Weighted Bootstrapping of Means

E. Haeusler,* D.M. Mason,† and M.A. Newton‡
University of Giessen, University of Delaware, and University of Washington

1 Introduction

Efron's bootstrap (Efron [6, 7]) and its many variants are becoming fundamental tools of statistical inference, as evidenced by the work of Hall [13], for example, and by the surveys of Hinkley [15], and Diciccio and Romano [5], and a number of interesting theoretical problems remain. One line of work still somewhat in its infancy concerns viewing the bootstrap as one procedure in a class of methods involving *random weighting* of observations. Here, we survey various random weighting methods and discuss what is known about their consistency and rates of consistency for estimating sampling distributions. We derive a scheme involving iid weights having a distribution determined by the data itself, and we compare this to Efron's bootstrap.

First to review Efron's original proposal, suppose that data $X_1, X_2, \ldots, X_n$ form a random sample from an unknown distribution F on the real line and that we wish to do inference about an unknown real-valued parameter $T(F)$ where T itself is known. (We concentrate on the mean $T(F) = \int x dF(x)$.) Typically, $T(F)$ is estimated from the data by $S_n(F_n)$ for some function S_n (often $S_n = T$) where F_n is the empirical distribution of the data; i.e. the distribution putting point mass $1/n$ at each realized X_i. It is reasonable to base inference about $T(F)$ on the distribution $H_n(\cdot, F)$ of $S_n(F_n) - T(F)$ or some scaled version of this difference, but $H_n(\cdot, F)$ is unknown because F is unknown. The usual approach in classical statistics to cope with this problem is to approximate $H_n(\cdot, F)$ by an asymptotic distribution function, based on an appropriate limit theorem from probability theory. An alternative is a *bootstrap* solution, which is always possible in principle; use $H_n(\cdot, F_n)$ as a surrogate for the unknown $H_n(\cdot, F)$. This is the conditional distribution of $S_n(\tilde{F}_n) - T(F_n)$ given the data where $\tilde{F}_n$ is the empirical distribution of a bootstrap sample; i.e. a sample of size n from F_n. Generally, $H_n(\cdot, F_n)$ cannot be computed analytically and therefore must be approximated by Monte Carlo methods.

It is well known that the empirical distribution of an individual bootstrap sample can be expressed in terms of the original data and an independent vector of multinomial weights:

*Supported partially by a Heisenberg stipend.
†Supported partially by NSF Grant DMS-9002731.
‡Supported by ONR contract N00014-91-J-1074 and by the Applied Physics Laboratory at the University of Washington.

$$(1.1) \qquad \tilde{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} M_{n,i}1[X_i \le t], \qquad -\infty < t < \infty$$

where

$$(1.2) \qquad (M_{n,1}, M_{n,2}, \ldots, M_{n,n}) \sim \text{Mult}_n\{n, (\tfrac{1}{n}, \tfrac{1}{n}, \ldots, \tfrac{1}{n})\}.$$

The vector in (1.2) can be simulated by throwing n objects into n equally likely cells. The bootstrap estimate $H_n(\cdot, F_n)$ hinges on the conditional distribution of $\tilde{F}_n$ and hence on the particular multinomial distribution above, among other things. This bootstrap estimate is a special case of estimates based on objects

$$(1.3) \qquad \tilde{G}_n(t) = \frac{1}{n}\sum_{i=1}^{n} Y_{n,i}1[X_i \le t], \qquad -\infty < t < \infty$$

where the vector $Y_n = (Y_{n,1}, \ldots, Y_{n,n})$ has some conditional distribution given the data. This conditional distribution is determined by the statistician and may be tailored to the particular problem at hand. For instance, by making the distribution of weights different from the multinomial above, perhaps allowing it to depend on the data or on the function T, we may get estimates of $H_n(\cdot, F)$ as good as or better than the standard bootstrap estimate. In this paper, such a random weighting scheme is referred to as a *weighted bootstrap*.

2 Weighted bootstrapping of means

In estimating the mean $T(F) = \int x dF(x)$ by the sample mean $T(F_n) = n^{-1}\Sigma_{i=1}^{n} X_i =: \bar{X}_n$, inference may be based upon $H_n(\cdot, F)$, the distribution function of

$$(2.1) \qquad \sqrt{n}(T(F_n) - T(F))/\sigma$$

where $\sigma^2 = \sigma^2(F)$ is the variance of X_1 assumed to be finite. When F is a normal distribution with unknown mean and variance, $H_n(\cdot, F)$ is known. In general, $H_n(\cdot, F)$ is unknown, in which case we can either use asymptotic theory or appeal to a bootstrap procedure. From (1.1), the corresponding bootstrap distribution $H_n(\cdot, F_n)$ is the conditional distribution (given data) of

$$(2.2) \qquad \sqrt{n}(T(\tilde{F}_n) - T(F_n))/\sigma_n = n^{-1/2}\sum_{i=1}^{n} M_{n,i}(X_i - \bar{X}_n)/\sigma_n$$

which is a centered and scaled *weighted bootstrapped mean*

$$(2.3) \qquad \tilde{X}_n = \frac{1}{n}\sum_{i=1}^{n} Y_{n,i}X_i \quad \text{with} \quad Y_{n,i} = M_{n,i}.$$

In (2.2), σ_n is an estimate of σ, for instance the sample variance

$$\sigma_n^2 = \sigma^2(F_n) = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

In this section, we discuss some weighted bootstrapped means different than the one in (2.3).

One way to get a weighted bootstrapped mean different than (2.3) is simply to change the bootstrap sample size. The weight vector $Y_n = (Y_{n,1}, \ldots, Y_{n,n})$ is still multinomial, but based on $m(n)$ throws into n equally likely cells instead of n throws (see after (1.2)).

Quenouille's [20] jackknife is another simple example of such a weighted bootstrapped mean. Here the weight vector $Y_n = (Y_{n,1}, \ldots, Y_{n,n})$ itself can only take on n possible values with equal probability - these values being the n-vectors having ones in $n-1$ spots and a zero in one spot.

The Bayesian bootstrap is yet another example. In an attempt to show the oddity (from a Bayesian perspective) of Efron's bootstrap, Rubin (1981) invented the Bayesian bootstrap - the direct Bayesian analogue of Efron's original proposal. The Bayesian bootstrap amounts to simulating a posterior Dirichlet process (Ferguson [9]) under a limiting improper prior on the distribution function F. In approximating the posterior distribution of the quantity in (2.1), the Bayesian bootstrap involves simulating the weighted bootstrapped mean

$$\text{(2.4)} \qquad \tilde{X}_n = \frac{1}{n}\sum_{i=1}^{n} Y_{n,i}X_i \quad \text{with} \quad Y_{n,i} = n(U_{n-1:i} - U_{n-1:i-1}),$$

where $U_{k:i}$ is the i^{th} order statistic of k independent uniform (0,1) random variables which are independent of the data. Here, $U_{n-1:0} = 0$ and $U_{n-1:n} = 1$. The weights are thus scaled one-spacings of $n-1$ uniform random variables.

For a given sample size n, the weight vector Y_n of the Bayesian bootstrap is equal in distribution to a vector of n independent exponential random variables normalized by their mean. Specifically,

$$\text{(2.5)} \qquad Y_{n,i} \stackrel{\mathcal{D}}{=} \frac{nV_i}{\sum_j V_i}$$

where V_i are iid with cdf $1 - e^{-v}$. In studying the asymptotic properties of the Bayesian bootstrap, Weng [26] suggests a procedure with weights like in (2.5) except that $V_i \sim_{iid}$ Gamma (1,4), which it turns out is equivalent to using four-spacings of $4n-1$ uniforms. The same suggestion was made earlier by Zheng and Tu [27] in another discussion of random weighting methods. Præstgaard [19] studies weights like those in (2.5) where no parametric assumptions are placed on the V_i. Rather, it is assumed $V_i \geq 0, EV_i = 1$ and the so called $L_{2,1}$ integrability condition

$$\text{(2.6)} \qquad \int_0^\infty \sqrt{P(\mid V_1 \mid > t)}\, dt < \infty$$

holds.

Independent and identically distributed weights are another possibility, which Dehling, Denker and Woyczynski [4] have investigated in the context of U-statistics with kernels having less than second moments. In Section 5, we propose a distribution for iid weights which gives an asymptotically accurate approximation to $H_n(\cdot, F)$ when $T(F)$ is the mean.

All the weight vectors described in this section are exchangeable; i.e. the joint distribution of their components (given the data) is invariant under permutation. A study of the class of bootstraps under exchangeable weighting schemes has been started by Mason and Newton [17].

Generalizing the Bayesian bootstrap in a different direction, Newton and Raftery [18] describe a weighted likelihood bootstrap for approximate simulation of Bayesian posteriors. The motivation is not to estimate the sampling distribution of a mean, however certain weighted bootstrapped averages are involved in the theory. An observation's contribution to a likelihood equation is randomly weighted as a way to simulate a posterior in a parametric or semiparametric model.

3 Consistency of weighted bootstrapped means

Let $X_1, X_2, \ldots \sim F$ with $\mu = EX$ and $\sigma^2 < \infty$ the variance of X_1. A weighted bootstrapped mean is called consistent if

$$(3.1)\quad \sup_{-\infty<t<\infty} \mid P\left(\sqrt{n}\frac{(\tilde{X}_n - \bar{X}_n\bar{Y}_n)}{\sigma_n\sigma_Y} \leq t \mid X_1, \ldots, X_n\right) - P\left(\sqrt{n}\frac{(\bar{X}_n-\mu)}{\sigma} \leq t\right) \mid \to 0,$$

almost surely, as $n \to \infty$, where $\sigma_Y^2 < \infty$ is the marginal variance of the weight $Y_{n,i}$ assumed to be independent of n and i, and $\bar{Y}_n$ is the average of these weights. Consistency of Efron's bootstrap was first established by Bickel and Freedman [1] assuming $0 < \sigma^2 < \infty$, while the most general consistency results available for the real line are described in Csörgő and Mason [2]. Lo [16] proved consistency of the Bayesian bootstrap. In the more general context of the Efron-type bootstrapped empirical process indexed by functions, consistency has been studied by Giné and Zinn [10]. Præstgaard [19] develops analogous results for weighted bootstraps with weights $Y_{n,i} = nV_i/\Sigma_j V_j$ for iid $V_j \geq 0$ satisfying the $L_{2,1}$ integrability condition (2.6) and having mean 1.

Consistency of exchangeably weighted bootstrapped means on the real line has been studied in Mason and Newton [17] using an interesting connection with the theory of rank statistics. When the weights $Y_{n,1}, \ldots, Y_{n,n}$ are exchangeable, and if $(R_{n,1}, \ldots, R_{n,n})$ is a random permutation of the integers from 1 to n (taking each permutation with equal probability $1/n!$ and independent of the data and the weights), then

$$(3.2)\qquad \sqrt{n}(\tilde{X}_n - \bar{X}_n\bar{Y}_n)/(\sigma_n e_n) \stackrel{\mathcal{D}}{=} \frac{1}{n}\sum_{i=1}^{n} Y_n, R_{n,i}(X_i - \bar{X}_n)/(\sigma_e e_n),$$

where σ_n^2 is as above and

$$e_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_{n,i} - \bar{Y}_n)^2.$$

Having introduced a third level of randomness through the *ranks* $R_{n,i}$ we can condition down not only on the data $X_1, \ldots, X_n$, but also on the weights $Y_{n,1}, \ldots,$

$Y_{n,n}$ themselves. What is left has randomness induced by the ranks alone, and in fact takes the form of a standardized linear rank statistic (see Hájek and Šidák [12], for example) for which much asymptotic theory exists. Using a theorem of Hájek [11], Mason and Newton [17] develop sufficient conditions on the weights and the data which ensure consistency of the weighted bootstrapped mean. For weighted bootstraps using iid weights, the Lindeberg-Lévy central limit theorem can be invoked to prove consistency.

Further developments along this line have been obtained by Einmahl and Mason [8]. As a part of a larger study of weighted approximations to exchangeable processes, they extended the results of Csörgő and Mason [2] to general weighted bootstrapped empirical processes in the Mason and Newton [17] context.

4 ASYMPTOTIC EXPANSIONS AND RATES OF CONSISTENCY

All the weighted bootstrapped means surveyed in Section 2 provide consistent estimates for the distribution function of $\bar{X}_n - \mu$ under fairly general conditions. Whether any particular weighted bootstrapped mean is useful in practice depends on properties which go beyond consistency. The property which we are going to discuss here is the rate of consistency. This involves Edgeworth expansions. To illustrate this we will first consider the nonparametric bootstrapped mean $\tilde{X}_n$ given in (2.3), which we denote $\bar{X}_n^*$ to facilitate comparisons.

Recall that for a non-lattice random variable X_1 with finite absolute third moment and variance $\sigma^2 > 0$, the one term Edgeworth expansion for the distribution of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ is given by

$$\text{(4.1)} \qquad \sup_{-\infty<t<\infty} \mid P(\sqrt{n}(\bar{X}_n - \mu)/\sigma \leq t) - \mathcal{E}_{n,1}(t) \mid = o(1/\sqrt{n}),$$

where

$$\mathcal{E}_{n,1}(t) = \Phi(t) - \frac{1}{6\sqrt{n}}\phi(t)(t^2 - 1)\beta,$$

with

$$\beta = E((X_1 - \mu)^3)/\sigma^3$$

being the skewness of X_1, and Φ and ϕ being the cdf and density respectively of a standard normal random variable. Singh [23] has shown that under the same conditions, the conditional distribution function of $\bar{X}_n^*$, given the data, when properly centered and normalized has the Edgeworth expansion, almost surely as $n \to \infty$

$$\text{(4.2)} \quad \sup_{-\infty<t<\infty} \mid P(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)/\sigma_n \leq t \mid X_1, \dots, X_n) - \mathcal{E}_{n,1}^*(t) \mid = o(1/\sqrt{n}),$$

where

$$\mathcal{E}_{n,1}^*(t) = \Phi(t) - \frac{1}{6\sqrt{n}}\phi(t)(t^2 - 1)\beta_n(\{X_i\}),$$

with

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

and

$$\beta_n(\{X_i\}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^3/\sigma_n^3,$$

being the sample variance and sample skewness respectively.

Notice that $\mathcal{E}_{n,1}$ and $\mathcal{E}^*_{n,1}$ are of exactly the same form except that β in $\mathcal{E}_{n,1}$ gets replaced in $\mathcal{E}^*_{n,1}$ by its empirical counterpart $\beta_n(\{X_i\})$. That is, centered population moments get replaced by centered sample moments. For this reason, $\mathcal{E}^*_{n,1}$ is called a one term empirical Edgeworth expansion. For the bootstrap error

$$r_n^*(t) = P(\sqrt{n}(\bar{X}_n - \mu)/\sigma \le t) - P(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)/\sigma_n \le t \mid X_1, \ldots, X_n),$$

we get from (4.1) and (4.2), uniformly in t, almost surely as $n \to \infty$,

$$r_n^*(t) = \frac{1}{6\sqrt{n}}\phi(t)(t^2 - 1)(\beta_n(\{X_i\}) - \beta) + o(1/\sqrt{n}),$$

so that because $\beta_n(\{X_i\}) \to \beta$ almost surely,

$$\sup_{-\infty<t<\infty} \mid r_n^*(t) \mid = o(1/\sqrt{n}) \tag{4.3}$$

almost surely as $n \to \infty$. For the error of the normal approximation

$$r_n(t) = P(\sqrt{n}(\bar{X}_n - \mu)/\sigma \le t) - \Phi(t)$$

we have by the Berry-Esseen theorem

$$\sup_{-\infty<t<\infty} \mid r_n(t) \mid = O(1/\sqrt{n}), \tag{4.4}$$

the rate being optimal in general. The two asymptotic results (4.3) and (4.4) indicate that for large sample sizes n, the distribution of the bootstrapped mean should often provide a closer approximation to the distribution function of the sample mean than the asymptotic normal distribution.

Thus we have seen that an Edgeworth expansion of the conditional distribution function of a bootstrapped mean as the one in (4.2) is crucial for gaining information about the rate of consistency of the procedure. This information is obtained by comparing this Edgeworth expansion to that of the distribution of the sample mean given in (4.1).

For the weighted bootstrapped mean with exchangeable weights, the rank statistics approach towards consistency, as sketched in Section 3, is also useful for obtaining Edgeworth expansions. A recent result of Schneller [22] on Edgeworth expansions for linear rank statistics is especially easy to apply. This is his Theorem 2.12. However, because of his technical condition (2.15), originally due to van Zwet [25], this approach via his theorem 2.12 *just fails* to include the original Singh [23] theorem quoted in (4.2) above. In order to satisfy his condition (2.15) one either has to impose conditions on the weights which do not hold for the multinomial weights of $\bar{X}_n^*$, or conditions on the distribution of X_1 which are stronger than X_1 being non-lattice.

For weights of the type given in (2.5) when $V_1, V_2, \ldots,$ form a sequence of positive random variables, one can obtain rather readily from part (a) of Theorem 2.12 of Schneller [22] the following result. Assume that V_1, V_2 are iid positive random variables such that for some $s > 24/7$

$$E \mid V_1 - EV_1 \mid^{2s} < \infty,$$

and further suppose that $\mid V_1 - V_2 \mid$ has a density which is bounded near zero. Also let $X_1, X_2, \ldots,$ be iid with finite third absolute moment, mean μ, positive variance σ^2 and third central moment μ_3. Then almost surely as $n \longrightarrow \infty$

$$\sup_t \mid P(Z_n \leq t \mid X_1, \ldots, X_n) - \Phi(t) - \tfrac{1}{6\sqrt{n}}\phi(t)(t^2 - 1)\Delta \mid = o(1/\sqrt{n}), \tag{4.5}$$

where Z_n is

$$Z_n = \sqrt{n}(\tilde{X}_n - \bar{X}_n \bar{Y}_n)/(\sigma_n e_n), \tag{4.6}$$

the weights $Y_{n,i} = nV_i/\Sigma_j V_j$, and

$$\Delta = E(V_1 - EV_1)^3/(Var(V_1))^{3/2}$$

is the skewness of V_1. In particular, whenever $\Delta = 1$ and X_1 is non-lattice, combining (4.5) with (4.1) yields

$$\sup_t \mid P(Z_n \leq t \mid X_1, \ldots, X_n) - P(\sqrt{n}(\bar{X}_n - \mu)/\sigma \leq t) \mid = o(1/\sqrt{n}),$$

almost surely as $n \longrightarrow \infty$.

If V_1 is chosen to be Gamma (1,4), then $\Delta = 1$. This agrees with a result of Weng [26] who proved the analogous result for the Gamma (1,4) case when, in Z_n, the e_n^2 as in (3.2) is replaced by 4, the variance of a Gamma (1,4) random variable.

Moreover, one can obtain from Schneller's theorem in much the same way, under the same conditions on $X_1, X_2, \ldots,$ and $V_1, V_2, \ldots,$ (but with Y_1 not necessarily assumed to be positive), the same result for weights of the form

$$Y_{n,i} = V_i. \tag{4.7}$$

This suggests taking a closer look at weighted bootstrapped means of the form

$$\bar{X}_n^Y = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i$$

where Y_i are iid weights which are independent of the sample.

Notice that given the sample $X_1, \ldots, X_n$, the mean $\bar{X}_n^Y$ is a weighted sum of iid random variables with nonrandom weights. This means that Edgeworth expansions for the distribution of the centered mean conditioned on the sample can be readily derived by straightforward modifications of classical proofs for

Edgeworth expansions for the distribution functions of sums of iid random variables. One result that is obtainable in this way is the one term expansion that reads as follows. If $\sigma^2 > 0$, $E \mid X_1 \mid^3 < \infty$, $\sigma_Y^2 > 0$, $E \mid Y_1 \mid^3 < \infty$ and either Y_1 or X_1 is non-lattice, then

$$(4.8) \qquad \sup_{-\infty<t<\infty} \mid P(\sqrt{n}(\bar{X}_n^Y - \bar{X}_n\bar{Y}_n)/(\sigma_n\sigma_Y) \leq t \mid X_1, \ldots, X_n) - \mathcal{E}_{n,1}^Y(t) \mid = o(1/\sqrt{n})$$

almost surely as $n \to \infty$, where

$$\mathcal{E}_{n,1}^Y(t) = \Phi(t) - \frac{1}{6\sqrt{n}}\phi(t)(t^2-1)\beta_n(\{X_i\})\beta^Y,$$

with $\beta_n(\{X_i\})$ being the sample skewness as before and

$$\beta^Y = E((Y_1 - EY_1)^3)/\sigma_Y^3$$

being the skewness of Y_1. From (4.1) and (4.8) we obtain for the error

$$r_n^Y(t) = P(\sqrt{n}(\bar{X}_n - \mu)/\sigma \leq t) - P(\sqrt{n}(\bar{X}_n^Y - \bar{X}_n\bar{Y}_n)/(\sigma_n\sigma_Y) \leq t \mid X_1, \ldots, X_n)$$

of the weighted bootstrapped mean $\bar{X}_n^Y$, that uniformly in t, almost surely as $n \to \infty$

$$r_n^Y(t) = \frac{1}{6\sqrt{n}}\phi(t)(t^2-1)(\beta_n(\{X_i\})\beta^Y - \beta) + o(1/\sqrt{n}),$$

whenever X_1 is also non-lattice. Consequently if $\beta^Y = 1$, then we have

$$(4.9) \qquad \sup_{-\infty<t<\infty} \mid r_n^Y(t) \mid = o(1/\sqrt{n})$$

almost surely as $n \to \infty$. This says that we obtain the same rate of consistency as for the ordinary bootstrap. As far as the consistency rate (4.9) is concerned, this rather arbitrary resampling scheme is indistinguishable from the nonparametric bootstrap!

Further information about the errors $r_n^*(t)$ and $r_n^Y(t)$ can be obtained if Edgeworth expansions of higher orders are taken into account. We study these expansions in the next section and we derive a distribution for iid weights which depends on the data.

5 A TWO-TERM EDGEWORTH EXPANSION FOR $\bar{X}_n^Y$

Firstly, we consider the two-term expansion for the sample mean $\bar{X}_n$. If $\sigma^2 > 0$, $E(X_1^4) < \infty$ and Cramér's condition

$$\limsup_{|t|\to\infty} \mid E\{e^{itX_1}\} \mid < 1$$

is satisfied, then

$$\sup_{-\infty<t<\infty} \mid P(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \leq t) - \mathcal{E}_{n,2}(t) \mid = o\left(\frac{1}{n}\right),$$

where

$$\mathcal{E}_{n,2}(t) = \Phi(t) \quad -\frac{1}{\sqrt{n}}\frac{1}{6}(t^2-1)\phi(t)\beta(X_1)$$
$$-\frac{1}{n}\frac{1}{24}(t^3-3t)\phi(t)\kappa(X_1)$$
$$-\frac{1}{n}\frac{1}{72}(t^5-10t^3+15t)\phi(t)\beta(X_1)^2$$

with $\beta(X_1)$ again being the skewness of X_1 and

$$\kappa(X_1) = \frac{1}{\sigma^4}E\{(X_1-\mu)^4\} - 3$$

being the kurtosis of X_1.

The two-term Edgeworth expansion $\mathcal{E}^*_{n,2}(t)$ for Efron's bootstrapped mean $\bar{X}^*_n$ has the same form with $\beta(X_1)$ replaced by the sample skewness

$$\beta_n(\{X_i\}) = \frac{1}{\sigma_n^3}\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^3$$

and $\kappa(X_1)$ replaced by the sample kurtosis

$$\kappa_n(\{X_i\}) = \frac{1}{\sigma_n^4}\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^4 - 3.$$

Therefore, we get the following two-term expansion for the bootstrap error:

$$\begin{aligned} r^*_n(t) \quad &= \frac{1}{\sqrt{n}}\frac{1}{6}(t^2-1)\phi(t)\{\beta_n(\{X_i\}) - \beta(X_1)\} \\ &+\frac{1}{n}\frac{1}{24}(t^3-3t)\phi(t)\{\kappa_n(\{X_i\}) - \kappa(X_1)\} \\ &+\frac{1}{n}\frac{1}{72}(t^5-10t^3+15t)\phi(t)\{\beta_n(\{X_i\})^2 - \beta(X_1)^2\} \\ &+o\left(\tfrac{1}{n}\right)\ a.s. \text{ uniformly in } -\infty < t < \infty. \end{aligned}$$

The two-term expansion for the weighted bootstrap error, if $E(Y_1^4) < \infty$ and Cramér's condition for Y_1 is satisfied, is given by

$$\begin{aligned} r^Y_n(t) \quad &= \frac{1}{\sqrt{n}}\frac{1}{6}(t^2-1)\phi(t)\{\beta_n(\{X_i\})\beta(Y_1) - \beta(X_1)\} \\ &+\frac{1}{n}\frac{1}{24}(t^3-3t)\phi(t)\{\tilde{\kappa}_n(\{X_i\})\kappa(Y_1) - \kappa(X_1)\} \\ &+\frac{1}{n}\frac{1}{72}(t^5-10t^3+15t)\phi(t)\{\beta_n(\{X_i\})^2\beta(Y_1)^2 - \beta(X_1)^2\} \\ &+o\left(\tfrac{1}{n}\right)\ a.s. \text{ uniformly in } -\infty < t < \infty \end{aligned}$$

where we use the notation $\beta_n(\{X_i\})$ and $\beta(X_1)$ for the sample skewness of $X_1, \ldots, X_n$ and the skewness of X_1 respectively. Moreover, $\kappa(Y_1)$ is the kurtosis of Y_1 and

$$\tilde{\kappa}_n(\{X_i\}) = \frac{1}{\sigma_n^4}\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^4.$$

The reason for having $\tilde{\kappa}$ comes from the fact that $\kappa_j(cX) = c^j \kappa_j(X)$ for the j^{th} cumulant of X. Therefore, if the weights Y_i satisfy $\beta(Y_1) = 1$ as before and in addition

$$(5.1) \qquad \tilde{\kappa}_n(\{X_i\})\kappa(Y_1) = \kappa_n(\{X_i\}),$$

then we have for every fixed t that

$$(5.2) \qquad n\{r_n^*(t) + r_n^Y(t)\} \xrightarrow{\mathcal{D}} N(0, r(t))$$

as $n \to \infty$, where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and $N(0, r(t))$ is a centered normal random variable with variance $r(t) \geq 0$. Since third order terms in the expansions of $r_n^*(t)$ and $r_n^Y(t)$ are of the order $1/n^{3/2}$, from such expansions, which we will not detail here, we obtain for some real constant $c(t)$

$$(5.3) \qquad n^{3/2}\{r_n^*(t) - r_n^Y(t)\} \to c(t)$$

in probability as $n \to \infty$. Multiplying (5.2) and (5.3) yields

$$n^{5/2}\{r_n^*(t)^2 - r_n^Y(t)^2\} \xrightarrow{\mathcal{D}} N(0, r(t)c(t)^2)$$

as $n \to \infty$ and hence (if $r(t)c(t)^2 > 0$) that

$$(5.4) \qquad P(\mid r_n^*(t) \mid > \mid r_n^Y(t) \mid) \xrightarrow{\mathcal{D}} \tfrac{1}{2}.$$

Thus, under appropriate regularity conditions for three-term Edgeworth expansions to hold, and if the weights are chosen so that $\beta(Y_1) = 1$ and (5.1) are satisfied, we see that the weighted bootstrapped mean and Efron's bootstrapped mean are asymptotically equivalent in the sense that the error for one version is greater than the error for the other with (asymptotic) probability 1/2. Now, condition (5.1) is clearly impossible for the weights Y_i which are independent of the observations X_i so that we have to allow for weights with distributions depending on $X_1, \ldots, X_n$. Formally, we have to consider the modified weighted bootstrapped mean

$$\bar{X}_n^Y = \frac{1}{n}\sum_{i=1}^{n} Y_{n,i}X_i,$$

where now for each n the weights $Y_{n,1}, \ldots, Y_{n,n}$ are iid given the observations $X_1, \ldots, X_n$, i.e. under the conditional probability $P(\cdot \mid X_1, \ldots, X_n)$. For these weights, we can indeed obtain the appropriate expansions of $\bar{X}_n^Y$. We give the details here only for the two-term expansion, the three-term expansions being analogous.

THEOREM 5.1. *If with probability one*

$$\{\mid Y_{n,1} \mid^4 : n \geq 1\}$$

is uniformly integrable under $P(\cdot \mid X_1, \ldots, X_n)$ *and*

$$\limsup_{|t|\to\infty} \sup_{n\geq 1} \mid E\{e^{itY_{n,1}} \mid X_1,\ldots,X_n\} \mid < 1,$$

then with probability one

$$\sup_{-\infty<t<\infty} \mid (\sqrt{n}\frac{\bar{X}_n^Y - \bar{Y}_n\bar{X}_n}{\sigma_Y\sigma_n} \leq t \mid X_1,\ldots,X_n) - \mathcal{E}_{n,2}^Y(t) \mid = o\left(\frac{1}{n}\right)$$

where $\sigma_Y^2 = Var(Y_{n,1} \mid X_1,\ldots,X_n)$ *and*

$$\begin{aligned}\mathcal{E}_{n,2}^Y(t) &= \Phi(t) - \tfrac{1}{\sqrt{n}}\tfrac{1}{6}(t^2-1)\phi(t)\beta_n(\{X_i\})\beta(Y_{n,1} \mid X_1,\ldots,X_n)\\ &-\tfrac{1}{n}\tfrac{1}{24}(t^3-3t)\phi(t)\tilde{\kappa}_n(\{X_i\})\kappa(Y_{n,1} \mid X_1,\ldots,X_n)\\ &-\tfrac{1}{n}\tfrac{1}{24}(t^5-10t^3+15t)\phi(t)\beta_n(\{X_i\})^2\beta(Y_{n,1} \mid X_1,\ldots,X_n)\end{aligned}$$

with $\beta(Y_{n,1} \mid X_1,\ldots,X_n)$ *and* $\kappa(Y_{n,1} \mid X_1,\ldots,X_n)$ *being the skewness and kurtosis of* $Y_{n,1}$ *under* $P(\cdot \mid X_1,\ldots,X_n)$.

It remains to find a distribution of $Y_{n,i}$ given $X_1,\ldots,X_n$ for which the terms up to order $\frac{1}{n}$ in the expansions of $r_n^Y(t)$ and $r_n^*(t)$ match up. Without loss of generality we can assume that $Y_{n,i}$ has conditional mean 0 and conditional variance 1. To match up the $1/\sqrt{n}$ terms, we must also have $\beta(Y_{n,1} \mid X_1,\ldots,X_n) = 1$. The remaining condition (5.1), now formulated for the weights $Y_{n,i}$ depending on $X_1,\ldots,X_n$ is equivalent to

$$E\{Y_{n,i}^4 \mid X_1,\ldots,X_n\} = 4 - 3\frac{\sigma_n^4}{\frac{1}{n}\sum_{i=1}^n (X_i-\bar{X}_n)^4} \in [1,4]$$

because $\sigma_n^4 \leq n^{-1}\sum_i (X_i-\bar{X}_n)^4$. Therefore, for any $c \in [1,4]$ we have to find a random variable Y_c such that

$$\text{(5.5)} \qquad E(Y_c) = 0 \quad E(Y_c^2) = 1 \quad E(Y_c^3) = 1 \quad E(Y_c^4) = c.$$

Because $E((u_1 + u_2Y_c + u_3Y_c^2)^2)$ is a nonnegative definite quadratic form in u_1, u_2, u_3, and therefore its matrix must have a nonnegative determinant, the first three moment constraints above imply that $E(Y_c^4) \geq 2$. Thus the above term-matching can be accomplished only for observations $X_1,\ldots,X_n$ for which

$$4 - 3\frac{\sigma_n^4}{\frac{1}{n}\sum_{i=1}^n (X_i-\bar{X}_n)^4} \geq 2$$

or equivalently

$$\text{(5.6)} \qquad \frac{n^{-1}\sum_{i=1}^n (X_i-\bar{X}_n)^4}{\sigma_n^4} \geq \frac{3}{2}.$$

A sufficient condition for (5.6) to hold at least asymptotically with probability one is the unimodality of the underlying distribution. If the distribution of X_1 is unimodal about μ, i.e. if the distribution function is convex on $(-\infty, \mu)$ and concave on (μ, ∞), then it follows from the Choquet-representation of unimodal distributions, cf. Dharmadhikari and Joag-dev [3], that

$$\frac{1}{\sigma^4} E\{(X_1 - \mu)^4\} \geq \frac{9}{5} > \frac{3}{2}$$

and so (5.6) is satisfied with probability one for all large n.

The random variable defined below satisfies (5.5) for $c \in [2,4]$.

$$(5.7) \qquad S_c = \begin{cases} \frac{1}{2}(1+\sqrt{4c-3}) & \text{with probability} \quad \frac{1}{c-1}\frac{\sqrt{4c-3}-1}{2\sqrt{4c-3}} \\ 0 & \quad ,, \qquad\qquad ,, \qquad \frac{c-2}{c-1} \\ \frac{1}{2}(1-\sqrt{4c-3}) & \quad ,, \qquad\qquad ,, \qquad \frac{1}{c-1}\frac{\sqrt{4c-3}+1}{2\sqrt{4c-3}} \end{cases}$$

This random variable is discrete and therefore does not satisfy Cramér's condition, so we can satisfy (5.5) by taking an appropriate convex combination of S_2 from above and the random variable T_4 having probability density function

$$(5.8) \qquad p(t) = \begin{cases} \frac{1}{100} & -3 \leq t < -2 \\ \frac{1}{40} & -2 \leq t < -1 \\ \frac{19}{30} & -1 \leq t < 0 \\ \frac{11}{60} & 0 \leq t < 1 \\ \frac{1}{12} & 1 \leq t < 2 \\ \frac{13}{200} & 2 \leq t < 3 \\ 0 & \text{else} \end{cases}$$

which satisfies (5.5) with $c = 4$.

6 A SIMULATION STUDY OF A WEIGHTED BOOTSTRAP

The considerations of Section 5 lead us to the following weighted bootstrap procedure. Given a data set $X_1, \ldots, X_n$, compute

$$c_n = 4 - 3\frac{\sigma_n^4}{n^{-1}\sum_i (X_i - \bar{X}_n)^4}.$$

If $c_n \leq 2$, then generate a multinomial vector of weights $(Y_{n,1}, \ldots, Y_{n,n})$ as in (1.2). If $c_n > 2$, generate $Y_{n,i}$ iid from the mixture

$$Y_{n,i} = \begin{cases} S_{2,i} & \text{with probability} \quad 2 - c_n/2 \\ T_{4,i} & \quad ,, \qquad\qquad ,, \qquad c_n/2 - 2 \end{cases}$$

where $S_{2,i}$ are independent copies of S_2 from (5.7) and $T_{4,i}$ are independent copies of T_4 from (5.8).

In a small simulation experiment, we have carried out this special weighted bootstrap and compared it with Efron's bootstrap. For four different cases (Figures 1 through 4 respectively) $nsim = 1000$ data sets of sample size n were generated. The data are independent Gamma $(1, a)$ random variables, i.e. with shape parameter a and scale parameter 1. Note that for a Gamma $(1, a)$ random variable with $a > 0$,

$$c_n \rightarrow_{a.s.} 4 - \frac{a}{a+2} > 2.$$

For each data set, both Efron's bootstrap and this special bootstrap were performed, each based on $nboot = 10000$ sampled weight vectors. To summarize these 1000 pairs of bootstrap distributions, we look only at boxplots of 5 quantiles of these distributions. The results are intriguing; in many cases, the median quantile from this admittedly strange bootstrap is closer to the true quantile of interest than the median quantile from Efron's bootstrap, although the variability is generally larger. As an aside we mention that in connection with nonparametric function estimation, Härdle and Mammen [14] have introduced the idea of wild bootstrapping which is somewhat in the same spirit as the strange bootstrap studied here.

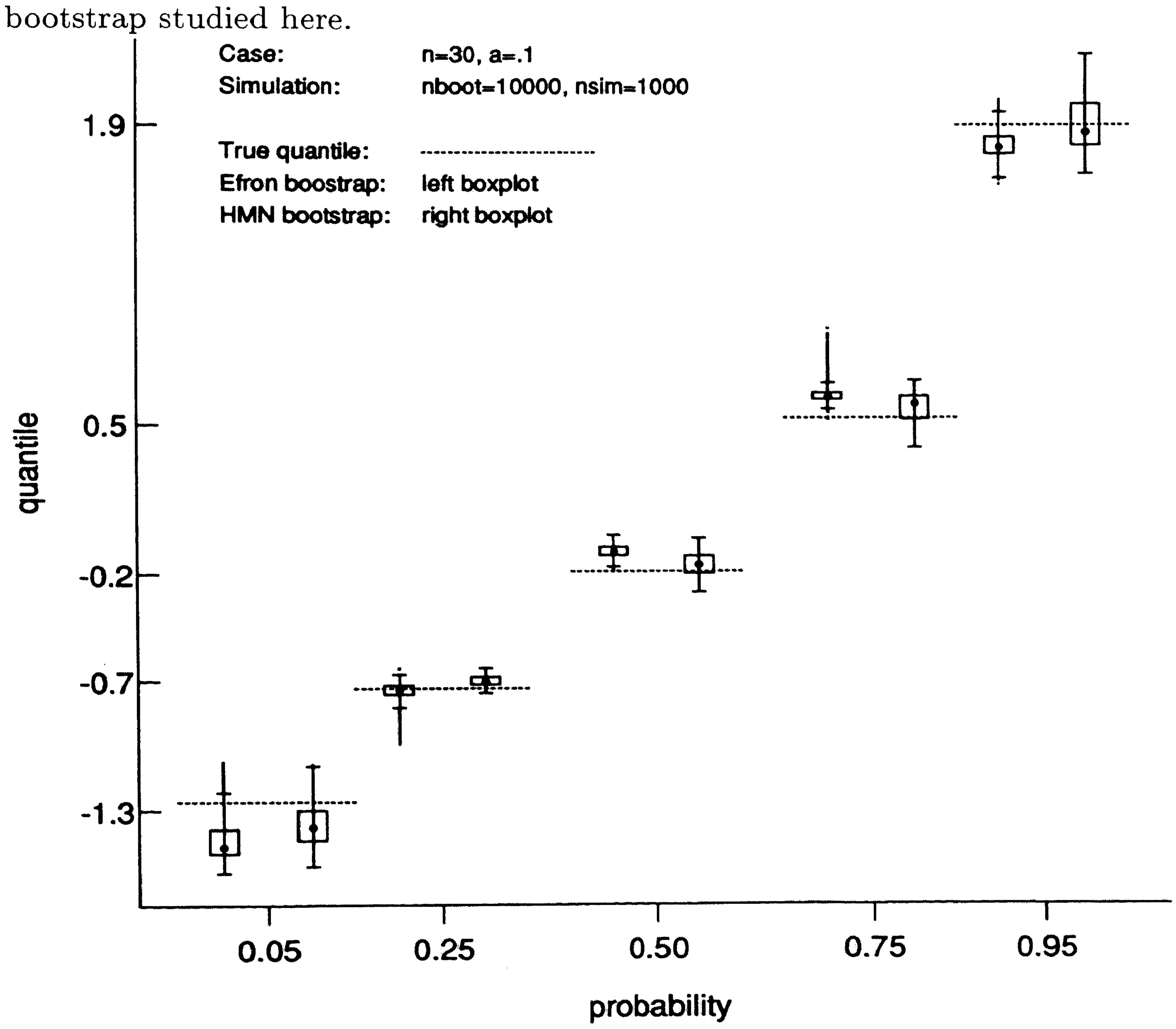

Figure 1.

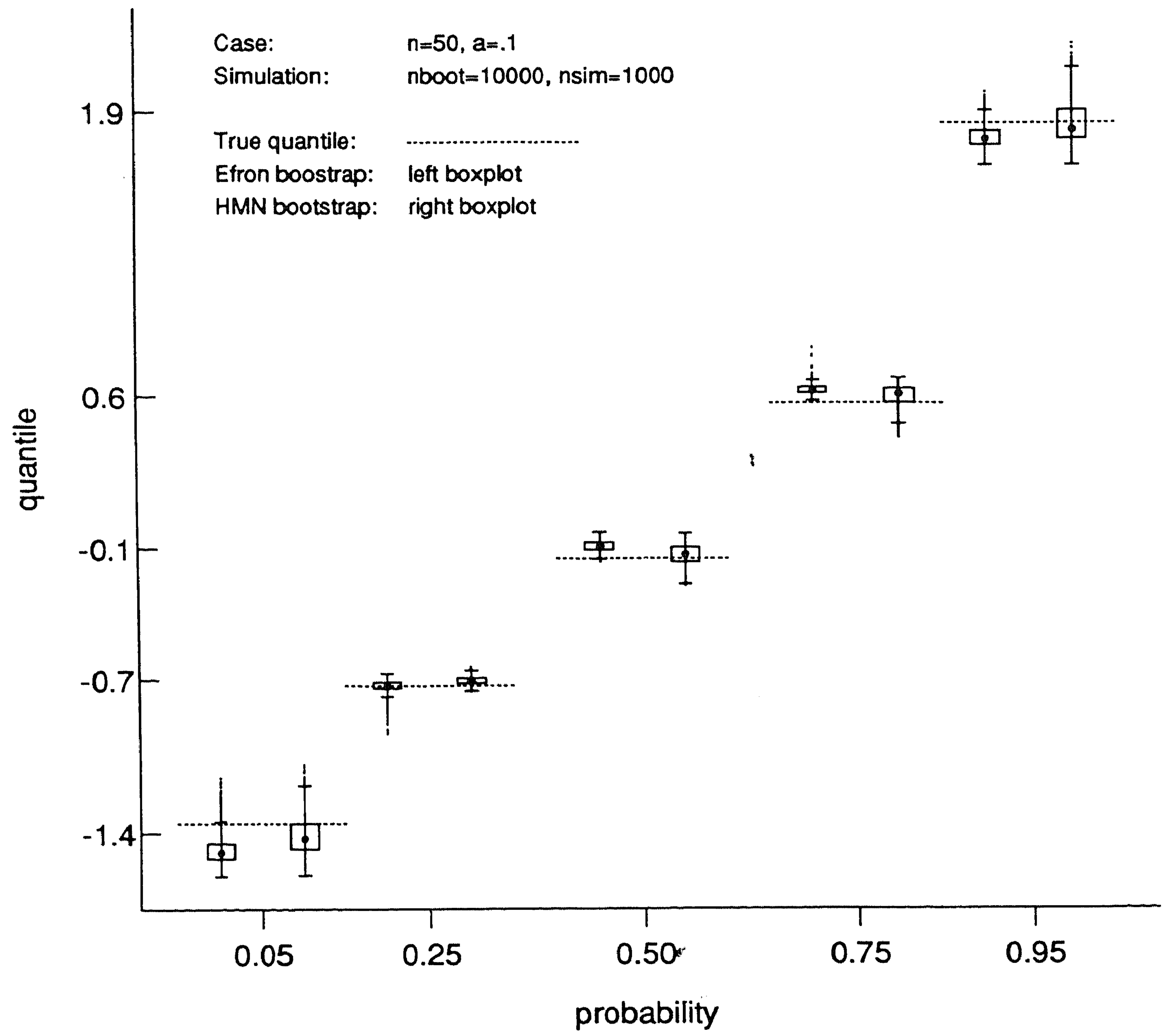

Figure 2.

REFERENCES

1. P. BICKEL, D.A. FREEDMAN (1981). Some asymptotic theory for the bootstrap. *Ann. Statist. 9*, 1196-1217.
2. S. CSÖRGŐ, D.M. MASON (1989). Bootstrapping empirical functions. *Ann. Statist. 17*, 1447-1471.
3. S.W. DHARMADHIKARI, K. JOAG-DEV (1988). *Unimodality, Convexity and Applications.* Boston: Academic Press.
4. H. DEHLING, H. DENKER, W.A. WOYCZINSKI (1990). Resampling U-statistics using p-stable laws. *J. Multivariate Anal. 34*, 1-13.
5. T.J. DICICCIO, J.P. ROMANO (1988). A review of bootstrap confidence intervals. *J.R. Statist. Soc. B 50*, 338-354.
6. B. EFRON (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist. 7*, 1-26.
7. B. EFRON (1982). The jackknife, the bootstrap and other resampling plans. In *Regional Conference Series in Applied Mathematics.* No. 38 Philadelphia: SIAM.
8. U. EINMAHL, D.M. MASON (1991). Approximations to permutation and exchangeable processes. *J. Theor. Probab.* To appear.
9. T. FERGUSON (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist. 1*, 209-230.

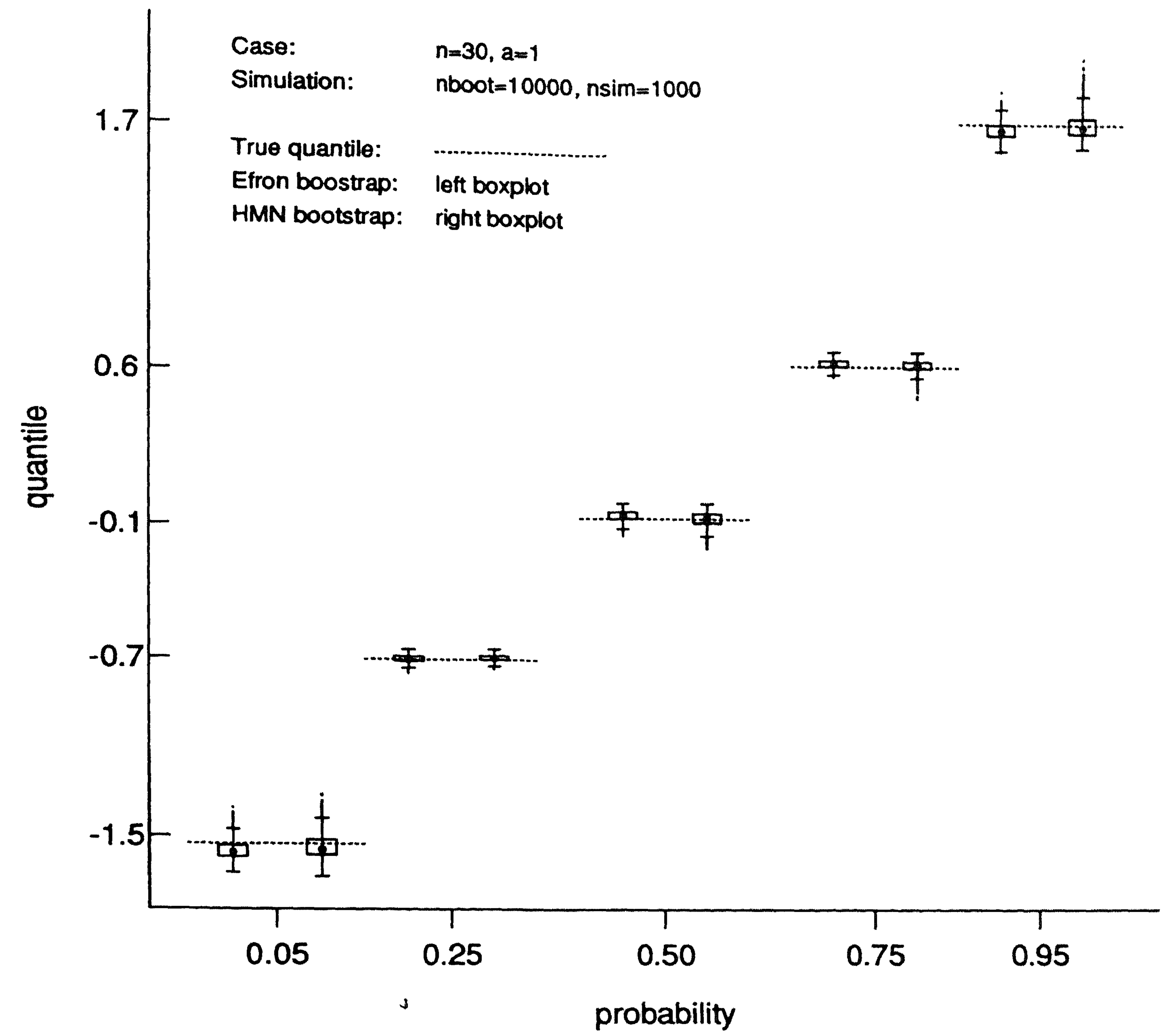

Figure 3.

10. E. GINÉ, J. ZINN (1990). Bootstrapping general empirical functions. *Ann. Prob. 18*, 851-869.
11. J. HÁJEK (1961). Some extensions of the Wald-Wolfowitz-Noether theorem. *Ann. Math. Statist. 32*, 506-523.
12. J. HÁJEK, Z. ŠIDÁK (1967). *Theory of Rank Tests*, Academic Press, New York.
13. P. HALL (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist. 16*, 927-953.
14. W. HÄRDLE, E. MAMMEN (1989). Comparing non-parametric versus parametric regression fits. *Discussion Paper A-177*, University of Bonn.
15. D.V. HINKLEY (1988). Bootstrap methods. *J.R. Statist. Soc. B 50*, 321-337.
16. A.Y. LO (1987). A large sample study of the Bayesian bootstrap. *Ann. Statist. 15*, 360-375.
17. D. MASON, M.A. NEWTON (1990). A rank statistics approach to the consistency of a general bootstrap, *Technical Report 190*, Department of Statistics, University of Washington.
18. M.A. NEWTON, A.E. RAFTERY (1991). Approximate Bayesian inference by the weighted likelihood bootstrap, *Technical Report 199*, Department of Statistics, University of Washington.

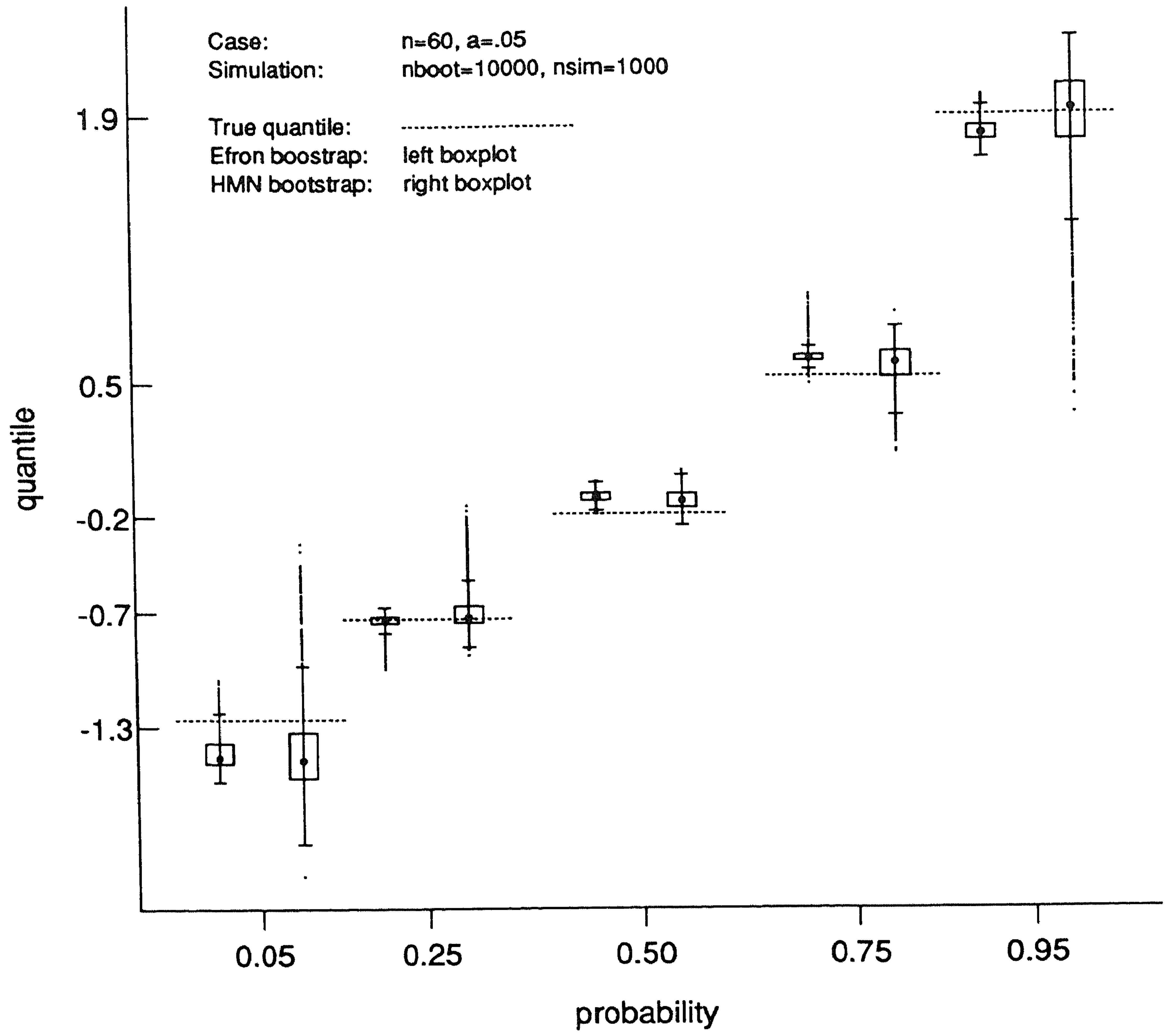

Figure 4.

19. J. PRÆSTGAARD (1990). Bootstrap with general weights and multiplier central limit theorems, *Technical Report 195*, Department of Statistics, University of Washington.
20. M.H. QUENOUILLE (1956). Notes on bias in estimation. *Biometrika 43*, 353-360.
21. D. RUBIN (1981). The Bayesian Bootstrap. *Ann. Statist. 9*, 130-134.
22. W. SCHNELLER (1989). Edgeworth expansion for the linear rank statistic. *Ann. Statist. 17*, 1103-1123.
23. K. SINGH (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist. 9*, 1186-1195.
24. J.W.H. SWANEPOEL (1986). A note on proving that the (modified) bootstrap works. *Common. Statist. - Theory Meth. 15*, 3193-3203.
25. W.R. VAN ZWET (1981). On the Edgeworth expansion for the simple linear rank statistic. In: *Colloq. Math. Soc. János Bolyai (B.V. Gnedenko, M.L. Puri, and I. Vince, eds.) 32*, 889-909, North-Holland, Amsterdam.
26. C.S. WENG (1989). On a second-order asymptotic property of the Bayesian bootstrap mean. *Ann. Statist. 17*, 705-710.
27. Z. ZHENG, D. TU (1988). Random weighting method in regression models. *Scientia Sinica Series A, XXXI*, 1442-1459.