THE BOOTSTRAP AND EMPIRICAL LIKELIHOOD

Tim C. Hesterberg, MathSoft Inc., 1700 Westlake Ave N., Seattle, WA 98109, timh@statsci.com

Key Words: Bootstrap Tilting, Importance Sampling.

Abstract:

We point out connections between the bootstrap and empirical likelihood, and indicate how the other presentations in the session "The Bootstrap and Empirical Likelihood" at the 1997 Joint Statistical Meetings fit into this framework.

1. Empirical Likelihood for a mean

The goal of this article is to point out connections between the bootstrap and empirical likelihood (EL). We consider only univariate statistics, where the connections are clearest.

We assume that $x_1, \ldots, x_n$ are an i.i.d. sample, and wish to test $H_0 : \mu = \mu_0$. The principle behind EL [11] is to restrict consideration to distributions with support on the observed data, then do maximum likelihood inference. This implies that we may describe a distribution in terms of the probabilities $w_1, \ldots, w_n$ the distribution assigns to the observed data. The maximum likelihood distribution satisfying H_0 is obtained by maximizing the likelihood $\prod_{i=1}^n w_i$ subject to the constraints

$$w_i \geq 0, \sum w_i = 1, \sum w_i x_i = \mu_0. \qquad (1)$$

The solution can be written in the form

$$w_i = \frac{c}{1 - t(x_i - \bar{x})} \qquad (2)$$

where c is a normalizing constant and t is a "tilting parameter". The choice $t = 0$ corresponds to uniform weights $w_i \equiv 1/n$, the unrestricted maximum likelihood estimate; the t that solves $\sum w_i x_i = \mu_0$ is found numerically. Figure 1 shows the resulting weights in an example where the unweighted sample mean is below μ_0; t is positive and larger weights are assigned to the larger values of x to make the weighted mean equal μ_0. Finally, H_0 is rejected in favor of $H_a : \mu \neq \mu_0$ if

$$-2\log(\prod w_i / \prod(1/n)) > \chi^2_\alpha, \qquad (3)$$

which gives asymptotically correct inferences under general conditions [11].

This work was partially funded by NIH SBIR 2R44CA67734-02.

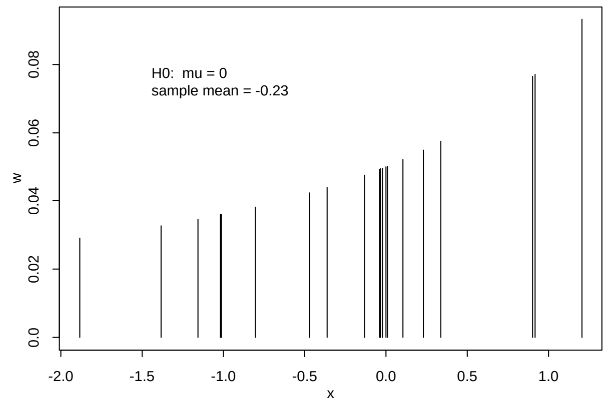

Figure 1: Empirical Likelihood Weights

The first connection with the bootstrap is provided by [15], who suggested using the bootstrap to to calibrate the EL procedure, rather than rely on the asymptotic arguments underlying (3). In another presentation in this session, Lee [9] suggest doing the same thing for a variation of EL. [5] use bootstrap estimation of Bartlett correction factors for EL, and [12, 14] use bootstrap calibration for likelihood ratio tests.

2. Bootstrap Tilting for a mean

We turn now to the bootstrap tilting (BT) test for a single mean. This is based on [3], though we consider hypothesis testing to emphasize connections with EL. One begins by choosing a least-favorable family; the choices in [3] correspond to either exponential tilting:

$$w_i = c\exp(t(x_i - \bar{x})) \qquad (4)$$

or (2), which we call "maximum likelihood tilting". As in EL, t is chosen to satisfy $\sum w_i x_i = \mu_0$. However, the decision to reject H_0 is based on a p-value estimated by weighted bootstrap sampling

$$\hat{p} = P^*_{\mathbf{w}}(\bar{X}^* > \bar{x})$$

where $P^*_{\mathbf{w}}$ indicates bootstrap sampling with probabilities $\mathbf{w} = (w_1, \ldots, w_n)$ and $\bar{X}^*$ is a bootstrap sample mean. Thus, where EL relies on asymptotic results, BT estimates the null distribution for $\bar{X}$, using the empirical maximum likelihood distribution in place of the unknown true distribution.

3. Bootstrap Tilting and Empirical Likelihood for nonlinear statistics

BT and EL extend to nonlinear statistics in similar ways. Suppose that the statistic of interest, θ, is symmetric in its arguments. Restricting consideration to distributions with support on the empirical data, we may express θ as a function of the probabilities, $\theta = \theta(\mathbf{w})$. Derivatives of θ with respect to the weights,

$$U_i(\mathbf{w}) = \lim_{\epsilon \to 0} \epsilon^{-1}(\theta((1-\epsilon)\mathbf{w} + \epsilon\delta_i) - \theta(\mathbf{w})) \quad (5)$$

are know variously as empirical influence function, infinitesimal jackknife, or score function. They may be calculated analytically or approximated numerically using a finite value of ϵ. These derivatives are used in (2) or (4) in place of $(x_i - \bar{x})$. One variation of both BT and EL involves evaluating the derivatives only once, at the vector of uniform weights

$$\mathbf{w}_0 = (1/n, \ldots, 1/n)$$

then tilting using one of:

$$w_i = \begin{cases} c\exp(tU_i(\mathbf{w}_0)) \\ c(1 - tU_i(\mathbf{w}_0))^{-1}. \end{cases}$$

Actually solving the optimizations underlying either EL or BT corresponds to tilting with

$$w_i = \begin{cases} c\exp(t(U_i(\mathbf{w}) - \overline{U})) \\ c(1 - t(U_i(\mathbf{w}) - \overline{U}))^{-1} \end{cases}$$

where the derivatives are continually updated. In any case, t is chosen to solve $\theta(\mathbf{w}) = \theta_0$.

[15] shows in the EL context that the derivatives need not be continually updated, that little is lost by alternating a small number of times between evaluating the derivatives and solving the constraints. This could be done for BT as well. [6] discusses methods in a bootstrap context that correspond to updating derivatives once, using either finite difference methods or linear regression with a single quadratic interaction term.

4. What kind of tilting?

Early BT work provides little basis for choosing between tilting methods. At one extreme is exponential tilting using derivatives evaluated once; this is fastest, and is used by [3] in the BT context and Lee [9] in EL with bootstrap calibration, in another presentation in this session. At the other extreme is maximum likelihood tilting, with updated derivatives. That the EL derivation results in ML tilting with updated derivatives provides some heuristic justification for using these variations in BT. There are substantive reasons for these choices as well. First, updating the derivatives at least once results in more conservative inferences—wider confidence intervals and smaller Type I error—which are usually more accurate in practice.

Second, using ML tilting also results in more conservative inferences. Note that the Taylor-series expansions of (2) in t has a quadratic term double that of (4). Figure 2 shows how the ML weights are larger at both extremes of the data (and smaller in the middle, because the weights are normalized). Bootstrap sampling using ML weights results in greater variance, hence wider confidence intervals and smaller type I error.

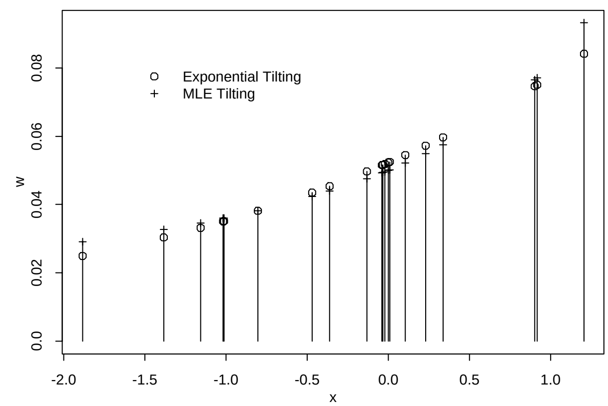

Figure 2: ML vs. exponential tilting

Furthermore, ML tilting has a nice unbiasedness property—when H_0 is true and $\mathbf{w}$ is obtained by ML tilting, then

$$E^*_{\mathbf{w}}(h(X^*)) = \sum w_i h(x_i) = E(h(X)) + O(n^{-2})$$

for any nonlinear function h, subject to regularity conditions ([7]). In particular, the variance of the sample mean under weighted bootstrap sampling biased by a factor of only $O(n^{-2})$, in contrast to n^{-1} for usual bootstrap sampling.

[2] obtained an important result, that BT confidence intervals are second-order correct (one-sided coverage errors are $O(n^{-1})$ under general circumstances, for both linear and nonlinear statistics. This is the same asymptotic accuracy as for other leading bootstrap procedures. We conjecture that BT inference using ML tilting and updated derivatives has *better finite-sample accuracy than other bootstrap intervals*, due to the small bias for the variance. This should be tested in simulation studies.

5. Importance Sampling Reweighting

A very efficient implementation of BT [3] does not in use involve weighted bootstrap sampling, but rather unweighted bootstrap sampling, adjusted by importance sampling reweighting.

In importance sampling, a single set of simulation observations from a single "design" distribution g may provide estimates under multiple "target" distributions f, using different weights. In terms of expected values,

$$\int \theta(x)f(x)dx = \int \theta(x)\frac{f(x)}{g(x)}g(x)dx$$

By sampling from g, and reweighting observations using the ratio $f(x)/g(x)$, we obtain weighted samples for f. ([7] discusses variations on this basic rule.) In the BT context, g corresponds to unweighted bootstrap sampling, and different f's to different values of the tilting parameter t and hence $\mathbf{w}$.

Importance sampling reweighting is used in bootstrap likelihood [1], parametric bootstrap recycling [10], Ventura's presentation in this session on nonparametric bootstrap recycling [13], and a bootstrap diagnostic procedure under investigation by Hinkley and me, which studies sensitivity of bootstrap distributions to changes in the underlying distribution by letting f change along a least-favorable family.

6. Other connections

In the final presentation in this session, Kitamura [8] applies EL ideas to the block bootstrap method for time series. Least-favorable families play a role in the derivation of some bootstrap inference procedures, including BC-a intervals [4] and automatic percentile intervals [2]. [1] studies bootstrap likelihood and EL, and discusses relative advantages of EL and bootstrap methods.

References

[1] A. C. Davison, D. V. Hinkley, and B. J. Worton. Bootstrap likelihoods. *Biometrika*, 79(1):113–130, 1992.

[2] T. J. DiCiccio and J. P. Romano. Nonparametric confidence limits by resampling methods and least favorable families. *International Statistical Review*, 58(1):59–76, 1990.

[3] B. Efron. Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics*, 9:139 – 172, 1981.

[4] B. Efron. Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association*, 82:171 – 200, 1987.

[5] P. Hall and B. La Scala. Methodology and algorithms of empirical likelihood. *International Statistical Review*, 1990.

[6] Tim C. Hesterberg. Tail-specific linear approximations for efficient bootstrap simulations. *Journal of Computational and Graphical Statistics*, 4(2):113–133, June 1995.

[7] Tim C. Hesterberg. Weighted average importance sampling and defensive mixture distributions. *Technometrics*, 37(2):185–194, 1995.

[8] Y. Kitamura. Empirical likelihood and the bootstrap for time series regressions. unpublished, Department of Economics, University of Minnesota, July 25 1997.

[9] S. M. S. Lee and G. A. Young. Bootstrap and improved nonparametric likelihood ratio confidence intervals. unpublished, Department of Statistics, U. of Hong Kong, July 1 1997.

[10] M. A. Newton and C. J. Geyer. Bootstrap recycling: A Monte Carlo alternative to the nested bootstrap. *Journal of the American Statistical Association*, 89(427):905–912, 1994.

[11] Art Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75:237–249, 1988.

[12] W. R. Schucany, W. H. Frawley, S. Wang, and H. L. Gray. Multivariate testing with positive orthant alternatives. unpublished, Southern Methodist University, received September 96 1996.

[13] V. Ventura. *Likelihood Inference by Monte Carlo Methods and Efficient Nested Bootstrapping.* PhD thesis, Oxford University, 1997. See 1997 Proceedings of the Statistical Computing Section of the ASA.

[14] S. Wang, W. A. Woodward, H. L. Gray, S. Wiechecki, and S. R. Sain. A new test for outlier detection from a multivariate mixture distribution. unpublished, Department of Statistics, Texas A&M University, received September 1996.

[15] A. T. A. Wood, K. A. Do, and B. M. Broom. Sequential linearization of empirical likelihood constraints with application to u- statistics. *Journal of Computational and Graphical Statistics*, 5(4):365–385, 1996.