ON CHOICE OF m FOR THE m OUT OF n BOOTSTRAP IN HYPOTHESIS TESTING

Peter J. Bickel and Jian-Jian Ren
University of California - Berkeley and University of Nebraska - Lincoln

Abstract. This paper considers the use of the m out of n bootstrap (Bickel, Göetze and van Zwet, 1994) in setting critical values for a quite general class of hypothesis testing problems. We show that the usual n out of n nonparametric bootstrap generally fails to estimate the null distribution of the test statistics, and that if the $m = o(n)$ out of n bootstrap is used to set the critical value of the tests, the procedure is asymptotically consistent. The critical issue of choice of m is considered and a method of selecting m is proposed. We show that the proposed method of selecting m is asymptotically consistent, and present some simulation results on the proposed method.

AMS 1991 subject classifications. 62G09, 62G10.
Key words and phrases. Doubly censored data, Hadamard differentiability, median, right censored data, statistical functional, trimmed mean.

May 25, 1996

Peter Bickel is a professor in Department of Statistics, University of California - Berkeley, 367 Evans Hall, Berkeley, CA 94720, USA. Jian-Jian Ren is an assistant professor in Department of Mathematics and Statistics, University of Nebraska - Lincoln, Lincoln, NE 68588-0323, USA. The first author's research was partially supported by NSF grant DMS95-04955, and the second author's research was partially supported by NSF grant DMS95-10376.

1. INTRODUCTION

It is logically clear but not always evident or appreciated that the usual nonparametric bootstrap (the n out of n bootstrap) should fail when one tries to estimate the distribution of test statistics under a semiparametric (restricted nonparametric) hypothesis and ignores the restrictions imposed by the hypothesis. For example, Freedman (1981) points out that in setting confidence intervals for the usual slope estimate for regression through the origin, one must resample not the residuals but the residuals centered at their mean. If one considers setting confidence bands as the dual of hypothesis testing, a moment's thought will show that not centering the residuals is tantamount to not imposing the model requirement that the expectation of the error is 0. For more recent examples, see Hārdle and Mammen (1993), Mammen (1992), Bickel, Gōetze and van Zwet (1994). Particularly, Bickel and Ren (1995) showed that the usual n out of n nonparametric bootstrap fails to estimate the null distribution of the Cramér-von Mises test statistics in goodness of fit tests with doubly censored data. They propose that one uses the m out of n bootstrap to set the critical value of the test and show that the proposed testing procedure is asymptotically consistent and has correct power against $\sqrt{n}$ - alternatives.

In this paper, we consider the use of the m out of n bootstrap (Bickel, Gōetze and van Zwet, 1994) in a general class of hypothesis testing problems and consider the critical issue of choice of m. The complete sample case, the right censored sample case and the doubly censored sample case are all considered in the same framework.

The paper is organized as follows: Section 2 shows that the n out of n bootstrap fails to estimate the null distribution of the test statistic for a general class of hypothesis testing problems, and proposes the use of the m out of n bootstrap in setting critical values in these tests. Section 2 also shows that generally, the proposed m out of n bootstrap method is asymptotically consistent for hypothesis testing problems. Proofs are sketched in the appendix. For a quite general class of hypothesis testing problems, Section 3 proposes a method of selecting m in the m out of n bootstrap testing procedure, and establishes the asymptotic consistency of the selection method with proofs deferred to the appendix. Section 4 presents some simulation results on the proposed method of selecting m, and Section 5 includes some concluding remarks.

2. The m OUT OF n BOOTSTRAP IN HYPOTHESIS TESTING

Since the two-tailed case can be studied analogously, we only consider the one-tailed testing problem:

$$H_0\colon \theta = 0 \qquad \text{vs} \qquad H_1\colon \theta > 0 \tag{2.1}$$

where $\theta = T(F)$, F is the underlying continuous distribution function (d.f.) and $T(\cdot)$ is a statistical functional. The following are some examples of the statistical functional $T(\cdot)$.

Example 1. Mean: If θ is the mean of F, then $T(F) = \int_{-\infty}^{\infty} x\,dF(x)$.

Example 2. α-trimmed mean: If θ is the α-trimmed mean of F, then

$$T(F) = \frac{1}{(1-2\alpha)} \int_{\alpha}^{1-\alpha} F^{-1}(x)\,dx.$$

Example 3. Median: If θ is the median of F, then $T(F) = F^{-1}(0.5)$.

If an independently and identically distributed (i.i.d.) sample: $X_1, \cdots, X_n$ from F is observed, the empirical d.f. $\hat{F}_n$ based on this sample is the *nonparametric maximum likelihood estimator* (NPMLE) of F. Since $\sqrt{n}[\hat{F}_n - F]$ weakly converges to a centered Gaussian process (see Shorack and Wellner, 1986), from Theorem II.8.1 (Andersen, Borgan, Gill and Keiding, 1993) we know that a Hadamard differentiable (or compact differentiable) functional $T(\cdot)$ implies

$$\sqrt{n}[T(\hat{F}_n) - T(F)] \xrightarrow{D} T'_F(G_F) = \mathrm{N}(0, \sigma_F^2), \qquad \text{as } n \to \infty \tag{2.2}$$

where $0 < \sigma_F^2 < \infty$, G_F is a centered Gaussian process and T'_F is the Hadamard derivative of $T(\cdot)$ at F. Hence, the test statistic for (2.1) is given by

$$T_n = \sqrt{n}\,T(\hat{F}_n). \tag{2.3}$$

Since incomplete data are frequently encountered in medical research and reliability research, we note that (2.2) and (2.3) also apply to the following censored data cases.

In a *right censored sample*:

$$V_i = \begin{cases} X_i, & \text{if } X_i < Y_i, \quad \delta_i = 1 \\ Y_i, & \text{if } X_i \geq Y_i, \quad \delta_i = 0 \end{cases} \tag{2.4}$$

where Y_i is the *right censoring variable* and is independent from X_i, the Kaplan-Meier estimator based on (V_i, δ_i), $i = 1, \cdots, n$, still denoted as $\hat{F}_n$, is the NPMLE of F.

In a *doubly censored sample* (Turnbull, 1974):

$$V_i = \begin{cases} X_i, & \text{if } Z_i \leq X_i < Y_i, \quad \delta_i = 1 \\ Y_i, & \text{if } X_i \geq Y_i, \quad \delta_i = 2 \\ Z_i, & \text{if } X_i < Z_i, \quad \delta_i = 3 \end{cases} \tag{2.5}$$

where X_i is independent from (Y_i, Z_i) with $P\{Y_i > Z_i\} = 1$ and Y_i and Z_i are the *right* and *left censoring variables*, respectively, the NPMLE $\hat{F}_n$ of F can be numerically computed (Mykland and Ren, 1994) based on (V_i, δ_i), $i = 1, \cdots, n$. One may note that the right censored case (2.4) is a special case of the doubly censored case (2.5).

Since the weak convergence of the NPMLE $\hat{F}_n$ for right censored data (2.4) and for doubly censored data (2.5) have been established (Gill, 1983; Gu and Zhang, 1993), respectively, we know that (2.2) follows from Theorem II.8.1 (Andersen, Borgan, Gill and Keiding, 1993), and thus (2.3) is the test statistic based on (2.4) or (2.5) for (2.1). For the rest of the paper, we refer (2.3) to the test statistics for (2.1) either with a complete i.i.d. sample or a right censored sample (2.4) or a doubly censored sample (2.5).

One should note that for censored data, the variance σ_F^2 in (2.2) also depends on the censoring variable distributions, which can be quite complicated and are usually not easy to estimate even under the null hypothesis. To set the critical value of the tests, one needs to estimate the unknown null distribution of the test statistic T_n given in (2.3). As expected, the usual nonparametric n out of n bootstrap fails in this case. We justify this as follows.

For the complete data case or the censored data case above, we first note the following under suitable conditions (Giné and Zinn, 1990; Bickel and Ren, 1995):

$$\sqrt{n}\,[\hat{F}_n - F] \xrightarrow{D} G_F \qquad \Rightarrow \qquad \sqrt{m}\,[\hat{F}_m^* - \hat{F}_n] \xrightarrow{D} G_F, \quad a.s. \tag{2.6}$$

where G_F is a centered Gaussian process, $\hat{F}_m^*$ is the NPMLE based on the bootstrap sample of size m with replacement, and $m \to \infty$, as $n \to \infty$. We also note that from Theorem II.8.1 (Andersen, Borgan, Gill and Keiding, 1993), the weak convergence of $\sqrt{n}\,[\hat{F}_n - F]$ and the Hadamard differentiability property of $T(\cdot)$ ensure

$$\sqrt{n}\,T(\hat{F}_n) = \sqrt{n}\,T(F) + T'_F(\sqrt{n}\,[\hat{F}_n - F]) + o_p(1), \qquad \text{as } n \to \infty \tag{2.7}$$

where T'_F is the Hadamard derivative of $T(\cdot)$ at F and is a linear functional. From (2.6), we have the tightness of $\sqrt{m}\,[\hat{F}_m^* - F]$ for $m = O(n)$, and thus from Fernholz (1983, Chapter 4),

$$\sqrt{m}\,T(\hat{F}_m^*) = \sqrt{m}\,T(F) + T'_F(\sqrt{m}\,[\hat{F}_m^* - F]) + o_p(1), \qquad \text{as } n \to \infty. \tag{2.8}$$

Hence, under the null hypothesis of (2.1), the n out of n bootstrap gives $m = n$ in (2.8) and thus from (2.6),

$$\begin{aligned}
\sqrt{n}\,T(\hat{F}_n^*) &= T'_F(\sqrt{n}\,[\hat{F}_n^* - F]) + o_p(1) \\
&= T'_F(\sqrt{n}\,[\hat{F}_n^* - \hat{F}_n]) + T'_F(\sqrt{n}\,[\hat{F}_n - F]) + o_p(1) \\
&\overset{D}{\approx} T'_F(G_F^*) + T'_F(G_F), \qquad \text{as } n \to \infty
\end{aligned}$$

where G_F^* and G_F are two independent centered Gaussian processes with the same covariance function. Evidently, $\sqrt{n}\,T(\hat{F}_n^*)$ does not have the same distribution as the null distribution which is $T'_F(G_F)$ in (2.2). So we have a bootstrap failure.

To see that the m out of n bootstrap gives correct estimate for the null distribution under H_0, one may note that if $m = o(n)$ in (2.8), then from (2.6) and the weak convergence of $\sqrt{n}[\hat{F}_n - F]$, we have that for $m \to \infty$,

$$\sqrt{m}T(\hat{F}^*_m) = T'_F(\sqrt{m}[\hat{F}^*_m - \hat{F}_n]) + \sqrt{m/n}T'_F(\sqrt{n}[\hat{F}_n - F]) + o_p(1) \overset{D}{\approx} T'_F(G_F). \tag{2.9}$$

Thus, we propose that one uses $m = o(n)$, $m \to \infty$, as the bootstrap sample size, and use C^*_α as the critical value of the test (2.1), where for $0 < \alpha < 1$, C^*_α is given by

$$P_n\{T^*_m \geq C^*_\alpha\} = \alpha. \tag{2.10}$$

This is called *the m out of n bootstrap method for hypothesis testing.*

In the next theorem, we show that generally, the proposed m out of n bootstrap method is asymptotically consistent for hypothesis testing problems. Proofs are deferred to the appendix. Let

$$\mathfrak{F}_0 = \{\text{collection of distribution functions}\}$$

and consider the test

$$H_0\text{: } F \in \mathfrak{F}_0 \tag{2.11}$$

where F is the underlying distribution from which a random sample $O_1, \cdots, O_n$ is drawn. In particular, one should note that $O_i = X_i$ for the complete i.i.d. sample case, and $O_i = (V_i, \delta_i)$ for the censored sample case with F to be the distribution of (X_i, Y_i) or (X_i, Y_i, Z_i). Suppose that $T_n = T_n(O_1, \cdots, O_n; F)$ is the test statistic for (2.11) and that H_0 is rejected for large T_n. Then, denoting

$$\mathfrak{L} = \{h\text{:}\mathbf{R} \to \mathbf{R};\ |h(x) - h(y)| \leq |x - y|,\ \|h\| \leq 1\}$$

and for $h \in \mathfrak{L}$,

$$\theta_m(F) \equiv E_F\{h(T_m(O_1, \cdots, O_m; F))\},$$

$$\theta_{m,n}(F) \equiv n^{-m} \sum_{r=1}^{m} \sum_{i_1 + \cdots + i_r = m,\, i_j \geq 0} \binom{n}{r} \binom{m}{i_1, \cdots, i_r} E_F\{h(T_m(O_1^{(i_1)}, \cdots, O_r^{(i_r)}; F))\},$$

where $O_i^{(j)} \equiv (O_i, \cdots, O_i)_{1 \times j}$,

THEOREM 1. *Let $m = o(n)$ with $m \to \infty$, as $n \to \infty$ and let C^*_α be given by* (2.10). *Assume:*

(a) under H_0, $T_n \xrightarrow{D} W_0$, as $n \to \infty$, where $P\{W_0 \geq C^0_\alpha\} = \alpha$ for a continuous W_0;

(b) for $F \notin \mathfrak{F}_0$, $T_n \xrightarrow{P} \infty$ and $T_m/T_n \xrightarrow{P} 0$, as $n \to \infty$;

(c) $\sup_{h \in \mathfrak{L}} |\theta_{m,n}(F) - \theta_m(F)| = o(1)$.

Then,

(i) For $0 < \alpha < 1$, $\lim_{n \to \infty} P\{T_n \geq C^*_\alpha \mid H_0\} = P\{W_0 \geq C^0_\alpha\} = \alpha$;

(ii) For alternatives H_n: $F = F_n$ such that for some $F_0 \in \mathfrak{F}_0$, $\{F_n\}$ are contiguous to

F_0, we have that under H_n,

$$C_\alpha^* \xrightarrow{P} C_\alpha^0, \qquad as\ n \to \infty.$$

Therefore, if under H_n, $T_n \xrightarrow{D} W_a$, we have

$$lim_{n\to\infty} P\{T_n \geq C_\alpha^* \mid H_n\} = P\{W_a \geq C_\alpha^0\};$$

(iii) For fixed alternatives H_1: $F = F_1 \notin \mathfrak{F}_0$,

$$lim_{n\to\infty} P\{T_n \geq C_\alpha^* \mid H_1\} = 1.$$

REMARK 1. Assumption (c) in Theorem 1 is to ensure that statistic T_m is not greatly affected by ties in its arguments. Bickel, Gõetze and van Zwet (1994) give some simple and easily verifiable sufficient conditions for (c).

REMARK 2. Theorem 1 includes the Theorem of Bickel and Ren (1995) as a special case. Theorem 1 also includes those Hadamard differentiable test statistics given by (2.3) as special cases, where we have that for H_0 given by (2.1),

$$\lim_{n\to\infty} P\{T_n \geq C_\alpha^0 \mid H_0\} = P\{T'_F(G_F) \geq C_\alpha^0 \mid H_0\} = \alpha.$$

and

$$\lim_{n\to\infty} P\{T_n \geq C_\alpha^* \mid H_n\} = P\{\delta + T'_{F_0}(G_{F_0}) \geq C_\alpha^0\}$$

for $\sqrt{n}T(F_n) \to \delta$ and $\| \sqrt{n}[F_n - F_0] - \Delta_0 \| \to 0$.

We note that to apply the above proposed method in practice, the critical issue is the choice of m. This is investigated in the next section.

3. CHOICE OF m.

In this section, we consider the test (2.1) and the test statistics T_n given by (2.3). To select an m in the proposed m out of n bootstrap testing procedure, we note that (2.7) and (2.8) hold when $T(\cdot)$ is Hadamard differentiable and for the NPMLE $\hat{F}_n$, $\sqrt{n}[\hat{F}_n - F]$ converges. Now from (2.9), we know that under H_0, we should choose m such that

$$T_m^* = \sqrt{m}T(\hat{F}_m^*) \overset{D}{\approx} T'_F(G_F). \tag{3.1}$$

Note that for any given F, (2.7) and (2.8) along with (2.6) give

$$\begin{aligned} T_m^* &= \sqrt{m}T(\hat{F}_m^*) = \sqrt{m}T(F) + T'_F(\sqrt{m}(\hat{F}_m^* - F)) + o_p(1) \\ &= \sqrt{m}T(F) + T'_F(\sqrt{m}(\hat{F}_m^* - \hat{F}_n)) + T'_F(\sqrt{m}(\hat{F}_n - F)) + o_p(1) \\ &= T'_F(\sqrt{m}(\hat{F}_m^* - \hat{F}_n)) + \sqrt{m/n}\Big\{\sqrt{n}T(F) + T'_F(\sqrt{n}(\hat{F}_n - F))\Big\} + o_p(1) \qquad (3.2) \\ &= T'_F(\sqrt{m}(\hat{F}_m^* - \hat{F}_n)) + \sqrt{m/n}T_n + o_p(1) \overset{D}{\approx} T'_F(G_F) + \sqrt{m/n}T_n. \end{aligned}$$

Hence, we would want to choose m such that the term $\sqrt{m/n}\,T_n$ in (3.2) is 'negligible'. This calculation does not take into account $1/\sqrt{m}$ terms coming from the Edgeworth expansion of the distribution of $\sqrt{m}\,T(\hat{F}_m^*)$ and implicitly assumes that $o_p(1)$ in (3.2) is of smaller order than $\sqrt{m/n}\,T_n$. Nevertheless, this crude first order attack gives reasonable answers. One may note that there are two points which should be taken into account when one sets a selection rule for m: (i) m should be small enough to have small $\sqrt{m/n}\,T_n$; (ii) m should be as large as possible to have (2.8). In this context, it is reasonable to consider the concept of '*small*' $\sqrt{m/n}\,T_n$ with respect to the $(1-\alpha)th$ percentile of T_m^*. We explain this concept as follows.

One may note that (3.2) suggests $C_\alpha^* \approx C_\alpha^0 + \sqrt{m/n}\,T_n$ under H_0, which implies that the Type I error of the test satisfies

$$P\{T_n \geq C_\alpha^* \mid H_0\} \approx P\{T'_F(G_F) \geq C_\alpha^0 + \sqrt{m/n}\,T_n \mid H_0\}.$$

Thus for $\epsilon > 0$ and $e_\epsilon = \min\{C_{\alpha-\epsilon}^0 - C_\alpha^0, C_\alpha^0 - C_{\alpha+\epsilon}^0\}$, if $\sqrt{m/n}\,|T_n| \leq e_\epsilon$, then asymptotically $|P\{T_n \geq C_\alpha^* \mid H_0\} - \alpha| \leq \epsilon$.

Given $\epsilon > 0$, we need to estimate e_ϵ in practice. From (3.2) and (2.7), we know that if $T(F) = O(1/\sqrt{n})$ (i.e., under $\sqrt{n}$-local alternatives), we may use sampling $m = \sqrt{n}$ (say) out of n to estimate e_ϵ; that is use e_ϵ^* to estimates e_ϵ, where

$$e_\epsilon^* = \min\{C_{\alpha-\epsilon}^* - C_\alpha^*, C_\alpha^* - C_{\alpha+\epsilon}^*\} \tag{3.3}$$

for C_α^* and $C_{\alpha\pm\epsilon}^*$ given by (2.10) with $m = \sqrt{n}$. Thus, we choose $\hat{m}$ such that

$$\sqrt{\hat{m}/n}\,|T_n| \leq e_\epsilon^* \quad \Leftrightarrow \quad \hat{m} = \Big(\frac{e_\epsilon^*}{T_n}\Big)^2 n. \tag{3.4}$$

This selection procedure of m for the m out of n bootstrap in hypothesis testing is summarized as follows.

SELECTION OF m FOR THE m OUT OF n BOOTSTRAP:

(MNB1) For a desirable $\epsilon > 0$, use bootstrap samples with size $m = \sqrt{n}$ to compute e_ϵ^* given in (3.3);

(MNB2) Use bootstrap samples with size $\hat{m}$ given by (3.4) to find the critical value $\hat{C}_\alpha^*$ which is given by (2.10) with $m = \hat{m}$;

(MNB3) Use the test statistic T_n given by (2.3) and the critical value $\hat{C}_\alpha^*$ found in (MNB2) to draw conclusions.

In the next theorem, we give the asymptotic results on the above proposed selection method for m. The proof is given in the appendix.

THEOREM 2. *For the test* (2.1) *with a complete i.i.d. sample or a right censored sample* (2.4) *or a doubly censored sample* (2.5), *if T_n in* (2.3) *is used as the test statistic* *and if* $\hat{C}^*_\alpha$ *by* (MNB1)-(MNB3) *is used as the critical value of the test, then* (2.6) *and a Hadamard differentiable* $T(\cdot)$ *at* F *imply that for any* $\epsilon > 0$,

(*i*) *for* $0 < \alpha < 1$, *under* H_0, *we have* $\hat{m} = O_p(n)$ *and*

$$| \lim_{n\to\infty} P\{T_n \geq \hat{C}^*_\alpha \mid H_0\} - \alpha \mid \leq \epsilon; \tag{3.5}$$

(*ii*) *for contiguous alternatives* H_n: $\theta = \theta_n = T(F_n)$, *where* $\| \sqrt{n}[F_n - F_0] - \Delta_0 \| \to 0$ *and* $\sqrt{n}\theta_n \to \delta$, *as* $n \to \infty$ *with* F_0 *satisfying* H_0 *and* T'_{F_0} *exists, we have* $\hat{m} = O_p(n)$ *and*

$$\begin{aligned} | \lim_{n\to\infty} P\{T_n \geq \hat{C}^*_\alpha \mid H_n\} - P\{\delta + T'_{F_0}(G_{F_0}) \geq C^0_\alpha\} | \\ \leq P\{C^0_{\alpha+\epsilon} \leq \delta + T'_{F_0}(G_{F_0}) \leq C^0_{\alpha-\epsilon}\}; \end{aligned} \tag{3.6}$$

(*iii*) *for fixed alternative* H_1: $\theta = \theta_1 > 0$, *we have* $\hat{m} = O_p(1)$ *and*

$$\lim_{n\to\infty} P\{T_n \geq \hat{C}^*_\alpha \mid H_1\} = 1. \tag{3.7}$$

REMARK 3. Theorem 2-(iii) shows that using $\hat{C}^*_\alpha$ by (MNB1)-(MNB3), the power of the m out of n bootstrap test approaches 1 faster than that using C^*_α, given by (2.10) with $m = o(n)$, as the critical value, because under H_1, (2.8) implies $C^*_\alpha = O_p(\sqrt{m})$ while $\hat{C}^*_\alpha = O_p(1)$. This is clearly supported by the simulation results presented in Section 4.

4. SIMULATION RESULTS.

In this section, we present some simulation results on the proposed method (MNB1)-(MNB3). We denote $N(\mu, \sigma^2)$ as a normal distribution with mean μ and variance σ^2, and Cauchy(μ) as a standard Cauchy distribution with median μ. For T_n given by (2.3), let $C^{(n)}_\alpha$, given by

$$P\{T_n \geq C^{(n)}_\alpha \mid H_0\} = \alpha, \tag{4.1}$$

be the true critical value of the test (2.1). We denote the power functions of the test (2.1) with the true critical value $C^{(n)}_\alpha$ given by (4.1), the m out of n bootstrap critical value $\hat{C}^*_\alpha$ given by (MNB2) and the m out of n bootstrap critical value C^*_α given by (2.10) with $m = \sqrt{n}$ as

$$P_0(t) = P\{T_n \geq C^{(n)}_\alpha \mid \theta = t\},$$

$$P_b(t) = P\{T_n \geq \hat{C}^*_\alpha \mid \theta = t\},$$

$$P_2(t) = P\{T_n \geq C^*_\alpha \mid \theta = t\},$$

respectively, where T_n is given by (2.3). All simulation results presented below are based on 300 runs for complete or right censored samples, and 100 runs (because it is very time consuming to conduct the simulation studies) for doubly censored samples or Cauchy samples of size $n = 1000$. In all the tests considered here, $\alpha = 0.05$ and $\epsilon = 0.025$

are used in (MNB1)-(MNB3).

In Figure 1, we compare the power curves of P_0, P_b and P_2 for a median test with a complete i.i.d. sample of size $n = 400$ from $N(\mu, 25)$. In our study, the true critical value $C_\alpha^{(n)}$ is obtained by the Monte Carlo method, all power curves are the average of the 300 simulation runs, and for each run the percentiles of P_0, P_b, P_2, $\hat{C}_\alpha^*$, C_α^* are obtained from 400 bootstrap samples. Figure 2 and Figure 3 compare P_0 and P_b for the same median test as Figure 1 with right censored data and doubly censored data, respectively.

P_0: -----; P_b: • • •; P_2: -----; $X \sim N(\mu, 25)$

Figure 1.

P_0: -----; P_b: • • •; $X \sim N(\mu, 25)$, $Y \sim N(3, 36)$

Figure 2.

P_0: -----; P_b: • • •; $X \sim N(\mu, 25)$, $Y \sim N(3, 36)$, $Z = 2Y/3 - 6.5$

Figure 3.

From Figure 1-3, it is clear that the over all performance of the power curves for median tests is excellent for either complete data or censored data with normal distributions, and Figure 1 clearly shows that the over all power by $\hat{m}$ obtained from (MNB1)-(MNB3) is much better than that by $m = \sqrt{n}$.

Figure 4, Figure 5 and Figure 6 compare P_0 and P_b for a 15%-trimmed mean test with complete data, right censored data and doubly censored data, respectively, where F is $N(\mu, 100)$. One may note that the NPMLE for censored data is used in Example 2 to compute the trimmed mean for right censored data or doubly censored data.

P_0: -----; P_b: -----; $X \sim N(\mu, 100)$

Figure 4.

P_0: -----; P_b: -----; $X \sim N(\mu, 100)$, $Y \sim N(6.5, 144)$

Figure 5.

P_0: -----; P_b: ------; $X \sim N(\mu, 100)$, $Y \sim N(6.5, 144)$, $Z = 2Y/3 - 14$

Figure 6.

Figure 4-6 show that the power curves of trimmed mean tests by (MNB1)-(MNB3) perform very well for either complete data or censored data.

In our simulation studies, we also considered distributions with heavy tails such as Cauchy distribution. In Figure 7 and Figure 8, we compare P_0 and P_b for the median test with a complete i.i.d. Cauchy sample of size $n = 400$ and $n = 1000$, respectively.

P_0: -----; P_b: -----; $X \sim$ Cauchy(μ)

Figure 7.

P_0: -----; P_b: -----; $X \sim$ Cauchy(μ)

Figure 8.

Figure 7 and Figure 8 show that the power curves by (MNB1)-(MNB3) also perform very well in the neighborhood of the null hypothesis for Cauchy distribution, but that the m out of n bootstrap test has lower power than the unattainable P_0. We do not have a good heuristic explanation. The corresponding attainable permutation test using symmetry of the null has power close to P_0. One may note that the over all performance of the power curves is better for the larger sample case $n = 1000$ than that for $n = 400$.

5. CONCLUSIONS.

It is shown that generally, the m out of n bootstrap testing procedure is asymptotically consistent for hypothesis testing problems. For a quite general class of testing problems, a method of selecting m for the m out of n bootstrap in hypothesis testing is proposed, and it is shown that the proposed method is asymptotically consistent. With a general formulation, the proposed method applies to complete data or right censored data or doubly censored data. Simulation studies show that the proposed method generally performs very well.

One may note that the proposed method of selecting m is determined by the analysis in certain types of situations, which along with some simulation results, such as Figure 7-8, suggest this is not the last word in choice of m. In particular, we critically use the $\sqrt{m}$-standardization of our test statistics even though we do not use explicit knowledge of its limit law. One may also note that we do not advocate using the m out of n bootstrap method if other easy and computationally efficient methods are available (for instance, resampling n observations from centered residuals may be applied for our examples here with complete i.i.d. samples or right censored samples), rather in cases such as those treated in Bickel and Ren (1995).

APPENDIX

Proof of Theorem 1. (i) It suffices to show that under H_0,

$$C_\alpha^* \xrightarrow{P} C_\alpha^0, \qquad \text{as } n \to \infty. \tag{A.1}$$

Let $Q_{m,F}$ be the d.f. of $T_m = T_m(O_1, \cdots, O_m; F)$ and $Q_{m,n}^*$ be the d.f. of $T_m^* = T_m(O_1^*, \cdots, O_m^*; \hat{F}_n)$, where $O_1^*, \cdots, O_m^*$ is the bootstrap sample from $\hat{F}_n$, the empirical distribution of the observed sample. From assumption (c) and the proof of Theorem 2 by Bickel, Gõetze and van Zwet (1994), we know

$$E_F d(Q_{m,n}^*, Q_{m,F}) = o(1), \tag{A.2}$$

where d is the bounded Lipschitz metric on probability distribution on the range space of T_m. Thus, the assumption (a) implies that under H_0, $d(Q_{m,n}^*, Q_0) \xrightarrow{P} 0$, as $n \to \infty$, where Q_0 is the d.f. of W_0. Hence, the continuity of Q_0 implies (A.1).

(ii) (A.1) and the continuity condition imply that under H_n, $C_\alpha^* \xrightarrow{P} C_\alpha^0$, as $n \to \infty$.

(iii) For any $x \in \mathbf{R}$ and $\rho > 0$, define a continuous and bounded function as below

$$h_\rho(t) = \begin{cases} 1, & t \le x - \rho \\ (x-t)/\rho, & x - \rho < t < x \\ 0, & t \ge x \end{cases}$$

then

$$\begin{aligned} P_n\{T_m^* < x\} &= \int I\{t < x\}\, dQ_{m,n}^*(t) \ge \int h_\rho(t)\, dQ_{m,n}^*(t) \\ &= \Big\{\int h_\rho(t)\, dQ_{m,n}^*(t) - \int h_\rho(t)\, dQ_{m,F}(t)\Big\} + \int h_\rho(t)\, dQ_{m,F}(t) \\ &\ge \Big\{\int h_\rho(t)\, dQ_{m,n}^*(t) - \int h_\rho(t)\, dQ_{m,F}(t)\Big\} + \int I\{t \le x - \rho\}\, dQ_{m,F}(t) \\ &= \Big\{\int h_\rho(t)\, dQ_{m,n}^*(t) - \int h_\rho(t)\, dQ_{m,F}(t)\Big\} + P\{T_m \le x - \rho\}. \end{aligned}$$

From (A.2) and the above, we know that under H_1,

$$\begin{aligned} \alpha = P_n\{T_m^* \ge C_\alpha^* \mid H_1\} &\le P\{T_m > C_\alpha^* - \rho \mid H_1\} + o_p(1) \\ &= P\{(T_m/T_n) > (C_\alpha^* - \rho)/T_n \mid H_1\} + o_p(1). \end{aligned}$$

Thus, assumption (b) implies $C_\alpha^*/T_n \xrightarrow{P} 0$, as $n \to \infty$. Hence,

$$P\{T_n \ge C_\alpha^* \mid H_1\} = P\{1 \ge (C_\alpha^*/T_n) \mid H_1\} \to 1, \qquad \text{as } n \to \infty. \qquad \square$$

Proof of Theorem 2. (i) First, from (3.2) we have that for $m = \sqrt{n}$,

$$\text{(A.3)} \qquad \sup_{-\infty < x < \infty} | P_n\{T_m^* \geq x \,|\, H_0\} - P\{T'_F(G_F) \geq x \,|\, H_0\} | \xrightarrow{P} 0, \qquad \text{as } n \to \infty,$$

which implies that under H_0,

$$\text{(A.4)} \qquad e_\epsilon^* \xrightarrow{P} e_\epsilon^0, \qquad \text{as } n \to \infty.$$

Therefore, we have $\hat{m} = O_p(n)$, as $n \to \infty$.

For the complete i.i.d. sample case, we know that from Shorack and Wellner (1986, page 108-109), if $\hat{m} \to \infty$, then $\sqrt{\hat{m}}\,[\hat{F}_{\hat{m}}^* - \hat{F}_n]$ weakly converges to G_F. For the censored case, from Gu and Zhang (1993), $\sqrt{\hat{m}}\,[\hat{F}_{\hat{m}}^* - \hat{F}_n]$ can be shown to be equivalent to a linear operator of an empirical processes (we skip the technical discussion for this part of the proof, whose idea can be found in Bickel and Ren, 1995), thus it also converges weakly. Hence, from the weak convergence of $\sqrt{n}\,[\hat{F}_n - F]$ and $\hat{m} = O(n)$, we know that

$$\sqrt{\hat{m}}\,[\hat{F}_{\hat{m}}^* - F] = \sqrt{\hat{m}}\,[\hat{F}_{\hat{m}}^* - \hat{F}_n] + \sqrt{\hat{m}/n}\left(\sqrt{n}\,[\hat{F}_n - F]\right)$$

is tight. Therefore, from Fernholz (1983, Chapter 4) and from (2.7), we have that under H_0, for $\hat{m} \to \infty$,

$$T_{\hat{m}}^* = \sqrt{\hat{m}}\,T(\hat{F}_{\hat{m}}^*) = \sqrt{\hat{m}}\,T(F) + T'_F(\sqrt{\hat{m}}\,(\hat{F}_{\hat{m}}^* - F)) + o_p(1)$$

$$\text{(A.5)} \qquad = T'_F(\sqrt{\hat{m}}\,(\hat{F}_{\hat{m}}^* - \hat{F}_n)) + \sqrt{\hat{m}/n}\,T_n + o_p(1) \overset{D}{\approx} T'_F(G_F) + \hat{d}_n$$

where $\hat{d}_n = \sqrt{\hat{m}/n}\,T_n$. Note that (A.5) implies that for $n \to \infty$,

$$\sup_{-\infty < x < \infty} | P_n\{T_{\hat{m}}^* - \hat{d}_n \geq x \,|\, H_0\} - P\{T'_F(G_F) \geq x \,|\, H_0\} | \xrightarrow{P} 0,$$

thus

$$\text{(A.6)} \qquad \sup_{-\infty < x < \infty} | P_n\{T_{\hat{m}}^* \geq x \,|\, H_0\} - P\{T'_F(G_F) \geq x - \hat{d}_n \,|\, H_0\} | \xrightarrow{P} 0.$$

Hence, we have

$$| \alpha - P\{T'_F(G_F) \geq \hat{C}_\alpha^* - \hat{d}_n \,|\, H_0\} | \xrightarrow{P} 0, \qquad \text{as } n \to \infty,$$

and the continuity of the distribution of $T'_F(G_F)$ gives

$$| C_\alpha^0 - \hat{C}_\alpha^* + \hat{d}_n | \xrightarrow{P} 0, \qquad \text{as } n \to \infty.$$

Since $|\hat{d}_n| \leq e_\epsilon^*$, from (A.4) we have that for $n \to \infty$,

$$\hat{C}_\alpha^* = C_\alpha^0 + \hat{d}_n + o_p(1) \leq C_\alpha^0 + e_\epsilon^* + o_p(1) = C_\alpha^0 + e_\epsilon^0 + o_p(1) \leq C_{\alpha-\epsilon}^0 + o_p(1),$$

$$\hat{C}_\alpha^* = C_\alpha^0 + \hat{d}_n + o_p(1) \geq C_\alpha^0 - e_\epsilon^* + o_p(1) = C_\alpha^0 - e_\epsilon^0 + o_p(1) \geq C_{\alpha+\epsilon}^0 + o_p(1).$$

Therefore, (3.5) follows from

$$P\{T_n \geq \hat{C}_\alpha^* \,|\, H_0\} \leq P\{T_n \geq C_{\alpha+\epsilon}^0 + o_p(1) \,|\, H_0\} \to P\{T'_F(G_F) \geq C_{\alpha+\epsilon}^0 \,|\, H_0\} = \alpha + \epsilon,$$

$$P\{T_n \geq \hat{C}_\alpha^* \,|\, H_0\} \geq P\{T_n \geq C_{\alpha-\epsilon}^0 + o_p(1) \,|\, H_0\} \to P\{T'_F(G_F) \geq C_{\alpha-\epsilon}^0 \,|\, H_0\} = \alpha - \epsilon.$$

(ii) First, from the contiguity condition and (A.4), we have that under H_n

$$e_\epsilon^* \xrightarrow{P} e_\epsilon^0, \quad \text{as } n \to \infty. \tag{A.7}$$

From the contiguity assumption, we have the weak convergence of $\sqrt{n}[\hat{F}_n - F_0]$ (see Shorack and Wellner, 1986, page 108-109 for complete data case; see discussion above (3.9) of Bickel and Ren, 1995, for the censored case), thus from Theorem II.8.1 (Andersen, Borgan, Gill and Keiding, 1993) we have

$$\sqrt{n}T(\hat{F}_n) = \sqrt{n}T(F_0) + T'_{F_0}(\sqrt{n}[\hat{F}_n - F_0]) + o_p(1), \qquad \text{as } n \to \infty, \tag{A.8}$$

$$\sqrt{n}T(F_n) = \sqrt{n}T(F_0) + T'_{F_0}(\sqrt{n}[F_n - F_0]) + o_p(1), \qquad \text{as } n \to \infty. \tag{A.9}$$

Note that (A.8) and (A.9) imply that under H_n, $T_n \xrightarrow{D} T'_{F_0}(G_{F_0}) + \delta$, as $n \to \infty$. Thus (A.7) implies $\hat{m} = O_p(n)$, as $n \to \infty$. From the tightness of $\sqrt{\hat{m}}[\hat{F}^*_{\hat{m}} - F_0]$ and from (A.8), we have that under H_n, as $n \to \infty$

$$\begin{aligned} T^*_{\hat{m}} &= \sqrt{\hat{m}}T(\hat{F}^*_{\hat{m}}) = \sqrt{\hat{m}}T(F_0) + T'_{F_0}(\sqrt{\hat{m}}(\hat{F}^*_{\hat{m}} - F_0)) + o_p(1) \\ &= T'_{F_0}(\sqrt{\hat{m}}[\hat{F}^*_{\hat{m}} - \hat{F}_n]) + \sqrt{\hat{m}/n}T_n + o_p(1) \overset{D}{\approx} T'_{F_0}(G_{F_0}) + \hat{d}_n. \end{aligned} \tag{A.10}$$

Hence, from (A.7) and (A.10) the rest of the proof follows line by line of the proof of (i).

(iii) Note that (2.7) and (2.8) hold in general for any F. Hence, we have $e^*_\epsilon = O_p(1)$ under H_1 and $T_n = O_p(\sqrt{n})$ under H_1. The proof follows from $\hat{m} = O_p(1)$ under H_1 and $\hat{C}^*_\alpha = O_p(1)$ under H_1. □

REFERENCES

Andersen, P.K., Borgan, Ø, Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag.

Bickel, P.J., Gõetze, F. and Van Zwet, W.R. (1994). Resampling fewer than n observations: gains, losses and remedies for losses. Preprint 94-084, Sonderforschungsbereich 343, Universität Bielefeld.

Bickel, P.J. and Ren, J. (1995). The m out of n bootstrap and goodness of fit tests with doubly censored data. *Robust Statistics, Data Analysis, and Computer Intensive Methods.* Lecture Notes in Statistics, Springer-Verlag. 35-47.

Fernholz, L.T. (1983). *Von Mises Calculus for Statistical Functional.* Lecture Notes in Statistics 19. Springer, New York. [1.3f, 2.1b, 2.3a].

Freedman, D.F. (1981). Bootstrapping regression models. *Ann. Statist.* 12, 1218-1228.

Gill, R.D. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* 11, 49-58.

Giné, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *Ann. Prob.* 18, 852-869.

Gu, M.G. and Zhang, C.H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.* Vol. 21, No. 2, 611-624.

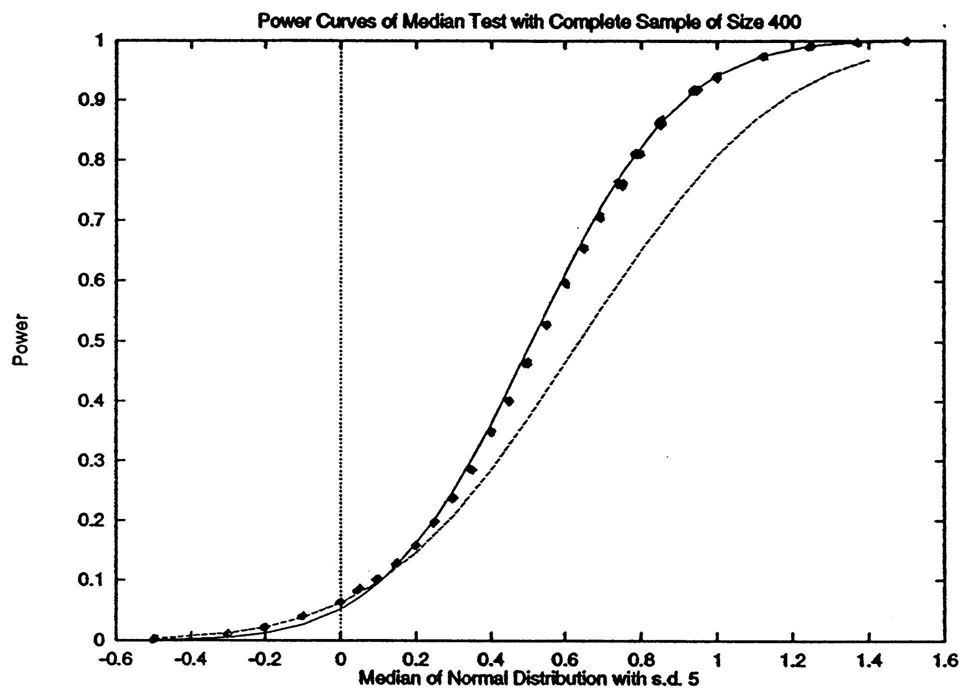
Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* 21, 1926-1947.

Mammen, E. (1992). *When Does Bootstrap Work?* Springer Verlag, New York.

Mykland, P. and Ren, J. (1994). Algorithms for computing the self-consistent and maximum likelihood estimators with doubly censored data. *Ann. Statist.* (To appear).

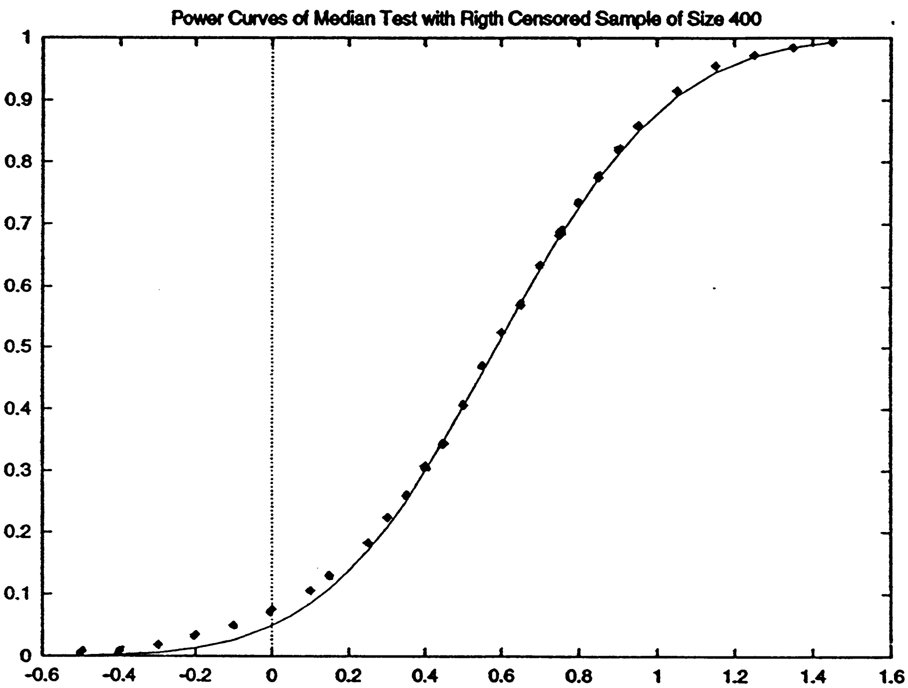
Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley & Sons, INC.

Turnbull, B.W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* 69, 169-173.



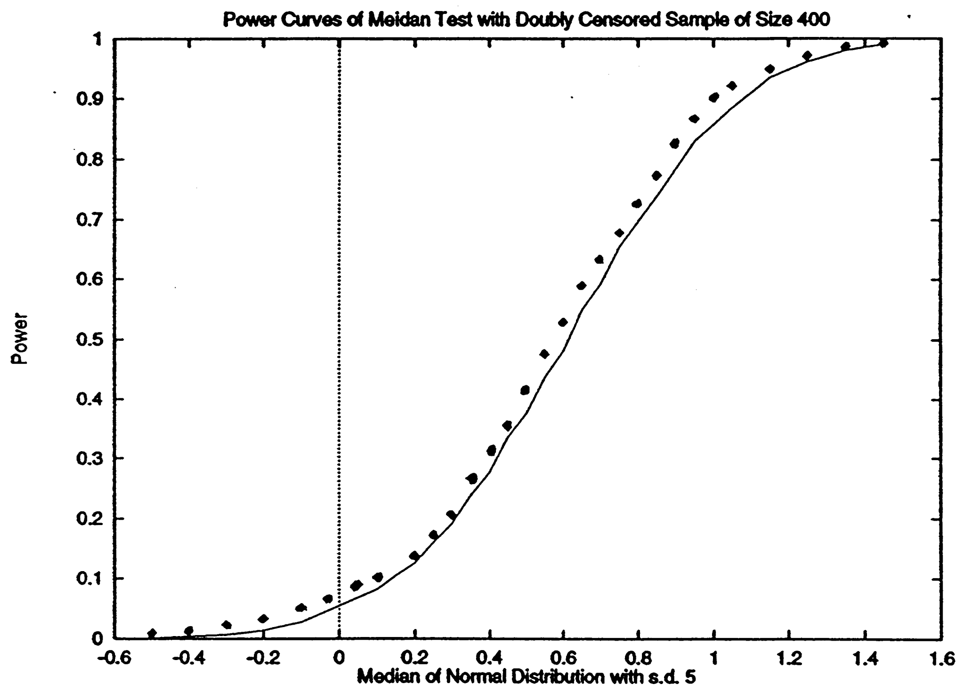
P_0: ——; P_b: • • •; P_2: -----; $X \sim \mathrm{N}(\mu, 25)$

Figure 1.



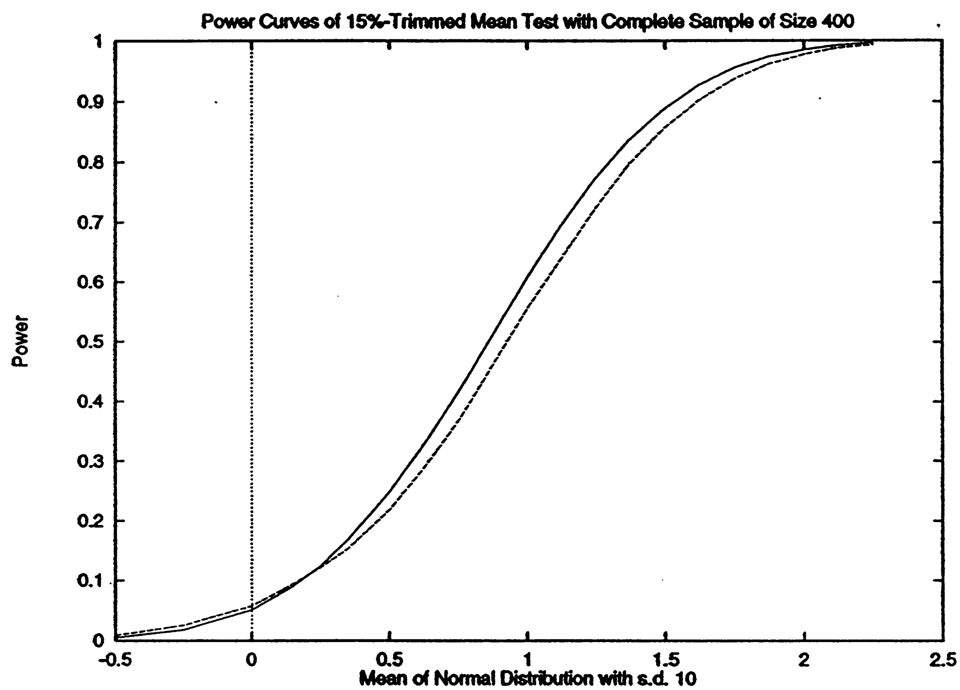
P_0: ——; P_b: • • •; $X \sim \mathrm{N}(\mu, 25)$, $Y \sim \mathrm{N}(3, 36)$

Figure 2.



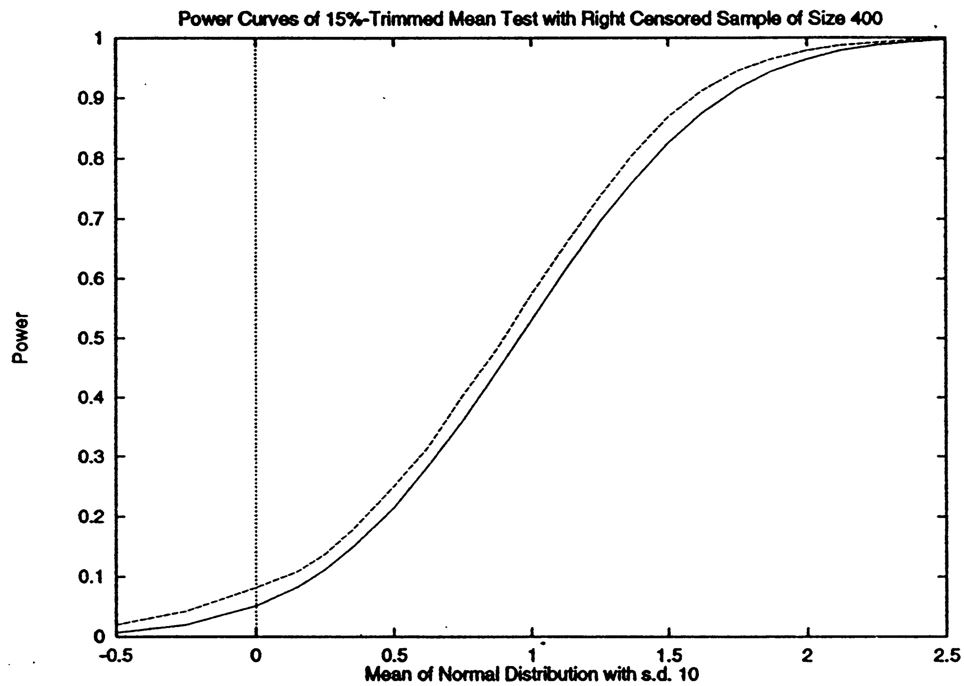
P_0: ——; P_b: • • •; $X \sim \mathrm{N}(\mu, 25)$, $Y \sim \mathrm{N}(3, 36)$, $Z = 2Y/3 - 6.5$

Figure 3.



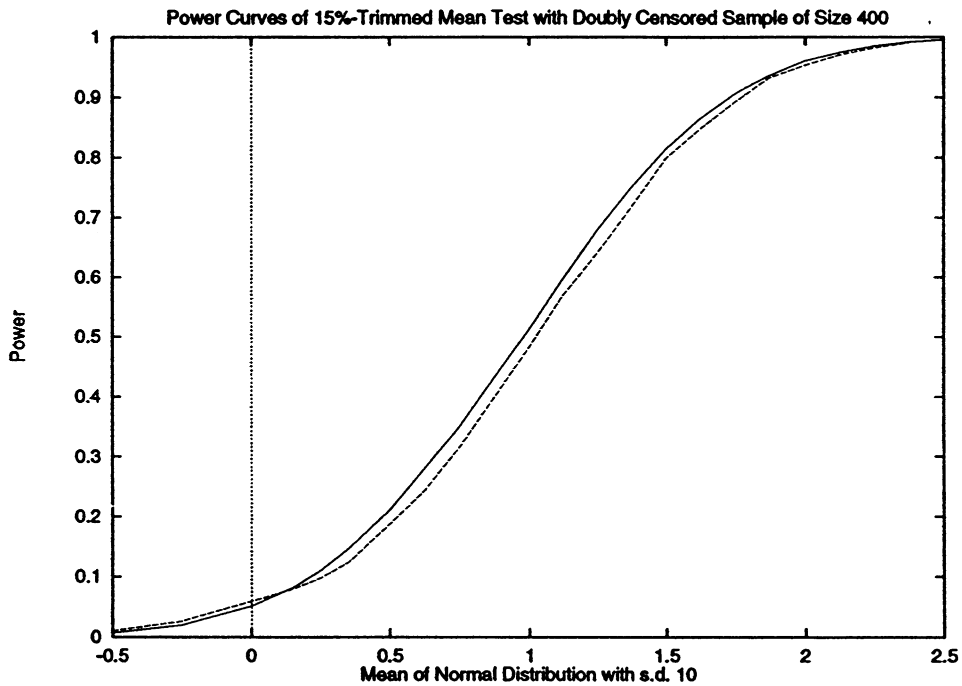
P_0: ——; P_b: ----; $X \sim \mathrm{N}(\mu, 100)$

Figure 4.



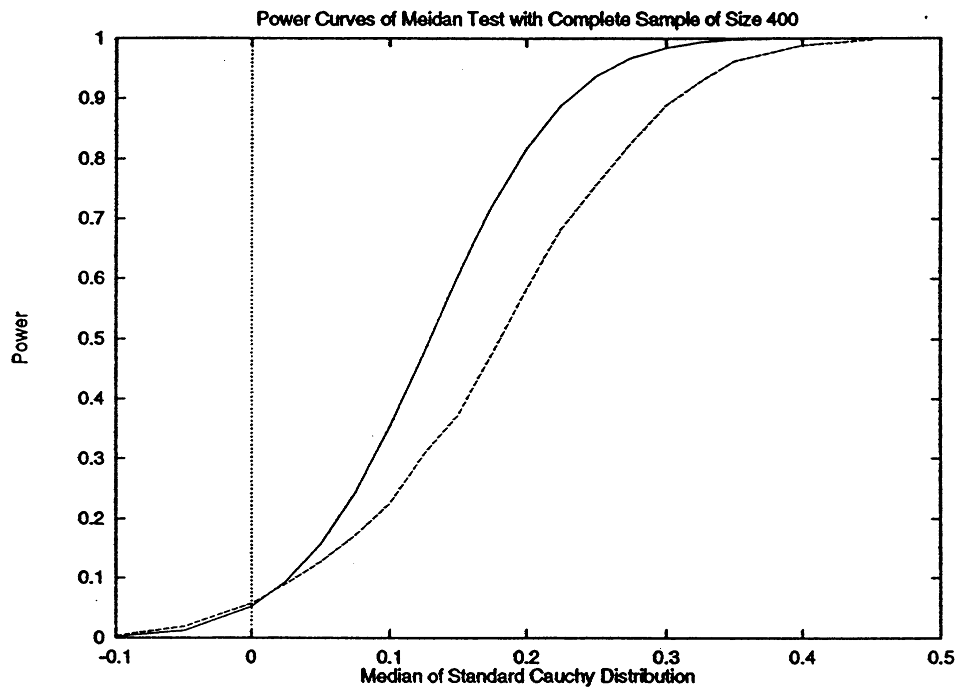
P_0: ——; P_b: -----; $X \sim \mathrm{N}(\mu, 100)$, $Y \sim \mathrm{N}(6.5, 144)$

Figure 5.



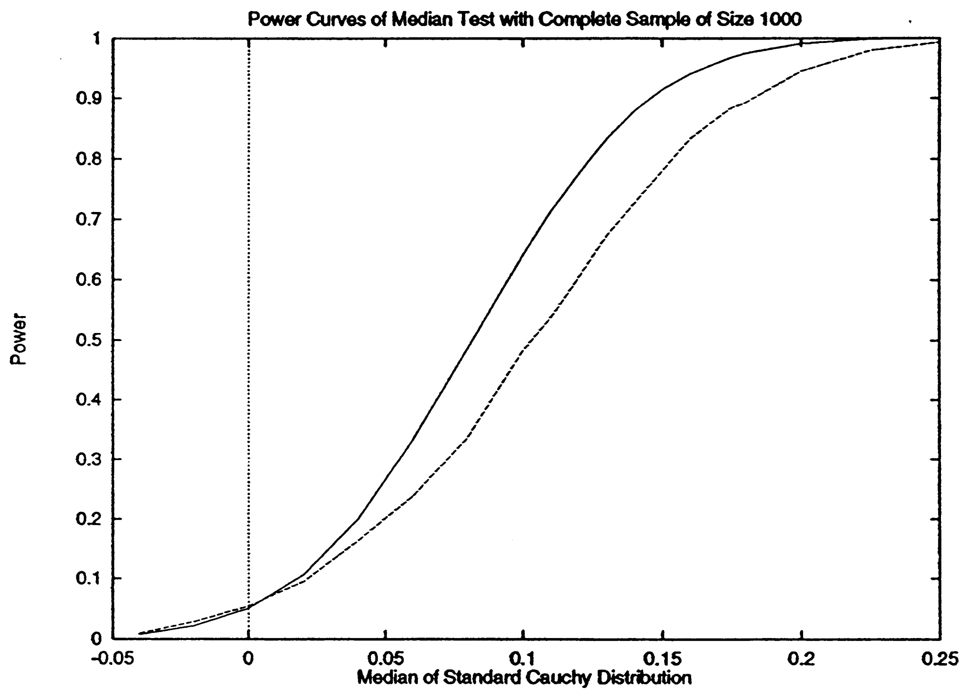
P_0: ——; P_b: ------; $X \sim \mathrm{N}(\mu, 100)$, $Y \sim \mathrm{N}(6.5, 144), Z = 2Y/3 - 14$

Figure 6.



P_0: —— ; P_b: ----- ; $X \sim$ Cauchy(μ)

Figure 7.



P_0: ——; P_b: - - - -; $X \sim \text{Cauchy}(\mu)$

Figure 8.