SMOOTHING THE BOOTSTRAP FOR USE WITH THE BINOMIAL DISTRIBUTION[1]

Peter Hall

University of North Carolina, Chapel Hill[2]

SUMMARY. It is shown practically and theoretically that a smoothed version of the bootstrap performs well when used to construct confidence intervals for binomial proportions. Smoothing enables rounding error to be minimized in a uniform manner. Without smoothing, the size of rounding error can almost double. Generalizations to other lattice distributions are indicated.

KEY WORDS. Binomial distributions, Bootstrap, Normal approximations, One-sided confidence intervals, Rounding error, Skewness.

---

[1]Research supported by USAF Grant No. F49620 82 C 0009.

[2]On leave from Australian National University.

1. INTRODUCTION

The purpose of this note is to show practically and theoretically that a certain smoothed version of Efron's (1979, 1982) bootstrap is a very satisfactory method of constructing confidence intervals for binomial proportions. A direct bootstrap approach performs poorly by comparison. Generalizations to other lattice distributions will be indicated.

As Beran (1984) points out, the general advantages and disadvantages of smoothing are not well understood at present. We suggest that in the case of a lattice distribution, some degree of smoothing is essential to ensure that coverage probabilities are close to the levels desired.

Let us briefly discuss the benefits of smoothing. It was shown in Hall (1982) that if we seek a one-sided binomial confidence interval with coverage probability α, we should be content if the true converge probability is actually

$$\alpha + \tfrac{1}{2}\, s_n(\alpha)(npq)^{-\frac{1}{2}}\phi(z) + o(n^{-\frac{1}{2}}) \qquad (1.1)$$

for a sequence of numbers $s_n(\alpha)$ in the interval $[-1,1]$, where $\phi = \Phi'$. Again, z is the solution of $\Phi(z)=\alpha$. The term of order $n^{-\frac{1}{2}}$ in (1.1) describes the type of rounding error inherent in sampling from the binomial distribution. Singh's (1981) result (1.7), and its proof, suggest that a certain version of the bootstrap will result in rounding errors up to *twice* those described by (1.1). Indeed, if we define

$$F(x) \equiv \mathrm{pr}\{n^{\frac{1}{2}}(\hat{p}_1 - p)/(pq)^{\frac{1}{2}} \le x\} \ ,$$

and its bootstrap approximation

$$\hat{F}(x) \equiv \text{pr}\{n^{\frac{1}{2}}(\hat{p}_2-\hat{p}_1)/(\hat{p}_1\hat{q}_1)^{\frac{1}{2}} \le x|\hat{p}_1\} \ ,$$

then the absolute error between $F(x)$ and $\hat{F}(x)$ can be as large as $(npq)^{-\frac{1}{2}}\phi(x)$, not just $\frac{1}{2}(npq)^{-\frac{1}{2}}\phi(x)$. This point is made explicit by Theorem 1 below. By smoothing we are able to reduce the worst-case error to its "optimal" size $\frac{1}{2}(npq)^{-\frac{1}{2}}\phi(x)$.

The bootstrap is essentially a first-order correction for skewness. An alternative skewness correction was suggested by Hall (1982), and was assessed using criteria similar to those employed here. As a rule, the smoothed bootstrap performs slightly *better* than Hall's correction, due we suggest to the fact that the bootstrap approximation holds *uniformly*; see our Theorem 2. The improvement is hardly detectable on the scoring scale in Hall (1982); for example, it follows from Table 1 below that the corrections score identically on that scale when n=10 or 20. However, a more detailed analysis using a finer grid of p-values puts the bootstrap ahead.

2. METHODOLOGY

We begin by stating a result which shows that jumps almost as large as *twice* the optimum described by (1.1), can occur using the raw bootstrap approximation. These jumps occur at any nonzero value of x, and for arbitrarily large n.

THEOREM 1. *Assume* $0<p<1$ *and* $p \neq \frac{1}{2}$. *For each* $\varepsilon > 0$ *and* $x \neq 0$,

$$\limsup_{n\to\infty} \text{pr}\{|\hat{F}(x) - F(x)| > (npq)^{-\frac{1}{2}}(1-\varepsilon)\phi(x)\} > 0.$$

A proof is given in Appendix 1. This theorem is close in spirit to result (1.7) of Singh (1981). However, Singh's proof depends crucially on varying the value of x for different N's and n's. We stress that x is held fixed in our Theorem 1, at a value such as 1.645 or 1.960, and so the theorem leads to pessimism about the performance of the raw bootstrap when used to construct confidence intervals.

To reduce rounding error we suggest interpolating at midpoints of histogram blocks. That is, we define $\hat{G}$ by

$$\hat{G}\{(m+\tfrac{1}{2}+u)/(n\hat{p}_1\hat{q}_1)^{\frac{1}{2}}\} \equiv (1-u)\ \hat{F}\{(m+\tfrac{1}{2})/(n\hat{p}_1\hat{q}_1)^{\frac{1}{2}}\} + u\ \hat{F}\{(m+\tfrac{3}{2})/(n\hat{p}_1\hat{q}_1)^{\frac{1}{2}}\}$$

for all integers m and all $u \in [0,1]$. This smoothed bootstrap approximant effectively halves the maximum rounding error. In fact for each x,

$$|\hat{G}(x) - F(x)| \le \tfrac{1}{2}(npq)^{-\frac{1}{2}}\phi(x) + O_p(n^{-1}) .$$

Therefore $\hat{G}$ holds more promise than $\hat{F}$ as a device for constructing confidence intervals.

The function $\hat{G}$ is continuous and *strictly* increasing, from zero to one, on $[-(N+\tfrac{1}{2})(n\hat{p}_1\hat{q}_1)^{-\frac{1}{2}},\ (n+\tfrac{1}{2}-N)(n\hat{p}_1\hat{q}_1)^{-\frac{1}{2}}]$, provided only that $\hat{p}_1$ is neither 0 nor 1. The probability that $\hat{p}_1=0$ or 1 is exponentially small. Given $\alpha \in (0,1)$, let t_α be the solution of $\hat{G}(t_\alpha)=\alpha$, and $z=z(\alpha)$ the solution of $\Phi(z)=\alpha$. Theorem 2 shows that the order of approximation of t_α to the true critical point is best possible, in the sense described by (1.1).

THEOREM 2. *As* $n \to \infty$,

$$\Pr\{n^{\frac{1}{2}}(\hat{p}_1-p)/(pq)^{\frac{1}{2}} \le t_\alpha\} = \alpha + \tfrac{1}{2}(npq)^{-\frac{1}{2}} s_n(\alpha)\phi(z) + o(n^{-\frac{1}{2}})$$

uniformly in α, *where* s_n *takes values only between* -1 *and* +1.

Appendix 2 contains the proof. A longer proof shows that the term $o(n^{-\frac{1}{2}})$ is actually $O(n^{-1})$.

To construct confidence intervals using Theorem 2, observe that for any $t \in (-\infty,\infty)$ the event $n^{\frac{1}{2}}(\hat{p}_1-p)/(pq)^{\frac{1}{2}} \le t$ is equivalent to $n^{\frac{1}{2}}(\hat{p}_1-p) \le f(t|\hat{p}_1)$, where

$$f(t|\hat{p}_1) \equiv [t\{n^{-1}t^2 + 4\hat{p}_1(1-\hat{p}_1)\}^{\frac{1}{2}} + (2\hat{p}_1-1)n^{-\frac{1}{2}}t^2]\{2(1+n^{-1}t^2)\}^{-1} .$$

Thus, the intervals $[\hat{p}_1 - nf(t_\alpha|\hat{p}_1),\infty)$ and $(-\infty,\hat{p}_1 - n^{-\frac{1}{2}}f(t_{1-\alpha}|\hat{p}_1)]$ both cover p with probability given by (1.1).

3. APPLICATION

Suppose we observe a value $N=n\hat{p}_1$ from a binomial Bi(n,p) distribution, and are given a confidence coefficient $\alpha \in (0,1)$. Using binomial tables or otherwise, calculate the integer M which satisfies

$$\pi_1 \equiv \text{pr}(N_2 \le M|\hat{p}_1) \le \alpha \le \text{pr}(N_2 \le M+1|\hat{p}_1) \equiv \pi_2 .$$

Derive t_α by interpolating between the value π_1 taken by $\hat{F}$ at $(M+\frac{1}{2}-n\hat{p}_1)/(n\hat{p}_1\hat{q}_1)^{\frac{1}{2}}$, and the value π_2 taken at $(M+\frac{3}{2}-n\hat{p}_1)/(n\hat{p}_1\hat{q}_1)^{\frac{1}{2}}$. Thus,

$$t_\alpha = \{M+\tfrac{1}{2}+(\alpha-\pi_1)(\pi_2-\pi_1)^{-1} - n\hat{p}_1\}/(n\hat{p}_1\hat{q}_1)^{\frac{1}{2}} .$$

The smoothed confidence interval $I_s \equiv [\hat{p}_1 - n^{-\frac{1}{2}}f(t_\alpha|\hat{p}_1),\infty)$ covers p with probability approximately α. Using the unsmoothed bootstrap, we would choose M* such that $\text{pr}(N_2 \le M^*|\hat{p}_1)$ is as close as possible to α, then

set $t_\alpha^* \equiv (M-n\hat{p}_1)/(n\hat{p}_1\hat{q}_1)^{\frac{1}{2}}$, and finally construct the confidence interval $I_u \equiv [\hat{p}_1 - n^{-\frac{1}{2}} f(t_\alpha^* | \hat{p}_1), \infty)$.

For a given value of p, the probabilities that I_s and I_u cover p may be computed exactly, after some tedious algebra, using binomial tables. They equal values of the binomial cumulative distribution function. In the cases n=10 and n=20 and α=0.95, the unsmoothed and smoothed bootstrap approximations are compared in Table 1 on the basis of how close the coverage probabilities are to the optimal probability, π. Here π equals that probability $\text{pr}(\hat{p}_1 - p \leq u_0)$ which satisfies

$$|\text{pr}(\hat{p}_1 - p \leq u_0) - 0.95| = \inf_u |\text{pr}(\hat{p}_1 - p \leq u) - 0.95| .$$

A tick in the row headed u(ν) means that the unsmoothed interval's coverage probability equals π when n=ν. Likewise for the smoothed interval in the row labeled s(ν).

Table 1. *Comparison of smoothed and unsmoothed bootstrap confidence intervals.*

20p	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
u(10)	✓	✓	✓	×	✓	×	×	✓	×	✓	×	✓	✓	×	×	×	×	×	×
s(10)	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	×	×	×	×
u(20)	✓	✓	✓	×	×	×	×	×	✓	✓	✓	✓	×	×	×	✓	×	×	×
s(20)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×

4. GENERALIZATION

Interpolation may be used in conjunction with the bootstrap to construct confidence intervals for the mean of any lattice-valued distribution.

However, the argument in the last paragraph of Section 2 does not apply generally. Rather, we should proceed as follows. Given a sample $X \equiv \{X_1,\ldots,X_n\}$ from a lattice distribution with span h taking values $\{0,\pm h,\pm 2h,\ldots\}$, define $\overline{X} \equiv n^{-1}\Sigma X_i$ and $s^2(X) \equiv n^{-1}\Sigma (X_i-\overline{X})^2$. Let $Y \equiv \{Y_1,\ldots,Y_n\}$ be a random sample with the bootstrap distribution: for each i, $pr(Y_i = X_j|X) = n^{-1}$, $1 \le j \le n$. Using simulation or otherwise, compute the distribution function

$$\hat{F}_1(x) \equiv pr\{n^{\frac{1}{2}}(\overline{Y}-\overline{X})/s(Y) \le x|X\} .$$

Define $\hat{G}_1$ by interpolating between values of $\hat{F}_1$ at midpoints of histogram blocks. Let t_α be the solution of $\hat{G}_1(t_\alpha)=\alpha$. If $E(X_1)=\mu$, $var(X_1)=\sigma^2$ and $E(|X_1|^{6+\eta})<\infty$ for some $\eta > 0$, then

$$pr\{n^{\frac{1}{2}}(\overline{X}-\mu)/s(X) \le t_\alpha\} = \alpha + \tfrac{1}{2}(h/\sigma n^{\frac{1}{2}})s_n(\alpha)\phi(z) + o(n^{-\frac{1}{2}}) \qquad (4.1)$$

uniformly in $\alpha \in (0,1)$, where s_n takes values only between -1 and 1. Confidence intervals for μ follow immediately. Coverage probabilities are given by an analogue of (1.1).

A proof of (4.1) uses the argument leading to result (22.5) of Bhattacharya and Rao (1976, p. 231), for the bivariate lattice-valued vector $(\Sigma X_i, \Sigma X_i^2)$.

If we used this technique in the binomial case we would derive the distribution of $n^{\frac{1}{2}}(\hat{p}_2-\hat{p}_1)/(\hat{p}_2\hat{q}_2)^{\frac{1}{2}}$, rather than that of the statistic $n^{\frac{1}{2}}(\hat{p}_2-\hat{p}_1)/(\hat{p}_1\hat{q}_1)^{\frac{1}{2}}$ whose distribution function is $\hat{F}$. However, the equivalence noted in the last paragraph of Section 2 implies that there can be no difference between the resulting confidence intervals.

Stimulating discussions with Raymond J. Carroll are gratefully acknowledged.

APPENDIX I. *Proof of Theorem 1.*

Let $g(x) \equiv [x]-x+\frac{1}{2}$. An extension of Singh's (1981) proof of his formula (1.6), and Theorem 6 of Petrov (1974, p. 171), give:

$$\hat{F}(x) - F(x) = (npq)^{-\frac{1}{2}}[g\{(n\hat{p}_1\hat{q}_1)^{\frac{1}{2}}x\} - g\{(npq)^{\frac{1}{2}}x\}]\phi(x) + O_p(n^{-1}).$$

The sequence $(npq)^{\frac{1}{2}}x - [(npq)^{\frac{1}{2}}x]$, $n \geq 1$, is dense in $(0,1)$, and so for each $\varepsilon > 0$ there exists an infinite sequence n_k such that $(n_kpq)^{\frac{1}{2}}x - [(n_kpq)^{\frac{1}{2}}x] \in (0,\varepsilon)$ for all k. This entails $g\{(n_kpq)^{\frac{1}{2}}x\} \in (-\varepsilon+\frac{1}{2},\frac{1}{2})$, and so Theorem 1 with ε replaced by 3ε will follow if we prove that

$$\limsup_{k \to \infty} \text{pr}[g\{(n_k\hat{p}_1\hat{q}_1)^{\frac{1}{2}}x\} \in (-\tfrac{1}{2},2\varepsilon-\tfrac{1}{2})] > 0.$$

Now, $(n\hat{p}_1\hat{q}_1)^{\frac{1}{2}} = (npq)^{\frac{1}{2}} + R$, where $R \equiv (\frac{1}{2}-p)n^{\frac{1}{2}}(\hat{p}_1-p)/(pq)^{\frac{1}{2}} + O_p(n^{-\frac{1}{2}})$. Also, $n^{\frac{1}{2}}(\hat{p}_1-p)/(pq)^{\frac{1}{2}} \to N(0,1)$ in distribution. Therefore the probability that $R-[R] \in (1-2\varepsilon,1-\varepsilon)$ is bounded away from zero as $n \to \infty$. If $R-[R] \in (1-2\varepsilon,1-\varepsilon)$ then since $(n_kpq)^{\frac{1}{2}} \in (m_k,m_k+\varepsilon)$ for some integer m_k, we must have $(n_k\hat{p}_1\hat{q}_1)^{\frac{1}{2}} \in (m-2\varepsilon,m)$ for some integer m. But this entails $g\{(n_k\hat{p}_1\hat{q}_1)^{\frac{1}{2}}\} \in (-\frac{1}{2},2\varepsilon-\frac{1}{2})$, and so we are done.

APPENDIX II. *Proof of Theorem 2.*

Let $C_1, C_2, \ldots$ be positive constants. The proof is in three steps.

Step (i). A modification of the proof of Petrov's (1974, p. 171) Theorem 6 shows that on the set $E \equiv \{|\hat{p}_1-p| \leq n^{-\frac{1}{4}}\}$, which satisfies $P(\tilde{E}) \leq C_1n^{-3/4}$, we have

$$\hat{F}(x) \leq \Phi(x) + (npq)^{-\frac{1}{2}}(1-2p)(1/6)(1-x^2)\phi(x) + C_2n^{-3/4} \qquad (A.1)$$

uniformly in x of the form $(m+\frac{1}{2})/(n\hat{p}_1\hat{q}_1)^{\frac{1}{2}}$, for integers m.

Step (ii). Let ψ have two continuous derivatives on $(-\infty,\infty)$, fix $a > 0$, and define ψ_1 by linearly interpolating between values taken by ψ at points $(m+\frac{1}{2})/a$, for integers m. Then $\|\psi-\psi_1\| \le a^{-2}\|\psi''\|$, where $\|\cdot\|$ is the sup norm. Therefore on E,

$$\hat{G}(x) \le \Phi(x) + (npq)^{-\frac{1}{2}}(1-2p)(1/6)(1-x^2)\phi(x) + C_3 n^{-3/4} \qquad (A.2)$$

for *all* x, using (A.1).

Step (iii). Let $\chi(y) \equiv (1-y^2)\phi(y)$ and $C_4 \equiv C_3 + (pq)^{-1}(\|\phi'\| + \|\chi'\|)$. Given $\alpha \in I \equiv (2C_4 n^{-3/4},\ 1-2C_4 n^{-3/4})$, let $\beta \equiv \alpha - C_4 n^{-3/4}$, $u \equiv z(\beta)$ and $v \equiv u-(npq)^{-\frac{1}{2}}(1-2p)(1/6)(1-u^2)$. For $n \ge n_0$ say,

$$\Phi(v) + (npq)^{-\frac{1}{2}}(1-2p)(1/6)(1-v^2)\phi(v) + C_3 n^{-3/4} \le \alpha$$

whenever $\alpha \in I$. Therefore by (A.2), $\hat{G}(v) \le \alpha$, implying that $t_\alpha \ge v$ and

$$\pi(\alpha) \equiv \mathrm{pr}\{n^{\frac{1}{2}}(\hat{p}_1-p)/(pq)^{\frac{1}{2}} \le t_\alpha\}$$

$$\ge \mathrm{pr}\{n^{\frac{1}{2}}(\hat{p}_1-p)/(pq)^{\frac{1}{2}} \le v\} - P(\tilde{E}).$$

Edgeworth-expand $\mathrm{pr}\{n^{\frac{1}{2}}(\hat{p}_1-p)/(pq)^{\frac{1}{2}} \le v\}$ using Petrov's (1974, p. 171) Theorem 6, and then Taylor-expand functions of v about u, to obtain:

$$\pi(\alpha) \ge \alpha + \tfrac{1}{2}(npq)^{-\frac{1}{2}}\, s_{n1}(\alpha)\phi\{z(\alpha)\} - C_5 n^{-3/4}$$

uniformly in $\alpha \in I$, where s_{nj} takes values only in $[-1,1]$. Likewise, $\pi(\alpha) \le \alpha + \tfrac{1}{2}(npq)^{-\frac{1}{2}} s_{n2}(\alpha)\phi\{z(\alpha)\} + C_5 n^{-3/4}$, whence $\pi(\alpha) = \alpha + \tfrac{1}{2}(npq)^{-\frac{1}{2}} s_{n3}(\alpha)\phi(z) + o(n^{-\frac{1}{2}})$ uniformly in $\alpha \in I$. Monotonicity of π allows this identity to be extended to all $\alpha \in (0,1)$.

REFERENCES

BERAN, R. (1984). Jackknife approximations to bootstrap estimates. *Ann. Statist.* 12, 101-118.

BHATTACHARYA, R.N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. New York: Wiley.

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*. Philadelphia: Society for Industrial and Applied Mathematics.

HALL, P. (1982). Improving the normal approximation when constructing one-sided confidence intervals for binomial or Poisson parameters. *Biometrika* 69, 647-652.

PETROV, V.V. (1974). *Sums of Independent Random Variables*. Berlin: Springer.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* 9, 1187-1195.