

# Modified confidence intervals for the Mahalanobis distance

Paul H. Garthwaite<sup>a\*</sup>, Fadlalla G. Elfadaly<sup>a,b</sup>, John R. Crawford<sup>c</sup>

<sup>a</sup>*School of Mathematics and Statistics, The Open University, MK7 6AA, UK;*

<sup>b</sup>*Statistics Department, Faculty of Economics & Political Science, Cairo University, Egypt;*

<sup>c</sup>*School of Psychology, King's College, University of Aberdeen, AB24 3FX, UK.*

## Abstract

Reiser (2001) proposes a method of forming confidence interval for a Mahalanobis distance that yields intervals which have exactly the nominal coverage, but sometimes the interval is  $(0, 0)$ . We consider the case where Mahalanobis distance quantifies the difference between an individual and a population mean, and suggest a modification that avoids implausible intervals.

*Keywords:* Credible interval; Noncentral F; Objective prior

## 1. Introduction

Suppose a set of measurements is made on an individual, recorded as a vector  $\mathbf{y}$ . For example, in a medical context where a patient might undergo a set of diagnostic tests, each component of  $\mathbf{y}$  might be the measurement on one test. In psychology, a person might perform a number of tests and  $\mathbf{y}$  could be the individual's profile of test scores. In commerce,  $\mathbf{y}$  might be a number of measurements made on a manufactured item. Mahalanobis distance can be used to calibrate the degree to which  $\mathbf{y}$  differs from the mean of a normative population. The squared Mahalanobis distance,  $\delta^2$  say, is defined by

$$\delta^2 = (\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}), \quad (1)$$

where  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are the mean and covariance matrix of the distribution of measurements from the normative population. The squared Mahalanobis distance is some-

---

\*Corresponding author.

*E-mail addresses:* paul.garthwaite@open.ac.uk(P.H. Garthwaite), f.elfadaly@open.ac.uk(F.G. Elfadaly), j.crawford@abdn.ac.uk(J.R. Crawford)

times called the *Mahalanobis index*. We assume that observations from the normative population follow a multivariate normal distribution.

A non-central  $F$  distribution can be used to form a confidence interval for  $\delta$  when  $\boldsymbol{\mu}$  and  $\Sigma$  are unknown and must be estimated from a normative sample. This method of forming a confidence interval for a Mahalanobis distance was proposed by Reiser (2001). It gives intervals that have the correct coverage so, for example, the proportion of 95% confidence intervals that will contain the true value of  $\delta$  is indeed 95 percent. However, as noted by Reiser (2001), *sometimes both the lower and upper limits of the confidence interval will equal 0*. Such intervals are (of course) amongst the 5 percent of intervals that do not contain  $\delta$ . Thus, while coverage equals the nominal level, there will be occasions when we know with absolute certainty that a confidence interval does not contain the true value of the quantity of interest. This is regrettable and the purpose of this note is to modify the method so that unreasonable interval estimates are never constructed.

In Section 2 we first give the unmodified method of forming a confidence interval and then develop our modification to the method. In Section 3 we examine the effect of the modification and concluding comments are given in Section 4.

## 2. Construction of interval estimators

We will refer to the individual whose profile is of interest as the case and members of the normative population as controls. Often the case has a distinguishing feature as when, for example, (s)he has received treatment for an illness while the controls are healthy volunteers. We assume a profile  $\mathbf{X}$  is a  $k$ -dimensional vector and that  $\mathbf{X} \sim \text{MVN}(\boldsymbol{\mu}, \Sigma)$  if  $\mathbf{X}$  is the profile of a control. We also assume that a sample of  $n$  controls has  $\bar{\mathbf{x}}$  as its sample mean and  $\mathbf{S}$  as its sample variance.

Denote the profile of the case by  $\mathbf{y}$ , where  $\mathbf{y}$  is nonrandom. Then

$$\hat{\delta}^2 = (\mathbf{y} - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{y} - \bar{\mathbf{x}}), \quad (2)$$

is the sample estimate of the squared Mahalanobis distance of the case. A general result [see, for example, Mardia et al. (1979), Exercise 3.5.1 and Theorem 3.5.2] is that if  $\mathbf{X} \sim \text{MVN}(\boldsymbol{\mu}, \Sigma)$  and  $\mathbf{M} \sim \text{Wishart}(\Sigma, n - 1)$ , then  $\{(n - k)/k\} \mathbf{X}^T \mathbf{M}^{-1} \mathbf{X} \sim$

$F_{k,n-k}(\boldsymbol{\mu}^T \boldsymbol{\Sigma} \boldsymbol{\mu})$ , where  $F_{k,n-k}(\boldsymbol{\mu}^T \boldsymbol{\Sigma} \boldsymbol{\mu})$  is a non-central  $F$ -distribution on  $k$  and  $n - k$  degrees of freedom with non-centrality parameter  $\boldsymbol{\mu}^T \boldsymbol{\Sigma} \boldsymbol{\mu}$ . From this it follows that

$$D^2 = \left[ \frac{n(n-k)}{(n-1)k} \right] \widehat{\delta}^2 \sim F_{k,n-k}(n\delta^2). \quad (3)$$

Exploiting this result, Reiser (2001) proposed the following method of forming confidence intervals for  $\delta$ . Let  $(\delta_l, \delta_u)$  denote the interval. Provided  $D^2$  is not too small,  $\delta_l$  and  $\delta_u$  are chosen to satisfy

$$Pr [F_{k,n-k}(n\delta_l^2) \leq D^2] = 1 - \frac{\alpha}{2} \quad (4)$$

and

$$Pr [F_{k,n-k}(n\delta_u^2) \leq D^2] = \frac{\alpha}{2}. \quad (5)$$

If  $D^2$  is so small that  $Pr[F_{k,n-k}(0) \leq D^2] \leq 1 - \frac{\alpha}{2}$ , then there is no  $\delta_l$  that satisfies (4) and  $\delta_l$  is instead set equal to 0. If  $Pr[F_{k,n-k}(0) \leq D^2] \leq \frac{\alpha}{2}$ , then there is no  $\delta_u$  that satisfies (5) and  $\delta_u$  is also instead set equal to 0.

Wunderlich et al., (2015) prove that the interval  $(\delta_l, \delta_u)$  is an exact  $1 - \alpha$  equal-tailed confidence interval for  $\delta$ . (This exactness result does not follow immediately from equation (3) because equations (4) and (5) do not always have a solution.) Nevertheless, the confidence intervals are not completely satisfactory as the interval will sometimes be  $(0, 0)$ , which, with probability 1, will not contain  $\delta$ . (Since  $\mathbf{y}$  is a continuous variate, with probability 1,  $\mathbf{y} \neq \boldsymbol{\mu}$  and  $\delta \neq 0$ .)

Moreover, any method of forming exact confidence intervals will sometimes give an interval of  $(0, 0)$  if intervals are determined by  $\mathbf{y} - \bar{\mathbf{x}}$  and  $\mathbf{S}$ . This follows from the definition of a confidence interval. If we repeatedly take samples of size  $n$  from the normative population then, for any specified case, for  $\alpha/2$  of the samples the upper limit of the  $1 - \alpha$  confidence interval for  $\delta$  must be below  $\delta$ . This holds even when the case's profile ( $\mathbf{y}$ ) is very close to  $\boldsymbol{\mu}$  and  $\delta \rightarrow 0$ . Given  $\mathbf{S}$ , let  $A(\mathbf{S})$  denote the set of values of  $\bar{\mathbf{x}} - \mathbf{y}$  for which the upper limit of the confidence interval is 0. Put  $\mathbf{w} = \bar{\mathbf{x}} - \mathbf{y}$ . As  $\mathbf{y}$  is a non-random vector,  $\mathbf{w} \sim \text{MVN}(\boldsymbol{\mu} - \mathbf{y}, \boldsymbol{\Sigma}/n)$ . Let

$$h(\mathbf{y}) = \int \left[ \int_{A(\mathbf{S})} |2\pi\boldsymbol{\Sigma}/n|^{-1/2} \exp\left\{-\frac{n}{2}(\mathbf{w} - (\boldsymbol{\mu} - \mathbf{y}))^T \boldsymbol{\Sigma}^{-1}(\mathbf{w} - (\boldsymbol{\mu} - \mathbf{y}))\right\} d\bar{\mathbf{x}} \right] f(\mathbf{S}) d\mathbf{S}, \quad (6)$$

where  $f(\mathbf{S})$  is the marginal distribution of  $\mathbf{S}$  [ $f(\mathbf{S})$  is proportional to a Wishart distribution] and the outside integral is over the range of values of  $\mathbf{S}$ . Then  $h(\mathbf{y})$  is the probability that the upper confidence limit is 0. We have that  $h(\mathbf{y})$  equals  $\alpha/2$  when  $\mathbf{y} = \boldsymbol{\mu}$  and, from equation (6), it is clear that  $h(\mathbf{y})$  is a continuous function of  $\mathbf{y}$ . Hence, for some  $\eta > 0$ , we have that  $h(\mathbf{y})$  exceeds, say,  $\alpha/4$  for  $\|\mathbf{y} - \boldsymbol{\mu}\| < \eta$ . Thus, for a range of profiles there is a non-negligible probability that the sample data will lead to a confidence interval of  $(0, 0)$ , and this result holds for any method of forming exact confidence intervals. To indicate the size of  $h(\mathbf{y})$  for the method proposed by Reiser, suppose  $\alpha = 0.05$ ,  $k = 3$  and  $n = 30$ . Then  $(\delta_l, \delta_u)$  is set equal to  $(0, 0)$  when  $D^2 < 0.071$  which, for example, happens with probability 0.016 if  $\Sigma$  is the  $3 \times 3$  identity matrix and  $\mathbf{y} - \boldsymbol{\mu} = (0.1, 0.1, 0.1)^T$ .

When  $D^2$  is small, a potentially attractive alternative to forming a confidence interval is to form a Bayesian credible interval. The definition of a credible interval does not involve repeated sampling. In principle *each*  $1 - \alpha$  credible interval contains the true value of the quantity of interest with probability  $1 - \alpha$ . Thus, a credible interval for  $\delta$  will not equal  $(0, 0)$ , as that is implausible. The challenge to forming a Bayesian credible interval is to find an appropriate prior distribution.

The vector  $\mathbf{y}$  could be the profile of an individual from the control population, rather than the profile of the case, so equation (1) defines  $\delta$  for an individual control as well as the case. Let  $\pi(\delta)$  be the prior distribution for  $\delta$  when  $\delta$  is the Mahalanobis distance of the case and let  $\psi(\delta)$  be the prior distribution when  $\delta$  is the Mahalanobis distance of a randomly selected control. We do not intend to specify  $\pi(\delta)$  but aim to place bounds on quantiles of its resulting posterior distribution through consideration of  $\psi(\delta)$ . As  $\delta$  measures the difference between  $\mathbf{y}$  and the mean of the controls, *a priori* we would expect  $\delta$  to generally be larger when it relates to the case than when it relates to a control. Hence, if  $\pi_q(\delta)$  and  $\psi_q(\delta)$  are the  $q$ th quantiles of  $\pi(\delta)$  and  $\psi(\delta)$ , respectively, it is reasonable to assume that  $\pi_q(\delta) \geq \psi_q(\delta)$  for any  $q \in (0, 1)$ . We shall assume that this relationship is passed on to the posterior distribution. Specifically, we shall assume that for any given  $\hat{\delta}^2 = (\mathbf{y} - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{y} - \bar{\mathbf{x}})$  and  $q \in (0, 1)$ ,

$$\pi_q(\delta | \hat{\delta}) \geq \psi_q(\delta | \hat{\delta}). \quad (7)$$

We do not know  $\pi(\delta)$ , so  $\pi(\delta | \widehat{\delta})$  is also unknown, but we can determine  $\psi(\delta | \widehat{\delta})$  (see below). Then  $\psi_{\alpha/2}(\delta | \widehat{\delta})$  and  $\psi_{1-\alpha/2}(\delta | \widehat{\delta})$  are lower bounds for the lower and upper endpoints of an equal-tailed  $(1 - \alpha/2)$  credible interval for  $\delta$ , where  $\delta$  is the Mahalanobis distance of the case.

Our modification to the Reiser's method of forming confidence intervals is straightforward: If an endpoint of a  $1 - \alpha$  confidence interval for the case is less than the lower bound of the corresponding endpoint of a  $1 - \alpha$  credible interval, set the endpoint equal to that lower bound. This treats the case like a control whenever the case resembles the average control more closely than the great majority of controls resemble their average. Usually, though, the modification is not applicable and the confidence interval is unchanged.

To implement the modification requires  $\psi(\delta | \widehat{\delta})$ . It is easier to work in terms of  $\lambda = \delta^2$  and  $\widehat{\lambda} = \widehat{\delta}^2$ , rather than  $\delta$  and  $\widehat{\delta}$ . Let  $\psi^*(\lambda)$  be the prior distribution that corresponds to  $\psi(\delta)$ . If we let  $\Lambda$  denote the Mahalanobis index of a control, then

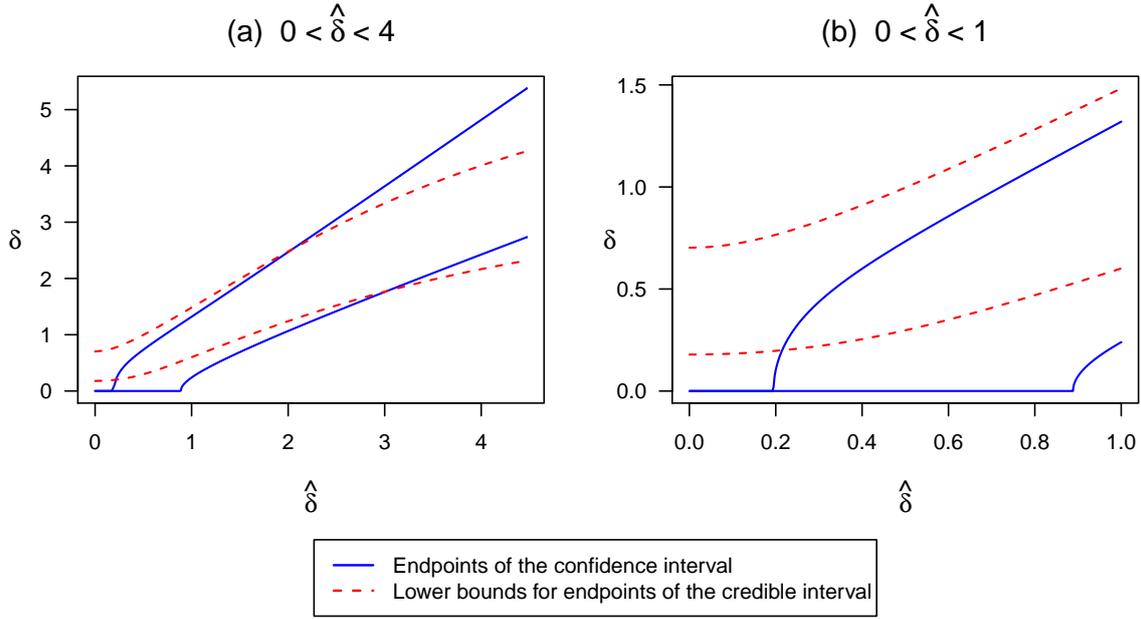
$$\Lambda = (\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_k^2, \quad (8)$$

where  $\chi_k^2$  is a central chi-square distribution on  $k$  degrees of freedom. Consequently, if  $\lambda$  is the value of  $\Lambda$  taken by a *control* that had been chosen at random then, *a priori*,  $\lambda \sim \chi_k^2$ . Thus  $\lambda \sim \chi_k^2$  is an objective prior distribution for a control's Mahalanobis index,  $\lambda$ , and we take  $\psi^*(\lambda)$  as this chi-square distribution.

The likelihood for  $\lambda$  follows from equation (3),  $[n(n-k)/\{(n-1)k\}]\widehat{\lambda} \sim F_{k, n-k}(n\lambda)$ . We multiply the prior distribution by the likelihood, compute the normalising constant through numerical integration, and hence obtain the posterior distribution,  $\psi^*(\lambda | \widehat{\lambda})$ . Further details are given in the appendix. Let  $\lambda_q$  denote the  $q$ th quantile of  $\psi^*(\lambda | \widehat{\lambda})$ . Then  $\lambda_q^{1/2}$  is a lower bound on the  $q$ th quantile of  $\pi(\delta | \widehat{\delta})$ , so  $\lambda_{\alpha/2}^{1/2}$  and  $\lambda_{1-\alpha/2}^{1/2}$  are lower bounds on the lower and upper endpoints of a central  $1 - \alpha$  posterior credible interval for the Mahalanobis distance ( $\delta$ ) of the case. Our modification sets  $\delta_l$  equal to  $\lambda_{\alpha/2}^{1/2}$  if  $\delta_l < \lambda_{\alpha/2}^{1/2}$  and puts  $\delta_u = \lambda_{1-\alpha/2}^{1/2}$  if  $\delta_u < \lambda_{1-\alpha/2}^{1/2}$ .

### 3. Effects of the modification

The two main features of the modified method are:



**Fig. 1.** Endpoints of the 95% confidence interval for  $\delta$  and lower bounds on endpoints of the corresponding credible interval are plotted against  $\hat{\delta}$  for a normative sample with  $k = 5$  and  $n = 25$ . The range of  $\hat{\delta}$  is 0–4 in plot (a) and 0–1 in plot (b).

1. The modification only takes effect when  $\hat{\delta}$  is small; when  $\hat{\delta}$  is a modest size (or bigger) the endpoints of the confidence interval are unchanged.
2. The method gives a plausible interval estimate of  $\delta$  for all values of  $\hat{\delta}$ .

Consequently the modified method has precisely the characteristics that were sought.

These features are illustrated in Figure 1, where we plot the 95% confidence interval endpoints and the credible interval lower bounds against  $\hat{\delta}$  for  $k = 5$  and  $n = 25$ . Plot (a) in the figure is for values of  $\hat{\delta}$  from 0 to 4, giving the broader picture, while (b) focuses on the important region where  $\hat{\delta}$  is small. The latter figure shows that the upper limit of the confidence interval drops from 0.44 to 0 as  $\hat{\delta}$  reduces from 0.3 to 0.19. When  $\hat{\delta} > 3.02$ , the confidence interval endpoints (solid lines) are above the corresponding credible interval lower bounds (dotted lines), so the confidence interval would only be modified for  $\hat{\delta} < 3.02$ , consistent with feature 1. Only the lower endpoint is changed for  $2.05 < \hat{\delta} < 3.02$ . Hence the modified interval estimate for  $\delta$  is given by (i) the two dotted lines in the figures for  $\hat{\delta} < 2.05$ , (ii) the lower dotted line and upper solid line for  $2.05 < \hat{\delta} < 3.02$ , and (iii) the two solid lines for

$3.02 < \hat{\delta}$ . These all seem reasonable interval estimates of  $\delta$ , consistent with feature 2. When  $\hat{\delta} = 0$ , the modified confidence interval for  $\delta$  is (0.18, 0.70). As  $\delta$  measures the difference between a particular individual's profile and the mean profile of the control population, this is decidedly more plausible than the unadjusted confidence interval of (0, 0).

We will refer to the values of  $\hat{\delta}$  at which an endpoint of the confidence interval and its lower bound are equal as *transition* values. The following are some other features of the modification.

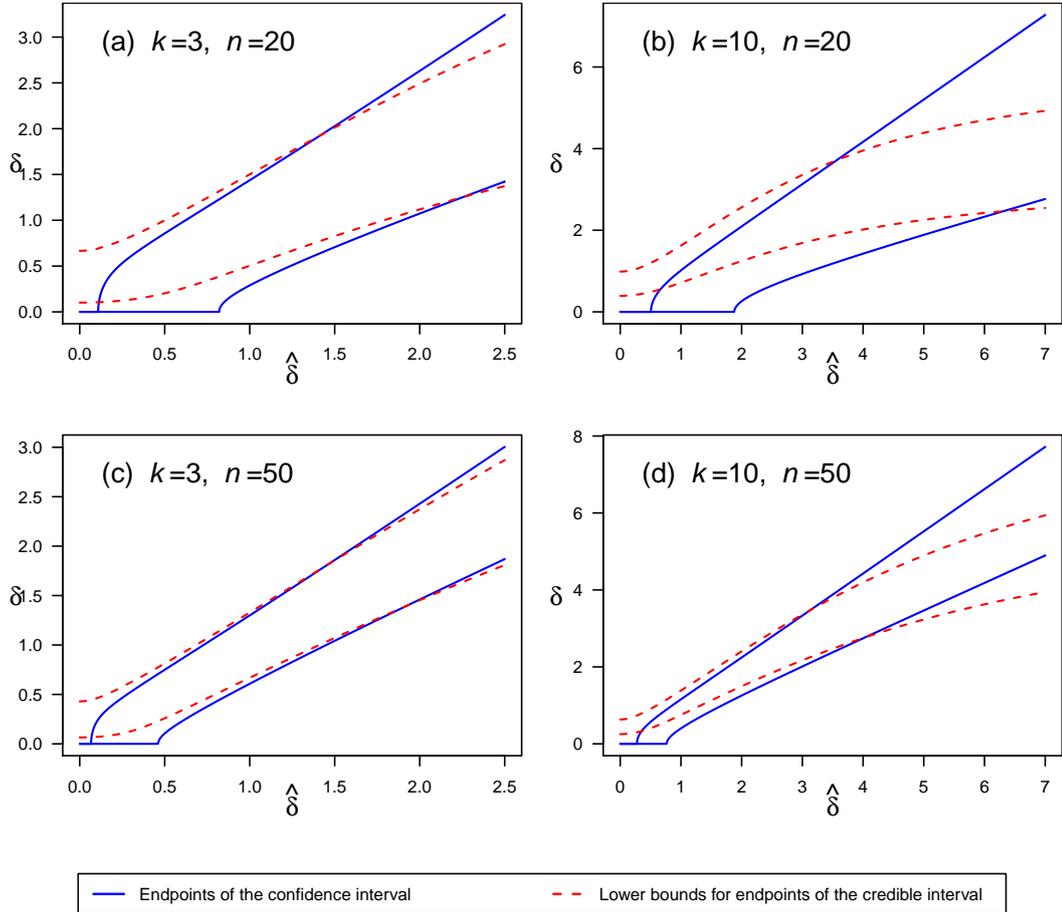
3. With the modified method (and also the unmodified method) each endpoint of the confidence interval is a continuous monotonic increasing function of  $\hat{\delta}$ . The modification does not introduce discontinuities at the transition values.
4. The effect of the modification on a confidence interval endpoint is small relative to the sensitivity of the endpoint to the value of  $\hat{\delta}$ .

These features of the modified method are illustrated in Table 1 and Figure 2, using  $n = 20$  and  $50$  with  $k = 3$  and  $10$ . In the table the effect of the modification is evaluated at half the transition value and 'equivalent change in  $\hat{\delta}$ ' is the amount by which  $\hat{\delta}$  must be increased to have the same effect. For example, for  $n = 20$  and  $k = 3$ , the modified confidence limit at 1.125 (half the transition value) is the same as the unmodified confidence limit at  $\hat{\delta} = 1.125 + 0.22$ . The change of 0.22 in  $\hat{\delta}$  is small, as are the other values for *equivalent change in  $\hat{\delta}$*  in Table 1. Figure 1 shows more generally that the effect of the modification is small, and it also illustrates that the modified method gives confidence intervals whose endpoints are continuous monotonic increasing functions of  $\hat{\delta}$ .

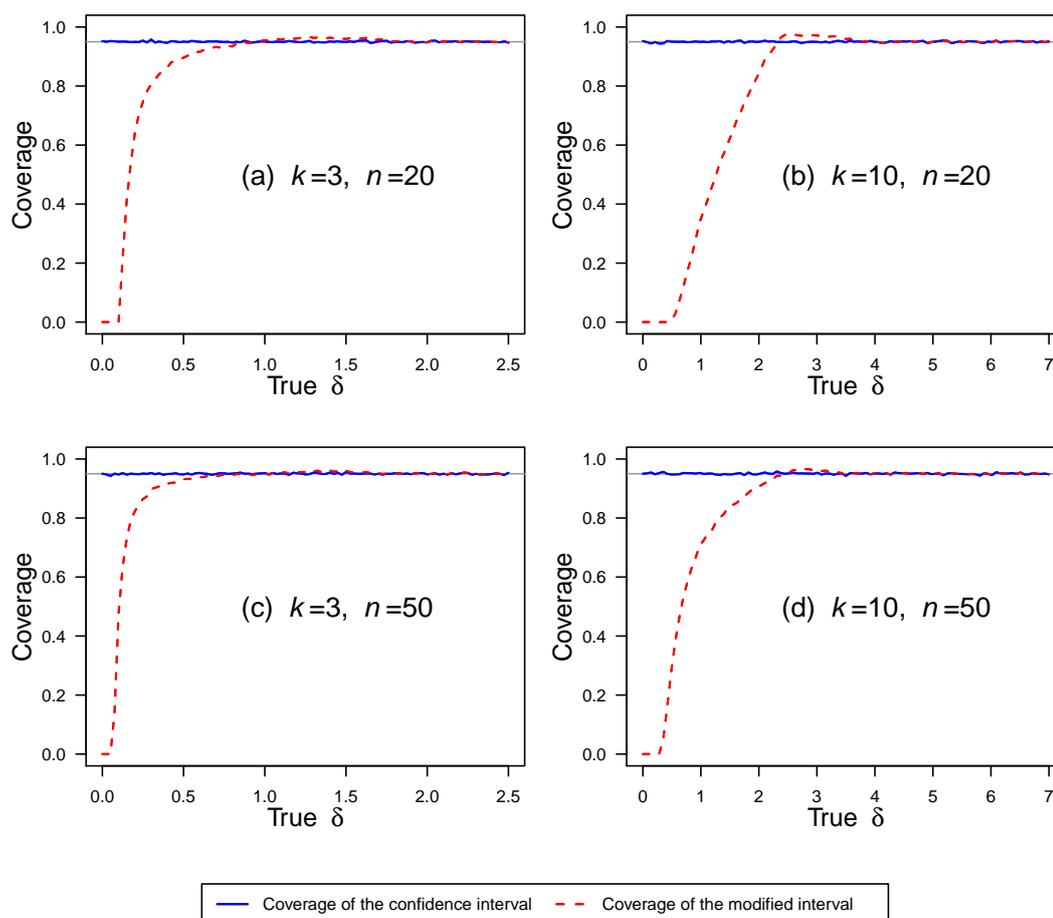
Two different sets of simulations were conducted to examine the coverage of confidence intervals formed using the modified and unmodified methods. In the first set, a value of  $\delta$  was fixed and 10,000 values were generated from  $F_{k,n-k}(n\delta^2)$ . Each generated value was multiplied by  $k(n-1)/\{n(n-k)\}$  to give 10,000 values of  $\hat{\delta}^2$  (c.f. equation (3)). From each  $\hat{\delta}$  a 95% confidence interval was calculated using both the modified and unmodified methods and the coverage of the intervals (i.e. the proportion that contained the true value of  $\delta$ ) was determined. This was done for

**Table 1** Transition values, the effect of the modification on confidence interval limits, and the increase in  $\hat{\delta}$  that would have the same effect.

	Lower limit				Upper limit			
	$n = 20$		$n = 50$		$n = 20$		$n = 50$	
	$k = 3$	$k = 10$	$k = 3$	$k = 10$	$k = 3$	$k = 10$	$k = 3$	$k = 10$
Transition value	2.25	6.31	1.91	4.03	1.41	3.56	1.45	3.13
Effect of modification	0.18	0.74	0.06	0.24	0.10	0.50	0.05	0.19
Equivalent change in $\hat{\delta}$	0.22	1.53	0.07	0.31	0.09	0.48	0.04	0.17



**Fig. 2.** Endpoints of the 95% confidence interval for  $\delta$  and lower bounds on endpoints of the corresponding credible interval are plotted against  $\hat{\delta}$  for (a)  $k = 3$ ,  $n = 20$ , (b)  $k = 10$ ,  $n = 20$ , (c)  $k = 3$ ,  $n = 50$ , and (d)  $k = 10$ ,  $n = 50$ .



**Fig. 3.** Coverage of 95% confidence intervals for the modified and unmodified methods, plotted against the parameter value  $\delta$ .

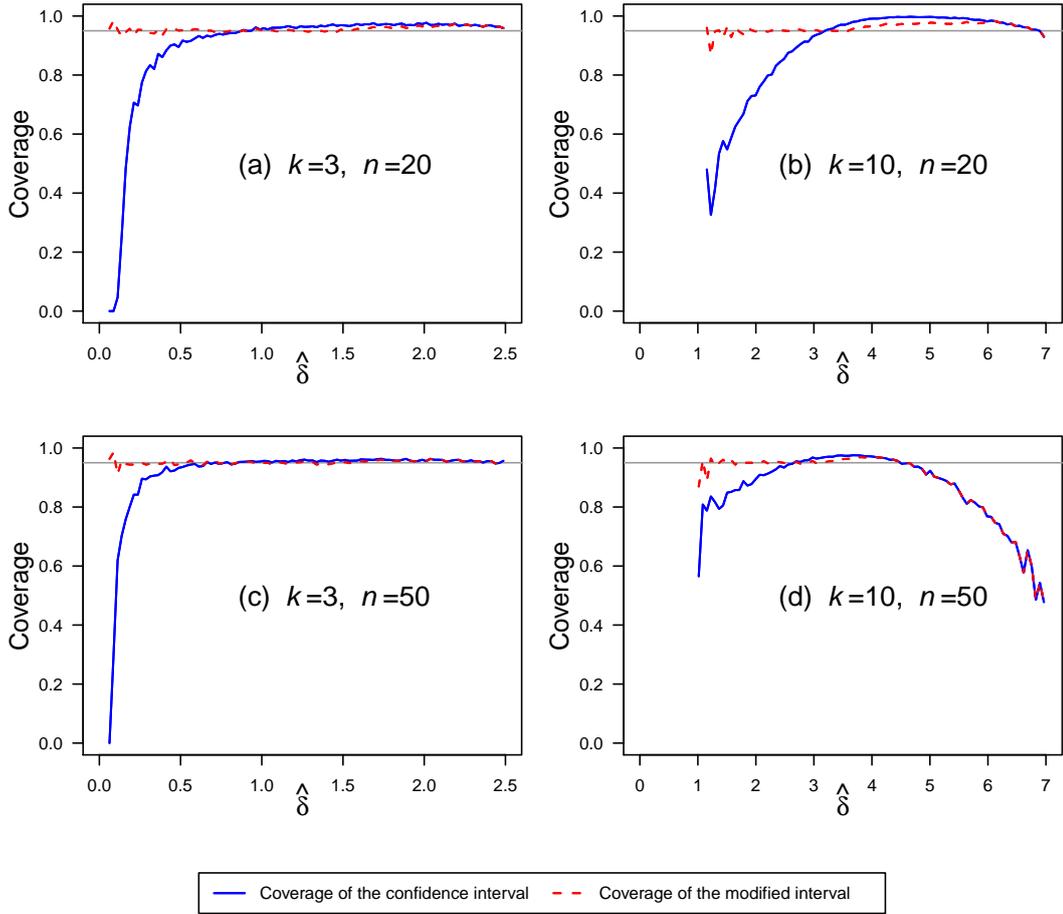
$k = 3, 10$  and  $n = 20, 50$  and  $100$  different values of  $\delta$ . Coverage is plotted against  $\delta$  for each combination of  $k$  and  $n$  in Figure 3. It can be seen that coverage is always equal to 0.95 with the unmodified method, apart from simulation error, consistent with theory. Coverage with the modified method can be much lower than 0.95 for small  $\delta$ , but is close to the nominal level when  $\delta$  is above 0.75 for  $k = 3$ , or above 2.25 for  $k = 10$ . To give perspective on these values, only 7.8% of controls have a value of  $\delta$  less than 0.75 when  $k = 3, n = 20$ , and only 2.1% of controls have  $\delta < 2.25$  when  $k = 10, n = 20$ . Thus coverage for the modified method is low only when the mean profile of controls is closer to the profile of the case than to the profile of the vast majority of controls.

For the second set of simulations, suppose a new treatment for a medical condition is tested by giving it to one patient with the condition (the case), while the controls are patients who have the condition but do not receive the new treatment. If the new treatment has no effect, then the profile of the case is from the same population as the profiles of the controls. The second set of simulations examines this situation and addresses the question:

“Suppose the case is picked at random from the controls and given a treatment that has no effect. If  $\hat{\delta}$  is the observed Mahalanobis distance of the case, what is the probability for the modified and unmodified methods that the 95% confidence interval will contain the true value,  $\delta$ ?”.

For a specified  $k$  and  $n$  ( $k = 3$  or  $10$ ;  $n = 20$  or  $50$ ), 500,000 values of  $\delta$  were generated from a  $\chi_k^2$  distribution (c.f. equation (1)). For each value of  $\delta$ , a random value from  $F_{k,n-k}(n\delta^2)$  was multiplied by  $k(n-1)/\{n(n-k)\}$  so as to obtain a single random value for  $\hat{\delta}^2$  (c.f. equation (3)). Based on  $\hat{\delta}$ , 95% confidence intervals for  $\delta$  were formed using the modified and unmodified methods and it was ascertained whether the true value of  $\delta$  lay in the intervals. So far, this set of simulations is similar to the first set. Indeed, if intervals were grouped according to the value of  $\delta$  (so the first group of intervals might arise from  $0.00 \leq \delta < 0.025$ , the second group from  $0.025 \leq \delta < 0.05$ , etc.) then a plot of the coverage for each group against the group's mid-point value of  $\delta$  would be the same as Figure 3, apart from simulation variation. However, for this second set of simulations the results were grouped according to the value of  $\hat{\delta}$ . The range of  $\hat{\delta}$  in each group was 0.025 for  $k = 3$  and 0.07 for  $k = 10$ . The coverages in the different groups are plotted against  $\hat{\delta}$  in Figure 4. (For very small  $\hat{\delta}$ , some groups contained very few items; coverages based on samples of fewer than 20 have not been plotted.)

For very small  $\hat{\delta}$  the unmodified method gives an unrealistic interval of  $(0, 0)$  and a coverage of 0. Figure 4 indicates that coverage of the method improves quickly as  $\hat{\delta}$  increases. However, its coverage is never better than the coverage of the modified method, which has excellent coverage for small  $\hat{\delta}$ . Hence Figure 4 supports the view that the modified method should be used to form interval estimates unless there are



**Fig. 4.** Coverage of 95% confidence intervals for the modified and unmodified methods when the case is from the same population as the controls, plotted against  $\hat{\delta}$ .

*a priori* reasons for expecting the average profile of the controls to be closer to the case's profile than to the profiles of most controls.

#### 4. Discussion

We should stress that the method developed in this paper is only appropriate when the quantity measured by the Mahalanobis distance is the discrepancy between an individual (the case) and the mean of a control population. In this situation the lower bounds of endpoints of a Bayesian credible interval can be obtained and these form principled lower bounds for interval estimates of the Mahalanobis distance. Mahalanobis distance is also commonly used as a measure of the distance between two population means or the distance between a population mean and some hypothesized

values. In these contexts a confidence interval of  $(0, 0)$  may be reasonable since, for example, two population would have identical population means if the characteristic that defines the populations has no bearing on the variables of interest. In these same contexts, a Bayesian analysis might use a prior distribution with a spike (point mass) of probability at 0 for equality of the population means, when the posterior credible interval could be  $(0, 0)$ , which would place no lower bounds on confidence interval endpoints.

The method developed here enables scientists to consistently report interval estimates that are sensible. Software implementing it is freely available from <http://users.mct.open.ac.uk/paul.garthwaite>. The software may be run over the web without down-loading it to one's computer, though the latter is an option.

### Acknowledgements

The work reported here was supported by a project grant from the Medical Research Council, grant number MR/J013838/1.

### Appendix

The prior distribution of  $\lambda$ ,  $\psi^*(\lambda)$ , is a central chi-square distribution on  $k$  degrees of freedom, so we have that

$$\psi^*(\lambda) = \frac{1}{\Gamma(k/2) 2^{k/2}} \lambda^{k/2-1} \exp(-\lambda/2), \quad \lambda \geq 0.$$

Since  $[n(n-k)/\{(n-1)k\}]\hat{\lambda}$  has a non-central  $F$  distribution on  $k$  and  $n-k$  degrees of freedom with non-centrality parameter  $n\lambda$ , the likelihood for  $\lambda$ , say  $L(\lambda; \hat{\lambda})$ , is

$$L(\lambda; \hat{\lambda}) = \sum_{r=0}^{\infty} \frac{(n\lambda/2)^r \exp(-n\lambda/2)}{B[(n-k)/2, k/2+r] r!} m^{k/2+r} \left( \frac{1}{1+m\hat{\lambda}} \right)^{n/2+r} \hat{\lambda}^{k/2+r-1}, \quad \lambda \geq 0,$$

where  $m = n/(n-1)$  and  $B[., .]$  is the beta function. Multiplying the prior distribution by the likelihood, we obtain the posterior distribution of  $\lambda$  as

$$\begin{aligned} \psi^*(\lambda|\hat{\lambda}) &= \frac{1}{c} \sum_{r=0}^{\infty} \frac{(n^r/2) (\lambda/2)^{k/2+r-1} \exp[-(n+1)\lambda/2]}{B[(n-k)/2, k/2+r] \Gamma(k/2) r!} \times \\ & m^{k/2+r} \left( \frac{1}{1+m\hat{\lambda}} \right)^{n/2+r} \hat{\lambda}^{k/2+r-1}, \quad \lambda \geq 0, \end{aligned} \quad (9)$$

where the normalizing constant,  $c$ , is obtained through numerical integration of (9) over  $0 < \lambda < \infty$ .

To compute the value of  $c$ , we use the Romberg's numerical integration method as a generalization of Simpson's and the trapezoidal quadrature rules (Press, 1989, p. 129). We divide the integration domain into two integrals over  $0 < \lambda < \hat{\lambda}$  and  $\hat{\lambda} < \lambda < \infty$ . A change of variables is applied to map the infinite range of the second improper integral to a finite one by inverting the integration boundaries. The  $q$ th quantile,  $\lambda_q$ , of the posterior distribution,  $\psi^*(\lambda|\hat{\lambda})$ , are then computed through a simple search procedure. Starting from  $\hat{\lambda}$  as an initial value, we search in both directions for the quantile  $\lambda_q$  such that  $|\int_{\lambda=0}^{\lambda_q} \psi^*(\lambda|\hat{\lambda})d\lambda - q| < 10^{-6}$ .

## References

- Mardia, K.V., Kent, J.T., Bibby, J.M., 1979. *Multivariate Analysis*. Academic Press, San Diego.
- Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1989. *Numerical Recipes in Pascal*. Cambridge University Press, New York.
- Reiser, B., 2001. Confidence intervals for the Mahalanobis distance. *Comm. Statist. Simulation Comput.* 30, 37-45.
- Wunderlich, A., Noo, F., Gallas, B.D. and Heiburn, M.E., 2015. Exact confidence intervals for channelized Hotelling observer performance in image quality studies. *IEEE Trans. Med. Imaging.* 34, 453-464.