

# BOOTSTRAP ASYMPTOTICS

*Rudolf Beran*<sup>1</sup>

Distinguished Professor Emeritus, Department of Statistics  
University of California, Davis, CA 95616, USA  
E-mail: [beran@wald.ucdavis.edu](mailto:beran@wald.ucdavis.edu)

The bootstrap, introduced by Efron (1979), merges simulation with formal model-based statistical inference. A statistical model for a sample  $X_n$  of size  $n$  is a family of distributions  $\{P_{\theta,n} : \theta \in \Theta\}$ . The parameter space  $\Theta$  is typically metric, possibly infinite-dimensional. The value of  $\theta$  that identifies the true distribution from which  $X_n$  is drawn is unknown. Suppose that  $\hat{\theta}_n = \hat{\theta}_n(X_n)$  is a consistent estimator of  $\theta$ . The bootstrap idea is:

(a) Create an artificial *bootstrap world* in which the true parameter value is  $\hat{\theta}_n$  and the sample  $X_n^*$  is generated from the fitted model  $P_{\hat{\theta}_n,n}$ . That is, the conditional distribution of  $X_n^*$ , given the data  $X_n$ , is  $P_{\hat{\theta}_n,n}$ .

(b) Act as if a sampling distribution computed in the fully known bootstrap world is a trustworthy approximation to the corresponding, but unknown, sampling distribution in the model world.

For example, consider constructing a confidence set for a parametric function  $\tau(\theta)$ , whose range is the set  $T$ . As in the classical pivotal method, let  $R_n(X_n, \tau(\theta))$  be a specified *root*, a real-valued function of the sample and  $\tau(\theta)$ . Let  $H_n(\theta)$  be the sampling distribution of the root under the model. The *bootstrap distribution* of the root is  $H_n(\hat{\theta}_n)$ , a random probability measure that can also be viewed as the conditional distribution of  $R_n(X_n^*, \tau(\hat{\theta}_n))$  given the sample  $X_n$ . An associated *bootstrap confidence set* for  $\tau(\theta)$ , of nominal coverage probability  $\beta$ , is then  $C_{n,B} = \{t \in T : R_n(X_n, t) \leq H_n^{-1}(\beta, \hat{\theta}_n)\}$ . The quantile on the right can be approximated, for instance, by Monte Carlo techniques. The intuitive expectation is that the coverage probability of  $C_{n,B}$  will be close to  $\beta$  whenever  $\hat{\theta}_n$  is close to  $\theta$ .

When does the bootstrap approach work? Bootstrap samples are perturbations of the data from which they are generated. If the goal is to probe how a statistical procedure performs on data sets similar to the one at hand, then repeating the statistical procedure on bootstrap samples stands to be instructive. An exploratory rationale for the bootstrap appeals intellectually when empirically supported probability models for the data are lacking. Indeed, the literature on “statistical inference” continues to struggle with an uncritical tendency to view data as a *random* sample from a statistical model *known* to the statistician apart

---

<sup>1</sup>Rudolf Beran was Department Chair at UC Davis (2003–07) and at UC Berkeley (1986–89). He received in 2006 the Memorial Medal of the Faculty of Mathematics and Physics, Charles University, Prague, in recognition of “distinguished and wide-ranging achievements in mathematical statistics, . . . devoted service to the international statistical community, and a long-lasting collaboration with Czech statisticians.” During 1997–99 he held an Alexander von Humboldt U.S. Senior Scientist Award at Heidelberg University. He has authored or co-authored over 100 papers in international journals and published lecture notes (with G. R. Ducharme) on *Asymptotic Theory for Bootstrap Methods in Statistics* (Publications CRM, Université de Montréal, 1991).

from parameter values. In discussing the history of probability theory, Doob (1972) described the mysterious interplay between probability models and physical phenomena: “But deeper and subtler investigations had to await until the blessing and curse of direct physical significance had been replaced by the bleak reliability of abstract mathematics.”

Efron (1979) and most of the subsequent bootstrap literature postulate that the statistical model  $\{P_{\theta,n}: \theta \in \Theta\}$  for the data is credible. “The bootstrap works” is taken to mean that bootstrap distributions, and interesting functionals thereof, converge in probability to the correct limits as sample size  $n$  increases. The convergence is typically established pointwise for each value of  $\theta$  in the parameter space  $\Theta$ . A template argument: Suppose that  $\Theta$  is metric and that (a)  $\hat{\theta}_n \rightarrow \theta$  in  $P_{\theta,n}$ -probability as  $n \rightarrow \infty$ ; (b) for any sequence  $\{\theta_n \in \Theta\}$  that converges to  $\theta$ ,  $H_n(\theta_n) \Rightarrow H(\theta)$ . Then  $H_n(\hat{\theta}_n) \Rightarrow H(\theta)$  in  $P_{\theta,n}$ -probability. Moreover, any weakly continuous functional of the bootstrap distribution converges in probability to the value of that functional at the limit distribution.

Such equicontinuity reasoning, in various formulations, is widespread in the literature on bootstrap convergence. For statistical models of practical interest, considerable insight may be needed to devise a metric on  $\Theta$  such that the template sufficient conditions both hold. Some early papers on bootstrap convergence after Efron (1979) are Bickel and Freedman (1981), Hall (1986), Beran (1987). Broader references are the books and monographs by Hall (1992), Mammen (1992), Efron and Tibshirani (1993), Davison and Hinkley (1997) and the review articles in the bootstrap issue of *Statistical Science* **18** (2003).

These references leave the impression that bootstrap methods often work, in the sense of correct pointwise asymptotic convergence or pointwise second-order accuracy, at every  $\theta$  in the parameter space  $\Theta$ . Counter-examples to this impression have prompted further investigations. One line of research has established necessary and sufficient conditions for correct pointwise convergence of bootstrap distributions as  $n$  tends to infinity (cf. Beran (1997), van Zwet and van Zwet (1999)).

In another direction, Putter (1994) showed: Suppose that the parameter space  $\Theta$  is complete metric and that (a)  $H_n(\theta) \Rightarrow H(\theta)$  for every  $\theta \in \Theta$  as  $n \rightarrow \infty$ ; (b)  $H_n(\theta)$  is continuous in  $\theta$ , in the topology of weak convergence, for every  $n \geq 1$ ; (c)  $\hat{\theta}_n \rightarrow \theta$  in  $P_{\theta,n}$ -probability for every  $\theta \in \Theta$  as  $n \rightarrow \infty$ . Then  $H_n(\hat{\theta}_n) \Rightarrow H(\theta)$  in  $P_{\theta,n}$ -probability for “almost all”  $\theta \in \Theta$ . The technical definition of “almost all” is a set of Baire category II. While “almost all”  $\theta$  may sound harmless, the failure of bootstrap convergence on a tiny set in the parameter space typically stems from non-uniform convergence of bootstrap distributions over neighborhoods of that set. When that is the case, pointwise limits are highly deceptive.

To see this concretely, let  $\hat{\theta}_{n,S}$  denote the James-Stein estimator for an unknown  $p$ -dimensional mean vector  $\theta$  on which we have  $n$  i.i.d. observations, each having a  $N(0, I_p)$  error. Let  $H_n(\theta)$  be the sampling distribution of the root  $n^{1/2}(\hat{\theta}_{n,S} - \theta)$  under this model. As  $n$  tends to infinity with  $p \geq 3$  fixed, we find (cf. Beran (1997)):

- (a) The natural bootstrap distribution  $H_n(\bar{X}_n)$ , where  $\bar{X}_n$  is the sample mean

vector, converges correctly almost everywhere on the parameter space, except at  $\theta = 0$ . A similar failure occurs for the bootstrap distribution  $H_n(\hat{\theta}_{n,S})$ .

(b) The weak convergences of the sampling distribution  $H_n(\theta)$  and of the two bootstrap distributions just described are *not* uniform over neighborhoods of the point of bootstrap failure,  $\theta = 0$ .

(c) The exact quadratic risk of the James-Stein estimator strictly dominates that of  $\bar{X}_n$  at every  $\theta$ , especially at  $\theta = 0$ . If the dimension  $p$  is held fixed, the region of substantial dominance in risk shrinks towards  $\theta = 0$  as  $n$  increases. The asymptotic risk of the James-Stein estimator dominates that of the sample mean only at  $\theta = 0$ . That the dominance is strict for every finite  $n \geq 1$  is missed by the non-uniform limit. Apt in describing non-uniform limits is George Berkeley's celebrated comment on infinitesimals: "ghosts of departed quantities."

In the James-Stein example, correct pointwise convergence of bootstrap distributions as  $n$  tends to infinity is an inadequate "bootstrap works" concept, doomed by lack of uniform convergence. The example provides a leading instance of an estimator that dominates classical counterparts in risk and fails to bootstrap naively. The message extends farther. Stein (1956, first section) already noted that multiple shrinkage estimators, which apply different shrinkage factors to the summands in a projective decomposition of the mean vector, are "better for most practical purposes." Stein (1966) developed multiple shrinkage estimators in detail. In recent years, low risk multiple shrinkage estimators have been constructed implicitly through regularization techniques, among them adaptive penalized least squares with quadratic penalties, adaptive submodel selection, or adaptive symmetric linear estimators. Naive bootstrapping of such modern estimators fails as it does in the James-Stein case.

Research into these difficulties has taken two paths: (a) devising bootstrap patches that fix *pointwise* convergence of bootstrap distributions as the number of replications  $n$  tends to infinity (cf. Beran (1997) for examples and references to the literature); (b) studying bootstrap procedures under asymptotics in which the dimension  $p$  of the parameter space increases while  $n$  is held fixed or increases. Large  $p$  bootstrap asymptotics turn out to be uniform over usefully large subsets of the parameter space and yield effective bootstrap confidence sets around the James-Stein estimator and other regularization estimators (cf. Beran (1995), Beran and Dümbgen (1998)). The first section of Stein (1956) foreshadowed the role of large  $p$  asymptotics in studies of modern estimators.

## References

- [1] Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74** 457–468.
- [2] Beran, R. (1995). Stein confidence sets and the bootstrap. *Statistica Sinica* **5** 109–127.
- [3] Beran, R. (1997). Diagnosing bootstrap success. *Annals of the Institute of Statistical Mathematics* **49** 1–24.
- [4] Beran, R. and Dümbgen, L. (1998). Modulation of estimators and confidence sets. *Annals of Statistics* **26** 1826–1856.

- [5] Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics* **9** 1196–1217.
- [6] Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press.
- [7] Doob, J. L. (1972). William Feller and twentieth century probability. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (L. M. Le Cam, J. Neyman, E. L. Scott, eds.) **II**, xv–xx. University of California Press, Berkeley and Los Angeles.
- [8] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7** 1–26.
- [9] Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- [10] Hall, P. (1986). On the bootstrap and confidence intervals. *Annals of Statistics* **14** 1431–1452.
- [11] Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*. Springer, New York.
- [12] Mammen, E. (1992). *When Does Bootstrap Work?* Lecture Notes in Statistics **77**. Springer, New York.
- [13] Putter, H. (1994). *Consistency of Resampling Methods*. Ph.D. dissertation, Leiden University.
- [14] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (J. Neyman, ed.) **I**, 197–206. University of California Press, Berkeley and Los Angeles.
- [15] Stein, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. In *Festschrift for Jerzy Neyman* (F. N. David, ed.) 351–364. Wiley, New York.
- [16] van Zwet, E. W. and van Zwet, W. R. (1999). A remark on consistent estimation. *Mathematical Methods of Statistics* **8** 277–284.