RM696

This Chapter is dedicated to the memory of Kesar Singh (1955–2012), pioneer of the mathematical foundations of bootstrap resampling.

[1]

THE SEQUENTIAL BOOTSTRAP

P.K. Pathak and C.R. Rao

July 26, 2012

Abstract

The main focus of this article is to carefully examine the information content of resampling techniques in bootstrap from a survey sampling point of view. Given an observed sample of size n, resampling for bootstrap involves n repeated trials of simple random sampling with replacement (SRSWR). It is well-known that SRSWR does not result in samples that are equally informative (Pathak, 1964). In 1997, Rao, Pathak and Koltchinskii introduced a sequential bootstrap resampling scheme in the literature, stemming from the observation made by Efron (1983) that the usual bootstrap samples are supported on average on approximately $0.632n$ of the original data points. The primary goal of our sequential bootstrap was to stablize the information content of bootstrapped samples as well as to investigate whether consistency and second-order correctness of Efron's simple bootstrap carried over to this sequential case. In Rao, Pathak and Koltchinskii (1997), we showed that the main empirical characteristics of the sequential bootstrap asymptotically differ from those of the corresponding simple bootstrap by a magnitude of $O(n^{-3/4})$. In Babu, Pathak and Rao (1999), we established the second-order correctness of the sequential bootstrap for a general class of

*C.R. Rao is Professor Emeritus, Department of Statistics, Pennsylvania State University, University Park, PA, 16802, and Research Professor, University at Buffalo, Buffalo, NY, 14260; P.K. Pathak is Professor, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824.

estimators under regularity conditions analogous to those for the simple bootstrap. In a separate paper, Shoemaker and Pathak (2001) carried out simulations similar to those of Mammen (1992), confirming the consistency of the sequential bootstrap. More recently, Jiménez-Gamero et al. (2006), Pino-Mejías et al. (2010), Pauly (2011) and others have investigated the sequential bootstrap and its variants both empirically as well as theoretically. These investigations indicate that the sequential bootstrap is likely to perform better than the simple bootstrap in small to moderate samples. We present an expositary account of these findings and discuss their potential applications.

KEY WORDS: bootstrap, sequential resampling, poisson sampling, information content, sampling viewpoint, asymptotic correctness, empirical measures, empirical processes, weak convergence, quantiles

1. Introduction

The bootstrap is a ubiquitous technique to estimate performance characteristics of complex statistical methods which often do not admit simple closed form solutions. Its wide-spread use in statistics was initiated by Bradley Efron (1979), its rigorous mathematical foundations were established by Kesar Singh (1981) and its underlying principles have their origins in randomization, sample surveys and Jackknife. In sample surveys terminology, its paradigm is akin to Mahalanobis' interpenetrating subsampling, as well as two-phase sampling (Hall, 2003). As an illustration, a two-phase sampling version of Efron's simple bootstrap can be paraphrased as follows. Consider a target population of finitely many units and for simplicity assume that its units take values in the real line. Let F, μ, and σ^2 respectively denote this population's distribution function, mean, and variance. Consider the problem of estimating these parameters. Let $\mathcal{S} = (X_1, X_2, \ldots, X_n)$ be a random sample from this population. It is easily seen that δ_{X_1} is an unbiased estimator of F with δ_x being the unit probability function at x. Similarly X_1 is an unbiased estimator of μ, and $(X_1 - X_2)^2/2$ is an unbiased estimator of σ^2. These estimators are evidently poor estimators. Given the observed sample $\mathcal{S}$, the Rao-Blackwellization of these estimators yields better estimators. These are:

$$F_n := E(\delta_{X_1}|\mathcal{S}) = n^{-1} \sum_{1 \le i \le n} \delta_{X_i} \tag{1.1}$$

$$\bar{X}_n := E(X_1|\mathcal{S}) = n^{-1} \sum_{1 \le i \le n} X_i \tag{1.2}$$

$$s_n^2 := E((X_1 - X_2)^2/2|\mathcal{S}) = (n-1)^{-1} \sum_{1 \le i \le n} (X_i - \bar{X}_n)^2 \tag{1.3}$$

The respective sampling errors of these estimators are:

$$(F_n - F),\ (\bar{X}_n - \mu),\ (s_n^2 - \sigma^2) \tag{1.4}$$

The simple bootstrap is a resampling technique for estimating such sampling errors and their distributions purely from the observed sample $\mathcal{S}$. In survey sampling terminology, given $\mathcal{S}$, this entails a second phase of sampling in which one selects a with replacement simple random sample of size n from the sample $\mathcal{S}$ from the first phase, resulting in the second phase sample:

$$\hat{\mathcal{S}} := (\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n) \tag{1.5}$$

The survey sampling paradigm for two-phase sampling is to consider $\mathcal{S}$ as a surrogate target population and the second-phase sample $\hat{\mathcal{S}}$ as the surrogate sample $\mathcal{S}$. In terms of the surrogate target population $\mathcal{S}$ and the surrogate sample $\hat{\mathcal{S}}$, the estimators for the sampling errors in the preceding equations are given by:

$$(\hat{F}_n - F_n),\ (\bar{X}_{\hat{n}} - \bar{X}_n),\ (\hat{s}_n^2 - s_n^2) \tag{1.6}$$

in which $\hat{F}_n$, $\bar{X}_{\hat{n}}$, $\hat{s}_n^2$ are respectively the emprical distribution, mean and variance based on the surrogate sample $\hat{\mathcal{S}}$.
The conditional distributions of these statistics given $\mathcal{S}$ furnish estimates of the sampling distribution of the corresponding statistics from phase one sampling.

Efron used repeated resamplings a large number of times (B) at the second phase to empirically estimate these distributions and referred to his method as "bootstrap resampling". The main purpose of our illustration is to point out that Efron's method has a great potential for extending his paradigm to more complex non-IID scenarios under the two-phase sampling framework with Rao-Blackwellization. For example, if the first-phase sampling is simple random sampling (without replacement) of size n from a population of size N then for the bootstrap sampling to go through, the second phase sampling needs to ensure that the surrogate image $\hat{F}$ of the underlynig population F is such that for each unit in $\hat{F}$, the probability of its inclusion in the second phase is $\approx (n/N)$ and for each pair of units in $\hat{F}$, its probability of inclusion is $\approx n(n-1)/N(N-1)$.

What bootstrap is: There are numerous versions of the bootstrap. It has been studied for virtually all forms of statistical inference, including point and interval estimations, tests of hypothesis, classification models, and cross-validation (Chernick, 2008). For readers unfamiliar with the bootstrap, a brief outline of Efron's simple bootstrap in the context of variance estimation is as follows:

1. Let F be a given population (distribution function) and let $\mu(F)$ be a given parameter to be estimated on the basis of an observed sample $\mathcal{S} := (X_1, X_2, \ldots, X_n)$ drawn from the population F. Suppose that the plug-in estimater $\mu(F_n) = T(X_1, X_2, \ldots, X_n)$, say, is used to estimate $\mu(F)$, in which F_n is the empirical distribution based on $\mathcal{S}$. Now to estimate the variance of $\mu(F_n)$, do

2. Draw observations $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n$ from $\mathcal{S}$ by simple random sampling with replacement (SRSWR), i.e. $(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n) \sim F_n$

3. Compute $\hat{T}_n = T(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n)$

4. Repeat Steps 2 and 3, B times, say 1000, and let $(\hat{T}_{n,1}, \hat{T}_{n,2}, \ldots, \hat{T}_{n,B})$ denote the respective estimates so computed.

5. The bootstrap estimate of the variance of $\mu(F_n)$ is:

$$v_{boot} = \frac{1}{B} \sum_{j=1}^{B} (\hat{T}_{n,j} - \frac{1}{B} \sum_{k=1}^{B} \hat{T}_{n,k})^2$$

We turn now to examine Efron's simple bootstrap in some detail from the viewpoint of information content. We will continue to use the notations introduced in preceding discussions and suppose that the plug-in estimator $\mu(F_n)$ is to be used as an estimator of $\mu(F)$.

Then, for a large class of functions μ, Efron's bootstrap resampling method (1979) provides a robust method of evaluating the performance characteristics of $\mu(F_n)$, solely on the basis of information derived (through randomization) from the observed sample $\mathcal{S}$. Consider the simple case when $\mu(F)$ is the population mean, $\mu(F_n)$ the corresponding sample mean, and consider the pivot.

$$\Pi_n = \sqrt{n}(\mu(F_n) - \mu(F))/\sqrt{\sigma^2(F_n)} \tag{1.7}$$

in which $\sigma^2(F_n)$ is the sample variance based on F_n.

The central limit theorem entails that the sampling distribution of Π_n can be approximated by the standard normal distribution. On the other hand, the bootstrap furnishes an alternative approach based on resampling to estimating the sampling distribution of Π_n from $\mathcal{S}$ more precisely. Generally speaking, the central limit approximation is accurate to $o(1)$, while resampling approximation is accurate to $o(n^{-1/2})$. For example, let G_n denote the distribution function of Π_n. Then the central limit theorem yields that

$$\|G_n - \Phi\| := \sup_x |G_n(x) - \Phi(x)| \quad = o(1) \tag{1.8}$$

where $\Phi(x)$ denotes the standard normal distribution function. The bootstrap approximation captures the skewness of the distribution G_n in the following sense:

$$\sqrt{n}\|G_n - H_n\| = o(1) \tag{1.9}$$

in which H_n represents a one-term Edgeworth expansion for G_n.

The simple bootstrap resampling scheme to approximating the distribution G_n is based on resampling by simple random sampling with replacement (SRSWR). Given the observed sample $\mathcal{S}$, let $\hat{\mathcal{S}}_n = (\hat{X}_1, \ldots, \hat{X}_n)$ be an SRSWR sample drawn from $\mathcal{S}$ in this manner. Let $\hat{F}_n$ be the empirical distribution based on $\hat{\mathcal{S}}_n$. Let

$$\hat{\Pi}_n = \sqrt{n}(\mu(\hat{F}_n) - \mu(F_n))/\sqrt{\sigma^2(\hat{F}_n)} \quad = \sqrt{n}(\hat{\mu}_n - \mu_n)/\hat{\sigma}_n, \quad \text{say,} \tag{1.10}$$

denote the pivot based on $\hat{\mathcal{S}}_n$. Then for large n, the conditional distribution of $\hat{\Pi}_n$ given $\mathcal{S}$ is close to that of Π_n. In practice, this conditional distribution of $\hat{\Pi}_n$ is simulated by repeated resampling from $\mathcal{S}$ by SRSWR of size n, a large number of times usually of order $O(1000)$. This observed frequency distribution (ensemble) of $\hat{\Pi}_n$ is referred to as the bootstrap distribution of the pivot Π_n. Thus for example, for large n, the quantiles from the bootstrap distribution of Π_n can be used to set up a confidence interval for μ based on the pivot Π_n.

Owing to the with replacement nature of SRSWR, not all of the observations in bootstrap samples $\hat{\mathcal{S}}_n$, say, are based on distinct units from $\mathcal{S}$. In fact, the information content of $\hat{\mathcal{S}}_n$, the set of observations from distinct units in $\hat{\mathcal{S}}_n$, is a random variable. Let ν_n denote the number of observations in $\hat{\mathcal{S}}_n$ that are based on distinct units from $\mathcal{S}$. Then

$$E(\nu_n) = n[1 - (1 - \frac{1}{n})^n] \simeq n(1 - e^{-1}) \simeq n(0.632) \tag{1.11}$$

$$\mathrm{Var}(\nu_n) = \mathrm{Var}(n - \nu_n) = n(1 - \frac{1}{n})^n + n(n-1)(1 - \frac{2}{n})^n - n^2(1 - \frac{1}{n})^{2n} \tag{1.12}$$

$$\simeq ne^{-1}(1 - e^{-1}) \tag{1.13}$$

In fact, the distribution of ν_n approaches a binomial distribution $b(n, p)$ with $p \simeq 0.632$, showing that the information content of a bootstrap sample, is approximately sixty-three percent of the information content of the observed sample $\mathcal{S}$. In what follows, we describe alternatives to the simple bootstrap that keep the information content of the bootstrap samples more stable.

2. A Sequential Bootstrap Resampling Scheme

Under this resampling scheme, we keep the information content of each bootstrap sample constant by requiring that the number of distinct observations in each sample be kept constant, equal to a preassigned value $\approx (0.632)n$ as follows:

To select a bootstrap sample, draw observations from $\mathcal{S}$ sequentially by SRSWR until there are $(m+1) \approx n(1-e^{-1})+1$ distinct observations in the observed bootstrap sample, the last observation in the bootstrap sample is discarded to ensure simplicity in technical details.

Thus an observed bootstrap sample has the form:

$$\hat{\mathcal{S}}_N = (\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_N) \tag{2.1}$$

in which $\hat{X}_1, \ldots, \hat{X}_N$ have $m \approx n(1-e^{-1})$ distinct observations from $\mathcal{S}$. The number of distinct observations in $\hat{\mathcal{S}}_N$ is precisely $[n(1-e^{-1})]$; it is no longer a random variable. The sample size N is a random variable with $E(N) \approx n$. The pivot based on $\hat{\mathcal{S}}_N$ is

$$\hat{\Pi}_N = \sqrt{N}(\mu(\hat{F}_N) - \mu(F_n)) \tag{2.2}$$

in which $\hat{F}_N$ denotes the empirical distribution based on the bootstrap sample $\hat{\mathcal{S}}_N$. For simplicity in exposition, we assume $\sigma^2(F) = 1$. Now for a comparative study of the pivot $\hat{\Pi}_N$ based on the sequential resampling approach, versus $\hat{\Pi}_n$ based on a bootstrap sample of fixed size n, it is necessary to estimate the order of magnitude of the random variable N. It is easily seen that N admits the following decomposition in terms of independent (geometric) random variables:

$$N = I_1 + I_2 + \ldots + I_m \tag{2.3}$$

in which $m = [n(1-e^{-1})]$, $I_1 = 1$ and for each k, $2 \leq k \leq m$, and for $j \geq 1$:

$$P(I_k = j) = (1 - \frac{k-1}{n})(\frac{k-1}{n})^{j-1} \tag{2.4}$$

Therefore

$$E(N) = n[\frac{1}{n} + \frac{1}{(n-1)} + \ldots + \frac{1}{(n-m+1)}] = n + O(1) \tag{2.5}$$

since $m = [n(1 - e^{-1})]$

Similarly, it is easily seen that

$$\text{Var}(N) = \sum_{k=1}^{m} \frac{n(k-1)}{(n-k+1)^2} = n(e-1) + O(1) \tag{2.6}$$

Thus

$$\frac{E(N-n)^2}{n^2} = \frac{(e-1)}{n} + O(\frac{1}{n^2}) \tag{2.7}$$

showing that $(N/n) \to 1$ in probability.

Further analysis similar to that of Mitra and Pathak (1984) can be used to show that

$$\frac{E(\hat{\Pi}_N - \hat{\Pi}_n)^2}{\text{Var}(\Pi_n)} \leq K\sqrt{\frac{\text{Var}(N)}{n^2}} = O(\frac{1}{\sqrt{n}}) \tag{2.8}$$

This implies that $\hat{\Pi}_n$ and $\hat{\Pi}_N$ are asymptotically equivalent. The preceding computations show that this sequential resampling plan, in addition to keeping the information content of bootstrap samples constant, also preserves its asymptotic correctness (consistency). In fact, our investigations show that under sequential forms of resampling, asymptotic correctness of the bootstrap procedure goes through only if the number m of distinct units in bootstrap samples satisfies: $m = n(1 - e^{-1}) + o(n)$. Consequently any other sequential resampling procedure for which first-order asymptotics goes through cannot be "significantly" different from ours in terms of its information content, e.g. consistency of variants of our sequential bootstrap by Jiménez-Gamero et al. (2004, 2006), Pino-Mejías et al. (2010) and Pauly (2011) is a consequence of this regularity condition. Incidentally, the measure of disparity in two estimators as given by (2.8) is a simple tool frequently used in sampling theory to establish asymptotic equivalence of two competing estimators.

A second approach for consistency of the sequential bootstrap can be based on the so-called Mallows metric for weak convergence (cf Bickel and Friedman, 1981): In a given class of distributions with a finite second moment, define a given sequence of distributions $\{F_n\}$ to converge in M-sense to F if and only if (a) F_n converges weakly to F and (b) $\int x^2 dF_n$ converges to $\int x^2 dF$. It is easily seen that this notion of convergence is induced by the following M-metric:

$$d^2(F, G) = \inf_{X \sim F, Y \sim G} E(X - Y)^2 \tag{2.9}$$

in which the infimum is taken over all pairs of random variables (X, Y) with given marginals F and G respectively (Mallows, 1972).

Under this metric, (2.9), it is easily shown that

$$d(\hat{\Pi}_n, \Pi_n) \leq d(F_n, F) \tag{2.10}$$

so that

$$d(\hat{\Pi}_n, \Phi) \leq d(\hat{\Pi}_n, \Pi_n) + d(\Pi_n, \Phi) \leq d(F_n, F) + d(\Pi_n, \Phi) \tag{2.11}$$

The first term on the right converges to zero by the law of large numbers (the Glivenko-Cantelli lemma), while the second term does so by the central limit theorem. This establishes the consistency of the simple bootstrap. An added complication that arises in sequential resampling scheme is that the bootstrap sample size N is now a random variable. Thus for the consistency of the sequential bootstrap to go through, one needs to show that $d(\hat{\Pi}_N, \Phi)$ can be made arbitrarily small; (2.11) is no longer directly applicable. Nevertheless, based on techniques similar to those of Pathak (1964) and Mitra and Pathak (1984) (Lemma 3.1 in Mitra and Pathak), it can be shown that:

$$d(\hat{\Pi}_N, \hat{\Pi}_n) = O(n^{-1/4}) \tag{2.12}$$

Consistency of the pivot $\hat{\Pi}_N$ follows from (2.11) and (2.12) and the triangle inequality. A limitation of the preceding two approaches is that they apply only to linear statistics and cannot be easily extended to more general statistics. A third and a more general approach is to treat pivots like $\hat{\Pi}_n$ and $\hat{\Pi}_N$ as random signed measures and study their convergence in the functional sense. In this functional setting, we now describe the key result which furnishes a rigorous justification of the sequential bootstrap for a large class of functionals.

3. Bootstrapping Empirical Measures With A Random Sample Size

For simplicity in exposition, we tacitly assume at the outset the existence of a suitable probability space where the random variables, functions etc. under study are well-defined (see Rao, Pathak and Koltchinskii, 1997). Now consider $\{X_n : n \geq 1\}$ a sequence of independent random elements with a common distribution P in a certain given space χ. Let P_n denote the empirical measure based on the sample $(X_1, \ldots, X_n)$ from P, i.e.,

$$P_n := n^{-1} \sum_{1 \leq i \leq n} \delta_{X_i} \tag{3.1}$$

with δ_x being the unit point mass at $x \in X$.

Let $\{\hat{X}_{n,j} : j \geq 1\}$ be a sequence of independent random elements with common distribution P_n. We refer to this sequence as a sequence of bootstrapped observations, or just a bootstrapped sequence. Given a number $N \geq 1$, let $\hat{P}_{n,N}$ denote the empirical measure based on the bootstrap sample $(\hat{X}_{n,1}, \ldots, \hat{X}_{n,N})$ of size N; i.e.

$$\hat{P}_{n,N} := N^{-1} \sum_{1 \leq i \leq N} \delta_{\hat{X}_{n,i}} \tag{3.2}$$

We refer to it as a bootstrapped empirical measure of size N. The main object of this section is to show that if (N/n) converges to one in probability, then the bootstrapped empirical measure $\hat{P}_{n,N}$ is at a distance of $o(n^{-1/2})$ from the bootstrapped empirical measure $\hat{P}_n := \hat{P}_{n,n}$, so that all of the $\sqrt{n}$ - asymptotic results for the classical bootstrap carry over to the sequential bootstrap with a random sample size.

To do so, define the random measure: $Z_n := \sqrt{n}(P_n - P),\ n \geq 1$, and for any class $\mathcal{F}$ of measurable functions $f : \chi \mapsto R^1$, let G_P be a mean zero Gaussian process indexed by $\mathcal{F}$, with covariance $E[f(X)g(X)] - Ef(X)Eg(X)$ for all $f, g \in \mathcal{F}$. Both Z_n and G_P can be viewed as being indexed by $\mathcal{F}$. We say that $\mathcal{F}$ is P-Donsker if Z_n converges in the Hoffman-Jϕrgensen weak sense to G_P in the space $\ell^\infty(\mathcal{F})$ of all uniformly bounded functions on the class $\mathcal{F}$. In this case we say that $\mathcal{F} \in CLT(P)$. The following result is well-known (Giné and Zinn, 1986). If $\mathcal{F} \in CLT(P)$, then

$$E^*\|\sum_{1\leq i\leq n}(\delta_{X_i} - P)\|_{\mathcal{F}} = O(n^{1/2}) \tag{3.3}$$

in which E^* is the so-called outer expectation.

The investigation of the classical bootstrap for the general empirical measures was initiated by P. Gaenssler (1987). A remarkable theorem due to Giné and Zinn (1990) is the equivalence of the following two conditions:

(a) $\mathcal{F} \in CLT(P)$;

(b) There exists a Gaussian process G_P defined on a certain probability space such that

$$\sqrt{n}(\hat{P}_n - P_n) \to G_P \quad \text{weakly in} \quad \ell^\infty(\mathcal{F}) \tag{3.4}$$

establishing the equivalence of the consistency of the simple bootstrap for the general empirical measures over a class $\mathcal{F}$ and the corresponding central limit theorem: $\mathcal{F} \in CLT(P)$. Praestgaard and Wellner (1993) studied more general versions of the bootstrap, including the one in which the size of bootstrap sample $m \neq n$.

We adopt the following terminology in the sequel. Given a sequence $\{\eta_n : n \geq 1\}$ of random variables and a sequence $\{a_n : n \geq 1\}$ of positive real numbers, we write

$$\eta_n = o_p(a_n) \tag{3.5}$$

iff, for each $\varepsilon > 0$,

$$\limsup_{n\to\infty} E^*(I(\{|\eta_n| \geq \varepsilon a_n\})) = 0 \tag{3.6}$$

We write

$$\eta_n = O_p(a_n) \tag{3.7}$$

iff

$$\lim_{c\to\infty} \limsup_{n\to\infty} E^*(I(\{|\eta_n| \geq c a_n\})) = 0 \tag{3.8}$$

It is easy to check that $\eta_n = O_p(a_n)$ if and only if $\eta_n = o_p(b_n)$, for any sequence b_n of positive numbers such that $a_n = o(b_n)$. An immediate consequence of Fubini's theorem is that $\eta_n = o_p(a_n)$ iff, for each $\varepsilon > 0$,

$$\mathbf{I}(\{|\eta_n| \geq \varepsilon a_n\}) = o_p(1) \tag{3.9}$$

i.e. the random element on the left hand side of (3.9) converges to zero in probability.

We now turn to the key results of this section that are instrumental in establishing consistency of various forms of sequential bootstrap resampling schemes in the literature.

Theorem 3.1. Let $\{a_n\}$ be a sequence of positive real numbers such that $a_n = O(n)$. Suppose that $\mathcal{F} \in CLT(P)$ and let

$$|N_n - n| = o_p(a_n) \tag{3.10}$$

Then

$$|\hat{P}_{n,N_n} - \hat{P}_n|_{\mathcal{F}} = o_p(\frac{a_n^{1/2}}{n}) \tag{3.11}$$

The proof of this theorem follows mainly from the Giné-Zinn theorem (1990), (3.4) and suitable triangle inequalities (cf Rao, Pathak and Koltchinskii (1997) for details).

Corollary 3.1.1: Let $\{a_n\}$ be a sequence of positive numbers such that $a_n = o(n)$. Suppose that $\mathcal{F} \in CLT(P)$ and $|N_n - n| = O_p(a_n)$. Then

$$|\hat{P}_{n,N_n} - \hat{P}_n|_{\mathcal{F}} = O_p(\frac{a_n^{1/2}}{n}) \tag{3.12}$$

For a proof, apply Theorem 3.1 to any sequence $\{b_n\}$ such that $a_n = o(b_n)$.

Corollary 3.1.2: Suppose that $\mathcal{F} \in CLT(P)$ and $|N_n - n| = o_p(n)$. Then

$$|\hat{P}_{n,N_n} - \hat{P}_n|_{\mathcal{F}} = o_p(n^{-1/2}) \tag{3.13}$$

Corollary 3.1.3: Under the conditions of Corollary 3.1.2,

$$|N_n^{1/2}(\hat{P}_{n,N_n} - P_n) - n^{1/2}(\hat{P}_n - P_n)|_{\mathcal{F}} = o_p(1) \tag{3.14}$$

Moreover, under the conditions of Theorem 3.1,

$$|N_n^{1/2}(\hat{P}_{n,N_n} - P_n) - n^{1/2}(\hat{P}_n - P_n)|_{\mathcal{F}} = o_p(\sqrt{\frac{a_n}{n}}) \tag{3.15}$$

and under the conditions of Corollary 3.1.2

$$\|N_n^{1/2}(\hat{P}_{n,N_n} - P_n) - n^{1/2}(\hat{P}_n - P_n)\|_{\mathcal{F}} = O_p(\sqrt{\frac{a_n}{n}}) \tag{3.16}$$

Indeed,

$$\|N_n^{1/2}(\hat{P}_{n,N_n} - P_n) - n^{1/2}(\hat{P}_n - P_n)\|_{\mathcal{F}} \tag{3.17}$$

$$\leq |N_n^{1/2} - n^{1/2}|(\|\hat{P}_{n,N_n} - P_n\|_{\mathcal{F}} + \|\hat{P}_n - P_n\|_{\mathcal{F}}) + n^{1/2}\|\hat{P}_{n,N_n} - \hat{P}_n\|_{\mathcal{F}}. \tag{3.18}$$

Under the conditions of Corollary 3.1.2,

$$|N_n^{1/2} - n^{1/2}| = n^{1/2}|(\frac{N_n}{n})^{1/2} - 1| = o_p(n^{1/2}) \tag{3.19}$$

Under the conditions of Theorem 3.1 we have

$$|N_n^{1/2} - n^{1/2}| = o_p(a_n n^{-1}) = o_p(\sqrt{\frac{a_n}{n}}) \tag{3.20}$$

The Giné-Zinn Theorem implies that $\{n^{1/2}\|\hat{P}_n - P_n\|_{\mathcal{F}}\}$ is stochastically bounded, and the result follows from Corollary 3.1.2 and Theorem 3.1.

4. Convergence Rates For The Sequential Bootstrap

In what follows we apply the results of Section 3 to the empirical measures based on the sequential bootstrap. We start with the following theorem summarizing the properties of the empirical measures in this case.

Theorem 4.1. Suppose that $\mathcal{F} \in CLT(P)$ and let $\hat{P}_{n,N_n}$ be the empirical measure based on a sequential bootstrap sample. Then

$$|\hat{P}_{n,N_n} - \hat{P}_n|_{\mathcal{F}} = O_p(n^{-3/4}) \tag{4.1}$$

and

$$|N_n^{1/2}(\hat{P}_{N_n,n} - P_n) - n^{1/2}(\hat{P}_n - P_n)|_{\mathcal{F}} = O_p(n^{-1/4}) \tag{4.2}$$

Indeed, in this case we have, by (2.7), $|N_n - n| = O_p(n^{1/2})$ and Corollaries 3.1.1 and 3.1.3 imply the result. In fact, a more careful analysis , e.g. by invoking the Mills ratio, allows one to replace O-terms by o-terms in these equations. It is worth noting that Equation (4.2) here implies the asymptotic equivalence of the sequential bootstrap and the simple bootstrap.

Illustrative Example: Consider the case $X = R^1$ and let $\mathcal{F} := \{I_{(-\infty,t]} : t \in R^1\}$ so that general empirical measures considered earlier turn out to be the classical empirical distribution functions. Denote

$$F(t) := P((-\infty,t]) = \int_{R^1} I_{(-\infty,t]} dP, \tag{4.3}$$

$$F_n(t) := P_n((-\infty,t]) = \int_{R^1} I_{(-\infty,t]} dP_n \tag{4.4}$$

and

$$\hat{F}_{n,N}(t) := \hat{P}_{n,N}((-\infty,t]) = \int_{R^1} I_{(-\infty,t]} d\hat{P}_{n,N} \tag{4.5}$$

We also use the abbreviation $\hat{F}_n := \hat{F}_{n,n}$ and $|\cdot|_\infty$ denotes the sup-norm.
Since, by the Kolmogorov-Donsker theorem, $\mathcal{F} \in CLT(P)$ for all Borel probability measures P on R^1 (where $\mathcal{F} = \{I_{(-\infty,t]} : t \in R^1\}$), we have the following result.

Theorem 4.2. For any distribution function F on R^1,

$$|\hat{F}_{n,N_n} - \hat{F}_n|_\infty = O_p(n^{-3/4}) \tag{4.6}$$

and

$$|N_n^{1/2}(\hat{F}_{n,N_n} - F_n) - n^{1/2}(\hat{F}_n - F_n)|_\infty = O_p(n^{-1/4}) \tag{4.7}$$

In case F is the uniform distribution on $[0,1]$, we get as a trivial corollary that the sequence of stochastic processes

$$\left\{N_n^{1/2}(\hat{F}_{n,N_n}(t) - F_n(t)) : t \in [0,1]\right\}_{n\geq 1} \tag{4.8}$$

converges weakly (in the space $\ell^\infty([0,1])$ or $D[0,1]$) to the standard Brownian bridge process

$$B(t) := w(t) - tw(1),\ t \in [0,1], \tag{4.9}$$

w being the standard Wiener process. More generally, if F is continuous, then the limit is the process $(B \circ F)(t) = B(F(t)),\ t \in R^1$. These facts easily imply justification of the sequential bootstrap for a variety of statistics θ_n, which can be represented as $\theta_n = T(F_n)$ with a compactly (Hadamard) differentiable functional T. More precisely, let T be a functional (or, more generally, an operator with values in a linear normed space) $G \mapsto T(G)$, defined on a set $\mathcal{G}$ of distribution functions G. Then T is supposed to be compactly differentiable at F tangentially to the space of all uniformly bounded and uniformly continuos functions (this is differentiability in a certain uniform sense and is much like uniform continuity on compact sets (cf Gill, 1989)). For such statistics, we have

$$T(\hat{F}_{n,N_n}) - T(\hat{F}_n) = o_p(n^{-1/2}), \tag{4.10}$$

proving the first order asymptotic correctness of the sequential resampling approach. These observations can be applied, for instance, to the operator $G \mapsto G^{-1}$, defined by

$$G^{-1}(t) := \inf\{x \in R^1 : G(x) \geq t\} \tag{4.11}$$

and taking a distribution function to its quantile function (see, e.g., Fernholz (1983) or Gill (1989) for compact differentiability of such operators). Specifically, if F is continuously differentiable at a point $x \in R^1$ with $F'(x) > 0$, then

$$|\hat{F}^{-1}_{n,N_n}(t) - \hat{F}^{-1}_n(t)| = o_p(n^{-1/2}) \tag{4.12}$$

where $t = F(x)$. In fact, we have obtained bounds for quantiles that are sharper than (4.12) (see Theorem 4.4 in Rao, Pathak and Koltchinskii, 1997). The different approaches that we have described so far show that the distance between the sequential and the usual bootstrap is at most of the order of $O(n^{-\frac{1}{4}})$. Although this entails consistency of the sequential bootstrap, it does not guarantee its second order correctness. We turn now to two approaches to estblishing second order correctness of the sequential bootstrap.

5. Second Order Correctness Of The Sequential Bootstrap

The proof of the second order correctness of the sequential bootstrap requires the Edgeworth expansion for dependent random variables. Along the lines of the Hall-Mammen work (1994), we first outline an approach based on cumulants. This approach assumes that a formal Edgeworth expansion is valid for pivot under the sequential bootstrap.

Let N_i denote the number of times the ith observation x_i from the original sample appears in the sequential bootstrap sample, $1 \leq i \leq n$. Then

$$N = N_1 + N_2 + \ldots + N_n \tag{5.1}$$

in which $N_1, N_2, \ldots$ are exchangeable random variables.

The probability distribution of N is given by

$$P(N = k) = \binom{n-1}{m} \Delta^m (\frac{x}{n})^k \tag{5.2}$$

for $k \geq m$, and in which Δ is the difference operator with unit increment.

The moment generating function of N is given by

$$M(t) = E(e^{tN}) = \binom{n-1}{m} \Delta^m \frac{n}{(n - xe^t)} \tag{5.3}$$

The second order correctness of the sequential bootstrap for linear statistics such as the sample sum is closely related to the behavior of the moments of the random variables $\{N_i : 1 \leq i \leq n\}$. Among other things, the asymptotic distribution of each N_i is Poisson with mean 1. In fact, it can be shown that

$$E(N_1 - 1)^{k_1} \ldots (N_i - 1)^{k_i} = \prod_{j=1}^{i} (e^\Delta - 1 - \Delta)(x-1)^{k_j} + O(\frac{1}{n}) \tag{5.4}$$

It follows from (5.4) that to order $O(n^{-1})$, the random variables $\{N_i : 1 \leq i \leq n\}$ are asymptotically independent. This implies that the Hall-Mammen type (1994) conditions for the second order correctness of the sequential bootstrap hold. This approach is based on the tacit assumption that formal Edgeworth type expansions go through for the sequential bootstrap. A rigorous justification of such an approach is unavailable in the literature at the present time. Another approach which bypasses this difficulty altogether entails a slight modification of the sequential bootstrap. It is based on the observation that each N_i in Equation (5.1) is approximately a Poisson variate subject to the constraint:

$$I(N_1 > 0) + I(N_2 > 0) + \ldots + I(N_n > 0) = m \approx n(1 - e^{-1}) \tag{5.5}$$

i.e. there are exactly m non-zero $N_i s$, $\quad 1 \leq i \leq n$. This observation enables us to modify the sequential bootstrap so that existing techniques on the Edgeworth expansion, such as those of Babu and Bai (1996), Bai and Rao (1991, 1992), Babu and Singh (1989) and others, can be employed. We refer to this modified version as the Poisson bootstrap.

The Poisson Bootstrap: The original sample $(x_1, \ldots, x_n)$ is assumed to be from $\mathbb{R}^k$ for greater flexibility. Let $\alpha_1, \ldots, \alpha_n$ denote n independent observations form $P(1)$, the Poisson distribution with unit mean. If there are exactly $m = [n(1 - e^{-1})]$ non-zero values among $\alpha_1, \ldots, \alpha_n$, take

$$\hat{S}_N = \{(x_1, \alpha_1), \ldots, (x_n, \alpha_n)\} \tag{5.6}$$

otherwise reject the αs and repeat the procedure. This is the conceptual definition. The sample size N of the Poisson bootstrap admits the representation:

$$N = \alpha_1 + \alpha_2 + \ldots + \alpha_n \tag{5.7}$$

in which $\alpha_1, \ldots, \alpha_n$ are IID Poisson variates with mean $= 1$ and with the added restriction that exactly m of the n αs are non-zero, i.e. $I(\alpha_1 > 0) + \ldots + I(\alpha_n > 0) = m$.

A simple way to implement the Poisson bootstrap in practice is to first draw a simple random sample without replacement (SRSWOR) of size m from the set of unit-indices

$\{1, 2, \ldots, n\}$, say $(i_1, i_2, \ldots, i_m)$. Then assign respectively to these $\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_m}$ values independently drawn from the truncated Poisson distribution with $\lambda = 1$ and left-truncated at $x = 0$ (R-syntax: qpois(runif(m, dpois(0,1), 1), 1)) and set $\alpha = 0$ for the remaining αs.

It can be shown that the moment generating function $M_N(t)$ of $N = \alpha_1 + \alpha_2 + \ldots + \alpha_n$ is (Theorem 2.1 in Babu, Pathak and Rao, 1999):

$$M_N(t) = \Big[\frac{(e^{(e^t-1)} - e^{-1})}{(1 - e^{-1})}\Big]^m \tag{5.8}$$

so that the distribution of N can be viewed as that of m IID random variables with a common moment generating function:

$$m(t) = \frac{(e^{(e^t-1)} - e^{-1})}{(1 - e^{-1})} \tag{5.9}$$

It is clear that $m(t)$ is the moment generating function of the Poisson distribution with location parameter $\lambda = 1$ and truncated at $x = 0$.

This modification of the sequential bootstrap enables us to develop a rigorous proof of the second order correctness in the sequential case. Now let $(X_1, \ldots, X_n)$ be IID random variables with mean μ and variance σ^2. We assume that X_1 is strongly non-lattice, i.e. it satisfies Cramér's condition:

$$\limsup_{|t| \to \infty} |E(exp(itX_1))| < 1 \tag{5.10}$$

Let $\{Y_j : j \geq 1\}$ be a sequence of IID Poisson random variables with mean 1. We now state three main results, furnishing a rigorous justification for the second order correctness of the sequential bootstrp. These results follow from conditional Edgeworth expansions for weighted means of multivariate random vectors (cf Babu, Pathak and Rao, 1999).

Theorem 5.1. Suppose that $E\|X_1\|^5 < \infty$ and that the characteristic function of X_1 satisfies Cramér's condition (5.10). If $m - n(1 - e^{-1})$ is bounded, then

$$(5.11)\quad P\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{n}(X_i - \bar{X})Y_i \le x s_n | T_n = m; X_1, \ldots, X_n\Big) - P\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - E(X_1)) \le x\sigma\Big)$$

$$= O_p(n^{-1})$$

uniformly in x, given $T_n := \sum_{i=1}^{n} I(Y_i > 0) = m$.

Smooth Functional Model: An extension of Theorem 5.1 to the multivariate case goes through to statistics which can be expressed as smooth fuctions of multivariate means. Now let $X_1, \ldots, X_n$ be a sequence of IID random vectors with mean μ and dispersion matrix Σ. Let Σ_n denote the corresponding sample dispersion matrix. Then the following results hold.

Theorem 5.2. Suppose that X_1 is strongly non-lattice and $E\|X_1\|^3 < \infty$. Let H be a 3-times continuously differentiable function in a neighborhood of μ. Let $l(y)$ denote the vector of first-order partial derivatives at y and suppose that $l(\mu) \ne 0$. If $m - n(1 - e^{-1})$ is bounded, then for almost all sample sequences $\{X_i : 1 \le i \le n\}$, we have

$$(5.12)\qquad \sqrt{n}\Big|P\Big(\frac{\sqrt{N}(H(N^{-1}\sum_{i=1}^{n} X_i Y_i) - H(\bar{X}_n))}{\sqrt{l'(\bar{X}_n)\Sigma_n l(\bar{X}_n)}} \le x | T_n = m; X_1, \ldots, X_n\Big)$$

$$-P\Big(\frac{\sqrt{n}(H(\bar{X}_n) - H(\mu))}{\sqrt{l'(\mu)\Sigma l(\mu)}} \le x\Big)\Big|_\infty = o(1)$$

in which $|\cdot|_\infty$ denotes the sup-norm over x.

The following result is well-suited for applications to studentized statistics.

Theorem 5.3. Let $\{X_i : 1 \le i \le n\}$ satisfy the conditions of Theorem 5.2. Suppose that the function H is 3-times continuously differentiable in the neighborhood of the origin and $H(0) = 0$. If $m - n(1 - e^{-1})$ is bounded, then for almost all sample sequences $\{X_i : 1 \le i \le n\}$, we have

$$\sqrt{n}\Big|P\Big(\frac{\sqrt{N}(H(N^{-1}\sum_{i=1}^{n}(X_i - \bar{X})Y_i))}{\sqrt{l'(0)\Sigma_n l(0)}} \le z|T_n = m; X_1, \ldots, X_n\Big) - P\Big(\frac{\sqrt{n}(H(\bar{X}_n - \mu))}{\sqrt{l'(0)\Sigma l(0)}} \le z\Big)\Big|_\infty = o(1) \tag{5.13}$$

For example, an immediate consequence of Theorem 5.3 is the second-order correctness of the following sequential bootstrap pivot:

$$\hat{\pi}_N = \sqrt{N}(\sum_{i=1}^{n}(X_i - \bar{X})Y_j)/s_n \tag{5.14}$$

given that $T_n := \sum_{i=1}^{n} I(Y_i > 0) = m$.

6. Concluding Remarks

The sequential bootstrap discussed here has its origins in the simple bootstrap of Bradley Efron and his observation that the information content of the simple bootstrap is supported on approximately $0.632n$ of the original data. It is well-known that owing to the with replacement nature of resampling in the simple bootstrap, the information content of its bootstrapped sample is a random variable. The sequential bootstrap keeps the information content of its bootstrapped samples at $0.632n$-level of the original sample size. This constancy of information in the sequential bootstrap samples ensures homogeneity of information and reduces the likelihood of information-deficient samples. The sequential bootstrap provides better breakdown points. The empirical characteristics of the sequential bootstrap are within $O(n^{-3/4})$ of the simple bootstrap and its second-order correctness goes through under existing regularity conditions. It is perhaps worthwhile to conclude this article with a final remark that recent studies in the literature (Pino-Mejías et al. (2010), Pauly (2011) and others) indicate that the sequential bootstrap is likely to perform better than the simple bootstrap in small to moderate size samples under a variety of applications commonly encountered in practice.

Acknowledgments

The authors are grateful to Professors Babu, Koltchinskii and Koul for their perusal and helpful remarks. This work was supported in part from NIH grants RC065888, RC060667 and RC060042.

References

Babu, G.J., Bai, Z.D.,1996. Mixture of global and local Edgeworth expansions and their applications. *J. Multivariate Analysis* 59, 282-307.

Babu, G.J., Pathak, P.K., Rao, C.R., 1999. Second-order correctness of the Poisson bootstrap. *Ann. Statist.*27, 1666–1683.

Babu, G.J., Singh, K., 1989. On Edgeworth expansions in the mixture cases. *Ann. Statist.* 17, 443–447.

Bai, Z.D., Rao, C.R., 1991. Edgeworth expansions of a function of sample means. *Ann. Statist.*19, 1295-1315.

Bai, Z.D., Rao, C.R., 1992. A note on the Edgeworth expansion for ratio of sample means. *Sankhyā A* 19, 309–322.

Bickel, P.J.,Freedman, D.A., 1981. Some asymptotic theory for the bootstrap. *Ann. Statist.* 9, 1196-1217.

Chernick, M.R., 2008. *Bootstrap Methods: A guide for practitioners and researchers, 2nd Edition.* Wiley, Hoboken, New Jersey.

Efron, B., 1979. Bootstrap methods: Another look at the jackknife. *Ann. Statist.* 7, 1-26.

Efron, B., 1982. *The Jackknife, the Bootstrap, and Other Resampling Plans.* SIAM, Philadelphia.

Efron, B., 1983. Estimating the error rate of a prediction rule: Improvement on cross-validation. *J. Am. Statist. Assoc.*, 78, 316-331.

Fernholz, L.T., 1983. Von Mises calculus for statistical functionals, *Lecture Notes in Statistics* 19, Springer-Verlag, New York.

Gaenssler, P., 1987. Bootstrapping empirical measures indexed by Vapnik-Červonenkis classes of sets. In *Probability Theory and Mathematical Statistics* (Yu. V. Prohorov, V.A. Statulevichius, V. Sazonov et al., Eds) 1, 467-481. VNU Science Press, Utrecht.

Gill, R., 1989. Non- and Semi-parametric Maximum Likelihood Estimators and the von Mises Method (Part 1). *Scand. J. Statist.* 16, 97-128.

Giné, E., Zinn, J., 1984. Some limit theorems for empirical processes. *Ann. Prob.* 12, 929-989.

Giné, E., Zinn, J., 1986. Lectures on the central limit theorem for empirical processes. *Lecture Notes in Mathematics* 1221, 50-113. Springer.

Giné, E., Zinn, J., 1990. Bootstrapping general empirical measures. *Ann. Prob.* 18, 851-869.

Hall, P., 2003. A short prehistory of the bootstrap. *Statistical Science* 18, 158–167.

Hall, P., Mammen, E., 1994. On general resampling algorithms and their performance in distribution estimation. *Ann. Statist.* 24, 2011–2030.

Jiménez-Gamero, M.D., Muñoz-Garcia J., Pino-Mejías, R., 2004. Reduced bootstrap for the median. *Statistica Sinica* 14, 1179-1198.

Jiménez-Gamero, M.D., Muñoz-Garcia J., Cubiles-de-la-Vega M.D., 2006. Consistency of the reduced bootstrap for sample means. *Stat. Probab. Lett.* 76, 689-697.

Mahalanobis, P.C., 1946. On large-scale sample surveys. *Philos. Trans. Roy. Soc. London Ser. B* 231, 329-451.

Mallows, C.L., 1972. A note on asymptotic joint normality. *Ann. Math. Stat.* 43, 508-515.

Mammen, E., 1992. *When Does Bootstrap Work?* New York: Springer-Verlag.

Mitra, S.K., Pathak, P.K., 1984. The nature of simple random sampling. *Ann. Statist.* 12, 1536-1542.

Pathak, P.K., 1964. Sufficiency in sampling theory. *Ann. Math. Statist.* 35, 795-808.

Pathak, P.K., 1964. On inverse sampling with unequal probabilities. *Biometrika* 51, 185-193.

Pauly, M., 2011. Consistency of the subsample bootstrap empirical process. *Statistics, iFirst* 45, 1-6.

Pino-Mejías, R., Jiménez-Gamero, M.D., González, A.E., 2010. A Monte Carlo comparison of three consistent bootstrap procedures. Personal Communication

Praestgaard, J., Wellner, J., 1993. Exchangeably weighted bootstraps of the general empirical process. *Ann. Prob.* 21, 2053-2086.

Rao, C.R., Pathak, P.K., Koltchinskii, V.I., 1997. Bootstrap by sequential resampling. *J. Statist. Plann. Inference* 64, 257–281.

Shoemaker, O.J., Pathak, P.K., 2001. The sequential bootstrap: a comparison with regular bootstrap. *Commun. Stat. Theory Methods* 30, 1661–1674.

Singh, K., 1981. On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* 9, 1187–1195.