Richardson Extrapolation and the Bootstrap

By

P.J. Bickel
Department of Statistics
University of California
Berkeley

and

J.A. Yahav
Department of Statistics
The Hebrew University
Jerusalem

Technical Report No. 71
July 1986
(revised September 1987)

Research supported by
Office of Naval Research contract N00014-80-C0163.

Department of Statistics
University of California
Berkeley, California

Richardson Extrapolation and the Bootstrap

By

P.J. Bickel
Department of Statistics
University of California
Berkeley, California 94720

and

J.A. Yahav
Department of Statistics
The Hebrew University
Jerusalem

AUTHOR'S FOOTNOTE

Peter J. Bickel is Professor of Statistics, University of California, Berkeley 97420. Joseph A. Yahav is Professor of Statistics, Hebrew University, Jerusalem, Israel. This work was partially supported by ONR contract N00014-80-C0163.

We are indebted to Persi Diaconis for referring us to Kuipers and Niederreiter (1978) enabling us to obtain a considerable simplification of our original proof of the theorem in the appendix. We also thank Adele Cutler for the programming of the simulations and other calculations of section 3.

ABSTRACT

Simulation methods, in particular Efron's (1979) bootstrap, are being applied more and more widely in statistical inference. Given data, $(X_1, \cdots, X_n)$, distributed according to P belonging to a hypothesized model **P** the basic goal is to estimate the distribution $\mathbf{L_P}$ of a function $T_n (X_1, \cdots, X_n, P)$. The bootstrap presupposes the existence of an estimate $\hat{P} (X_1, \cdots, X_n)$ and consists of estimating $\mathbf{L_P}$ by the distribution $\mathbf{L}_n^*$ of $T_n(X_1^*, \cdots, X_n^*, \hat{P})$ where $(X_1^*, \cdots, X_n^*)$ is distributed according to $\hat{P}$. The method is particularly of interest when $\mathbf{L}_n^*$, though known in principle, is realistically only computable by simulation.

Such computation can be expensive if n is large and T_n is very complex - see for instance the multivariate goodness of fit tests of Beran and Millar (1985). Even when application of the bootstrap to a single data set is not excessively expensive, Monte Carlo studies of the bootstrap are another matter.

We propose a method based on the classical ideas of Richardson extrapolation for reducing the computational cost inherent in bootstrap simulations and Monte Carlo studies of the bootstrap by doing the simulations for statistics based on two smaller sample sizes.

We study theoretically which ratio of the two small samples sizes is apt to give us best results. We show how our method works for approximating the χ^2, t and smoothed binomial distributions and for setting bootstrap percentile confidence intervals for the variance of a normal distribution with mean 0.

KEY WORDS: cost of computation, Edgeworth, approximation.

Richardson Extrapolation and the Bootstrap

P.J. BICKEL and J.A. YAHAV*

1. INTRODUCTION

Let L_n^*, as in the abstract, be the bootstrap distribution of a statistic $T_n(X_1, \ldots, X_n, P)$. With knowledge of particular features of $\mathbf{L}_n^*$ various devices such as importance sampling can be used to reduce the number r of Monte Carlo replications needed to compute (or rather estimate) L_n^* closely. The total cost of computation for a simulation is proportional to $c(n)r$ where $c(n)$, the cost of computing T_n, usually rises at least linearly with n and often faster. In this note we explore a way of reducing $c(n)$ rather than r. To fix ideas suppose T_n is univariate and let F_n^* be the distribution function of $\mathbf{L}_n^*$. For most statistics T_n of interest, it is either known or plausible to conjecture that F_n^* tends to a limit A_0 in probability

$$F_n^*(x) = A_0(x) + o_p(1) \tag{1.1}$$

for all x and often uniformly in x as well. Examples, see, for instance, Bickel and Freedman (1981), are the usual pivots for parameters $\theta(F)$ when $X_1, \cdots, X_n$ are i.i.d. F and $\hat{P} \longleftrightarrow \hat{F}$ is the empirical distribution. Thus if $T_n = \sqrt{n}(\theta(\hat{F}) - \theta(F))$ then $A_0 = N(0, \sigma^2(F))$ under mild conditions, and if $T_n = \sqrt{n}\dfrac{(\theta(\hat{F}) - \theta(F))}{\sigma(\hat{F})}$ then

$A_0 = \mathrm{N}(0,1)$. A_0 can also be known to exist but not be readily computable. For example let $T_n = \sqrt{n}\, \sup_x |\hat{F}(x) - F(x)|$ with F possibly discrete, a situation discussed in Bickel and Freedman (1981). Even more, an asymptotic expansion in powers of $n^{-1/2}$ is known to be true in some cases and reasonable to conjecture in many others. That is,

$$F_n^*(x) = A_0(x) + \sum_{j=1}^{k} n^{-j/2} A_j(x) + O_P\left(n^{\frac{-(k+1)}{2}}\right). \tag{1.2}$$

The most important special cases arise when A_0 is normal and the expansion (1.2) is of Edgeworth type. Examples of such expansions appear in the context of the bootstrap in Singh (1981), Bickel and Freedman (1981), Abramowitz and Singh (1985) etc. Expansions for the distributions F_n of statistics $T_n(X_1, \cdots, X_n)$ under fixed F have been extensively studied - see for example Bhattacharya and Ranga Rao (1976).

In this context, our proposal is to calculate $F_{n_1}, \cdots, F_{n_{k+1}}$ where,

$$n_1 + \cdots + n_{k+1} = b \ll n. \tag{1.3}$$

We use the F_{n_j} to approximate F_n. This procedure is classically used in numerical analysis, where it is called Richardson extrapolation, as a way of approximating F_∞. Our application of these ideas differs in that,

i) We are interested in F_n, not F_∞

ii) F_∞ is sometimes known, as in the Edgeworth case, and can be used to improve the approximation

iii) We are interested in the design problem of selecting the n_j subject to the "budget" constraint (1.3).

The use of our method in the bootstrap context just involves putting * on the F_{n_j}, F_n. We develop the method in detail in the next section and give explicit solutions to three formulations of the design problem for $k=1$. Finally in section 3, we test our method on approximations of known F_n as well as some bootstrap examples. The results are very encouraging.

2. EXTRAPOLATION

Throughout this section (I-K) will refer to Isaacson and Keller (1966). Write $t = n^{-1/2}$, $0 < t \leq 1$. We are given a sequence of distribution functions $F_n \stackrel{\Delta}{=} G_t$ and write,

$$G_t = P_t + \Delta_t. \tag{2.1}$$

$$P_t = A_0 + \sum_{j=1}^{k} t^j A_j.$$

The argument in the functions G_t, A_j plays no role in our discussion and is omitted. We calculate $G_{t_0}, \cdots, G_{t_k}$, $t < t_0 < \cdots < t_k$. If $\Delta_t = 0$ for $t, t_0, \cdots, t_k$ we obtain G_t perfectly from the G_{t_j} by using the Lagrange interpolating polynomial, (I-K p.188)

$$\hat{G}_t = \sum_{j=0}^{k} G_{t_j} \phi_{k,j}(t) \tag{2.2}$$

$$\phi_{k,j}(t) = \Pi_{i \neq j}[(t-t_i)/(t_j-t_i)].$$

In particular for the only case we study in detail, $k = 1$,

$$\hat{G}_t = (t_1-t_0)^{-1}[(t_1-t)G_{t_0}+(t-t_0)G_{t_1}]. \tag{2.3}$$

We consider three classes for Δ depending on a parameter M

$$\mathbf{D}_1 = \{\Delta : \frac{d^{k+1}\Delta_t}{dt^{k+1}} \textit{ exists} \text{ and } \sup_t \left|\frac{d^{k+1}\Delta_t}{dt^{k+1}}\right| \le M\}.$$

Since Δ is only defined at the points $n^{-1/2}$, $n = 1,2,...$ we interpret $\Delta \in \mathbf{D}_1$ as applying to some smooth function agreeing with Δ at all points $n^{-1/2}$. Our other two classes make no smoothness assumptions on Δ.

$$\mathbf{D}_2 = \{\Delta : \sup_t t^{-(k+1)}|\Delta_t| \le M\}$$

$$\mathbf{D}_3 = \{\Delta : 0 \le t^{-(k+1)}\Delta_t \le M \text{ for all } t>0 \textit{ or } -M \le t^{-(k+1)}\Delta_t \le 0 \text{ for all } t>0\}.$$

For fixed $t, t_0, \cdots, t_k$ we define the error of approximation by,

$$E_i(t,t_0,\cdots,t_k) = \sup\{|\hat{G}_t - G_t| : \Delta \in \mathbf{D}_i\},\ 1 \le i \le 3.$$

We want to minimize E_i subject to a fixed budget b

$$\sum_{j=0}^{k} t_j^{-2} = b. \tag{2.4}$$

If t_j satisfy (2.4) and $b \to \infty$ then $t_0 \to 0$.

We claim that,

$$E_1 \sim \frac{M}{(k+1)!}\prod_{j=0}^{k}(t_j-t) \tag{2.5}$$

$$E_2 \sim M\left\{\sum_{j=0}^{k}|\phi_{k,j}(t)|\,t_j^{k+1}+t^{k+1}\right\} \tag{2.6}$$

$$E_3 \sim M\left\{[\sum_{j=0}^{k}[\phi_{k,j}(t)]_+ t_j^{k+1}] \vee [\sum_{j=0}^{k}[\phi_{k,j}(t)]_- t_j^{k+1}]+t^{k+1}\right\} \tag{2.7}$$

where $a_+ = a \vee 0$, $a_- = -(a \wedge 0)$. To check (2.5) apply theorem 1 p.190 (I-K) according to which,

$$G_t - \hat{G}_t = [(k+1)!]^{-1} \prod_{i=0}^{k} (t - t_i) \frac{d^{k+1} G_t}{dt^{k+1}} (\xi) \tag{2.8}$$

where $t < \xi < t_k$. Note that $\frac{d^{k+1}}{dt^{k+1}} P_t = 0$. To check (2.6), (2.7) note that, interpolation is linear so that

$$\hat{G}_t = \hat{P}_t + \hat{\Delta}_t .$$

Since $P_t = \hat{P}_t$, we have

$$G_t - \hat{G}_t = \Delta_t - \hat{\Delta}_t$$

and (2.6), (2.7) follow from (2.2). From (2.5), E_1 is minimized subject to (2.3) as $b \to 0$ by

$$t_0 = \cdots = t_k = \sqrt{\frac{(k+1)}{b}} . \tag{2.9}$$

The allocation (2.9) is, of course, not feasible since the t_j must be distinct. However the clear moral is that if the error term Δ is sufficiently smooth the n_j should be chosen as nearly equal to each other as possible.

This is analogous to the prescription appearing in the leave one out jackknife. The argument for doing so in that situation rather than leaving more out has more to do with the polynomially increasing number of subsets that need be considered. This conclusion is clearly valid not just under (2.4) but under any reasonable symmetric side condition on $t_0, \cdots, t_k$. If we suppose $t = o(t_0)$, i.e. the budget is much smaller than n we can simplify (2.6) to,

$$E_2 \sim M(\Pi_{j=0}^{k} t_j) \sum_{j=0}^{k} t_j^k [\Pi_{i<j}(t_j - t_i)\Pi_{i>j}(t_i - t_j)]^{-1} \tag{2.10}$$

and (2.7) to

$$E_3 \sim M(\Pi_{j=0}^{k} t_j) \sum_{j=0}^{k} t_j^k \{[\Pi_{i \neq j}(t_j - t_i)]_+ \wedge [\Pi_{i \neq j}(t_j - t_i)]_-\}^-. \tag{2.11}$$

Evidently (2.10) is minimized asymptotically by

$$t_j^{-2} = \lambda_j^2 b$$

where $\lambda_j > 0$,

$$\sum_{j=0}^{k} \lambda_j^2 = 1 \tag{2.12}$$

and $\lambda_0, \cdots, \lambda_k$ minimize,

$$(\Pi_{j=0}^{k} \lambda_j)^{-1} \sum_{j=0}^{k} [\lambda_j \Pi_{i<j}(\lambda_i - \lambda_j)\Pi_{i>j}(\lambda_j - \lambda_i)]^{-1} \tag{2.13}$$

subject to (2.12). This minimization can be carried out, in principle, for any k. The explicit solutions for the cases we are primarily concerned with, E_2 and E_3 for $k = 1$ are as follows, if we ignore the restriction that the $\lambda_j^2 b$ are integers. For E_2

$$\lambda_0^2 = 1 - \lambda_1^2 = .89 \tag{2.14}$$

or more specifically

$$\lambda_0 = \cos\left[\frac{1}{2}(\sin^{-1}\frac{1}{\omega_0})\right]$$

where $\omega_0 = \dfrac{1+\sqrt{5}}{2} = 1.6180$ is the unique positive root of,

$$\omega^3 - 2\omega - 1 = 0.$$

To see this note that for $k = 1$, (2.13) is just $(\lambda_0\lambda_1)^{-1}(\lambda_0-\lambda_1)^{-1}(\lambda_0+\lambda_1)$. Substitute $\lambda_0 = \cos\theta$ to get as objective,

$$2(1+\sin 2\theta)(\cos 2\theta \sin 2\theta)^{-1}$$

and then substitute $\sin 2\theta = \sqrt{1-\nu^2} = \frac{1}{\omega}$. Similarly for $k = 1$,

$$E_3 \sim M\frac{t_0 t_1^2}{(t_1-t_0)}$$

and a similar minimization leads to,

$$\lambda_0^2 = \frac{1}{2}\left[1+\frac{1}{\sqrt{2}}\right] = .85. \qquad (2.15)$$

In all these cases $E_j = O(b^{-(k+1)/2})$.

Here are two examples of $\{F_n\}$ for which we will check our approach and which belong to $\mathbf{D}_1$ and $\mathbf{D}_3$ respectively.

1. The Gamma Family:

Let F_n be the distribution of $(S_n - n)(2n)^{-1/2}$ where S_n has the χ_n^2 distribution. Evidently, we can define G_t for $t > 0$ by

$$G_t(x) = \Gamma^{-1}(v^{-1})\lambda^v \int_0^{x_t} e^{-\lambda s} s^{v-1} ds \tag{2.16}$$

where

$$x_t = xv^{1/2} + \frac{v}{\lambda}$$

$v = 2t^{-2}$, $\lambda = \frac{1}{2}$. It is not hard using standard Stirling type expansions for Γ and its derivatives to show by writing

$$G_t(x) = \frac{e^{-v} v^{v-1/2}}{\Gamma(v)} \int_{-\sqrt{v}}^{x} [e^{-uv^{-1/2}}(1 + uv^{-1/2})]^v (1 + uv^{-1/2})^{-1} du$$

that $A_0 = \Phi$, the standard normal distribution and that G_t has bounded derivatives of all orders in t. So $\Delta \in \mathbf{D}_1$ here for all k. Evidently, taking $\lambda = \frac{1}{2}$ plays no role and this observation applies to the standardized gamma family in general.

2. The binomial distribution with continuity correction

Let F_n be the distribution of $(S_n - n_p)/\sqrt{npq}$ convoluted with the uniform distribution on $(-\frac{1}{2\sqrt{npq}}, \frac{1}{2\sqrt{npq}})$ where S_n has a binomial (n,p) distribution, $q = 1-p$, $0 < p < 1$. It is well known that, see for example Feller (1971) p.540, F_n is of the form,

$$F_n(x) = \Phi(x) + n^{-1/2}A_1(x) + O(n^{-1}).$$

However, if we analyze the remainder term further, by Theorem 23.1 p.238 of Bhattacharya and Ranga Rao it is of the form,

$$F_n(x) - \Phi(x) - n^{-1/2}A_1(x) = n^{-1}\left[\int_{-\frac{1}{2}}^{\frac{1}{2}} uS_1(np + x\sigma\sqrt{n} - u)du\right] \cdot \qquad (2.15)$$

$$\cdot P(x,\sigma) + o(n^{-1})$$

where $\sigma = \sqrt{pq}$

$$S_1(t) = t - \frac{1}{2},\ 0 < t < 1$$

$$S_1(t+1) = S_1(t).$$

Check that

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} uS_1(v-u)du = -\frac{S_1^2}{2}(x + \frac{1}{2}). \qquad (2.16)$$

We claim that unless $x = 0$ and p is rational the sequence $\{S_1(np + \sqrt{n}\,\sigma x + \frac{1}{2})$ is uniformly distributed modulo 1, i.e. $\#\{h : S_1(np + \sqrt{n}\,\sigma x + \frac{1}{2}) \le t,\ n \le N\}/N \to t + \frac{1}{2}$ as $N \to \infty$, $-\frac{1}{2} < t < \frac{1}{2}$. A proof is sketched in the appendix.

Hence as $n \to \infty$ the coefficient of n^{-1} in (2.15) ranges over an interval $[0, \frac{1}{8}]$ or $[-\frac{1}{8}, 0]$ and comes arbitrarily close to all values in the interval. So $\{F_n\}$ belongs to D_3 for $k = 1$.

Notes: 1). In a wide class of examples including the two we have discussed A_0 is known. Then if (2.1) holds for $k = r+1$ we can improve our estimate using only k sample sizes and still have an error which is $O(b^{-(r+1)/2})$. We define $Q_t = \frac{G_t - A_0}{t}$ and use the estimate

$$G_t^* = A_0 + t\hat{Q}_t$$

where $\hat{Q}_t$ is defined by (2.2) with $k = r$. In particular for $r = 1$, the allocations (2.9), (2.14), (2.15) give errors which are $O(b^{-3/2})$. We study this approximation also by simulation in the next section.

2). In some cases, for example F_n the t distribution with n degrees of freedom, the series for F_n is in powers of n^{-1}. It is easy in this case also to obtain the optimal choice of $\frac{t_0}{t_1}$ for $\mathbf{D}_2$, $\mathbf{D}_3$ that is, for (2.4) replaced by,

$$t_0^{-1} + t_1^{-1} = b.$$

We find for $\mathbf{D}_2$

$$n_j = \rho_j b, \quad \rho_1 = 1 - \rho_0,$$
$$\rho_0 = .5(1+\sqrt{3}) = .79 \tag{2.17}$$

and for $\mathbf{D}_3$

$$\rho_0 = .75. \tag{2.18}$$

If as would usually be the case in applications the A_j, t are unknown it would seem safer to use the approximation for $t = n^{-1/2}$.

3). An undesirable feature of our approach is that no a posteriori estimate of the error actually incurred is available. If t_1 is small and $\Delta \in \mathbf{D}_1$ we can get an estimate by increasing our budget. We add $\tilde{t}^{-2} \neq t_j^{-2}$, $j = 0,1$ units and calculate $G_{\tilde{t}}$. Now, by (2.8)

$$G_u - \hat{G}_u = \frac{1}{2}\frac{d^2\Delta}{dt^2}(\xi)(t_1 - t_0)^{-1}(u - t_0)(u - t_1) \tag{2.19}$$

where $t < \xi < t_1$, for any $t \le u \le t_1$. If t_1 is small we expect the coefficient $\dfrac{d^2\Delta}{dt^2}$ in (2.19) to be stable, so that we obtain

$$|G_t - \hat{G}_t| \propto |(t - t_0)(t - t_1)(s - t_0)^{-1}(s - t_1)^{-1}|\, |G_{\tilde{t}} - \hat{G}_{\tilde{t}}|. \tag{2.20}$$

If $\Delta \in \mathbf{D}_2$ or $\mathbf{D}_3$ no realistic estimate of the error presents itself. However, suppose as may be shown to be true in example 2 that, if $0 < \lambda_1 < \cdots < \lambda_k < 1$, $a_1, \cdots, a_k$ are real, $n \to \infty$, $s_j = [\lambda_j n]^{-1/2}$,

$$\#(\Delta_{s_j} \le s_j^2 a_j : 1 \le j \le k)/n \to \prod_{j=1}^{k} G(a_j). \tag{2.21}$$

That is $s_1^{-2}\Delta_{s_1}, \cdots, s_k^{-2}\Delta_{s_k}$ are asymptotically independently distributed with common distribution G. This is, of course, a poor approximation if λ_j, λ_{j+1} are too close and we are prevented from using (2.21) for the purposes of design. However if we increase our budget so as to permit calculation of G at $t_0, t_1, \cdots, t_l$ $k \ge 2$ and assume (2.1) it is natural to consider the estimate $\hat{G}_t^l$

$$\hat{G}_t^l = \hat{A}_0^l + t\hat{A}_1^l \tag{2.22}$$

where $\hat{A}_0^l$, $\hat{A}_1^l$ are the weighted fixed least squares estimates of A_0, A_1,

$$\hat{A}_1^l = \sum_{i=0}^{l}(t_i - \hat{t})G_{t_i}\sigma_i^{-2} / \sum_{i=0}^{l}(t_i - \hat{t})^2\sigma_i^{-2} \tag{2.23}$$

$$\hat{A}_0^l = \sum_{i=0}^{l} G_{t_i}\frac{\sigma_i^{-2}}{W} - \hat{A}_1^l\hat{t} \tag{2.24}$$

where

$$\sigma_i = t_i^{\,2}$$

$$W = \sum_{i=0}^{l}\sigma_i^{-2}$$

$$\hat{t} = \sum_{i=0}^{l} t_i\sigma_i^{-2}/W.$$

The error, $G_t - \hat{G}_t^l$ can be estimated by

$$\{W^{-1} + t^2[\Sigma(t_i - \hat{t})^2\sigma_i^{-2}]^{-1}\sum_{i=0}^{l}(G_{t_i} - \hat{A}_0^l - \hat{A}_1^l t_i)^2\sigma_i^{-2}\}^{1/2}. \tag{2.25}$$

The range of validity of the approximations (2.20) and (2.25) needs to be investigated by simulation.

3. COMPUTATION AND SIMULATION

In this section we study the actual performance of our approximations in the examples of section 2. We also study the performance of the approximation for student's t distribution where the expansion is in powers of $\frac{1}{n}$. Finally in table 6 we provide the results of a bootstrap simulation where we illustrate the operating characteristics of confidence bounds based on a Richardson extrapolation approximation compared with those based on a full bootstrap.

a) <u>χ_n^2 approximation</u>

We computed the Richardson extrapolation for $\frac{\chi_n^2(\alpha)-n}{\sqrt{2n}}$, α = 10%, 90%, 95%, 99%, where $\chi_n^2(\alpha)$ is the α^{th} percentile of the χ_n^2 distribution, and compared it the Fisher square root approximation applied to the quantiles.

$$\frac{\chi_n^2(\alpha)-n}{\sqrt{2n}} \propto Z(\alpha)+\frac{Z^2(\alpha)}{2\sqrt{2n}}$$

where $Z(\alpha)$ is the standard normal α-percentile. We used n = 50, 100 b = 15, 20, 30 and $1-\lambda = \frac{n_0}{b} = .1,.2,.25,.40$ where $n_0 < n_1$, $n_0+n_1 = b$. We note,

(i) The approximation improves as b and n increase.

(ii) The allocation $\lambda = .6$ is best, as expected.

(iii) For $n_0+n_1 = 15,20$ and all λ the Richardson extrapolation is essentially as good as Fisher's approximation for the .9, .1 percentiles and still gives the same two significant figures as Fisher's does for the .95, .99 percentiles.

(iv) For $n_0+n_1 = 30$ it is better in all cases save one where the results are virtually equivalent. The $\lambda = .6$ allocation seems to give nearly 3 significant figures.

Table 1. Richardson extrapolation 2.3 for χ_n^2

	percentiles			
n = 50	10	90	95	99
true values	-1.2311	1.3167	1.7505	2.6154
Fisher app.	-1.1995	1.3637	1.7802	2.5969
	(0.0317)	(0.0470)	(0.0297)	(-0.0185)
$n_0+n_1 = 15$				
n_0 n_1				
1, 14	-1.2557	1.3572	1.7995	2.6686
	(-0.0246)	(0.0404)	(0.0490)	(0.0532)
3, 12	-1.2528	1.3417	1.7796	2.6446
	(-0.0217)	(0.0250)	(0.0291)	(0.0292)
4, 11	-1.2511	1.3391	1.7764	2.6410
	(-0.0199)	(0.0224)	(0.0259)	(0.0256)
6, 9	-1.2493	1.3367	1.7735	2.6377
	(-0.0182)	(0.0200)	(0.0230)	(0.0223)
$n_0+n_1 = 20$				
n_0, n_1				
2, 18	-1.2481	1.3374	1.7747	2.6400
	(-0.0169)	(0.0207)	(0.0242)	(0.0246)
4, 16	-1.2448	1.3319	1.7680	2.6324
	(-0.0137)	(0.0152)	(0.0175)	(0.0170)
5, 15	-1.2439	1.3307	1.7665	2.6307
	(-0.0127)	(0.0139)	(0.0160)	(0.0153)
8, 12	-1.2424	1.3289	1.7644	2.6285
	(-0.0113)	(0.0122)	(0.0139)	(0.0131)

n = 100				
true values	-1.2475	1.3080	1.7212	2.5319
Fisher app.	-1.2235	1.3397	1.7406	2.5176
	(0.0239)	(0.0317)	(0.0193)	(-0.0143)
$n_0+n_1 = 20$				
n_0 n_1				
2, 18	-1.2740	1.3399	1.7585	2.5694
	(-0.0265)	(0.0319)	(0.0373)	(0.0374)
4, 16	-1.2687	1.3314	1.7481	2.5576
	(-0.0212)	(0.0234)	(0.0268)	(0.0257)
5, 15	-1.2671	1.3294	1.7457	2.5550
	(-0.0197)	(0.0214)	(0.0245)	(0.0231)
8, 12	-1.2649	1.3267	1.7424	2.5515
	(-0.0174)	(0.0187)	(0.0212)	(0.0196)
$n_0+n_1 = 30$				
n_0 n_1				
3, 27	-1.2628	1.3252	1.7410	2.5508
	(-0.0154)	(0.0172)	(0.0197)	(0.0189)
6, 24	-1.2592	1.3205	1.7354	2.5448
	(-0.0117)	(0.0125)	(0.0142)	(0.0129)
7, 23	-1.2585	1.3198	1.7345	2.5439
	(-0.0110)	(0.0117)	(0.0133)	(0.0120)
12, 18	-1.2569	1.3180	1.7324	2.5418
	(-0.0095)	(0.0100)	(0.0112)	(0.0099)

In table 2 we exhibit the Richardson extrapolation results for the χ_n^2 distribution using the knowledge of the limit as $n \to \infty$, as in note 1. That is, we use the expansion

$$\frac{\chi_n^2(\alpha)-n}{\sqrt{2n}} = Z(\alpha)+A_1\frac{1}{\sqrt{n}}+A_2\frac{1}{n}+o_P\left[\frac{1}{n}\right]$$

$$or\ \sqrt{n}\left[\frac{\chi_n^2(\alpha)-n}{\sqrt{2n}}-Z(\alpha)\right] = A_1+A_2\frac{1}{\sqrt{n}}+o_P\left[\frac{1}{\sqrt{n}}\right]$$

where $Z(\alpha)$ is the α percentile of the standard normal. A_1, A_2 are estimated using $\chi_{n_0}^2$ and $\chi_{n_1}^2$. The results are extremely good for both $n = 50$ and $n = 100$ (which we omit). The extrapolation, even for $n_0+n_1 = 15$ and $\lambda = .9$ gives three significant figures for all percentiles. For $n_0+n_1 = 30$ it often gives five significant figures.

Table 2: Richardson extrapolation (Note 1) for χ_{50}^2.

	percentiles			
$n = 50$	10	90	95	99
true values	-1.2311	1.3167	1.7505	2.6154
Fisher app.	-1.1995	1.3637	1.7802	2.5969
	(0.0317)	(0.0470)	(0.0297)	(-0.0185)
$n_0+n_1 = 15$				
n_0 n_1				
1, 14	-1.2289	1.3165	1.7510	2.6178
	(0.0022)	(-0.0002)	(0.0005)	(0.0024)
3, 12	-1.2306	1.3168	1.7510	2.6172
	(0.0006)	(0.0001)	(0.0005)	(0.0018)
4, 11	-1.2307	1.3168	1.7509	2.6171
	(0.0004)	(0.0001)	(0.0005)	(0.0017)
6, 9	-1.2308	1.3168	1.7509	2.6169
	(0.0003)	(0.0001)	(0.0004)	(0.0016)
$n_0+n_1 = 20$				
n_0 n_1				

2, 18	-1.2305	1.3167	1.7509	2.6168
	(0.0006)	(0.0000)	(0.0004)	(0.0015)
4, 16	-1.2309	1.3168	1.7508	2.6166
	(0.0002)	(0.0001)	(0.0003)	(0.0012)
5, 15	-1.2309	1.3168	1.7508	2.6165
	(0.0002)	(0.0001)	(0.0003)	(0.0011)
8, 12	-1.2310	1.3168	1.7508	2.6164
	(0.0001)	(0.0001)	(0.0003)	(0.0010)
$n_0+n_1 = 30$				
$n_0\ n_1$				
3, 27	-1.2310	1.3167	1.7507	2.6161
	(0.0002)	(0.0000)	(0.0002)	(0.0007)
6, 24	-1.2311	1.3167	1.7506	2.6159
	(0.0001)	(0.0000)	(0.0002)	(0.0005)
7, 23	-1.2311	1.3167	1.7506	2.6159
	(0.0001)	(0.0000)	(0.0001)	(0.0005)
12, 18	-1.2311	1.3167	1.7506	2.6159
	(0.0000)	(0.0000)	(0.0001)	(0.0005)

Student's t distribution has an expansion in powers of $\frac{1}{n}$. The Richardson extrapolation (2.3) with $\frac{1}{\sqrt{n}}$ gave no improvement over the ordinary normal approximation as expected. In table 3 we present the Richardson extrapolation to the t distribution as in note 2. and compare these results to the normal approximation. We looked at the same values of n, b, λ, α for approximation to $t_n(\alpha)$, the α^{th} percentile of the t distribution with n degrees of freedom. For $\lambda = .6$ $b = 30$ the approximation is valid to 3 significant figures for $n = 100$ in all but one case and improves on the normal approximation.

Table 3 : Richardson extrapolation for the t distribution.

	percentiles			
n = 50	10	90	95	99
true values	-1.2987	1.2987	1.6759	2.4033
normal app.	-1.2816	1.2816	1.6449	2.3263
	(0.0171)	(-0.0171)	(-0.0310)	(-0.0770)
$n_0+n_1 = 15$				
$n_0\ n_1$				
3, 12	-1.2849	1.2849	1.6376	2.2099
	(0.0138)	(-0.0138)	(-0.0383)	(-0.1934)
4, 11	-1.2878	1.2878	1.6462	2.2595
	(0.0110)	(-0.0110)	(-0.0298)	(-0.1438)
6, 9	-1.2900	1.2900	1.6526	2.2947
	(0.0087)	(-0.0087)	(-0.0233)	(-0.1086)
$n_0+n_1 = 20$				
$n_0\ n_1$				
4, 16	-1.2922	1.2922	1.6584	2.3198
	(0.0065)	(-0.0065)	(-0.0175)	(-0.0835)
5, 15	-1.2933	1.2933	1.6614	2.3357
	(0.0055)	(-0.0055)	(-0.0146)	(-0.0676)
8, 12	-1.2945	1.2945	1.6649	2.3535
	(0.0042)	(-0.0042)	(-0.0111)	(-0.0498)
n = 100				
true values	-1.2901	1.2901	1.6602	2.3642
normal app.	-1.2816	1.2816	1.6449	2.3263
	(0.0085)	(-0.0085)	(-0.0153)	(-0.0379)

$n_0 + n_1 = 20$

$n_0\ n_1$				
4, 16	-1.2818	1.2818	1.6378	2.2577
	(0.0083)	(-0.0083)	(-0.0224)	(-0.1065)
5, 15	-1.2831	1.2831	1.6417	2.2785
	(0.0070)	(-0.0070)	(-0.0185)	(-0.0857)
8, 12	-1.2848	1.2848	1.6463	2.3018
	(0.0053)	(-0.0053)	(-0.0139)	(-0.0624)

$n_0 + n_1 = 30$

$n_0\ n_1$				
6, 24	-1.2869	1.2869	1.6520	2.3274
	(0.0031)	(-0.0031)	(-0.0082)	(-0.0368)
7, 23	-1.2873	1.2873	1.6530	2.3321
	(0.0028)	(-0.0028)	(-0.0072)	(-0.0321)
12, 18	-1.2880	1.2880	1.6550	2.3414
	(0.0020)	(-0.0020)	(-0.0053)	(-0.0228)

Tables 4,5 give the Richardson extrapolation for the continuity corrected binomial distribution. That is we define

$$B_n(s) = \sum_{k=0}^{[s]} \binom{n}{k} p^k (1-p)^{n-k} + (s-[s]) \binom{n}{[s]+1} p^{[s]+1} (1-p)^{n-1-[s]}$$

and $Q_n(u) = B_n(np + u\sqrt{np(1-p)})$. We approximated the percentiles $Q_n^{-1}(\alpha)$ for n, b, λ as before with $p = .2, .4$. Note that,

(i) The $\lambda = .75$ allocation seems to work best but differs little from $\lambda = .8, .9$. On the other hand $\lambda = .6$ is poorer. This is in agreement with our theory for class $\mathbf{D}_3$. For $p = .2$, $n = 50; 100$ $b = 15; 20$ the $\lambda = .75$ allocation does as well as the normal. For $b = 20; 30$ it's better, typically giving one more significant figure. For $p = .4$, it is in general poorer though far from terrible.

This is understandable since for $p = .5$, $A_1 = 0$ and the extrapolation is adding noise to the normal approximation.

Table 4 : Richardson extrapolation for the Binomial distribution with $p = .2$

	percentiles			
$n = 50$	10	90	95	99
true values	-1.2591	1.3125	1.7177	2.4900
normal app.	-1.2816	1.2816	1.6449	2.3263
	(-0.0225)	(-0.0309)	(-0.0728)	(-0.1637)
$n_0+n_1 = 15$				
n_0 n_1				
1, 14	-1.2689	1.2591	1.6071	2.4969
	(-0.0097)	(-0.0533)	(-0.1106)	(0.0068)
3, 12	-1.2392	1.3702	1.6743	2.6561
	(0.0199)	(0.0577)	(-0.0434)	(0.1661)
4, 11	-1.1692	1.2861	1.5821	2.3349
	(0.0900)	(-0.0264)	(-0.1356)	(-0.1551)
6, 9	-1.1679	1.4169	1.6882	2.5377
	(0.0913)	(0.1044)	(-0.0295)	(0.0477)
$n_0+n_1 = 20$				
n_0 n_1				
2, 18	-1.2182	1.3060	1.6984	2.4728
	(0.0409)	(-0.0065)	(-0.0193)	(-0.0172)
4, 16	-1.2751	1.2595	1.7357	2.4304
	(-0.0160)	(-0.0530)	(0.0180)	(-0.0597)
5, 15	-1.2724	1.3224	1.7362	2.5814
	(-0.0133)	(0.0099)	(0.0185)	(0.0914)

8, 12	-1.1082	1.2538	1.8704	2.8400
	(0.1509)	(-0.0587)	(0.1527)	(0.3500)
n = 100				
true values	-1.2733	1.3036	1.6922	2.4351
normal app.	-1.2816	1.2816	1.6449	2.3263
	(-0.0083)	(-0.0220)	(-0.0473)	(-0.1088)
$n_0+n_1 = 20$				
n_0 n_1				
2, 18	-1.2111	1.2835	1.6845	2.4144
	(0.0622)	(-0.0202)	(-0.0078)	(-0.0207)
4, 16	-1.2699	1.2280	1.7121	2.3686
	(0.0033)	(-0.0757)	(0.0199)	(-0.0665)
5, 15	-1.2638	1.3054	1.7208	2.5622
	(0.0094)	(0.0018)	(0.0286)	(0.1271)
8, 12	-1.0651	1.2220	1.8840	2.8864
	(0.2082)	(-0.0816)	(0.1918)	(0.4513)
$n_0+n_1 = 30$				
n_0 n_1				
3, 27	-1.2644	1.3313	1.6692	2.4688
	(0.0089)	(0.0277)	(-0.0231)	(0.0337)
6, 24	-1.2233	1.3226	1.6628	2.4172
	(0.0500)	(0.0190)	(-0.0294)	(-0.0179)
7, 23	-1.2386	1.2849	1.7134	2.3774
	(0.0347)	(-0.0187)	(0.0211)	(-0.0577)
12, 18	-1.1641	1.3332	1.4957	2.4281
	(0.1092)	(0.0296)	(-0.1966)	(-0.0070)

Table 5 : Richardson extrapolation for the Binomial distribution with p = .4.

n = 50	percentiles 10	90	95	99
true values	-1.2776	1.2882	1.6694	2.3720
normal app.	-1.2816	1.2816	1.6449	2.3263
	(-0.0040)	(-0.0066)	(-0.0245)	(-0.0457)
$n_0+n_1 = 15$				
n_0 n_1				
1, 14	-1.2692	1.2656	1.6423	2.4690
	(0.0084)	(-0.0226)	(-0.0271)	(0.0971)
3, 12	-1.1991	1.3197	1.6443	2.3799
	(0.0785)	(0.0315)	(-0.0252)	(0.0079)
4, 11	-1.2562	1.1647	1.6847	2.2989
	(0.0214)	(-0.1235)	(0.0153)	(-0.0731)
6, 9	-1.2007	1.0239	1.8705	2.4680
	(0.0770)	(-0.2643)	(0.2010)	(0.0961)
$n_0+n_1 = 20$				
n_0 n_1				
2, 18	-1.2344	1.2776	1.6229	2.4184
	(0.0432)	(-0.0105)	(-0.0466)	(0.0464)
4, 16	-1.2823	1.2781	1.6526	2.3583
	(-0.0047)	(-0.0101)	(-0.0168)	(-0.0137)
5, 15	-1.2702	1.2816	1.6411	2.3911
	(0.0074)	(-0.0066)	(-0.0283)	(0.0191)
8, 12	-1.3054	1.2832	1.7722	2.5967

	(-0.0278)	(-0.0049)	(0.1027)	(0.2247)
$n = 100$				
true values	-1.2811	1.2892	1.6619	2.3475
normal app.	-1.2816	1.2816	1.6449	2.3263
	(-0.0005)	(-0.0076)	(-0.0170)	(-0.0212)
$n_0+n_1 = 20$				
n_0 n_1				
2, 18	-1.2167	1.2576	1.5951	2.4224
	(0.0644)	(-0.0316)	(-0.0668)	(0.0749)
4, 16	-1.2738	1.2589	1.6383	2.3349
	(0.0073)	(-0.0303)	(-0.0236)	(-0.0126)
5, 15	-1.2674	1.2767	1.6183	2.4029
	(0.0138)	(-0.0125)	(-0.0436)	(0.0554)
8, 12	-1.3107	1.2668	1.7943	2.6437
	(-0.0295)	(-0.0224)	(0.1324)	(0.2962)
$n_0+n_1 = 30$				
n_0 n_1				
3, 27	-1.2317	1.2922	1.6644	2.3581
	(0.0494)	(0.0030)	(0.0025)	(0.0106)
6, 24	-1.2379	1.2613	1.6580	2.3519
	(0.0432)	(-0.0279)	(-0.0039)	(0.0043)
7, 23	-1.2601	1.3154	1.6403	2.3348
	(0.0210)	(0.0262)	(-0.0216)	(-0.0128)
12, 18	-1.2439	1.2762	1.6676	2.3576
	(0.0373)	(-0.0130)	(0.0057)	(0.0101)

In table 6 we show the results for the bootstrap experiment. The "population" is $\sigma^2\chi_1^2$ and we are interested in a confidence bound for σ. We study the unadjusted bootstrap i.e. the percentiles of the bootstrap distribution of $\sqrt{\bar{X}_n}$ where $\bar{X}$ is the sample mean. For $n = 50, 100, 500$ we took 500 samples of size n from χ_1^2. For each sample we took 1,000 bootstrap samples and computed the .05 .1 and .95 percentiles of the bootstrap distribution of $\sqrt{\bar{X}_n}$ for sample size n_0, n_1, n. We study the behavior of the "90%" lower confidence bound and the "90%" confidence interval i.e. the .1 percentile and the interval between the .95 and .05 percentiles.

For each n we count the number of times the population parameter falls inside the confidence set, out of the 500 samples. We computed the average and S.D of the rescaled lower bound, i.e. of

$$\sqrt{n}\,(1 - G_n^{*-1}(.1))$$

and of the rescaled interval i.e.

$$I_n^*(.9) = \sqrt{n}\,(G_n^{*-1}(.95) - G_n^{*-1}(.05))$$

where $G_n^{*-1}(\alpha)$ is the α percentile of the bootstrap distribution of $\sqrt{\bar{X}}$. Table 6 shows clearly that Richardson extrapolation is a good approximation to the full bootstrap and is not very sensitive to the allocation of n_0, n_1. The last entry gives estimated computing times on Sun workstations at Berkeley. The linear saving in the sample size we expected is confirmed.

Table 6 : A Bootstrap experiment.

n	n_0	n_1	lower bound count	interval count	rescaled C.bound ave	rescaled C.bound sd	rescaled length ave	rescaled length sd	time (cpusecs)
50	full	Bootstrap	462	443	0.83732	.007231	2.23093	.018770	1603
50	2	18	455	439	0.84251	.007652	2.26337	.019407	680
50	4	16	468	449	0.83286	.007090	2.23915	.018084	680

100	full	Bootstrap	457	445	0.85825	.005957	2.25543	.015018	3171
100	2	18	459	438	0.85736	.006190	2.27469	.014191	688
100	4	16	472	446	0.85685	.006639	2.26594	.014139	686
500	full	Bootstrap	453	453	0.89302	.003029	2.31675	.070418	15754
500	5	45	454	448	0.88568	.004092	2.31589	.009186	1665
500	10	40	454	455	0.89668	.004200	2.33916	.086705	1666

4. APPENDIX: theory for example 2

We establish the claim asserted in example 2 in the form of a theorem.

Theorem: $[an + b\sqrt{n}\,]$ is uniformly distributed (u.d.) mod 1 unless $b = 0$ and a is rational.

Proof: We refer repeatedly to the text of Kuipers and Niederreiter (K-N). Suppose a is irrational. Note that

$$a(n+1) + b\sqrt{n+1} - an - b\sqrt{n} = a + b\,0(n^{-1/2}) \to a$$

as $n \to \infty$. By theorem 3.3 of (K-N), $an + b\sqrt{n}$ is u.d. mod 1.

If a is rational we apply

Lemma: Let b_n be a sequence such that $\{b_{sj+k}\}_{j \ge 1}$ is u.d. mod1 for $s \neq 0$, $0 \le k \le s$. Then if a is rational, $a = \frac{r}{s}$, $an + b_n$ is u.d. mod 1.

Proof: Check Weyl's criterion (K-N). Let $n = ms$. Then

$$\left|\frac{1}{n}\sum_{l=1}^{n} e^{2\pi i h(a_l + b_l)}\right| = \left|\frac{1}{ms}\sum_{j=0}^{m-1}\sum_{k=0}^{s-1} e^{2\pi i h\left[r\frac{k}{s} + b_{sj+k}\right]}\right| \tag{4.1}$$
$$\le \frac{1}{s}\sum_{k=0}^{s-1}\left|\frac{1}{m}\sum_{j=0}^{m-1} e^{2\pi i h b_{sj+k}}\right| \to 0$$

as $m \to \infty$ by Weyl's criterion applied to $\{b_{sj+k}\}_{j \ge 1}$. If $n = ms + b$, $0 < b < s$, the difference from (4.1) is at most $\frac{b}{ms} \to 0$. The lemma follows by Weyl's criterion.

Q.E.D.

Put $b_n = b\sqrt{n}$. If $b > 0$ $b_{s(j+1)+k} - b_{sj+k}$ is decreasing to 0 in j since $\sqrt{x}$ is concave. Moreover,

$$j(b_{s(j+1)+k} - b_{sj+k}) = \Omega(j^{1/2}) \to \infty.$$

By Fejer's theorem (K-N Theorem 2.5) $\{b_{sj+k}\}$ is u.d. mod 1 and the theorem follows.

5. REFERENCES

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

Beran, R. and Millar, P. W. (1986), "Confidence sets for a multivariate distribution," *Annals of Statistics,* 14, 431-443.

Bhattacharya, R. and Ranga Rao, R. (1976), "Normal approximation and asymptotic expansions," J. Wiley, New York.

Bickel, P. J. and Freedman, D. A. (1981), "Some asymptotic theory for the Bootstrap," *Annals of Statistics,* 9, 1196-1217.

Efron, B. (1979), "Bootstrap methods: Another look at jacknife," *Annals of Statistics,* 7, 1-26.

Feller, W. (1971), "An introduction to probability theory and its applications," J. Wiley, New York.

Isaacson, E. and Keller, H. B. (1966), "Analysis of numerical methods," J. Wiley, New York

Kuipers, L. and Niederreiter, H. (1979), "Uniform distribution or sequences," J. Wiley and sons.

Singh, K. (1981),"On asymptotic accuracy of Efron's bootstrap," *Annals of Statistics,* 9, 1187-1195.