STEPHEN T. BUCKLAND, PAUL H. GARTHWAITE
AND HOWARD G. LOVELL

*Scottish Agricultural Statistics Service, University of Aberdeen, and
University of Aberdeen*

This paper is very thorough and complete, and we do not feel that we can add materially to the subject matter covered. However, we would like to take this opportunity to widen the scope of the discussion in two different directions. First, we would like to say a few words in favour of ordinary percentile or "backwards" confidence intervals relative to the more sophisticated methods preferred by Hall. Second, we believe that for circumstances in which better confidence intervals are desired, it is often possible to apply a method that is exact apart from Monte Carlo variation, so that we need not consider progressive corrections to improve our approximations. We discuss two such methods, one based on Robbins–Monro search and the other on "permutation intervals." Neither requires us to assume that the "bootstrap world," in which parameters are estimated, is equivalent to the "real world," in which they are not.

**1. Complex applications.** In the Introduction, Hall notes that the percentile ("backwards") method was used in more than half of the cases in nontechnical statistical work and the hybrid method in most of the rest. One advantage of the percentile method is that it always provides confidence limits that lie within the permissible parameter range, provided the estimation procedure does not allow out-of-range estimates. By contrast, if a distribution has a finite bound to the permissible parameter space and the percentile confidence limit is close to this bound, the hybrid method will often give an interval that extends well beyond the bound for a skew distribution. When a distribution is skew, but with scale independent of location, the hybrid method is exact and the percentile method is not. However, we believe that such distributions are rare and that, in practice, the percentile method will normally be preferable to the hybrid method, as it is for the example of Table 2. The methods preferred by Hall also fail to ensure that the permissible parameter space is not violated.

A further advantage of the percentile interval is that, in our experience, it generates intervals similar in length to a central interval. By contrast, Hall's preferred methods may lead to intervals considerably longer, as noted by Hall and indicated in the example of Table 2. The improvement in coverage error may therefore be offset by the unnecessary increase in interval length.

In practice, skew distributions with at least one finite bound to the parameter space are common. If a true confidence limit is close to a bound, second- and third-order corrections may be unimportant relative to higher-order corrections; the distribution conditional on the estimated parameter(s) $\hat{\theta}$ may be very different from that conditional on true parameter(s) $\theta$. When implementation is feasible, we prefer instead to avoid considerations of second-, third-, ... order corrections and to use one of the following approaches:

**2. "Exact" bootstrap confidence intervals using Robbins–Monro search.** Where we require parametric bootstrap confidence limits for a single parameter and we have no nuisance parameters, we may select some estimate of, say, the upper limit and generate one or more bootstrap replications on the assumption that the estimate is the value of the unknown parameter. We then repeat the process for a new estimate of the upper limit. As we generate more bootstrap replications, we can use them to improve our estimate of the confidence limit. The Robbins–Monro search procedure [Robbins and Monro (1951)] maximises the efficiency of the search. Full details on the implementation of the method appear in Garthwaite and Buckland (1988); here, we give an outline of the concept only.

The Robbins–Monro process is used to search for each confidence limit separately. The procedure locates itself at the "best" estimate of a limit, generates a single bootstrap replication, finds the new best estimate in the light of the additional information and moves to it. Consequently, it steps from one best estimate of the limit to another. A "steplength constant" governs the magnitude of steps and if its optimal value is known, the procedure is fully efficient. In practice, the steplength constant must be estimated, and independent searches are necessary for the upper and lower limits, so that the procedure generally requires rather more than double the bootstrap replications to achieve the same nominal precision as the percentile method [Garthwaite and Buckland (1988)].

Suppose we have data $x_i$, $i = 1, \dots, n$, from which a parameter $\theta$ is estimated by $\hat{\theta}$. Suppose further that the data come from a single parameter density $f(x; \theta)$ of known form. Under relatively general conditions [Garthwaite and Buckland (1988)], the following search procedure converges to an exact $100(1 - 2\alpha)\%$ interval as the number of steps tends to infinity.

Let $u_m$ be the estimate of the upper confidence limit from $m$ steps. Generate a bootstrap sample of size $n$ from $f(x; u_m)$ and from this sample, estimate $\theta$ by $\hat{\theta}_m$, say. Set

$$u_{m+1} = \begin{cases} u_m - c\alpha/m, & \text{if } \hat{\theta}_m \geq \hat{\theta}, \\ u_m + c(1 - \alpha)/m, & \text{if } \hat{\theta}_m < \hat{\theta}. \end{cases}$$

The constant $c$ is the steplength constant. If $u_m$ is currently equal to the upper $100\alpha^*\%$ point [i.e., $\alpha^* = P(\hat{\theta}_m < \hat{\theta})$], the expected distance we step is $\{\alpha^* c(1 - \alpha)/m\} - \{(1 - \alpha^*)c\alpha/m\}$. This expression is zero for $\alpha^* = \alpha$, in which case $u_m$ is equal to the required upper confidence limit. The expression is positive when $\alpha^* > \alpha$ and negative for $\alpha^* < \alpha$, so that each step reduces the expected distance from the solution. The procedure continues for a predetermined number of steps or until convergence occurs with acceptable precision. An independent search yields the lower confidence limit estimate. Further details are given by Garthwaite and Buckland (1988).

The preceding method is based on the usual definition of a confidence interval and uses the relationship that exists between equal-tailed confidence intervals and two-sided hypothesis tests. For multivariate problems, an obvious way of implementing the method to set limits for a single parameter would be to replace nuisance parameters by their sample-data estimates. This would be moving toward the definition of a bootstrap confidence interval used by Hall, where resamples are generated from a density in which all parameters are replaced by sample estimates, including the parameter for which the interval is being constructed.

If we can define the expectations of the observations as functions of the parameter for which we require a confidence interval, then the preceding method may be used in conjunction with the nonparametric bootstrap, where we sample with replacement from the deviations of the observations from the expectations, estimated using the current best estimate of the confidence limit. In other words, at each step we update our estimates of the expectations of the observations, calculate the observed deviations from these estimates and sample with replacement from the deviations to provide a bootstrap replication. Otherwise, we proceed as before. Again, if there are nuisance parameters, we can choose to estimate them and apply the procedure conditional on those estimates.

**3. Exact permutation intervals.** Another way to exploit the relationship between hypothesis testing and confidence intervals is to use randomisation or permutation tests to generate intervals. If we enumerate all permutations, we obtain an exact confidence interval. If, for computational efficiency, we sample with replacement from the permutations, the method is exact apart from Monte Carlo variation.

Permutation (randomisation) tests are included in many elementary nonparametric textbooks. We use the permutation test for matched pairs to illustrate briefly the method. Suppose that an experiment is performed to compare the effects of two treatments. Subjects are matched in pairs as closely as possible and then one member of each pair is assigned to each treatment. A test is to be performed of the null hypothesis $H_0$ that there is no difference between the effects of the two treatments.

Suppose that observations are on an interval scale and the numerical difference for a particular pair is $|d|$. If $H_0$ is true, the observed difference $Y_a - Y_b$ could be either $+d$ or $-d$, depending only on the random allocation of treatments to subjects. The two signs are equally likely to appear. For an entire sample of $n$ pairs of observations, there are $2^n$ possible patterns of signed differences. Therefore there are $2^n$ ways in which the sum of the differences $\Sigma d_i$ might be constructed by chance (although that does not mean that there are $2^n$ different totals, since some totals may be constructed in several ways). The rejection region for the two-sided permutation test of size $2\alpha$ consists of the $200\alpha\%$ most extreme totals (irrespective of sign).

We may now define a permutation interval. The $100(1 - 2\alpha)\%$ permutation interval for a parameter $\theta$ consists of all values $\theta_0$ which would not be rejected if they were the subject of the null hypothesis $H_0$: $\theta = \theta_0$ and the alternative was two-sided, i.e., $H_1$: $\theta \neq \theta_0$. For example, a permutation interval for the mean difference in the case of matched pairs of observations is constructed by finding all the values of $\theta$ for which $H_0$: $\mu_A - \mu_B = \theta$ would not be rejected in favour of $H_1$: $\mu_A - \mu_B \neq \theta$, where $\mu_A$ and $\mu_B$ denote the expectation of $y_A$ and $y_B$, respectively.

This procedure is distribution-free in the sense that it does not depend upon the shape of the parent distribution(s) from which the observations were drawn. It does, of course, depend upon the (necessarily symmetric) permutation distribution generated from the sample data.

However, the value being estimated is a population value and, whatever procedure is used, random sampling is desirable. With matched pairs, if the two members of each pair were selected in the same way and given equal probabilities of the two possible assignments to treatments, it may seem plausible to treat the sample of differences as more representative of the (hypothetical) population of differences than the two individual samples are of their parent populations.

Permutation intervals are obtainable for parameters other than the mean. All that is necessary is that they can be estimated by sample statistics having unique values with each permutation of the data. Thus intervals can be constructed for variances and medians, but not for modes. The search for the confidence limit again renders the method inefficient in more than one dimension. If there is a single parameter of interest, as before we might choose to condition on the estimates of the nuisance parameters.

## REFERENCES

GARTHWAITE, P. H. and BUCKLAND, S. T. (1988). Generating Monte Carlo confidence intervals by the Robbins–Monro process. Unpublished.

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22** 400–407.

STEPHEN T. BUCKLAND
SASS, ABERDEEN UNIT
MACAULAY LAND USE RESEARCH INSTITUTE
CRAIGIEBUCKLER
ABERDEEN AB9 2QJ
UNITED KINGDOM

PAUL H. GARTHWAITE AND HOWARD G. LOVELL
DEPARTMENT OF STATISTICS
UNIVERSITY OF ABERDEEN
EDWARD WRIGHT BUILDING
DUNBAR STREET
OLD ABERDEEN AB9 2TY
UNITED KINGDOM

## THOMAS J. DICICCIO AND JOSEPH P. ROMANO

*Stanford University*

We congratulate Hall for a most stimulating paper. Hall has presented bootstrappers with a useful framework within which to compare resampling methods.

Before getting to the main topic of our discussion, we would like to raise two issues involving smoothing and uniformity. Outside of the obvious problem that a stable estimate of variance may be difficult to obtain, one may question the extent to which the results contribute to a complete theory of confidence intervals. It is known, for example, that outside of the "smooth function" model, the percentile-$t$ method may be inconsistent. For instance, bootstrap confidence intervals for functionals of a density (based on percentile-$t$ or other proposed methods) will generally be inconsistent unless resampling is performed from an