



Necessary and Sufficient Conditions for Weak Consistency of the Median of Independent but not Identically Distributed Random Variables

Ivan Mizera; Jon A. Wellner

The Annals of Statistics, Vol. 26, No. 2 (Apr., 1998), 672-691.

Stable URL:

<http://links.jstor.org/sici?sici=0090-5364%28199804%2926%3A2%3C672%3ANASCFW%3E2.0.CO%3B2-6>

The Annals of Statistics is currently published by Institute of Mathematical Statistics.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ims.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

**NECESSARY AND SUFFICIENT CONDITIONS FOR WEAK
CONSISTENCY OF THE MEDIAN OF INDEPENDENT
BUT NOT IDENTICALLY DISTRIBUTED
RANDOM VARIABLES**

BY IVAN MIZERA¹ AND JON A. WELLNER²

Comenius University and University of Washington

Necessary and sufficient conditions for the weak consistency of the sample median of independent, but not identically distributed random variables are given and discussed.

1. Consistency of the sample median. For each $n = 1, 2, \dots$, suppose that $X_{n1}, X_{n2}, \dots, X_{nn}$ are independent random variables with distribution functions $F_{n1}, F_{n2}, \dots, F_{nn}$. Let F_n denote the empirical distribution function of the X_{ni} 's:

$$F_n(x) \equiv \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, x]}(X_{ni}),$$

and let \bar{F}_n be the average distribution function

$$\bar{F}_n \equiv \frac{1}{n} \sum_{i=1}^n F_{ni}.$$

For any distribution function G , let G^{-1} be the left-continuous inverse of G defined by $G^{-1}(u) \equiv \inf\{x: G(x) \geq u\}$, $0 < u < 1$. Throughout this paper, unless otherwise noted, we call $G^{-1}(1/2)$ *the median of G* , even when there is a nondegenerate interval $[m_0, m_1]$ of *median points* in the sense that $P_G(Y \leq m) \geq 1/2$ and $P_G(Y \geq m) \geq 1/2$ for $m \in [m_0, m_1]$ (see also the discussion after Examples 6 and 7 in Section 3).

The problem is to give necessary and sufficient conditions for weak consistency of the sample median $F_n^{-1}(1/2)$: under what conditions on the F_{ni} 's does it hold that

$$(1.1) \quad F_n^{-1}\left(\frac{1}{2}\right) - \bar{F}_n^{-1}\left(\frac{1}{2}\right) \rightarrow_P 0?$$

The sufficiency part of the problem has been studied by several authors, mainly for the i.i.d. case, the case when F_{ni} are equal to F for all $i = 1, 2, \dots, n$ and all $n \geq 1$: either for the median alone, starting perhaps from Kolmogorov (1931), or in the more general framework of M -estimation, as in Huber (1981).

Received April 1996; revised June 1997.

¹Supported by Fonds National de la Recherche Scientifique, by the Banque National de Belgique and by the Slovak GAS Grant 1/4196/97.

²Research supported in part by NSF Grant DMS-93-06809.

AMS 1991 subject classifications. Primary 60F05, 62G30; secondary 62G20, 62E20.

Key words and phrases. Median, weak consistency, empirical distribution function, majorization.

In the general, “non-i.i.d.” case, the relevant references are Weiss (1969), Sen (1968, 1970), Stigler (1976), Shorack and Wellner (1986) and Portnoy (1991).

However, it seems that so far no sufficient and *necessary* conditions for (1.1) have been established. In this context, a comparison with the older brother of the median in the realm of location estimation—the sample mean—comes to mind: here the research on laws of large numbers have been crowned by theorems giving necessary and sufficient conditions, for the i.i.d. as well as “non-i.i.d.” cases; see, for instance, Petrov (1995).

Consistency of the sample median has many statistical applications—for i.i.d. as well as for non-identically distributed observations. For an interesting application of medians to filtering in a setting involving dependence as well as non-identically distributed observations, see Moore and Jorgenson (1993). For an applied, operational side, sufficient conditions are vital: they provide the fuel needed to proceed further. The impact of necessary conditions is different. For instance, the first author was brought to the problem through the alignment considerations for the runs tests of randomness under a heteroscedasticity hypothesis: the sample median is a natural aligning estimator for rank tests based on signs—and invariance properties of these determine their use in nonhomogeneous situations [see Dufour, Hallin and Mizera (1995)]. Thus, necessary conditions may be viewed more as negative results: they outline the scope and limitations of situations where consistency is needed.

2. Necessary and sufficient conditions. It turns out that in the special case when all F_{ni} 's have common median $\xi_{1/2}$, that is, when

$$(2.1) \quad F_{ni}^{-1}\left(\frac{1}{2}\right) = \xi_{1/2} \quad \text{for } i = 1, 2, \dots, n, \quad n = 1, 2, \dots,$$

the problem has a neat solution. For fixed $\varepsilon > 0$, define

$$a_n(\varepsilon) \equiv E\left[\mathbb{F}_n\left(\bar{F}_n^{-1}\left(\frac{1}{2}\right) + \varepsilon\right)\right] = \bar{F}_n\left(\bar{F}_n^{-1}\left(\frac{1}{2}\right) + \varepsilon\right)$$

and

$$b_n(\varepsilon) \equiv E\left[\mathbb{F}_n\left(\bar{F}_n^{-1}\left(\frac{1}{2}\right) - \varepsilon\right)\right] = \bar{F}_n\left(\bar{F}_n^{-1}\left(\frac{1}{2}\right) - \varepsilon\right).$$

Note that $b_n(\varepsilon) \leq \frac{1}{2} \leq a_n(\varepsilon)$.

THEOREM 1. *Suppose that $X_{n1}, X_{n2}, \dots, X_{nn}$ are independent random variables with distribution functions $F_{n1}, F_{n2}, \dots, F_{nn}$, all with a common median $\xi_{1/2}$. A necessary and sufficient condition for weak consistency of the sample median is that*

$$(2.2) \quad \sqrt{n}(a_n(\varepsilon) - \frac{1}{2}) \rightarrow \infty \quad \text{and} \quad \sqrt{n}\left(\frac{1}{2} - b_n(\varepsilon)\right) \rightarrow \infty$$

holds for all $\varepsilon > 0$.

PROOF. This will be proved in Section 4 via Theorem 3, which shows that the statement is a consequence of the more general Theorem 2. The basic method involves rewriting events concerning sample quantiles in terms

of events concerning the empirical distribution function (see, e.g., Kiefer (1970)]. □

Condition (2.2) of Theorem 1 remains sufficient in general without assuming the equality of medians condition (2.1), and in fact condition (2.2) is also necessary under a mild nondegeneracy hypothesis: see Theorem 3 below. Furthermore, this basic result carries over in an obvious way to a general t th quantile; see Section 4.

Although Theorem 1 can provide a satisfactory answer in many practical cases, the general problem is of interest too—in parallel to the classical laws of large numbers: for example, in the case of the weak law of large numbers, Feller [(1971), pages 235 and 565] gives conditions (now known as the weak- L_1 condition) under which there exist constants μ_n such that the sequence $\{\bar{X}_n\}$ of sample means satisfies $\bar{X}_n - \mu_n \rightarrow 0$ in probability. Thus, we would like to consider the problem of consistency *without assuming* (2.1).

Such a problem, however, is a more delicate one. The difficulty may lie in the fact that under the “non-i.i.d.” case, condition (2.1) fails for an arbitrary sequence of deterministic (and hence independent) random variables (such a sequence would be forced to be constant in the i.i.d. case). Nevertheless, it turns out that the “purely deterministic” situation enjoys the same level of tractability as the “purely stochastic” one. It is the borderline “not deterministic, not stochastic” behavior which causes problems.

For fixed $\varepsilon > 0$, define $c_n(\varepsilon)$ and $d_n(\varepsilon)$ to be nonnegative numbers such that

$$\begin{aligned} c_n^2(\varepsilon) &\equiv n \operatorname{Var} \left[\mathbb{F}_n \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) + \varepsilon \right) \right] \\ &= \frac{1}{n} \sum_{i=1}^n F_{ni} \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) + \varepsilon \right) \left(1 - F_{ni} \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) + \varepsilon \right) \right) \end{aligned}$$

and

$$\begin{aligned} d_n^2(\varepsilon) &\equiv n \operatorname{Var} \left[\mathbb{F}_n \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) - \varepsilon \right) \right] \\ &= \frac{1}{n} \sum_{i=1}^n F_{ni} \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) - \varepsilon \right) \left(1 - F_{ni} \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) - \varepsilon \right) \right). \end{aligned}$$

Note that

$$(2.3) \quad 0 \leq c_n(\varepsilon) \leq \frac{1}{2} \quad \text{and} \quad 0 \leq d_n(\varepsilon) \leq \frac{1}{2}.$$

We adopt the following conventions: $a/0 = \infty$ if $a > 0$; $0/0 = 0$.

THEOREM 2. *Suppose that $X_{n1}, X_{n2}, \dots, X_{nn}$ are independent random variables with distribution functions $F_{n1}, F_{n2}, \dots, F_{nn}$.*

A sufficient condition for weak consistency of the sample median is that

$$(2.4) \quad \sqrt{n} \frac{a_n(\varepsilon) - \frac{1}{2}}{c_n(\varepsilon)} \rightarrow \infty \quad \text{and} \quad \sqrt{n} \frac{\frac{1}{2} - b_n(\varepsilon)}{d_n(\varepsilon)} \rightarrow \infty$$

holds for all $\varepsilon > 0$.

A necessary condition for weak consistency of the sample median can be formulated as follows: The sample median is not consistent, if, for some $\varepsilon > 0$:

(i) there is a subsequence $\{n'\}$ such that

$$(2.5a) \quad \sqrt{n'}c_{n'}(\varepsilon) \rightarrow \infty \quad \text{and} \quad \sqrt{n'} \frac{a_{n'}(\varepsilon) - \frac{1}{2}}{c_{n'}(\varepsilon)} = O(1)$$

or

$$(2.5b) \quad \sqrt{n'}d_{n'}(\varepsilon) \rightarrow \infty \quad \text{and} \quad \sqrt{n'} \frac{\frac{1}{2} - b_{n'}(\varepsilon)}{d_{n'}(\varepsilon)} = O(1);$$

(ii) or, there is a $K > 0$ and a subsequence $\{n'\}$ such that

$$(2.6a) \quad \sqrt{n'}c_{n'}(\varepsilon) \rightarrow K \quad \text{and} \quad \sqrt{n'} \frac{a_{n'}(\varepsilon) - \frac{1}{2}}{c_{n'}(\varepsilon)} < K$$

or

$$(2.6b) \quad \sqrt{n'}d_{n'}(\varepsilon) \rightarrow K \quad \text{and} \quad \sqrt{n'} \frac{\frac{1}{2} - b_{n'}(\varepsilon)}{d_{n'}(\varepsilon)} < K;$$

(iii) or, there is a subsequence $\{n'\}$ such that

$$(2.7a) \quad \sqrt{n'}c_{n'}(\varepsilon) \rightarrow 0 \quad \text{and} \quad \sqrt{n'} \frac{a_{n'}(\varepsilon) - \frac{1}{2}}{c_{n'}(\varepsilon)} = O(1)$$

or

$$(2.7b) \quad \sqrt{n'}d_{n'}(\varepsilon) \rightarrow 0 \quad \text{and} \quad \sqrt{n'} \frac{\frac{1}{2} - b_{n'}(\varepsilon)}{d_{n'}(\varepsilon)} = O(1);$$

The proof is postponed to Section 4.

COROLLARY 1. *If, under the assumptions of Theorem 2, for any subsequence it follows that*

$$(2.8) \quad \begin{aligned} \sqrt{nc_n(\varepsilon)} &\rightarrow \infty \quad \text{whenever} \quad a_n(\varepsilon) \rightarrow \frac{1}{2} \quad \text{and} \\ \sqrt{nd_n(\varepsilon)} &\rightarrow \infty \quad \text{whenever} \quad b_n(\varepsilon) \rightarrow \frac{1}{2}, \end{aligned}$$

then the sample median is consistent if and only if (2.4) holds for every $\varepsilon > 0$.

PROOF. The necessary part is a consequence of (2.5) and the fact that if a_n or b_n is bounded away from $1/2$, then (2.4) holds due to (2.3). The sufficiency part follows directly from Theorem 2. \square

The final theorem shows that Theorem 1 is a special case of Theorem 2.

THEOREM 3. *Suppose that $X_{n1}, X_{n2}, \dots, X_{nn}$ are independent random variables with distribution functions $F_{n1}, F_{n2}, \dots, F_{nn}$. If there are a, b such that*

$$(2.9) \quad 0 < a \leq F_{ni}(\bar{F}_n^{-1}(\frac{1}{2})) \leq b < 1 \quad \text{for all } i = 1, 2, \dots, n,$$

then (2.8) holds and condition (2.2) is equivalent to condition (2.4).

PROOF. The proof is for $a_n(\varepsilon)$ and $c_n(\varepsilon)$; for $b_n(\varepsilon)$ and $d_n(\varepsilon)$ it is analogous. Note again that if $\liminf a_n(\varepsilon) > 1/2$, both (2.2) and (2.4) hold; hence in proving the equivalence we can also restrict to the case when $a_n(\varepsilon) \rightarrow 1/2^+$, but then (2.9) entails

$$F_{ni}(\bar{F}_n^{-1}(\frac{1}{2}) + \varepsilon) \geq a,$$

hence

$$\begin{aligned} c_n^2(\varepsilon) &= \frac{1}{n} \sum_{i=1}^n F_{ni} \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) + \varepsilon \right) \left(1 - F_{ni} \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) + \varepsilon \right) \right) \\ &\geq \frac{1}{n} \sum_{i=1}^n a \left(1 - F_{ni} \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) + \varepsilon \right) \right) = a(1 - a_n(\varepsilon)) \rightarrow \frac{1}{2}a > 0, \end{aligned}$$

showing that $c_n(\varepsilon)$ is bounded away from zero; thus, in this case (2.2) and (2.4) are equivalent and (2.8) holds. \square

Note that (2.1) is a special case of (2.9).

3. Examples, corollaries and remarks. Here are some examples illustrating the general case. The presentation is condensed; a more thorough treatment can be found in Mizera and Wellner (1996). We use the symbol δ_x for the point probability concentrated at x . For simplicity, we consider in Examples 1–5 only odd $n = 2k + 1$ (for even n we could put $X_{ni} = 0$ almost surely for all i , if desired).

EXAMPLE 1. Let

$$X_{ni} =_d \begin{cases} \delta_{-1}, & \text{for } i = 1, 2, \dots, k, \\ \delta_0, & \text{for } i = k + 1, \\ \delta_1, & \text{for } i = k + 2, k + 3, \dots, n. \end{cases}$$

This is a purely deterministic case. The difference between $\mathbb{F}_n^{-1}(1/2)$ and $\bar{F}_n^{-1}(1/2)$ is identically 0, condition (2.2) does not hold, but condition (2.4) holds trivially.

In the deterministic case, including Example 1, condition (2.4) is always satisfied, since $c_n(\varepsilon) = d_n(\varepsilon) = 0$.

EXAMPLE 2. Replace $X_{n, k+1}$ in Example 1 by a random variable uniformly distributed in $[-1, 1]$. Now we have, for positive $\varepsilon < 1$,

$$\sqrt{n}c_n(\varepsilon) = \frac{1}{2}\sqrt{1 - \varepsilon^2} \equiv K_\varepsilon > 0$$

and

$$\sqrt{n} \frac{a_n(\varepsilon) - \frac{1}{2}}{c_n(\varepsilon)} = n \frac{\varepsilon/(2n)}{\frac{1}{2}\sqrt{1 - \varepsilon^2}} = \frac{\varepsilon}{\sqrt{1 - \varepsilon^2}}.$$

This is less than K_ε for small ε , so Theorem 2 yields inconsistency—as can also be checked directly.

EXAMPLE 3. Replace $X_{n,k+1}$ in Example 1 by a random variable uniformly distributed in $[-1/n, 1/n]$. Then (2.4) holds, and the median is consistent by Theorem 2. Note that $\sqrt{nc_n(\varepsilon)}$ and $\sqrt{nd_n(\varepsilon)}$ converge to 0, but (2.7) fails. [Actually, $c_n(\varepsilon)$ and $d_n(\varepsilon)$ are zero for large n , and condition (2.4) holds trivially again. A more sophisticated variation of this example could be produced by letting $X_{n,k+1}$ shrink to 0 in a more “smooth” way; for instance, setting its distribution to $N(0, 1/n)$.]

The results of Theorem 2 reveal some general features of the possible behavior in the “independent, but not identically distributed” paradigm. Theorem 2 suggests distinguishing three principal situations:

Case 1. *Stochastic.* The variance of n times the empirical distribution function \mathbb{F}_n at the points $r_n^\pm \equiv \bar{F}_n^{-1}(1/2) \pm \varepsilon$ is unbounded. This corresponds to the situation of (2.5) and partially also (2.6). In these cases, generally speaking, the problem of consistency can be decided in terms of “macroparameters,” the mean and variance of the empirical distribution function. A typical representative is the i.i.d. case.

Case 2. *Quasideterministic.* The variance of $n\mathbb{F}_n(r_n^\pm)$ degenerates to zero—in our setting (2.7). A typical representative is the purely deterministic case. Here consistency again can be, basically, decided in the terms of mean and variance of the empirical distribution function.

Case 3. *“Chaotic” or “pseudostochastic.”* The remaining case—the behavior is erratic and unpredictable; in fact, as shown by Examples 4 and 5, the problem of consistency is undecidable in terms of the mean and variance of the empirical distribution function.

EXAMPLE 4. Let $X_{ni} =_d (1 - 1/k)\delta_{-1} + (1/k)\delta_0$, for $i = 1, 2, \dots, k$, $X_{ni} =_d \delta_0$ for $i = k+1$, and $X_{ni} =_d (1/k)\delta_0 + (1-1/k)\delta_1$ for $i = n-k+1, n-k+2, \dots, n$. For $\varepsilon \in (0, 1)$,

$$a_n(\varepsilon) = \frac{k+2}{n} = \frac{1}{2} + \frac{3}{2n}, \quad b_n(\varepsilon) = \frac{k-1}{n} = \frac{1}{2} - \frac{3}{2n}$$

and $c_n^2(\varepsilon) = d_n^2(\varepsilon) = (1/n)(1 - 1/k)$. Hence $\sqrt{nc_n(\varepsilon)} = \sqrt{nd_n(\varepsilon)} \rightarrow 1$, and

$$\sqrt{n} \frac{a_n(\varepsilon) - \frac{1}{2}}{c_n(\varepsilon)} = \sqrt{n} \frac{\frac{1}{2} - b_n(\varepsilon)}{d_n(\varepsilon)} = \frac{\frac{3}{2}}{\sqrt{1 - 1/k}} \rightarrow \frac{3}{2}.$$

Theorem 2 is inconclusive here, but $\mathbb{F}_n^{-1}(1/2) = 0 = \bar{F}_n^{-1}(1/2)$ with probability 1 for all n , and the sample median is consistent.

EXAMPLE 5. Let $p_n = (1/2)(1 - k^{-1/2})$, and define

$$X_{ni} =_d \begin{cases} \delta_{-1}, & \text{for } i = 1, 2, \dots, k-3, \\ p_n\delta_{-1} + (1 - p_n)\delta_1, & \text{for } i = k-2, k-1, \\ \delta_0, & \text{for } i = k, k+1, k+2, \\ (1 - p_n)\delta_{-1} + p_n\delta_1, & \text{for } i = k+3, k+4, \\ \delta_1, & \text{for } i = k+5, k+6, \dots, n. \end{cases}$$

Computing $a_n(\varepsilon)$, $b_n(\varepsilon)$, $c_n(\varepsilon)$ and $d_n(\varepsilon)$ for this example, we discover they are the same as in Example 4 for all ε . Theorem 2 is again inconclusive, but the median is inconsistent! To see this, just note that $k + 1 \geq n/2$ and

$$\begin{aligned} P[\text{card}\{i: X_{ni} = -1\} = k + 1] &= P[\text{card}\{i: X_{ni} = 1\} = k + 1] \\ &= p_n^2(1 - p_n)^2 \rightarrow \frac{1}{16} > 0. \end{aligned}$$

The last two examples show why the oddity of restricting to odd sequences appears.

EXAMPLE 6. Let $X_{n1}, X_{n2}, \dots, X_{nn}$ be the first n terms of a sequence formed by setting, with probability 1, $X_1 = -1$, $X_2 = 1$, $X_3 = -1$, $X_4 = 1, \dots$. The sample median is equal to -1 for all n and it is consistent.

EXAMPLE 7. The same as Example 6, but now let $X_1 = 1$, $X_2 = -1$, $X_3 = 1$, $X_4 = -1, \dots$. The sample median is equal to 1 for all odd n and is equal to -1 for all even n , and is inconsistent.

Note that, for even n , we have in both examples the same $a_n(\varepsilon)$, $b_n(\varepsilon)$, $c_n(\varepsilon)$ and $d_n(\varepsilon)$ for all $\varepsilon > 0$. Problems of this kind come from the possible existence of multiple median points—points where \mathbb{F}_n is equal to $1/2$. (This is also the reason for adopting the convention $0/0$: it serves to rule out median points which do not move toward the median, as defined in Section 1.) For the *sample* median, this effect is restricted to even n . In the general case, a possible way is outlined in the Introduction: we could adopt a set-valued definition of the median. Consistency then could be defined through some concept of set convergence: upper or lower limits [we follow the terminology of Aubin and Frankowska (1990)].

The upper-limit consistency expresses that any sample median sequence (i.e., any possible sequence of sample median points) approaches the sequence of population medians: the distance between a *sequence* of sample medians and *the sets* of population medians converges to zero in probability. If the population median remains fixed, as in Theorem 1, then all accumulation points in probability of any sequence of sample medians lie in the population median set.

The lower-limit consistency is a more stringent one, requiring that the Hausdorff distance between *the sets* of sample medians and *the sets* of population medians goes to zero in probability. In our opinion, the upper-limit consistency satisfies all needs of applications; technical complications brought by the lower-limit one are not counterbalanced by increase of its practical utility or specific interpretation.

In this vein, extensions of the present results covering the case of non-unique population median of the sequence of distributions \bar{F}_n could be considered. Analogous results (concerning the upper convergence) could be derived, with a_n 's, b_n 's, c_n 's and d_n 's carefully redefined and proofs, going along the same lines, revised. We, however, sacrifice such a development, preferring

rather to maintain clarity of the exposition and readability of proofs, with a hope that in a need of possible generalizations the relevant ideas will be transparent enough.

Actually, the outlined path was followed in the (simpler in this respect) i.i.d. case. First, in the i.i.d. case conditions (2.2) are satisfied if and only if

$$(3.1) \quad F(\xi_{1/2} + \varepsilon) > \frac{1}{2} \quad \text{and} \quad F(\xi_{1/2} - \varepsilon) < \frac{1}{2}$$

for all $\varepsilon > 0$. Note that (3.1) holds for all $\varepsilon > 0$ if and only if F is not flat at $\xi_{1/2} = F^{-1}(1/2)$, and this holds if and only if $F^{-1}(u)$ is continuous at $1/2$. This is the usual condition for consistency of $\mathbb{F}_n^{-1}(1/2)$ as an estimator of $F^{-1}(1/2)$ [see, e.g., Huber (1981), page 54, or Serfling (1980), page 75]. This also follows from the representation of the empirical quantile function $\{\mathbb{F}_n^{-1}(t): 0 < t < 1, n \geq 1\} =_d \{F^{-1}(\mathbb{G}_n^{-1}(t)): 0 < t < 1, n \geq 1\}$, where $\mathbb{G}_n^{-1}(t)$ is the quantile function of i.i.d. uniform(0, 1) random variables [see Shorack and Wellner (1986), pages 4, 5 and 637]. Second, if we adopt the broader viewpoint of upper-limit consistency, we can say that *in the i.i.d. case the sample median is always weakly consistent*: any sample median sequence still converges to the interval of all possible median points of F [see, e.g., Mizera (1993)].

We have seen that, for the i.i.d. case, the conditions of Theorem 1 can be simply verified. Another application of Theorem 1 is to *heteroscedastic models*. Suppose F_0 is a fixed absolutely continuous distribution function with median 0 and with a bounded density f_0 such that if $x \in [-\lambda, \lambda]$, then $f(x) \geq L$ for some $\lambda > 0$ and $L > 0$. Suppose that $F_{ni}(x) = F_0(c_{ni}x)$ for $1 \leq i \leq n$ and $n \geq 1$ for nonnegative constants $\{c_{ni}\}$. Note that the median of F_0 is unique and (2.1) holds. We want to express consistency of the sample median through the “empirical distribution” G_n of scaling constants c_{ni} at stage n : $G_n(t) = n^{-1} \sum_{i=1}^n 1_{[c_{ni} \leq t]}$. The question may be of interest outside the realm of location estimation: just think of F_{ni} ’s as distributions of possible disturbances in regression or autoregressive models, say; heteroscedastic models are particularly popular in these types of applications. These links deserve further exploration, but we know the answer in the location case.

Note first that $\bar{F}_n(x) = \int_0^\infty F_0(tx) dG_n(t)$. If $G_n \rightarrow_d G$, where G is not degenerate at 0, then $\bar{F}_n(x) \rightarrow \int_0^\infty F_0(tx) dG(t)$. If F_0 is strictly increasing, then the limit is strictly increasing too, so the condition of Theorem 1 holds. On the other hand, if $G_n \rightarrow_d \delta_0$, the distribution with all its mass at 0, the condition for Theorem 1 can hold or fail depending on the rate at which the sequence of distributions G_n degenerates to 0. Mizera and Wellner [(1996), Example 7] show that, when $\sqrt{n}(G_n - \delta_0) \rightarrow \Delta$ in the sense of uniform convergence, condition (2.2) fails and hence consistency fails by Theorem 1. More generally, let $\Phi_c(x)$ be a function from $(0, \infty)$ to $(0, \infty)$ equal to $1/c$ if $x \leq c$ and $1/x$ if $x \geq c$. Hallin and Mizera (1996) derived the following corollary of Theorem 1: a necessary and sufficient condition for the consistency of the sample median is that $\sqrt{n} \int_0^\infty \Phi_c(t) dG_n(t) \rightarrow \infty$ for some (equivalently, for all) $c > 0$. The special case of this result follows from the asymptotic normality result of Sen (1968). See also Hallin and Mizera (1996, 1997) for the extensions of this result and Theorem 1 to consistency rates and general M -estimators.

Of course the methods developed here for the median carry over straightforwardly to an arbitrary fixed t th quantile. Suppose, for instance, that $0 < t < 1$ and $F_{ni}^{-1}(t) = \xi_t$ for all $1 \leq i \leq n$ and $n \geq 1$. Then

$$F_n^{-1}(t) - \bar{F}_n^{-1}(t) \rightarrow_p 0$$

if and only if

$$\sqrt{n}(a_n(\varepsilon) - t) \rightarrow \infty \quad \text{and} \quad \sqrt{n}(t - b_n(\varepsilon)) \rightarrow \infty$$

for every $\varepsilon > 0$ where

$$a_n(\varepsilon) \equiv \bar{F}_n(\bar{F}_n^{-1}(t) + \varepsilon) \quad \text{and} \quad b_n(\varepsilon) \equiv \bar{F}_n(\bar{F}_n^{-1}(t) - \varepsilon).$$

4. Proofs for Theorems 1 and 2. The necessity parts of these proofs are based on majorization arguments. To make these more transparent, we introduce partial sum functions, also called Lorenz functions [for this and more background on majorization, see Marshall and Olkin (1979)]. A function from $[0, n]$ to $[0, n]$ is called *piecewise linear* if it is continuous and linear on every interval $[k - 1, k]$, $k = 1, 2, \dots, n$. We also say that P is a piecewise linear function with *turning points* $(k_1, x_1), (k_2, x_2), \dots, (k_m, x_m)$, if it is continuous, $P(k_1) = x_1, P(k_2) = x_2, \dots, P(k_m) = x_m$ and linear on $[0, n] \setminus \{k_1, k_2, \dots, k_m\}$. The right endpoint $(n, P(n))$ and, for aesthetic reasons, also the left endpoint $(0, 0)$ are always mentioned among the turning points when the function is specified.

Given a sequence $0 \leq q_1 \leq q_2 \leq \dots \leq q_n$, its *partial sum function* S_q is a piecewise linear function from $[0, n]$ to $[0, n]$ such that

$$S_q(x) = (x - [x]) \sum_{i=1}^{[x]} q_i + ([x] - x) \left(\sum_{i=1}^{[x]} q_i - q_{[x]} \right).$$

In other terms, S_q is a piecewise linear function with turning points $(0, 0), (1, q_1), (2, q_1 + q_2), \dots, (n, \sum_{i=1}^n q_i)$. Clearly, S_q is nondecreasing and convex. Conversely, any nondecreasing convex piecewise linear function S from $[0, n]$ to $[0, n]$ is a partial sum function for some sequence $0 \leq q_1 \leq q_2 \leq \dots \leq q_n$. Such a sequence is *majorized* by a sequence $0 \leq \tilde{q}_1 \leq \tilde{q}_2 \leq \dots \leq \tilde{q}_n$ if and only if

$$(4.1) \quad S_{\tilde{q}}(n) = S_q(n) \quad \text{and} \quad S_{\tilde{q}}(x) \leq S_q(x)$$

for all $x \in [0, n]$; that is,

$$\sum_{j=1}^i \tilde{q}_{nj} \leq \sum_{j=1}^i q_{nj}$$

for all $i = 1, 2, \dots, n$, with equality for n . [An equivalent and more usual definition of majorization involves the n -vectors \tilde{q} and q arranged in descending order and the family of reverse inequalities. See Marshall and Olkin (1979), page 9. For brevity, we say also that S_q is majorized by $S_{\tilde{q}}$, if (4.1) holds; that is, the graph of $S_{\tilde{q}}$ does not exceed that of S_q and they have common endpoints.

We use Gleser's (1975) refinement of Hoeffding's (1956) inequality.

LEMMA 1. Let Y_i and \tilde{Y}_i be, respectively, independent random variables with Bernoulli(q_i) and Bernoulli(\tilde{q}_i) distributions. If the sequence $0 \leq q_1 \leq q_2 \leq \dots \leq q_n \leq 1$ is majorized by the sequence $0 \leq \tilde{q}_1 \leq \tilde{q}_2 \leq \dots \leq \tilde{q}_n \leq 1$, then

$$P\left[\sum_{i=1}^n Y_i \geq \lambda\right] \geq P\left[\sum_{i=1}^n \tilde{Y}_i \geq \lambda\right],$$

provided $\lambda \geq \lceil \sum_{i=1}^n q_i \rceil + 2$.

As already noted, Theorem 1 is a consequence of Theorem 2. The basic idea behind the necessity part of the direct proof is contained in Figure 1. For the details of the direct proof see Mizera and Wellner (1996).

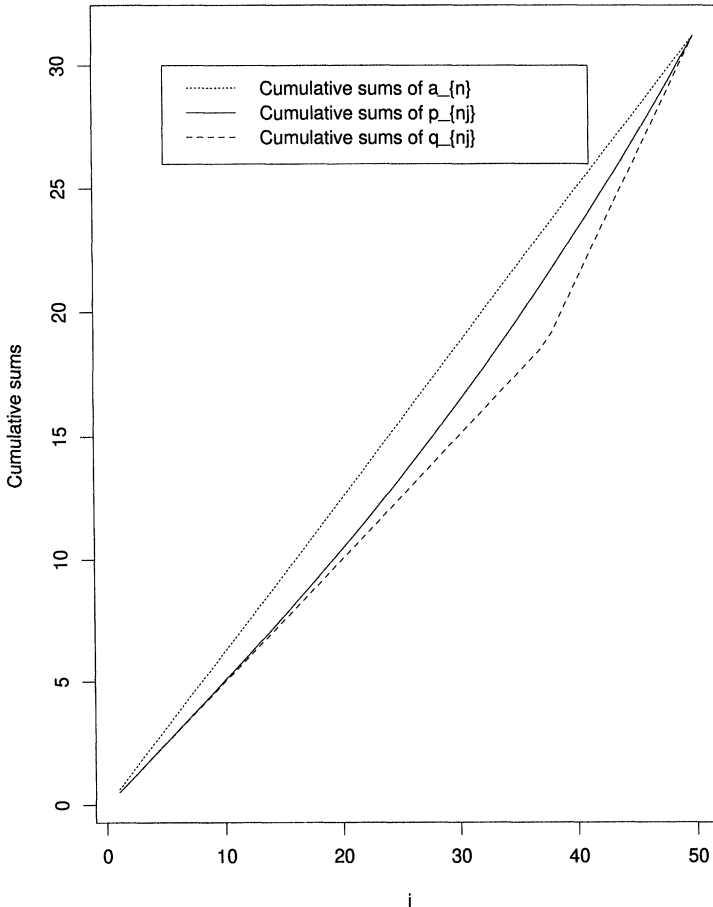


FIG. 1. Majorization plot.

Now we give several lemmas to prepare for the proof of Theorem 2. Given a partial sum function S_q , we define its *entropy* to be

$$H(S_q) = \sum_{i=1}^n q_i(1 - q_i).$$

Due to concavity of $x(1 - x)$, we have that $H(S_q) \geq H(S_{\tilde{q}})$ whenever S_q is majorized by $S_{\tilde{q}}$.

LEMMA 2. *Let S_q be a partial sum function with $S_q(n) = \alpha$. If $S_q(k) \leq \beta$, then*

$$H(S_q) \leq \alpha - \frac{\beta^2}{k} - \frac{(\alpha - \beta)^2}{n - k}.$$

PROOF. Just observe that if $S_q(k) \leq \beta$, then the graph of S_q does not exceed the graph of piecewise linear function R with turning points $(0, 0)$, (k, β) , (n, α) by convexity of S_q . Then

$$\begin{aligned} H(S_q) &\leq H(R) = k \frac{\beta}{k} \left(1 - \frac{\beta}{k}\right) + (n - k) \frac{\alpha - \beta}{n - k} \left(1 - \frac{\alpha - \beta}{n - k}\right) \\ &= \alpha - \frac{\beta^2}{k} - \frac{(\alpha - \beta)^2}{n - k}, \end{aligned}$$

since R is majorized by S_q . \square

To overcome the “defect of 2” in Gleser’s inequality given in Lemma 1, which seems to be substantial [see Gleser (1975) for details], we need also the following lemma.

LEMMA 3. *Let $1 \geq q_{mi} \geq \eta_m > 0$ for all $i = 1, 2, \dots, m$ and $m = 1, 2, \dots$. Suppose that Y_{mi} are independent random variables with Bernoulli(q_{mi}) distributions for $i = 1, \dots, m$. If*

$$\frac{\sum_{i=1}^m q_{mi}(1 - q_{mi})}{\eta_m^2} = O(1),$$

then

$$\lim_{m \rightarrow \infty} P \left[\sum_{i=1}^m Y_{mi} = m \right] > 0.$$

PROOF. Note first that the sequence

$$P \left[\sum_{i=1}^m Y_{mi} = m \right] = \prod_{i=1}^m q_{mi}$$

is nonincreasing and bounded by 0 from below; hence we can really speak about the limit. The inequality $\log x \geq (x - 1)/x$ for $0 < x \leq 1$ yields

$$\begin{aligned} \log \prod_{i=1}^m q_{mi} &= \sum_{i=1}^m \log q_{mi} \geq \sum_{i=1}^m \frac{q_{mi} - 1}{q_{mi}} \\ &= \sum_{i=1}^m -\frac{q_{mi}(1 - q_{mi})}{q_{mi}^2} \geq -\frac{\sum_{i=1}^m q_{mi}(1 - q_{mi})}{\eta_m^2}. \end{aligned}$$

The statement follows, since the last term is bounded away from $-\infty$. \square

PROOF OF THE SUFFICIENT CONDITION OF THEOREM 2. Let $\varepsilon > 0$. We begin by rewriting

$$\begin{aligned} &P \left[\left| \mathbb{F}_n^{-1} \left(\frac{1}{2} \right) - \bar{F}_n^{-1} \left(\frac{1}{2} \right) \right| > \varepsilon \right] \\ &\leq P \left[\mathbb{F}_n^{-1} \left(\frac{1}{2} \right) > \bar{F}_n^{-1} \left(\frac{1}{2} \right) + \varepsilon \right] + P \left[\mathbb{F}_n^{-1} \left(\frac{1}{2} \right) < \bar{F}_n^{-1} \left(\frac{1}{2} \right) - \varepsilon \right] \\ &\leq P \left[\mathbb{F}_n \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) + \varepsilon \right) < \frac{1}{2} \right] + P \left[\mathbb{F}_n \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) - \varepsilon \right) \geq \frac{1}{2} \right] \\ &= P \left[E\mathbb{F}_n \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) + \varepsilon \right) - \mathbb{F}_n \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) + \varepsilon \right) > \alpha_n(\varepsilon) - \frac{1}{2} \right] \\ &\quad + P \left[\mathbb{F}_n \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) - \varepsilon \right) - E\mathbb{F}_n \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) - \varepsilon \right) \geq \frac{1}{2} - b_n(\varepsilon) \right] \\ &\leq P \left[\left| \mathbb{F}_n \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) + \varepsilon \right) - E\mathbb{F}_n \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) + \varepsilon \right) \right| > \alpha_n(\varepsilon) - \frac{1}{2} \right] \\ &\quad + P \left[\left| \mathbb{F}_n \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) - \varepsilon \right) - E\mathbb{F}_n \left(\bar{F}_n^{-1} \left(\frac{1}{2} \right) - \varepsilon \right) \right| \geq \frac{1}{2} - b_n(\varepsilon) \right] \\ &\leq \frac{\text{Var}(\mathbb{F}_n(\bar{F}_n^{-1}(\frac{1}{2}) + \varepsilon))}{(\alpha_n(\varepsilon) - \frac{1}{2})^2} + \frac{\text{Var}(\mathbb{F}_n(\bar{F}_n^{-1}(\frac{1}{2}) - \varepsilon))}{(\frac{1}{2} - b_n(\varepsilon))^2} \\ &= \frac{c_n^2(\varepsilon)}{n(\alpha_n(\varepsilon) - \frac{1}{2})^2} + \frac{d_n^2(\varepsilon)}{n(\frac{1}{2} - b_n(\varepsilon))^2}, \end{aligned}$$

the last inequality follows from two applications of Chebyshev's inequality (note that $P[|X - EX| \geq \eta] \leq \text{Var } X/\eta^2$ remains valid for $\text{Var } X = 0$, if $\eta > 0$). Under (2.4), the last terms converge to 0, completing the proof. \square

PROOF OF THE NECESSARY CONDITION OF THEOREM 2. The proof is for (2.5b), (2.6b) and (2.7b); for (2.5a), (2.6a) and (2.7a) it is analogous and symmetric. The hypothesis implies that, for some ε and some subsequence,

$$(4.2) \quad \sqrt{n} \frac{\left(\frac{1}{2} \right) - b_n(\varepsilon)}{d_n(\varepsilon)} = O(1).$$

Note that this automatically entails that $b_n(\varepsilon) \rightarrow 1/2^-$. We start from the elementary inequalities

$$\begin{aligned}
 P\left[\left|\mathbb{F}_n^{-1}\left(\frac{1}{2}\right) - \bar{\mathbb{F}}_n^{-1}\left(\frac{1}{2}\right)\right| > \varepsilon\right] &\geq P\left[\bar{\mathbb{F}}_n^{-1}\left(\frac{1}{2}\right) - \mathbb{F}_n^{-1}\left(\frac{1}{2}\right) > \varepsilon\right] \\
 (4.3) \qquad \qquad \qquad &\geq P\left[\mathbb{F}_n\left(\bar{\mathbb{F}}_n^{-1}\left(\frac{1}{2}\right) - \varepsilon\right) > \frac{1}{2}\right] = P\left[\sum_{i=1}^n Y_{ni} > \frac{1}{2}n\right],
 \end{aligned}$$

where Y_{ni} are independent random variables with Bernoulli(p_{ni}) distribution, with $p_{ni} = \mathbb{F}_{ni}\left(\bar{\mathbb{F}}_n^{-1}\left(\frac{1}{2}\right) - \varepsilon\right)$. Note that

$$\sum_{i=1}^n p_{ni} = nb_n(\varepsilon) \quad \text{and} \quad \sum_{i=1}^n p_{ni}(1 - p_{ni}) = nd_n^2(\varepsilon).$$

We shall suppose, without restricting generality, that $p_{n1} \leq p_{n2} \leq \dots \leq p_{nn}$. If we succeed in finding a subsequence of probabilities appearing in the last term of (4.3) such that its \liminf is bounded away from zero, we succeed in proving the inconsistency of the whole sequence—and hence in proving the statement.

In the sequel, we pass sometimes to a subsequence; however, we keep the same indexing, to avoid tedious notation. We also drop the argument of ε for b_n and d_n .

Suppose first that there is a subsequence such that $\sqrt{n}d_n \rightarrow \infty$, satisfying (2.5b). The random variables $Z_{ni} = Y_{ni} - p_{ni}$ have zero expectation and satisfy the Liapunov condition

$$\begin{aligned}
 \frac{\sum_i E|Z_{ni}|^3}{(\sum_i \text{Var } Z_{ni})^{3/2}} &= \frac{\sum_i p_{ni}^3(1 - p_{ni}) + (1 - p_{ni})^3 p_{ni}}{(\sum_i p_{ni}(1 - p_{ni}))^{3/2}} \\
 &= \frac{\sum_i p_{ni}(1 - p_{ni})(p_{ni}^2 + (1 - p_{ni})^2)}{(\sum_i p_{ni}(1 - p_{ni}))^{3/2}} \leq \frac{2 \sum_i p_{ni}(1 - p_{ni})}{(\sum_i p_{ni}(1 - p_{ni}))^{3/2}} \\
 &= 2\left(\sum_i p_i(1 - p_i)\right)^{-1/2} = 2(nd_n^2)^{-1/2} \rightarrow 0.
 \end{aligned}$$

In view of (4.2), we have, for n large,

$$\begin{aligned}
 P\left[\sum_{i=1}^n Y_{ni} > \frac{1}{2}n\right] &= P\left[\sum_{i=1}^n \frac{Y_{ni} - p_{ni}}{\sqrt{\sum_i p_{ni}(1 - p_{ni})}} > \sqrt{n} \frac{1/2 - b_n}{d_n}\right] \\
 &\geq P\left[\sum_{i=1}^n \frac{Z_{ni}}{\sqrt{\sum_i \text{Var } Z_{ni}}} \geq M\right] \rightarrow \int_M^\infty \phi(x) dx > 0,
 \end{aligned}$$

by the Liapunov central limit theorem; $\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ is the standard normal density function. Combining this with (4.3) implies that the median is inconsistent in this case.

Now, for the case of (2.6b), the heuristic outline of the proof is as follows (Figure 2 seems to be helpful). Condition (2.6b) ensures that, asymptotically, the entropy of the partial sum function S_n of p_{ni} 's is not too small in comparison with the difference between $n/2$ and nb_n . As a consequence, the possible

S_n lying too close to the minimum entropy boundary (P in Figure 2) are ruled out. That particularly means that S_n starts to be nonzero before crossing the boundary of $n/2$; in other words, there are more than $n/2$ nonzero p_{ni} 's (otherwise, there is no hope for inconsistency). [The sharpness of (2.6b) is illustrated by the piecewise linear function which is zero until $n/2$ and then linearly ascends to (n, nb_n) : it has entropy approximately equal to nb_n .] However, having enough positive p_{ni} 's does not mean automatically inconsistency (as shown in Examples 4 and 5). The entropy condition helps again, since it admits only those S_n 's which ascend steeply enough, after being equal to zero. Thus, after ruling out the low entropy cases, the majority of the remaining possible S_n 's is majorized by a partial sum function such that it is easily seen to yield inconsistency (Q or Q_Γ in Figure 2). The inconsistency for S_n 's then follows via Gleser's inequality. Due to its, already mentioned, "defect of 2," not all of the S_n 's could be treated this way: for these, Lemma 3 applies.

To start with details, suppose that there is a subsequence $\sqrt{n}d_n \rightarrow K > 0$ satisfying (2.6b). Such a sequence contains only a finite number of $d_n = 0$, so that we can pass to a subsequence containing no $d_n = 0$. We can pass further to a subsequence such that

$$(4.4) \quad \sqrt{n} \frac{(\frac{1}{2}) - b_n}{d_n} \rightarrow M < K.$$

As a consequence, $n(1/2 - b_n) \rightarrow L = MK$. Now, let

$$[[nb_n]] \equiv \begin{cases} [nb_n] - 1, & \text{if } \frac{1}{2}n - L \text{ is an integer and } nb_n \geq \frac{1}{2}n - L, \\ [nb_n], & \text{otherwise.} \end{cases}$$

Let $\Delta_n = nb_n - [[nb_n]]$. Here, the following possibilities can occur: L and $2L$ is not an integer, that is, the fractional part of $n/2 - L$ is never zero; then since $n/2 - L - nb_n \rightarrow 0$, we have

$$(4.5) \quad \liminf_{n \rightarrow \infty} \Delta_n > \Delta > 0.$$

Or, $n/2 - L$ is integer for odd or even n (note that the possibilities are exclusive), but there is (respectively for odd or even n) a subsequence of nb_n approaching $n/2 - L$ from below; more precisely, $n/2 - L - nb_n$ is decreasing and nonzero. Then again (4.5) holds. All these possibilities are denoted as case (a) (Figure 2a). Note that, in this case, we can pass to a subsequence such that $\Delta_n < 1$ (n should be large, and odd or even n should be reduced to a subsequence if necessary). Hence

$$(4.6) \quad \limsup_{n \rightarrow \infty} \Delta_n \leq 1.$$

In the remaining case, denoted as case (b) (Figure 2b), $n/2 - L$ is integer for odd or even n , and, respectively for odd or even n , we can pick only (recall that, from every convergent sequence, a monotone subsequence can be picked) a subsequence of nb_n approaching $n/2 - L$ from above; that is, $n/2 - L - nb_n$ is increasing, possibly zero. Then we reduce, respectively, odd or even n to this subsequence; for those n , we have $\Delta_n \geq 1$, hence (4.5) holds, and, since $n/2 -$

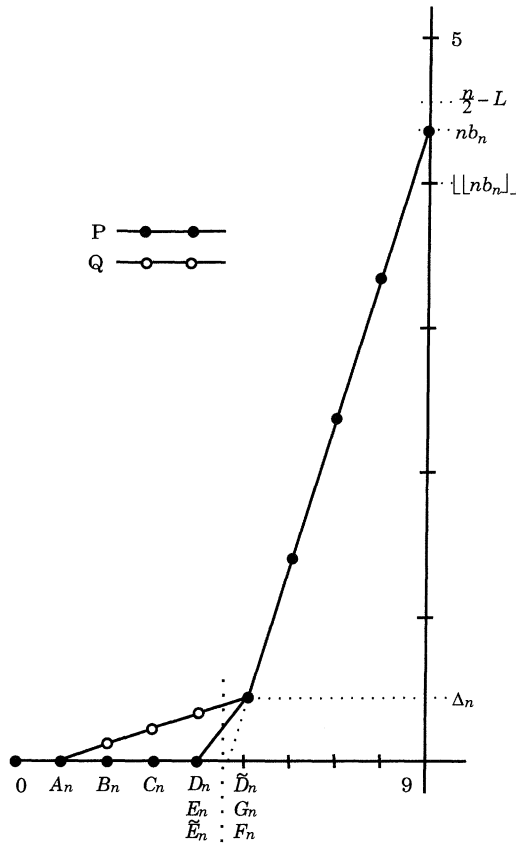


FIG. 2a. Case (i), since $L = 0$ and $n = 9$ is odd, and case (a), since $n/2 - L$ is not integer.

$L - nb_n \rightarrow 0$, we have also (4.6). For remaining n (even or odd, respectively), we have (4.5) and (4.6) due to the same reason as in case (a), since $n/2 - L$ is not integer. Summing up, we can pass to a subsequence such that $\Delta_n \geq \Delta > 0$ for all n , and such that either (a) $\Delta_n < 1$ for all n or (b) $\Delta_n \rightarrow 1^+$, $\Delta_n < 2$ for all n .

Let $F_n = n - \lfloor \lfloor nb_n \rfloor \rfloor$, let $E_n = F_n - 1$. Let $\tilde{E}_n = E_n$ in case (a) (see Figure 2a); let $\tilde{E}_n = E_n - 1$ in case (b) (see Figure 2b). For n odd, let $\tilde{D}_n = G_n = \lceil n/2 \rceil$, $D_n = \lfloor n/2 \rfloor$; for n even, let $\tilde{D}_n = n/2 + 1$, $D_n = n/2 - 1$, $G_n = n/2$. In all cases, let $A_n = D_n - 3$, $B_n = D_n - 2$, $C_n = D_n - 1$. Note that, in any case, we have

$$(4.7) \quad 4 \leq E_n - A_n + 1 = F_n - A_n \leq L + 6,$$

due to (4.6) (see again Figure 2).

Returning to (4.3), we consider the partial sum function S_n of p_{ni} 's. Since all $p_{ni} \geq 0$, we have $S_n(x) \geq 0$ for all x ; since all $p_{ni} \leq 1$, the graph of S_n

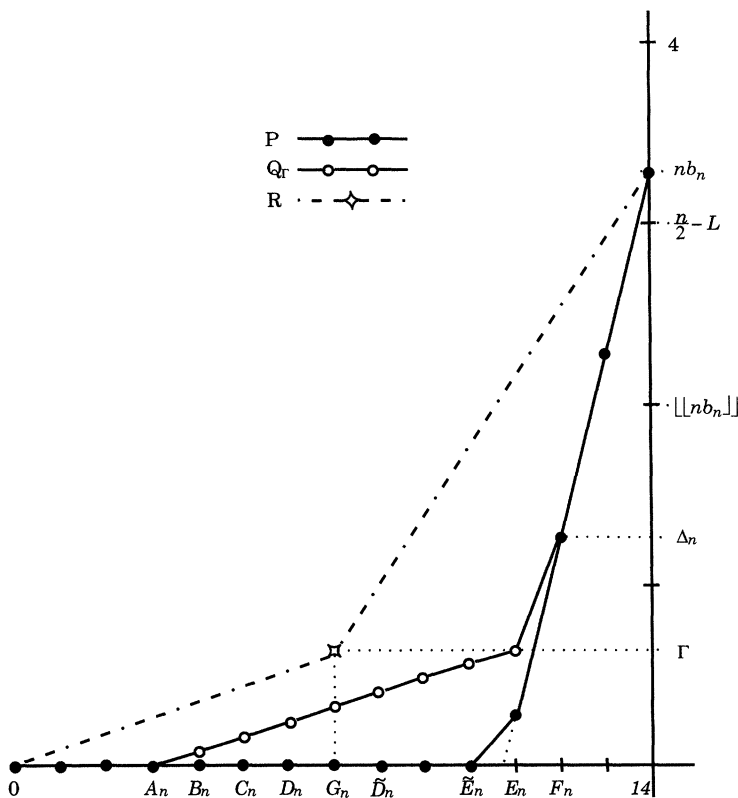


FIG. 2b. Cases (b) and (ii): $L = 4$, $n = 14$ and nb_n approaches $n/2 - L$ from above. All the symbols are explained in the text, except for R , which is a model example for R_m (or R from Lemma 2), bounding partial sum functions passing below the point (G_n, Γ) .

lies above (possibly touching) the line with slope 1, passing through (n, nb_n) . Consequently, S_n is majorized by the piecewise linear function P with turning points $(0, 0)$, $(\tilde{E}_n, 0)$, $(\tilde{E}_n + 1, nb_n - \lfloor nb_n \rfloor)$ and (n, nb_n) (see Figure 2). Note that $\tilde{E}_n + 1 \leq F_n$.

Now, we have again two cases—for the whole current subsequence (n is considered large enough): (i) $G_n = F_n$ (this happens if $L = 0$ and n is odd, as in Figure 2a); (ii) $G_n \neq F_n$ (as in Figure 2b). In case (i), let Q be the piecewise linear function with turning points $(0, 0)$, $(\tilde{E}_n, 0)$, (F_n, Δ_n) and (n, nb_n) —the partial sum function of the sequence of probabilities

$$q_{ni} = \begin{cases} 0, & \text{for } i = 1, 2, \dots, A_n, \\ \frac{\Delta_n}{F_n - A_n}, & \text{for } i = B_n, B_n + 1, \dots, E_n, \\ 1, & \text{for } i = F_n + 1, F_n + 2, \dots, n. \end{cases}$$

Note that (4.6) and (4.7) ensure that $\Delta_n/(F_n - A_n) \leq 1$. If S_n is majorized by Q for infinitely many n , we have by Lemma 1, for all these n ,

$$\begin{aligned} P\left[\sum_{i=1}^n Y_{ni} > \frac{1}{2}n\right] &\geq P\left[\sum_{i=1}^n \text{Bernoulli}(p_{ni}) \geq \frac{1}{2}n + 3\right] \\ &\geq P\left[\sum_{i=1}^n \text{Bernoulli}(q_{ni}) \geq \frac{1}{2}n + 3\right] \\ &\geq \left(\frac{\Delta_n}{F_n - A_n}\right)^{F_n - A_n} \geq \left(\frac{\Delta}{L + 6}\right)^{L+6} > 0, \end{aligned}$$

the last inequalities due to (4.5) and (4.6).

In case (ii), we shall show first that we can pass to a subsequence such that

$$(4.8) \quad S_n(G_n) > \Gamma > 0.$$

If (4.8) fails, then we have $\lim_{n \rightarrow \infty} S_n(G_n) = 0$; then, given any $m = 1, 2, \dots$, there is an $N(m) \geq m$ such that $S_{N(m)} \leq 1/m$; that is, $S_{N(m)} \leq R_m$, where R_m is the piecewise linear function with turning points $(0, 0)$, $(G_m, 1/m)$, $(N(m), N(m)b_{N(m)})$. Since $K^2 = \lim_{n \rightarrow \infty} nd_n^2$, Lemma 2 yields

$$\begin{aligned} K^2 &= \lim_{m \rightarrow \infty} N(m)d_{N(m)}^2 = \lim_{m \rightarrow \infty} \sum_{i=1}^{N(m)} p_{N(m), i}(1 - p_{N(m), i}) \\ &= \lim_{m \rightarrow \infty} H(S_{N(m)}) \leq \lim_{m \rightarrow \infty} H(R_m) \end{aligned}$$

and

$$H(R_m) = N(m)b_{N(m)} - \frac{m^{-2}}{G_{N(m)}} - \frac{(N(m)b_{N(m)} - m^{-2})^2}{N(m) - G_{N(m)}} \rightarrow MK,$$

since $N(m)b_{N(m)} - N(m)/2 \rightarrow -L$, $b_{N(m)} \rightarrow 1/2$, $G_{N(m)} \sim N(m)/2$ and due to (4.6). We have obtained that $K^2 \leq MK$, a contradiction to (4.4).

Hence, in case (ii) there is a $\Gamma > 0$, independent of n , such that (4.8) holds. Note that, due to (4.7) and (4.5), we can choose Γ such that

$$(4.9) \quad \Delta_n - \Gamma > \frac{\Gamma}{E_n - A_n}$$

for all n . We have then $\Delta_n - \Gamma \leq 1$ for large n , due to (4.6); we pass again to a subsequence to have it for all n . This and (4.9) ensure that the piecewise linear function Q_Γ with turning points $(0, 0)$, $(A_n, 0)$, (E_n, Γ) , (F_n, Δ_n) and (n, nb_n) is the partial sum function of the sequence of probabilities

$$q_{ni} = \begin{cases} 0, & \text{for } i = 1, 2, \dots, A_n, \\ \frac{\Gamma}{E_n - A_n}, & \text{for } i = B_n, B_n + 1, \dots, E_n, \\ \Delta_n - \Gamma, & \text{for } i = F_n, \\ 1, & \text{for } i = F_n + 1, F_n + 2, \dots, n \end{cases}$$

(see Figure 2b). Again, if S_n is for infinitely many n majorized by Q_Γ , we have by Lemma 1, for all these n ,

$$\begin{aligned} P\left[\sum_{i=1}^n Y_{ni} > \frac{1}{2}n\right] &\geq P\left[\sum_{i=1}^n \text{Bernoulli}(p_{ni}) \geq \frac{1}{2}n + 3\right] \\ &\geq P\left[\sum_{i=1}^n \text{Bernoulli}(q_{ni}) \geq \frac{1}{2}n + 3\right] \\ &\geq (\Delta_n - \Gamma)\left(\frac{\Gamma}{E_n - A_n}\right)^{E_n - A_n} \geq \left(\frac{\Gamma}{L + 6}\right)^{L+6} > 0, \end{aligned}$$

the last inequalities due to (4.5), (4.6) and (4.9). Note that in both cases (i) and (ii), Gleser's inequality was applicable since

$$\frac{1}{2}n + 3 \geq \frac{1}{2}n + 3 - L \geq \lceil nb_n \rceil + 2.$$

Hence, we obtained inconsistency, if, for infinitely many n , S_n is majorized by Q , in case (i), or by Q_Γ , in case (ii).

Thus, it remains only to show what happens if S_n is only for finitely many n majorized by Q_Γ (or Q). Then, for infinitely many n , the graph of Q_Γ (or Q) exceeds that of S_n . We pass to the corresponding subsequence; case (ii) is treated first. Note that in this case we have not only (4.7), but also

$$(4.10) \quad 4 \leq E_n - A_n.$$

We have $S_n(B_n) < Q_\Gamma(B_n)$ or $S_n(C_n) < Q_\Gamma(C_n)$ or $S_n(D_n) < Q_\Gamma(D_n)$. Since $S_n(G_n) \geq \Gamma$ [recall (4.8)], either

$$(4.11) \quad S_n(G_n) - S_n(B_n) \geq \Gamma - \Gamma \frac{1}{E_n - A_n}$$

and consequently

$$(4.12) \quad S_n(G_n) - S_n(D_n) \geq \frac{1}{3}\Gamma \left(1 - \frac{1}{E_n - A_n}\right) \geq \frac{1}{4}\Gamma,$$

in view of (4.10); or

$$(4.13) \quad S_n(G_n) - S_n(C_n) \geq \Gamma - \Gamma \frac{2}{E_n - A_n}$$

and similarly

$$(4.14) \quad S_n(G_n) - S_n(D_n) \geq \frac{1}{2}\Gamma \left(1 - \frac{2}{E_n - A_n}\right) \geq \frac{1}{4}\Gamma;$$

or directly

$$(4.15) \quad S_n(G_n) - S_n(D_n) \geq \Gamma - \Gamma \frac{3}{E_n - A_n} \geq \frac{1}{4}\Gamma,$$

for the same reason (see Figure 2b). Hence,

$$(4.16) \quad \frac{1}{4}\Gamma \leq p_{n, G_n} \leq p_{n, G_{n+1}} \leq \dots \leq p_{nn}.$$

In case (i), we proceed in an entirely similar way: as in (4.11)–(4.15), with Q instead of Q_Γ , F_n instead of E_n , Δ instead of Γ and using (4.7) instead of (4.10). In both cases we arrive to (4.16); then Lemma 3 concludes that

$$\liminf_{n \rightarrow \infty} P \left[\sum_{i=1}^n Y_{ni} > \frac{1}{2}n \right] \geq \lim_{n \rightarrow \infty} P \left[\sum_{i=G_n-1}^n Y_{ni} = n - G_n + 1 \right] > 0,$$

since

$$\frac{\sum_{i=G_n-1}^n p_{ni}(1-p_{ni})}{\left(\frac{1}{9}\right)\Gamma} \leq \frac{\sum_{i=1}^n p_{ni}(1-p_{ni})}{\left(\frac{1}{9}\right)\Gamma} = O(1),$$

with Δ instead of Γ in case (i). Thus, we have obtained inconsistency—and finished the proof for (2.6b).

Finally, suppose that there is a subsequence of odd integers satisfying (2.7b). If it contains an infinite number of n such that $b_n = 1/2$, we have for all these n that sample median equals $\bar{F}_n^{-1}(\frac{1}{2}) - \varepsilon$ —for n odd, multiple median points are not possible. This yields inconsistency. Thus, suppose there is only a finite number of n such that $b_n = 0$; pass to a subsequence such that all $b_n < 1/2$. Such a subsequence cannot have $d_n = 0$ for infinitely many n , since this would result in a subsequence containing ∞ infinitely many times, contradicting (2.7b). Hence, we can pass to a subsequence containing only nonzero d_n .

The rest of the proof proceeds along the same lines as that for (2.6b). Note that in this case we have $K = M = L = 0$, hence $\tilde{D}_n = F_n$: thus, case (b) cannot occur, since n is odd. Since $G_n = \tilde{D}_n$ for n odd, we have only case (i). Then either S_n is not exceeded by Q infinitely many times (then Lemma 1 applies) or it is infinitely many times exceeded by Q at A_n or B_n , but not exceeded at G_n , the case covered by Lemma 3. \square

Acknowledgments. We are indebted to Marc Hallin for posing to us the problem leading to Theorem 1. The help of Yihui Zhan and Mária Benešová in preparation of the figures is also gratefully acknowledged.

REFERENCES

- AUBIN, J.-P. and FRANKOWSKA, H. (1990). *Set-Valued Analysis*. Birkhäuser, Boston.
- DUFOUR, J.-M., HALLIN, M. and MIZERA, I. (1995). Generalized runs tests for heteroscedastic time series. *J. Nonparametr. Statist.* To appear.
- FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*. Wiley, New York.
- GLESER, L. J. (1975). On the distribution of the number of successes in independent trials. *Ann. Probab.* **3** 182–188.
- HALLIN, M. and MIZERA, I. (1996). Sample heterogeneity and the asymptotics of M -estimators. Preprint IS-P 1996-15 (n. 49), Inst. Statistique, Univ. Libre de Bruxelles. (Available via <http://www.dcs.fmph.uniba.sk/~mizera/writings.html>.)
- HALLIN, M. and MIZERA, I. (1997). Unimodality and the asymptotics of M -estimators. In *L₁-Statistical Procedures and Related Topics* (Y. Dodge, ed.) 47–56. IMS, Hayward, CA.
- HOEFFDING, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* **27** 713–721.
- HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 713–721.

- HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.
- KIEFER, J. (1970). Old and new methods for studying order statistics and sample quantiles. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 349–357. Cambridge Univ. Press, Cambridge.
- KOLMOGOROV, A. N. (1931). Method of median in the theory of errors. *Mat. Sb.* **38** 47–50. [Reprinted in English in *Selected Works of A. N. Kolmogorov* (A. N. Shirayev, ed.) Kluwer, Dordrecht, 1991.]
- MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.
- MIZERA, I. (1993). Weak continuity and identifiability of M -functionals. Ph.D. dissertation, Charles Univ., Prague. (In Slovak.)
- MIZERA, I. and WELLNER, J. A. (1996). Necessary and sufficient conditions for weak consistency of the median of independent but not identically distributed random variables. Technical Report 306, Dept. Statistics, Univ. Washington, Seattle. (Available via <http://www.stat.washington.edu/jaw/jaw.research.available.html>.)
- MOORE, A. W., JR. and JORGENSEN, J. W. (1993). Median filtering for removal of low-frequency background drift. *Analytical Chemistry* **65** 188–191.
- PETROV, V. V. (1995). *Limit Theorems of Probability Theory*. Clarendon, Oxford.
- PORTNOY, S. (1991). Asymptotic behavior of regression quantiles in nonstationary, dependent cases. *J. Multivariate Anal.* **38** 100–113.
- SEN, P. K. (1968). Asymptotic normality of sample quantiles for m -dependent processes. *Ann. Math. Statist.* **39** 1724–1730.
- SEN, P. K. (1970). A note on order statistics for heterogeneous populations. *Ann. Math. Statist.* **41** 2137–2139.
- SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- STIGLER, S. M. (1976). The effect of sample heterogeneity on linear functions of order statistics with applications to robust estimation. *J. Amer. Statist. Assoc.* **71** 956–960.
- WEISS, L. (1969). The asymptotic distribution of quantiles from mixed samples. *Sankhyā Ser. A* **31** 313–318.

DEPARTMENT OF PROBABILITY
AND STATISTICS, MLYNSKÁ DOLINA
COMENIUS UNIVERSITY
SK-84215 BRATISLAVA
SLOVAKIA
E-MAIL: mizera@fmph.uniba.sk

DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
BOX 354322
SEATTLE, WASHINGTON 98195-4322
E-MAIL: jaw@stat.washington.edu