Statistical Science
2003, Vol. 18, No. 2, 175–184

The Impact of the Bootstrap on Statistical Algorithms and Theory

Rudolf Beran

Abstract. Bootstrap ideas yield remarkably effective algorithms for realizing certain programs in statistics. These include the construction of (possibly simultaneous) confidence sets and tests in classical models for which exact or asymptotic distribution theory is intractable. Success of the bootstrap, in the sense of doing what is expected under a probability model for data, is not universal. Modifications to Efron's definition of the bootstrap are needed to make the idea work for modern procedures that are not classically regular.

Key words and phrases: Confidence sets, error in coverage probability, double bootstrap, simultaneous confidence sets, local asymptotic equivariance, convolution theorem.

1. INTRODUCTION

Statistics is the study of algorithms for data analysis. Both the nature of the study and the nature of the algorithms have shifted with time. For the first half of the twentieth century, the primary tools available to statisticians were mathematics, logic and mechanical calculators. Advances in probability theory directed statisticians toward probability models for data and toward discussions of abstract principles that take such models for granted. In retrospect, the diverse statistical theories offered by Wald's (1950) *Statistical Decision Functions*, Fisher's (1956) *Statistical Methods and Scientific Inference* and Savage's (1954) *The Foundations of Statistics* share a common reliance on probability models. Throughout this period, statistical computation remained laborious. Both Fisher's (1930) *Statistical Methods for Research Workers* and Quenouille's (1959) *Rapid Statistical Calculations* exhibit how strongly computational environment influences the structure of statistical procedures.

After 1960, Prohorov's results on weak convergence of probability measures led to sustained development of asymptotic theory in statistics. Notable achievements by about 1970 were the clarification of what is meant by asymptotic optimality; the recognition that maximum likelihood estimators behave badly in some regular parametric models but can be patched in various ways; the unexpected discovery of superefficient estimators, such as the Hodges example, for which asymptotic risk undercuts the information bound on sets of Lebesgue measure zero; and the remarkable discovery, through the James–Stein estimator, that superefficient estimators for parameters of sufficiently high dimension can dominate the maximum likelihood estimator (MLE) globally. Modern estimators in multiparametric or nonparametric models, such as those generated by adaptive penalized maximum likelihood, use this insight to improve substantially on the risk of maximum likelihood estimators.

By mid-century, mathematical logicians investigating the notion of proof had greatly refined the concept of algorithm (Berlinski, 2001). As realized in subsequent decades through digital computers, programming languages, displays, printers and numerical linear algebra, stable algorithms influenced statistical practice. A much wider range of statistical procedures, numerical and graphical, became computationally feasible. Details of the performance of a statistical procedure in case studies or in repeated pseudo-random realizations from probability models began to supplement or supplant results from asymptotic theory, such as rates of convergence and asymptotic minimax bounds.

The fundamental distinctions among data, probability model, pseudo-random numbers and algorithm returned to prominence. The extent to which deter-

Rudolf Beran is Professor of Statistics, University of California, Davis, California 95616 (e-mail: beran@wald.ucdavis.edu).

ministic pseudo-random sequences can imitate properties of random variables received greater mathematical attention (Knuth, 1969, Chapter 3). Tukey's (1970) *Exploratory Data Analysis* gave a purely algorithmic account of various statistical procedures. Although the exposition of the book made no overt use of either computers or probability models, the author remarked, "Today's understanding of how well elementary techniques work owes much to both mathematical analysis and experimentation by computer." The design of the S language and its implementations through the commercial S-Plus and the later open-source R were further stages in the algorithmic development of statistics.

One consequence of these trends was growing differentiation of statistical interests. Some statisticians continued to use probability models to analyze the behavior of statistical procedures, some focused on devising computationally effective statistical algorithms, whether or not inspired by a probability model and some sought to analyze data, much of which does not constitute a random sample. Dialog among these subgroups was limited. Important efforts at rapprochement in the areas of time series analysis and robustness occurred in the 1970s, but did not overcome the difficulty of linking algorithmic and probability model formulations of statistics.

Such was the backdrop to the publication of Efron's (1979) article on the bootstrap. Unlike many innovations, the bootstrap quickly gained broad attention, in part because it raised matters of immediate interest to each statistical subgroup listed above. The reception of the bootstrap may be contrasted with the long neglect of subsampling, which was pioneered by Gosset in Student (1908a, b). It was studies of the bootstrap that revived research into subsampling (Politis, Romano and Wolf, 1999). In the historical context of the 1970s, some statisticians realized that statistical procedures based largely on analytical manipulations do not provide an effective technology for handling data sets that arise in the computer age. Monte Carlo algorithms for approximating bootstrap distributions offered a remarkably intuitive way to estimate complex sampling distributions that depend on unknown parameters. Mathematically interesting in their own right, such Monte Carlo techniques reflected the growing role of algorithms and computational experiments in statistics. Indeed, in the future, the performance of a statistical procedure under probability models would remain important, but would not be the sole theoretical arbiter of its success.

Studies of bootstrap procedures have brought out several points:

- Bootstrap algorithms provide an effective and intuitive way to realize certain programs in statistics. These include the construction of (possibly simultaneous) confidence sets, tests and prediction regions in classical models for which exact or asymptotic distribution theory is intractable.
- Success of the bootstrap, in the sense of doing what is expected under a probability model for data, is not universal. Modifications to Efron's (1979) definition of the bootstrap are needed to make the idea work for estimators that are not classically regular.
- A probability model is a mathematical construct that seeks to approximate salient relative frequencies in data. In the application of bootstrap techniques to data analysis, the pertinence and limitations of probability modeling are questions. For a discussion of answers in the earlier context of spectrum analysis, see Section 17 of Brillinger and Tukey (1984).

This article develops the first two of these points. General background and bootstrap bibliographies can be found in Beran and Ducharme (1991), Davison and Hinkley (1997), Efron and Tibshirani (1993), Hall (1992) and Mammen (1992).

2. BOOTSTRAP CONFIDENCE SETS

Consider a statistical model in which a sample X_n of size n has joint probability distribution $P_{\theta,n}$, where $\theta \in \Theta$ is unknown. The parameter space Θ is an open subset of a metric space, whether of finite or infinite dimension. Of interest is the parameter $\tau = T(\theta)$, where T is a specified function on Θ. The unknown value of τ lies in $\mathcal{T} = T(\Theta)$. A confidence set C for τ is a random subset of $\mathcal{T}$ that depends on the sample X_n such that the *coverage probability* $P_{\theta,n}[C \ni \tau]$ is as close as possible to a specified design level β. We write β in place of the conventional $1 - \alpha$ with the understanding that β is close to 1.

We pursue confidence set constructions based on a real-valued *root* $R_n(X_n, \tau)$, that has a left-continuous c.d.f. under the model that is $H_n(x, \theta) = P_{\theta,n}[R_n(X_n, \tau) < x]$. The proposed confidence set for τ is $C = \{t \in \mathcal{T} : R_n(X_n, t) \le c(\beta)\}$, a form suggested by classical confidence sets that are based on a pivot with a distribution that does not depend on unknown parameters. The crux of the matter is to choose the critical value $c(\beta)$ so as to bring the coverage probability close to β.

If θ were known and the c.d.f. $H_n(x, \theta)$ were continuous in x, an oracular choice of critical value would be $c(\beta) = H_n^{-1}(\beta, \theta)$, the largest βth quantile of the c.d.f. Evidently

$$C_n = \{t \in \mathcal{T} : R_n(X_n, t) \le H_n^{-1}(\beta, \theta)\} \quad (1)$$
$$= \{t \in \mathcal{T} : H_n(R_n, \theta) \le \beta\}$$

has exact coverage probability β, because the distribution of $H_n(R_n, \theta)$ is Uniform(0, 1).

This oracle construction is, in fact, useful when the root is a pivot, because in that case the c.d.f. $H_n(\cdot, \theta)$ does not depend on the unknown θ. Outside this very restrictive setting, it is natural to seek asymptotic constructions of the critical value. Let $\hat{\theta}_n$ denote an estimator for θ that is consistent in the metric d on $\Theta: d(\hat{\theta}_n, \theta) \to 0$ in probability as $n \to \infty$. An intuitive solution is to estimate the c.d.f. $H_n(\cdot, \theta)$ by $\hat{H}_B(\cdot) = H_n(\cdot, \hat{\theta}_n)$, which is called the *bootstrap c.d.f.* The corresponding *bootstrap confidence set* is then

$$C_B = \{t \in \mathcal{T} : R_n(X_n, t) \le \hat{H}_B^{-1}(\beta)\} \quad (2)$$
$$= \{t \in \mathcal{T} : \hat{H}_B(R_n) \le \beta\}.$$

As introduced above, the bootstrap distribution is a random probability measure, which may be usefully reinterpreted as a conditional distribution. Consider the *bootstrap world* in which the true parameter is $\hat{\theta}_n$, so that τ is replaced by $\hat{\tau}_n = T(\hat{\theta}_n)$, and observe an artificial sample X_n^* that is drawn from the fitted model $P_{\hat{\theta}_n,n}$. In other words, the conditional distribution of X_n^* given X_n is $P_{\hat{\theta}_n,n}$. Then $\hat{H}_B(\cdot)$ is the conditional c.d.f. of $R_n(X_n^*, \hat{\tau}_n)$ given X_n. This formulation is the starting point for Monte Carlo algorithms that approximate the bootstrap distribution and confidence set.

A basic question is whether the coverage probability of bootstrap confidence set C_B converges to the desired level β as $n \to \infty$. The following, almost self-evident, proposition provides a template for checking this.

TEMPLATE A. *Suppose that, for every $\theta \in \Theta$,*

(i) $d(\hat{\theta}_n, \theta) \to 0$ *in $P_{\theta,n}$ probability as $n \to \infty$.*

(ii) (Triangular array convergence.) *For any sequence $\{\theta_n \in \Theta\}$ such that $d(\theta_n, \theta)$ converges to 0, the c.d.f.s $\{H_n(\cdot, \theta_n)\}$ converge weakly to $H(\cdot, \theta)$.*

(iii) *The limit c.d.f. $H(x, \theta)$ is continuous in x.*

Let $\|\cdot\|$ denote supremum norm on the real line. Then $\|\hat{H}_B - H(\cdot, \theta)\| \to 0$ in $P_{\theta,n}$ probability as $n \to \infty$ and, for $\beta \in (0, 1)$,

$$\lim_{n\to\infty} P_{\theta,n}[C_B \ni \tau] = \beta. \quad (3)$$

The skill in using this result lies in devising the metric d. Condition (i) is easier to check if the metric is weak, while condition (ii) is more likely to hold if the metric is strong. The following two examples illustrate an important general point: whether directly or through its Monte Carlo approximation, the intuitive confidence set C_B solves some classical parametric and nonparametric confidence set problems where asymptotic distributions are intractable. In such problems, the conditions of Template A can often be checked (we omit details) and C_B therefore has correct asymptotic coverage probability.

EXAMPLE 2.1 (Kolmogorov–Smirnov confidence band). The sample consists of n i.i.d. random variables, each of which has unknown c.d.f. F. The problem is to construct a fixed-width confidence band for F using the root $R_n(X_n, F) = n^{1/2}\|\hat{F}_n - F\|$, where $\hat{F}_n$ is the empirical c.d.f. and the norm is supremum norm. Here $\theta = \tau = F$. Let $\hat{\theta}_n = \hat{F}$. In contrast to the usual direct asymptotic construction, C_B has correct asymptotic coverage probability whether or not F is continuous.

EXAMPLE 2.2 (Confidence double cone for an eigenvector). The sample consists of n i.i.d. random k vectors, each having an unknown distribution P with finite mean vector and covariance matrix $\Sigma(P)$. Here $\theta = P$. The problem, stemming from principal component analysis, is to construct a confidence double cone for the eigenvector (actually eigenaxis) τ that is associated with the largest eigenvalue of $\Sigma(P)$. The root is $R_n(X_n, \tau) = n(1 - |\hat{\tau}_n'\tau|)$, where $\hat{\tau}_n$ is the sample eigenvector associated with the largest eigenvalue of the sample covariance matrix. Let $\hat{\theta}_n = \hat{P}_n$ denote the empirical distribution of the sample. Although the asymptotic distribution of the root lacks workable analytic expression, C_B has correct asymptotic coverage probability if the largest eigenvalue of $\Sigma(P)$ is unique.

When the root is not a pivot, the coverage probability of C_B at sample size n usually differs from the design goal β. We consider structurally how properties of the root influence the error in coverage probability. To this end, suppose that the c.d.f. $H_n(x, \theta)$ admits an asymptotic expansion

$$H_n(x, \theta) \quad (4)$$
$$= H(x, \theta) + n^{-r_0/2}h(x, \theta) + O\big(n^{-(r_0+1)/2}\big)$$

that holds locally uniformly in θ. In the expansion, r_0 is positive, the limit c.d.f. $H(\cdot, \theta)$ is continuous and strictly monotone on its support, and $h(\cdot, \theta)$ is continuous. Moreover, suppose that $\hat{\theta}_n$ is a root-n consistent estimator of θ. We distinguish two cases, according to whether or not the limit c.d.f. $H(\cdot,\theta)$ depends on θ. Heuristic reasoning suggests the following conclusions:

CASE AP (Asymptotic pivot). Suppose that the limit c.d.f. does not depend on θ, so that $H(x,\theta) = H(x)$. Then typically the bootstrap confidence set C_{B} has coverage probability $\beta + O(n^{-(r_0+1)/2})$. In contrast, the asymptotic confidence set C_{A} that is constructed like C_{B} but uses critical value $H^{-1}(\beta)$ has coverage probability $\beta + O(n^{-r_0/2})$. In this case, confidence set C_{B} has asymptotically smaller error in converge probability than does confidence set C_{A}. This is an important advantage to the bootstrap confidence set that goes beyond relative ease of construction.

CASE NAP (Not asymptotic pivot). Suppose that the limit c.d.f. $H(\cdot,\theta)$ depends on θ and the c.d.f. $J_n(\cdot,\theta)$ of $H_n(R_n,\hat{\theta}_n)$ admits the expansion

$$\begin{aligned} &J_n(x,\theta) \\ &\quad = U(x) + n^{-r_1/2}u(x,\theta) + O\big(n^{-(r_1+1)/2}\big), \end{aligned} \tag{5}$$

where $U(\cdot)$ is the Uniform(0, 1) c.d.f. Then typically both C_{B} and C_{A}—the asymptotic analog of C_{B} that uses critical value $H^{-1}(\beta,\hat{\theta}_n)$—have coverage probability $\beta + O(n^{-r_1/2})$.

Examples 2.1 (with discontinuous F) and 2.2 illustrate NAP situations where construction of C_{A} is difficult while Monte Carlo approximation of C_{B} is straightforward. The next two examples show bootstrap confidence set C_{B} in its preferred AP setting.

EXAMPLE 2.3 (Behrens–Fisher confidence interval). The sample is the union of two independent subsamples. Subsample i contains n_i i.i.d. random variables, each having a $N(\mu_i,\sigma_i^2)$ distribution. Here $\theta = (\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)$. The problem is to construct a confidence interval for $\tau = \mu_1 - \mu_2$. The Behrens–Fisher root is $R_n(X_n,\tau) = |(\hat{\mu}_1 - \hat{\mu}_2) - \tau|/[\hat{\sigma}_1^2/n_1 + \hat{\sigma}_2^2/n_2]^{1/2}$, where $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ are the sample mean and variance. Let $\hat{\theta}_n = (\hat{\mu}_1,\hat{\mu}_2,\hat{\sigma}_1^2,\hat{\sigma}_2^2)$. Suppose that $n_1 = \lfloor n\lambda \rfloor$ and $n_2 = n - n_1$ with $0 < \lambda < 1$. While the exact distribution of the root depends on the unknown ratio σ_1^2/σ_2^2, the limit distribution as $n \to \infty$ is the folded-over standard normal. Further asymptotic analysis reveals that, in this AP example, C_{B} is asymptotically as accurate as Welch's confidence interval for the Behrens–Fisher problem. Unlike the latter, C_{B} requires no algebraic intervention by the user.

EXAMPLE 2.4 (Likelihood ratio confidence set). The sample consists of n i.i.d. random variables drawn from a distribution that belongs to a canonical exponential family indexed by θ in $\mathbb{R}^k$. Let $\tau = \theta$, let $L_n(\theta)$ be the log-likelihood function and let $\hat{\theta}_n$ be the maximum likelihood estimator of θ. The root $R_n(X_n,\theta) = 2[L_n(\hat{\theta}_n) - L_n(\theta)]$ has a chi-squared asymptotic distribution with k degrees of freedom, an instance of AP. Deeper asymptotic analysis shows that C_{B} indirectly accomplishes the Bartlett adjustment to the chi-squared asymptotics. That C_{B} automatically matches the Bartlett adjustment for standard likelihood ratio confidence sets associated with the multivariate normal model is a powerful result that has not been much exploited.

On reflection, the bootstrap itself provides an intuitive way to change an NAP root into an AP root. In the setting of Template A, the distribution of the transformed root

$$R_{n,1}(X_n,\tau) = \hat{H}_{\mathrm{B}}[R_n(X_n,\tau)] = H_n(R_n,\hat{\theta}_n) \tag{6}$$

converges to the Uniform(0, 1) distribution. Consequently, $R_{n,1}$ is an AP root if its c.d.f. has a suitable asymptotic expansion. The transformation that maps R_n into $R_{n,1}$ is called *prepivoting*. Let $H_{n,1}(x,\theta) = P_{\theta,n}[R_{n,1}(X_n,\tau) < x]$ and let $\hat{H}_{\mathrm{B},1} = H_{n,1}(\cdot,\hat{\theta}_n)$ be its bootstrap estimator. The bootstrap confidence set determined by $R_{n,1}(X_n,\tau)$ is then

$$\begin{aligned} C_{\mathrm{B},1} &= \{t \in \mathcal{T} : \hat{H}_{\mathrm{B},1}[R_{n,1}(X_n,t)] \le \beta\} \\ &= \{t \in \mathcal{T} : R_n(X_n,t) \le \hat{H}_{\mathrm{B}}^{-1}[\hat{H}_{\mathrm{B},1}^{-1}(\beta)]\}. \end{aligned} \tag{7}$$

The construction of $C_{\mathrm{B},1}$ involves two bootstrap worlds. In the *first* bootstrap world, as already described, the true parameter is $\hat{\theta}_n$, so that τ is replaced by $\hat{\tau}_n = T(\hat{\theta}_n)$, and we observe an artificial sample X_n^* with a conditional distribution, given X_n, of $P_{\hat{\theta}_n,n}$. Let $\theta_n^* = \theta(X_n^*)$ be the reestimation of θ based on this sample from the first bootstrap world. In the *second* bootstrap world, the true parameter is θ_n^*, so that τ is replaced by $\tau_n^* = T(\theta_n^*)$, and we observe an artificial sample X_n^{**} with a conditional distribution, given X_n and X_n^*, of $P_{\theta_n^*,n}$. Then:

- $\hat{H}_{\mathrm{B}}$ is the conditional c.d.f. of $R_n^* = R_n(X_n^*,\hat{\tau}_n)$, given X_n.
- $\hat{H}_{\mathrm{B},1}$ is the conditional distribution of $R_{n,1}^* = R_{n,1}(X_n^*,\hat{\tau}_n)$, given X_n. Moreover,

$$\begin{aligned} R_{n,1}^* &= H_n(R_n^*,\hat{\theta}_n^*) \\ &= P_{\theta_n^*,n}[R_n^{**} < R_n^* | X_n, X_n^*], \end{aligned} \tag{8}$$

where $R_n^{**} = R_n(X_n^{**}, \tau^*)$.

From this we see that a double nested Monte Carlo algorithm serves to approximate $C_{B,1}$. The inner level of the algorithm replicates X_n^* to approximate $\hat{H}_B$. For each replicate of X_n^*, the outer level replicates X_n^{**} to approximate $\hat{H}_{B,1}$.

Working with the second bootstrap world is often called double bootstrapping and confidence sets such as $C_{B,1}$ that rely on the second bootstrap world are called double-bootstrap confidence sets. Hall (1992) developed a mathematical setting where double bootstrapping, in several asymptotically equivalent forms, reduces coverage probability error asymptotically.

As noted above, prepivoting typically transforms a NAP root into an AP root that generates bootstrap confidence sets with smaller error in coverage probability than bootstrap confidence sets based on the NAP root. In addition, prepivoting an AP root typically reduces the asymptotic order of the second term in its asymptotic distribution. Consequently, bootstrap confidence sets based on prepivoted AP roots also often enjoy asymptotically smaller error in coverage probability than those based on the initial root. This is an argument for continued prepivoting. However, the computational burden discourages more than double bootstrapping and the asymptotic theory just outlined may not capture all factors that influence coverage probability at finite sample sizes.

EXAMPLE 2.3 (Continued). Under the normal model, Welch's analytic confidence interval and the bootstrap confidence interval C_B based on the Behrens–Fisher root both have error in coverage probability of order n^{-2}. The double-bootstrap confidence interval $C_{B,1}$ has error in coverage probability that is of order n^{-3}. Simulations with normal pseudo-random samples indicate the reality of these improvements at remarkably small sample sizes.

EXAMPLE 2.5 (Nonparametric t statistic). Studentizing is another way to transform a root to AP. Suppose that the sample is i.i.d. with unknown c.d.f. F that has finite mean $\mu(F)$ and finite variance. Here $\theta = F$ and of interest is $\tau = \mu(F)$. As the root, take the one-sided t statistic $n^{1/2}(\hat{\mu} - \mu(F))/\hat{\sigma}$, using the standard mean and variance estimators. Let $\hat{\theta}_n = \hat{F}_n$ be the empirical c.d.f. of the sample. The error in the coverage probability of C_B is usually of order n^{-1} when F is not symmetric. Simulations reveal that actual errors in coverage probability at moderate sample sizes depend strongly on the shape of F and may be considerable.

Specific references for this section are Bickel and Freedman (1981), Beran and Srivastava (1985) and Beran (1987, 1988b).

3. BOOTSTRAP SIMULTANEOUS CONFIDENCE SETS

The two-step bootstrap, to be distinguished from double bootstrap, is an intuitive technique that greatly extends the Tukey and Scheffé methods for constructing simultaneous confidence sets. As well, it achieves tighter asymptotic control of coverage probabilities than do constructions based on the Bonferroni inequality. Retaining the notation of Section 2, suppose that the parameter $\tau = T(\theta)$ has *components* $\{\tau_u = T_u(\theta) : u \in U\}$. The index set U is a metric space. For each u, let C_u denote a confidence set for the component τ_u. By simultaneously asserting the confidence sets $\{C_u : u \in U\}$, we obtain a simultaneous confidence set C for the components $\{\tau_u\}$.

We assume that the components $\{\tau_u\}$ are logically similar. The problem is to construct the confidence sets $\{C_u\}$ in such a way that

$$P_{\theta,n}[C_u \ni \tau_u] \quad \text{is the same } \forall\, u \in U \tag{9}$$

and

$$P_{\theta,n}[C_u \ni \tau_u \;\forall\, u \in U] = P_{\theta,n}[C \ni \tau] = \beta. \tag{10}$$

Property (9) is called *balance*. It reflects our wish that the confidence set C treat the logically similar components τ_u in an even-handed way while controlling the simultaneous coverage probability (10).

EXAMPLE 3.1 (Scheffé's method). Suppose that the sample of size n has an $N(D\gamma, \sigma^2 I_n)$ distribution, where the vector γ is $p \times 1$ and the matrix D has rank p. The unknown parameter $\theta = (\gamma, \sigma^2)$ is estimated by $\hat{\theta} = (\hat{\gamma}, \hat{\sigma}^2)$ from least squares theory. Let U be a subspace of $\mathbb{R}^p$ that has dimension q. The components of interest are the linear combinations $\{\tau_u = u'\gamma : u \in U\}$. Let $\hat{\sigma}_u^2 = u'(D'D)^{-1}u\hat{\sigma}^2$. Scheffé's simultaneous confidence set C_S for the linear combinations $\{u'\gamma\}$ is the simultaneous assertion of the component confidence intervals

$$\begin{aligned} C_{S,u} = \{u'\gamma : |u'(\hat{\gamma} - \gamma)|/\hat{\sigma}_u \\ \le q^{1/2}\sigma F_{q,n-p}^{1/2}(\beta)\}, \quad u \in U. \end{aligned} \tag{11}$$

The coverage probabilities of $\{C_{S,u}\}$ are clearly equal, while the overall coverage probability of C_S is β (Miller, 1966, Chapter 2). The F-distribution theory that underlies (11) breaks down if U is not a subspace or if the normal model is not pertinent.

Consider the general situation where $R_{n,u} = R_{n,u}(X_n, \tau_u)$ is a root for the component τ_u. Let $\mathcal{T}_u$ and $\mathcal{T}$ denote, respectively, the ranges of $\tau_u = T_u(\theta)$ and $\tau = T(\theta)$. Every point in $\mathcal{T}$ can be written in the component form $t = \{t_u : u \in U\}$. The simultaneous confidence sets to be considered are

$$C = \{t \in \mathcal{T} : R_{n,u}(X_n, t_u) \le c_u(\beta)\ \forall u \in U\}. \tag{12}$$

The problem is to devise critical values $\{c_u(\beta)\}$ so that, to a satisfactory approximation, C is balanced and has simultaneous coverage probability β for $\{\tau_u\}$.

Let $H_{n,u}(\cdot, \theta)$ and $H_n(\cdot, \theta)$ denote the left-continuous cumulative distribution functions of $R_{n,u}$ and of $\sup_{u\in U} H_{n,u}(R_{n,u}, \theta)$, respectively. If θ were known and the two c.d.f.s just defined were continuous in their first arguments, an oracular choice of critical values for the component confidence sets would be $c_u(\beta) = H_{n,u}^{-1}[H_n^{-1}(\beta, \theta), \theta]$. The oracular component confidence set

$$\begin{aligned} C_u &= \{t_u \in \mathcal{T}_u : R_{n,u}(X_n, t_u) \le c_u(\beta)\} \\ &= \{t_u \in \mathcal{T}_u : H_{n,u}(R_{n,u}, \theta) \le H_n^{-1}(\beta, \theta)\} \end{aligned} \tag{13}$$

has coverage probability $H_n^{-1}(\beta, \theta)$ for τ_u. The oracular simultaneous confidence set C, defined through (12), has coverage probability β for τ by definition of $H_n(\cdot, \theta)$.

To approximate this oracle construction in bootstrap terms is straightforward. Suppose that $\hat{\theta}_n$ is a consistent estimator of θ. The bootstrap estimators of $H_{n,u}(\cdot, \theta)$ and $H_n(\cdot, \theta)$ are, respectively, $\hat{H}_{\mathrm{B},u} = H_{n,u}(\cdot, \hat{\theta})$ and $\hat{H}_{\mathrm{B}} = H_n(\cdot, \hat{\theta})$. Define the bootstrap critical values

$$\hat{c}_{\mathrm{B},u}(\beta) = \hat{H}_{\mathrm{B},u}^{-1}[\hat{H}_{\mathrm{B}}^{-1}(\beta)]. \tag{14}$$

Define a bootstrap confidence set for τ_u by

$$C_{\mathrm{B},u} = \{t_u \in \mathcal{T}_u : R_{n,u}(X_n, t_u) \le \hat{c}_{\mathrm{B},u}(\beta)\}. \tag{15}$$

Simultaneously asserting these confidence sets generates the following bootstrap simultaneous confidence set for the $\{\tau_u\}$:

$$C_{\mathrm{B}} = \{t \in \mathcal{T} : R_{n,u}(X_n, t_u) \le \hat{c}_{\mathrm{B},u}(\beta)\ \forall u \in U\}. \tag{16}$$

The definition of the simultaneous confidence set C_{B} involves only the first bootstrap world. Indeed, let $\hat{\tau}_{n,u} = T_u(\hat{\theta}_n)$. Then $\hat{H}_{\mathrm{B},u}$ and $\hat{H}_{\mathrm{B}}$ are just the conditional distributions, given X_n, of $R_{n,u}(X_n^*, \hat{\tau}_{n,u})$ and of $\sup_{u\in U} H_{n,u}[R_{n,u}(X_n^*, \hat{\tau}_{n,u}), \hat{\theta}_n]$. Thus, a Monte Carlo approximation to the bootstrap critical values (14) requires only one round of bootstrap sampling. Computation of the supremum over U may require further approximations when the cardinality of U is not finite. In practice, the case of a finite number of components $\{\tau_u\}$ is both approachable and important.

The following template outlines a way to check whether C_{B} has correct asymptotic coverage probability for $\tau = \{\tau_u\}$ and whether it is asymptotically balanced. Suppose that Θ is an open subset of a metric space with metric d and that U is also a metric space. Let $\mathcal{C}(U)$ be the set of all continuous bounded functions on U, metrized by supremum norm. Assume that $\mathcal{C}(U)$ is topologically complete and that the processes $R_n(X_n, \theta) = \{R_n(X_n, T_u(\theta)) : u \in U, n \ge 1\}$ indexed by u have sample paths in $\mathcal{C}(U)$. We observe that boundedness of the sample paths can be assured by making a continuous, bounded and strictly monotone transformation of $\{R_{n,u}\}$ and that this transformation of the roots does not affect the bootstrap simultaneous confidence set C_{B}.

TEMPLATE B. *Suppose that, for every $\theta \in \Theta$,*

(i) *$d(\hat{\theta}_n, \theta) \to 0$ in $P_{\theta,n}$ probability as $n \to \infty$.*

(ii) (Triangular array convergence.) *There exists a process $R(\theta) = \{R_u(\theta) : u \in U\}$ with sample paths in $\mathcal{C}(U)$ such that if $\lim_{n\to\infty} d(\theta_n, \theta) = 0$, then $\mathcal{L}[R_n(X_n, \theta_n)|P_{\theta_n,n}]$ converges weakly in $\mathcal{C}(U)$ to $\mathcal{L}[R(\theta)]$, which has separable support.*

(iii) *The family $\{H_u(x, \theta) : u \in U\}$, where $H_u(x, \theta)$ denotes the c.d.f. of $R_u(\theta)$, is equicontinuous in x. For every $\varepsilon > 0$ there exists $\delta > 0$ such that $\sup_{u\in U} |H_u(x, \theta) - H_u(y, \theta)| < \varepsilon$ whenever $|x - y| < \delta$.*

(iv) *The c.d.f. $H(x, \theta)$ of $\sup_{u\in U} H_u[R_u(\theta), \theta]$ is continuous and strictly monotone in x. Then, for $\beta \in (0, 1)$,*

$$\lim_{n\to\infty} P_{\theta,n}[C_{\mathrm{B}} \ni \tau_u] = \beta \tag{17}$$

and

$$\lim_{n\to\infty} \sup_{u\in U} |P_{\theta,n}[C_{\mathrm{B},u} \ni \tau_u] - H^{-1}(\beta, \theta)| = 0. \tag{18}$$

Moreover, $\hat{H}_{\mathrm{B}}^{-1}(\beta)$ converges in $P_{\theta,n}$ probability to $H^{-1}(\beta, \theta)$.

If the conditions for Template B are met, C_{B} has asymptotic coverage probability β, C_{B} is asymptotically balanced and each $C_{\mathrm{B},u}$ has coverage probability that is estimated asymptotically by $\hat{H}_{\mathrm{B}}^{-1}(\beta)$. Using this template may require skillful choice of the metrics on Θ and U. The following two examples illustrate the scope of bootstrap simultaneous set C_{B}.

EXAMPLE 3.1 (Continued). If U is a q-dimensional subspace of $\mathbb{R}^p$, then the bootstrap confidence set C_{B} coincides with Scheffé's exact construction described earlier. This simplification occurs because $R_{n,u}$ and R_n turn out to be pivots. By Template B, simultaneous confidence set C_{B} remains asymptotically valid if U is any closed subset of $\mathbb{R}^p$, including the important case of a finite set. Template B continues to apply when the linear model errors are i.i.d. with mean zero and finite variance, their distribution P being unknown. In this context, $\theta = (\gamma, P)$ and $\hat{\theta} = (\hat{\gamma}, \hat{P})$, where $\hat{P}$ is the empirical distribution of the residuals, recentered to have average 0.

EXAMPLE 3.2 (Simultaneous confidence cones for mean directions). The sample consists of r independent directional subsamples of sizes $\{n_u : 1 \le u \le r\}$. The observations in subsample u are i.i.d. unit vectors in $\mathbb{R}^3$ drawn from an unknown spherical distribution P_u. The mean direction of P_u is the unit vector $m(P_u) = \mu(P_u)/|\mu(P_u)|$, where $\mu(P_u)$ is the expectation of P_u, assumed to be a nonnull vector. We wish to construct simultaneous confidence cones for the r mean directions $\{m(P_u) : 1 \le u \le r\}$. Here $\theta = (P_1, P_2, \ldots, P_r)$, $\hat{\theta}$ is the corresponding set of empirical distributions, $U = \{1, 2, \ldots, r\}$ and $\tau_u = m(P_u)$. Let $\hat{m}_u$ denote the resultant of the uth subsample, normalized to unit length. Suitable roots are $R_{n,u} = n_u[1 - \hat{m}_u' m(P_u)]$ for $1 \le u \le r$. Template B is applicable. Confidence set C_{B} consists of r simultaneous confidence cones for the mean directions $\{m(P_u)\}$. The half angle of each cone varies so as to achieve asymptotic balance.

The ability of C_{B} to handle implicitly the asymptotic distribution theory needed for balanced simultaneous confidence sets is just the start of the story. At finite sample sizes, simultaneous confidence set C_{B} usually suffers error in its overall coverage probability and lack of balance among the marginal coverage probabilities of its constituent confidence sets $\{C_{\mathrm{B},u}\}$. A suitable use of prepivoting can reduce *both* types of error asymptotically—a task that would rarely be possible by direct analytical approaches. Let

$$S_{n,u} = \hat{H}_{\mathrm{B}}[\hat{H}_{\mathrm{B},u}(R_{n,u})] \tag{19}$$

and apply the construction of C_{B} to the transformed roots $\{S_{n,u}\}$ to obtain the simultaneous confidence set $C_{\mathrm{B},1}$.

In greater detail, let $K_{n,u}(\cdot, \theta)$ and $K_n(\cdot, \theta)$ be the left-continuous c.d.f.s of $S_{n,u}$ and $\sup_{u \in U} K_{n,u}(S_{n,u}, \theta)$, respectively. The corresponding bootstrap estimates of these c.d.f.s are $\hat{K}_{\mathrm{B},u} = K_{n,u}(\cdot, \hat{\theta}_n)$ and $\hat{K}_{\mathrm{B}} = K_n(\cdot, \hat{\theta}_n)$. Confidence set $C_{\mathrm{B},1}$ has the form (16) with the critical values $\{\hat{c}_{\mathrm{B},u}(\beta)\}$ replaced by the refined critical values

$$\hat{c}_{\mathrm{B},1,u}(\beta) = \hat{H}_{\mathrm{B},u}^{-1}[\hat{H}_{\mathrm{B}}^{-1}[\hat{K}_{\mathrm{B},u}^{-1}[\hat{K}_{\mathrm{B}}^{-1}(\beta)]]]. \tag{20}$$

The interpretation of $\hat{H}_{\mathrm{B},u}$ and $\hat{H}_{\mathrm{B}}$ as conditional c.d.f.s in the first bootstrap world was discussed following (16). The second bootstrap world yields an analogous interpretation of $\hat{K}_{\mathrm{B},u}$ and $\hat{K}_{\mathrm{B}}$ that serves as a basis for Monte Carlo approximation of the double-bootstrap critical values in (20). Currently available asymptotic analysis of the refined simultaneous confidence set $C_{\mathrm{B},1}$ is mostly heuristic. Rigorous analysis in a few examples supports the promise of $C_{\mathrm{B},1}$ for reducing errors in balance and in overall coverage probability.

Specific references for this section are Beran (1988a, 1990) and Beran and Fisher (1998).

4. BOOTSTRAP TESTS

The scope of hypothesis testing is limited by the narrowness of its formulation. The abuse of testing theory in problems for which it is not designed, such as model selection, makes the point. Nevertheless, we consider briefly the impact of bootstrap distributions on the construction of tests.

The *test statistic* approach constructs a bootstrap null distribution for a chosen test statistic. In the setting of Section 2, let Θ_0 denote a subset of Θ. Consider the problem of testing the composite null hypothesis $\theta \in \Theta_0$ against the alternative that this is not so. The test statistic is $V_n = V_n(X_n)$, large values of V_n being evidence against the null hypothesis. For $\theta \in \Theta_0$, let $H_n(x, \theta) = P_{\theta,n}[V_n < x]$ denote the left-continuous c.d.f. of V_n. Let $\hat{\theta}_n$ be an estimator of Θ that is consistent under the null hypothesis and that has values restricted to Θ_0. The bootstrap null distribution is then $\hat{H}_{\mathrm{B}} = H(\cdot, \hat{\theta}_n)$ The bootstrap test of nominal level α rejects the null hypothesis if $V_n > \hat{H}_{\mathrm{B}}^{-1}(1 - \alpha)$.

Intuitively we expect that $\lim_{n \to \infty} P_{\theta,n}[V_n > \hat{H}_{\mathrm{B}}^{-1}(1 - \alpha)] = \alpha$ for every $\theta \in \Theta_0$. In this event, we say that the asymptotic rejection probability of the bootstrap test is α for every θ in null hypothesis set Θ_0. Template A with Θ replaced by Θ_0 provides a way to verify this property in particular examples. The condition that the possible values of $\hat{\theta}_n$ lie in Θ_0 is not needed for the argument under the null hypothesis, but is important for the power of the test.

EXAMPLE 4.1 (Minimum Kolmogorov distance tests). The sample consists of n i.i.d. random variables, each of which has unknown c.d.f. F. Under the null hypothesis, F belongs to a specified parametric family of c.d.f.s $\{G_\xi : \xi \in \Xi\}$, where Ξ is an open subset of $\mathbb{R}^k$. Let $\hat{F}_n$ be the empirical c.d.f. and let the test statistic be $V_n = n^{1/2} \inf_{\xi \in \Xi} \|\hat{F}_n - G_\xi\|$. In this problem, $\theta = F$ and $\hat{\theta}_n = G_{\hat{\xi}_n}$, where $\hat{\xi}_n$ minimizes $\|\hat{F}_n - G_\xi\|$ over $\xi \in \Xi$. The bootstrap test just outlined has asymptotic rejection probability α for every distribution F in the parametric null hypothesis. This bootstrap construction solves a problem that is difficult for analytical approaches. Moreover, it may be extended to minimum distance tests in higher dimensions that involve empirical measures on half-spaces or other Vapnik–Cervonenkis classes.

Outside the simplest settings, constructing a plausible bootstrap null distribution may encounter difficulties because there is more than one way to construct a fitted model that satisfies the composite null hypothesis. Example 4.2 below is an instance. Inverting a pertinent bootstrap confidence set offers a clearer path in such cases.

The *confidence set* approach considers the testing problem where, under the null hypothesis, $T(\theta) = \tau_0$ for given function T and specified value τ_0. We construct a bootstrap confidence set of nominal level $1 - \alpha$ for $\tau = T(\theta)$ using a root $R_n(X_n, \tau)$. The test rejects the null hypothesis if this bootstrap confidence set does not contain τ_0. Under Template A for bootstrap confidence sets, the asymptotic rejection probability of this bootstrap test is α for every θ that satisfies the null hypothesis. In this approach, the estimator $\hat{\theta}_n$ used to construct the bootstrap distribution is not required to take values in Θ_0—a substantial practical advantage in examples where restricting $\hat{\theta}_n$ to Θ_0 can be done in several ways.

EXAMPLE 4.2 (Testing equality of two mean directions). The sample consists of two independent subsamples, the observations in subsample j being n_j i.i.d. random unit vectors in $\mathbb{R}^3$ drawn from an unknown spherical distribution P_j. The problem is to test equality of the two mean directions $m(P_1)$ and $m(P_2)$, defined in Example 3.2. For unit column vectors d_1 and d_2 in $\mathbb{R}^3$, $\rho(d_1, d_2) = (d_1' d_2, (d_1 \times d_2)')'$ is a unit column vector in $\mathbb{R}^4$ that represents, in one-to-one manner, the rotation which takes d_1 into d_2 in the plane determined by those two vectors. Suppose that $n_1 = \lfloor n\lambda \rfloor$ and $n_2 = n - n_1$ with $0 < \lambda < 1$. Let $\tau = \rho(m(P_1), m(P_2))$ and $\hat{\tau}_n = \rho(\hat{m}_1, \hat{m}_2)$ using notation from Example 3.2. A suitable root is $R_n(X_n, \tau) = n(1 - \hat{\tau}_n' \tau)$. The test rejects if bootstrap confidence set C_B for τ contains the vector $(1, 0, 0, 0)'$ and has asymptotic rejection probability α under the null hypothesis that the two mean directions are equal.

The discussion in Section 2 of double bootstrapping may be extended to bootstrap tests of either variety. The discussion in Section 3 of simultaneous confidence sets has implications for simultaneous bootstrap tests, but these seem not to have been explored.

Specific references for this section are Beran (1986, 1988b), Beran and Millar (1986), Beran and Fisher (1998) and Bickel and Ren (2001).

5. CLASSICAL CHARACTER OF THE BOOTSTRAP

Bootstrap constructions of asymptotically valid confidence sets and tests are an important algorithmic success. A deep link between correct convergence of bootstrap distributions and the properties of classically regular estimators in regular statistical models strongly constrains the scope of this success. Investigation of this link returns us to the theoretical milieu in which the bootstrap arose.

Suppose that the sample X_n consists of n i.i.d. random variables that have joint distribution $P_{\theta,n}$. Here θ is an unknown element of parameter space Θ, which is an open subset of $\mathbb{R}^k$. Of interest is $T(\theta) \in \mathbb{R}^m$, where T is a differentiable function with $m \times k$ derivative matrix $\nabla T(\theta)$ of full rank. For any $\theta_0 \in \Theta$ and every $h \in \mathbb{R}^k$, let $L_n(h, \theta_0)$ denote the log-likelihood ratio of the absolutely continuous part of $P_{\theta_0 + n^{-1/2}h,n}$ with respect to $P_{\theta_0,n}$. The model is *locally asymptotically normal* (LAN) at θ_0 if there exists a random column vector $Y_n(\theta_0)$ and a nonsingular symmetric matrix $I(\theta_0)$ such that:

(a) Under $P_{\theta_0,n}$, $L_n(h_n, \theta_0) - h'Y_n(\theta_0) - 2^{-1}h' \cdot I(\theta_0)h = o_p(1)$ for every $h \in \mathbb{R}^k$ and every sequence $h_n \to h$;

(b) $\mathcal{L}[Y_n(\theta_0) | P_{\theta_0,n}] \Rightarrow N(0, I(\theta_0))$.

The LAN property is possessed by smoothly parametrized exponential families and other regular parametric models.

For any sequence of estimators $\{\hat{\tau}_n\}$ of $T(\theta)$, let $H_n(\theta)$ denote $\mathcal{L}[n^{1/2}(\hat{\tau}_n - T(\theta)) | P_{\theta,n}]$. The estimators $\{\hat{\tau}_n\}$ are *locally asymptotically equivariant* (LAE) at θ_0 if, for every $h \in \mathbb{R}^k$ and every sequence $h_n \to h$, $H_n(\theta_0 + n^{-1/2}h_n) \Rightarrow H(\theta_0)$.

Estimators $\{\hat{\tau}_{n,\mathrm{E}}\}$ that are classically efficient for $T(\theta_0)$ in a model that is LAN at θ_0 satisfy

$$\hat{\tau}_{n,\mathrm{E}} = T(\theta_0) + n^{-1/2} \nabla T(\theta_0) I^{-1}(\theta_0) Y_n(\theta_0) + o_p(1) \tag{21}$$

under $P_{\theta_0,n}$ and are LAE at θ_0. The distributions $\{\mathcal{L}[n^{1/2}(\hat{\tau}_{n,\mathrm{E}} - T(\theta_0)|P_{\theta_0}]\}$ converge weakly to the $N(0, \Sigma_T(\theta_0))$ distribution, where $\Sigma_T(\theta_0) = \nabla T(\theta_0) I^{-1}(\theta_0)[\nabla T(\theta_0)]'$. In estimation theory, the precise role of the estimators $\{\hat{\tau}_{n,\mathrm{E}}\}$, which include (possibly emended) MLEs, is revealed by Hájek's convolution theorem.

CONVOLUTION THEOREM. Let $K_n(\theta) = \mathcal{L}[(n^{1/2} \cdot (\hat{\tau}_n - \hat{\tau}_{n,\mathrm{E}}), Y_n(\theta))|P_{\theta,n}]$. Suppose that the model is LAN and that $H_n(\theta_0) \Rightarrow H(\theta_0)$ as $n \to \infty$. The following two statements are equivalent:

- The estimator sequence $\{\hat{\tau}_n\}$ is LAE at θ_0 with limit distribution $H(\theta_0)$.
- For every $h \in \mathbb{R}^k$ and every sequence $h_n \to h$, $K_n(\theta_0 + n^{-1/2}h_n) \Rightarrow D(\theta_0) \times N(0, I(\theta_0))$ for some distribution $D(\theta_0)$ such that $H(\theta_0) = D(\theta_0) * N(0, \Sigma_T(\theta_0))$.

The local asymptotics just described have global implications. Suppose that $H_n(\theta) \Rightarrow H(\theta)$ for every $\theta \in \Theta$. Then there exists a Lebesgue null set E in Θ such that the estimators $\{\hat{\tau}_n\}$ are LAE at every $\theta \in \Theta - E$. It follows that the limit distribution $H(\theta)$ must have convolution structure for almost every $\theta \in \Theta$. Moreover, if $H(\theta)$ and $\Sigma_T(\theta)$ are both continuous in θ, the former in the topology of weak convergence, then $H(\theta)$ has convolution structure for *every* θ. This is often the case for classical parametric estimators of the maximum likelihood or minimum distance type. These considerations surrounding the convolution theorem have an elegant consequence that concerns superefficiency. For loss function $w[n^{1/2}(\hat{\tau}_n - T(\theta_0))]$, where w is any symmetric, subconvex, continuous and nonnegative function, no estimator sequence $\{\hat{\tau}_n\}$ can have smaller asymptotic risk at θ_0 than $\{\hat{\tau}_{n,\mathrm{E}}\}$ unless θ_0 belongs in the null set of non-LAE points of $\{\hat{\tau}_n\}$.

A deep connection exists between the convolution theorem and correct convergence of bootstrap distributions. Let $\{\hat{\theta}_n\}$ be estimators of θ. The parametric bootstrap estimator of distribution $H_n(\theta)$ is $\hat{H}_{\mathrm{B}} = H_n(\hat{\theta}_n)$. Let $J_n(\theta)$ denote $\mathcal{L}[n^{1/2}(\hat{\theta}_n - \theta)|P_{\theta,n}]$.

BOOTSTRAP CONVERGENCE THEOREM. *Assume the hypotheses for the convolution theorem. As $n \to \infty$, suppose that $J_n(\theta_0) \Rightarrow J(\theta_0)$, a limit distribution that has full support on $\mathbb{R}^k$. The following two statements are then equivalent*:

- *The estimator sequence $\{\hat{\tau}_n\}$ is LAE at θ_0 with limit distribution $H(\theta_0)$.*
- *$\hat{H}_{\mathrm{B}} \Rightarrow H(\theta_0)$ in $P_{\theta_0,n}$ probability as $n \to \infty$.*

The practical significance of this result is that ordinary bootstrapping will not work as desired at parameter values where the estimator being bootstrapped is not classically regular. Important instances of irregular estimators are Stein shrinkage estimators and their relatives, which include adaptive penalized least squares estimators and adaptive model-selection estimators. To deal with possibly non-LAE estimators requires modifications of the bootstrap idea. Among these are the m out of n bootstrap, subsampling and the multiparametric bootstrap.

Specific references for this section are Le Cam and Yang (1990), Beran (1997), Bickel and Freedman (1981), Politis, Romano and Wolf (1999) and Beran and Dümbgen (1998).

ACKNOWLEDGMENT

This research was supported in part by National Science Foundation Grant DMS-03-00806.

REFERENCES

BERAN, R. (1986). Simulated power functions. *Ann. Statist.* **14** 151–173.

BERAN, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74** 457–468.

BERAN, R. (1988a). Balanced simultaneous confidence sets. *J. Amer. Statist. Assoc.* **83** 679–686.

BERAN, R. (1988b). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *J. Amer. Statist. Assoc.* **83** 687–697.

BERAN, R. (1990). Refining bootstrap simultaneous confidence sets. *J. Amer. Statist. Assoc.* **85** 417–426.

BERAN, R. (1997). Diagnosing bootstrap success. *Ann. Inst. Statist. Math.* **49** 1–24.

BERAN, R. and DUCHARME, G. (1991). *Asymptotic Theory for Bootstrap Methods in Statistics.* Les Publications CRM, Univ. Montréal.

BERAN, R. and DÜMBGEN, L. (1998). Modulation of estimators and confidence sets. *Ann. Statist.* **26** 1826–1856.

BERAN, R. and FISHER, N. I. (1998). Nonparametric comparison of mean directions or mean axes. *Ann. Statist.* **26** 472–493.

BERAN, R. and MILLAR, P. W. (1986). Confidence sets for a multivariate distribution. *Ann. Statist.* **14** 431–443.

BERAN, R. and SRIVASTAVA, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *Ann. Statist.* **13** 95–115.

BERLINSKI, D. (2001). *The Advent of the Algorithm.* Harcourt, New York.

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

BICKEL, P. J. and REN, J.-J. (2001). The bootstrap in hypothesis testing. In *State of the Art in Probability and Statistics*: *A Festschrift for Willem R. van Zwet* 91–112. IMS, Hayward, CA.

BRILLINGER, D. R. and TUKEY, J. W. (1984). Spectrum analysis in the presence of noise: Some issues and examples. In *The Collected Works of John W. Tukey* (D. R. Brillinger, ed.) **2** 1001–1141. Wadsworth, Monterey, CA.

DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application.* Cambridge Univ. Press.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

EFRON, B. and TIBSHIRANI, R. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, New York.

FISHER, R. A. (1930). *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh.

FISHER, R. A. (1956). *Statistical Methods and Scientific Inference*. Hafner, New York.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer, New York.

KNUTH, D. E. (1969). *The Art of Computer Programming* **2**. Addison–Wesley, Reading, MA.

LE CAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics*: *Some Basic Concepts.* Springer, New York.

MAMMEN, E. (1992). *When Does Bootstrap Work*? *Asymptotic Results and Simulations. Lecture Notes in Statist.* **77**. Springer, New York.

MILLER, R. (1966). *Simultaneous Statistical Inference.* McGraw-Hill, New York.

POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. Springer, New York.

QUENOUILLE, M. H. (1959). *Rapid Statistical Calculations*: *A Collection of Distribution-Free and Easy Methods of Estimation and Testing*. Griffin, London.

SAVAGE, L. J. (1954). *The Foundations of Statistics*. Wiley, New York.

STUDENT (1908a). The probable error of a mean. *Biometrika* **6** 1–25.

STUDENT (1908b). Probable error of a correlation coefficient. *Biometrika* **6** 302–310.

TUKEY, J. W. (1970). *Exploratory Data Analysis* (limited preliminary edition). Addison–Wesley, Reading, MA. (Regular edition published in 1977.)

WALD, A. (1950). *Statistical Decision Functions*. Wiley, New York.