INTRODUCTION TO THE BOOTSTRAP

Suppose we have observed $X_1, ..., X_n$ (not necessarily real numbers, they can be in any space S) and suppose for simplicity that the n observations are all different. We can form the empirical measure $P_n = \frac{1}{n}\sum_{j=1}^n \delta_{X_j}$ where $\delta_x(A) := 1_A(x) := 1$ if $x \in A$ and 0 otherwise, for any subset $A \subset S$. Let P be the unknown probability distribution, from which we assume X_j are i.i.d. We would like to estimate some functional $T(P)$. Here "functional" just means a function whose domain is a relatively abstract space, in this case a space of probability measures P, including P_n and P_n^B defined below. For P defined on the real line, an example of a functional $T(P)$ is the median of P, defined as the midpoint of the interval of medians if the interval does not reduce to a point. For P defined on any space, another example of a real-valued functional is $T(P) = E_P f := \int f\,dP$ where f is a bounded function which in measure-theoretic terms is measurable, or in probability terms is a random variable with respect to P, so that since it is bounded, its expectation is well-defined, as is its variance.

We can give a point estimate $T(P_n)$ of $T(P)$ but we'd like to know how uncertain the estimate is, for example to give a confidence interval for $T(P)$ if T is real-valued, without knowing anything more about P than the observations $X_1, ..., X_n$ summarized in P_n. For example, we don't assume that P belongs to any particular parametric family.

What the bootstrap does is to resample from the given sample, i.e. to take $X_1^B, ..., X_K^B$ i.i.d. (P_n). Here we're sampling "with replacement" from the original sample. In general one could consider different values of the *bootstrap sample size* K, but the default choice, to be used here, will be $K = n$. Thus from $X_1^B, ..., X_n^B$ we can form the *bootstrap empirical measure*

$$P_n^B \; := \; \frac{1}{n}\sum_{k=1}^n \delta_{X_k^B}.$$

We will be interested not in the (unconditional) distribution of P_n^B but in its conditional distribution given P_n. To estimate this distribution one can use a Monte Carlo method: one repeats the resampling some large number R of times (R may be called the number of bootstrap replications), giving R i.i.d. values of P_n^B all for the same P_n, from which one can estimate the conditional distribution of $T(P_n^B)$ given P_n. This

Date: 18.465, revised March 28, 2015.

requires altogether Rn i.i.d. samples from a given P_n and, to find the sampling distribution of a functional $T(P_n^B)$, R separate evaluations of this functional. The intent of the method is, to get a confidence set (such as a confidence interval if T is real-valued) for the unknown $T(P)$, we'd like to know the distribution of $T(P'_n) - T(P)$, where P'_n is a random empirical measure from P (and so to be distinguished from the observed P_n from which P_n^B is sampled) and we estimate this from the conditional distribution of $T(P_n^B) - T(P_n)$ given P_n, which we can observe.

When is equality in distribution preserved? When random variables X and Y have the same distribution or in other words are equal in distribution, we will write $X =_d Y$ (X is equal in distribution to Y). It follows that if c is any constant, then $X + c =_d Y + c$ and $cX =_d cY$. But for example let X and Y be i.i.d. $N(0, 1)$. Then $X =_d Y$ and $X =_d X$, but $X + X \neq_d X + Y$ because $X + X = 2X$ is $N(0, 4)$ while $X + Y$ is $N(0, 2)$.

Need for computation. For reasonably large R (and n), the bootstrap is a computer-intensive method. The availability of computers made possible the invention of the bootstrap by Efron (1979), see also the exposition by Efron and Tibshirani (1993). For example, the paper by Suzuki and Shimodaira (2006), 3d page, mentions a bootstrap calculation taking over 7 hours on one processor, or 24 minutes on 20 parallel processors.

High-level probability theory of the bootstrap. Let's see what can be said about the bootstrap from a theoretical viewpoint. The bootstrap has good properties for suitable sample means or collections of them. Let g be a function on S. Then $\int g dP_n$ is just the observed sample mean of $g(X_j)$. We know that if g is a random variable with finite variance, so that $\int g(x)^2 dP(x) < \infty$, then by the central limit theorem, $\sqrt{n} \int g d(P_n - P)$ converges in distribution as $n \to \infty$ to a normal variable with mean 0 and the same variance as g. Moreover if $g_1, ..., g_k$ are random variables each with finite variance, then the random vector $\sqrt{n}\{\int g_i d(P_n - P)\}_{i=1}^k$ converges in distribution to a vector, say $\{G_P(g_i)\}_{i=1}^k$, with k-variate normal distribution having mean 0 and the same covariance matrix as that of the g_i for P. Since P is unknown, it's useful to take the functions g_i to be bounded, so we can be sure that finite means, variances and covariances will exist. The convergence to normality holds uniformly over some infinite families of functions, for example on the real line, over the set of all indicator functions $1_{(-\infty,x]}$ for all x, as shown in the KMT (Komlós-Major-Tusnády) theorem, made more precise by Bretagnolle and Massart. General conditions on a family of functions for such uniform central limit theorems to hold are given for example in van der Vaart and Wellner (1996) and Dudley (2014). Moreover, Giné and Zinn (1990) proved under general conditions that if the uniform central limit theorem holds over a family $\mathcal{F}$ of functions, then it holds also for the bootstrapped empirical process $\sqrt{n}(P_n^B - P_n)$ conditional on P_n, in probability as $n \to \infty$. Expositions are given in van der Vaart and Wellner (1996, §3.6) and Dudley (2014, §§9.2–9.4).

We saw however that even in the most classical case of empirical distribution functions, the Bretagnolle-Massart theorem didn't give a fast enough rate of convergence to be of direct practical use, and that quantiles for the supremum norm of classical empirical processes (Kolmogorov statistics) seemed to converge to their limits at a $1/\sqrt{n}$ rate rather than the $(\log n)/\sqrt{n}$ rate given by the KMT theorem. In general, still less is known about the speed of convergence of empirical processes to their limiting Gaussian processes. In some cases the convergence is known to be slow. For example, in Euclidean space $\mathbb{R}^d$, let B_d be the collection of all closed balls $B(x,r) := \{y : |y-x| \leq r\}$ for all possible $x \in \mathbb{R}^d$ and all $r > 0$. Let P be the uniform distribution on the unit cube $I^d := \{x : 0 \leq x_i \leq 1,\ 1 \leq i \leq d\}$. It is known that $\sqrt{n}(P_n - P)$ converges in distribution with respect to uniform convergence over B_d to G_P, but but Beck (1985) showed that for $d \geq 2$ convergence is no faster than at the rate $O(n^{-1/(2d)})$. Rather the Giné–Zinn theorem gives us some overall reassurance that the bootstrap works asymptotically rather generally.

A functional with general, in fact possibly infinite-dimensional, values, is as follows: let $\mathcal{F}$ be a class of bounded measurable functions, and let $T(P) := \{\int f\, dP : f \in \mathcal{F}\}$. Such functionals arise in the Giné–Zinn theorem mentioned previously. In this course we'll be concerned at least for the time being with real-valued functionals.

Definition. For a real-valued functional T and for $X_1, \dots, X_n$ i.i.d. (P) with empirical measure P_n, we'll say that the bootstrap *is valid* for T and P if there exists some $t > 0$ such that as $n \to \infty$:
(a) The distribution of $n^t(T(P_n') - T(P))$ converges to that of a finite valued, non-degenerate random variable Y, where non-degenerate means that $P(Y = 0) < 1$;
(b) The conditional distribution of $n^t(T(P_n^B) - T(P_n))$ given P_n converges to that of the same Y as $n \to \infty$, in probability with respect to P_n, where the last phrase means that as $n \to \infty$, the probability that

P_n is such that the given conditional distribution is close to that of Y approaches 1.

Remarks. If part (a) of the definition holds for some $t > 0$, then it holds only for that t, because if $0 < s < t < u$ then $n^s(T(P_n) - T(P)) \to 0$ in probability, and $n^u(T(P_n) - T(P))$ is not bounded in probability, so it cannot converge in distribution.

If (a) holds, then most often in practice, $t = 1/2$. For example let $T(P) = \int g\,dP$ for some bounded function g which is a random variable with respect to P, with variance $\sigma^2 > 0$ depending on P. Then part (a) of the definition holds with $t = 1/2$, as $\sqrt{n}(T(P_n) - T(P))$ converges in distribution to $N(0, \sigma^2)$ by the central limit theorem. In this case the bootstrap is valid, as the conditional distribution of $\sqrt{n}(T(P_n^B) - T(P_n))$ given P_n does converge to the same limiting distribution, in probability with respect to P_n (or so it seems, by the Lindeberg triangular arrays central limit theorem), but the bootstrap is not helpful or needed. One can estimate σ^2 by $s_g^2 = \frac{1}{n-1}\sum_{j=1}^n (g(X_j) - \overline{g})^2$ where $\overline{g} = \int g dP_n$ and and apply the central limit theorem directly.

The bootstrap — basic properties. Suppose again for convenience that $X_1, ..., X_n$ are all distinct. The probability that a given observation, say X_j, is omitted from the bootstrap sample, i.e. $X_k^B \neq X_j$ for all $k = 1, ..., n$, is $\left(1 - \frac{1}{n}\right)^n$, which converges to $1/e$ as $n \to \infty$. Thus, on average, for large n, a fraction about $1/e$ of the original observations are omitted from the bootstrap sample. Further, as $n \to \infty$, if n_j is the number of times X_j is selected in one bootstrap sample of size n, then $(n_1, ..., n_n)$ have a multinomial $(n; 1/n, ..., 1/n)$ distribution, so the marginal distribution of each n_j is binomial$(n, 1/n)$ which converges as $n \to \infty$ to a Poisson(1) distribution, i.e. $\Pr(n_j = k) \to 1/(ek!)$ as $n \to \infty$ for each $k = 0, 1, ...$.

Since the X_j are all different, each choice of $n_1, \ldots, n_n$ gives a different value of P_n^B. The number of possible choices of integers $n_j \geq 0$ such that $\sum_{j=1}^n n_j = n$ is $\binom{2n-1}{n}$, as is known from basic combinatorics. [It can be seen as follows: consider the set of all strings of $2n-1$ characters consisting of n 1's and $n-1$ 0's. There are clearly $\binom{2n-1}{n}$ such strings. There is a one-to-one correspondence between such strings and choices of n_j as follows. Let n_1 be the number of 1's before the first 0, let n_j be the number of 1's between the $(j-1)$st and jth 0's for $j = 2, \ldots, n-1$, and let n_n be the number of 1's after the last 0.]

As $n \to \infty$, it can be seen via Stirling's formula that $\binom{2n-1}{n}$ is asymptotic to $4^n n^b C$ for some b and some constant C, where the dominant factor 4^n grows geometrically with n. So, unless n is rather small, it's not practicable to find the exact distribution of $T(P_n^B)$ given P_n, as one would have to compute T at roughly 4^n different P_n^B's. One would also need to compute, for each possible $n_1, \ldots, n_n$, the multinomial probability $\binom{n}{n_1,\ldots,n_n} n^{-n}$. So there is a need for bootstrap sampling.

In such sampling we do R bootstrap replications for some large enough R. Specifically, let X_{ki}^B be i.i.d. (P_n) for $k = 1, ..., n$ and $i = 1, ..., R$. Let

$$P_{ni}^B := \frac{1}{n} \sum_{k=1}^{n} \delta_{X_{ki}^B}$$

for $i = 1, ..., R$. Thus we have R independent copies of P_n^B. We can form the R i.i.d. random variables $T_i := T(P_{ni}^B)$.

The bootstrap for real-valued X_j, order statistics, and quantiles. For X_j real, The bootstrap sample has its own order statistics $X_{(k)}^B$, $k = 1, ..., n$, which for given P_n have a discrete distribution. As will be seen in PS6, the probability distributions of these order statistics can be evaluated in terms of binomial distributions. So, it's unnecessary actually to do bootstrap sampling in these cases as we have the exact distribution. As to be found in PS6 one can get approximate bootstrap $100(1-\alpha)\%$ confidence intervals for quantiles. (They can only be approximate because of the discrete distribution of the bootstrap order statistics.) But as will also be seen in PS6, one can directly get nonparametric confidence intervals for quantiles without the bootstrap. Then one can compare the bootstrap and non-bootstrap confidence intervals to see if they agree exactly or approximately. If they do, they can give further reassurance of the validity of the bootstrap, even if it is not really needed in this case.

For extreme order statistics, however, the bootstrap may not be valid.

Example. Consider the functional $T(P) = \sup\{x : \; P((-\infty, x]) = 0\}$. Then $T(P_n) = X_{(1)}$ and $T(P_n^B) = X_{(1)}^B$. We will have $X_{(1)}^B = X_{(j_1)}$ for some j_1. Suppose that $X_1, \ldots, X_n$ are all distinct. Then $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$. By definition of bootstrap sampling, $\Pr(j_1 \geq j) = \left(\frac{n-(j-1)}{n}\right)^n$ for $j = 1, \ldots, n$, which converges as $n \to \infty$ to e^{1-j}. It follows that $\lim_{n\to\infty} \Pr(j_1 = j) = e^{1-j} - e^{-j} = q^{j-1}p$ where $q = 1/e$ and $p = 1 - q$. Thus the distribution of j_1 converges to a geometric(p) distribution, and we have the asymptotic distribution of $X_{(1)}^B$ in terms of the $X_{(j)}$.

For simplicity, let P be $U[0, 1]$, so that $T(P) = 0$. The following is known: for the order statistics $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ from $U[0, 1]$, define $X_{(0)} = 0$, $X_{(n+1)} = 1$, and $s_j = X_{(j)} - X_{(j-1)}$ for $j = 1, \dots, n+1$. Then for each $n \geq 1$, the joint distribution of the spacings $\{s_j\}_{j=1}^{n+1}$ equals that of $\{Y_j/S_{n+1}\}_{j=1}^{n+1}$ where $Y_1, \dots, Y_{n+1}$ are i.i.d. standard exponential random variables and $S_{n+1} = \sum_{i=1}^{n+1} Y_i$. A reference for this is Shorack and Wellner (2009, §8.2, Proposition 1 p. 335).

Since $EY_j = 1$ for each j, by the law of large numbers, $S_{n+1}/(n+1) \to 1$ as $n \to \infty$, and so $S_{n+1} \sim n+1 \sim n$. It follows that $n(T(P_n) - T(P)) = n(X_{(1)} - 0) = ns_1$ converges in distribution to standard exponential, so part (a) in the definition of bootstrap validity holds with $t = 1$.

However, $n(T(P_n^B) - T(P_n)) = n(X_{(1)}^B - X_{(1)})$ equals 0 with probability converging to $p > 0$, so it does not have the same limiting distribution as in part (a), and the bootstrap is not valid for this T and P. There would be a similar failure for the same T and any P with a density f such that $f(x)$ approaches a positive limit as $x \downarrow a$ for some a and $f(x) = 0$ for $x < a$, such as $U[a,b]$ or the distribution of $a + X$ where X has an exponential (λ) density. Here a might be unknown and we might want to estimate it.

The bootstrap and standard errors. Recall that if $X_1, \dots, X_n$ are i.i.d. with finite mean μ, variance σ^2 and standard deviation σ, then for $\overline{X} = (X_1 + \cdots + X_n)/n$ we have $E\overline{X} = \mu$ and $\mathrm{Var}(\overline{X}) = \sigma^2/n$, so the standard deviation of $\overline{X}$ is $\sigma/\sqrt{n}$, which is called the *standard error of the mean.* It can be estimated by $\widehat{SE} = s_X/\sqrt{n}$ where $s_X^2 = \frac{1}{n-1}\sum_{j=1}^n (X_j - \overline{X})^2$. For n large enough, by the central limit theorem, $\sqrt{n}(\overline{X} - \mu)$ is approximately $N(0, \sigma^2)$, so we can get approximate $100(1-\alpha)\%$ confidence intervals for μ with endpoints $\overline{X} \pm \widehat{SE} z_{\alpha/2}$ where $P(Z \geq z_\beta) = \beta$ for a $N(0,1)$ variable Z.

Applying the same idea to the bootstrap (Efron and Tibshirani, Chapter 6), where now T is general or complicated enough that we cannot treat it as directly as we can for means or quantiles, but we can calculate $T(P_n)$, suppose we observe a given P_n and take R i.i.d. bootstrap samples giving P_{ni}^B, $i = 1, \dots, R$, recall $T_i = T(P_{ni}^B)$, take the sample mean

$$\overline{T} = \frac{1}{R}\sum_{i=1}^R T_i$$

and sample variance $(s_T^B)^2 = \frac{1}{R-1}\sum_{i=1}^R (T_i - \overline{T})^2$. One can get an approximate confidence interval for $T(P)$ for the unknown P as described below.

For the R bootstrap library "boot," one begins with a sample y such as $(X_1, ..., X_n)$. For example, suppose the functional T to be bootstrapped is the median. One can create a bootstrap object which may be called y.boot by

```
> library(boot)
> set.seed(101)
> y.boot = boot(y,function(x,i) median(x[i]), R = 1000)
```

Here i indexes bootstrap replications and will run from 1 to R. If one then types "y.boot" one gets output with labels in the first line: "original bias std.error" and numbers under each. *This output relates mainly to the normal-based confidence intervals, which Venables and Ripley, and I, de-emphasize.* The first number "original" is just $T(P_n)$, in this case the sample median of the sample y. The second number, "bias," equals $\overline{T} - T(P_n)$, recalling that $\overline{T}$ is the sample mean of the T_i.

"Standard error" means the standard deviation of a statistic, or an estimate of it. Sometimes, and especially when called "standard error of the mean," it means standard deviation of a sample mean, or an estimate of it, namely, for i.i.d. random variables T_i with standard deviation σ, the standard deviation of $\overline{T}$ is $\sigma/\sqrt{R}$, estimated by $s_T^B/\sqrt{R}$. In this situation, the relevant statistic is an individual $T_i = T(P_{ni}^B)$. Assuming that $T(P'_n) - T(P)$ is approximately normally distributed with mean μ (not necessarily 0 in general) and standard deviation σ, one would estimate σ by the sample standard deviation s_T^B of the T_i.

In a simple "toy" example $y = (0, 1, 3)$, the mean of the bootstrap sample median m^B is 1.2593, and its median is 1, so that the true bias of $\overline{T}$ is 0.2593. Applying "boot" with $R = 1000$, the estimated bias R gave was 0.246. The true standard deviation of m^B is 1.1086 and R gave the estimated "standard error" of 1.1067. Dividing by $\sqrt{R}$ would give something much smaller. Of course, one would like n much larger than 3 so that the bootstrap would become valid and approximate normality might hold. The example was chosen just to check the meaning of the outputs of some of R's bootstrap functions.

The normal-based confidence intervals from the bootstrap work as follows. Assume as in general with the bootstrap that the conditional distribution of $T(P_n^B) - T(P_n)$ given P_n is approximately the same as the distribution of $T(P'_n) - T(P)$ and now moreover, that this distribution is approximately $N(\mu, \sigma^2)$ for some μ and some $\sigma > 0$. One then estimates μ by $\widehat{\mu} = \overline{T} - T(P_n)$ which is the "bias," and σ by the sample standard deviation of the T_i which is the "standard error." If $T(P_n) - T(P)$ as a random variable has approximately this distribution, then a point estimate of $T(P)$ is $T(P_n) - \widehat{\mu}$. (This may be somewhat surprising since one might have thought $T(P_n)$ itself was the natural point estimator of $T(P)$.)

Besides what is displayed at first, a bootstrap object such as y.boot above actually has much more information in it, including the order statistics of the T_i, $T_{(1)} \leq T_{(2)} \leq \cdots \leq T_{(R)}$. The R command boot.ci gives a choice of confidence intervals in which those of this form are called "normal" in the output, abbreviated "norm" in the command as in

> boot.ci(y.boot,conf=c(0.90,0.95),type = c("norm", "basic", "perc"))
where "basic" and "perc," as we'll see below, use the order statistics $T_{(i)}$. In PS6 you can see how they behave in some cases.

Using the bootstrap order statistics. An idea seemingly better than the standard error approach in bootstrapping is to use not only the sample mean and variance of the bootstrap observations T_i but all the information in them, via their order statistics $T_{(1)} \leq T_{(2)} \leq \cdots \leq T_{(R)}$. From these one can estimate quantiles of the distribution of the T_i for the given P_n. For $0 < q < 1$ the qth sample quantile of the T_i is defined as $T_{(\lceil Rq \rceil)}$ if Rq is not an integer, where $\lceil x \rceil$ is defined as the least integer $\geq x$, or as $\frac{1}{2}[T_{(Rq)} + T_{(Rq+1)}]$ if Rq is an integer, as in the familiar case of the sample median where $q = 1/2$. For the similar case of Monte Carlo sampling, where we have some N instead of R, recall that in finding quantiles for the dip statistics, the Hartigans used $N = 9999$ and Maechler used $N = 10^6 + 1$ so that Nq is not an integer for any of the q's used. Whereas, for $R = 1000$ as suggested in the bootstrap, Rq will be an integer for the usual values of q.

One can get a nonparametric (as opposed to normal-based) $100(1 - \alpha)\%$ confidence interval for $T(P_n^B)$ given P_n as $[L, U]$ where L is the $\alpha/2$ sample quantile of the T_i and U is the $1-(\alpha/2)$ quantile. This is called the "percentile" confidence interval, also in the R output, abbreviated "perc" in the command as above. It's an observed confidence interval for $T(P_n^B)$. Then subtracting $T(P_n)$ from both endpoints, we get such a confidence interval for $T(P_n^B) - T(P_n)$ conditional on P_n.

Let P_n' be (again) a general empirical measure for the given P, as opposed to the observed P_n from which P_n^B are sampled. Certainly $T(P_n') =_d T(P_n)$, and since $T(P)$ and $\sqrt{n}$ are constants,

$$\sqrt{n}(T(P_n') - T(P) =_d \sqrt{n}(T(P_n) - T(P)).$$

If the bootstrap works, then the left side of the given equation is approximately equal in distribution to

$$(\sqrt{n}(T(P_n^B) - T(P_n))|P_n).$$

But we saw earlier that one has to be careful in algebraic manipulations with quantities equal in distribution. If there are conditional probabilities involved, still more care may be needed. Davison and Hinkley (1997 or earlier) proposed to plug in $T(P_n)$ in place of $T(P_n')$. That would give

$$1-\alpha \sim \Pr\left(L - T(P_n) \le T(P_n) - T(P) \le U - T(P_n)\right)$$
$$= \Pr\left(2T(P_n) - U \ \le\ T(P) \le 2T(P_n) - L\right),$$

a proposed confidence interval for $T(P)$.

Namely, as $[L - T(P_n), U - T(P_n)]$ is an observed $1-\alpha$ confidence interval for $T(P_n^B) - T(P_n)$ given P_n, if we take it as an approximate $1-\alpha$ confidence interval for $T(P_n') - T(P)$, and then by the plug-in for $T(P_n) - T(P)$, then the inequalities

$$L - T(P_n) \le T(P_n) - T(P) \le U - T(P_n)$$

are equivalent by simple additions and subtractions to

$$2T(P_n) - U \le T(P) \le 2T(P_n) - L$$

and so give us as an approximate $1-\alpha$ confidence interval for $T(P)$,

$$[2T(P_n) - U, 2T(P_n) - L].$$

This is the interval called "basic" in the R output and command (not abbreviated). On p. 136 Venables and Ripley say "the intervals based on normality are not adequate" (because of asymmetry) and that the "basic" intervals are preferable to ("more rational" than) the percentile intervals as confidence intervals for $T(P)$.

Drawbacks of the basic interval. It is not clear that the plug-in $P_n' = P_n$ is valid. One needs to do experimental (Monte Carlo) checking to see how well different intervals work. Davison and Hinkley (inventors of the basic interval) did so themselves in their 1997 book. They found that the basic interval worked no better than the percentile interval and that neither worked as well as some other bootstrap-based intervals, such as the BC_a (bias-corrected, accelerated interval, implemented in R "boot" as "bca").

Suppose that $T(\cdot) > 0$, so that $0 < T_L < T_U < \infty$. Use of the basic interval can lead to strange results if $T(P) > 2T(P_n)$, or equivalently $T(P_n) < T(P)/2$, as can happen with probability > 0 in case $T(P)$ is the variance of P (see PS6 problem 5). So it seems the basic interval should not be recommended.

Phylogenetic trees. The paper by Efron, Halloran and Holmes (1996) treats an application of the bootstrap which has become rather popular.

They mention that application of the bootstrap to phylogenetic trees had been proposed earlier by Felsenstein (1985). I found from the Web of Science that the Felsenstein paper has been cited over 15,000 times by other papers through October 2010, or 27,170 times through March 2015 according to Google Scholar, a very large impact. The paper by Efron et al. had been cited over 570 times. It proposes some improvements to the method which have been incorporated in the R package pvclust (Suzuki and Shimodaira, 2006). We will not be using that in our first week on the bootstrap. We will get to it later.

Notes. In the title of Shorack and Wellner (1986, repub. 2009), "Empirical Processes" meant classical empirical processes $\sqrt{n}(F_n - F)$ except in the last Chapter 26. In Chapter 26, and in van der Vaart and Wellner (1996), it means general empirical processes $\sqrt{n}(P_n - P)$.

REFERENCES

Beck, J. (1985). Lower bounds on the approximation of the multivariate empirical process. *Z. Wahrscheinlichkeitsth. verw. Geb.* **70**, 289-306. A supplementary exposition suggesting some revisions of the proof and filling in details, from a seminar in the Fall of 2006, is posted at www-math.mit.edu/∼rmd/Research/beck85.pdf

Brown, L. D., Cai, T. T., and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science* **16**, 101–133.

Davison, A. C., and Hinkley, D. V. (1997), *Bootstrap Methods and their Applications*, Cambridge University Press.

Dudley, R. M. (2014). *Uniform Central Limit Theorems.* Cambridge University Press, second edition.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**, 1-26.

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

Efron, B., Halloran, E., and Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proc. Nat. Acad. USA* **93**, 7085–7090, 13429–13434. Available from the PNAS website with authors' "corrections incorporated" [the latter page range gave "additions and corrections"].

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783-791.

Giné, E., and Zinn, J. (1990). Bootstrapping general empirical measures. *Annals of Probability* **18**, 851-869.

Shorack, G., and Wellner, J. A. (1986, 2009). *Empirical Processes with Applications to Statistics*, series "*Classics in Applied Mathematics*." Society for Industrial and Applied Math. (SIAM). The 2009 edition is a reprint of the Wiley (1986) edition but has in addition a list, pp. 901–915, of (a few major, mostly minor) Errata in the previous edition.

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540-1542.

van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. (I've heard that a second edition is in the works, but it may not have appeared yet, 3/15.)