The Annals of Statistics
1998, Vol. 26, No. 5, 1719–1732

BREAKDOWN THEORY FOR BOOTSTRAP QUANTILES[1]

BY KESAR SINGH

Rutgers University

A general formula for computing the breakdown point (in robustness) for the tth bootstrap quantile of a statistic T_n is obtained. The answer depends on t and the breakdown point of T_n. Since the bootstrap quantiles are vital ingredients of bootstrap confidence intervals, the theory has implications pertaining to robustness of bootstrap confidence intervals. For certain L and M estimators, a robustification of bootstrap is suggested via the notion of Winsorization.

1. Introduction. Consider the 10% trimmed mean $T(0.1)$ (i.e., 5% trimming each side) on a random sample of size $n = 20$. If there is one outlier in the upper side, that is, $X_{(n)}$ is extraordinarily large, $T_{0.1}$ stays unaffected due to the trimming. Now, suppose a bootstrap sample of size 20 is drawn from this sample. The outlier $X_{(n)}$ could appear one time, two times or, in the most extreme case, 20 times in the bootstrap sample. Consider the resampling distribution of the bootstrap trimmed mean $T^*(0.1)$. If $X_{(n)}$ appears only one time in the bootstrap sample, $T^*(0.1)$ is free of it. If it appears more than one time, $T^*(0.1)$ will be influenced by the outlier. The chances for the event that $T^*(0.1)$ is free of the outlier is

$$P(\mathrm{Bin}(20, 0.05) \le 1) = p_0 \quad (\text{say}),$$

which is about 73.6%. This means that if $X_{(n)}$ converged to $+\infty$, $100\,(1 - p_0)\%$ of all the $T^*(0.1)$ will converge to $+\infty$ as a consequence. In other words, the bootstrap quantiles Q_t^* of $T^*(0.1)$, where t ranges from 0 to 1, will go to ∞ for all $t > p_0 \approx 0.736$. In the terminology of the celebrated concept of breakdown in robustness, $T(0.1)$ has upper breakdown (UB) $= 0.1$, meaning that at least 10% (i.e., 2 out of 20) of the data have to go to ∞ in order to carry $T(0.1)$ to ∞. Asymptotically, though, this breakdown is 5%. The above reasoning takes us to the following conclusions:

1. The UB for Q_t^* is $= 0.05$ for $t > p_0$.
2. The UB for Q_t^* is ≥ 0.1 for $t \le p_0$.

Thus the lower bootstrap quantiles are more robust than the upper ones in terms of going to ∞. Of course a parallel reasoning can be given for the lower breakdown (LB).

Received November 1996; revised March 1998.

[1] Supported in part by NSF grants.

AMS 1991 *subject classifications*. 62G09, 62G15.

Key words and phrases. Bootstrap, quantiles, breakdown in robustness, L and M estimators, Winsorization.

There is a miniscule, though positive, probability that $T^*(0.1) = X_{(n)}$. As a consequence of that, $\widehat{\mathrm{SE}}^*$, that is, the bootstrap-based estimated standard error, has $\mathrm{UB} = (1/n)$ (i.e., 0 as $n \to \infty$) and it remains so for all the trimming proportions as well as for the sample median. A recent article by Stromberg (1997) presents this phenomenon and some resampling-based robust estimators of SE in a multivariate setting.

In this paper, we shall first present the foregoing discussion in the form of a general theorem. Then we present a device to boost the breakdown of bootstrap quantiles as well as that of $\widehat{\mathrm{SE}}^*$ while retaining the same asymptotic distribution for certain robust statistics. The device is simply to Winsorize the data, with a suitable proportion, prior to bootstrapping. Winsorization at a level 2α, for some $0 < \alpha < \frac{1}{2}$, replaces upper $100\alpha\%$ data by the $(1-\alpha)$th sample quantile and the lower $100\alpha\%$ data by the αth sample quantile. In modern statistics, Winsorization is usually cited in the context of the S.E. of a trimmed mean [see page 366 of Lehmann (1983)]. In some cases, trimming the data prior to bootstrapping works as well (see Remark 4 in Section 4). Formal results on improved breakdown and unaltered asymptotics are presented for trimmed type L-statistics, the sample median and M-estimators [see Huber (1964), (1981)] with breakdown approximately $\frac{1}{2}$. A version of scale-invariant M-estimators is also considered. We later remark on the normalized and Studentized bootstrap statistics. The case of multivariate estimators is also discussed briefly.

2. The breakdown formula. Let T_n be a statistic based on a random sample of size n. Let b denote its UB, that is, nb is the smallest number of observations that needs to go to $\pm\infty$ in order to force T_n to go to $+\infty$. Here nb is an integer between 1 and n. It is assumed here that the minimum number of outliers which cause UB (i.e., $T_n \to \infty$) are either all in the upper side of the sample or all in the lower side, but not some in the upper side and some in the lower side. Of course, this is almost always the case, though counter-examples can be constructed. For t between 0 and 1, let Q_t^* denote the tth quantile of the bootstrap distribution of T_n^*, that is,

$$Q_t^* = \min\{x\colon P_B(T_n^* \le x) \ge t\}.$$

The following theorem states a formula for b_t, the UB for Q_t^*.

THEOREM 1. *The UB b_t for Q_t^* is the* min p, *with np as an integer between* 1 *and n, such that*

$$P(\mathrm{Bin}(n,p) \ge nb) \ge 1-t$$

(b is the UB of T_n).

Let us fix a t and ponder the UB of Q_t^*. As b increases, $P(\mathrm{Bin}(n,p) \ge nb)$ decreases, when p is held fixed. This entails that it would take larger values of p to make this binomial probability exceed $(1-t)$. Thus, for a fixed t, higher UB of T_n means higher UB of Q_t^*. Now, let us fix a statistic T_n with

UB $= b$ and let t move upwardly toward 1. It will take smaller values of p to make $P(\text{Bin}(n, p) \geq nb)$ exceed $1 - t$. This means that UB of Q_t^* will decrease as t moves outwardly toward 1. The conclusion thus is that it pays to start out with a robust T_n at any level of t. Furthermore, given a T_n, it pays to stay away from extreme quantiles, that is, t near 0 and 1, in choosing Q_t^*-based inferences, considering the lower breakdown LB of Q_t^* also.

With the sample size $n = 10$, Table 1 contains the values for nb_t with the choices of nb as 2, 3, 5 and t as 0.5, 0.75, 0.9, 0.99. As the table displays, for a fixed b, b_t decreases as t increases, and for a fixed t, b_t increases as b increases. These phenomena perfectly agree with the above discussion.

PROOF OF THEOREM 1. In order to prevent the breakdown of a T_n^*, the corresponding bootstrap sample should have the number of upper outliers less than nb. If Q_t^* has to break down, it means that the proportion of nonbreakdown class of T_n^* is less than t. This implies that b_t is equal to the min p, with np as an integer between 1 and n, such that

$$P(\text{Bin}(n, p) < bn) < t;$$

this is equivalent to the statement of the theorem. Similar reasoning is given when T_n goes to $+\infty$ due to lower outliers. □

When T_n is a scale statistic, like the S.D. and the interquartile range, the UB can occur due to nb^* upper outliers or nb^{**} lower outliers. In such a case, if $b = \min(b^*, b^{**})$ and $b^* \neq b^{**}$, one appeals to the fact that b_t given by Theorem 1 is a monotonic function of b to make the theory work.

An asymptotic formula. Let us recall that b and b_t are UB (upper breakdown) for T_n and its tth bootstrap quantile, respectively. For any fixed t in $(0, 1)$, the following expansion holds:

$$b_t = b - \frac{z_t\sqrt{b(1-b)}}{\sqrt{n}} + O\left(\frac{1}{n}\right), \tag{2.1}$$

where $\Phi(z_t) = t$. A notable feature in this expansion is that the lead term in the right side is free from t. The second-order term is monotonically decreasing in t. Just the opposite will be found in the case of LB (the lower breakdown).

TABLE 1
(nb_t)

nb \ t	**0.5**	**0.75**	**0.9**	**0.99**
2	2	1	1	1
3	3	2	2	1
5	5	4	3	2

PROOF. To prove the expansion, one takes

$$p = b - n^{-1/2}[b(1-b)]^{1/2} z_t + cn^{-1}$$

and obtains a Berry–Esseen type bound for the normal approximation of the binomial c.d.f., uniform in c belonging to a fixed compact set. For all c and n large enough, $P(\mathrm{Bin}(n, p) < nb)$ is less than t, and for all c small enough and n large enough, this binomial probability is greater than t. The expansion thus follows. □

3. Winsorizing prior to bootstrapping. In robust estimation, typically the data values in the exterior have limited or no influence on the estimator. Therefore, some of the exterior data can be altered in order to boost the breakdown and hence the robustness of bootstrap quantiles. This can be done while keeping the bootstrap distribution more or less the same. This robustification of bootstrap will be demonstrated in this section on certain robust L and M estimators.

Let the order statistics of the original data be denoted by $X_{(1)}, \dots, X_{(n)}$. The empirical and quantile process of the X-data is defined as follows:

$$F_n(x) = (\text{number of } X_i \le x)/n, \qquad -\infty < x < \infty;$$

$$F_n^{-1}(t) = \min\{x\colon F_n(x) \ge t\}, \qquad 0 \le t \le 1.$$

For some fixed α between 0 and $\frac{1}{2}$, define the 2α-trimmed mean

$$L_n = \frac{1}{1-2\alpha}\int_{\alpha}^{1-\alpha} F_n^{-1}(t)\,dt.$$

Thus in essence, one is trimming (approximately) $100\alpha\%$ data from each side and computing the mean on the rest. For some β between 0 and $\frac{1}{2}$, let us define the Winsorization of the β-fraction of the X-data from each end. Let $l = [n\beta]$ = largest integer less than or equal to $n\beta$. Let

$$X_i^* = \begin{cases} X_{(l+1)}, & \text{if } X_i \le X_{(l)}, \\ X_{(n-l)}, & \text{if } X_i \ge X_{(n-l+1)}, \\ X_i, & \text{otherwise.} \end{cases}$$

The X_i^* are the Winsorized data.

Specifically for bootstrapping a trimmed mean L_n defined above, the proposal is to fix some α, β such that $0 < \beta \le \alpha < \frac{1}{2}$ and resample from the X^*-data instead of the X-data. In effect, this resampling is equivalent to the following: let $Y_1, Y_2, \dots, Y_n$ be random draws with replacement from the original X-data. Define (recall that $l = [n\beta]$)

$$Y_i^* = \begin{cases} X_{(l+1)}, & \text{if } Y_i \le X_{(l)}, \\ X_{(n-l)}, & \text{if } Y_i \ge X_{(n-l+1)}, \\ Y_i, & \text{otherwise.} \end{cases}$$

It is easily seen that Y_i^* are random draws from $\{X_i^*\}$. Let us further define

$$L_n^* = \frac{1}{1-2\alpha}\int_\alpha^{1-\alpha} G_n^{-1}(t)\,dt,$$

$$L_n^{**} = \frac{1}{1-2\alpha}\int_\alpha^{1-\alpha} H_n^{-1}(t)\,dt,$$

where $G_n(\cdot)$ and $H_n(\cdot)$ are the empirical c.d.f. of the bootstrap data $\{Y_i\}$ and $\{Y_i^*\}$, respectively.

Let b_t and b_t^* denote the UB (upper breakdown) of the tth quantile of L_n^* and L_n^{**} respectively, under the bootstrap distribution. Then one has the following theorem.

THEOREM 2. (a) $b_t^* = \max(\beta', b_t)$, *where* $\beta' = ([n\beta]+1)/n$. *Thus* $b_t^* \geq \beta$. *A similar result holds for the corresponding LB.*

(b) *If* $\beta < \alpha$, *the bootstrap probability of the event* $L_n^* \neq L_n^{**}$ *goes to* 0, *exponentially fast as* $n \to \infty$, *a.s.*

(c) *Assume that F has a density, bounded below and above, in neighborhoods of* $F^{-1}(\alpha)$ *as well as* $F^{-1}(1-\alpha)$. *If* $\alpha = \beta$, *one has*

$$L_n^* - L_n^{**} = O_p(n^{-1}\log n)$$

in bootstrap probability, a.s.

The proof of the theorem is deferred to the Appendix.

The UB of L_n itself is $([\alpha n]+1)/n$. The UB of the tth quantile of L_n^* can be as low as $1/n$, for t near 1. Thus $b_t^* \geq \beta$ for all t is a genuine improvement in the robustness. Parts (b) and (c) essentially assert that $\sqrt{n}(L_n^* - L_n^{**}) \to 0$ in bootstrap probability. Thus, normalized L_n^* and L_n^{**} will have the same limiting distribution. Hence, L_n^{**} can replace L_n^*.

The *sample median* is the limiting case of L_n defined above as $\alpha \to \frac{1}{2}$ and it is not covered by Theorem 2. However, the robustification by Winsorizing prior to the bootstrap works just fine, as in the case of L_n. In the median case UB and LB both are $n/2$ when n is even, and $(n+1)/2$ when n is odd. Any Winsorizing factor $\beta < \frac{1}{2}$ is acceptable in theory for the median case. In practice, though, β should be chosen well below $\frac{1}{2}$, say less than or equal to $\frac{1}{4}$. This theory for trimmed means extends, without requiring any additional efforts, to the following class of scale functionals:

$$\begin{aligned} W_n &= \int_{1/2}^{1-\alpha} F_n^{-1}(t)\,dt - \int_\alpha^{1/2} F_n^{-1}(t)\,dt \\ &= \int_{1/2}^{1-\alpha}\left[F_n^{-1}(t) - F_n^{-1}(1-t)\right]dt. \end{aligned}$$

The robustification device stays the same. For the scale functional interquartile range,

$$IQR = F_n^{-1}\left(\tfrac{3}{4}\right) - F_n^{-1}\left(\tfrac{1}{4}\right),$$

one could Winsorize at some level $\beta < \frac{1}{4}$ for the robustification purpose without affecting the asymptotics. In the foregoing discussion, one could use a general weight $w(t)$ such that $\int_{\alpha}^{1-\alpha} w(t)\,dt = 1$ in the location case and $\int_{\alpha}^{1-\alpha} w(t)\,dt = 0$ in the scale case.

Now, we turn to certain m-estimators; specifically, the class considered is defined as follows: let g be a monotonically increasing function from $\mathbb{R} \to \mathbb{R}$, such that $g(-x) = -g(x)$. For a positive constant c, define $g_c(\cdot)$ as

$$g_c(x) = \begin{cases} g(-c), & \text{if } x \le -c, \\ g(x), & \text{if } -c \le x \le c, \\ g(c), & \text{if } x \ge c. \end{cases}$$

The m-estimator θ_n is defined as the unique solution of the equation

$$\sum_1^n g_c(X_i - \theta_n) = 0.$$

The corresponding parameter θ is a solution of

$$E_F g_c(X - \theta) = 0,$$

where F is the underlying population. The solution θ of the above equation is assumed to be unique [see Huber (1964) for details on m-estimation].

Elementary arguments show that

$$b = \text{UB of } \theta_n = \begin{cases} \dfrac{n}{2} + 1, & \text{when } n \text{ is even}, \\ \dfrac{n+1}{2}, & \text{when } n \text{ is odd}. \end{cases}$$

The same hold for LB and thus the breakdown of θ_n is $\frac{1}{2}$, in limit. However, for the reasons explained earlier, the breakdown of its bootstrap quantiles can be as low as $1/n$. Winsorization is proposed here, too, in order to raise the breakdown of the corresponding Q_t^*, the tth bootstrap quantile of θ_n.

Let $Y_1, \dots, Y_n$ be a bootstrap sample. For a positive $d \ge c$, let us define

$$Y_i^* = \begin{cases} \theta_n - d, & \text{if } Y_i \le \theta_n - d, \\ \theta_n + d, & \text{if } Y_i \ge \theta_n + d, \\ Y_i, & \text{otherwise}. \end{cases}$$

Let θ_n^* and θ_n^{**} be the m-estimators based on the samples $\{Y_i\}$ and $\{Y_i^*\}$, respectively. We assume the following regularity conditions:

1. The function g has a bounded continuous derivative on the interval $[-c - \varepsilon, c + \varepsilon]$, for $\varepsilon > 0$.
2. The c.d.f. F has a nonzero, bounded density near the points $\theta - c$ and $\theta + c$.

THEOREM 3. (a) *If b_t and b_t^* denote the UB for the tth bootstrap quantiles of θ_n^* and θ_n^{**}, respectively, then*

$$b_t^* = \max(b, b_t),$$

where b is the UB of θ_n and b_t is given by Theorem 1. *Thus $b_t^* \geq \frac{1}{2}$. A similar result holds for LB.*

(b) *The bootstrap probability of $\{\theta_n^* \neq \theta_n^{**}\}$ decays exponentially fast if $d > c$, a.s.*

(c) *In the case $c = d$, a shrinkage occurs which causes inconsistency. Assume in addition that the population is symmetric. Let $\lambda_1 = F(\theta - c)g'(-c) = [1 - F(\theta + c)]g'(c)$ and $\lambda_2 = \int_{\theta-c}^{\theta+c} g'(x)\,dF(x)$. Then*

$$\theta_n^{**} - \theta_n = (\theta_n^* - \theta_n)\big(\lambda + o_p(1)\big)$$

in bootstrap probability (*a.s.*), *where $\lambda = \lambda_2/(\lambda_1 + \lambda_2)$. If $F(\theta - c) = 1 - F(\theta + c) = \alpha$ and $g(x) = x$, then $\lambda = 1 - \alpha/(1 - \alpha)$, which is approximately equal to* 1 *if α is very small.*

The proofs are in the Appendix.

We consider now the scale-invariant version of the m-estimator, when the scale is estimated separately. See Carroll (1978) for asymptotics on such m-estimators. Let W_n be a robust scale functional with UB $= \rho = \rho(n)$ satisfying $1 \leq n\rho \leq (n/2) + 1$. Consider the solution θ_n of the equation

$$\sum_{i=1}^{n} g_c\left(\frac{X_i - \theta_n}{W_n}\right) = 0. \tag{3.1}$$

Clearly θ_n is scale invariant if W_n has proper invariance. We impose a fairly nonrestrictive condition on W_n:

$$\frac{W_n}{R_n} \leq \text{a constant}, \tag{3.2}$$

where R_n denotes the range of the data $\{X_i\}$.

Under condition (3.2), we show now that the UB of θ_n, defined by (3.1), is ρ. Let us write $\theta_n = W_n \xi_n$ where ξ_n is the solution of

$$\sum_{i=1}^{n} g_c\left(\frac{X_i}{W_n} - \xi_n\right) = 0. \tag{3.3}$$

Under condition (3.2), the solution ξ_n of (3.3) stays bounded away from 0, in the positive side of the real line, as the upper $n\rho$ data tend to $+\infty$. Consequently, $\theta_n = W_n \xi_n$ moves towards $+\infty$. One can argue easily that $n\rho$ is the minimum number of data that can move θ_n to $+\infty$. Thus the UB of θ_n is ρ and Theorem 1 is applicable with $T_n = \theta_n$ and $b = \rho$. All the scale functionals of the form

$$W_n = \int F_n^{-1}(t)\,dB(t)$$

with $\int_0^1 dB(t) = 0$ satisfy condition (3.2). So does the popular scale functional

$$W_n = \text{med}\{|X_i - M_n|\},$$

where M_n is the median of the original data. This W_n has limiting breakdown $\frac{1}{2}$. This particular choice of the scale functional is one of the most suitable choices for our purpose.

It should be mentioned here that while studying the LB of θ_n defined by (3.1), one has to consider the minimum number of lower data that must go to $-\infty$ in order to send W_n to $+\infty$.

The recommended Winsorization in the case of scale-invariant m-estimators is described as follows:

$$Y_i^* = \begin{cases} Y_i, & \text{if } \left|\dfrac{Y_i - \theta_n}{W_n}\right| \le d, \\ \theta_n + W_n d, & \text{if } \dfrac{Y_i - \theta_n}{W_n} > d, \\ \theta_n - W_n d, & \text{if } \dfrac{Y_i - \theta_n}{W_n} < -d, \end{cases}$$

where $d > c$ [see the definition of the function $g_c(\cdot)$]. In order to obtain an exponential bound for $P(\theta_n^* \ne \theta_n^{**})$, one would need a large deviation-type bound on the scale functional W_n. Such a bound typically holds for robust scale functionals.

The improved UB for the Q_t^* of the scale-invariant m-estimator θ_n is given by

$$\max(\rho, b_t).$$

Here b_t denotes the UB of Q_t^* of θ_n prior to the Winsorization, as given by Theorem 1. The inconsistency in the case $c = d$ is quite general and hence it is recommended that d is kept greater than c (perhaps $d \approx 1.5c$).

4. Miscellaneous remarks. Let us recall that Q_t^* denotes the tth bootstrap quantile of a statistic T_n, UB$(\cdot)$ and LB$(\cdot)$ are the upper and lower breakdown of the statistics within parentheses.

REMARK 1. *Some implications*. Consider the one-sided, percentile-method based, confidence intervals of the type $[Q_t^*, \infty)$ or $(-\infty, Q_t^*]$. A breakdown of Q_t^* in either direction could be regarded as the breakdown of such an interval. Thus, one could utilize Theorem 1 to compute the breakdown of a one-sided C.I. Consider now an interval of the type $[Q_\alpha^*, Q_{1-\alpha}^*]$, $0 < \alpha < \frac{1}{2}$. Lower breakdown of Q_α^* or the upper breakdown of $Q_{1-\alpha}^*$ could render this two-sided interval useless. Thus the breakdown of the latter interval is given by

$$\min\{\mathrm{LB}(Q_\alpha^*), \mathrm{UB}(Q_{1-\alpha}^*)\}. \tag{4.1}$$

A robust measure of scale of the sampling distribution of T_n, could be defined as $[Q_{1-\alpha}^* - Q_\alpha^*]$, $0 < \alpha < \frac{1}{2}$. One could similarly argue that (4.1) can be regarded as a breakdown of this scale statistics, too. The most commonly used scale statistic is

$$E_B(T_n^* - T_n)^2, \tag{4.2}$$

where T_n^* is the statistic T_n, computed on a bootstrap sample. Since (4.2) involves the most extreme quantile Q_t^*, its breakdown is usually $1/n$, even if that of T_n is $\frac{1}{2}$ [see Stromberg (1997)]. If the breakdown of Q_t^*, for all t and that of T_n is greater than or equal to β, then the same holds for (4.2). Thus the Winsorization techniques of the earlier section can be called upon to raise the breakdown of (4.2), while retaining its consistency.

REMARK 2. *Normalized statistics*. Consider the normalized statistic $\sqrt{n}(T_n - T_F)$. The corresponding bootstrap statistic is $\sqrt{n}(T_n^* - T_n)$. The related bootstrap quantiles are $\sqrt{n}(Q_t^* - T_n)$. We observe here that

$$\text{UB}\big(\sqrt{n}\,(Q_t^* - T_n)\big) \geq \min\{\text{UB}(Q_t^*), \text{LB}(T_n)\} \tag{4.3}$$

and the analogous inequality holds in the LB case. The observation (4.3) is based on the reasoning that even if Q_t^* is dragged out to ∞, $Q_t^* - T_n$ may refuse to follow suit (i.e., when T_n itself goes to ∞).

REMARK 3. *Studentized statistics*. A Studentized statistic is of the form $t_n = (T_n - T_F)/\widehat{\text{SE}}$, where $\widehat{\text{SE}}$ is the estimated standard error of T_n, obtained using the bootstrap or otherwise. The bootstrap statistic which corresponds to t_n is clearly $t_n^* = (T_n^* - T_n)/\widehat{\text{SE}}^*$, where $\widehat{\text{SE}}^*$ is precisely $\widehat{\text{SE}}$ computed on a bootstrap sample. Following the same reasoning as in the normalized case (i.e., Remark 2), one can deduce that the UB of the sth quantile of t_n^* is greater than or equal to min{UB of the sth quantile of T_n^*, LB(T_n)}. It is assumed here that upper or lower outliers do not cause $\widehat{\text{SE}}$ to approach 0. Other possible reasons for $\widehat{\text{SE}}$ to go to 0 are excluded from consideration. However, this conclusion lacks substance, at least in the case of studentized mean, that is, $t_n = \sqrt{n}(\overline{X} - \mu)/s_n$. To carry this t_n to $+\infty$, one would need to drive 100% of the data toward $+\infty$. Thus, the sth quantile of t_n^* has UB given by

$$\min\{p: B(n, p) = n\} > 1 - s.$$

The resulting number is generally greater than $\frac{1}{2}$. A much more relevant breakdown of t_n occurs when just one data value goes to ∞. Then, the t-statistics approximately equal $+1$, which is entirely independent of the data at hand! It is not clear what the consequence of this odd phenomenon is on the quantiles of t_n^*. It should be a worthwhile project to study the breakdown of bootstrap based tests along the lines of existing breakdown-related literature for test [see He, Simpson and Portnoy (1990), Ylvisacker (1977)].

REMARK 4. *Trimming prior to bootstrapping*. In the case of the one-dimensional sample median, another way to robustify bootstrap would be to trim symmetrically prior to resampling, instead of Winsorizing. Fix a $0 < \beta < \frac{1}{2}$. Trim off $[\beta n]$ observation from each end and then resample from the remaining $n - 2[n\beta]$ data values, in order to learn about the sampling distribution of the sample median. It turns out that one needs to lower the bootstrap sample size to

$$m = n(1 - 2\beta)^2 \tag{4.4}$$

in order to preserve consistency. To see (4.4), consider the limiting distribution function of the empirical c.d.f. of the remaining $n - 2[n\beta]$ data, after the trimming. The median of this truncated limiting population is the same as the median M of the original population; however, the population density at M is inflated by a factor $1/(1 - 2\beta)$. The asymptotic variance of the sample median is given by $[4nf^2(M)]^{-1}$. Thus, the bootstrap sample size m should be changed to (4.4), in order to nullify the change in the population density at M. In the author's opinion, Winsorizing is preferable.

The resulting UB of the tth bootstrap quantile is equal to $[n\beta]/n + b_t(n - [2n\beta])$ where $b_t \equiv b_t(n)$ is given by Theorem 1.

REMARK 5. *Multivariate estimators*. Let T_n be p-variate estimator, $p \geq 2$. Let b be a fraction such that *at least* bn observations need to go to ∞ in order to cause the breakdown, $\|T_n\| \to \infty$. In the cloud of all possible bootstrap-vector statistics T_n^*, let us define a tth centrality-quantile (CQ_t) as a vector T_n^* such that $100t\%$ of the T_n^* vectors are more central than the T_n^* under consideration. To measure centrality, one could use Tukey's depth or some other depth [see Liu and Singh (1993)]. Now, the same arguments which led to Theorem 1 imply the following:

$$\text{the breakdown of } CQ_t \geq \min\{p\colon P(\mathrm{Bin}(n, p) \geq nb) > 1 - t\}.$$

It should be pointed out here that as $\|T_n^*\| \to \infty$, its centrality tends toward its minimum value, and hence it becomes more and more of an outwardly extreme CQ_t. The Winsorizing idea, discussed in Section 3, for robustizing the bootstrap in the case of m-estimators, extends in a straightforward manner for multivariate m-estimators.

APPENDIX

PROOF OF THEOREM 2 Part (a). If the number of outliers present in the upper side is less than or equal to $[n\beta]$, then the bootstrap statistic stays untouched by these outliers. If this number goes beyond $[n\beta]$, then suddenly all the upper outliers become effective. The statement in (a) is based on this logic.

Part (b). Here, we are assuming $\beta < \alpha$. In order for L_n^{**} to be different from L_n^*, one of the following must occur:

$$G_n^{-1}(\alpha) \leq X_{(l+1)}, \qquad l = [n\beta], \tag{A1}$$

$$G_n^{-1}(1 - \alpha) \geq X_{(n-l)}. \tag{A2}$$

Let us recall that F_n, G_n, H_n are the empirical c.d.f. of $\{X_i\}$, $\{Y_i\}$ and $\{Y_i^*\}$, respectively, and $\{Y_i^*\}$ are the Winsorized bootstrap data as prescribed in Section 3. Because if A_1 and A_2 do not hold then $G_n^{-1}(t) = H_n^{-1}(t)$ for all $\alpha \leq t \leq 1 - \alpha$; which means $L_n^* = L_n^{**}$.

Consider A1:

$$G_n^{-1}(\alpha) \le X_{(l+1)},$$
$$\Rightarrow \alpha \le G_n\big(X_{(l+1)}\big).$$

The bootstrap mean of this right-hand side equals

$$F_n\big(X_{(l+1)}\big) = \beta \pm \frac{1}{n}.$$

Thus A1 is a subset of

$$\sup_x |G_n(x) - F_n(x)| \ge \alpha - \beta - \frac{1}{n},$$

which is a large deviation-type event and it is well known that its probability goes to 0, exponentially. One could deduce it from the famous DKW inequality [see Dvoretzky, Kiefer and Wolfowitz (1956) and Massart (1990)]. The DKW inequality states that if $\xi_1, \xi_2, \dots, \xi_n$ are i.i.d. observations from a population with c.d.f. η and $\eta_n(\cdot)$ is the empirical c.d.f., then for any $d > 0$,

$$P\Big(\sqrt{n}\ \sup_x |\eta_n(x) - \eta(x)| > d\Big) \le 2\exp(-2d^2).$$

A similar result is proved for A2.

Part (c). Here, we have $\alpha = \beta$. Let us define

$$\alpha_n = \sup\{t\colon H_n^{-1}(t) > G_n^{-1}(t)\}.$$

The result in this segment of the theorem hinges on the following two claims:

(A3) $$P_B\big(\alpha_n > \alpha + cn^{-1/2}(\log n)^{1/2}\big) \to 0$$

for c large enough, a.s.

(A4) With $I_n = [\alpha, \alpha + cn^{-1/2}(\log n)^{1/2}]$,

$$P_B\Big(\sup_{t \in I_n} |G_n^{-1}(t) - H_n^{-1}(t)| > c' n^{-1/2}(\log n)^{1/2}\Big) \to 0,$$

a.s. for some c' depending upon c.

Similar results are derived around the upper end of the interval $[\alpha, 1-\alpha]$. From these results, it is concluded that

$$P_B\left\{\int_\alpha^{1-\alpha} |G_n^{-1}(t) - H_n^{-1}(t)|\, dt > cn^{-1}\log n\right\} \to 0$$

a.s. for some c large enough and thus the theorem follows. It remains to establish A3 and A4:

$$\begin{aligned}
&\alpha_n > \alpha + cn^{-1/2}(\log n)^{1/2} \\
&\quad \Rightarrow H_n^{-1}\left(\alpha + cn^{-1/2}(\log n)^{1/2}\right) > G_n^{-1}\left(\alpha + cn^{-1/2}(\log n)^{1/2}\right) \\
\text{(A3)} \qquad &\quad \Rightarrow H_n^{-1}\left(\alpha + cn^{-1/2}(\log n)^{1/2}\right) \le X_{(l+1)}, \\
&\qquad l = [n\alpha] = [n\beta] \\
&\quad \Rightarrow \alpha + cn^{-1/2}(\log n)^{1/2} \le H_n\left(X_{(l+1)}\right) = G_n\left(X_{(l+1)}\right),
\end{aligned}$$

the bootstrap mean of which $F_n(X_{(l+1)}) = \alpha \pm 1/n$. Thus, the event in A3 is a subset of

$$\sup|G_n(x) - F_n(x)| > cn^{-1/2}(\log n)^{1/2};$$

on which DKW can be applied to derive the desired conclusion.

(A4) In a neighborhood of α, with the bootstrap probability going to 1 exponentially,

$$H_n^{-1}(t) \ge G_n^{-1}(t).$$

As a consequence,

$$\sup_{I_n} |G_n^{-1}(t) - H_n^{-1}(t)| \le H_n^{-1}\left(\alpha + cn^{-1/2}(\log n)^{1/2}\right) - G_n^{-1}(\alpha).$$

Also,

$$H_n^{-1}\left(\alpha + cn^{-1/2}(\log n)^{1/2}\right) \le \max\left\{X_{(l+1)}, G_n^{-1}\left(\alpha + cn^{-1/2}(\log n)^{1/2}\right)\right\}.$$

Thus,

$$\begin{aligned}
&\sup_{I_n} |G_n^{-1}(t) - H_n^{-1}(t)| \\
&\quad \le |X_{(l+1)} - G_n^{-1}(\alpha)| + |G_n^{-1}\left(\alpha + cn^{-1/2}(\log n)^{1/2}\right) - G_n^{-1}(\alpha)| \\
&\quad \le |X_{(l+1)} - G_n^{-1}(\alpha)| + |G_n^{-1}(\alpha) - F_n^{-1}(\alpha)| \\
&\qquad + |G_n^{-1}\left(\alpha + cn^{-1/2}(\log n)^{1/2}\right) - F_n^{-1}\left(\alpha + cn^{-1/2}(\log n)^{1/2}\right)| \\
&\qquad + |F_n^{-1}\left(\alpha + cn^{-1/2}(\log n)^{1/2}\right) - F_n^{-1}(\alpha)|.
\end{aligned}$$

The last term above is further written as

$$\begin{aligned}
&|F_n^{-1}(\alpha) - F^{-1}(\alpha)| \\
&\quad + |F_n^{-1}\left(\alpha + cn^{-1/2}(\log n)^{1/2}\right) - F^{-1}\left(\alpha + cn^{-1/2}(\log n)^{1/2}\right)| \\
&\quad + |F^{-1}(\alpha) - F^{-1}\left(\alpha + cn^{-1/2}(\log n)^{1/2}\right)|.
\end{aligned}$$

Now, everything above is in terms of $G_n^{-1}(\cdot) - F_n^{-1}(\cdot)$, $F_n^{-1}(\cdot) - F^{-1}(\cdot)$ and $F^{-1}(\cdot)$. A standard set-inequality argument in conjunction with the DKW inequality is applied to finish off this proof. The details are omitted. □

PROOF OF THEOREM 3 Part (a). Until θ_n itself breaks down, all the outliers are totally ineffective, due to the Winsorization at $\theta_n \pm c$. However, as soon as $\theta_n \to \infty$, all the upper outliers become effective and then the formula in Theorem 1 for b_t applies. Thus, one has $b_t^* = \max(b_t, b)$. A similar logic applies in the lower case.

Part (b). Here, the bootstrap data $\{Y_i\}$ are Winsorized at $d > c$. The Winsorized data are denoted by $\{Y_i^*\}$. We begin by noting that $\theta_n^* = \theta_n^{**}$ if

$$\theta_n^* \pm c \text{ is contained in } \theta_n \pm d.$$

As a consequence,

$$\begin{aligned}\{\theta_n^* \neq \theta_n^{**}\} &\subseteq \{\theta_n^* + c \geq \theta_n + d\} \cup \{\theta_n^* - c \leq \theta_n - d\}\\ &= \{|\theta_n^* - \theta_n| \geq d - c > 0\}.\end{aligned}$$

This is a large deviation event in the bootstrap probability, in terms of θ_n^* as an estimator of θ_n. This can be converted into a large deviation event in terms of a sample mean, with bounded summands, as follows: for a $\delta > 0$,

$$\{\theta_n^* \geq \theta_n + \delta\} \subseteq \left\{\sum_1^n g_c(Y_i - \theta_n - \delta) \geq 0\right\}.$$

Let us look at the bootstrap mean of the summand $g_c(Y_i - \theta_n - \delta)$. Clearly,

$$E_B g_c(Y_1 - \theta_n - \delta) = \frac{1}{n}\sum_1^n g_c(X_i - \theta_n - \delta) < 0.$$

Since $\theta_n \to \theta$, a.s., $E_B g_c(Y_1 - \theta_n - \delta) \leq -\delta' < 0$ for all large n, a.s. The random variables $g_c(Y_i - \theta_n - \delta)$ are uniformly bounded. Thus it takes a standard asymptotic bound to conclude that $P_B(\theta_n^* \geq \theta_n + \delta)$ dwindles exponentially fast. One can treat $P_B(\theta_n^* \leq \theta_n - \delta)$ similarly.

Part (c). Here, we take $c = d$. Consider the case when $\theta_n^* < \theta_n$. The other case, $\theta_n^* > \theta_n$, is handled similarly. When $\{Y_i^*\}$ replace $\{Y_i\}$, the total change that occurs in $n^{-1}\sum g_c(Y_i - \theta_n^*)$ is

$$+F(\theta - c)g'(-c)(\theta_n - \theta_n^*) + o_p^*(1) = \lambda_1(\theta_n - \theta_n^*) + o_p^*(1)$$

(o_p^* refers to bootstrap probability). To counter this change, so that the average remains 0, one has to move θ_n^* upward, in the direction of θ_n. By the time θ_n^* is taken all the way to θ_n, one has already exceeded the needed correction. This explains the shrinkage phenomenon.

Let us attempt to measure this shrinkage. Suppose, θ_n^* is moved up by an amount $\delta_n < (\theta_n - \theta_n^*)$. Then, $n^{-1}\sum g_c(Y_i^* - \theta_n^*)$ moves down by the amount (using the assumed regularity conditions)

$$\left[F(\theta - c)g'(-c) + \int_{\theta - c}^{\theta + c} g'\, dF + o_p^*(1)\right]\delta_n = \left[\lambda_1 + \lambda_2 + o_p^*(1)\right]\delta_n.$$

Thus the balance occurs when

$$\left[\lambda_1 + \lambda_2 + o_p^*\right]\delta_n = \left[\lambda_1 + o_p^*(1)\right](\theta_n - \theta_n^*).$$

If $(\theta_n^{**} - \theta_n) = A_n(\theta_n^* - \theta_n)$, then $\delta_n = (1 - A_n)(\theta_n - \theta_n^*)$. Thus the balancing equation becomes, with $A_n = \lambda + o_p^*(1)$,

$$(\lambda_1 + \lambda_2)(1 - \lambda)(\theta_n - \theta_n^*) = \lambda_1(\theta_n - \theta_n^*).$$

Cancelling $(\theta_n - \theta_n^*)$ from both sides, one has

$$(1 - \lambda) = \lambda_1/(\lambda_1 + \lambda_2)$$

or

$$\lambda = 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{\lambda_2}{\lambda_1 + \lambda_2}.$$

If $F(\theta - c) = 1 - F(\theta + c) = \alpha$ and $g(x) = x$, one has $\lambda_1 = \alpha$ and $\lambda_2 = 1 - 2\alpha$. Therefore,

$$\lambda = \frac{1 - 2\alpha}{1 - \alpha} = 1 - \frac{\alpha}{1 - \alpha}. \qquad \square$$

Acknowledgments. The comments of an Associate Editor and the referees helped improve the presentation.

REFERENCES

CARROLL, R. J. (1978). On almost sure expansion for *M*-estimators. *Ann. Statist.* **6** 314–318.

DVORETZKY, A., KIEFER, J. C. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

HE, X., SIMPSON, D. G. and PORTNOY, S. L. (1990). Breakdown resistance of tests. *J. Amer. Statist. Assoc.* **85** 446–452.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York (transferred to Wadsworth, 1991).

LIU, R. and SINGH, K. (1993). A quality index based on data-depth. *J. Amer. Statist. Assoc.* **88** 252–260.

MASSART, P. (1990). The tight constant in the Dvoretzky, Kiefer, Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283.

STROMBERG, A. J. (1997). Robust covariance estimates based on resampling. *J. Statist. Plann. Inference* **57** 321–334.

YLVISAKER, D. (1977). Test resistance. *J. Amer. Statist. Assoc.* **72** 551–556.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
HILL CENTER FOR THE MATHEMATICAL SCIENCES
BUSCH CAMPUS
PISCATAWAY, NEW JERSEY 08855
E-MAIL: kesar@stat.rutgers.edu