Bootstrap with Kernel Smoothing for Discrete Data

by

Rudy Guerra, Alan M. Polansky, and William R. Schucany
Department of Statistical Science
Southern Methodist University
Dallas, TX 75275-0332

Bootstrap with Kernel Smoothing for Discrete Data

Rudy Guerra, Alan M. Polansky and William R. Schucany

Department of Statistical Science
Southern Methodist University
Dallas, Texas 75275-0332, U.S.A.

SUMMARY

The smoothed bootstrap paradigm involves replacing the empirical distribution function F_n with a smoothed version. Thus far, this idea has been mainly considered for continuous data arising from a distribution F with density f. In this paper a class of smoothed bootstraps for discrete data is presented. Varieties of these proposed resampling methods are then compared with the standard bootstrap in a simulation study of percentile method and bootstrap-t confidence intervals. The Monte Carlo simulation involves samples from two large real populations obtained from toxicological research. Some versions of the smoothed bootstrap yield a worthwhile small-sample improvement in coverage for this application to a ratio estimator with integer-valued data. The potential gains from this methodology are evident, as are the dangers from oversmoothing.

Key words: Density estimation; Confidence interval; Simulation; Proportion; Teratology.

1 Introduction

The bootstrap, introduced by Efron (1979), is a general method for estimating measures of statistical accuracy. An elementary application involves estimating the standard error $\sigma(F)$ of an estimate $\hat{\theta}(X_1, ..., X_n)$ of an unknown parameter $\theta(F)$, where $X_1, ..., X_n$ is a random sample from an unknown probability distribution F. The standard bootstrap estimates $\sigma(F)$ by substituting F_n, the empirical distribution function, for F. Usually $\sigma(F_n)$ does not have an analytically closed form but can be estimated by resampling as follows. Obtain a bootstrap sample $X_1^*, ..., X_n^*$ of n independent draws from F_n, and evaluate $\hat{\theta}^* = \hat{\theta}(X_1^*, ..., X_n^*)$. Independently repeat this process a large number (B) of times, obtaining bootstrap replicates of $\hat{\theta}^*$. The sample standard error based on the B bootstrap replicates is then an estimate for $\sigma(F_n)$. A readable description of the topic is available in the monograph by Efron and Tibshirani (1993).

Smoothing the bootstrap calls for replacing the empirical distribution by a smoothed version. The motivation is that bootstrap resamples not be restricted to only those values in the original sample. In the continuous case F_n is replaced by $\hat{F}_h$, the distribution associated with estimating the underlying density f by a kernel density estimate $\hat{f}_h$, where the smoothing parameter h determines the amount of smoothing. The evidence of substantial improvements in mean square error due to resampling from $\hat{F}_h$ has not been overwhelming. For a recent review of these ideas see De Angelis and Young (1992); also see Efron(1979, 1982), Silverman and Young (1987), and Hall, DiCiccio, and Romano (1989). As an example

of one study that does not focus on mean squared error see Banks (1988).

One version of a smoothed bootstrap for discrete data has been investigated by Frangos and Schucany (1995) (FS). The model for the bivariate integer data (x_i, n_i) that they consider leads to a two-stage resampling procedure for the bootstrap. The first step draws a random sample of pairs (x_i^*, n_i^*) with replacement from (x_i, n_i) as with the standard bootstrap. The second step assumes a parametric form for the conditional distribution of X given n, and thus draws a random sample (x_i^{**}, n_i^{**}), where x_i^{**} is distributed according to an estimated conditional distribution given that $n_i^{**} = n_i^*$. The final bootstrap sample is comprised of the (x_i^{**}, n_i^{**}). An important feature of this scheme is that the pairing of (x_i, n_i) is retained, and by not imposing a model at the first step there is still a significant nonparametric component to the methodology. The second stage is the "smoothing" step: The sample space of x_i is known to be $0, 1, ..., n_i$ and it may be desirable to resample from this full range. Knowledge of the integer lattice with finite support distinguishes this case from the continuous. This is the basis for expecting some improvements in the discrete case by reassigning some probability to the *empty* cells which would otherwise not be resampled.

In this paper we replace the parametric smoothing imposed by FS with a nonparametric construction, namely, estimating a discrete conditional distribution with a discrete kernel density estimator. Section 2 discusses two possible kernel estimators, while Section 3 uses these estimators for bootstrapping. Section 4 compares the kernel based smoothed bootstrap with the standard bootstrap via coverage and average lengths of the associated confidence intervals, and Section 5 contains some concluding remarks.

2

2 Estimators for Probability Mass Functions

Smooth estimators for probability mass functions have been considered by several authors. Smoothing of observed relative frequencies in multinomial situations is the subject of early articles by Good (1965), Fienberg and Holland (1973), and Stone (1974). Kernel methods have been considered by Aitchison and Aitken (1976), and more recently by Titterington (1980), Hall (1981), Wang and Van Ryzin (1981), Bowman, Hall and Titterington (1984), and Hall and Titterington (1987).

In the present article we investigate the kernel methods suggested by Titterington (1980) and Wang and Van Ryzin (1981) as representative of two classes of linear smoothers with kernels appropriate for ordered categories and distinctly different methods for choosing the weights. As in the continuous case, we expect that the choice of kernel shape matters much less than does the selection of smoothing parameter. (See Wang and Van Ryzin (1981) page 302 for their simulation experience with "more sophisticated weights".)

Let $\mathcal{D}$ denote a sample of n independent observations $x_1, ..., x_n$ from the probability mass function $p(\cdot)$ defined on the finite discrete sample space $\mathcal{S} = \{e_1, e_2, \ldots, e_k\}$. A general kernel estimator of the probability function is

$$\hat{p}(y \mid h, \mathcal{D}) = \frac{1}{n} \sum_{i=1}^{n} K(y, x_i \mid h), \qquad y \varepsilon \mathcal{S}, \tag{1}$$

where $K(\cdot, x \mid h)$ is a kernel function and h a smoothing parameter determining the degree

of smoothing. An equivalent representation is linear in the cell frequencies

$$\hat{p}(y \mid h, \mathcal{D}) = \sum_{j=1}^{k} r_j K_j(y \mid h), \tag{2}$$

where k is the cardinality of $\mathcal{S}$, r_j the observed relative frequency corresponding to cell j ("cell" j in this context is equivalent to a realization whose value is e_j) and $K_j(y \mid h)$ the value of the kernel at cell j. The property that the kernels sum to 1.0 implies

$$\sum_{y \in \mathcal{S}} \hat{p}(y \mid h, \mathcal{D}) = 1. \tag{3}$$

In view of (2), Titterington (1980) proposes a general kernel estimator of the k-dimensional vector of probabilities

$$p^*(C) = C^T r, \tag{4}$$

where $r^T = (r_1, ..., r_k)$ and C is a $k \times k$ matrix with each row representing k nonnegative weights summing to unity. By structuring the linear smoothing matrix C as

$$C = I + (1 - h)G, \tag{5}$$

where $G_{ii} = -1, G_{ij} \geq 0, i \neq j$, and $G1 = 0$ for specific choices of G_{ij} it is possible to obtain some well-known kernel functions. In addition, if the smoothing parameter h is obtained by a minimum mean squared error (MSE) criterion, Titterington (1980) shows that the optimal value h^* satisfies

$$1 - h^* = -tr(VG)/tr(G^T pp^T G + VGG^T), \tag{6}$$

where $V = n^{-1}(\Lambda - pp^T), \Lambda = diag(p_1, ..., p_k)$. In practice, r_j is substituted for p_j in (6) to yield $\hat{h}$, an estimate of h^*. As one example of unlimited range smoothing (URS), we

investigate the G_{ij} from Titterington (1980), for fixed i

$$G_{ij} = \begin{cases} 2j/\{i(k-1)\} & , \ j < i \\ 2(k+1-j)/\{(k+1-i)(k-1)\} & , \ j > i. \end{cases} \quad (7)$$

These weights decline *linearly* on each row of G as j moves away from $j = i$. Consequently, when we use (7) we denote the smoother by URSL. To study weights that fall off *quadratically* we consider

$$G_{ij} = \begin{cases} 6j^2/\{i(k-1)(2i-1)\} & , \ j < i \\ 6(k-j+1)^2/\{(k-i+1)(k-1)(2(k-i)+1)\} & , \ j > i, \end{cases} \quad (8)$$

which yields a smoother to be denoted by URSQ. In addition, we define and consider limited range smoothers (LRS) by setting $G_{ij} = 0$ for all $|i - j| > w$, for some integer w greater than 1.

Wang and Van Ryzin (1981) present a class of estimators for probability functions defined on the integers J. Following their notation, let p_i denote the mass at $i \varepsilon J$ and S_0 denote an interval on the real line containing the origin. Their estimate of p_i, $i \varepsilon J$, is the infinite linear combination

$$p^*(i \mid s) = \sum_{j=-\infty}^{\infty} r_j W(s, i, j), \quad (9)$$

where $W(s, i, j)$ is a discrete window weight function [cf. Parzen (1962)] defined on $S_0 \times J \times J$, with parameter s. As in (2), the estimators are weighted averages of the observed relative frequencies r. Many commonly used estimators are special cases of (9).

The weight function recommended by Wang and Van Ryzin (1981) and that we employ

here is an LRS with uniform weight function over w symmetrically neighboring cells

$$W(s,i,j) = \begin{cases} \frac{1}{2}s/w & , \quad |i-j| = 1, ..., w \\ 1-s & , \quad i = j \\ 0 & , \quad |i-j| \geq w+1 \end{cases} \tag{10}$$

where w is an integer ≥ 1, and $s \varepsilon S_0 = [0,1]$. Even though it is not monotone decreasing, it accounts for ordered categories by using only neighboring cells. For this class of "uniform w" weights an analytically closed form optimal solution $\hat{s}$, based on minimizing global MSE, is given in Table 1 of Wang and Van Ryzin (1981). For example, when $w = 1$, LRS is essentially identical to (4) and (6) with an appropriate G_{ij}. They also discuss choices of s for a local MSE criterion.

For this paper we consider probability functions on the discrete, finite sample space $\mathcal{S} = \{0, 1, ..., m\}$. It is therefore necessary to modify the Wang and Van Ryzin estimator, since it is defined over the complete set of integers. To this end we devise a boundary kernel that folds all of the estimated probability, that would otherwise be beyond the boundaries of the sample space, back onto the endpoints. That is, we assign the total mass associated with $p^*(j|\hat{s}), j < 0$, to $p^*(0|\hat{s})$ and with $p^*(j|\hat{s}), j > m$, to $p^*(m|\hat{s})$. Taking $w = 1$ in (10), i.e., the uniform 1 weight function, our modified estimator becomes

$$p^*(i|\hat{s}) = \begin{cases} (1 - \frac{1}{2}\hat{s})r_0 + \frac{1}{2}\hat{s}r_1 & , \quad i = 0 \\ \frac{1}{2}\hat{s}r_{i-1} + (1 - \hat{s})r_i + \frac{1}{2}\hat{s}r_{i+1} & , \quad 0 < i < m \\ (1 - \frac{1}{2}\hat{s})r_m + \frac{1}{2}\hat{s}r_{m-1} & , \quad i = m \end{cases} \tag{11}$$

6

3 Smoothed Bootstrapping

The specific application that we use to illustrate the proposed approach consists of setting a confidence interval on a mean rate of integer counts based on M independent pairs (x_i, n_i), where n_i is a positive integer, and conditional on n_i the sample space of the random variate X_i is $\{0, 1, 2, ..., n_i\}$. We assume that (X_i, n_i) and (X_j, n_j) are identically distributed conditional on $n_i = n_j$. We will be concerned with bootstrap interval estimation of $\mu = E[X]/E[n]$ based on the pooled ratio estimate $\hat{\mu} = \sum_{i=1}^{M} X_i / \sum_{i=1}^{M} n_i$. This requires a bootstrap distribution of replicates that we specify below. An important aspect of proper resampling is that the simulation preserve the stochastic assumptions stated above. For $b = 1, ..., B$ repeat the following steps.

1. Draw a (standard) bootstrap sample (x_i^*, n_i^*) of size M from the empirical joint distribution of the original data (x_i, n_i).

2. Draw a (smoothed) bootstrap sample (x_i^{**}, n_i^{**}) of size M, where $n_i^{**} = n_i^*$ and x_i^{**} is distributed according to a smoothed version of the conditional distribution of $X|n_i^*$ as in Section 2.

3. Calculate the pooled ratio estimate, $\hat{\mu}^b = \sum_{i=1}^{M} x_i^{**} / \sum_{i=1}^{M} n_i^{**}$.

The empirical distribution of the B bootstrap replicates $\{\hat{\mu}^b\}$ may then be used to construct confidence intervals.

The intervals we consider are based on the percentile (Efron, 1982; Efron and Tibshirani, 1993) and studentized (Abramovitch and Singh, 1985) methods. The percentile method yields a $1 - 2\alpha$ central confidence interval

$$[\hat{H}^{-1}(\alpha), \hat{H}^{-1}(1 - \alpha)], \tag{12}$$

where $\hat{H}$ is an estimate of the bootstrap distribution function, $\hat{H}(t) = \#\{\hat{\mu}^b \leq t\}/B$. The bootstrap-t determines quantiles as in (12), but in step 3 it replicates studentized quantities $(\hat{\mu}^b - \hat{\mu})/\widehat{SE}^b$, where $\widehat{SE}^b$ is an estimate of the standard error of $\hat{\mu}^b$. The standard error for each resample is estimated with the jackknife as in Gladen (1979) and Frangos and Schucany (1995).

The construction of a smoothed version of the conditional distribution $X|n$ is as follows. For each distinct n in the original data, collect all pairs (x_i, n_i) such that $n_i = n$. The original data is thus partitioned into mutually exclusive sets, one for each distinct n. Independently, each data set, denoted by $\mathcal{D}_n$, is then used to construct a smooth estimator for the conditional probability function associated with that value of n. (It is easy to see that this two-stage sampling applied to the unsmoothed $\mathcal{D}_n$ is equivalent to ordinary resampling of pairs.) Note that this approach does not require the same degree of smoothing within each $\mathcal{D}_n$, thereby allowing for a rich variety of patterns among the sets. We do, however, require that the smoothed estimators among the $\mathcal{D}_n$ be of the same type, for example all uniform-1 weight function estimators, but with possibly different smoothing parameters, s_n.

8

The values of x are ordered and any reasonable probability function estimator should incorporate this fact. The Wang and Van Ryzin (1981) class of uniform w weight estimators by definition take account of the natural order in the conditional sample space $\{0, 1, ..., n\}$. The general kernel estimator of Titterington (1980) is also capable of accounting for order by taking G_{ij} in (5) to be decreasing as $|i - j|$ is increasing, $i \neq j$, such as in (7).

Ideally, one wishes to resample from a distribution that is smoother in some sense, but differs as little as possible in other important features. However, the proposed smoothing will change some moments of the resampling distribution, and may introduce bias into the bootstrap estimate. In order to minimize the potential bias one may want to resample from a "smooth" distribution with the same mean and variance as the standard resampling distribution. Unfortunately, the additional desire to preserve the integer lattice support makes the problem considerably more difficult than the continuous case. Indeed, the optimization associated with such constrained optimal smoothing appears to be an integer programming problem. Titterington (1980) has noted similar problems. More discussion of the difficulty of rescaling is presented in the next section.

Both types of estimators handle the degenerate case the same. Specifically, if for a fixed n' only one value, say x_0, is observed in the sample $\mathcal{D}_{n'}$, then the smoothing parameter is selected so that mass 1 is placed at x_0. This corresponds to $h = 1$ in (5) and $s = 0$ in (9).

Although the main emphasis of this paper is on nonparametric smoothing, our simulation study compares this technique with the "semiparametric" approach of Frangos and Schucany

(1995). In step 2 of the above bootstrap procedure the FS resample (x_i^{**}, n_i^{**}) is drawn by

taking $n_i^{**} = n_i^*$ and $x_i^{**}|n_i^{**}$ from a binomial distribution $B(n_i^*, p_i^*)$ with $p_i^* = x_i^*/n_i^*$.

4   A Simulation Study: Small Sample Behavior

To investigate the performance of the various bootstrap procedures on real data we have

defined a study population using toxicological data from Table 3 of Luning *et al.* (1966).

This dataset was also studied by Gladen (1979) to compare jackknife estimates and tests to

methods based on distributional models. The population consists of 604 litters with data

pairs (x, n), where x denotes the number of affected mice out of n total fetuses from one

female. Figure 1 is a plot of frequencies of affected mice, conditional on the total number

of fetuses. The proportion of affected fetuses (μ) in the population is 0.25258 and we

consider confidence intervals for this parameter. Another large population of control mice

($\mu \approx 0.0996$) was also studied. The results presented here are representative of those in the

control population. Carr and Portier (1993) compare the ordinary bootstrap with other

methods in the more difficult problem of dose-response modeling. Here we are studying the

relative performance of various refined resampling procedures at a single fixed dose.

The simulations were performed on a RISC System 6000 at Southern Methodist

University using C code and NAG routines. For each sample size (number of litters) and

estimation procedure there were 1000 independent iterations, each yielding one confidence

interval from $B = 1000$ bootstrap resamples. The nominal 90% intervals are summarized

by the actual coverage and average length across the 1000 iterations. The standard errors for the coverages close to 90% are just less that 1%.

Table 1 summarizes results at 4 sample sizes. The standard bootstrap (STAN), which simply eliminates step 2 of the previous section, significantly undercovers at $M=10$ and 15. A dramatic increase in actual coverage is achieved by the binomial method (BIN), but at a cost of excessively long intervals relative to the standard bootstrap. This significant overcoverage is a noteworthy difference from those found by FS. The uniform-1 LRS is a good compromise. It corrects the actual coverage with average interval lengths considerably smaller than BIN and not much longer than STAN. Similar results were found when the method of Titterington (1980) was chosen to be a limited range smoother as described in Section 2.

To what should we attribute this difference in FS? Although these conditional distributions exhibit no obvious departures from slightly overdispersed binomials, the BIN method behaved significantly different than it did on the FS simulated beta-binomial distributions. This particular semiparametric bootstrap may be somewhat sensitive to the actual model, since the intervals were consistently too wide. This suggests that there are subtle differences between this mouse population and the joint distribution imposed by conditionally mixing binomials at each litter size. The FS method matches the litter-to-litter variation well, but generates too much variability in the ratio estimator. Indeed for fixed binomial populations with $p = 0.25$ at every litter size, the BIN bootstrap covered 97% at $M = 15, 30$. On the other hand, the uniform-1 weight function estimator (LRS), being akin to a nearest neigh-

bor procedure, has a smoothing effect only at and near those points of the sample space where data have been observed.

The URSL, the linear choice of G specified by (7), yields relatively long intervals, but also obtains poor coverage as M increases. The long intervals can be explained by reasoning similar to that given above for the binomial method. The poor coverage results from bias introduced in the ratio estimator for the smoothed samples. Less smoothing, represented by URSQ was a step in the right direction. However, it was not sufficient to rescue URS entirely. We also experimented with relocating and rescaling the smoothed resamples in an effort to control this bias, as is often done in the continuous case. In this specific application, however, standard approaches can yield exactly the same intervals as the standard bootstrap since one might be actually controlling the numerator of $\hat{\mu}^b$ by imposing the desired moments. More sophisticated approaches to controlling the bias are needed and certainly warrant further investigation.

The bootstrap-t corrected the coverage but the intervals at the smaller sample sizes were longer than the satisfactory LRS intervals. It may be worth noting that LRS also represents less computing than the bootstrap-t, which estimates a standard error for each resample. The bootstrap-t did not produce desirable results when we used it after either LRS, URSL, or URSQ smoothing.

What one sees is that the best of the various intervals was the percentile method based on LRS. The satisfactory improvement from LRS relative to STAN (unsmoothed) appears

to be due to a tradeoff of less skewness for more spread. This is consistent with the general idea that the percentile method works best on a symmetric pivot. Choosing the extent of smoothing is critical. The window $w = 1$ was successful, but the unlimited range was disastrous here.

5 Discussion

Previous investigations of bootstrap smoothing consider continuous data and focus on mean squared error performance to judge the relative merits of a smoothed bootstrap. For the most part the results indicate no global preference for the smoothed bootstrap, De Angelis and Young (1992). In this paper we depart from these investigations in that we consider discrete data and assess the merits of a smoothed bootstrap through confidence intervals. One reason to anticipate a better payoff than with continuous distributions, is that one may capitalize on the *known* integer lattice. In both situations the choice of smoothing parameters may be a delicate matter. However, the important distinction between the two is that, especially for integer-valued random variables, one may more faithfully prescribe the support for the resamples. These discrete smoothers impose a structure that constitutes fundamentally different information about the nature of the range of the random variable.
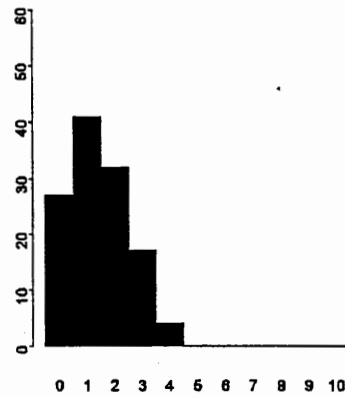
We have shown that employing a smoothed version of the empirical distribution, when using the bootstrap in confidence interval construction, improves the coverage in small samples without a substantially increased interval length relative to the standard bootstrap.

At least in the present situation, the best procedure, namely the uniform-1 estimator, does not actually yield extensive smoothing. In the uniform-1 method the mass for an observation is smeared to only those values immediately above and below that observed point. These results agree with those of Wang and Van Ryzin (1981), who have noted that the uniform-1 method performs "approximately as well or better than the other more sophisticated weight functions." The key feature may be that the support of the kernel is limited to $\pm w$. The two less successful competitors evidently do too much smoothing over the full range of x_i.
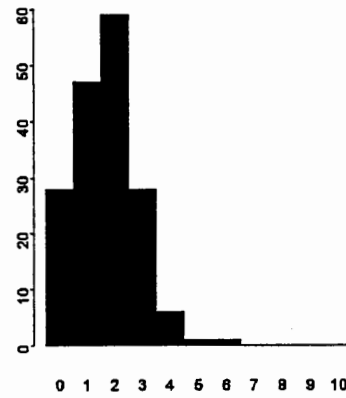
There is a distinct difference between these kernels and those with compact support in the continuous case, where the smoothing parameter controls the window width. Here one may fix the range, for example at $w = 1$, and then select the amount of smoothing to only those neighboring cells. Ideally, the two would be chosen simultaneously. We suspect that it would also be better to select the smoothing parameter(s) from a criterion directly related to coverage. This warrants additional research. One complication pursuant to this idea that the relevant Edgeworth expansions derived by Hall (1992) do not hold for the lattice case. In general, however, Young (1994) and a number of discussants stress the importance of finite-sample practicalities of workable bootstraps.

Figure 1: Conditional Frequencies of Affected Mice in the Luning Dataset
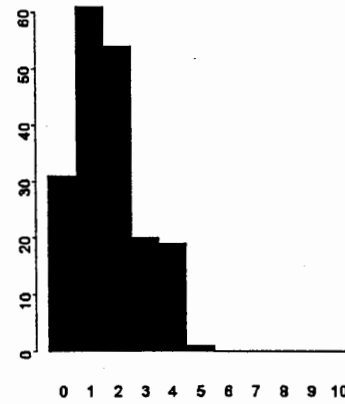
Litter Size = 5 (121 Observations)
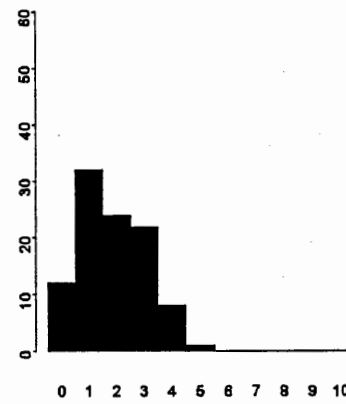
Litter Size = 6 (170 Observations)
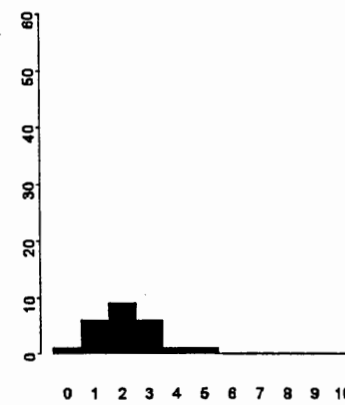
Litter Size = 7 (186 Observations)

Litter Size = 8 (99 Observations)

Litter Size = 9 (24 Observations)

Litter Size = 10 (4 Observations)
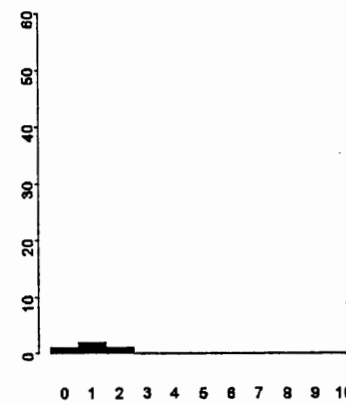
Table 1: Bootstrap Confidence Intervals for a Proportion

M	Percentile					Bootstrap-t			
	STAN	BIN	LRS	URSL	URSQ	STAN	LRS	URSL	URSQ
10	84.5	96.4	89.6	85.2	85.8	90.1	91.4	79.0	80.2
	.170	.237	.185	.233	.213	.209	.210	.194	.198
15	86.6	96.1	90.2	78.7	82.1	91.5	89.4	69.8	73.3
	.144	.195	.154	.195	.178	.163	.160	.148	.151
30	88.6	96.2	92.0	61.0	68.4	90.7	88.0	52.2	57.7
	.104	.139	.112	.140	.128	.110	.109	.100	.102
50	89.6	96.6	91.0	41.8	48.8	91.1	88.4	45.5	50.1
	.082	.108	.087	.107	.099	.084	.083	.077	.079

The first entry is the empirical coverage in percent (SE $\approx$ 1.0) and the second entry is average length of the confidence interval. The nominal coverage is $1 - 2\alpha = 0.90$.

References

Abramovitch, L. and Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *The Annals of Statistics* **13**, 116-132.

Aitchison, J. and Aitken, C.G.G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika* **63**, 413-20.

Banks, D.L. (1988). Histospline smoothing the Bayesian bootstrap. *Biometrika* **75**, 673-84.

Bowman, A.W., Hall, P., and Titterington, D.M. (1984). Cross-validation in nonparametric estimation of probabilities and probability densities. *Biometrika* **71**, 341-51.

Carr, G. and Portier, C. (1993). An evaluation of some methods for fitting dose-response models to quantal-response developmental toxicology data. *Biometrics* **49**, 779-791.

De Angelis, D. and Young, G.A. (1992). Smoothing the bootstrap. *International Statistical Review* **60**, 45-56.

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

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

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

Fienberg, S.E. and Holland, P.W. (1973). Simultaneous estimation of multinomial cell probabilities. *Journal of the American Statistical Association* **68**, 683-91.

Frangos, C.C. and Schucany, W.R. (1995). Improved bootstrap confidence intervals in toxicological experiments. to appear *Communications in Statistics A, Theory and Methods* **25**, 3.

Gladen, B. (1979). The use of the jackknife to estimate proportions from toxicological data in the presence of litter effects. *Journal of the American Statistical Association* **74**, 278-83.

Good, I.J. (1965). *The Estimation of Probabilities*. Cambridge, MA: MIT Press.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hall, P. (1981). On nonparametric multivariate binary discrimination. *Biometrika* **68**, 287-94.

Hall, P., DiCiccio, T.J., and Romano, J.P. (1989). On smoothing and the bootstrap. *The Annals of Statistics* **17**, 692-704.

Hall, P. and Titterington, D.M. (1987). On smoothing sparse multinomial data. *Australian Journal of Statistics* **29**, 19-37.

Luning, K.G., Sheridan, W., Ytterborn, K.H., and Gullberg, U. (1966). The relationship between the number of implantations and the rate of intra-uterine death in mice. *Mutation Research* **3**, 444-451.

Parzen, E. (1962). On estimation of a probability function and its mode. *The Annals of Mathematical Statistics* **33**, 1065-76.

Silverman, B.W. and Young, G.A. (1987). The bootstrap: To smooth or not to smooth? *Biometrika* **74**, 469-79.

Stone, M. (1974). Cross-validation and multinomial prediction. *Biometrika* **61**, 509-15.

Titterington, D.M. (1980). A comparative study of kernel-based density estimates for categorical data. *Technometrics* **22**, 259-68.

Wang, M-C and Van Ryzin, J. (1981). A class of smooth estimators for discrete distribution. *Biometrika* **68**, 301-9.

Young, G.A. (1994). Bootstrap: More than just a stab in the dark? (with discussion). *Statistical Science* **9**. 382-415.