedwyl [19]. It contains six measurements that were made on 100 genuine bank notes and 100 forged bank notes. The measurements were: *length* (length of bank note), *left* (width of note, measured on its left side), *right* (width of note, measured on the right), *bottom* (width of margin at the bottom), *top* (width of margin at the top) and *diagonal* (length of the image diagonal). All variables were measured in millimetres.
(2) *Male and female athletes*: Data on 102 male and 100 female athletes were collected at the Australian Institute of Sport [20]. For our study, we considered the following nine measurements on each athlete: *Wt* (weight in kg), *Ht* (height in cm), *RCC* (red cell count), *Hg* (haemoglobin), *Hc* (hematocrit), *WCC* (white cell count), *Ferr* (plasma ferritin concentration), *Bfat* (% body fat) and *SSF* (sum of skin fold thickness).
(3) *Tibetan skulls*: Data reported in Morant [21] were collected from south-western and eastern districts of Tibet. Five measurements (all in millimetres) were made on each of 32 skulls: *Length* (greatest length of skull), *Breadth* (greatest horizontal breadth of skull), *Height* (height of skull), *Fheight* (upper face length) and *Fbreadth* (face breadth between outermost points of cheekbones). The first 17 skulls came from graves in Sikkim and the neighbouring area of Tibet, and the remaining 15 were from a battlefield in the Lhasa district.
(4) *Psychological measurements*: Beall [22] gives data on 32 men and 32 women. Four psychological measurements were made on each person: *Pi* (pictorial inconsistencies), *Tr* (tool recognition), *Pb* (paper from board) and *Vo* (vocabulary).
(5) *Flea beetles*: Lubischew [23] gives data on two species of flea beetles (*Haltica oleracea* and *Haltica carduorum*). Four measurements were made on each flea: *Dt* (distance of transverse groove from posterior border of prothorax (microns)), *Le* (length of elytra (0.01 mm)), *Ls* (length of second antenatal joint (microns)) and *Lt* (length of third antenatal joint (microns)).

Table 1 summarizes some key features of the data sets: their sizes, the number of variables considered here and the range of correlations between variables. It can be seen that the data sets vary in size from quite small (15) to moderately large (102) and each set contains between 4 and 9 variables. The correlations between variables vary from small to moderate in most data sets, with the largest correlation in a data set generally lying between 0.62 and 0.72. The Athletes data set is an exception with some correlations above 0.95.

4.2. Simulation procedure for the bank note data set

To simplify explanation, we first focus on the bank notes data set. We calculated the mean and covariance of the 100 genuine bank note measurements and took these as the mean ($\boldsymbol{\mu}$)

and variance ($\boldsymbol{\Sigma}$) of an MVN population distribution describing genuine bank notes. We also took one of the fake bank notes and examined how its measurements ($\boldsymbol{x}$) distinguish it from the genuine bank notes. To this end, we calculated $(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$, the MI (squared MD) between $\boldsymbol{x}$ and $\boldsymbol{\mu}$, and then applied the Garthwaite–Koch partition to evaluate the contributions of individual variables. Through simulation, we investigated different ways of forming confidence intervals for these contributions (W_i^2) and their percentages ($100\% \times W_i^2/\sum W_j^2$). We examined various sample sizes (20, 50, 80, 100 and 200) and 10 fake bank notes (the first 10 in the data set). For simplicity, we describe the simulation procedure for samples of size 20.

We generated one data sample of size 20 from an MVN($\boldsymbol{\mu}, \boldsymbol{\Sigma}$) distribution and then generated 1000 bootstrap resamples from this data sample. Each resample was a random sample of size 20 drawn with replacement from the data sample. We took one of the first 10 fake bank notes and calculated the MI between that bank note and the mean of the resample, using the estimated covariance matrix of the resample, $\widehat{\boldsymbol{\Sigma}}$. Then estimates of $W_1, \ldots, W_6$ from this resample were calculated using the Garthwaite–Koch partition. This gives 1000 estimates of each W_i. From each bootstrap sample, 25 second-level bootstrap samples were generated so as to determine (approximate) standard errors of the estimates. The standard errors were needed for the studentized pivotal method and Method B. After generating the bootstrap samples and second-level samples, 95% confidence intervals for individual contributions and their percentages were calculated using the methods discussed in Section 3.

As noted above, the Garthwaite–Koch partition was also applied to $(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$. This yielded 'true values' for individual contributions and their percentages, and we determined which confidence intervals covered their target values. The procedure was repeated 1000 times by generating 1000 data samples, from which we estimated coverage probabilities of the confidence intervals for each variable's contributions and their percentage contributions.

4.3. Simulation procedures for other data sets

Simulation procedures for the Athletes data set and the Tibetan skulls data set were the same as for the bank note data set. For the Athletes data set, the male athletes took the role of the genuine bank notes so their sample mean and covariance matrix became the mean and variance of the population distribution. The first 10 female athletes took the role of the first 10 fake bank notes, so the MIs from each of these female athletes to the mean of the men were the quantity of interest. For the Tibetan skulls data set, the 17 skulls from the Sikkim area took the role of the genuine bank notes while the first 10 skulls from the Lhasa district took the role of the fake bank notes.

These simulations all concern the MI between an individual and a mean. However, the MI between two means is also of importance, so with the last two data sets we examined the MI that underlies Hotelling's two-sample T^2 test. From each data set, two population MVN distributions were constructed that had different means, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, but the same covariance matrix, $\boldsymbol{\Sigma}$. For the psychological test data set, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ were set equal to the sample means for men and women, and $\boldsymbol{\Sigma}$ was equated to their pooled sample covariance matrix. Two sample groups, each of size n, were generated from the population distributions and resamples of size n were generated by sampling with replacement from each group separately. The MI between the means of the resamples for men and women was calculated using the pooled covariance matrix of the resamples as $\widehat{\boldsymbol{\Sigma}}$. Estimates of the W_i were evaluated using the Garthwaite–Koch partition and confidence intervals were constructed in the same way as with the bank notes data. Sample sizes of $n = 20, 50, 80, 100$ and 200 were examined. This simulation procedure was also followed with the Flea beetles data set; $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ were set equal to the sample means of *Haltica oleracea* and *Haltica carduorum*, respectively, and $\boldsymbol{\Sigma}$ to their pooled sample covariance matrix.

4.4. Results

Table 2 illustrates the output that was obtained for a single simulation. It is from the simulations of contributions of individual variables for the bank note data and gives results for the studentized pivotal method for the first fake bank note with samples of size 20. Various 95% confidence intervals were constructed, with different nominal coverages in each

Table 2. Coverages (%) of confidence intervals formed by the studentized pivotal method for the contributions of individual variables to the MI of the first fake bank note with samples of size 20. Coverages of tail areas and median widths of confidence intervals are also given.

	First three variables											
	Length				Left				Right			
Tails[a]	CCI[b]	CLT[c]	CRT[d]	Width[e]	CCI[b]	CLT[c]	CRT[d]	Width[e]	CCI[b]	CLT[c]	CRT[d]	Width[e]
(0.0, 5.0)	98.3	0.0	1.7	20.8	91.1	0.0	8.9	24.9	94.7	0.0	5.3	26.0
(0.5, 4.5)	98.2	0.4	1.4	22.1	91.8	0.1	8.1	28.5	94.7	0.4	4.9	28.1
(1.0, 4.0)	98.1	0.6	1.3	23.7	91.8	0.4	7.8	32.2	94.6	0.7	4.7	31.5
(1.5, 3.5)	97.6	1.3	1.1	25.6	91.8	1.0	7.2	37.7	94.2	1.5	4.3	36.0
(2.0, 3.0)	97.5	1.5	1.0	28.0	92.2	1.1	6.7	44.4	93.6	2.5	3.9	41.9
(2.5, 2.5)	97.0	2.0	1.0	31.2	92.5	1.6	5.9	56.3	93.3	3.1	3.6	52.9
(3.0, 2.0)	96.2	2.8	1.0	35.6	92.8	1.8	5.4	78.8	93.5	3.4	3.1	71.2
(3.5, 1.5)	96.5	3.1	0.4	43.0	92.8	2.3	4.9	120.3	93.6	3.7	2.7	97.9
(4.0, 1.0)	96.0	3.9	0.1	60.3	93.3	2.7	4.0	234.6	93.8	4.1	2.1	181.9
(4.5, 0.5)	95.5	4.4	0.1	117.3	94.0	3.1	2.9	720.0	94.7	4.3	1.0	563.7
(5.0, 0.0)	95.0	5.0	0.0	2545.8	96.8	3.2	0.0	2776.5	95.6	4.4	0.0	2419.1
Shortest	97.0	1.3	1.7	20.7	90.8	0.3	8.9	24.9	93.4	1.3	5.3	25.9
	Last three variables											
	Bottom				Top				Diagonal			
Tails[a]	CCI[b]	CLT[c]	CRT[d]	Width[e]	CCI[b]	CLT[c]	CRT[d]	Width[e]	CCI[b]	CLT[c]	CRT[d]	Width[e]
(0.0, 5.0)	96.0	0.0	4.0	32.8	96.9	0.0	3.1	31.0	99.0	0.0	1.0	26.7
(0.5, 4.5)	96.6	0.1	3.3	31.7	97.1	0.1	2.8	30.2	98.8	0.3	0.9	26.4
(1.0, 4.0)	96.7	0.3	3.0	31.6	97.2	0.3	2.5	30.6	98.9	0.6	0.5	26.8
(1.5, 3.5)	97.0	0.5	2.5	31.7	97.0	0.8	2.2	31.0	98.5	1.1	0.4	27.3
(2.0, 3.0)	96.5	1.5	2.0	32.4	96.8	1.3	1.9	31.3	97.6	2.0	0.4	28.1
(2.5, 2.5)	96.1	2.1	1.8	33.5	96.7	1.7	1.6	32.5	97.4	2.2	0.4	29.6
(3.0, 2.0)	96.1	2.6	1.3	35.1	97.1	1.9	1.0	34.1	97.0	2.7	0.3	31.2
(3.5, 1.5)	96.7	2.6	0.7	36.6	96.6	2.6	0.8	36.0	96.4	3.4	0.2	33.3
(4.0, 1.0)	96.6	3.0	0.4	39.6	96.3	3.3	0.4	39.8	96.0	4.0	0.0	37.1
(4.5, 0.5)	96.0	3.9	0.1	45.9	96.0	4.0	0.0	46.4	95.8	4.2	0.0	46.0
(5.0, 0.0)	95.8	4.2	0.0	6137.0	95.6	4.4	0.0	5531.1	95.4	4.6	0.0	2901.6
Shortest	94.3	2.0	3.7	30.7	96.0	1.1	2.9	29.3	97.1	1.9	1.0	25.7

[a]The nominal coverage in (left-tail, right-tail).
[b]Coverage of confidence interval.
[c]Coverage of left-tail.
[d]Coverage of right-tail.
[e]Median width of confidence intervals.

tail that add to 5%. Each row of the table gives a different confidence interval, with the nominal coverages in each tail shown in the first column. The actual coverages of confidence intervals are given in the columns headed CCI (*C*overage of *C*onfidence *I*nterval), while the coverages in tails are given in CLT (left tail) and CRT (right tail). The table also gives the median width of intervals for each combination of tail probabilities. The shortest 95% confidence interval was identified for each variable in each of the 1000 samples and results for these shortest intervals are recorded in a separate row. The bank note data set has six variables and results are presented separately for each of these.

The coverages of the confidence intervals are a little low for the variable,'left', and a little high with some intervals for the variable 'diagonal', but otherwise the coverages are reasonably close to the target coverage of 95%. For this fake bank note, the shortest interval was always close to the interval that had 0.0% and 0.5% nominal coverages in the left and right tails, respectively, and far different from the equal-tailed interval. This asymmetry in the nominal tail coverages was also found with the other fake bank notes that were examined and marked asymmetry was also found with most variables in each of the other data sets. Greater nominal coverage in the upper tail thus seems a trait of the shortest confidence interval for individual contributions to an MI. Comparison of the equal-tailed and shortest confidence intervals reveals substantial variation in their relative lengths, with the shortest confidence interval only a little shorter than the equal-tailed interval for some variables (such as 'bottom') and much larger for others – the equal-tailed interval is more than twice the width of the shortest interval for the variables 'left' and 'right'.

Condensed results for all data sets and each sample size are presented in Tables 3 and 4. They give the average coverage across variables for the equal-tailed and shortest 95% confidence intervals. For the bank note data, each average is based on 60 separate coverages, as the 6 variables and 10 fake bank notes gave 60 MDs. Averages are based on 90, 50, 4 and 4 coverages for the Athletes, Skulls, Psychological tests and Flea beetles data, respectively. (With the last two data sets we examined a single difference between two means rather than the MDs for 10 items.) The tables also give a comparison of the width of intervals relative to the width of intervals given by Method A. Specifically, for each MD, the median width of the equal-tailed and shortest 95% confidence intervals were calculated for each bootstrap method and divided by the median width of the corresponding intervals given by Method A. Averages of these ratios are presented in brackets in the tables. (Hence, for example, each average that is given for the Athletes data is based on 90 ratios.) Table 3 gives average coverages and average width ratios for the contributions of variables and Table 4 gives averages for the proportion of the MI attributed to each variable.

To meet the definition of a confidence interval, a method must be conservative rather than liberal, so the coverage of its confidence intervals should preferably be above the nominal value of 95%, rather than below it. In Tables 3 and 4, average coverages that achieve at least the nominal level are given in bold-face type. It is readily seen that Methods A and B (the new methods) almost always achieve their nominal coverage, while the other methods do not. The obvious question is whether the new methods achieve their higher coverage at the expense of giving wider confidence intervals. Looking at Table 3, the bias-corrected percentile method does typically give narrower intervals than Method A; its width ratios (in brackets) are almost always less than or equal to 1.00. However, its average coverages are too far below 95% for the method to be preferred to alternatives. The other methods typically give wider intervals than method A – much wider in the cases of the pivotal methods,

Table 3. Average coverage (%) of 95% confidence intervals for individual contributions given by the percentile, bias-corrected percentile, non-studentized and studentized pivotal methods, and Methods A and B.

Sample size	Data set	Percentile method	Bias-corrected percentile	Non-student. pivotal	Studentized pivotal	Method A	Method B
Equal tailed intervals							
20	Bank notes	80.9 (2.01)	90.7 (0.84)	87.9 (3.24)	92.7 (1.63)	**97.7** (1.00)	94.4 (0.68)
20	Athletes	89.1 (1.69)	91.8 (0.76)	86.2 (5.24)	**95.4** (3.28)	**99.3** (1.00)	**99.2** (1.97)
20	Skulls	85.4 (1.70)	91.0 (0.94)	87.3 (4.62)	91.6 (2.28)	**97.7** (1.00)	94.1 (0.80)
20	Psychol. tests	91.7 (1.24)	91.2 (0.98)	82.9 (7.49)	92.7 (4.80)	**97.2** (1.00)	**96.5** (1.06)
20	Flea beetles	88.7 (1.30)	94.6 (1.08)	94.9 (1.66)	**96.1** (1.05)	**97.0** (1.00)	**95.9** (1.05)
50	Bank notes	88.3 (1.23)	91.5 (0.98)	87.5 (5.86)	92.7 (3.89)	**96.1** (1.00)	**95.1** (1.02)
50	Athletes	92.1 (1.27)	94.0 (0.99)	88.8 (6.43)	**95.6** (3.70)	**98.6** (1.00)	**97.8** (0.99)
50	Skulls	90.8 (1.18)	91.8 (1.00)	87.6 (5.55)	93.9 (3.77)	**96.7** (1.00)	**96.2** (1.03)
50	Psychol. tests	94.3 (1.09)	92.1 (0.97)	85.6 (6.47)	93.5 (4.63)	**96.2** (1.00)	**96.7** (1.07)
50	Flea beetles	92.6 (1.09)	94.9 (1.02)	**95.2** (1.16)	93.5 (1.00)	**95.0** (1.00)	**95.7** (1.06)
80	Bank notes	91.0 (1.13)	92.2 (0.98)	88.8 (5.42)	94.3 (3.69)	**95.9** (1.00)	**96.2** (1.04)
80	Athletes	93.3 (1.15)	93.8 (1.00)	89.6 (5.51)	**95.2** (3.19)	**97.3** (1.00)	**97.0** (1.01)
80	Skulls	92.0 (1.11)	92.1 (1.00)	88.3 (5.19)	94.3 (3.50)	**95.8** (1.00)	**96.4** (1.06)
80	Psychol. tests	94.5 (1.05)	92.7 (0.98)	87.9 (6.51)	93.7 (5.14)	**96.0** (1.00)	**96.6** (1.05)
80	Flea beetles	93.3 (1.06)	94.9 (1.01)	**95.2** (1.09)	**95.1** (1.03)	94.8 (1.00)	**95.9** (1.06)
100	Bank notes	91.5 (1.10)	92.1 (0.99)	88.9 (5.22)	93.9 (3.68)	**95.7** (1.00)	**95.7** (1.05)
100	Athletes	93.5 (1.12)	94.1 (1.00)	89.8 (4.96)	**95.1** (2.96)	**96.9** (1.00)	**96.6** (1.02)
100	Skulls	92.9 (1.09)	92.8 (1.00)	89.1 (4.89)	94.7 (3.31)	**96.0** (1.00)	**96.6** (1.05)
100	Psychol. tests	**95.2** (1.04)	92.0 (0.97)	88.7 (6.00)	94.3 (5.45)	**96.0** (1.00)	**96.7** (1.07)
100	Flea beetles	93.5 (1.04)	94.2 (1.01)	94.8 (1.07)	**95.3** (1.04)	94.7 (1.00)	**95.9** (1.06)
200	Bank notes	93.6 (1.05)	93.5 (0.99)	90.7 (4.35)	94.7 (3.00)	**95.9** (1.00)	**95.8** (1.06)
200	Athletes	94.0 (1.05)	94.1 (1.00)	90.7 (3.66)	94.9 (2.19)	**95.5** (1.00)	**96.1** (1.04)
200	Skulls	94.0 (1.04)	93.5 (0.99)	90.8 (3.63)	94.9 (2.48)	**95.5** (1.00)	**96.3** (1.06)
200	Psychol. tests	**95.2** (1.02)	91.9 (0.99)	89.7 (6.19)	94.2 (4.67)	**95.9** (1.00)	**96.8** (1.07)
200	Flea beetles	94.6 (1.02)	**95.0** (1.00)	**95.5** (1.03)	**96.1** (1.05)	94.7 (1.00)	**96.3** (1.06)
Shortest intervals							
20	Bank notes	92.0 (1.91)	89.4 (0.81)	94.3 (1.80)	92.8 (1.16)	**98.9** (1.00)	**96.1** (0.72)
20	Athletes	**96.8** (1.68)	89.8 (0.75)	92.4 (2.31)	93.4 (1.86)	**99.7** (1.00)	**99.9** (1.80)
20	Skulls	93.8 (1.64)	89.6 (0.91)	93.8 (2.33)	92.3 (1.49)	**98.3** (1.00)	**95.3** (0.83)
20	Psychol. tests	**95.3** (1.21)	90.3 (0.94)	92.3 (3.18)	92.3 (2.20)	**96.7** (1.00)	**96.8** (1.04)
20	Flea beetles	91.4 (1.27)	94.0 (1.05)	**97.4** (1.38)	**96.2** (1.01)	94.5 (1.00)	94.0 (1.04)
50	Bank notes	93.3 (1.20)	91.0 (0.94)	93.0 (2.67)	92.5 (1.94)	**96.6** (1.00)	**96.2** (1.01)
50	Athletes	**95.5** (1.27)	92.7 (0.97)	93.9 (2.93)	**95.1** (1.90)	**98.0** (1.00)	**97.4** (0.99)
50	Skulls	93.9 (1.16)	91.4 (0.97)	93.3 (2.61)	93.7 (1.90)	**96.1** (1.00)	**96.1** (1.02)
50	Psychol. tests	**95.2** (1.08)	91.3 (0.95)	92.8 (3.07)	92.7 (2.22)	**95.7** (1.00)	**96.2** (1.06)
50	Flea beetles	93.5 (1.07)	94.2 (1.01)	**96.6** (1.07)	92.2 (1.00)	93.8 (1.00)	94.2 (1.05)
80	Bank notes	93.8 (1.12)	91.7 (0.96)	93.3 (2.54)	94.0 (1.89)	**96.2** (1.00)	**96.6** (1.02)
80	Athletes	**95.1** (1.15)	92.7 (0.99)	93.9 (2.59)	94.6 (1.72)	**96.5** (1.00)	**96.4** (1.00)
80	Skulls	93.9 (1.09)	91.5 (0.98)	93.3 (2.46)	93.9 (1.80)	**95.3** (1.00)	**96.2** (1.04)
80	Psychol. tests	**95.2** (1.04)	91.9 (0.96)	93.4 (2.94)	93.6 (2.25)	**95.9** (1.00)	**96.2** (1.04)
80	Flea beetles	93.9 (1.04)	94.1 (1.00)	**95.3** (1.03)	94.3 (1.02)	94.1 (1.00)	**95.4** (1.05)
100	Bank notes	93.5 (1.09)	91.8 (0.97)	93.2 (2.48)	93.9 (1.87)	**95.7** (1.00)	**96.0** (1.04)
100	Athletes	94.9 (1.11)	93.2 (0.99)	94.1 (2.40)	94.3 (1.63)	**96.2** (1.00)	**96.1** (1.01)
100	Skulls	94.2 (1.07)	92.3 (0.98)	93.7 (2.35)	94.1 (1.74)	**95.5** (1.00)	**96.3** (1.04)
100	Psychol. tests	**95.1** (1.04)	90.1 (0.97)	92.9 (2.87)	93.5 (2.41)	**95.3** (1.00)	**96.1** (1.07)
100	Flea beetles	93.3 (1.03)	93.9 (1.00)	**95.2** (1.03)	94.5 (1.04)	93.9 (1.00)	**95.3** (1.06)
200	Bank notes	94.6 (1.04)	93.2 (0.98)	93.9 (2.14)	94.6 (1.63)	**95.8** (1.00)	**95.9** (1.05)
200	Athletes	94.4 (1.05)	93.4 (1.00)	94.2 (1.93)	94.5 (1.38)	94.8 (1.00)	**95.3** (1.03)
200	Skulls	94.4 (1.03)	93.0 (0.98)	94.0 (1.93)	94.7 (1.51)	**95.1** (1.00)	**95.9** (1.05)
200	Psychol. tests	**95.5** (1.02)	90.9 (0.99)	92.8 (2.76)	92.3 (2.11)	**95.5** (1.00)	**96.4** (1.06)
200	Flea beetles	94.4 (1.02)	94.1 (1.00)	94.9 (1.01)	**95.6** (1.05)	94.1 (1.00)	**95.8** (1.06)

Average of the ratio of the median widths of intervals relative to the median widths of intervals given by Method A are shown in brackets. Population distributions are multivariate normal.

Table 4. Average coverage (%) of 95% confidence intervals for percentage of contributions given by the percentile, bias-corrected percentile, non-studentized and studentized pivotal methods, and Methods A and B.

Sample size	Data set	Percentile method	Bias-corrected percentile	Non-student pivotal	Studentized pivotal	Method A	Method B
Equal tailed intervals							
20	Bank notes	92.5 (0.54)	88.4 (0.50)	81.3 (1.45)	91.3 (1.22)	**97.6** (1.00)	94.8 (0.65)
20	Athletes	94.2 (0.64)	84.7 (0.62)	80.1 (2.40)	93.9 (2.38)	**97.2** (1.00)	**97.4** (1.15)
20	Skulls	91.4 (0.60)	86.5 (0.57)	82.1 (1.44)	90.7 (1.26)	**96.1** (1.00)	92.8 (0.71)
20	Psychol. tests	93.7 (0.76)	88.1 (0.71)	82.2 (1.98)	91.9 (1.80)	**97.5** (1.00)	**97.0** (0.93)
20	Flea beetles	93.3 (0.77)	92.3 (0.78)	94.5 (1.06)	94.9 (0.84)	**98.2** (1.00)	**96.5** (0.93)
50	Bank notes	93.4 (0.77)	89.8 (0.73)	85.4 (3.06)	92.2 (2.40)	**97.0** (1.00)	**95.7** (0.92)
50	Athletes	94.7 (0.74)	90.7 (0.74)	86.6 (3.25)	94.7 (2.39)	**98.1** (1.00)	**97.2** (0.93)
50	Skulls	93.2 (0.80)	89.9 (0.77)	86.0 (2.49)	93.1 (2.09)	**96.6** (1.00)	**95.8** (0.94)
50	Psychol. tests	94.1 (0.91)	90.4 (0.85)	85.5 (3.05)	92.3 (2.51)	**96.3** (1.00)	**96.7** (1.03)
50	Flea beetles	94.0 (0.91)	93.6 (0.92)	**95.4** (1.03)	92.7 (0.94)	**95.7** (1.00)	**96.0** (1.02)
80	Bank notes	93.7 (0.86)	90.8 (0.84)	87.2 (3.59)	93.7 (2.69)	**96.6** (1.00)	**96.4** (0.99)
80	Athletes	94.5 (0.84)	91.5 (0.84)	88.2 (3.48)	94.4 (2.34)	**97.0** (1.00)	**96.6** (0.96)
80	Skulls	93.3 (0.88)	90.6 (0.86)	87.5 (2.95)	93.6 (2.36)	**95.8** (1.00)	**96.2** (1.01)
80	Psychol. tests	**95.0** (0.94)	91.6 (0.90)	88.1 (3.63)	93.4 (3.02)	**96.1** (1.00)	**96.5** (1.03)
80	Flea beetles	94.5 (0.94)	94.2 (0.95)	**95.3** (1.01)	94.2 (0.99)	**95.8** (1.00)	**95.9** (1.04)
100	Bank notes	93.6 (0.90)	90.9 (0.87)	87.6 (3.71)	93.6 (2.84)	**96.0** (1.00)	**96.1** (1.02)
100	Athletes	94.6 (0.87)	92.2 (0.87)	88.7 (3.41)	94.3 (2.27)	**96.7** (1.00)	**96.3** (0.98)
100	Skulls	93.9 (0.91)	91.6 (0.89)	88.4 (3.03)	93.9 (2.40)	**95.9** (1.00)	**96.3** (1.02)
100	Psychol. tests	**95.2** (0.95)	91.4 (0.91)	88.8 (3.70)	93.5 (3.49)	**96.3** (1.00)	**96.4** (1.04)
100	Flea beetles	93.9 (0.96)	93.5 (0.96)	94.6 (1.01)	**95.4** (1.00)	94.5 (1.00)	**96.7** (1.03)
200	Bank notes	**95.0** (0.95)	93.1 (0.93)	90.2 (3.60)	94.5 (2.62)	**96.2** (1.00)	**96.1** (1.04)
200	Athletes	94.2 (0.94)	92.8 (0.94)	90.3 (2.96)	94.5 (1.97)	**95.3** (1.00)	**95.9** (1.02)
200	Skulls	94.5 (0.95)	93.0 (0.93)	90.6 (2.90)	94.6 (2.13)	**95.5** (1.00)	**96.1** (1.04)
200	Psychol. tests	**95.6** (0.97)	91.9 (0.96)	90.0 (4.40)	94.1 (3.60)	**96.0** (1.00)	**96.8** (1.05)
200	Flea beetles	94.1 (0.98)	94.0 (0.98)	94.8 (1.00)	**95.9** (1.03)	94.5 (1.00)	**96.2** (1.04)
Shortest intervals							
20	Bank notes	93.5 (0.59)	87.0 (0.54)	89.6 (1.26)	91.4 (1.04)	**98.4** (1.00)	**95.6** (0.66)
20	Athletes	93.5 (0.67)	82.7 (0.67)	86.5 (2.11)	92.4 (2.14)	**96.0** (1.00)	**95.9** (1.14)
20	Skulls	91.4 (0.65)	85.0 (0.62)	88.3 (1.34)	90.5 (1.14)	**97.4** (1.00)	93.4 (0.73)
20	Psychol. tests	93.3 (0.80)	86.5 (0.74)	89.4 (1.78)	90.7 (1.49)	**96.9** (1.00)	**96.4** (0.94)
20	Flea beetles	90.6 (0.80)	89.9 (0.81)	**95.1** (1.03)	93.4 (0.86)	**96.3** (1.00)	94.3 (0.95)
50	Bank notes	94.0 (0.80)	89.2 (0.76)	91.2 (2.06)	91.6 (1.59)	**97.9** (1.00)	**96.4** (0.93)
50	Athletes	93.9 (0.79)	89.4 (0.79)	92.0 (2.31)	94.3 (1.67)	**97.7** (1.00)	**96.7** (0.94)
50	Skulls	93.0 (0.83)	88.9 (0.80)	91.5 (1.88)	92.4 (1.52)	**96.5** (1.00)	**95.7** (0.95)
50	Psychol. tests	93.8 (0.92)	89.4 (0.86)	91.6 (2.24)	91.0 (1.83)	**95.7** (1.00)	**95.7** (1.03)
50	Flea beetles	92.4 (0.91)	92.5 (0.92)	**95.5** (0.98)	91.2 (0.94)	94.7 (1.00)	94.6 (1.02)
80	Bank notes	93.9 (0.88)	90.2 (0.85)	91.8 (2.18)	93.4 (1.68)	**97.0** (1.00)	**96.7** (0.99)
80	Athletes	93.5 (0.86)	90.4 (0.86)	92.8 (2.24)	93.9 (1.56)	**96.5** (1.00)	**95.9** (0.97)
80	Skulls	93.1 (0.89)	89.7 (0.87)	92.0 (2.00)	92.8 (1.57)	**95.5** (1.00)	**95.9** (1.01)
80	Psychol. tests	94.9 (0.95)	90.6 (0.91)	92.7 (2.43)	92.9 (1.95)	**95.9** (1.00)	**95.9** (1.02)
80	Flea beetles	93.8 (0.95)	93.6 (0.95)	**95.3** (0.98)	93.2 (0.99)	**95.0** (1.00)	**95.4** (1.04)
100	Bank notes	93.6 (0.91)	90.2 (0.88)	92.0 (2.20)	93.5 (1.72)	**95.9** (1.00)	**96.2** (1.01)
100	Athletes	93.9 (0.89)	91.3 (0.89)	93.2 (2.14)	93.7 (1.51)	**96.1** (1.00)	**95.7** (0.98)
100	Skulls	93.5 (0.92)	90.8 (0.89)	92.6 (1.98)	93.3 (1.56)	**95.6** (1.00)	**96.1** (1.01)
100	Psychol. tests	94.9 (0.96)	90.0 (0.92)	92.4 (2.44)	92.6 (2.07)	**96.0** (1.00)	**95.9** (1.04)
100	Flea beetles	93.2 (0.95)	92.8 (0.96)	94.5 (0.99)	94.7 (1.00)	94.0 (1.00)	**96.0** (1.03)
200	Bank notes	94.8 (0.96)	92.5 (0.94)	93.5 (2.01)	94.4 (1.57)	**96.0** (1.00)	**96.0** (1.04)
200	Athletes	93.6 (0.94)	92.2 (0.94)	93.7 (1.83)	94.3 (1.34)	94.8 (1.00)	**95.2** (1.02)
200	Skulls	94.0 (0.96)	92.2 (0.94)	93.6 (1.79)	94.1 (1.44)	**95.0** (1.00)	**95.8** (1.03)
200	Psychol. tests	**95.1** (0.98)	90.6 (0.96)	92.7 (2.48)	92.6 (2.00)	**95.7** (1.00)	**96.4** (1.05)
200	Flea beetles	93.5 (0.98)	93.5 (0.98)	94.0 (0.99)	**95.5** (1.03)	94.0 (1.00)	**95.6** (1.05)

Average of the ratio of the median widths of intervals relative to the median widths of intervals given by Method A are shown in brackets. Population distributions are multivariate normal.

but also slightly wider with both the percentile method and Method B. Hence, Method A clearly has better results than other methods in Table 3, with Method B a close second.

While the pattern of average coverages in Table 4 is similar to that in Table 3, its pattern of width ratios is a little different. In Table 4, the pivotal methods still typically give much wider intervals than the new methods and the bias-corrected method (which gives poor coverage) still has the narrowest intervals, but now the percentile method gives slightly narrower intervals than the new methods, and differences between the widths of the new methods favour Method A less consistently. Nevertheless, taking both coverage and interval width into account, Method A is again the best method, with Method B a very close second.

Theoretical results about asymptotic properties have not been derived for the new methods (A and B) of forming confidence intervals. To explore their behaviour for larger sample sizes, further simulations were conducted with these methods that were identical to those reported above, but for sample sizes of 500 and 1000. Results are presented in Table 5. It can be seen that the coverage is always close to the nominal level of 95%, especially for the larger sample size of 1000. These results suggest that coverage will tend to 95% as sample size increases. The coverages for Method A are generally closer to 95% than those of Method B, albeit by marginal amounts.

As Method A seems the best method, we examined more closely how it might be used. Specifically, we compared the shortest confidence intervals that it gave with the equal-tailed intervals it gave. For each MD, the width of the equal-tailed 95% interval was divided by the width of the shortest 95% confidence intervals. The frequency distributions of these ratios differed substantially depending upon whether or not the shortest confidence interval was a one-sided interval (with 0 as its lower endpoint). Table 6 gives the relative frequency distributions when the shortest interval is *not* one-sided. It can be seen the equal-tailed interval is generally only slightly longer than the shortest interval, especially when the contribution of each variable (rather than percentage contribution) is the quantity of interests, when the equal-tailed interval is seldom more than 5% longer than the shortest interval. As people are unfamiliar with interpreting confidence intervals that are neither equal-tailed

Table 5. Average coverage (%) of 95% equal-tailed and shortest confidence intervals for Method A and Method B, for sample sizes of 500 and 1000.

	Equal tailed intervals				Shortest intervals			
	Method A		Method B		Method A		Method B	
Data set	500	1000	500	1000	500	1000	500	1000
Contribution of individual variables								
Bank notes	95.5	95.3	96.3	96.2	95.2	94.9	96.0	95.9
Athletes	95.6	95.2	96.5	96.1	95.1	94.7	95.8	95.6
Skulls	95.5	95.3	96.4	96.1	95.0	94.9	96.0	95.8
Psychol. tests	95.5	95.3	96.4	96.5	95.4	94.1	96.1	95.2
Flea beetles	95.5	94.6	96.4	95.9	94.9	94.2	95.9	95.7
Percentage contribution of variables								
Bank notes	95.7	95.0	96.6	95.9	95.4	94.6	96.3	95.6
Athletes	95.7	95.3	96.4	96.2	95.0	94.7	95.8	95.7
Skulls	95.7	95.2	96.5	96.1	95.2	94.8	96.1	95.8
Psychol. tests	95.2	95.1	96.2	96.3	94.9	93.9	95.8	95.1
Flea beetles	95.7	95.1	96.4	95.9	95.3	94.4	96.1	95.5

Table 6. Relative frequency distribution (%) for the width ratio of confidence intervals given by equal-tailed interval relative to shortest interval using Method *A* for shortest intervals that are not one-sided.

		Width ratio						
Sample size	Data set	1.0–1.05	1.05–1.1	1.1–1.25	1.25–1.5	1.5–1.75	1.75–2.0	2 and over
Contribution of individual variables								
20	Bank notes	97.62	2.38	0.00	0.00	0.00	0.00	0.00
	Athletes	87.80	4.88	0.00	7.32	0.00	0.00	0.00
	Skulls	97.84	2.08	0.09	0.00	0.00	0.00	0.00
	Psychol. tests	95.21	4.50	0.29	0.00	0.00	0.00	0.00
	Flea beetles	93.06	6.53	0.41	0.00	0.00	0.00	0.00
100	Bank notes	96.61	3.22	0.17	0.00	0.00	0.00	0.00
	Athletes	88.86	10.55	0.59	0.00	0.00	0.00	0.00
	Skulls	94.36	5.41	0.22	0.00	0.00	0.00	0.00
	Psychol. tests	91.52	8.12	0.35	0.00	0.00	0.00	0.00
	Flea beetles	99.10	0.90	0.00	0.00	0.00	0.00	0.00
Percentage contribution of variables								
20	Bank notes	86.64	11.82	1.54	0.00	0.00	0.00	0.00
	Athletes	93.32	6.11	0.57	0.00	0.00	0.00	0.00
	Skulls	54.81	22.13	21.31	1.75	0.00	0.00	0.00
	Psychol. tests	71.77	18.02	9.67	0.54	0.00	0.00	0.00
	Flea beetles	86.14	12.21	1.60	0.05	0.00	0.00	0.00
100	Bank notes	93.66	5.80	0.54	0.00	0.00	0.00	0.00
	Athletes	81.94	16.33	1.73	0.00	0.00	0.00	0.00
	Skulls	89.15	9.22	1.60	0.03	0.00	0.00	0.00
	Psychol. tests	93.24	6.52	0.23	0.00	0.00	0.00	0.00
	Flea beetles	98.68	1.33	0.00	0.00	0.00	0.00	0.00

Table 7. Relative frequency distribution (%) for the width ratio of confidence intervals given by equal-tailed interval relative to shortest interval using Method *A* for shortest intervals that are one-sided.

		Width ratio						
Sample size	Data set	1.0–1.05	1.05–1.1	1.1–1.25	1.25–1.5	1.5–1.75	1.75–2.0	2 and over
Contribution of individual variables								
20	Bank notes	2.69	12.94	30.48	31.53	16.88	4.36	1.13
	Athletes	0.41	3.54	22.76	51.00	18.66	2.92	0.70
	Skulls	4.42	11.47	36.39	35.55	10.33	1.53	0.31
	Psychol. tests	4.53	11.62	46.94	35.93	0.97	0.00	0.00
	Flea beetles	13.39	32.70	50.87	3.04	0.00	0.00	0.00
100	Bank notes	1.86	9.38	58.25	30.33	0.17	0.00	0.00
	Athletes	3.95	15.33	58.67	21.89	0.16	0.00	0.00
	Skulls	2.66	10.85	57.27	29.01	0.21	0.00	0.00
	Psychol. tests	1.86	9.71	54.75	33.61	0.07	0.00	0.00
	Flea beetles*	–	–	–	–	–	–	–
Percentage contribution of variables								
20	Bank notes	7.59	13.96	35.59	35.72	6.64	0.48	0.03
	Athletes	4.49	7.34	44.83	40.48	2.74	0.10	0.01
	Skulls	6.40	13.35	40.93	33.07	5.63	0.55	0.07
	Psychol. tests	3.66	8.87	41.82	41.89	3.73	0.04	0.00
	Flea beetles	9.97	21.95	56.15	11.67	0.26	0.00	0.00
100	Bank notes	1.51	7.62	53.30	36.84	0.74	0.00	0.00
	Athletes	2.87	12.24	57.31	27.06	0.51	0.01	0.00
	Skulls	2.02	8.81	53.55	34.95	0.68	0.00	0.00
	Psychol. tests	2.67	8.70	51.58	36.42	0.63	0.00	0.00
	Flea beetles*	–	–	–	–	–	–	–

* Shortest confidence intervals for the flea beetles data were never one-sided for this sample size.

nor one-sided, there will seldom be much justification for presenting a shortest confidence interval if it is not one-sided.

Table 7 gives the relative frequency distributions when the shortest interval *is* one-sided. Now the equal-tailed interval is 10–50% wider than the one-sided interval in most cases. We earlier noted that a one-sided confidence interval for a squared quantity is attractive when 0 is contained in an equal-tailed interval for the un-squared quantity, which typically happens when the shortest interval is one-sided. Hence, when the shortest confidence interval is one-sided, there seems good reason to report it in preference to an equal-tailed interval.

Tables 6 and 7 only present results for sample sizes of 20 and 100. Results for sample sizes of 50, 80 and 200 were also produced but, for brevity, are not presented here because results did not vary appreciably with sample size.

5. Simulation study: skew distributions

In the last section, we compared transformation and examined their performance when the underlying population distributions were multivariate normal. Here we extend that work and examine the sensitivity of its results to departures from normal distributions. Specifically, we take the same population distributions as before and use the sinh–arcsinh transformation [24] to construct skew population distributions that retain features of the original distributions – each variable keeps its mean and variance and the correlation structure is broadly similar. Details are given in the next subsections.

5.1. Simulation procedure for skew distributions: bank note data

We will refer to the bank note data constructed here as the skew-genuine and skew-fake bank notes, to distinguish them from the genuine and fake bank notes from which we start. As before, we calculated the mean and covariance of the 100 genuine bank note measurements and took these as the mean ($\boldsymbol{\mu}$) and variance ($\boldsymbol{\Sigma}$) of an MVN distribution. We then generated 100,000 observations from an MVN($\mathbf{0}, \boldsymbol{\Sigma}$) distribution. Denote these observations as $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{100\,000}$ and put $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{im})^{\mathrm{T}}$. The sinh–arcsinh was then applied separately to each component of each $\boldsymbol{y}$, putting

$$y^{\#}_{i,j} = \sinh[\delta^{-1}\{\sinh^{-1}(y_{ij}) + \epsilon\}] \tag{20}$$

for $i = 1, \ldots, 100\,000$; $j = 1, \ldots, m$.

Let $Y^{\#}_j$ denote a scalar variable whose sample values are $y^{\#}_{1,j}, y^{\#}_{2,j}, \ldots, y^{\#}_{100\,000,j}$. The parameters ϵ and δ in Equation (20) respectively affect the skewness and tailweight of the distribution of $Y^{\#}_j$. For the bank note data, we set $\epsilon = 1.0$ and $\delta = 0.8$. Let $\bar{y}^{\#}_j$ and $s(y^{\#}_j)$ denote the sample mean and sample standard deviation of $Y^{\#}_j$. Also, let $\mu_{(j)}$ denote the jth component of $\boldsymbol{\mu}$ and $\sigma^2_{(j)}$ denote the jth diagonal element of $\boldsymbol{\Sigma}$. For $i = 1, \ldots, 100\,000$; $j = 1, \ldots, m$ put

$$x^{\#}_{i,j} = \mu_{(j)} + \sigma_{(j)}\{y^{\#}_{i,j} - \bar{y}^{\#}_j\}/s(y^{\#}_j) \tag{21}$$

and $\boldsymbol{x}^{\#}_i = (x^{\#}_{i,1}, \ldots, x^{\#}_{i,m})^{\mathrm{T}}$.

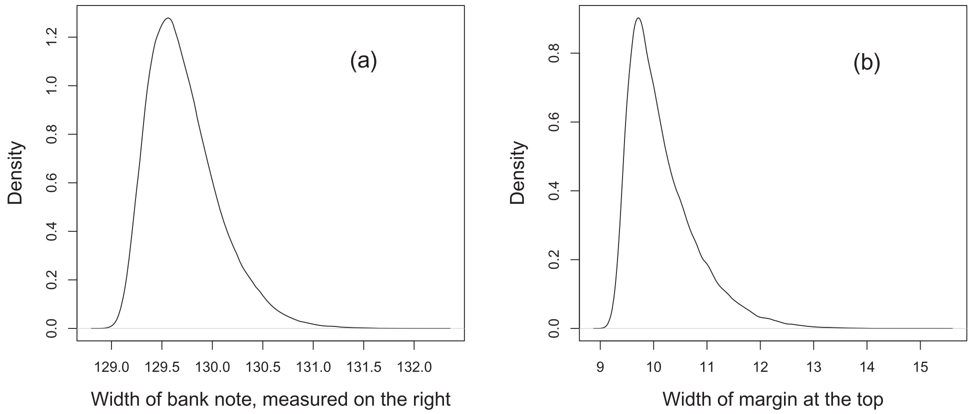

Figure 2. Probability density functions with coefficients of skewness of (a) 0.981 and (b) 1.410.

We suppose that the complete population of skew-genuine bank notes consists of 100 000 notes and that $\boldsymbol{x}_i^{\#}$ is the vector of measurements on the ith note ($i = 1, \ldots, 100\,000$). There are only two differences between the simulation method used now and the simulation method used in Section 4.

(1) In Section 4, sample data sets were generated from MVN$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Here, $\boldsymbol{x}_1^{\#}, \ldots, \boldsymbol{x}_{100\,000}^{\#}$ are treated as the population and a sample data set is generated by sampling without replacement from $\boldsymbol{x}_1^{\#}, \ldots, \boldsymbol{x}_{100\,000}^{\#}$.
(2) The distribution of the skew-fake bank notes should be similar in shape to that of the skew-genuine bank notes, but with a different mean. Let $\boldsymbol{\mu}^{\diamond} = (\mu_{(1)}^{\diamond}, \ldots, \mu_{(m)}^{\diamond})^{\mathrm{T}}$ denote the sample mean of the 100 fake bank notes and let $\boldsymbol{t} = (t_1, \ldots, t_m)^{\mathrm{T}}$ denote the deviations from this mean for one fake bank note. Analogous to Equations (20) and (21), for $j = 1, \ldots, m$ put

$$t_j^{\#} = \sinh[\delta^{-1}\{\sinh^{-1}(t_j) + \epsilon\}] \tag{22}$$

and

$$v_j^{\#} = \mu_{(j)}^{\diamond} + \sigma_{(j)}\{t_j^{\#} - \bar{y}_j^{\#}\}/s(y_j^{\#}) \tag{23}$$

where $\sigma_{(j)}, \bar{y}_j^{\#}$ and $s(y_j^{\#})$ take the same values as in Equation (21). We take $(v_1^{\#}, \ldots, v_m^{\#})^{\mathrm{T}}$ as the vector of values of the skew-fake bank note. In the simulations, a skew-fake bank note is constructed from each of the first 10 fake bank notes.

For the population of skew-genuine bank notes, the Pearson's moment coefficients of skewness for the six measurements were 1.016, 1.020, 0.981, 1.384, 1.410 and 1.150. Figures 2(a) and 2(b) show the marginal probability density functions (*p.d.f.*s) of the third and fifth measurements. It can be seen that the *p.d.f.*s have clear skewness.

5.2. Simulation procedure for skew distributions: other data sets

The simulation procedure used to construct skew distributions for the bank note data set was also used for the Athletes data and the Tibetan skull data. As in the earlier study, the

Table 8. Parameters ϵ and δ and the Pearson's moment coefficient of skewness for each variable in each data set.

Data set	ϵ	δ	Skewness (kurtosis)
Bank notes	1.0	0.8	1.02 (4.44), 1.02 (4.56), 0.98 (4.37), 1.38 (5.41), 1.41 (5.62), 1.15 (4.84)
Athletes	0.8	0.9	1.57 (5.50), 1.57 (5.55), 0.73 (3.76), 1.24 (4.82), 1.48 (5.30), 1.44 (5.19), 1.56 (5.36), 1.52 (5.43), 1.58 (5.53)
Skulls	0.5	0.9	1.24 (4.67), 1.24 (4.71), 1.22 (4.65), 1.22 (4.69), 1.24 (4.73)
Psychol. tests	0.4	0.8	1.29 (5.53), 1.33 (5.46), 1.35 (5.61), 1.33 (5.54)
Flea beetles	0.6	0.9	1.40 (5.11), 1.39 (5.03), 1.38 (5.04), 1.39 (5.01)

The Pearson's moment coefficient of kurtosis for each variable is given in parentheses.

male athletes or the skulls from the Sikkim area took the role of the genuine bank notes, while the first 10 female athletes or the first 10 skulls from the Lhasa district took the role of the fake bank notes.

For the Psychological test and Flea beetle data sets, in which two means are compared, $\mathbf{\Sigma}$ was set equal to the pooled sample covariance matrix and two populations of 100 000 skew-data were constructed. For the Psychology test, for one population $\mu_{(j)}$ in Equation (21) was set equal to the jth component of the sample mean for men and for the other population it was obtained from the sample mean for women. For the Tibetan skulls, $\mu_{(j)}$ was taken as the jth component of either the sample mean for skulls from the Sikkim area (for one population) or the sample mean for skulls from the Lhasa district (for the other population).

In applying the sinh–arcsinh transformation, the parameters ϵ and δ were varied across our five data sets so as to vary the degree of skewness and thickness of tails. Table 8 shows the values chosen for ϵ and δ and lists the Pearson's moment coefficients of skewness and kurtosis for each variable in the data sets. (The Pearson's moment coefficient of kurtosis is 3 for a normal distribution and larger for heavier tails.)

5.3. Results: skew distributions

Tables 9 and 10 give results for the skew population distributions that are equivalent to results presented in Tables 3 and 4 for the MVN population distributions. Table 9 gives average coverages of intervals for the contributions of individual variables and hence corresponds to Table 3. Table 10 gives similar results for the proportion of the MI attributable to each variable and so corresponds to Table 4. As before, the median widths of intervals given by each method were compared with the median widths of intervals given by Method A. The average of these ratios are given in brackets in the tables. Average coverages above the nominal level of 95% are again shown in bold-face type.

The motivation for these simulations was to examine the robustness of results to departures from normality in the population distributions. Hence we focus on comparing Tables 9 and 10 with Tables 3 and 4. Regarding average coverage, the same main features in Tables 3 and 4 were also found in Tables 9 and 10. Specifically:

- Methods A and B almost always achieve the nominal coverage.
- With other methods, the average coverage is typically below the nominal level.
- In particular, the average coverage of the bias-corrected percentile method is often well below the nominal level.

Table 9. Average coverage (%) of 95% confidence intervals for individual contributions given by the percentile, bias-corrected percentile, non-studentized and studentized pivotal methods, and Methods A and B.

Sample size	Data set	Percentile method	Bias-corrected percentile	Non-student pivotal	Studentized pivotal	Method A	Method B
Equal tailed intervals							
20	Bank notes	78.0 (2.11)	89.4 (0.80)	85.6 (2.87)	90.2 (1.38)	**96.8** (1.00)	92.8 (0.63)
20	Athletes	85.9 (1.76)	90.5 (0.73)	84.6 (4.38)	**95.3** (2.76)	**97.7** (1.00)	**97.7** (2.16)
20	Skulls	84.1 (1.90)	89.1 (0.90)	84.1 (3.92)	90.7 (1.93)	**98.6** (1.00)	94.9 (0.80)
20	Psychol. tests	88.8 (1.34)	89.7 (0.99)	78.5 (7.24)	91.6 (4.66)	**96.3** (1.00)	**95.6** (1.10)
20	Flea beetles	84.8 (1.47)	92.9 (1.15)	92.7 (1.81)	**95.3** (1.14)	**95.9** (1.00)	**95.6** (1.09)
50	Bank notes	85.6 (1.30)	89.6 (0.99)	84.4 (5.56)	91.7 (3.55)	**95.9** (1.00)	94.5 (1.04)
50	Athletes	89.0 (1.37)	93.0 (1.03)	88.0 (5.43)	94.7 (2.85)	**97.2** (1.00)	**96.1** (1.03)
50	Skulls	89.2 (1.30)	89.9 (1.04)	85.7 (5.81)	93.2 (3.66)	**96.5** (1.00)	**95.9** (1.11)
50	Psychol. tests	92.6 (1.13)	91.2 (0.96)	82.6 (7.00)	93.4 (4.86)	**95.6** (1.00)	**96.6** (1.10)
50	Flea beetles	89.7 (1.14)	93.4 (1.04)	93.0 (1.20)	93.9 (1.01)	94.6 (1.00)	**95.4** (1.08)
80	Bank notes	89.2 (1.18)	90.9 (0.99)	86.0 (5.51)	93.0 (3.68)	**95.4** (1.00)	**95.4** (1.08)
80	Athletes	90.1 (1.21)	93.2 (1.02)	88.8 (4.17)	94.5 (2.49)	**95.9** (1.00)	**95.3** (1.04)
80	Skulls	91.0 (1.18)	90.9 (1.03)	87.1 (5.42)	93.7 (3.54)	**95.9** (1.00)	**96.1** (1.11)
80	Psychol. tests	93.9 (1.07)	92.0 (0.96)	85.8 (6.31)	93.1 (4.82)	**95.6** (1.00)	**96.4** (1.10)
80	Flea beetles	91.9 (1.08)	94.4 (1.02)	93.8 (1.11)	94.6 (1.04)	94.8 (1.00)	**95.9** (1.07)
100	Bank notes	90.3 (1.14)	91.4 (0.99)	86.9 (5.17)	93.6 (3.59)	**95.5** (1.00)	**95.8** (1.09)
100	Athletes	90.6 (1.16)	93.3 (1.01)	89.3 (3.78)	94.6 (2.36)	**95.4** (1.00)	**95.2** (1.05)
100	Skulls	92.3 (1.14)	91.9 (1.02)	88.7 (5.15)	94.5 (3.59)	**95.9** (1.00)	**96.5** (1.11)
100	Psychol. tests	94.3 (1.05)	92.6 (0.95)	87.5 (5.63)	94.1 (4.84)	**95.3** (1.00)	**96.7** (1.09)
100	Flea beetles	93.1 (1.06)	94.5 (1.02)	94.4 (1.09)	**95.2** (1.05)	94.8 (1.00)	**96.0** (1.08)
200	Bank notes	92.2 (1.07)	92.1 (0.99)	88.7 (4.37)	94.2 (3.07)	94.9 (1.00)	**95.9** (1.09)
200	Athletes	92.1 (1.07)	93.5 (1.00)	90.8 (2.93)	94.3 (2.17)	94.3 (1.00)	**95.0** (1.06)
200	Skulls	93.9 (1.07)	92.7 (1.01)	90.5 (4.42)	94.8 (3.05)	**95.6** (1.00)	**96.5** (1.09)
200	Psychol. tests	**95.3** (1.03)	91.6 (0.96)	88.0 (5.71)	93.7 (4.58)	**95.3** (1.00)	**96.6** (1.08)
200	Flea beetles	94.2 (1.03)	**95.3** (1.01)	**95.0** (1.04)	**96.1** (1.05)	**95.7** (1.00)	**96.5** (1.07)
Shortest intervals							
20	Bank notes	90.4 (2.04)	88.3 (0.79)	93.7 (1.63)	91.3 (1.04)	**99.2** (1.00)	**95.5** (0.68)
20	Athletes	**96.4** (1.75)	89.5 (0.73)	92.3 (2.02)	93.6 (1.66)	**99.8** (1.00)	**99.9** (1.97)
20	Skulls	93.2 (1.82)	87.8 (0.87)	91.6 (2.04)	91.0 (1.33)	**98.6** (1.00)	**96.2** (0.83)
20	Psychol. tests	93.3 (1.31)	89.5 (0.95)	91.8 (3.16)	92.5 (2.27)	**96.1** (1.00)	**96.3** (1.07)
20	Flea beetles	89.5 (1.43)	92.2 (1.10)	**95.6** (1.49)	**95.4** (1.10)	94.1 (1.00)	94.2 (1.06)
50	Bank notes	90.9 (1.27)	89.6 (0.94)	91.6 (2.56)	92.0 (1.87)	**95.8** (1.00)	**95.3** (1.02)
50	Athletes	94.3 (1.36)	92.7 (1.00)	93.9 (2.62)	**95.3** (1.71)	**97.6** (1.00)	**97.0** (1.02)
50	Skulls	92.9 (1.26)	89.8 (1.00)	91.9 (2.75)	92.9 (1.96)	**95.5** (1.00)	**95.7** (1.08)
50	Psychol. tests	93.8 (1.10)	90.5 (0.93)	91.3 (2.96)	92.1 (2.29)	94.9 (1.00)	**96.2** (1.08)
50	Flea beetles	91.6 (1.11)	93.0 (1.02)	**95.1** (1.09)	92.8 (1.00)	93.1 (1.00)	94.2 (1.06)
80	Bank notes	92.6 (1.16)	90.9 (0.96)	92.0 (2.55)	93.0 (1.91)	**95.3** (1.00)	**95.7** (1.06)
80	Athletes	93.8 (1.19)	93.1 (0.99)	94.2 (2.21)	94.9 (1.52)	**95.8** (1.00)	**95.7** (1.02)
80	Skulls	93.3 (1.15)	90.8 (1.00)	92.5 (2.61)	93.4 (1.88)	**95.0** (1.00)	**95.8** (1.08)
80	Psychol. tests	94.6 (1.07)	91.3 (0.95)	92.9 (2.92)	92.9 (2.25)	**95.1** (1.00)	**96.0** (1.09)
80	Flea beetles	92.7 (1.06)	93.7 (1.01)	**95.0** (1.04)	93.6 (1.03)	93.8 (1.00)	94.7 (1.07)
100	Bank notes	93.1 (1.12)	91.4 (0.97)	92.4 (2.46)	93.6 (1.88)	**95.3** (1.00)	**95.9** (1.07)
100	Athletes	93.6 (1.14)	93.1 (0.99)	94.5 (2.05)	94.8 (1.46)	**95.4** (1.00)	**95.5** (1.03)
100	Skulls	94.1 (1.12)	91.4 (0.99)	93.1 (2.49)	93.9 (1.87)	**95.1** (1.00)	**96.2** (1.09)
100	Psychol. tests	94.9 (1.05)	91.4 (0.93)	93.4 (2.66)	93.5 (2.13)	94.7 (1.00)	**95.9** (1.08)
100	Flea beetles	93.5 (1.05)	94.1 (1.01)	**95.0** (1.03)	94.7 (1.04)	93.7 (1.00)	**95.3** (1.07)
200	Bank notes	93.3 (1.06)	91.8 (0.98)	93.0 (2.20)	94.0 (1.70)	94.7 (1.00)	**95.7** (1.07)
200	Athletes	93.8 (1.06)	93.3 (0.99)	94.7 (1.70)	94.0 (1.36)	94.6 (1.00)	**95.3** (1.05)
200	Skulls	94.4 (1.06)	92.1 (0.99)	93.5 (2.22)	94.2 (1.68)	94.8 (1.00)	**96.1** (1.08)
200	Psychol. tests	**95.2** (1.02)	90.2 (0.95)	92.3 (2.69)	92.1 (2.09)	**95.1** (1.00)	**96.1** (1.07)
200	Flea beetles	94.5 (1.02)	**95.0** (1.00)	**95.5** (1.01)	**95.9** (1.05)	**95.1** (1.00)	**96.2** (1.06)

Average of the ratio of the median widths of intervals relative to the median widths of intervals given by Method A are shown in brackets. Population distributions are skew.

Table 10. Average coverage (%) of 95% confidence intervals for percentage of contributions given by the percentile, bias-corrected percentile, non-studentized and studentized pivotal methods, and Methods A and B.

Sample size	Data set	Percentile method	Bias-corrected percentile	Non-student pivotal	Studentized pivotal	Method A	Method B
Equal tailed intervals							
20	Bank notes	92.6 (0.58)	86.5 (0.53)	79.0 (1.48)	88.8 (1.26)	**96.6** (1.00)	93.2 (0.68)
20	Athletes	93.2 (0.68)	83.2 (0.68)	79.8 (2.45)	93.0 (2.48)	**96.5** (1.00)	**96.9** (1.14)
20	Skulls	90.8 (0.63)	83.2 (0.58)	79.4 (1.42)	89.2 (1.30)	**96.0** (1.00)	92.9 (0.81)
20	Psychol. tests	93.3 (0.70)	86.6 (0.64)	79.2 (1.79)	90.2 (1.67)	**96.5** (1.00)	**95.8** (0.95)
20	Flea beetles	92.9 (0.74)	91.4 (0.75)	92.7 (1.03)	93.9 (0.87)	**97.3** (1.00)	**96.4** (0.94)
50	Bank notes	92.9 (0.72)	88.2 (0.69)	83.0 (2.72)	90.8 (2.14)	**96.2** (1.00)	**95.1** (0.93)
50	Athletes	93.5 (0.70)	89.7 (0.72)	86.1 (2.73)	93.7 (1.93)	**96.8** (1.00)	**95.9** (0.93)
50	Skulls	92.2 (0.74)	88.0 (0.71)	84.9 (2.17)	92.3 (1.80)	**96.2** (1.00)	**95.3** (0.95)
50	Psychol. tests	94.4 (0.86)	89.6 (0.80)	83.0 (2.98)	91.7 (2.47)	**95.4** (1.00)	**96.6** (1.06)
50	Flea beetles	92.4 (0.88)	91.6 (0.89)	93.6 (1.01)	92.3 (0.92)	94.8 (1.00)	**95.5** (1.04)
80	Bank notes	93.5 (0.82)	89.6 (0.79)	85.2 (3.28)	92.3 (2.53)	**95.8** (1.00)	**95.6** (1.02)
80	Athletes	92.9 (0.78)	90.6 (0.79)	87.6 (2.63)	93.5 (1.80)	**95.6** (1.00)	**95.1** (0.96)
80	Skulls	93.2 (0.83)	89.8 (0.80)	87.0 (2.62)	92.8 (2.10)	**95.9** (1.00)	**95.8** (1.01)
80	Psychol. tests	94.7 (0.91)	91.0 (0.86)	86.3 (3.42)	92.6 (2.88)	**95.6** (1.00)	**96.7** (1.07)
80	Flea beetles	94.2 (0.92)	93.9 (0.92)	94.9 (1.00)	94.1 (0.97)	**95.3** (1.00)	**96.5** (1.05)
100	Bank notes	94.0 (0.85)	90.6 (0.83)	86.2 (3.41)	93.0 (2.62)	**95.7** (1.00)	**96.0** (1.04)
100	Athletes	93.2 (0.82)	91.3 (0.82)	88.3 (2.60)	93.8 (1.79)	**95.3** (1.00)	**95.1** (0.99)
100	Skulls	93.8 (0.86)	90.6 (0.84)	88.2 (2.76)	93.5 (2.23)	**95.9** (1.00)	**96.1** (1.03)
100	Psychol. tests	**95.2** (0.93)	91.6 (0.86)	87.4 (3.47)	93.8 (3.06)	**95.4** (1.00)	**96.6** (1.07)
100	Flea beetles	94.2 (0.94)	93.8 (0.94)	94.8 (1.00)	94.6 (0.99)	**95.1** (1.00)	**96.2** (1.06)
200	Bank notes	94.0 (0.93)	91.4 (0.91)	88.3 (3.49)	93.9 (2.60)	**95.0** (1.00)	**96.0** (1.07)
200	Athletes	93.1 (0.90)	92.3 (0.90)	90.3 (2.37)	93.7 (1.82)	94.4 (1.00)	**95.0** (1.03)
200	Skulls	94.6 (0.93)	92.1 (0.91)	90.3 (3.08)	94.4 (2.37)	**95.7** (1.00)	**96.5** (1.05)
200	Psychol. tests	**95.9** (0.96)	91.3 (0.92)	88.2 (4.13)	93.7 (3.55)	**96.1** (1.00)	**96.9** (1.07)
200	Flea beetles	**95.0** (0.97)	94.8 (0.97)	94.9 (1.00)	**95.6** (1.03)	**95.4** (1.00)	**96.2** (1.06)
Shortest intervals							
20	Bank notes	93.2 (0.62)	84.9 (0.57)	87.2 (1.29)	89.1 (1.08)	**97.6** (1.00)	93.5 (0.69)
20	Athletes	92.9 (0.72)	81.2 (0.73)	85.0 (2.18)	91.6 (2.29)	**96.5** (1.00)	**96.9** (1.12)
20	Skulls	90.2 (0.66)	80.9 (0.61)	84.2 (1.33)	88.1 (1.19)	**95.0** (1.00)	92.1 (0.80)
20	Psychol. tests	93.0 (0.75)	85.2 (0.68)	88.0 (1.62)	89.7 (1.43)	**95.6** (1.00)	**95.1** (0.96)
20	Flea beetles	89.9 (0.78)	88.7 (0.79)	93.5 (1.01)	92.5 (0.89)	94.8 (1.00)	94.1 (0.95)
50	Bank notes	92.9 (0.76)	87.3 (0.73)	89.3 (1.91)	90.4 (1.51)	**96.2** (1.00)	**95.3** (0.94)
50	Athletes	92.8 (0.75)	88.8 (0.77)	92.0 (2.06)	94.2 (1.51)	**97.0** (1.00)	**96.0** (0.94)
50	Skulls	91.5 (0.78)	86.3 (0.75)	89.8 (1.76)	91.2 (1.45)	**95.6** (1.00)	94.9 (0.97)
50	Psychol. tests	93.7 (0.88)	87.9 (0.81)	90.8 (2.13)	90.4 (1.80)	**95.0** (1.00)	**95.9** (1.05)
50	Flea beetles	91.2 (0.88)	90.3 (0.89)	93.7 (0.97)	90.7 (0.92)	93.5 (1.00)	94.2 (1.03)
80	Bank notes	93.4 (0.84)	89.0 (0.81)	90.4 (2.09)	92.1 (1.64)	**95.8** (1.00)	**95.8** (1.02)
80	Athletes	92.3 (0.82)	89.8 (0.83)	92.8 (1.88)	93.8 (1.37)	**95.7** (1.00)	**95.1** (0.97)
80	Skulls	92.3 (0.85)	88.4 (0.83)	91.2 (1.92)	92.1 (1.54)	**95.2** (1.00)	**95.3** (1.02)
80	Psychol. tests	94.2 (0.93)	89.5 (0.87)	92.3 (2.36)	91.8 (1.90)	**95.3** (1.00)	**96.1** (1.07)
80	Flea beetles	93.4 (0.92)	92.9 (0.93)	**95.0** (0.97)	93.3 (0.97)	94.3 (1.00)	**95.5** (1.05)
100	Bank notes	93.8 (0.87)	89.7 (0.84)	91.3 (2.10)	92.8 (1.67)	**95.7** (1.00)	**96.1** (1.04)
100	Athletes	92.7 (0.84)	90.6 (0.85)	93.4 (1.80)	94.1 (1.33)	**95.6** (1.00)	**95.3** (0.99)
100	Skulls	93.1 (0.88)	89.2 (0.85)	91.8 (1.94)	92.6 (1.58)	**95.3** (1.00)	**95.6** (1.03)
100	Psychol. tests	94.8 (0.94)	90.2 (0.87)	92.8 (2.20)	92.7 (1.85)	**95.1** (1.00)	**96.1** (1.06)
100	Flea beetles	93.3 (0.94)	92.9 (0.94)	94.9 (0.98)	94.0 (0.99)	94.5 (1.00)	**95.6** (1.06)
200	Bank notes	93.6 (0.94)	90.8 (0.92)	92.2 (2.05)	93.6 (1.61)	94.7 (1.00)	**95.7** (1.06)
200	Athletes	93.3 (0.91)	91.8 (0.91)	94.1 (1.61)	93.6 (1.30)	94.8 (1.00)	**95.3** (1.03)
200	Skulls	94.0 (0.94)	91.1 (0.92)	93.0 (1.96)	93.5 (1.56)	**95.1** (1.00)	**96.0** (1.05)
200	Psychol. tests	**95.2** (0.97)	90.2 (0.92)	92.1 (2.40)	92.0 (1.95)	**95.6** (1.00)	**96.7** (1.07)
200	Flea beetles	94.6 (0.97)	94.3 (0.97)	94.7 (0.99)	**95.2** (1.03)	94.9 (1.00)	**95.9** (1.06)

Average of the ratio of the median widths of intervals relative to the median widths of intervals given by Method A are shown in brackets. Population distributions are skew.

Similarly, with respect to the width of intervals, the results in Tables 9 and 10 show the same basic patterns as in Tables 3 and 4.

- In Table 9 (as in Table 3), the bias-corrected percentile method typically gives narrower confidence intervals than Method A (at the expense of having poor coverage). The pivotal methods generally give much wider interval than Method A and the percentile method and Method B give slightly wider intervals than Method A.
- In Table 10 (as in Table 4), the pivotal methods still typically give much wider intervals than the new methods, the percentile method gives slightly narrower intervals than the new methods, and differences in widths of the new methods favour Method A. Again, the bias-corrected method has the narrowest intervals, but with poor coverage.

The only noteworthy difference between results with the skew distributions and those with the MVN distributions is that average coverages were slightly smaller with the skew distributions. With the new methods, average coverages were still almost always above the nominal level, so average coverages for these methods were better with the skew distributions than with the MVN distributions; the other methods gave average coverages that were usually below the nominal level for the MVN distributions, so these were further below the nominal level for the skew distributions.

With the MVN distributions, the overall conclusion was that Method A had better results than other methods, with Method B a close second. This conclusion also holds for the skew population distributions. In general, the results found with the MVN population distributions were robust to the introduction of skewness and higher kurtosis.

6. Concluding comments

Motivation for this paper is the potential importance of the Garthwaite–Koch partition (as illustrated, for example, in the work of Rogers [9]) and the need of a method for forming confidence intervals for the quantities it yields. In the simulations, the two new methods almost always gave confidence intervals whose coverage was conservative, while the coverages of the four standard methods that were examined were generally liberal. Nevertheless, the widths of the intervals given by the new methods tended to be much smaller than those given by the non-studentized and studentized pivotal methods, and similar in size to those of the percentile method. The only method that gave appreciably narrower intervals than the new methods was the bias-corrected percentile, but the coverage of its intervals was typically well below the nominal 95% level. These results held both for MVN population distributions and for skew population distributions with heavy tails, suggesting some robustness of these results to departures from normality. Consequently, in this study the new methods clearly outperformed the standard methods.

In the study, Method A performed marginally better than Method B – there was little to choose between them in terms of their coverages but Method A tended to give slightly narrower intervals. Method A is also computationally a little simpler and a little faster than Method B (unlike Method B, it does not require second-level bootstrap sampling), so it is the method we recommend for constructing bootstrap confidence intervals for both the contributions and the percentage contributions determined by the Garthwaite–Koch partition.

The widths of equal-tailed and shortest confidence intervals given by Method A were compared. It was found the shortest interval was generally not markedly narrower than the equal-tailed interval when the shortest interval was a two-sided confidence interval, but differences tended to be much greater when the shortest interval was a one-sided confidence interval. In the present context with squared quantities, the shortest interval has further merit when it is one-sided, as it gives coherence with the interval for the un-squared quantity. Hence, reporting the shortest interval in preference to the equal-tailed interval should be strongly considered when the shortest interval is one-sided.

The good performance of the new methods begs two obvious questions: 'Why did the new methods perform well for this application?' and 'For what types of application are they likely to be useful?' Regarding the first question, none of the pivotal quantities used in this paper are exactly pivotal (under the strict definition of a pivotal quantity), but the simulations indicate that the W_i variables are closer to giving pivotal quantities than $\log W_i^2$ or $\log(W_i^2/\sum_{j=1}^m W_j^2)$. Hence the new methods performed better than the non-studentized and studentized pivotal methods. Also, pivotal methods will outperform percentile methods when conditions do not hold for the latter to work well *and* the pivotal methods use good pivotal quantities. So it seems that theW_i are reasonably good pivotal quantities for the application of interest here.

Regarding the secondquestion of when the methods will be useful, the benefit of the new methods is that they enable the use of a broader range of pivotal quantities than can be used with standard bootstrap methods. Thus the question may be rephrased as 'When will broadening the range of pivotal quantities prove advantageous?' Davison and Hinkley [16] note the importance of variance stabilization in choosing the quantity (θ) to bootstrap. They write (p. 111), 'Experience suggests that bootstrap methods for confidence limits and significance tests ... are most effective when θ is essentially a location parameter, which is approximately induced by a variance-stabilizing transformation'. This suggests that the new methods should be considered when the quantity of interest (such as W_i^2 or $W_i^2/\sum_{j=1}^m W_j^2$) can be constructed from simpler quantities that are essentially location parameters. For example, the methods should be considered if the quantity of interest can be expressed as a function of the means of several inter-related variables that have been scaled to have unit sample variances.

We believe that both Method A and Method B should prove useful in practice. While Method A performed marginally better than method B in this application, this will not always be the case. The difference between the two methods is similar to the difference between the non-studentized and studentized pivotal methods: Method A uses pivotal quantities whose variances may fluctuate across samples, while Method B uses pivotal quantities that have been standardized to have consistent sample variances. This has little benefit in the present application because the variances of the pivotal quantities used by Method A do not vary appreciably across bootstrap samples (they derive from the Mahalnobis distance – a scale invariant quantity). Further research is needed to evaluate the methods in other applications.

Acknowledgments

We are grateful to a referee for constructive comments that improved this paper and to Chris Jones for advice about the sinh–arcsinh transformation. The first author was supported by a PhD studentship from the Open University.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] Mahalanobis PC. On tests and measures of group divergence. J Asiat Soc Bengal. 1930;26(4):541–588.
[2] Liu H, Weng Q. Enhancing temporal resolution of satellite imagery for public health studies: a case study of West Nile Virus outbreak in Los Angeles in 2007. Remote Sens Environ. 2012;117:57–71.
[3] Martens H, Naes T. Multivariate calibration. Chichester: John Wiley & Sons; 1991.
[4] Jouan-Rimbaud D, Massart DL, Saby CA, et al. Determination of the representativity between two multidimensional data sets by a comparison of their structure. Chemometr Intell Lab Syst. 1998;40(2):129–144.
[5] Shah NK, Gemperline PJ. Combination of the Mahalanobis distance and residual variance pattern recognition techniques for classification of near-infrared reflectance spectra. Anal Chem. 1990;62(5):465–470.
[6] Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol. 1933;24(6):417–441.
[7] Clark JD, Dunn JE, Smith KG. A multivariate model of female black bear habitat use for a geographic information system. J Wildlife Manage. 1993;57(3):519–526.
[8] Garthwaite PH, Koch I. Evaluating the contributions of individual variables to a quadratic form. Aust N Z J Stat. 2016;58(1):99–119.
[9] Rogers DJ. Dengue: recent past and future threats. Philos T Roy Soc B. 2015;370. issue 1665. Available from: http://dx.doi.org/10.1098/rstb.2013.0562.
[10] Madansky A, Olkin I. Approximate confidence regions for constraint parameters. Stanford University. Department of Statistics; 1969.
[11] Reiser B. Confidence intervals for the Mahalanobis distance. Comm Stat Simul Comput. 2001;30(1):37–45.
[12] Radhakrisnan R. Estimating Mahalanobis's distance using Bayesian analysis. Comm Stat Theory Methods. 1984;13(20):2583–2600.
[13] Efron B. Bootstrap methods: another look at the jackknife. Ann Statist. 1979;7(1):1–26.
[14] Ukoumunne OC, Davison AC, Gulliford MC, et al. Non-parametric bootstrap confidence intervals for the intraclass correlation coefficient. Stat Med. 2003;22(24):3805–3821.
[15] Efron B, Tibshirani RJ. An introduction to the Bootstrap. New York: Chapman & Hall; 1993.
[16] Davision AC, Hinkley DV. Bootstrap methods and their application. Cambridge: Cambridge University Press; 1997.
[17] Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat Med. 2000;19(9):1141–1164.
[18] Datta GS, Mukerjee R. Probability matching priors: higher order asymptotics. New York: Springer; 2004.
[19] Flury B, Riedwyl H. Multivariate statistics: a practical approach. Cambridge: Cambridge University Press; 1988.
[20] Cook RD, Weisberg S. An introduction to regression graphics. New York: Wiley; 1994.
[21] Morant GM. A first study of the Tibetan skull. Biometrika. 1923;14(3/4):193–260.
[22] Beall G. Approximate methods in calculating discriminant functions. Psychometrika. 1945;10(3):205–217.
[23] Lubischew AA. On the use of discriminant functions in taxonomy. Biometrics. 1962;18(4):455–477.
[24] Jones MC, Pewsey A. Sinh–arcsinh distributions. Biometrika. 2009;96(4):761–780.