*The Annals of Probability*
1990, Vol. 18, No. 3, 1284–1341

# THE ASYMPTOTIC JOINT DISTRIBUTION OF SELF-NORMALIZED CENSORED SUMS AND SUMS OF SQUARES

BY MARJORIE G. HAHN,[1] JIM KUELBS[2] AND DANIEL C. WEINER[3]

*Tufts University, University of Wisconsin and Boston University*

Empirical versions of appropriate centering and scale constants for random variables which can fail to have second or even first moments are obtainable in various ways via suitable modifications of the summands in the partial sum. This paper discusses a particular modification, called censoring (which is a kind of winsorization), where the (random) number of summands altered tends to infinity but the proportion altered tends to zero as the number of summands increases. Some analytic advantages inherent in this approach allow a fairly complete probabilistic and empirical theory to be developed, the latter involving the study of *studentized* or *self-normalized* sums. In particular, the joint asymptotic distributions of the empirically censored quantities of center and scale are determined as well as precise criteria for convergence to each of the allowable limit laws. Applications to the Feller class and domains of attraction are also considered.

**1. Introduction.** Let $X, X_1, X_2, \ldots$ be independent identically distributed (i.i.d.) random variables with distribution function $F(x) = P(X \le x)$ and partial sums $S_n = \sum_{i=1}^n X_i$. When $E(X^2) = \infty$, it is well known that there are centerings $\mu_n$ and deterministic normalizations (scalings) $\sigma_n$ such that for some nondegenerate $\beta$,

$$\mathscr{L}\left(\frac{S_n - \mu_n}{\sigma_n}\right) \to \beta, \tag{1.1}$$

iff $\beta$ is a stable probability measure, i.e., $\beta = \mathscr{L}(S_\alpha)$, where $S_\alpha$ is a stable random variable of index $\alpha$, $0 < \alpha \le 2$. The distributions $F$ for which $\mu_n$ and $\sigma_n$ exist so that (1.1) holds are said to be in the domain of attraction of a stable law of index $\alpha$, and we denote this by writing $F \in \mathrm{DA}(\alpha)$.

Necessary and sufficient conditions for $F \in \mathrm{DA}(\alpha)$ are classical [e.g., Feller (1971)] and it is obvious that any $\tau_n \sim \sigma_n$ can be used in (1.1). When $0 < \alpha < 2$ and $F \in \mathrm{DA}(\alpha)$, there are three classical ways to determine normalizations

Received March 1988; revised June 1989.
[1]Supported in part by NSF Grants DMS-85-02361 and DMS-87-02878.
[2]Supported in part by NSF Grant DMS-85-21586.
[3]Supported in part by NSF Grants DMS-86-03188 and DMS-88-96217.
*AMS* 1980 *subject classifications.* 60F05, 62G05, 62G30.
*Key words and phrases.* Asymptotic normality, self-normalized sums, studentization, censoring, winsorizing, Feller class, stochastic compactness, tightness, center and scale constants, infinite variance, domain of attraction.

$\tau_n \sim \sigma_n$ so that (1.1) holds. They are

$$\text{(1.2a)} \qquad \text{(the tail equation)} \quad \tau_n = \sup\{t \geq 0 \colon nP(|X| > t) \geq 1\},$$

$$\text{(1.2b)} \qquad \begin{gathered}\text{(the truncated second moment equation)}\\ \tau_n = \sup\{t \geq 0 \colon nEX^2 I(|X| \leq t) \geq t^2\},\end{gathered}$$

$$\text{(1.2c)} \qquad \begin{gathered}\text{(the censored second moment equation)}\\ \tau_n = \sup\{t \geq 0 \colon nE(X^2 \wedge t^2) \geq t^2\}.\end{gathered}$$

The term censoring refers to the replacement of $X$ by the censored quantity $(|X| \wedge t)\text{sgn}(X)$ as opposed to the truncated quantity $XI(|X| \leq t)$. Furthermore, in (1.2b) and (1.2c), for all but the first few $n$, $\tau_n$ actually satisfies the resulting equation with equality and the solutions suffice when $\alpha = 2$ as well.

In pursuing a more robust weak convergence theory when $E(X^2) = \infty$, a number of recent investigations have examined weak limit theorems for self-normalized or studentized partial sums suitably centered. A fundamental paper in this area is Logan, Mallows, Rice and Shepp (1973) and further references are contained therein. Another method is via trimmed sums studied recently by several authors. Ours is a third approach which combines some ideas from each of these. [For a related work more closely connected to trimming and winsorizing, see Hahn, Kuelbs and Weiner (1989a).]

To obtain self-normalized limit theorems, it is natural to examine normalizers (empirical scale quantities) $\hat{\tau}_n$ obtained from the empirical version of (1.2a), (1.2b) or (1.2c) when the distribution function $F$ is replaced by the empirical distribution function $F_n = (1/n)\Sigma_{j=1}^{n}\delta_{X_j}$. If the censoring equation is used, then the empirical version of $\tau_n$ is

$$\text{(1.3)} \qquad \hat{\tau}_n \equiv \sup\left\{t \geq 0 \colon n\int \frac{x^2 \wedge t^2}{t^2} dF_n(x) \geq 1\right\} = \left(\sum_{j=1}^{n} X_j^2\right)^{1/2}$$

and the corresponding version of the centerings $\mu_n$ can be defined by

$$\text{(1.4)} \qquad \hat{\mu}_n \equiv \int (|X| \wedge \hat{\tau}_n)\text{sgn}(x)\, dF_n(x) = \frac{1}{n}\sum_{j=1}^{n} (|X_j| \wedge \hat{\tau}_n)\text{sgn}(X_j).$$

Since $\hat{\tau}_n \geq \max_{j \leq n}|X_j|$, notice that

$$n\hat{\mu}_n = \sum_{j=1}^{n} (|X_j| \wedge \hat{\tau}_n)\text{sgn}(X_j) = \sum_{j=1}^{n} X_j = S_n$$

and the paper by Logan, Mallows, Rice and Shepp (1973) relates to the limiting behavior of the self-normalized sums

$$\text{(1.5)} \qquad \mathscr{L}\left(\frac{S_n - n\mu_n}{\hat{\tau}_n}\right) = \mathscr{L}\left(\frac{n(\hat{\mu}_n - \mu_n)}{\hat{\tau}_n}\right).$$

In fact, if $1 < \alpha \leq 2$, they assume $E(X) = 0$ and set all centerings equal to zero. When $0 < \alpha \leq 1$, they again use centerings equal to zero along with the assumption that $X$ is strictly stable for $\alpha = 1$. With these centering assumptions in place, they then prove that all limit laws have a subgaussian tail which depends in a complicated way on the perhaps unknown parameter $\alpha$. Furthermore, the limit laws have densities which have infinite discontinuities at $\pm 1$, so they are far less familiar than the classical normal density. Finally, the results do not include distributions $F$ outside the domain of attraction of some stable law, although the quantities in (1.3) and (1.4) are, in fact, appropriate in a much wider setting.

Consequently, we seek an alternative approach which attempts to address some of these problems. It is highly desirable to keep the normal or "nice" functions of it as the limiting distribution. The basic reason for nonnormality above is due to the existence of terms of excessively large magnitude as noted in Lévy's classic theorem of 1937. Thus, the basic idea is to somehow delete or otherwise neutralize their effect in $S_n$. Notice that since $\hat{\tau}_n \geq \max_{j \leq n}|X_j|$, all of the large terms in $\hat{\mu}_n$ remain unaltered. One way of improving this situation is to reduce the empirical censoring levels $\hat{\tau}_n$ in (1.3) and similarly their probabilistic analogues $\tau_n$ in (1.2c). This can be achieved by considering scale quantities $a_n$ determined by

$$nE\left(\frac{X^2 \wedge a_n^2}{a_n^2}\right) = r_n, \tag{1.6}$$

with $r_n > 1$ instead of the equation resulting from (1.2c). Once the normalizations are determined, they are used to determine centerings $\gamma_n$ given by

$$\gamma_n = E(|X| \wedge a_n)\operatorname{sgn}(X). \tag{1.7}$$

Our study then concentrates on the corresponding empirical versions $\hat{a}_n$ and $\hat{\gamma}_n$ of $a_n$ and $\gamma_n$ defined through the empirical distribution function by

$$\begin{aligned} r_n &= \frac{1}{\hat{a}_n^2}\sum_{j=1}^{n}\left(X_j^2 \wedge \hat{a}_n^2\right), \\ \hat{\gamma}_n &= \frac{1}{n}\sum_{j=1}^{n}(|X_j| \wedge \hat{a}_n)\operatorname{sgn}(X_j). \end{aligned} \tag{1.8}$$

Since $\hat{\gamma}_n$ is defined through $\hat{a}_n$, a detailed study of $\hat{\gamma}_n$ entails a detailed study of the joint distribution of $(\hat{\gamma}_n, \hat{a}_n)$.

If $r_n$ is bounded, there is typically no net improvement with respect to asymptotic normality. [Compare Mori (1984) and Maller (1982).] Hence we must consider $r_n \to \infty$. Furthermore, in order to study scale and centrality quantities which in the limit involve the entire distribution, it is necessary that $\lim_{n\to\infty} r_n/n = 0$. [However, the case $\lim_{n\to\infty} r_n/n = c > 0$ á la Stigler (1973), is also of interest and is being considered elsewhere.] Moreover, $r_n$ can be interpreted as a bound on the number of excessive terms which require modification [see Remark 3.5(ii)]. It is worth noting that the tail equation or truncated second moment equation analogues of (1.6) would serve as the basis of two other approaches. The first corresponds to that of trimmed or winsorized sums when the summands are ranked by magnitude [see, e.g., Griffin and Pruitt (1987) and Hahn, Kuelbs and Weiner (1989a)] while the latter truncated sums correspond to the approach taken in Hahn and Kuelbs (1988). For further discussion and comparison see Hahn, Kuelbs and Weiner (1989c).

The purpose of this paper is to pursue this alternative approach using the censoring equation analogue (1.6) of (1.2c). It rectifies most of the previously mentioned problems. The censoring equation has been selected for three reasons. First, the normalizations determined by the tail equation (1.2a), unlike those determined by (1.2b) and (1.2c), are not suitable for normalizing the full sums for variables in the domain of partial attraction of the normal. Second, the approach based on the truncated second moment equation (1.2b) encounters serious analytic difficulties in its mathematical implementation. Finally, the equations defining the parameters and their empirical versions in the censored case are analytically nice enough to allow for a complete set of self-normalized results.

Some of the analytic techniques used in the paper resemble or were inspired by Griffin and Pruitt (1987), (1989), while the stochastic integral representations used throughout the paper were certainly influenced and inspired by work in, for example, Csörgő, Horváth and Mason (1986) and Csörgő, Haeusler and Mason (1988a), (1988b). Our joint approach has precedent in the work of Davis and Resnick (1984).

*Organization and results.* The major results of this paper are Theorem 5.1 and Theorem 5.50. Theorem 5.1 gives the joint asymptotic distribution of $(\hat{\gamma}_n, \hat{a}_n)$. Theorem 5.50 gives precise criteria for convergence to each of the limit laws allowed by Theorem 5.1.

The censoring distribution parameters of scale and center are defined in Section 2. This requires the introduction of some auxiliary functions and notation to be used throughout the remainder of the paper. The analogous empirical versions are discussed in Section 3. If $F$ is symmetric about the origin, Theorem 3.6 establishes the universal asymptotic normality of the self-normalized or studentized quantity $n\hat{\gamma}_n/(\sqrt{r_n}\,\hat{a}_n)$. The symmetry assumption cannot be relaxed in general by merely considering $n(\hat{\gamma}_n - \gamma_n)/(\sqrt{r_n}\,\hat{a}_n)$. (See the paragraph before Corollary 5.70.) One goal of this paper is to identify when this assumption can be relaxed.

Since the centerings are determined by the normalizers (or scale constants), the latter must be studied first. The heart of the paper is Section 4 which gives the basic probabilistic results and a fairly complete detailed study of the empirical scale quantities. In particular, Theorem 4.7 characterizes both tightness and stochastic compactness of the properly normalized empirical scales

$$\sqrt{r_{n_k}}\left(\frac{\hat{a}_{n_k}}{a_{n_k}} - 1\right), \tag{1.9}$$

for any subsequence of integers $n_k \to \infty$. All possible subsequential limiting distributions are specified. Further, Proposition 4.34 determines precisely when there is subsequential convergence to a specific limiting distribution. Necessary and sufficient conditions for the quantities in (1.9) to be tight (or stochastically compact) with only mean zero normal limits are given in Proposition 4.35. Finally, Corollary 4.29 determines that tightness (respectively, stochastic compactness) of the quantities in (1.9) holds universally for every sequence $r_n \to \infty$ with $r_n/n \to 0$ and any $n_k \to \infty$ if and only if $X$ is in the Feller class (resp., $X$ is in the Feller class but outside the domain of partial attraction of the normal).

Section 5 focuses on the self-normalized results. A fairly complete theory of the joint asymptotic behavior of $\hat{\gamma}_n$ and $\hat{a}_n$ is provided. Theorem 5.1 includes a characterization of tightness of the properly normalized vectors, a determination of the form of all subsequential limits, conditions for subsequential convergence to a given limit law of allowable type whose support is not all of $\mathbf{R}^2$ and precisely when the support of the limit laws is a curve in $\mathbf{R}^2$ rather than all of $\mathbf{R}^2$. Crucial to the proof of this theorem is Proposition 4.3, of which Theorem 5.1 can be regarded as an empiricalization. General necessary and sufficient conditions for convergence to a given limit law are provided in Theorem 5.50. A characterization of tightness and stochastic compactness with only mean zero bivariate normal limits appears in Corollary 5.59, while Corollary 5.64 provides necessary and sufficient conditions for convergence to a bivariate normal limit. The marginal behavior of $\hat{\gamma}_n$ is finally deduced in Corollaries 5.68 and 5.70. Empirical determination of the limiting covariance matrix is achieved in Theorem 5.81.

The final section of the paper features applications and examples. For variables in the domain of attraction of a stable law, joint asymptotic normality (with specified limiting covariance) always occurs and other results greatly simplify. The Feller class [where (4.27) holds] is also considered. A broad class of tractable examples inside the Feller class is introduced which leads to nonsingular joint asymptotic normality. Finally, an example is included where even simple consistency, $\hat{a}_n/a_n \to 1$ in probability, fails.

## 2. Notation and preliminaries.

*Auxiliary functions and properties.* Let $X$ be a nondegenerate random variable with cumulative distribution function $F$. Fix $t > 0$. The joint and one-sided tails of $X$ will be denoted, respectively, by

$$G(t) \equiv P(|X| > t), \qquad G^+(t) \equiv P(X > t) \quad \text{and} \quad G^-(t) \equiv P(X < -t).$$

The $p$th moments of truncated and censored random variables for any positive integer $p$ will be designated by

$$\begin{aligned} \tilde{M}(p,t) &\equiv E(X^p I(|X| \le t)), \\ M(p,t) &\equiv E((|X| \wedge t)\operatorname{sgn}(X))^p \\ &= \begin{cases} \tilde{M}(p,t) + t^p G(t), & \text{if } p \text{ is even}, \\ \tilde{M}(p,t) + t^p\{G^+(t) - G^-(t)\}, & \text{if } p \text{ is odd}. \end{cases} \end{aligned} \tag{2.1}$$

It will also be convenient to have a notation for the above quantities normalized by a relevant power of $t$,

$$(2.2)\qquad \begin{aligned} \tilde{m}(p,t) &\equiv t^{-p}\tilde{M}(p,t), \\ m(p,t) &\equiv t^{-p}M(p,t) \\ &= \begin{cases} \tilde{m}(p,t) + G(t), & \text{if } p \text{ is even}, \\ \tilde{m}(p,t) + \{G^+(t) - G^-(t)\}, & \text{if } p \text{ is odd}. \end{cases} \end{aligned}$$

Recall that integration by parts allows the $p$th moment to be related back to the joint tail or one-sided tails as

$$(2.3)\qquad M(p,t) = \begin{cases} \int_0^t py^{p-1}G(y)\,dy, & \text{if } p \text{ is even}, \\ \int_0^t py^{p-1}\{G^+(y) - G^-(y)\}\,dy, & \text{if } p \text{ is odd}. \end{cases}$$

Write $M(p,t) = t^p m(p,t)$. By (2.3) and the product rule of differentiation, off a countable set,

$$t^p\frac{\partial}{\partial t}m(p,t) = \begin{cases} pt^{p-1}(G(t) - m(p,t)), & \text{if } p \text{ is even}, \\ pt^{p-1}(G^+(t) - G^-(t) - m(p,t)), & \text{if } p \text{ is odd}. \end{cases}$$

Recalling (2.2), it follows that for that $a > 0$,

$$m(p,t) = m(p,a) - p\int_a^t \frac{\tilde{m}(p,y)}{y}\,dy.$$

Letting $a \downarrow 0$, using the monotone convergence theorem on the positive and negative parts of $\tilde{m}(p,\cdot)$ and using dominated convergence to evaluate $m(p,0) = m(p,0+)$ yields

$$(2.4)\qquad m(p,t) = \begin{cases} P(X \neq 0) - p\int_0^t \frac{\tilde{m}(p,y)}{y}\,dy, & \text{if } p \text{ is even}, \\ P(X > 0) - P(X < 0) - p\int_0^t \frac{\tilde{m}(p,y)}{y}\,dy, & \text{if } p \text{ is odd}. \end{cases}$$

Now let $\Delta = \Delta_F \equiv \inf\{y > 0\colon G(y) < P(X \neq 0)\}$. Then $m(2,s) = P(X \neq 0)$ for $0 \le s \le \Delta$. Moreover, since $\tilde{m}(2,t) > 0$ for all $t > \Delta$, it is clear that $m(2,\cdot)$ decreases strictly and continuously on $(\Delta,\infty)$. Dominated convergence implies that $\lim_{t\to\infty} m(2,t) = 0$ while, in fact, $m(2,t) > 0$ for every $t \ge 0$. Consequently, $m(2,\cdot)$ has a well-defined, continuous inverse on $(0, P(X \neq 0))$.

A final observation that will be used several times is that

$$(2.5)\qquad \text{if } z_n > y_n,\ z_n \sim y_n, \quad \text{then } m(2,z_n) \sim m(2,y_n).$$

This follows because both

$$m(2,z_n) \le m(2,y_n)$$

and

$$m(2, z_n) = z_n^{-2} M(2, z_n) \geq z_n^{-2} M(2, y_n) = \left( \frac{y_n}{z_n} \right)^2 m(2, y_n) \sim m(2, y_n).$$

Similar reasoning yields the analogue of (2.5) for $m(p, \cdot\,)$ when $p \geq 2$ is even.

*Center and scale constants.* Fix a real sequence $\{r_n\}$ satisfying

$$0 < r_n \to \infty, \qquad \frac{r_n}{n} \to 0. \tag{2.6}$$

Since $X$ is assumed to be nondegenerate, $P(X \neq 0) > 0$. Thus, by the discussion following (2.4), there is, for each $n$ sufficiently large, a uniquely defined distribution scale parameter $a_n$ given implicitly by the equation

$$nm(2, a_n) = r_n. \tag{2.7}$$

Alternatively, if $d$ denotes the inverse for $m(2, \cdot\,)$, then $a_n = d(r_n/n)$. Recall that $d(1/n)$ provides classical scalings in the context of the usual central limit theorem for sums. One should view $r_n$ as a bound on the expected number of terms of excessive magnitude to be modified when censoring. [See Remark 3.5(ii).] These scale parameters $a_n$ satisfy

$$r_n = nG(a_n) + n\tilde{m}(2, a_n) \geq nG(a_n -) \geq nG(a_n). \tag{2.8}$$

Furthermore, the scale parameters can be used to define the centering constants $\gamma_n$ by

$$\gamma_n = M(1, a_n) = E((|X| \wedge a_n)\operatorname{sgn}(X)). \tag{2.9}$$

Because $r_n/n \to 0$ we also have $a_n \to \infty$. Thus, if $E|X| < \infty$, $\gamma_n \to EX$. If $EX^2 < \infty$, the additional relation $a_n^2 \sim (n/r_n)EX^2$ clearly holds.

**3. Self-normalization and universal asymptotic normality for symmetric distributions.** The underlying c.d.f., $F$, is often approximated via the empirical distribution function (e.d.f.) $F_n = n^{-1}\sum_{j=1}^n \delta_{X_j}$. The empirical versions of the probabilities and integrals introduced in (2.1) and (2.2) will be designated with a subscript $n$. In particular, the three quantities of most interest for self-normalized results are

$$\begin{aligned} nG_n(t) &= \#\big(\{j \leq n : |X_j| > t\}\big), \\ nM_n(1, t) &= \sum_{j=1}^n (|X_j| \wedge t)\operatorname{sgn}(X_j), \\ nm_n(2, t) &= \sum_{j=1}^n \frac{X_j^2 \wedge t^2}{t^2}. \end{aligned} \tag{3.1}$$

Fix a sequence $\{r_n\}$ as in (2.6). Then, empiricalizing (2.7) and (2.9) leads to

$$nm_n(2, \hat{a}_n) = r_n \tag{3.2}$$

and

$$\hat{\gamma}_n = M_n(1, \hat{a}_n). \tag{3.3}$$

The empirical version $\hat{a}_n$ of the scale constant $a_n$ will, for convenience, be termed the *scale estimator* throughout this article. It is natural to call $\hat{\gamma}_n$ the *empirically censored mean* and

$$n\hat{\gamma}_n = \sum_{j=1}^{n} (|X_j| \wedge \hat{a}_n)\operatorname{sgn}(X_j)$$

the *empirically censored sum*. Technically, $n\hat{\gamma}_n$ is not a winsorized sum because the censoring levels $\hat{a}_n$ are not order statistics. However, $\hat{\gamma}_n$ can be computed directly from knowledge of the summands ordered by magnitudes. We are interested in the self-normalized quantity

$$\frac{n(\hat{\gamma}_n - \gamma_n)}{\sqrt{r_n}\,\hat{a}_n} = \frac{\sum_{j=1}^{n}\{(|X_j| \wedge \hat{a}_n)\operatorname{sgn}(X_j) - \gamma_n\}}{\left(\sum_{j=1}^{n}(X_j^2 \wedge \hat{a}_n^2)\right)^{1/2}}.$$

First, it needs to be shown that (3.2) uniquely defines the estimator $\hat{a}_n$. Let $\Delta_n = \Delta_{F_n}$ [see the statement following (2.4)] and notice that $m_n(2, \Delta_n) = G_n(0)$ while $m_n(2, \cdot\,)$ decreases strictly and continuously to zero on $(\Delta_n, \infty)$. Then (3.2) can be solved uniquely for a positive value of $\hat{a}_n$ if and only if $G_n(0) > r_n/n$. For definiteness, define $\hat{a}_n = 0$ on the event $[G_n(0) \le r_n/n]$. We will show that

$$\hat{a}_n \to \infty \quad \text{a.s.} \tag{3.4}$$

In particular, $\hat{a}_n$ is almost surely eventually positive and hence uniquely defined by (3.2).

Let $C > 0$. By the strong law of large numbers,

$$m_n(2, C) = C^{-2}\frac{1}{n}\sum_{j=1}^{n} X_j^2 \wedge C^2 \to m(2, C) > 0 \quad \text{a.s.}$$

By monotonicity of $m_n(2, \cdot\,)$, $\hat{a}_n \le C$ implies $m_n(2, C) \le r_n/n$. Since $r_n/n \to 0$, it follows that

$$P(\hat{a}_n \le C \text{ infinitely often}) \le P\left(m_n(2, C) \le \tfrac{1}{2}m(2, C) \text{ infinitely often}\right) = 0.$$

[Note that $G_n(0) = m_n(2, 0) \ge m_n(2, C) > \frac{1}{2}m(2, C) > 0$ for all sufficiently large $n$, almost surely, so that in particular $\hat{a}_n \ge C > 0$ is uniquely defined.]

3.5 Remarks. (i) If $P(X = 0) = 0$, then $G_n(0) \doteq 1$ a.s. In this case, it is enough to require that $r_n/n < 1$ in order to insure that for each $n \ge 1$, $\hat{a}_n$ is strictly positive and uniquely defined by (3.2). In particular, this is the case if $F$ is continuous.

(ii) Let $Y_{n1} \ge Y_{n2} \ge \cdots \ge Y_{nn} \ge 0$ denote the order statistics for the sample of magnitudes $\{|X_j|: j \le n\}$ arranged in descending order, breaking ties by priority of index (or any other method). Now $nG_n(Y_{nj}-) \ge j$. Also, by continuity and (3.2), the analogue of (2.8) holds,

$$\#\{j \le n: |X_j| \ge \hat{a}_n\} = nG_n(\hat{a}_n-) \le nm_n(2, \hat{a}_n-) = nm_n(2, \hat{a}_n) = r_n,$$

on the event $[\hat{a}_n > 0]$. In particular, $\hat{a}_n \ge Y_{nr_n}$.

If $F$ is assumed to be nondegenerate and symmetric about the origin, the empirically censored sums when self-normalized can be shown to always be asymptotically normal. [Note that under symmetry, $\gamma_n \equiv 0$ in (2.9).] The universality of the following asymptotic normality result is suggestive of the relative insensitivity of the censoring method to heaviness of the tail of the distribution.

3.6. THEOREM. *Assume $F$ is nondegenerate and symmetric about the origin and $\{r_n\}$ satisfies* (2.6). *Then*

$$\mathscr{L}\left(\frac{n\hat{\gamma}_n}{\sqrt{r_n}\,\hat{a}_n}\right) = \mathscr{L}\left(\frac{n}{\sqrt{r_n}}m_n(1, \hat{a}_n)\right)$$
$$= \mathscr{L}\left(\frac{\sum_{j=1}^n(|X_j| \wedge \hat{a}_n)\operatorname{sgn}(X_j)}{\sqrt{\sum_{j=1}^n\left(X_j^2 \wedge \hat{a}_n^2\right)}}\right) \to \mathscr{N}(0,1).$$

PROOF. Assume $\{X_j: j \ge 1\}$ are defined on the probability space $(\Omega, \mathscr{F}, P)$. Let $\{\varepsilon_j\}$ be i.i.d. on a space $(\Omega', \mathscr{F}', P')$ with $P(\varepsilon_j = 1) = P(\varepsilon_j = -1) = \frac{1}{2}$. Let $\Omega'' = \Omega \times \Omega'$, $\mathscr{F}'' = \mathscr{F} \times \mathscr{F}'$, $P'' = P \times P'$. Since $\hat{a}_n$ is a function of $\{|X_1|, \dots, |X_n|\}$ alone and $X$ is symmetric, $\mathscr{L}(\varepsilon_1|X_1|, \varepsilon_2|X_2|, \dots, \varepsilon_n|X_n|, \hat{a}_n) = \mathscr{L}(X_1, X_2, \dots, X_n, \hat{a}_n)$. Let

$$A = \liminf_{n\to\infty}[\hat{a}_n > 0] = [\hat{a}_n > 0 \text{ for all sufficiently large } n].$$

Then $A \in \mathscr{F}$ and, by (3.4), $P(A) = 1$. [We recall, by (2.4), that $(n/\sqrt{r_n})m_n(1, \hat{a}_n)$ is defined by continuity even at $\hat{a}_n = 0$, if we interpret it as

$$\sqrt{n}\left((m_n(1, \hat{a}_n))/\sqrt{m_n(2, \hat{a}_n)}\right)$$

when $\hat{a}_n = 0$.]

We will show that for every $t$,

$$\phi_n(t) \equiv E\exp\left\{it\sum_{j=1}^n \frac{|X_j| \wedge \hat{a}_n}{\sqrt{r_n}\,\hat{a}_n}\operatorname{sgn}(X_j)\right\} \to e^{-t^2/2} \quad \text{as } n \to \infty.$$

Fix $t$. For $\omega \in \Omega$, consider

$$\lambda_n(\omega) = \int_{\Omega'}\exp\left\{it\sum_{j=1}^n\left(\frac{|X_j(\omega)| \wedge \hat{a}_n(\omega)}{\sqrt{r_n}\,\hat{a}_n(\omega)}\varepsilon_j(\omega')\right)\right\}dP'(\omega').$$

Then

$$\phi_n(t) = \int_\Omega \lambda_n(\omega)\,dP(\omega) = \int_A \lambda_n(\omega)\,dP(\omega).$$

Once it is shown that $\lambda_n(\omega) \to e^{-t^2/2}$ for every $\omega \in A$, an application of the bounded convergence theorem completes the proof.

Fix $\omega \in A$ and let $b_{nj}(\omega) = 1/(\sqrt{r_n}\,\hat{a}_n(\omega))(|X_j(\omega)| \wedge \hat{a}_n(\omega))$. By (3.2), since $\omega \in A$, for all sufficiently large $n$, $0 \le b_{nj}(\omega) \le 1/\sqrt{r_n} \to 0$ and $\Sigma_{j=1}^n b_{nj}^2(\omega) = 1$. Therefore,

$$\lambda_n(\omega) = \int_{\Omega'} \exp\left\{it \sum_{j=1}^{n} b_{nj}(\omega)\varepsilon_j(\omega')\right\} dP'(\omega') \equiv E\exp\{itT_n(\omega)\},$$

where, $\omega$ having been fixed, we may regard $T_n(\omega)$: $(\Omega', \mathscr{F}', P') \to \mathbf{R}$ as the row sum of the rowwise $P'$-independent triangular array $\{b_{nj}(\omega)\varepsilon_j(\omega')$: $j \le n$, $n \ge 1\}$. Since $\{\varepsilon_j\}$ is a Rademacher sequence on $(\Omega', \mathscr{F}', P')$, it is immediate that for each fixed $\omega \in A$, $\mathscr{L}(T_n(\omega)) \to \mathscr{N}(0,1)$. Thus, $\lambda_n(\omega) \to e^{-t^2/2}$ for each $\omega \in A$ and the proof is complete. □

**4. Limit theorems: Deterministic censoring/normalization; the scale estimator.** *Fix a sequence* $\{r_n\}$ *of positive numbers such that*

$$r_n \to \infty \quad \text{and} \quad r_n/n \to 0, \tag{4.1}$$

*unless otherwise noted* (specifically, in Corollary 4.29 and Example 4.40). Naturally, the asymptotic properties of $\hat{a}_n$ and $\hat{\gamma}_n$ will depend in part on the asymptotic behavior of the corresponding parameters $a_n$ and $\gamma_n$. Recall that eventually $\hat{a}_n$ and $\hat{\gamma}_n$ are determined by the equations

$$\hat{\gamma}_n = M_n(1, \hat{a}_n) \quad \text{and} \quad \hat{a}_n^2 = \frac{n}{r_n} M_n(2, \hat{a}_n).$$

The joint behavior of the empirically censored quantities $M_n(1, \hat{a}_n)$ and $(n/r_n)M_n(2, \hat{a}_n)$ will be derived from the joint behavior of the corresponding deterministically censored sums $M_n(1, a_n)$ and sums of squares $(n/r_n)M_n(2, a_n)$. The following proposition examines the joint asymptotic behavior of these latter quantities in complete generality. Following its proof we turn our attention to the properties of the scale estimator $\hat{a}_n$. The joint asymptotic behavior of the pair $(\hat{\gamma}_n, \hat{a}_n)$, which is our major objective in this paper, will be considered in Section 5, in particular in our main result, Theorem 5.1. Let

$$a^2(X) = \begin{cases} \operatorname{Var}(X)/EX^2, & \text{if } EX^2 < \infty, \\ 1, & \text{if } EX^2 = \infty. \end{cases} \tag{4.2}$$

4.3. Proposition. *The sequence* $\{(n/\sqrt{r_n})(m_n(1, a_n) - m(1, a_n), m_n(2, a_n) - m(2, a_n))\}$ *is tight in* $\mathbf{R}^2$ *and every subsequential limit is mean zero bivariate normal with (possibly degenerate) covariance matrix of the form*

$$\Sigma = \begin{pmatrix} a^2(X) & b \\ b & c^2 \end{pmatrix}, \tag{4.4}$$

*where*

$$\begin{cases} 0 < a^2(X) \le 1 \text{ and } b = c^2 = 0, & \text{if } EX^2 < \infty, \\ a^2(X) = 1, 0 \le c^2 \le 1 \text{ and } b^2 \le c^2, & \text{if } EX^2 = \infty. \end{cases}$$

*Convergence in distribution occurs along subsequences for which both*

$$\begin{aligned} &\frac{n}{r_n} m(3, a_n) \to b, \\ &\frac{n}{r_n} m(4, a_n) \to c^2. \end{aligned} \tag{4.5}$$

*Finally, if $X$ is asymptotically symmetric, i.e.,*

$$\lim_{t \to \infty} \frac{G^+(t) - G^-(t)}{G(t)} = 0, \tag{4.6}$$

*then for every subsequential limit law, $b = 0$.*

PROOF. Using the Cramér–Wold device, consider for $u$ and $v$ in $\mathbf{R}$,

$$\begin{aligned} &\frac{n}{\sqrt{r_n}}\big(u(m_n(1, a_n) - m(1, a_n)) + v(m_n(2, a_n) - m(2, a_n))\big) \\ &\quad = \sum_{j=1}^{n} \left( u \frac{(|X_j| \wedge a_n)\operatorname{sgn}(X_j) - E\big((|X_j| \wedge a_n)\operatorname{sgn}(X_j)\big)}{a_n \sqrt{r_n}} \right. \\ &\qquad\qquad \left. + v \frac{\big(X_j^2 \wedge a_n^2\big) - E\big(X_j^2 \wedge a_n^2\big)}{a_n^2 \sqrt{r_n}} \right) \\ &\quad \equiv \sum_{j=1}^{n} y_{nj}. \end{aligned}$$

The rowwise i.i.d. triangular array $\{y_{nj}: j \le n\}$ is infinitesimal and the only possible limits are normal since

$$|y_{nj}| \le \frac{2u}{\sqrt{r_n}} + \frac{2v}{\sqrt{r_n}} \to 0 \quad \text{as } n \to \infty.$$

Moreover $Ey_{nj} = 0$, so the weak convergence will follow upon establishing convergence of the variances. Now

$$\begin{aligned} n \operatorname{Var}(y_{nj}) &= u^2 \frac{n}{a_n^2 r_n} \operatorname{Var}((|X| \wedge a_n)\operatorname{sgn}(X)) + v^2 \frac{n}{a_n^4 r_n} \operatorname{Var}\big(X^2 \wedge a_n^2\big) \\ &\quad + \frac{2uv}{a_n^3 r_n} n \operatorname{Cov}\big((|X| \wedge a_n)\operatorname{sgn}(X), X^2 \wedge a_n^2\big). \end{aligned}$$

First assume that $EX^2 < \infty$. Then $a_n^2 \sim (n/r_n)EX^2$, while $\mathrm{Var}((|X| \wedge a_n)\mathrm{sgn}(X)) \to \mathrm{Var}\, X$. Moreover, if $EX^4 < \infty$, notice that

$$\frac{n}{a_n^4 r_n}\mathrm{Var}\left(X^2 \wedge a_n^2\right) \le \frac{n}{a_n^4 r_n}E\left(X^4 \wedge a_n^4\right) \sim \frac{n}{a_n^4 r_n}EX^4$$
$$= O\left(\frac{n}{(n/r_n)^2 r_n}\right) = O\left(\frac{r_n}{n}\right) = o(1).$$

But if $EX^4 = \infty$, then $EX^2 < \infty$ implies that

$$\frac{n}{a_n^4 r_n}E\left(X^4 \wedge a_n^4\right) = \frac{n}{a_n^4 r_n}\int_0^{a_n} 4y^3 G(y)\,dy \quad \text{[by (2.3)]}$$
$$= \frac{n}{a_n^4 r_n} o\left(\int_0^{a_n} 4yM(2,y)\,dy\right)$$
$$= o\left(\frac{n}{a_n^4 r_n}a_n^2\right) = o\left(\frac{n}{a_n^2 r_n}\right) = o(1).$$

The second equality utilizes the fact that $EX^2 < \infty$ implies

$$\limsup_{t\to\infty}\left(G(t)/m(2,t)\right) = 0,$$

so that $y^2 G(y) = o(M(2,y))$. The covariance terms now tend to zero by application of the Cauchy–Schwarz inequality.

Thus, without loss of generality, it may be assumed that $EX^2 = \infty$. In this case both

$$\frac{n}{a_n^2 r_n}\mathrm{Var}((|X| \wedge a_n)\mathrm{sgn}(X)) \sim \frac{n}{a_n^2 r_n}E\left(X^2 \wedge a_n^2\right) = \frac{n}{r_n}m(2,a_n) = 1$$

and

$$\frac{n}{a_n^4 r_n}\mathrm{Var}\left(|X|^2 \wedge a_n^2\right) \sim \frac{n}{a_n^4 r_n}E\left(\left(|X|^2 \wedge a_n^2\right)^2\right) = \frac{n}{r_n}m(4,a_n) \le \frac{n}{r_n}m(2,a_n) = 1,$$

using (2.7). By selecting subsequences it may be assumed that

$$\left(n_k/r_{n_k}\right)m\left(4,a_{n_k}\right) \to c^2 \le 1.$$

Turning to the covariance term,

$$\frac{n}{a_n^3 r_n}\mathrm{Cov}\left((|X| \wedge a_n)\mathrm{sgn}(X), X^2 \wedge a_n^2\right)$$
$$= \frac{n}{a_n^3 r_n}\left(E\left(\left(|X|^3 \wedge a_n^3\right)\mathrm{sgn}(X)\right) - E((|X| \wedge a_n)\mathrm{sgn}(X))E\left(X^2 \wedge a_n^2\right)\right)$$
$$= \frac{n}{r_n}\left(m(3,a_n) - m(1,a_n)m(2,a_n)\right).$$

Notice that $(n/r_n)m(1,a_n)m(2,a_n) = m(1,a_n) \to 0$, whereas

$$\left|\frac{n}{r_n}m(3,a_n)\right| \le \frac{n}{r_n}a_n^{-3}E\left(|X|^3 \wedge a_n^3\right) \le \frac{n}{r_n}m(2,a_n) = 1.$$

Thus, by selecting further subsequences, it can be assumed that $(n'_k/r_{n'_k})m(3, a_{n'_k}) \to b$ where $b^2 \le c^2$, by Cauchy–Schwarz. The form of (4.4) and subsequential mean zero asymptotic normality are thereby verified.

Finally, if $E|X|^3 < \infty$, then $EX^2 < \infty$ and so $b = 0$. But if $E|X|^3 = \infty$, condition (4.6) together with (2.3) implies that

$$\begin{aligned}\frac{n}{r_n}m(3, a_n) &= \frac{n}{r_n a_n^3}\int_0^{a_n} 3y^2\{G^+(y) - G^-(y)\}\, dy\\ &= \frac{n}{r_n a_n^3} o\left(\int_0^{a_n} 3y^2 G(y)\, dy\right)\\ &= o\left(\frac{n}{r_n a_n^3} E\left(|X|^3 \wedge a_n^3\right)\right) = o\left(\frac{n}{r_n} m(2, a_n)\right) = o(1). \qquad \square\end{aligned}$$

The remainder of this section will be devoted to the behavior of the scale estimator $\hat{a}_n$. As we will see, to relate $\hat{\gamma}_n$ to $M_n(1, a_n)$, it will be necessary in general to have much more than simple consistency of $\hat{a}_n$, i.e., $\hat{a}_n/a_n \to_p 1$. Specifically, we will require that $\{\sqrt{r_n}(\hat{a}_n/a_n - 1)\}$ be tight. The following theorem provides the necessary and sufficient analytic condition for this tightness and also provides the criterion for simple consistency, along a given subsequence. Additionally, a description of the resulting subsequential limit laws is given. After the proof, this analytic condition will be discussed further and a very general but much more intuitive and classical sufficient condition will be provided. Finally, asymptotic normality and precise conditions for convergence to a given limit law will be considered.

Recall that a sequence is *stochastically compact* if it is tight and every subsequential limit is nondegenerate (i.e., not concentrated at a point).

Most of the following results concern subsequences. This allows not only greater generality for determining the precise criteria for convergence of the appropriate variables, but also increases the flexibility for application of these results here and elsewhere. *Thus, fix a sequence of integers $n_k \to \infty$ unless otherwise noted* (specifically in Corollary 4.29 and Example 4.40).

4.7. THEOREM. *Given $0 < \rho_{n_k} \to \infty$ such that $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$, the following two conditions are equivalent*:

$$P\left(\left|\frac{\hat{a}_{n_k} - a_{n_k}}{a_{n_k}}\sqrt{r_{n_k}}\right| \ge \rho_{n_k}\right) \to 0, \tag{4.8}$$

$$\lim_{k\to\infty} \frac{n_k}{\sqrt{r_{n_k}}}\left\{m\left(2, a_{n_k}\left(1 - \frac{\rho_{n_k}}{\sqrt{r_{n_k}}}\right)\right) - m(2, a_{n_k})\right\} = \infty. \tag{4.9}$$

*In particular,* $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ *is tight if and only if* (4.9) *holds for* *every* $\rho_{n_k} \to \infty$ *such that* $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$, *while* $\hat{a}_{n_k}/a_{n_k} \to 1$ *in probability if and only if* (4.9) *holds for* *some* $\rho_{n_k} \to \infty$ *such that* $\rho_{n_k}/\sqrt{r_{n_\kappa}} \to 0$. *Every subsequential limiting distribution of* $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ *has the form*

$$\mathscr{L}(\Psi^{-1}(Z)), \tag{4.10}$$

*where* $Z \sim \mathscr{N}(0, c^2)$ *with* $0 \le c^2 \le 1$ *as in* (4.5) *and* $\Psi$ *is convex, strictly increasing with* $\Psi(0) = 0$, $\Psi(\infty) = \infty$, $\Psi(-\infty) = -\infty$ *and* $2(1 - c^2) \le \Psi' \le 2$. *The sequence* $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ *possesses no degenerate subsequential limit laws if and only if*

$$\liminf_{k\to\infty} \frac{n_k}{r_{n_k}} m(4, a_{n_k}) > 0. \tag{4.11}$$

*The sequence* $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ *is stochastically compact if and only if* (4.11) *holds and* (4.9) *holds whenever* $\rho_{n_k} \to \infty$ *and* $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$.

PROOF. As the proof is rather lengthy and quite technical, it is divided into several parts.

STEP 1: General facts. Here are derived certain general asymptotic identities which will provide the key to the analysis of the scale estimator $\hat{a}_n$. These culminate in (4.16).

A fundamental consequence of (3.2), (2.7) and (2.4) is the following identity which holds on the event $[\hat{a}_n > 0]$ [and thus which holds, due to (3.4), with probability tending to 1]:

$$\begin{aligned}
B_n &\equiv \frac{n}{\sqrt{r_n}}(m_n(2, a_n) - m(2, a_n)) \\
&= \frac{n}{\sqrt{r_n}}\left(m_n(2, a_n) - \frac{r_n}{n}\right) \\
&= \frac{n}{\sqrt{r_n}}(m_n(2, a_n) - m_n(2, \hat{a}_n)) \qquad (4.12) \\
&= \frac{2n}{\sqrt{r_n}}\int_{a_n}^{\hat{a}_n} \frac{\tilde{m}_n(2, s)}{s}\, ds \\
&= \frac{2n}{r_n}\int_0^{((\hat{a}_n - a_n)/a_n)\sqrt{r_n}} \tilde{m}_n\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\left(1 + \frac{u}{\sqrt{r_n}}\right)^{-1} du,
\end{aligned}$$

where the last line employs the change of variable $u = ((s - a_n)/a_n)\sqrt{r_n}$ and we interpret $\int_\alpha^\beta = -\int_\beta^\alpha$ when $\beta < \alpha$.

Analysis can proceed because the behavior of $\{B_n\}$ is completely determined by Proposition 4.3. In all that follows, we can and do restrict to the event

$[\hat{a}_n > 0]$. Because all claims will be made at the level of weak convergence, no loss of generality results, due to $P(\hat{a}_n > 0) \to 1$ via (3.4).

Fix $0 < \rho_n \to \infty$, such that $\rho_n = o(\sqrt{r_n})$. Now clearly with probability tending to 1, $\tilde{m}_n(2, \cdot\,) > 0$ on $(a_n(1 - \rho_n/\sqrt{r_n}), \infty)$, so on this ray, the rightmost member of (4.12) is (with probability tending to one) a strictly increasing function of $(\hat{a}_n - a_n/\hat{a}_n)\sqrt{r_n}$. Thus, (4.12) shows that (again with probability tending to 1) $((\hat{a}_n - a_n)/a_n)\sqrt{r_n} \geq \rho_n$ if and only if

$$\begin{aligned}
B_n &\geq \frac{2n}{r_n}\int_0^{\rho_n} \tilde{m}_n\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\frac{du}{1 + u/\sqrt{r_n}} \\
&= \frac{2n}{r_n}\int_0^{\rho_n} \tilde{m}\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\frac{du}{1 + u/\sqrt{r_n}} \\
&\quad + \frac{2n}{r_n}\int_0^{\rho_n}\left\{\tilde{m}_n\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\right. \\
&\qquad \left. - \tilde{m}\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\right\}\frac{du}{1 + u/\sqrt{r_n}} \qquad (4.13)\\
&= \frac{n}{\sqrt{r_n}}\left\{m(2, a_n) - m\left(2, a_n\left(1 + \frac{\rho_n}{\sqrt{r_n}}\right)\right)\right\} \\
&\quad + \frac{2n}{r_n}\int_0^{\rho_n}\left\{\tilde{m}_n\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\right. \\
&\qquad \left. - \tilde{m}\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\right\}\frac{du}{1 + u/\sqrt{r_n}} \\
&\equiv \tau_n + T_n.
\end{aligned}$$

Similarly, $((\hat{a}_n - a_n)/a_n)\sqrt{r_n} \leq -\rho_n$ if and only if

$$\begin{aligned}
B_n &\leq \frac{n}{\sqrt{r_n}}\left\{m(2, a_n) - m\left(2, a_n\left(1 - \frac{\rho_n}{\sqrt{r_n}}\right)\right)\right\} \\
&\quad - \frac{2n}{r_n}\int_{-\rho_n}^{0}\left\{\tilde{m}_n\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\right. \qquad (4.14)\\
&\qquad \left. - \tilde{m}\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\right\}\frac{du}{1 + u/\sqrt{r_n}} \\
&\equiv -v_n + U_n.
\end{aligned}$$

We will show that $|T_n| + |U_n| \to_p 0$, which reveals the relative dominance of the terms $\tau_n$ and $v_n$ introduced in (4.13)–(4.14). Noting that $(1 \pm \rho_n/\sqrt{r_n})^{-1} \to 1$, for large $n$,

$$
\begin{aligned}
E(|T_n| + |U_n|) &\le \frac{2n}{r_n} E\left( \int_{-\rho_n}^{\rho_n} \left| \tilde{m}_n\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right) \right.\right. \\
&\qquad \left.\left. - \tilde{m}\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\right| \left(1 + \frac{u}{\sqrt{r_n}}\right)^{-1} du \right) \\
&= \frac{2n}{r_n} \int_{-\rho_n}^{\rho_n} E\left| \tilde{m}_n\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right) \right. \\
&\qquad \left. - \tilde{m}\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\right| \left(1 + \frac{u}{\sqrt{r_n}}\right)^{-1} du \\
&\qquad\qquad \text{(by Fubini's theorem)} \\
&\le \frac{2n}{r_n} \int_{-\rho_n}^{\rho_n} \sqrt{\operatorname{Var} \tilde{m}_n\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)} \left(1 + \frac{u}{\sqrt{r_n}}\right)^{-1} du \\
&\qquad\qquad \text{(by Cauchy–Schwarz)} \\
&\le \frac{3}{r_n} \int_{-\rho_n}^{\rho_n} \sqrt{n\tilde{m}\left(4, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)}\, du \\
&\qquad\qquad \left[\text{since} \left(1 \pm \frac{\rho_n}{\sqrt{r_n}}\right)^{-1} \to 1\right] \\
&\le \frac{3}{r_n} \int_{-\rho_n}^{\rho_n} \sqrt{nm\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)}\, du \\
&\le \frac{6\rho_n}{r_n} \sqrt{nm\left(2, a_n\left(1 - \frac{\rho_n}{\sqrt{r_n}}\right)\right)} \\
&\sim \frac{6\rho_n}{r_n} \sqrt{nm(2, a_n)} \quad \text{[by (2.5)]} \\
&\sim \frac{6\rho_n}{\sqrt{r_n}} \to 0.
\end{aligned} \tag{4.15}
$$

By Markov's inequality, $|T_n| + |U_n| \to_p 0$.

We may now rewrite (4.13)–(4.14) in the form

$$(4.16)\qquad \begin{aligned} &\frac{\hat{a}_n - a_n}{a_n}\sqrt{r_n} \ge \rho_n \Leftrightarrow B_n + o_p(1) \ge \tau_n, \\ &\frac{\hat{a}_n - a_n}{a_n}\sqrt{r_n} \le -\rho_n \Leftrightarrow B_n + o_p(1) \le -v_n. \end{aligned}$$

The asymptotic identities (4.16), valid for every $\rho_n \to \infty$ such that $\rho_n / \sqrt{r_n} \to 0$, are the key to all the subsequent analyses of $\hat{a}_n$.

STEP 2: Equivalence of (4.8) and (4.9). Fix $\rho_{n_k} \to \infty$ such that $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$.

First, suppose (4.9) holds. Then the quantity $v_{n_k}$ in (4.14) and (4.16) satisfies $v_{n_k} \to \infty$. This forces the quantity $\tau_n$ in (4.13) and (4.16) to obey $\tau_{n_k} \to \infty$: Using (2.3) and a change of variables similar to that employed in (4.12), we have, along $n = n_k \to \infty$,

$$(4.17)\qquad \begin{aligned} \tau_n &= \int_0^{\rho_n} \frac{2n}{r_n} \tilde{m}\left(2, a_n\left(1 + \frac{y}{\sqrt{r_n}}\right)\right)\left(1 + \frac{y}{\sqrt{r_n}}\right)^{-1} dy \\ &= \int_0^{\rho_n} \frac{2n}{r_n} a_n^{-2} \tilde{M}\left(2, a_n\left(1 + \frac{y}{\sqrt{r_n}}\right)\right)\left(1 + \frac{y}{\sqrt{r_n}}\right)^{-3} dy \\ &\ge \int_0^{\rho_n} \frac{2n}{r_n} a_n^{-2} \tilde{M}(2, a_n)\left(1 + \frac{y}{\sqrt{r_n}}\right)^{-3} dy \\ &\ge \rho_n \frac{n}{r_n} \tilde{m}(2, a_n) \\ &\ge \frac{1}{2}\int_{-\rho_n}^{0} \frac{n}{r_n} \tilde{m}(2, a_n)\left(1 + \frac{x}{\sqrt{r_n}}\right)^{-1} dx \\ &\ge \frac{1}{2}\int_{-\rho_n}^{0} \frac{n}{r_n} \tilde{m}\left(2, a_n\left(1 + \frac{x}{\sqrt{r_n}}\right)\right)\left(1 + \frac{x}{\sqrt{r_n}}\right) dx \\ &\ge \frac{1}{4}\int_{-\rho_n}^{0} \frac{n}{r_n} \tilde{m}\left(2, a_n\left(1 + \frac{x}{\sqrt{r_n}}\right)\right)\left(1 + \frac{x}{\sqrt{r_n}}\right)^{-1} dx \\ &\qquad\qquad \left[\text{since } \rho_n = o\left(\sqrt{r_n}\right)\right] \\ &= \frac{1}{8} v_n \to \infty, \end{aligned}$$

where we have twice used $\tilde{M}(2, \cdot\,)$ nondecreasing and $\rho_n / \sqrt{r_n} \to 0$.

It follows, then, from (4.16), that

$$P\left(\left|\frac{\hat{a}_n - a_n}{a_n}\sqrt{r_n}\right| > \rho_n\right) \le P\big(B_n + o_p(1) \ge \tau_n\big) + P\big(B_n + o_p(1) \le -v_n\big) \to 0$$

using the tightness of $\{B_n\}$ obtained in Proposition 4.3. Thus, (4.8) holds.

Now suppose (4.8) holds. In order to verify (4.9), we must show that $v_n$ of (4.14) and (4.16) obeys $v_{n_k} \to \infty$. Given $\{n'_k\} \subset \{n_k\}$, restrict to a further subsequence $\{n''_k\} \subset \{n'_k\}$, along which (4.5) holds and also along which

$$v_n = \frac{n}{\sqrt{r_n}}\left\{m\left(2, a_n\left(1 - \frac{\rho_n}{\sqrt{r_n}}\right)\right) - m(2, a_n)\right\} \to v \in [0, \infty].$$

We claim that $v = \infty$, establishing (4.9). So suppose $v < \infty$. First, assume $c^2 > 0$ in (4.5). Then let $n = n''_k \to \infty$ and observe that $\mathscr{N}(0, c^2)$ is a nondegenerate normal distribution; in particular, a continuous one. By assumption (4.8), Proposition 4.3 and (4.16), we have

$$0 = \lim_{n\to\infty} P\left(\frac{\hat{a}_n - a_n}{a_n}\sqrt{r_n} \le -\rho_n\right) = \lim_{n\to\infty} P(B_n \le -v_n) = P(Z \le -v),$$

where $\mathscr{L}(Z) = \mathscr{N}(0, c^2)$. This contradiction would force $v = \infty$ if $c^2 > 0$. Therefore, suppose that $c^2 = 0$. We would have, for every $x \in \mathbf{R}$ along the appropriate subsequence, by (2.2) and the analogue of (2.5) for $m(4, \cdot\,)$,

$$\frac{n}{r_n} G\left(a_n\left(1 + \frac{x}{\sqrt{r_n}}\right)\right) \le \frac{n}{r_n} m\left(4, a_n\left(1 + \frac{x}{\sqrt{r_n}}\right)\right) \sim \frac{n}{r_n} m(4, a_n) \to 0.$$

Utilizing (2.5), (2.7) and (2.2), it follows that for every $x$,

$$\frac{n}{r_n}\tilde{m}\left(2, a_n\left(1 + \frac{x}{\sqrt{r_n}}\right)\right) \to 1. \tag{4.18}$$

Then via (2.4), the definition (4.14) of $v_n$ and the bounded convergence theorem, for every $R > 0$ and sufficiently large $n$ in the appropriate subsequence,

$$\begin{aligned} \infty > v + 1 \ge v_n &= \int_{-\rho_n}^{0} 2\frac{n}{r_n}\tilde{m}\left(2, a_n\left(1 + \frac{x}{\sqrt{r_n}}\right)\right)\left(1 + \frac{x}{\sqrt{r_n}}\right)^{-1} dx \\ &\ge \int_{-R}^{0} 2\frac{n}{r_n}\tilde{m}\left(2, a_n\left(1 + \frac{x}{\sqrt{r_n}}\right)\right)\left(1 + \frac{x}{\sqrt{r_n}}\right)^{-1} dx \qquad (4.19) \\ &\to \int_{-R}^{0} 2\,dx = 2R, \end{aligned}$$

a contradiction when $R$ is large. Therefore, $v = \infty$ and (4.9) is established.

STEP 3: Criteria for consistency and tightness. It is clear that tightness of $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ is equivalent to (4.8) holding for *every* $\rho_{n_k} \to \infty$; it is no real restriction to consider only $\rho_{n_k} \to \infty$ such that $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$. As for consistency, clearly $\hat{a}_{n_k}/a_{n_k} \to 1$ in probability if and only if for *some* $\sigma_{n_k} \to \infty$, we have $P(|\sigma_{n_k}(\hat{a}_{n_k}/a_{n_k} - 1)| > 1) \to 0$. Now without loss of generality we need consider only $\sigma_{n_k} = o(\sqrt{r_{n_k}})$. Letting $\rho_{n_k} = \sqrt{r_{n_k}}/\sigma_{n_k}$, we see that $\hat{a}_{n_k}/a_{n_k} \to 1$ in probability if and only if (4.8) holds for *some* $\rho_{n_k} \to \infty$ such that $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$.

STEP 4: Form of the limit laws. Assume that $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ is tight. Fix $\rho_{n_k} \to \infty$ such that $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$. In all the following, restrict attention to the subsequence $\{n_k\}$ but suppress the subscript $k$. Define

$$B_{n,1} \equiv \int_0^{((|(\hat{a}_n - a_n)/a_n|\sqrt{r_n}) \wedge \rho_n)\operatorname{sgn}(\hat{a}_n - a_n)} \frac{2n}{r_n} \tilde{m}_n\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\left(1 + \frac{u}{\sqrt{r_n}}\right)^{-1} du,$$

$$B_{n,2} \equiv \int_0^{((|(\hat{a}_n - a_n)/a_n|\sqrt{r_n}) \wedge \rho_n)\operatorname{sgn}(\hat{a}_n - a_n)} \frac{2n}{r_n} \tilde{m}\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\left(1 + \frac{u}{\sqrt{r_n}}\right)^{-1} du,$$

$$B_{n,3} \equiv \int_0^{((\hat{a}_n - a_n)/a_n)\sqrt{r_n}} \frac{2n}{r_n} \tilde{m}\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\left(1 + \frac{u}{\sqrt{r_n}}\right)^{-1} du.$$

Recalling the last member of (4.12), $P(B_n \neq B_{n,1}) \leq P(|(\hat{a}_n - a_n)/a_n|\sqrt{r_n} > \rho_n) \to 0$. Also,

$$E|B_{n,1} - B_{n,2}| \leq E\int_{-\rho_n}^{\rho_n} \frac{2n}{r_n}\left|\tilde{m}_n\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right) - \tilde{m}\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\right| \times\left(1 + \frac{u}{\sqrt{r_n}}\right)^{-1} du \to 0,$$

by the computation in (4.15). Finally,

$$P(B_{n,2} \neq B_{n,3}) \leq P\left(\left|\frac{\hat{a}_n - a_n}{a_n}\sqrt{r_n}\right| \geq \rho_n\right) \to 0.$$

Therefore, $B_n - B_{n,3} \to_p 0$.

To clarify the exposition and aid in the Helly selection procedure to follow, define new functions $f_n\colon \mathbf{R} \to [0, \infty)$ by setting

$$f_n(x) = \begin{cases} \dfrac{2n}{r_n}\tilde{m}\left(2, a_n\left(1 - \dfrac{\rho_n}{\sqrt{r_n}}\right)\right)\left(1 - \dfrac{\rho_n}{\sqrt{r_n}}\right)^{-1}, & \text{if } -\infty < x < -\rho_n, \\ \dfrac{2n}{r_n}\tilde{m}\left(2, a_n\left(1 + \dfrac{x}{\sqrt{r_n}}\right)\right)\left(1 + \dfrac{x}{\sqrt{r_n}}\right)^{-1}, & \text{if } -\rho_n \leq x < \rho_n, \\ \dfrac{2n}{r_n}\tilde{m}\left(2, a_n\left(1 + \dfrac{\rho_n}{\sqrt{r_n}}\right)\right)\left(1 + \dfrac{\rho_n}{\sqrt{r_n}}\right)^{-1}, & \text{if } \rho_n \leq x < \infty. \end{cases} \tag{4.20}$$

In particular, $B_{n,3} = \int_0^{((\hat{a}_n - a_n)/a_n)\sqrt{r_n}} f_n(x)\,dx + o_p(1)$.

Since $\tilde{m}(2, s) = m(2, s) - G(s)$ is the difference of two nonincreasing functions, each $f_n$ is of bounded variation. Uniform boundedness follows from

$$\begin{aligned} 0 \le f_n(x) &\le \frac{2n}{r_n} \tilde{M}\left(2, a_n\left(1 + \frac{\rho_n}{\sqrt{r_n}}\right)\right) a_n^{-2}\left(1 - \frac{\rho_n}{\sqrt{r_n}}\right)^{-3} \\ &\sim \frac{2n}{r_n} \tilde{m}\left(2, a_n\left(1 + \frac{\rho_n}{\sqrt{r_n}}\right)\right) \\ &\le \frac{2n}{r_n} m\left(2, a_n\left(1 + \frac{\rho_n}{\sqrt{r_n}}\right)\right) \\ &\sim \frac{2n}{r_n} m(2, a_n) = 2, \end{aligned} \tag{4.21}$$

recalling (2.5) and the fact that $\tilde{M}(2, \cdot\,)$ is nondecreasing.

Given $\{n'_k\} \subset \{n_k\}$, apply Helly selection [for the most useful version, see, e.g., Taylor (1985), page 398] to find $\{n''_k\} \subset \{n'_k\}$ and $f$ of bounded variation such that

$$\forall x: \quad \lim_{n''_k \to \infty} f_{n''_k}(x) = f(x). \tag{4.22}$$

Define

$$B_{n,4} \equiv \int_0^{((\hat{a}_n - a_n)/a_n)\sqrt{r_n}} f(x)\, dx.$$

We claim that, along $\{n''_k\}$, $B_{n,3} - B_{n,4} \to 0$ in probability. Given $\varepsilon > 0$, choose $R$ so that for all large $n = n''_k$, $P(|(\hat{a}_n - a_n)/a_n|\sqrt{r_n} > R) < \varepsilon$. Then, using the bounded convergence theorem, choose $K$ so large that for $k \ge K$,

$$\int_{-R}^{R} |f_{n''_k}(x) - f(x)|\, dx < \varepsilon.$$

Now, off a set of probability at most $\varepsilon$, for $k \ge K$,

$$|B_{n''_k,4} - B_{n''_k,3}| \le \int_{-R}^{R} |f_{n''_k}(x) - f(x)|\, dx < \varepsilon,$$

establishing the claim. Define

$$\Psi(x) \equiv \int_0^x f(s)\, ds. \tag{4.23}$$

Then, choose a further subsequence $\{n'''_k\} \subset \{n''_k\}$ along which $\mathscr{L}(B_n) \to \mathscr{N}(0, c^2)$ as in Proposition 4.3. [In particular, assume (4.5) holds.] Along $\{n'''_k\}$,

$$\mathscr{L}\left(\Psi\left(\frac{\hat{a}_n - a_n}{a_n}\sqrt{r_n}\right)\right) = \mathscr{L}(B_{n,4}) \sim \mathscr{L}(B_n) \to \mathscr{N}(0, c^2). \tag{4.24}$$

To establish the form (4.10), it remains to consider the properties of $\Psi$ in order to utilize (4.24). We claim that $f$ in (4.22) and (4.23) is nondecreasing, whence $\Psi$ is convex. Obviously, $\Psi(0) = 0$. Fix $x \le y$. Then for large $n$,

$$
\begin{aligned}
f_n(x) &= \frac{2n}{r_n}\tilde{m}\left(2, a_n\left(1 + \frac{x}{\sqrt{r_n}}\right)\right)\left(1 + \frac{x}{\sqrt{r_n}}\right)^{-1} \\
&\le \frac{2n}{r_n}\tilde{M}\left(2, a_n\left(1 + \frac{y}{\sqrt{r_n}}\right)\right)a_n^{-2}\left(1 + \frac{x}{\sqrt{r_n}}\right)^{-3} \\
&= \frac{2n}{r_n}\tilde{m}\left(2, a_n\left(1 + \frac{y}{\sqrt{r_n}}\right)\right)\left(1 + \frac{y}{\sqrt{r_n}}\right)^{-1}\left(\frac{1 + y/\sqrt{r_n}}{1 + x/\sqrt{r_n}}\right)^3 \\
&= f_n(y)\left(\frac{1 + y/\sqrt{r_n}}{1 + x/\sqrt{r_n}}\right)^3.
\end{aligned} \tag{4.25}
$$

Letting $n \to \infty$ along $\{n_k''\}$ gives $f(x) \le f(y)$.

From (4.21), $0 \le f(x) \le 2$. But along $\{n_k''\}$, (2.5) and its analogue for $m(4, \cdot\,)$, show that for $x$,

$$
\begin{aligned}
2 \ge f(x) &= \lim f_n(x) \\
&= \lim \frac{2n}{r_n}\left\{m\left(2, a_n\left(1 + \frac{x}{\sqrt{r_n}}\right)\right) - G\left(a_n\left(1 + \frac{x}{\sqrt{r_n}}\right)\right)\right\} \\
&= 2 - 2\lim \frac{n}{r_n}G\left(a_n\left(1 + \frac{x}{\sqrt{r_n}}\right)\right) \\
&\ge 2 - 2\lim \frac{n}{r_n}m(4, a_n) = 2(1 - c^2),
\end{aligned} \tag{4.26}
$$

where in the last equality (4.5) is used.

In (4.24), restrict to a further subsequence such that $\mathscr{L}(((\hat{a}_n - a_n)/a_n)\sqrt{r_n}) \to \mu$, say, and let $Y \sim \mu$. By continuity of $\Psi$, (4.24) implies $\Psi(Y) \sim \mathscr{N}(0, c^2)$. If $c^2 > 0$, $\mathscr{N}(0, c^2)$ is fully supported on all of $\mathbf{R}$. Thus, necessarily $\Psi(\infty) = \infty$ and $\Psi(-\infty) = -\infty$, since $\Psi$ is nondecreasing. Also, $f = \Psi'$ is nondecreasing by (4.25). Thus if for some $x_0$, we had $f(x_0) = 0$, it would follow that $f \equiv 0$ on $(-\infty, x_0]$, whence $\Psi(-\infty) = \Psi(x_0) > -\infty$, a contradiction. Thus $f = \Psi' > 0$ and $\Psi$ is strictly increasing. Hence $\Psi^{-1}$ exists and is continuous. We have $Z \equiv \Psi(Y) \sim \mathscr{N}(0, c^2)$ whence $Y \sim \Psi^{-1}(Z)$, as required.

If, however, $c^2 = 0$, (4.26) shows that in (4.22), $f \equiv 2$, so that in (4.24), we have $\Psi(x) = 2x$. Thus along the appropriate subsequence, $((\hat{a}_n - a_n)/a_n)\sqrt{r_n} \to 0$ in probability, and certainly here, $\mathscr{L}(0) = \mathscr{L}(\Psi^{-1}(Z))$ with $Z \sim \mathscr{N}(0, 0)$, since $\Psi(0) = 0$. Moreover $\Psi(x) = 2x$ clearly has the required properties. Thus, form (4.10) for the limit laws is fully established.

STEP 5: Degenerate limits and stochastic compactness. Suppose (4.11) fails. We can restrict to a subsequence on which $(n/r_n)m(4, a_n) \to 0$. By restricting to further subsequences, the argument involving $c^2 = 0$ at the end of Step 4 reveals that $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ has a degenerate subsequential limiting distribution.

Now suppose (4.11) holds. To show that every subsequential limit is nondegenerate, we may as well assume that $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ is tight. Given any subsequence of $\{n_k\}$, we can restrict to a further subsequence where (4.24) holds for $\Psi$ as in the statement of the theorem and along which $(n/r_n)m(4, a_n) \to c^2$. By (4.11), $c^2 > 0$. From (4.24) follows, $\mathscr{L}(((\hat{a}_n - a_n)/a_n)\sqrt{r_n}) \to \mathscr{L}(\Psi^{-1}(Z))$ with $Z \sim \mathscr{N}(0, c^2)$. We claim that since $\mathscr{N}(0, c^2)$ has support exactly $\mathbf{R}$, so does $\mathscr{L}(\Psi^{-1}(Z))$. (In particular, this limit law is nondegenerate.) Let $-\infty < \alpha < \beta < \infty$. Then $\Psi' = f \geq f(\alpha) > 0$ on $(\alpha, \beta)$, so that the strictly increasing property of $\Psi$ gives

$$P(\Psi^{-1}(Z) \in (\alpha, \beta)) = P(\Psi(\alpha) < Z < \Psi(\beta)) > 0,$$

since $\Psi(\beta) - \Psi(\alpha) \geq f(\alpha)(\beta - \alpha) > 0$ and $\mathscr{L}(Z) = \mathscr{N}(0, c^2)$ has an everywhere strictly positive density. Hence no subsequential limits of $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ are degenerate. Recalling that stochastic compactness is simply tightness with no degenerate limits, we have completed the proof of Theorem 4.7. □

The reduced scale quantities $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ and the analytic tightness criterion (4.9) for them depend on the given pair of sequences $\{n_k\}, \{r_{n_k}\}$. It is natural to ask under what conditions (4.8) and (4.9), or (4.11), will hold for *every* pair $\{n\}, \{r_n\}$. The answers are very classical ones. Moreover, when these classical conditions hold, various results in the sequel can be simplified and/or strengthened.

Following recent usage, a random variable $X$ (equivalently, its distribution $F$) is said to be in the *Feller class* provided

$$\Lambda = \limsup_{t \to \infty} \frac{G(t)}{m(2, t)} < 1. \tag{4.27}$$

Feller (1967) showed that (4.27) is necessary and sufficient for the partial sums $S_n = X_1 + \cdots + X_n$ derived from $F$ to be affinely stochastically compact, i.e., for there to exist shifts $\delta_n$ and normalizations $d_n$ such that the sequence $\{\mathscr{L}((S_n - \delta_n)/d_n)\}$ is stochastically compact. In particular, the Feller class includes the domain of attraction of every stable law and thus is quite large.

Recall that Lévy (1937) proved that $X$ (distributed according to $F$) is *outside the domain of partial attraction of the normal* (i.e., for no sequences $\{n_k\}, \{\delta_k\}, \{d_k\}$ can $\mathscr{L}((S_{n_k} - \delta_k)/d_k) \to N(0,1))$ if and only if

$$\lambda = \liminf_{t \to \infty} \frac{G(t)}{m(2, t)} > 0. \tag{4.28}$$

4.29. COROLLARY. *$\{((\hat{a}_n - a_n)/a_n)\sqrt{r_n}\}$ is tight for every $r_n \to \infty$ such that $r_n/n \to 0$, if and only if $X$ is in the Feller class. $\{((\hat{a}_n - a_n)/a_n)\sqrt{r_n}\}$ is stochastically compact for every such $\{r_n\}$ if and only if $X$ is in the Feller class but outside the domain of partial attraction of the normal.*

PROOF. Assume $X$ is in the Feller class. Fix $r_n \to \infty$, $r_n/n \to 0$ and then $\rho_n \to \infty$, $\rho_n/\sqrt{r_n} \to 0$. Since $a_n \to \infty$, there exists $0 < \beta < 1 - \Lambda$ such that via (4.27),

$$\begin{aligned}
&\frac{n}{\sqrt{r_n}}\left\{m\left(2, a_n\left(1 - \frac{\rho_n}{\sqrt{r_n}}\right)\right) - m(2, a_n)\right\} \\
&\quad = \frac{2n}{r_n}\int_{-\rho_n}^{0} \tilde{m}\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\frac{du}{1 + u/\sqrt{r_n}} \\
&\quad \geq \frac{2n}{r_n}\int_{-\rho_n}^{0} \beta m\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\frac{du}{1 + u/\sqrt{r_n}} \\
&\quad \geq \frac{\beta n}{r_n} m(2, a_n)\rho_n = \beta\rho_n \to \infty.
\end{aligned}$$

Thus, (4.9) holds for the full sequence $\{n_k = k\}$, for every $\rho_n \to \infty$ such that $\rho_n/\sqrt{r_n} \to 0$.

If in addition (4.28) holds, note that

$$\liminf_{n\to\infty} \frac{n}{r_n} m(4, a_n) \geq \liminf_{n\to\infty} \frac{n}{r_n} G(a_n) = \liminf_{n\to\infty} \frac{G(a_n)}{m(2, a_n)} \geq \lambda > 0,$$

so that (4.11) holds as well on $\{n_k = k\}$.

Suppose that $X$ is not in the Feller class. We construct $\{r_n\}$ and appropriate $\{\rho_n\}$ so that (4.9) fails along every subsequence. Define $d(\cdot)$ implicitly by $tm(2, d(t)) = 1$. Since (4.27) fails and $d(\cdot)$ is continuous,

$$1 = \limsup_{s\to\infty} \frac{G(s)}{m(2, s)} = \limsup_{t\to\infty} \frac{G(d(t))}{m(2, d(t))} = \limsup_{t\to\infty} tG(d(t)).$$

Thus, there are $t_k \to \infty$ with $t_k G(d(t_k)) \to 1$. By selecting a subsequence if necessary, it may be assumed that $t_{k+1} > t_k + 2$. Define the sequence $\{r_n\}$ by

$$r_n = \frac{n}{t_k} \quad \text{if } [t_k^2] \leq n < [t_{k+1}^2], \tag{4.30}$$

where $[\cdot]$ denotes the integer part of $x$. Notice that $r_n \geq [t_k^2]/t_k \sim t_k \to \infty$, whereas $n/r_n \equiv t_k$ on the block $[t_k^2] \leq n < [t_{k+1}^2]$, so that $n/r_n \to \infty$. [It is important here that we have never required that $\{r_n\}$ or $\{n/r_n\}$ be integer sequences.] As always, $nm(2, a_n) = r_n$, which implies that $a_n = d(n/r_n)$ so that

$$\lim_{n\to\infty} \frac{n}{r_n} G(a_n) = \lim_{k\to\infty} t_k G(d(t_k)) = 1.$$

Thus, $(n/r_n)\tilde{m}(2, a_n) \to 0$. Let $\rho_n = ((n/r_n)\tilde{m}(2, a_n))^{-1} \wedge r_n^{1/3}$, so that $\rho_n \to \infty$ but $\rho_n = o(\sqrt{r_n})$. Then

$$\begin{aligned}
&\frac{n}{\sqrt{r_n}}\left\{m\left(2, a_n\left(1 - \frac{\rho_n}{\sqrt{r_n}}\right)\right) - m(2, a_n)\right\} \\
&\quad = \frac{2n}{r_n}\int_{-\rho_n}^{0} \tilde{m}\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)\frac{du}{1 + u\sqrt{r_n}} \\
&\quad = \frac{2n}{r_n}\int_{-\rho_n}^{0} \tilde{M}\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right)a_n^{-2}\frac{du}{\left(1 + u/\sqrt{r_n}\right)^3} \\
&\quad \le \frac{3n}{r_n}\int_{-\rho_n}^{0} \tilde{M}(2, a_n)a_n^{-2}\,du \\
&\quad = \frac{3n}{r_n}\tilde{m}(2, a_n)\rho_n \le 3.
\end{aligned}$$

Since (4.9) fails (for every subsequence), by Theorem 4.7 $\{((\hat{a}_n - a_n)/a_n)\sqrt{r_n}\}$ has no convergent subsequence.

Suppose, finally, that $X$ is outside of the domain of partial attraction of the normal, so that (4.28) fails. Choose a subsequence on which $(n/r_n)G(xa_n) \to 0$ for every $x > 0$ [as in the proof of Lévy's (1937) theorem, page 113]. Then (2.7) and standard arguments show that on this subsequence,

$$\frac{n}{r_n} m(4, a_n) = \frac{n}{r_n}\int_0^1 4s^3 G(sa_n)\,ds \to 0.$$

Thus (4.11) fails. Theorem 4.7 now implies that $\{((\hat{a}_n - a_n)/\hat{a}_n)\sqrt{r_n}\}$ has a degenerate subsequential limit. □

It is important to note that (4.27) is not necessary if $\{r_n\}$ is fixed. See Example 4.40.

Next we derive necessary and sufficient conditions for convergence in distribution of $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ to a given law of the form in (4.10). Specifically, given a number $0 \le c^2 \le 1$ and a strictly increasing convex function $\Psi$ with $\Psi(0) = 0$, $\Psi(-\infty) = -\infty$, $\Psi(\infty) = \infty$ and $2(1 - c^2) \le \Psi' \le 2$, we identify exactly when $\mathscr{L}(((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}) \to \mathscr{L}(\Psi^{-1}(Z))$, where $Z \sim N(0, c^2)$.

Let $v = \mathscr{L}(\Psi^{-1}(Z))$. First consider the case of $v$ degenerate. Necessarily $c^2 = 0$. If $\mathscr{L}(((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}) \to v = \delta_0$, then consider a subsequence along which $(n/r_n)m(4, a_n) \to c_0^2$ and $f_n(x) \to f_0(x)$. By tightness and (4.24), along this subsequence $\mathscr{L}(((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}) \to \mathscr{L}(\Psi_0^{-1}(Z_0))$, where $\Psi_0(x) = \int_0^x f_0(s)\,ds$ and $Z_0 \sim N(0, c_0^2)$, as in Theorem 4.7. Therefore $\Psi_0^{-1}(Z_0)$ is degenerate, so necessarily $c_0^2 = 0$. Consequently, $\lim_{k\to\infty}(n_k/r_{n_k})m(4, a_{n_k}) = c_0^2 = 0$.

Conversely, if $\lim_{k\to\infty}(n_k/r_{n_k})m(4, a_{n_k}) = c^2 = 0$, (4.26) shows that $f_n(x) \to f(x) = 2$ so $\Psi(x) = 2x$. The argument surrounding (4.19) gives (4.9) and hence, by Theorem 4.7, tightness of $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$. But then, Proposition 4.3 and (4.24) give $2(((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}) \to_p 0$. Thus if $v$ is degenerate, $\mathscr{L}(((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}) \to v$ if and only if

$$\lim_{k\to\infty} \frac{n_k}{r_{n_k}} m(4, a_{n_k}) = 0. \tag{4.31}$$

If $v = \mathscr{L}(\Psi^{-1}(Z))$ is nondegenerate, a unique representation for the limit will be required. Now certainly $c^2 > 0$. Let $\Psi_i$, $i = 1, 2$, be strictly monotone and convex with $\Psi_i(0) = 0$, $\Psi_i(\infty) = \infty$, $\Psi_i(-\infty) = -\infty$ and $0 < \Psi' \le 2$. Suppose $Z_i \sim N(0, c_i^2)$ with $0 < c_i^2 \le 1$ and $\mathscr{L}(\Psi_1^{-1}(Z_1)) = \mathscr{L}(\Psi_2^{-1}(Z_2))$. Then

$$\begin{aligned}\mathscr{L}(Z_1) &= \mathscr{L}\big(\Psi_1(\Psi_1^{-1}(Z_1))\big) = \mathscr{L}\big(\Psi_1(\Psi_2^{-1}(Z_2))\big)\\ &= \mathscr{L}\big(\Psi_1(\Psi_2^{-1}(Z_1 c_2/c_1))\big).\end{aligned}$$

The strict monotonicity of $w(t) = \Psi_1(\Psi_2^{-1}(c_2 t/c_1))$ leads to $w(t) \equiv t$, which is equivalent to $\Psi_1/c_1 = \Psi_2/c_2$. To see this, let $q$ be the density of $\mathscr{L}(Z_1)$. Then for every $t$,

$$\begin{aligned}\int_{-\infty}^{t} q(s)\,ds &= P(Z_1 \le t) = P(w(Z_1) \le w(t))\\ &= P(Z_1 \le w(t)) = \int_{-\infty}^{w(t)} q(s)\,ds,\end{aligned}$$

whence $\int_t^{w(t)} q(s)\,ds \equiv 0$. But $q(s) > 0$ for every $s$ and thus $w(t) - t \equiv 0$. Therefore the expression $\Psi/c$ is invariant among representations for $v$ of the form (4.10). Thus given nondegenerate $v = \mathscr{L}(\Psi^{-1}(Z))$, $Z \sim N(0, c^2)$ and $\Psi$ as in Theorem 4.7, we can uniquely represent $v$ in the canonical form $v = \mathscr{L}(\Psi_0^{-1}(Z_0))$, where $Z_0 \sim N(0, 1)$ and $\Psi_0 = \Psi/c$.

Now we can determine convergence criteria for a given nondegenerate $v$ in canonical form $v = \mathscr{L}(\Psi_0^{-1}(Z_0))$. For the necessity part, suppose $\mathscr{L}(((\hat{a}_{n_k} - a_{n_k})/\hat{a}_{n_k})\sqrt{r_{n_k}}) \to v$. By Theorem 4.7, (4.11) holds, We claim, in addition, that

$$\lim_{k\to\infty} \sqrt{\frac{n_k}{r_{n_k}}}\, \frac{\tilde{m}\left(2, a_{n_k}\left(1 + x/\sqrt{r_{n_k}}\right)\right)}{\sqrt{m(4, a_{n_k})}} = \frac{1}{2}\Psi_0'(x) \tag{4.32}$$

for all but countably many $x$ (recalling the convexity of $\Psi_0$): Given any subsequence of $\{n_k\}$, restrict further so that $(n/r_n)m(4, a_n) \to c^2 > 0$, using (4.11). Restrict further so that (4.22) holds, for some $f$, where we recall (4.20). As in the proof of Theorem 4.7, it follows that on this restricted sequence, $\mathscr{L}(((\hat{a}_n - a_n)/\hat{a}_n)\sqrt{r_n}) \to \mathscr{L}(\Psi^{-1}(Z))$ where $Z \sim N(0, c^2)$ and $\Psi' = f$. By uniqueness of the representation of $v$, $\Psi/c = \Psi_0$, so $\Psi'(x) = c\Psi_0'(x)$ off a countable set. Thus, (4.32) holds on this restricted subsequence and a standard subsequence argument now shows (4.32) holds on the original subsequence $\{n_k\}$. For convenience, we note that in view of (4.11), (2.5) and (2.7), equation (4.32) is equivalent to

$$\text{(4.33)} \quad \lim_{k\to\infty} \left\{ \frac{n_k}{r_{n_k}} m(4, a_{n_k}) \right\}^{-1/2} \left( 1 - \frac{n_k}{r_{n_k}} G\left( a_{n_k}\left( 1 + \frac{x}{\sqrt{r_{n_k}}} \right) \right) \right) = \frac{1}{2}\Psi_0'(x),$$

off a countable set.

Thus, (4.11) and (4.33) are necessary for $\mathscr{L}(((\hat{a}_{n_k} - a_{n_k})/\hat{a}_{n_k})\sqrt{r_{n_k}}) \to v$.

We claim (4.11) and (4.33) also suffice for $\mathscr{L}(((\hat{a}_{n_k} - a_k)/a_{n_k})\sqrt{r_{n_k}}) \to v$. Restrict to a subsequence where $(n/r_n)m(4, a_n) \to c^2 > 0$ and define $\Psi = c\Psi_0$. Now (4.32) leads to (4.22) with $f = \Psi'$, off a countable set. On this restricted subsequence, it will follow from (4.24), etc., that on this subsequence $\mathscr{L}(((\hat{a}_n - a_n)/a_n)\sqrt{r_n}) \to \mathscr{L}(\Psi^{-1}(Z))$ with $Z \sim N(0, c^2)$, provided $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ can be shown to be tight. This will be done by verifying (4.9). Note that if $\rho_n \to \infty$, $\rho_n = o(\sqrt{r_n})$, then using properties of $\Psi_0$,

$$v_n = \int_{-\rho_n}^{0} f_n(x)\,dx \ge \int_R^0 f_n(x)\,dx \to \int_R^0 f(x)\,dx = -\Psi(-R) = c\Psi_0(-R) \to \infty,$$

where first $n$ and then $R$ have tended to infinity. Thus (4.9) holds. Tightness now follows from Theorem 4.7. Finally, note $\mathscr{L}(\Psi^{-1}(Z)) = v$, independent of the particular subsequence, thereby establishing the following proposition.

4.34. PROPOSITION. *Fix integers $n_k \to \infty$. Then*

$$\mathscr{L}\left(\left((\hat{a}_{n_k} - a_{n_k})/a_{n_k}\right)\sqrt{r_{n_k}}\right) \to \delta_0$$

*if and only if* (4.31) *holds. Given nondegenerate $v$ as in* (4.10) *with canonical representing convex function $\Psi_0$, $\mathscr{L}(((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}) \to v$ if and only if both* (4.11) *and* (4.33) *hold.*

Now turn to the asymptotic normality problem for subsequences of $\{((\hat{a}_n - a_n)/a_n)\sqrt{r_n}\}$. Assume (4.9) holds. Then $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ is tight with subsequential limit laws of form $\mathscr{L}(\Psi^{-1}(Z))$ where $Z \sim \mathscr{N}(0, c^2)$ as in (4.10). Under what condition will every such resulting law be normal?

4.35. PROPOSITION. *The sequence* $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ *is tight with only mean zero normal or degenerate subsequential limit laws if and only if*

$$\limsup_{k\to\infty} \frac{n_k}{r_{n_k}} G(a_{n_k}) < 1,$$

$$\forall x: \quad \lim_{k\to\infty} \frac{n_k}{r_{n_k}} \left\{ G\left( a_{n_k}\left(1 + \frac{x}{\sqrt{r_{n_k}}}\right)\right) - G(a_{n_k}) \right\} = 0. \tag{4.36}$$

*Stochastic compactness of* $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ *with only mean zero normal subsequential limits holds if and only if* (4.36) *and* (4.11) *both hold. Convergence to* $\mathscr{N}(0, \kappa^2)$, $\kappa^2 \geq 0$, *occurs along subsequences where* (4.36) *holds and*

$$\frac{r_n m(4, a_n)}{n\tilde{m}(2, a_n)^2} \to 4\kappa^2.$$

PROOF. Restrict to a subsequence of $\{n_k\}$ along which (4.22) and (4.24) hold. If $c^2 = 0$, then the computation in (4.26) shows $f$ is constant and hence $\Psi$ is linear. If $c^2 > 0$, then for $Z \sim \mathscr{N}(0, c^2)$, $\mathscr{L}(\Psi^{-1}(Z))$ is nondegenerate normal exactly when $\Psi^{-1}$ (hence $\Psi$) is linear. [This was argued following (4.31) since $\Psi$ is strictly increasing.] Linearity of $\Psi$ is equivalent to constancy of $f$ (since $f$ is nondecreasing). But, necessary and sufficient that every Helly selection $f$ as in (4.22) will be constant is

$$\forall x: \quad \lim_{k\to\infty} \frac{n_k}{r_{n_k}} \left\{ \tilde{m}\left(2, a_{n_k}\left(1 + \frac{x}{\sqrt{r_{n_k}}}\right)\right) - \tilde{m}(2, a_{n_k}) \right\} = 0, \tag{4.37}$$

or utilizing (2.2), (2.5) and (2.7), its equivalent

$$\forall x: \quad \lim_{k\to\infty} \frac{n_k}{r_{n_k}} \left\{ G\left( a_{n_k}\left(1 + \frac{x}{\sqrt{r_{n_k}}}\right)\right) - G(a_{n_k}) \right\} = 0. \tag{4.38}$$

Indeed, in the notation of (4.20), (4.37)–(4.38) are equivalent to

$$\forall x: \quad \lim_{n_k\to\infty} \left( f_{n_k}(x) - f_{n_k}(0)\right) = 0. \tag{4.39}$$

Under (4.39), along a subsequence where $f_n(0) \to \xi$, we will have $\forall x$: $f_n(x) \to f(x) \equiv \xi$. If $\xi > 0$, then (4.9) follows using the argument surrounding (4.19). But if $\xi = 0$, by a diagonalization argument and bounded convergence it is possible to produce $\rho_n \to \infty$, $\rho_n = o(\sqrt{r_n})$ such that along a further subsequence,

$$\frac{n}{\sqrt{r_n}} \left( m\left(2, a_n\left(1 - \frac{\rho_n}{\sqrt{r_n}}\right)\right) - m(2, a_n)\right) = \int_{-\rho_n}^{0} f_n(x)\, dx \to 0,$$

whence (4.9) fails. Thus (4.9) [under (4.39)] is equivalent to

$$\Xi = \liminf_{k\to\infty} f_{n_k}(0) > 0.$$

To see this, use the argument surrounding (4.19). Now

$$\limsup_{k\to\infty} (n_k/r_{n_k})G(a_{n_k}) = 1 - \frac{1}{2}\Xi.$$

Hence (4.9) and (4.37)–(4.38) can be combined into the compact asymptotic normality criterion (4.36) and the proposition is proved. □

Given $r_n \to \infty$, $r_n/n \to 0$, it may be asked whether (4.36) and (4.11) (the stochastically compact asymptotic normality criteria) can hold for the full sequence $\{n_k = k\}$ for a random variable $X$ not in the Feller class. The answer is affirmative as the next example shows.

4.40. EXAMPLE. Let $X$ be a Doeblin universal random variable; that is, $X$ belongs to the domain of partial attraction of every infinitely divisible law [Doeblin (1940)]. Necessarily $X$ is not in the Feller class by a result of Pruitt (1983). By considering an infinitely divisible law with no normal component and a continuous, infinite Lévy measure, we see that there are $t_k \to \infty$, $t_{k+1} > t_k + 2$, $c^2 > 0$ and $0 < \xi < 1$ such that

$$\lim_{y\to 1} \overline{\underline{\lim}}_{k\to\infty} t_k G(yd(t_k)) = 1 - \xi, \qquad (4.41)$$

$$\lim_{k\to\infty} t_k m(4, d(t_k)) = c^2,$$

where $\overline{\underline{\lim}}$ denotes each (in turn) of $\overline{\lim}$ and $\underline{\lim}$, and where we recall that $tm(2, \overline{d(t)}) = 1$ for large $t > 0$. Construct $r_n \to \infty$, $r_n/n \to 0$ as in (4.30). Then $d(t_k) = a_n$ for $[t_k^2] \le n < [t_{k+1}^2]$. We claim (4.36) and (4.11) hold, as was to be demonstrated. But (4.41) gives $(n/r_n)m(4, a_n) \to c^2 > 0$ which is (4.11) and $(n/r_n)G(a_n) \to 1 - \xi < 1$ which is the first part of (4.36). For the rest, choose $0 < y_1 < 1 < y_2 < \infty$ and write, for given $x$ and large $n$,

$$\frac{n}{r_n}\left| G\left(a_n\left(1 + \frac{x}{\sqrt{r_n}}\right)\right) - G(a_n)\right| \le \frac{n}{r_n}\{G(y_1 a_n) - G(y_2 a_n)\}. \qquad (4.42)$$

Letting $n \to \infty$ and then $y_1 \uparrow 1$, $y_2 \downarrow 1$ and utilizing (4.41), we obtain the rest of (4.36).

Utilizing Proposition 4.35, it can in fact be shown that

$$\mathscr{L}\left(((\hat{a}_n - a_n)/a_n)\sqrt{r_n}\right) \to N(0, c^2/4\xi^2)$$

along the full sequence $\{n\}$, despite the highly irregular analytic properties inherent in the universal distribution $\mathscr{L}(X)$.

**5. Self-normalized and empirical limit theorems.** Once again fix a sequence $\{r_n\}$ as in (4.1) and integers $n_k \to \infty$, unless otherwise noted. In this section the complete joint asymptotic behavior of $\hat{\gamma}_n$ and $\hat{a}_n$ will be determined. Note that the quantity $(n/\hat{a}_n\sqrt{r_n})(\hat{\gamma}_n - \gamma_n)$ is the self-normalized expression

$$\sum_{j=1}^{n} \frac{(|X_j| \wedge \hat{a}_n)\operatorname{sgn}(X_j) - E(|X| \wedge a_n)\operatorname{sgn}(X)}{\left(\sum_{j=1}^{n} X_j^2 \wedge \hat{a}_n^2\right)^{1/2}},$$

which assumes a studentized form.

Our main result is Theorem 5.1, which gives criteria for tightness of the naturally self-normalized joint estimators and characterizes the resulting subsequential limit laws. Criteria for convergence to a given limit law are given in Theorem 5.50. In particular, normal limits and criteria for convergence to them are considered in Corollaries 5.59, 5.64, 5.68 and 5.70. Our results providing for the empirical determination of *standard* normal limiting distributions are presented in Theorem 5.81.

We begin with our main result dealing with tightness of the naturally self-normalized joint estimators and characterizing the resulting subsequential limit laws. The naturalness of the self-normalizations is suggested by Theorems 4.7 and 3.6. (See also the discussion immediately preceding Corollary 5.70.) The subsequential limit laws have support on all of $\mathbf{R}^2$ except in two special cases. As will be seen, these cases are intimately related to $X$ belonging to the domain of partial attraction of the normal or of the ordinary Poisson law with parameter 1.

5.1. THEOREM. (i) *The sequence*

$$\left\{\mathscr{L}\left(\frac{n_k}{\hat{a}_{n_k}\sqrt{r_{n_k}}}(\hat{\gamma}_{n_k} - \gamma_{n_k}), \frac{\hat{a}_{n_k} - a_{n_k}}{\hat{a}_{n_k}}\sqrt{r_{n_k}}\right)\right\} \tag{5.2}$$

*is tight in* $\mathbf{R}^2$ *if and only if* (4.9) *holds for every* $\rho_{n_k} \to \infty$ *such that* $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$. *Every subsequential limit is of the form*

$$\mathscr{L}\left(Z_1 + \Phi(\Psi^{-1}(Z_2)), \Psi^{-1}(Z_2)\right), \tag{5.3}$$

*where* $(Z_1, Z_2)$ *is mean zero bivariate normal with covariance*

$$\Sigma = \begin{pmatrix} a^2(X) & b \\ b & c^2 \end{pmatrix}, \tag{5.4}$$

*with* $0 \le c^2 \le 1$, $b^2 \le c^2$ *and* $c^2 = 0$ *when* $a^2(X) < 1$ *and* $a^2(X)$ *is as in* (4.2). $\Psi$ *is a strictly increasing convex function with range* $(-\infty, \infty)$ *such that* $\Psi(0) = 0$, $2(1 - c^2) \le \Psi' \le 2$ *and* $\Phi = \Phi^+ + \Phi^-$, *where* $\Phi^{\pm}$ *are nondecreasing concave functions with* $\Phi^{\pm}(0) = 0$ *and* $(\Phi^+)' + (\Phi^-)' = 1 - \frac{1}{2}\Psi'$. *A limit law represented by* (5.3) *and* (5.4) *is supported on all of* $\mathbf{R}^2$ *when* $\det \Sigma \ne 0$ *and is supported on a curve in* $\mathbf{R}^2$ *when* $\det \Sigma = 0$.

(ii) *The sequence* (5.2) *will have subsequential limits of the form* (5.3) *whose support is not all of* $\mathbf{R}^2$, *if and only if along some subsequence* $\{n'\}$ *of* $\{n_k\}$, *for each* $0 < x < \infty$,

$$\lim_{n' \to \infty} \frac{n'}{r_{n'}} G^+(xa_{n'}) = \alpha I_{(0,c)}(x),$$
$$\lim_{n' \to \infty} \frac{n'}{r_{n'}} G^-(xa_{n'}) = \beta I_{(0,c)}(x), \tag{5.5}$$

*where* $\alpha\beta = 0$ *and either*
(a) $b^2 = c^2 = 0$ *and* $\alpha = \beta = 0$
*or*
(b) $b^2 = c^2 > 0$ *and* $\alpha + \beta = c^{-2}$.
*In either case, the corresponding limit law for* (5.2) *is concentrated on a curve in* $\mathbf{R}^2$.

PROOF OF THEOREM 5.1(i). For clarity the proof is divided into several steps.

STEP 1: Tightness of (5.2) implies (4.9) holds whenever $\rho_{n_k} \to \infty$ and $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$. If (5.2) is tight, the second marginal $\{(1 - a_{n_k}/\hat{a}_{n_k})\sqrt{r_{n_k}}\}$ is tight, and thus $\hat{a}_{n_k}/a_{n_k} \to_p 1$. Hence $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ is tight and so (4.9) holds, for every $\rho_{n_k} \to \infty$ such that $\rho_k/\sqrt{r_{n_k}} \to 0$, by Theorem 4.7.

STEP 2: (4.9) holding whenever $\rho_{n_k} \to \infty$ and $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$, implies tightness of (5.2). When (4.9) holds for even *some* $\rho_{n_k} \to \infty$ such that $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$, we have $\hat{a}_n/a_n \to_p 1$ by Theorem 4.7. Thus the two sequences

$$\left\{\frac{n_k}{\hat{a}_{n_k}\sqrt{r_{n_k}}}(\hat{\gamma}_{n_k} - \gamma_{n_k})\right\}$$

and

$$\left\{\frac{n_k}{a_{n_k}\sqrt{r_{n_k}}}(\hat{\gamma}_{n_k} - \gamma_{n_k})\right\}$$

are asymptotically equivalent. From (2.3), we have the fundamental

$$\begin{aligned}\frac{n}{a_n\sqrt{r_n}}(\hat{\gamma}_n - \gamma_n) &= \frac{n}{a_n\sqrt{r_n}}(M_n(1, \hat{a}_n) - M(1, a_n)) \\ &= \frac{n}{a_n\sqrt{r_n}}(M_n(1, a_n) - M(1, a_n)) \\ &\quad + \frac{n}{a_n\sqrt{r_n}}\int_{a_n}^{\hat{a}_n}\{G_n^+(s) - G_n^-(s)\}\,ds.\end{aligned} \tag{5.6}$$

Proposition 4.3 establishes the asymptotic normality of the first term. To seek tightness of the second term and (later) identify the joint limits, we will proceed very much as in the proof of Theorem 4.7.

After a change of variables, $s = a_{n_k}(1 + u/\sqrt{r_{n_k}})$, tightness of $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$ (as guaranteed by Theorem 4.7) can be used to reduce the study of

$$\frac{n_k}{a_{n_k}\sqrt{r_{n_k}}} \int_{a_{n_k}}^{\hat{a}_{n_k}} \{G_{n_k}^+(s) - G_{n_k}^-(s)\}\, ds \tag{5.7}$$

to the study of the asymptotically equivalent quantity (whose form is useful below)

$$\begin{aligned} \int_0^{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}} \frac{n_k}{r_{n_k}} \Bigg\{ G^+\left(a_{n_k}\left(1 + \frac{u}{\sqrt{r_{n_k}}}\right)\right) \\ - G^-\left(a_{n_k}\left(1 + \frac{u}{\sqrt{r_{n_k}}}\right)\right)\Bigg\} \frac{du}{1 + u/\sqrt{r_{n_k}}}. \end{aligned} \tag{5.8}$$

The replacement of $G_{n_k}^{\pm}$ by $G^{\pm}$ in (5.7) is accomplished by an argument analogous to that in (4.15) and utilizes

$$\begin{aligned} \operatorname{Var}\frac{n}{r_n} G_n^{\pm}\left(a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right) &\le \frac{n}{r_n^2} G^{\pm}\left(a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right) \\ &\le \frac{n}{r_n^2} G\left(a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right) \\ &\le \frac{n}{r_n^2} m\left(2, a_n\left(1 + \frac{u}{\sqrt{r_n}}\right)\right). \end{aligned}$$

To see that (5.7) is tight due to the tightness of $\{((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}\}$, simply note that given $\rho_{n_k} \to \infty$ with $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$, the integrand of (5.8) is eventually dominated uniformly in $|u| \le \rho_{n_k}$ by

$$\begin{aligned} 2\frac{n_k}{r_{n_k}} G\left(a_{n_k}\left(1 + \frac{u}{\sqrt{r_{n_k}}}\right)\right) &\le 2\frac{n_k}{r_{n_k}} G\left(a_{n_k}\left(1 - \frac{\rho_{n_k}}{\sqrt{r_{n_k}}}\right)\right) \\ &\le 2\frac{n_k}{r_{n_k}} m\left(2, a_{n_k}\left(1 - \frac{\rho_{n_k}}{\sqrt{r_{n_k}}}\right)\right) \sim 2, \end{aligned}$$

utilizing (2.7) and (2.5) as usual. Since $P(|((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}| > \rho_{n_k}) \to 0$, the quantity in (5.7) is, with probability tending to 1, dominated in magnitude by the tight quantity $3|((\hat{a}_{n_k} - a_{n_k})/a_{n_k})\sqrt{r_{n_k}}|$.

STEP 3: Characterization of the limit laws. To ease the notation, in this step we suppress the subscripts $k$, but it is to be understood that here all sequences are either $\{n_k\}$ or subsequences thereof. Assume hereafter that (4.9) holds whenever $\rho_n \to \infty$ and $\rho_n / \sqrt{r_n} \to 0$.

Choose $0 < \rho_n \to \infty$ such that $\rho_n / \sqrt{r_n} \to 0$, and construct two sequences of functions $\{g_n^{\pm}\}$ by

$$(5.9) \quad g_n^{\pm}(y) = \begin{cases} \dfrac{n}{r_n}\left(G^{\pm}\left(a_n\left(1 - \dfrac{\rho_n}{\sqrt{r_n}}\right)\right)\right)\left(1 - \dfrac{\rho_n}{\sqrt{r_n}}\right)^{-1}, & \text{if } y < -\rho_n, \\ \dfrac{n}{r_n}\left(G^{\pm}\left(a_n\left(1 + \dfrac{y}{\sqrt{r_n}}\right)\right)\right)\left(1 + \dfrac{y}{\sqrt{r_n}}\right)^{-1}, & \text{if } -\rho_n \le y < \rho_n, \\ \dfrac{n}{r_n}\left(G^{\pm}\left(a_n\left(1 + \dfrac{\rho_n}{\sqrt{r_n}}\right)\right)\right)\left(1 + \dfrac{\rho_n}{\sqrt{r_n}}\right)^{-1}, & \text{if } y \ge \rho_n. \end{cases}$$

Applying Helly selection (twice) to the sequences (5.9), nonincreasing limit functions $g^{\pm}$ are obtained such that the quantity (5.8) can be approximated in probability along appropriate subsequences by the quantity

$$(5.10) \qquad \int_0^{((\hat{a}_n - a_n)/a_n)\sqrt{r_n}} \{g^+(y) - g^-(y)\}\, dy \equiv \Phi\left(\frac{\hat{a}_n - a_n}{a_n}\sqrt{r_n}\right),$$

where

$$(5.11) \qquad \begin{aligned} \Phi(x) &\equiv \int_0^x \{g^+(y) - g^-(y)\}\, dy \\ &= \int_0^x g^+(y)\, dy - \int_0^x g^-(y)\, dy \equiv \Phi^+(x) - \Phi^-(x). \end{aligned}$$

[This approximation is accomplished via tightness of $\{((\hat{a}_n - a_n)/a_n)\sqrt{r_n}\}$ and an application of the bounded convergence theorem similar to that following (4.22).] Concavity of each of $\Phi^{\pm}$ in (5.11) follows from each of $g^{\pm}$ nonincreasing. Note that

$$(5.12) \qquad \begin{aligned} &g_n^+(y) + g_n^-(y) \\ &= \frac{n}{r_n} G\left(a_n\left(1 + \frac{y}{\sqrt{r_n}}\right)\right)\left(1 + \frac{y}{\sqrt{r_n}}\right)^{-1} \\ &= \frac{n}{r_n}\left\{m\left(2, a_n\left(1 + \frac{y}{\sqrt{r_n}}\right)\right) - \tilde{m}\left(2, a_n\left(1 + \frac{y}{\sqrt{r_n}}\right)\right)\right\}\left(1 + \frac{y}{\sqrt{r_n}}\right)^{-1} \\ &= \frac{n}{r_n} m\left(2, a_n\left(1 + \frac{y}{\sqrt{r_n}}\right)\right)\left(1 + \frac{y}{\sqrt{r_n}}\right)^{-1} - \frac{1}{2} f_n(y), \end{aligned}$$

recalling (4.20). Let $n \to \infty$ along the appropriate subsequence. Then, via (2.5), (2.7) and (4.22), $g^+ + g^- = 1 - \frac{1}{2}f$, i.e., $(\Phi^+)' + (\Phi^-)' = 1 - \frac{1}{2}\Psi'$.

Restrict to a further subsequence so that the approximation in probability of (5.7) by the term (5.10) is valid and also

$$\Psi\left(\frac{\hat{a}_n - a_n}{a_n}\sqrt{r_n}\right) - B_n \to_p 0,$$

where $B_n$ is as in (4.12). This uses the argument leading to (4.24). Let $A_n = (n/a_n\sqrt{r_n})\{M_n(1, a_n) - M(1, a_n)\}$. Using (5.6), we then have (using absolute continuity of $\Psi^{-1}$ and $\Phi$)

$$\begin{aligned}(5.13)\qquad &\left(\frac{n}{a_n\sqrt{r_n}}(\hat{\gamma}_n - \gamma_n), \frac{\hat{a}_n - a_n}{a_n}\sqrt{r_n}\right)\\ &\quad = \left(A_n + \Phi(\Psi^{-1}(B_n)), \Psi^{-1}(B_n)\right) + o_p(1).\end{aligned}$$

Restrict further so that (4.5) holds. Then Proposition 4.3 asserts that $\mathscr{L}(A_n, B_n) \to \mathscr{N}(0, \Sigma)$. Letting $(Z_1, Z_2) \sim \mathscr{N}(0, \Sigma)$, (5.13) yields the form in (5.3) immediately.

STEP 4: The support of (5.3) is all of $\mathbf{R}^2$ (resp., a curve in $\mathbf{R}^2$) precisely when $\det \Sigma$ is nonzero (resp., zero). Given a law of the form in (5.3), etc., suppose $\det \Sigma \neq 0$. Then the support of $\mathscr{L}(Z_1, Z_2)$ is exactly $\mathbf{R}^2$. The transformation $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ given by

$$(5.14)\qquad T(u,v) = \left(u + \Phi(\Psi^{-1}(v)), \Psi^{-1}(v)\right)$$

is absolutely continuous with (locally) absolutely continuous inverse

$$(5.15)\qquad T^{-1}(x,y) = (x - \Phi(y), \Psi(y)).$$

[The local absolute continuity of $\Psi^{-1}$ is an easy consequence of $0 < \Psi'$ nondecreasing and the absolute continuity of $\Psi$; cf. Kestelman (1937).] It follows that $\mathscr{L}(T(Z_1, Z_2))$ has support exactly $\mathbf{R}^2$.

Now suppose $\det \Sigma = 0$. If $c^2 = 0$, then $\operatorname{Var} Z_2 = 0$, so $Z_2 = 0$ a.s., whence $\mathscr{L}(T(Z_1, 0))$ has support $\{(x, 0)\colon x \in \mathbf{R}\}$ in $\mathbf{R}^2$. (Note that $Z_1$ is never degenerate unless $X$ itself is degenerate.) If $c^2 > 0$, necessarily (by the proof of Proposition 4.3) $EX^2 = \infty$ and so $Z_1 \sim \mathscr{N}(0, 1)$. Recall that $a^2(X) = 1$ if $EX^2 = \infty$. Thus $0 = \det \Sigma = c^2 - b^2$ and $\mathscr{L}(Z_1, Z_2) = \mathscr{L}(Z_1, cZ_1)$ has support $\{(x, cx)\colon x \in \mathbf{R}^2\}$. It follows that $\mathscr{L}(T(Z_1, Z_2))$ has support

$$\begin{aligned}&\left\{\left(x + \Phi(\Psi^{-1}(cx)), \Psi^{-1}(cx)\right)\colon x \in \mathbf{R}\right\}\\ (5.16)\qquad &\quad = \left\{\left(\frac{1}{c}\Psi(t) + \Phi(t), t\right)\colon t \in \mathbf{R}\right\}.\end{aligned}$$

In summary, the law of the form in (5.3) either has support $\mathbf{R}^2$ (when $\det \Sigma \neq 0$) or its support is a smooth curve, i.e., the image of the real line in $\mathbf{R}^2$ under an absolutely continuous transformation with locally absolutely continuous inverse (when $\det \Sigma = 0$).

Our proof of Theorem 5.1(ii) depends upon the introduction of certain auxiliary measures designed to play the role of the traditional Lévy measures in situations where the latter may not be available. Since the new technique involved is fairly self-contained and, in addition, may be of some independent interest, it is convenient to introduce these auxiliary measures (called *pseudo-Lévy measures*) and collect a few useful facts about them before completing the proof of Theorem 5.1.

5.17. *Pseudo-Lévy measures.* The scale equation $nm(2, a_n) = r_n$ generally provides normalizations suitable for partial sums of length $t_n = [n/r_n] \sim n/r_n$. Define two functions $\Gamma_n^{\pm}$ on $(0, \infty)$ by

$$\Gamma_n^{\pm}(x) = \frac{n}{r_n} G^{\pm}(xa_n). \tag{5.18}$$

When $X$ belongs to the Feller class, the normalized sums

$$\left\{ \frac{1}{a_n} \sum_{j \le t_n} (X_j - \gamma_n) \right\} \tag{5.19}$$

are stochastically compact. In this case, for every subsequence of $\{\Gamma_n^{\pm}\}$ there are right-continuous functions $\Gamma^{\pm}$ and a further subsequence along which $\Gamma_n^{\pm} \to \Gamma^{\pm}$ at every continuity point of the latter. The measure $\mu$ on $\mathbf{R} \setminus \{0\}$ described by

$$\mu(x, \infty) = \Gamma^+(x), \qquad \mu(-\infty, -x) = \Gamma^-(x), \qquad x > 0,$$

is a Lévy measure. Traditional computations show that along the appropriate subsequence,

$$\begin{aligned}
1 &= \lim \frac{n}{r_n} m(2, a_n) \\
&= -\lim \frac{n}{r_n} \int (x^2 \wedge 1)\, dG(xa_n) \\
&\ge \int_{\mathbf{R} \setminus \{0\}} (x^2 \wedge 1)\, d\mu(x), \\
&\quad \lim \frac{n}{r_n} m(3, a_n) \\
&= -\lim \frac{n}{r_n} \int (|x|^3 \wedge 1) \operatorname{sgn}(x)\, d(G^+(xa_n) - G^-(xa_n)) \\
&= \int_{\mathbf{R} \setminus \{0\}} (|x|^3 \wedge 1) \operatorname{sgn}(x)\, d\mu(x), \\
\lim \frac{n}{r_n} m(4, a_n) &= -\lim \frac{n}{r_n} \int (x^4 \wedge 1)\, dG(xa_n) \\
&= \int_{\mathbf{R} \setminus \{0\}} (x^4 \wedge 1)\, d\mu(x).
\end{aligned} \tag{5.20}$$

[Note that strict inequality holds in the first line when $F \in DA(2)$.]

However, when $X$ is not in the Feller class, (5.19) may not even possess any subsequential limit distributions, much less be tight. We will nevertheless require a computationally useful analogue of the Lévy measure $\mu$ for which essential properties such as (5.20) hold. But the measures we require need only be concentrated on $[-1, 1] \setminus \{0\}$. As we will see, the equation $(n/r_n)m(2, a_n) = 1$ affords us some control in (5.18), at least for $0 < x \le 1$.

Define functions $\Omega_n^{\pm}$ on $[0, \infty)$ by

$$\Omega_n^{\pm}(x) = (x^2 \wedge 1)\frac{n}{r_n}G^{\pm}(xa_n). \tag{5.21}$$

In order to provide for relative compactness of $\{\Omega_n^{\pm}\}$ on $[0, \infty)$ and for Helly-selected convergent subsequences, note that $\Omega_n^{\pm}(0) = 0$ and each $\Omega_n^{\pm}$ is right-continuous and of bounded variation on $[0, \infty)$ because $G^{\pm}(\cdot)$ is nonincreasing. For $0 < x \le 1$,

$$\begin{aligned}\Omega_n^{\pm}(x) &\le x^2\frac{n}{r_n}G(xa_n) \le x^2\frac{n}{r_n}m(2, xa_n) = \frac{n}{r_n a_n^2}M(2, a_n x)\\ &\le \frac{n}{r_n a_n^2}M(2, a_n) = \frac{n}{r_n}m(2, a_n) = 1,\end{aligned} \tag{5.22}$$

while for $1 \le x < \infty$,

$$\Omega_n^{\pm}(x) \le \frac{n}{r_n}G(xa_n) \le \frac{n}{r_n}G(a_n) \le \frac{n}{r_n}m(2, a_n) = 1. \tag{5.23}$$

To obtain Helly compactness, it is enough to restrict to $[0, 1]$, since on $[1, \infty)$, each of $\{\Omega_n^{\pm}\}$ are monotone and uniformly bounded due to (5.23).

On $[0, 1]$, write

$$\begin{aligned}\Omega_n^{+}(x) &= \frac{n}{r_n a_n^2}(xa_n)^2 G^{+}(xa_n)\\ &= \frac{n}{r_n a_n^2}E\Big(\big(X^2 \wedge (a_n x)^2\big)I(X \ge 0)\Big) - \frac{n}{r_n a_n^2}EX^2 I(0 \le X \le a_n x),\end{aligned}$$

which reveals $\Omega_n^{+}$ as the difference of two nonnegative, nondecreasing functions, each of which is uniformly bounded, as in (5.22), by 1. A similar representation shows that $\{\Omega_n^{-}\}$ also has uniformly bounded total variation.

Thus $\{\Omega_n^{\pm}\}$ is Helly compact. Along a weakly convergent subsequence, there are right-continuous functions $\Omega^{\pm}$ of bounded variation on $[0, \infty)$ such that $\Omega_n^{\pm}(x) \to \Omega^{\pm}(x)$ at every continuity value $x \ge 0$ of $\Omega^{\pm}$.

Now define $\Gamma^{\pm}(x) = \Omega^{\pm}(x)/x^2$ for $x > 0$ and put $\Gamma = \Gamma^{+} + \Gamma^{-}$ and $\Omega = \Omega^{+} + \Omega^{-}$. Finally, define the functions $\tilde{\Gamma}^{\pm}$ on $(0, \infty)$ by

$$\tilde{\Gamma}^{\pm}(x) = \begin{cases}\Gamma^{\pm}(x), & \text{if } 0 < x < 1,\\ 0, & \text{if } x \ge 1.\end{cases} \tag{5.24}$$

The measure $\tilde{\mu}$ on $\mathbf{R} \setminus \{0\}$ induced by $\tilde{\Gamma}^{\pm}$ is determined by

$$\tilde{\mu}(x, \infty) = \tilde{\Gamma}^{+}(x), \qquad \tilde{\mu}(-\infty, -x) = \tilde{\Gamma}^{-}(x), \qquad x > 0, \tag{5.25}$$

and is concentrated in $[-1,1]\setminus\{0\}$. We shall designate $\tilde{\mu}$ as a pseudo-Lévy measure. Its tails $\tilde{\Gamma}^{\pm}$ are faithful to the original tails $\Gamma_n^{\pm}$ along the subsequence only on $(0,1]$, but unlike genuine Lévy measures, pseudo-Lévy measures such as $\tilde{\mu}$ are always available with the Lévy-like properties desired below.

We need the following facts to hold along the subsequence along which $\Gamma_n^{\pm}\to\Gamma^{\pm}$:

$$\int_{\mathbf{R}\setminus\{0\}} x^2\,d\tilde{\mu}(x)\le 1, \tag{5.26}$$

$$\frac{n}{r_n}m(3,a_n)\to\int_{\mathbf{R}\setminus\{0\}} x^3\,d\tilde{\mu}(x), \tag{5.27}$$

$$\frac{n}{r_n}m(4,a_n)\to\int_{\mathbf{R}\setminus\{0\}} x^4\,d\tilde{\mu}(x). \tag{5.28}$$

To prove (5.26), let $\varepsilon>0$. Then (along the subsequence) (2.7) and (2.3) imply

$$\begin{aligned}1=\frac{n}{r_n}\int_0^1 2xG(xa_n)\,dx&\ge\frac{n}{r_n}\int_\varepsilon^1 2xG(xa_n)\,dx\\ &\to\int_\varepsilon^1 2x\Gamma(x)\,dx=-\int_{(\varepsilon,1]}x^2\,d\Gamma(x)+\Gamma(1)-\varepsilon^2\Gamma(\varepsilon).\end{aligned}\tag{5.29}$$

Since $\varepsilon^2\Gamma(\varepsilon)=\Omega(\varepsilon)\le 1$ by (5.22), we can let $\varepsilon\downarrow 0$ in (5.29) and use monotone convergence. Therefore, $\Omega(0+)=\lim_{\varepsilon\downarrow 0}\Omega(\varepsilon)=\lim_{\varepsilon\downarrow 0}\varepsilon^2\Gamma(\varepsilon)$ exists. If $\Omega(0+)>0$, for some $0<\delta<1$,

$$1\ge\int_0^1 2x\Gamma(x)\,dx=\int_0^1 2x^2\Gamma(x)\frac{dx}{x}\ge\int_0^\delta\Omega(0+)\frac{dx}{x}=\infty, \tag{5.30}$$

a contradiction. Thus, $\Omega(0+)=0$. Then (5.29) implies that

$$\begin{aligned}1\ge\int_0^1 2x\Gamma(x)\,dx&=-\int_{(0,1)}x^2\,d\Gamma(x)+\Gamma(1-)\\ &=\int_{\mathbf{R}\setminus\{0\}}x^2\,d\tilde{\mu}(x),\end{aligned}$$

the desired bound.

For (5.27), let $\varepsilon>0$. Then

$$\begin{aligned}\frac{n}{r_n}m(3,a_n)&=\frac{n}{r_n}\int_0^1 3x^2\{G^+(xa_n)-G^-(xa_n)\}\,dx\\ &=3\int_0^\varepsilon\{\Omega_n^+(x)-\Omega_n^-(x)\}\,dx+3\int_\varepsilon^1\{\Omega_n^+(x)-\Omega_n^-(x)\}\,dx\\ &\equiv w_1(n,\varepsilon)+w_2(n,\varepsilon).\end{aligned}\tag{5.31}$$

Now $|w_1(n, \varepsilon)| \le 3\varepsilon$ by (5.22), whereas

$$
\begin{aligned}
w_2(n, \varepsilon) &\to_{n\to\infty} 3\int_{\varepsilon}^{1}\{\Omega^+(x) - \Omega^-(x)\}\,dx \\
&\to_{\varepsilon\downarrow 0} \int_0^1 \{\Omega^+(x) - \Omega^-(x)\}\,dx
\end{aligned}
\tag{5.32}
$$

by two applications of bounded convergence. Therefore, along the appropriate subsequence,

$$
\begin{aligned}
\frac{n}{r_n} m(3, a_n) &\to 3\int_0^1 x^2\{\Gamma^+(x) - \Gamma^-(x)\}\,dx \\
&= -\int_{(0,1]} x^3\, d\{\Gamma^+(x) - \Gamma^-(x)\} + \Gamma^+(1) - \Gamma^-(1) \\
&= -\int_{(0,1)} x^3\, d\{\Gamma^+(x) - \Gamma^-(x)\} + \Gamma^+(1-) - \Gamma^-(1-) \\
&= \int_{\mathbf{R}\setminus\{0\}} x^3\, d\tilde{\mu}(x),
\end{aligned}
\tag{5.33}
$$

via integration by parts and the fact that $\lim_{x\downarrow 0} x^3\Gamma^{\pm}(x) = \lim_{x\downarrow 0} x\Omega^{\pm}(x) = 0$.

Finally, (5.28) follows similarly, noting

$$
\begin{aligned}
\frac{n}{r_n} m(4, a_n) &\to \int_0^1 4x^3\Gamma(x)\,dx = -\int_{(0,1)} x^4\, d\Gamma(x) + \Gamma(1-) \\
&= \int_{\mathbf{R}\setminus\{0\}} x^4\, d\tilde{\mu}(x).
\end{aligned}
\tag{5.34}
$$

Now the proof of Theorem 5.1 can be completed.

Proof of Theorem 5.1(ii). Given a limit law $v$ represented by (5.3)–(5.4) which is not supported on all of $\mathbf{R}^2$, restrict to a subsequence where (5.2) converges to $v$ and where (4.5) holds. Restrict further so that $\Gamma_n^{\pm} \to \Gamma^{\pm}$ as in the discussion of pseudo-Lévy measures. Our goal is to show that at least one of $\Gamma^{\pm}$ vanishes identically and that if $c^2 > 0$ in (4.5), the other of $\Gamma^{\pm}$ vanishes on $[c, \infty)$ but holds constant at $c^{-2}$ on $(0, c)$. This will verify (5.5), etc., when we show that $c^2 = 0$ forces both of $\Gamma^{\pm}$ to vanish.

Let $\tilde{\mu}$ be the pseudo-Lévy measure corresponding to $\Gamma^{\pm}$ as in (5.24)–(5.25). By (5.27)–(5.28) and (4.5),

$$
b = \int x^3\, d\tilde{\mu}(x), \qquad c^2 = \int x^4\, d\tilde{\mu}(x).
\tag{5.35}
$$

Since $v$ is degenerate, $\det \Sigma = 0$. If $c^2 = 0$, (5.35) shows that $\tilde{\mu}$ has no mass. Via (5.24), both functions $\Gamma^{\pm}$ must vanish identically and (5.5)(a) holds.

So assume $c^2 > 0$. Then $a^2(X) = 1$ so that $b^2 = c^2 > 0$. But from (5.35), the Cauchy–Schwarz inequality and (5.26),

$$\begin{aligned} b^2 &\le \left(\int |x|^3 \, d\tilde{\mu}(x)\right)^2 \le \int x^2 \, d\tilde{\mu}(x) \int x^4 \, d\tilde{\mu}(x) \\ &= c^2 \int x^2 \, d\tilde{\mu}(x) \le c^2 = b^2. \end{aligned} \tag{5.36}$$

The first inequality in (5.36) would be strict unless either $\tilde{\mu}(0,\infty) = 0$ or $\tilde{\mu}(-\infty, 0) = 0$. In either case, (5.24)–(5.25) force one of $\Gamma^{\pm}$ to vanish identically. Suppose it is $\Gamma^-$ which vanishes, which is to say $b > 0$. (The other case is similar.)

The second inequality in (5.36) (the Cauchy–Schwarz) would be strict unless for some $s > 0$,

$$x^4 = sx^3, \quad \text{a.e.-}\tilde{\mu}. \tag{5.37}$$

Since $c^2 > 0$, we must have $s > 0$. Thus $\tilde{\mu}$ is concentrated entirely on $\{s\}$. Necessarily $s \le 1$ due to (5.24)–(5.25). Thus $\Gamma^+(\mathrm{x}) = 0$ for $x \ge s$ and $\Gamma^+(x) = \tilde{\mu}(\mathbf{R} \setminus \{0\}) \equiv j$ for $0 < x < s$.

Finally, equality throughout (5.36) forces

$$1 = \int x^2 \, d\tilde{\mu}(x) = s^2 j \tag{5.38}$$

and thus

$$c^2 = \int x^4 \, d\tilde{\mu}(x) = s^4 j = s^2(s^2 j) = s^2, \tag{5.39}$$

revealing $s = c$ and $j = c^{-2}$. This corresponds to (5.5)(b) and completes the proof of Theorem 5.1. □

5.40. REMARK. To facilitate interpretation of the conditions (5.5), fix a subsequence $\{n'_k\}$ for which they hold and let $m_k = [n_k / r_{n_k}]$. Then there are centerings $\beta_k$ such that either $\mathscr{L}((S_{m_k} - \beta_k)/a_{n'_k}) \to \mathscr{N}(0,1)$ or $\mathscr{L}((S_{m_k} - \beta_k)/a_{n'_k}) \to \mathrm{Pois}(\alpha\delta_c + \beta\delta_{-c})$, the law of an ordinary Poisson random variable with parameter $c^2$, or the law of its negative. Under (5.5), to verify these claims it is necessary only to check the variances condition in the general central limit theorem. But, along the subsequence,

$$1 \ge \frac{n}{r_n} \int_0^{\varepsilon} 2xG(xa_n)\, dx + \frac{n}{r_n} \int_{\varepsilon}^{1} 2xG(xa_n)\, dx.$$

When $c^2 = 0$, letting $n \to \infty$ gives $(n/r_n)\int_{\varepsilon}^1 2xG(xa_n)\,dx \to 0$ and thus $n\,\mathrm{Var}\{(X/a_n) I(|X| \le \varepsilon a_n)\} \to 1$ for each $\varepsilon > 0$. If $c^2 > 0$, then for $0 < \varepsilon < 1$,

$$\lim_{n\to\infty} \frac{n}{r_n} \int_{\varepsilon}^{1} 2xG(xa_n)\, dx = \int_{\varepsilon}^{1} 2x\Gamma(x)\, dx = (c^2 - \varepsilon^2)\frac{1}{c^2} \to 1$$

as $\varepsilon \downarrow 0$. It follows that

$$\lim_{\varepsilon \downarrow 0} \varlimsup_{n \to \infty} n \operatorname{Var}\left\{ \frac{X}{a_n} I(|X| \le \varepsilon a_n) \right\} = 0,$$

as required.

Since the DPA of a Poisson is contained in the DPA of the normal [Gnedenko and Kolmogorov (1954), pages 189–190], the conditions (5.5) can only occur when $X \in$ DPA of the normal. This will be useful when studying full sequential results later. [See also (6.8).]

Next, we consider criteria for the convergence of (5.2) to a given limit law of the allowable form as specified by Theorem 5.1. Because these limit laws are generally nonnormal, certain canonical unique representations of them will be required first in order to develop the necessary conditions for convergence. The establishment of these bivariate canonical representations is more involved than was the case for their second (i.e., scale portion) marginals as considered in the proof of Proposition 4.34. The considerable space required is justified, however, by our consequent ability to simply derive the correct necessary conditions. In the study of a related problem similar to ours but involving trimmed sums [Griffin and Pruitt (1989); Csörgő, Haeusler and Mason (1988a)], necessary conditions for convergence to a given limit law of the allowable type are incomplete due to the (present) lack of suitably unique canonical representations for these limit laws.

Define the class of limit laws $\mathscr{C}$ on $\mathbf{R}^2$ by

$$\mathscr{C} = \Bigg\{ v\colon \text{there exists } \{X_j\} \text{ nondegenerate i.i.d., integers } n_k \to \infty$$

$$\text{and numbers } r_{n_k} \to \infty \text{ such that } r_{n_k}/n_k \to 0 \text{ and} \tag{5.41}$$

$$\mathscr{L}\left( \frac{n_k}{\hat{a}_{n_k}\sqrt{r_{n_k}}} (\hat{\gamma}_{n_k} - \gamma_{n_k}), \frac{\hat{a}_{n_k} - a_{n_k}}{\hat{a}_{n_k}} \sqrt{r_{n_k}} \right) \to v \Bigg\}.$$

Fix $v \in \mathscr{C}$. Let $v$ have first and second marginals $v_1$ and $v_2$, respectively. Since canonical representations for $v$ will only be required when $v_2 \ne \delta_0$, first assume this is the case. In particular, since such a limit can arise only from $\{X_j\}$ with $EX_j^2 = \infty$ (via Propositions 4.3 and 4.34) it suffices to assume that $a^2(X) = 1$ for every representation (5.3)–(5.4) for $v$.

When $v$ is represented in the form (5.3)–(5.4) as described in Theorem 5.1, we will write $v = [\Sigma, \Psi, \Phi]$. It is understood that the matrix $\Sigma$ and the functions $\Psi$ and $\Phi$ obey all the properties listed in Theorem 5.1. When the decomposition $\Phi = \Phi^+ - \Phi^-$ is required, we will write $v = [\Sigma, \Psi, \Phi; \Phi^+, \Phi^-]$.

Suppose $v$ has two representations, $v = [\Sigma_i, \Psi_i, \Phi_i]$, $i = 1, 2$, where

$$\Sigma_i = \begin{pmatrix} 1 & b_i \\ b_i & c_i^2 \end{pmatrix}.$$

Since $v_2 \neq \delta_0$, $c_i^2 > 0$, the proof of Proposition 4.34 shows $\Psi_i/c_i \equiv \Psi_0$ is invariant. Let $\beta_i = b_i/c_i$ and then

$$\Pi_i = \begin{pmatrix} 1 & \beta_i \\ \beta_i & 1 \end{pmatrix}.$$

We have $v = [\Pi_i, \Psi_0, \Phi_i]$, as may be seen by replacing $Z_2$ by $Z_3 = Z_2/c$ and $\Psi$ by $\Psi_0$ in (5.3), so that

$$(Z_1, Z_3) \sim \mathscr{N}\left(0, \begin{pmatrix} 1 & b/c \\ b/c & 1 \end{pmatrix}\right).$$

Without loss of generality, assume $\beta_1^2 \geq \beta_2^2$.

Define transformations $T_i\colon \mathbf{R}^2 \to \mathbf{R}^2$ by $T_i(x, y) = (x + \Phi_i(\Psi_0^{-1}(y)), \Psi_0^{-1}(y))$. Then $T_i$ is continuous and has continuous inverse. We have

$$\mathscr{L}(T_1(W_1, W_2)) = \mathscr{L}(T_2(Y_1, Y_2)), \tag{5.42}$$

where $(W_1, W_2) \sim \mathscr{N}(0, \Pi_1)$ and $(Y_1, Y_2) \sim \mathscr{N}(0, \Pi_2)$. Thus

$$\mathscr{L}\big(T_2^{-1}(T_1(W_1, W_2))\big) = \mathscr{L}(Y_1, Y_2).$$

If $\tilde{\Phi} = \Phi_1 - \Phi_2$, then

$$\mathscr{L}\big(W_1 + \tilde{\Phi}(\Psi_0^{-1}(W_2)), W_2\big) = \mathscr{N}(0, \Pi_2). \tag{5.43}$$

Using elementary properties of the bivariate normal distributions, write $W_1 = \beta_1 W_2 + \sqrt{1 - \beta_1^2}\, W_3$ and $Y_1 = \beta_2 Y_2 + \sqrt{1 - \beta_2^2}\, Y_3$, where $W_2, W_3, Y_2, Y_3$ are i.i.d. standard normal variables. Using independence, rewrite (5.43) as

$$\begin{aligned}
&\mathscr{N}\left(0, \begin{pmatrix} 1 - \beta_1^2 & 0 \\ 0 & 0 \end{pmatrix}\right) * \mathscr{L}\big(\beta_1 W_2 + \tilde{\Phi}(\Psi_0^{-1}(W_2)), W_2\big) \\
&\quad = \mathscr{N}\left(0, \begin{pmatrix} 1 - \beta_1^2 & 0 \\ 0 & 0 \end{pmatrix}\right) * \mathscr{N}\left(0, \begin{pmatrix} \beta_1^2 & \beta_2 \\ \beta_2 & 1 \end{pmatrix}\right) \\
&\quad = \mathscr{N}\left(0, \begin{pmatrix} 1 - \beta_1^2 & 0 \\ 0 & 0 \end{pmatrix}\right) * \mathscr{N}\left(0, \begin{pmatrix} \beta_1^2 - \beta_2^2 & 0 \\ 0 & 0 \end{pmatrix}\right) * \mathscr{N}\left(0, \begin{pmatrix} \beta_2^2 & \beta_2 \\ \beta_2 & 1 \end{pmatrix}\right),
\end{aligned} \tag{5.44}$$

where in the last line the fact $\beta_1^2 \geq \beta_2^2$ has been used. Since the characteristic function of any normal (even degenerate) is never-vanishing, the common normal factor in (5.44) can be cancelled to obtain

$$\begin{aligned}
&\mathscr{L}\big(\beta_1 W_2 + \tilde{\Phi}(\Psi_0^{-1}(W_2)), W_2\big) \\
&\quad = \mathscr{N}\left(0, \begin{pmatrix} \beta_1^2 - \beta_2^2 & 0 \\ 0 & 0 \end{pmatrix}\right) * \mathscr{N}\left(0, \begin{pmatrix} \beta_2^2 & \beta_2 \\ \beta_2 & 1 \end{pmatrix}\right).
\end{aligned} \tag{5.45}$$

The left member of (5.45) is supported on a curve in $\mathbf{R}^2$, whereas if $\beta_1^2 - \beta_2^2 > 0$, the right member will have support on all of $\mathbf{R}^2$. Thus $\beta_1^2 = \beta_2^2 \equiv \beta_0^2$, where always $\beta_0 \geq 0$ is selected. Thus

$$\mathscr{L}(\beta_1 W_2 + \tilde{\Phi}(\Psi_0^{-1}(W_2)), W_2) = \mathscr{N}\left(0, \begin{pmatrix} \beta_2^2 & \beta_2 \\ \beta_2 & 1 \end{pmatrix}\right) = \mathscr{L}(\beta_2 Y_2, Y_2). \tag{5.46}$$

For the last member of (5.46), the law of the ratio of the first marginal variable to the second is the point mass at $\beta_2$. For the same ratio applied to the first member of (5.46),

$$\mathscr{L}\left(\beta_1 + \frac{\tilde{\Phi}(\Psi_0^{-1}(W_2))}{W_2}\right) = \delta_{\beta_2}$$

and then

$$\mathscr{L}\left(\frac{\tilde{\Phi}(\Psi_0^{-1}(W_2))}{W_2}\right) = \delta_{\beta_2 - \beta_1}. \tag{5.47}$$

By continuity of $\tilde{\Phi}$ and elementary properties of $\Psi_0$ and $\mathscr{L}(W_2)$, it is easily shown that necessarily $\tilde{\Phi}(\Psi_0^{-1}(x)) \equiv (\beta_2 - \beta_1)x$, so that $\tilde{\Phi} = (\beta_2 - \beta_1)\Psi_0$ and thus $\Phi_1 = \Phi_2 + (\beta_2 - \beta_1)\Psi_0$. Now $\beta_1^2 = \beta_2^2$ implies either $\beta_1 = \beta_2$ and $\Phi_1 \equiv \Phi_2$, or $\beta_1 = -\beta_2$ and $\Phi_1 = \Phi_2 + 2\beta_2\Psi_0$.

By reversing the preceding development it may be shown that if

$$v = \left[\begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix}, \Psi_0, \Phi\right],$$

then also

$$v = \left[\begin{pmatrix} 1 & -\beta \\ -\beta & 1 \end{pmatrix}, \Psi_0, \Phi + 2\beta\Psi_0\right].$$

To summarize, if $v \in \mathscr{C}$ with $v_2 \neq \delta_0$, then among representations

$$v = \left[\begin{pmatrix} 1 & b \\ b & c^2 \end{pmatrix}, \Psi, \Phi\right]$$

the objects $\Psi_0 = \Psi/c$ and $\beta_0 = \sqrt{b^2/c^2}$ are invariant. There are at most two allowable $\Phi$, and $v$ admits at most two canonical representations

$$\left[\begin{pmatrix} 1 & \beta_0 \\ \beta_0 & 1 \end{pmatrix}, \Psi_0, \Phi_P\right] \quad \text{and} \quad \left[\begin{pmatrix} 1 & -\beta_0 \\ -\beta_0 & 1 \end{pmatrix}, \Psi_0, \Phi_N = \Phi_P + 2\beta_0\Psi_0\right],$$

where $P$ stands for positively correlated and $N$ for negatively correlated, in reference to the bivariate normal law used in representing $v$. The function $\Phi_P$ is unique and $v$ will possess a unique canonical representation if and only if $\beta_0 = 0$.

It is now convenient to introduce another abbreviation. Henceforth write $v = (\beta_0, \Psi_0, \Phi_P)$ when

$$(5.48)\quad v = \left[\begin{pmatrix} 1 & \beta_0 \\ \beta_0 & 1 \end{pmatrix}, \Psi_0, \Phi_P\right] = \left[\begin{pmatrix} 1 & -\beta_0 \\ -\beta_0 & 1 \end{pmatrix}, \Psi_0, \Phi_N = \Phi_P + 2\beta_0\Psi_0\right].$$

Finally, we consider the invariance of the decompositions $\Phi_P = \Phi_P^+ - \Phi_P^-$ and $\Phi_N = \Phi_N^+ - \Phi_N^-$ [cf. (5.11)]. These decompositions are unique only when $c^2$ is fixed among representations

$$\left[\begin{pmatrix} 1 & b \\ b & c^2 \end{pmatrix}, \Psi, \Phi; \Phi^+, \Phi^-\right].$$

Indeed, given $0 < c^2 \le 1$, $\Phi_P^\pm$ will be determined by the equations (which hold due to Theorem 5.1)

$$\begin{aligned} &\Phi_P^\pm(0) = 0, \\ (5.49)\qquad &\Phi_P^+ - \Phi_P^- = \Phi_P, \\ &(\Phi_P^+)' + (\Phi_P^-)' = 1 - \tfrac{1}{2}\Psi' = 1 - \tfrac{1}{2}c\Psi_0', \end{aligned}$$

where the last equation holds off a countable set, with similar determining conditions for $\Phi_N^\pm$. Thus the indexed collections $\{\Phi_{P,c^2}^\pm\colon 0 < c^2 \le 1\}$ and $\{\Phi_{N,c^2}^\pm\colon 0 < c^2 \le 1\}$ are uniquely determined by system (5.49) and its counterpart for $\Phi_N^\pm$. It will occasionally be convenient to augment the canonical representation for $v$ by writing $v = (\beta_0, \Psi_0, \Phi_P; \{\Phi_{P,c}^\pm\})$, where of course $\Phi_{P,c}^+$ determines $\Phi_{P,c}^-$ and then $\Phi_{N,c}^\pm$.

Now consider the degenerate case, $v \in \mathscr{C}$ with $v_2 = \delta_0$. Because in the representation (5.3)–(5.4) for $v$, $\Psi^{-1}(0) = 0$, necessarily

$$v = N\left(0, \begin{pmatrix} a^2(X) & 0 \\ 0 & 0 \end{pmatrix}\right).$$

The convergence criteria can now be stated.

5.50. THEOREM. *Let $v \in \mathscr{C}$.*

(i) *If $v_2 = \delta_0$, write*

$$v = N\left(0, \begin{pmatrix} a^2 & 0 \\ 0 & 0 \end{pmatrix}\right).$$

*Then*

$$\mathscr{L}\left(\frac{n_k}{\hat{a}_{n_k}\sqrt{r_{n_k}}}(\hat{\gamma}_{n_k}-\gamma_{n_k}),\ \frac{\hat{a}_{n_k}-a_{n_k}}{\hat{a}_{n_k}}\sqrt{r_{n_k}}\right)\to v$$

*if and only if*

$$\lim_{k\to\infty}\frac{n_k}{r_{n_k}}m(4,a_{n_k})=0 \quad \text{and} \quad a^2(X)=a^2. \tag{5.51}$$

(ii) *If* $v_2\neq\delta_0$, *write* $v=(\beta_0,\Psi_0,\Phi_P)$. *Then*

$$\mathscr{L}\left(\frac{n_k}{\hat{a}_{n_k}\sqrt{r_{n_k}}}(\hat{\gamma}_{n_k}-\gamma_{n_k}),\ \frac{\hat{a}_{n_k}-a_{n_k}}{\hat{a}_{n_k}}\sqrt{r_{n_k}}\right)\to v$$

*if and only if each of the following holds except at most for countably many values of* $x$:

$$\liminf_{k\to\infty}\frac{n_k}{r_{n_k}}m(4,a_{n_k})>0; \tag{5.52}$$

$$\lim_{k\to\infty}\sqrt{\frac{n_k}{r_{n_k}}}\,\frac{\tilde{m}\left(2,a_{n_k}\left(1+x/\sqrt{r_{n_k}}\right)\right)}{m(4,a_{n_k})}=\frac{1}{2}\Psi_0'(x); \tag{5.53}$$

$$\lim_{k\to\infty}\frac{n_k}{r_{n_k}}\frac{m^2(3,a_{n_k})}{m(4,a_{n_k})}=\beta_0^2; \tag{5.54}$$

*for each subsequence of* $\{n_k\}$ *on which* $\sqrt{\dfrac{n}{r_n}}\dfrac{m(3,a_n)}{\sqrt{m(4,a_n)}}\to\beta_0$,

$$\frac{n}{r_n}\left(G^+\left(a_n\left(1+\frac{x}{\sqrt{r_n}}\right)\right)-G^-\left(a_n\left(1+\frac{x}{\sqrt{r_n}}\right)\right)\right)\to\Phi_P'(x); \tag{5.55}$$

*for each subsequence of* $\{n_k\}$ *on which* $\sqrt{\dfrac{n}{r_n}}\dfrac{m(3,a_n)}{\sqrt{m(4,a_n)}}\to-\beta_0$,

$$\frac{n}{r_n}\left(G^+\left(a_n\left(1+\frac{x}{\sqrt{r_n}}\right)\right)-G^-\left(a_n\left(1+\frac{x}{\sqrt{r_n}}\right)\right)\right)\to\Phi_N'(x). \tag{5.56}$$

(iii) *Moreover, when* (5.52)–(5.54) *hold, so do the following off a countable set: Given a subsequence of* $\{n_k\}$ *on which* $(n/r_n)m(4,a_n)\to c^2\in(0,1]$, *and* *determining* $\{\Phi_{P,c}^{\pm}\}$ *and* $\{\Phi_{N,c}^{\pm}\}$ *via* (5.49),

*for every further subsequence such that* (5.55) *holds*,

$$\frac{n}{r_n}G^{+}\left(a_n\left(1+\frac{x}{\sqrt{r_n}}\right)\right) \to (\Phi_{P,c}^{+})'(x),$$
$$(5.57) \qquad \frac{n}{r_n}G^{-}\left(a_n\left(1+\frac{x}{\sqrt{r_n}}\right)\right) \to (\Phi_{P,c}^{-})'(x);$$

*for every further subsequence such that* (5.56) *holds*,

$$\frac{n}{r_n}G^{+}\left(a_n\left(1+\frac{x}{\sqrt{r_n}}\right)\right) \to (\Phi_{N,c}^{+})'(x),$$
$$(5.58) \qquad \frac{n}{r_n}G^{-}\left(a_n\left(1+\frac{x}{\sqrt{r_n}}\right)\right) \to (\Phi_{N,c}^{-})'(x).$$

PROOF. Let $\mu_n = \mathscr{L}((n/a_n\sqrt{r_n})(\hat{\gamma}_n - \gamma_n), ((\hat{a}_n - a_n)/a_n)\sqrt{r_n})$ and denote the marginals of $\mu_n$ by $\mu_n^i$, $i = 1, 2$. We separate the cases:

(i) If $\mu_{n_k} \to v$, then $\mu_{n_k}^2 \to v_2 = \delta_0$ so that $(n_k/r_{n_k})m(4, a_{n_k}) \to 0$ by Proposition 4.34. Inspection of (5.5)–(5.8) makes it clear that

$$\mathscr{L}\Big(\Big(n_k/a_{n_k}\sqrt{r_{n_k}}\Big)\Big(M_{n_k}(1, a_{n_k}) - M(1, a_{n_k})\Big)\Big) \sim \mu_{n_k}^1 \to \mathscr{N}(0, a^2).$$

Then Proposition 4.3 forces $a^2(X) = a^2$ and thus (5.51) holds.

Conversely, suppose (5.51) holds. Then $\mu_{n_k}^2 \to \delta_0$ by Proposition 4.34 and again (5.5)–(5.8) lead to $\mu_{n_k}^1 \sim \mathscr{L}((n_k/a_{n_k}\sqrt{r_{n_k}})(M_{n_k}(1, a_{n_k}) - M(1, a_{n_k}))) \to \mathscr{N}(0, a^2(X))$ by Proposition 4.3. That $\mu_{n_k} \to v$ follows from these simple marginal convergences is an easy consequence of $v_2 = \delta_0$.

(ii) First suppose that $\mu_{n_k} \to v$. Then $\mu_{n_k}^2 \to v_2$, so that (5.52) and (5.53) hold by Proposition 4.34. Given any subsequence of $\{n_k\}$, restrict to a further subsequence such that (4.5) holds and such that $g_n^{\pm}(y) \to g^{\pm}(y)$ for every $y$, recalling (5.9). Letting $\Phi(x) = \int_0^x g(y)\,dy$, we see that $\Phi'(x) = g(x)$ off of a countable set. Moreover, the argument leading up to (5.13) yields, along this restricted subsequence,

$$v \sim \mu_n \to \left[\begin{pmatrix} 1 & b \\ b & c^2 \end{pmatrix}, \Psi, \Phi\right],$$

where $\Psi = c\Psi_0$. Invariance forces $b^2/c^2 = \beta_0^2$, so that (5.54) holds, due to the separate boundedness of $(n/r_n)m(3, a_n)$ and $(n/r_n)m(4, a_n)$. If $b \ge 0$, then invariance also forces $\Phi = \Phi_P$, and if $b \le 0$, then $\Phi = \Phi_N$.

Conversely, assume (5.52)–(5.56). In particular, $\mu_{n_k}^2 \to v_2$, by Proposition 4.34, so that (4.9) holds by Theorem 4.7. By Theorem 5.1, $\{\mu_{n_k}\}$ is tight. Given any subsequence of $\{n_k\}$, restrict to a further subsequence on which $\mu_n \to \lambda$, say, and then restrict further so that (4.5) holds. By (5.54), $b^2/c^2 = \beta_0^2$. The argument leading to (5.13) in the proof of Theorem 5.1 can be modified slightly to give, on this restricted subsequence,

$$\lambda \sim \mu_n \to \left[\begin{pmatrix} 1 & b \\ b & c^2 \end{pmatrix}, \Psi, \Phi\right],$$

where $\Psi = c\Psi_0$ by (5.54) and where, if $b \geq 0$, $\Phi = \Phi_P$ by (5.55), while if $b \leq 0$, $\Phi = \Phi_N$ by (5.56). Thus

$$\lambda = \left[\begin{pmatrix} 1 & \beta_0 \\ \beta_0 & 1 \end{pmatrix}, \Psi_0, \Phi_P\right] \quad \text{or} \quad \lambda = \left[\begin{pmatrix} 1 & -\beta_0 \\ -\beta_0 & 1 \end{pmatrix}, \Psi_0, \Phi_N\right]$$

and in either case $\lambda = v$. Thus, $\mu_{n_k} \to v$, as desired.

(iii) Assume (5.52)–(5.54) and suppose that for some subsequence of $\{n_k\}$, $(n/r_n)m(4, a_n) \to c^2 \in (0, 1]$. Given a further subsequence where (5.55) holds, Helly-select a further subsequence and nonincreasing nonnegative functions $h^{\pm}$ such that, on this subsequence,

$$\frac{n}{r_n} G^{\pm}\left(a_n\left(1 + \frac{x}{\sqrt{r_n}}\right)\right) \to h^{\pm}(x),$$

for every $x$. Putting $H^{\pm}(x) = \int_0^x h^{\pm}(s)\,ds$, it follows from (5.11) that $H^+ - H^- = \Phi_P$, while (5.53), (5.12) and (2.5) lead to $H^+ + H^- = 1 - \frac{1}{2}c\Psi_0$. Since the solution of system (5.49) for $\Phi_{P,c}^{\pm}$ is unique and $H^{\pm}$ also satisfies (5.49), it follows that $H^+ = \Phi_{P,c}^+$ and $H^- = \Phi_{P,c}^-$. The same argument works when (5.56) holds instead of (5.55), and the proof of Theorem 5.50 is complete. □

Turning to the question of asymptotic normality, under what conditions will *every* subsequential limit of (5.2) be bivariate normal? Recall that a sequence of distributions on $\mathbf{R}^2$ is *stochastically compact* in $\mathbf{R}^2$ if it is tight and if no subsequential limit is supported on a proper hyperplane (i.e., line or point) of $\mathbf{R}^2$.

5.59. Corollary. *The sequence*

$$\left\{\mathscr{L}\left(\frac{n_k}{\hat{a}_{n_k}\sqrt{r_{n_k}}}(\hat{\gamma}_{n_k} - \gamma_{n_k}), \frac{\hat{a}_{n_k} - a_{n_k}}{\hat{a}_{n_k}}\sqrt{r_{n_k}}\right)\right\} \tag{5.60}$$

*is tight with only mean zero bivariate normal subsequential limits if and only if* (4.36) *holds. The sequence* (5.60) *is stochastically compact in* $\mathbf{R}^2$ *with only (nondegenerate) mean zero bivariate normal subsequential limits if and only if* (4.36) *holds and for no subsequence of* $\{n_k\}$ *do both* (4.5) *and* (5.5) *hold.*

Proof. Let $\{\mu_k\}$ denote the sequence (5.60), and denote the marginals of $\mu_k$ by $\mu_k^i$, $i = 1, 2$. If $\{\mu_k\}$ is tight with only normal limits, the same is true of $\{\mu_k^2\}$ and thus (4.36) holds by Proposition 4.35. Also, if a limit of $\{\mu_k\}$ is normal but concentrated on a curve in $\mathbf{R}^2$, that curve will be a line. Hence if $\{\mu_k\}$ is stochastically compact with only normal limits, Theorem 5.1 guarantees that for no subsequence of $\{n_k\}$ will (4.36), (4.5) and (5.5) all hold.

Conversely, assume (4.36). Then (4.9) holds whenever $\rho_{n_k} \to \infty$ and $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$, as in the proof of Proposition 4.35. By Theorem 5.1, $\{\mu_k\}$ is tight. Given any subsequence, restrict so that (4.5) holds and $\mu_k \to v$, say. If $v_2 = \delta_0$, then $v$ is obviously normal by Theorem 5.1. Otherwise $v_2 \neq \delta_0$, so take $v = (\beta_0, \Psi_0, \Phi_P)$. Now apply Theorem 5.50. Then (5.53) holds and so $\Psi_0'$ is constant due to (2.5) and the proof of Proposition 4.35. Thus, $\Psi_0$ is linear. But, recalling the notation (5.9), relations (2.5) and (5.12) and using the nonincreasing nature of $g_n^{\pm}$,

$$
\begin{aligned}
&|(g_n^+(y) - g_n^-(y)) - (g_n^+(0) - g_n^-(0))| \\
&\quad = |(g_n^+(y) - g_n^+(0)) - (g_n^-(y) - g_n^-(0))| \\
&\quad \le |g_n^+(y) - g_n^+(0)| + |g_n^-(y) - g_n^-(0)| \\
&\quad = |(g_n^+(y) + g_n^-(y)) - (g_n^+(0) + g_n^-(0))| \\
&\quad = |(1 + o(1) - \tfrac{1}{2}f_n(y)) - (1 - \tfrac{1}{2}f_n(0))| \\
&\quad = \tfrac{1}{2}|f_n(y) - f_n(0)| + o(1).
\end{aligned}
\tag{5.61}
$$

Letting $n \to \infty$ along the appropriate subsequence, the final term in (5.61) tends to zero for every $y$ due to constancy of $\Psi_0'$. It follows that each of $\Phi_P'$ and $\Phi_N'$ must be constant, by appealing to (5.55)–(5.56). Hence, $\Psi_0$ linear implies that both $\Phi_P$ and $\Phi_N$ are linear. Thus $(\beta_0, \Psi_0, \Phi_P)$ is also normal. Finally, its support can be a proper hyperplane of $\mathbf{R}^2$ only when (4.5) and (5.5) hold along an appropriate subsequence, by the normality and Theorem 5.1. □

Before giving the normal convergence criteria, we need some facts concerning the representation of normal laws in $\mathscr{C}$. If $v \in \mathscr{C}$ is normal and $v_2 = \delta_0$, then

$$
v = \mathscr{N}\left(0, \begin{pmatrix} a^2 & 0 \\ 0 & 0 \end{pmatrix}\right)
$$

for unique $0 < a^2 \le 1$.

If $v \in \mathscr{C}$ and $v_2 \neq \delta_0$, write $v = (\beta_0, \Psi_0, \Phi_P)$ uniquely. If $v = \mathscr{N}(0, \Pi)$, say, then both $\Psi_0$ and $\Phi_P$ are linear, as may be deduced from the proof of Corollary 5.59. Take $\Psi_0(x) = x/\xi$ and $\Phi_P(x) = \theta x$, where $\xi > 0$ by strict monotonicity of $\Psi_0$. Defining

$$
A_P = \begin{pmatrix} 1 & \theta\xi \\ 0 & \xi \end{pmatrix}
\tag{5.62}
$$

yields $v = \mathscr{L}(A_P(Z_1, Z_3)^t)$, where

$$
(Z_1, Z_3) \sim \mathscr{N}\left(0, \begin{pmatrix} 1 & \beta_0 \\ \beta_0 & 1 \end{pmatrix}\right).
$$

Thus, standard covariance computations yield

$$\Pi_0 = \operatorname{Cov} A_p(Z_1, Z_3)^t = A_p\begin{pmatrix} 1 & \beta_0 \\ \beta_0 & 1 \end{pmatrix} A_P^t.$$

Of course, $\Phi_N(x) = \Phi_P(x) + 2\beta_0\Psi_0(x) = (\theta + 2\beta_0/\xi)x$ and

$$\Pi = A_N\begin{pmatrix} 1 & -\beta_0 \\ -\beta_0 & 1 \end{pmatrix} A_N^t,$$

where

$$A_N = \begin{pmatrix} 1 & \theta\xi + 2\beta_0 \\ 0 & \xi \end{pmatrix}.$$

However, the numbers $\theta$, $\xi$ and $\beta_0$ with $|\theta| \le 1$, $\xi > 0$ and $0 \le \beta_0 \le 1$ such that

$$\Pi_0 = \begin{pmatrix} 1 & \theta\xi \\ 0 & \xi \end{pmatrix}\begin{pmatrix} 1 & \beta_0 \\ \beta_0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \theta\xi & \xi \end{pmatrix}, \tag{5.63}$$

are unique, lest the representation $v = \mathscr{N}(0, \Pi_0) = (\beta_0, \Psi_0(x) = x/\xi, \Psi_P(x) = \theta x)$ be nonunique. Moreover, given $\Pi_0$, the matrix $A_P$ in (5.62) can be determined uniquely by (5.63) and the bounds on $\theta$, $\xi$ and $\beta_0$, via ordinary (if slightly tedious) algebra.

5.64. COROLLARY. *Let $v \in \mathscr{C}$ with $v = \mathscr{N}(0, \Pi_0)$.*

(i) *If*

$$\Pi_0 = \begin{pmatrix} a^2 & 0 \\ 0 & 0 \end{pmatrix}$$

*with $0 < a^2 \le 1$, then*

$$\mathscr{L}\left(\frac{n_k}{\hat{a}_{n_k}\sqrt{r_{n_k}}}(\hat{\gamma}_{n_k} - \gamma_{n_k}), \frac{\hat{a}_{n_k} - a_{n_k}}{\hat{a}_{n_k}}\sqrt{r_{n_k}}\right) \to v$$

*if and only if* (5.51) *holds.*

(ii) *If*

$$\Pi_0 = A_P\begin{pmatrix} 1 & \beta_0 \\ \beta_0 & 1 \end{pmatrix} A_P^t$$

*with $A_P$ as in* (5.62), *then*

$$\mathscr{L}\left(\frac{n_k}{\hat{a}_{n_k}\sqrt{r_{n_k}}}(\hat{\gamma}_{n_k} - \gamma_{n_k}), \frac{\hat{a}_{n_k} - a_{n_k}}{\hat{a}_{n_k}}\sqrt{r_{n_k}}\right) \to v$$

*if and only if* (4.36), (5.52) *and* (5.54) *hold as well as the following*:

(5.65)
$$\lim_{k\to\infty}\sqrt{\frac{n_k}{r_{n_k}}}\,\frac{\tilde{m}(2,a_{n_k})}{\sqrt{m(4,a_{n_k})}}=\frac{\xi}{2}.$$

*On every subsequence of* $\{n_k\}$ *such that* $\sqrt{\frac{n}{r_n}}\frac{m(3,a_n)}{\sqrt{m(4,a_n)}}\to\beta_0$,

(5.66)
$$\frac{n}{r_n}(G^+(a_n)-G^-(a_n))\to\theta.$$

*On every subsequence of* $\{n_k\}$ *such that* $\sqrt{\frac{n}{r_n}}\frac{m(3,a_n)}{\sqrt{m(4,a_n)}}\to-\beta_0$,

(5.67)
$$\frac{n}{r_n}(G^+(a_n)-G^-(a_n))\to\theta+2\beta_0/\xi.$$

PROOF. (i) was part of Theorem 5.50. (ii) is a direct consequence of the invariance in representing normal $v\in\mathscr{C}$ and by a direct application of Corollary 5.59 and Theorem 5.50. □

Inspection of the correlation coefficient of the normal distributions $N(0,\Pi_0)$ in Corollary 5.64 sheds light on the asymptotic dependence between the reduced quantities $(n/\hat{a}_n\sqrt{r_n})(\hat{\gamma}_n-\gamma_n)$ and $((\hat{a}_n-a_n)/a_n)\sqrt{r_n}$. The importance of the quantity $\beta_0$ there (and its relative $b$ in Proposition 4.3), which is essentially determined by censored third moments, recalls the role played by the classical quantity $EX^3$. Particularly important is the case of asymptotic symmetry [i.e., (4.6)] as might be surmised from Theorem 3.6. Recall the definition (4.2) for the quantity $a^2(X)$.

5.68. COROLLARY. *Assume* (4.6) *and* (4.9) *hold for every* $\rho_{n_k}\to\infty$ *such that* $\rho_{n_k}/\sqrt{r_{n_k}}\to 0$. *Then*

(5.69)
$$\mathscr{L}\left(\frac{n_k}{\hat{a}_{n_k}\sqrt{r_{n_k}}}(\hat{\gamma}_{n_k}-\gamma_{n_k})\right)\to\mathscr{N}(0,a^2(X)).$$

*Furthermore*,

$$\frac{n_k}{\hat{a}_{n_k}\sqrt{r_{n_k}}}(\hat{\gamma}_{n_k}-\gamma_{n_k})\quad\textit{and}\quad\frac{\hat{a}_{n_k}-a_{n_k}}{\hat{a}_{n_k}}\sqrt{r_{n_k}}$$

*are asymptotically independent*.

PROOF. Recalling (5.9), asymptotic symmetry (4.6) forces $g_n^+-g_n^-\to 0$. Since (4.9) holds, Theorem 5.1 guarantees that every limit (5.3) of the sequence (5.2) has the form $\mathscr{L}(Z_1,\Psi^{-1}(Z_2))$, where

$$\mathscr{L}(Z_1,Z_2)=\mathscr{N}\left(0,\begin{pmatrix}a^2(X)&0\\0&c^2\end{pmatrix}\right).$$

Both claims of the corollary now follow immediately. □

The behavior of the first marginal, that is, the studentized quantity $(n(\hat{\gamma}_n - \gamma_n))/\hat{a}_n\sqrt{r_n}$, is of independent interest. When $X$ is symmetric, Proposition 4.3 and Theorem 3.6 suggest that generally the integral term in (5.6) tends to zero in probability (and thus plays no essential role in the weak convergence study). Now, under (4.9), Corollary 5.68 guarantees exactly this behavior under asymptotic symmetry, regardless of the tail of $\mathscr{L}(X)$. However, if $X$ is almost surely nonnegative and also outside the domain of partial attraction of the normal, Proposition 4.3 and (5.6)–(5.8) make it easy to see that consistency ($\hat{a}_n/a_n \to 1$ in probability) and tightness of $(n/\hat{a}_n\sqrt{r_n})(\hat{\gamma}_n - \gamma_n)$ in (5.6) force tightness of $((\hat{a}_n - a_n)/a_n)\sqrt{r_n}$ and hence also the validity of the tail-controlling condition (4.9). For general asymmetric $X$, therefore, the precise necessary and sufficient conditions for tightness and/or asymptotic normality of $(n/\hat{a}_n\sqrt{n})(\hat{\gamma}_n - \gamma_n)$ will involve quantification of the delicate playoff between heaviness of tails versus balance of tails (near-symmetry) in expressions such as (5.6) and (5.8). For brevity, we only consider asymptotic normality for $(n/\hat{a}_n\sqrt{r_n})(\hat{\gamma}_n - \gamma_n)$ under conditions sufficient for its tightness without supplementary tail-balance assumptions. The general case is being investigated elsewhere.

5.70. Corollary. *Assume* (4.9) *holds whenever* $\rho_{n_k} \to \infty$ *and* $\rho_{n_k}/\sqrt{r_{n_k}} \to 0$. *Then every subsequential limiting distribution of* $(n_k/\hat{a}_{n_k}\sqrt{r_{n_k}})(\hat{\gamma}_{n_k} - \gamma_{n_k})$ *will be normal or degenerate, if and only if for all* $y$,

$$\lim_{k\to\infty}\left\{\frac{G^+\left(a_{n_k}\left(1+y/\sqrt{r_{n_k}}\right)\right) - G^-\left(a_{n_k}\left(1+y/\sqrt{r_{n_k}}\right)\right)}{\tilde{m}\left(2, a_{n_k}\left(1+y/\sqrt{r_{n_k}}\right)\right)} - \frac{G^+(a_{n_k}) - G^-(a_{n_k})}{\tilde{m}(2, a_{n_k})}\right\} = 0. \tag{5.71}$$

Proof. Inspection of limit laws of the form in (5.3) shows that the desired normality/degeneracy occurs exactly when every pair $\Psi, \Phi$ resulting from Helly selections satisfies $\Phi \circ \Psi^{-1}$ is linear. This amounts to every quotient $\Phi'/\Psi'$ being constant, since by Theorem 5.1, $\Psi' > 0$. Recalling (4.18), (4.20) and (4.22) along with (5.9) and (5.11) [where $g^+, g^-$ are Helly selections from (5.9)] establishes the corollary, since (5.71) is simply the condition that every such quotient $(g^+ - g^-)/f$ be constant. □

The remainder of this section will be devoted to the statistical problem of empirically determining the limiting covariances $\Pi_0$ in the setting of Corollary 5.64. The emphasis will be on obtaining full sequential convergence to the *standard* normal (in $\mathbf{R}$ or $\mathbf{R}^2$) for empirical normalizations of $(\hat{\gamma}_n - \gamma_n)$ and $(\hat{\gamma}_n - \gamma_n, \hat{a}_n - a_n)$, respectively, under appropriate conditions, leading to an invariance principle. In this way our main empirical result, Theorem 5.81

represents the analogue in the censored setting of the classical results which hold under finite second or fourth moments.

For the remainder of this section let $n_k = k$ in order to consider the full sequence. *Assume the asymptotic normality condition* (4.36) *holds*—again, along the full sequence.

Now along subsequences on which the limit of (5.2) is of the form of Corollary 5.64(i) (i.e., degenerate normal) it is not possible to estimate the (nonexistent) quantity $\beta_0$. It is therefore desirable and convenient to have an alternative, common representation for the covariances $\Pi_0$ in Corollary 5.64(i) and (ii) in terms of quantities that can be empirically determined in the limit. Temporarily viewing $\beta_0$ as $b/c$ and $\xi$ as $c/\mu$ in Corollary 5.64(ii), we can rewrite

$$\Pi_0 = \begin{pmatrix} 1 & \theta/\mu \\ 0 & 1/\mu \end{pmatrix} \begin{pmatrix} a^2(X) & b \\ b & c^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \theta/\mu & 1/\mu \end{pmatrix}, \tag{5.72}$$

where $b$, $\theta$, $\mu$ and $c^2$ are subsequential limits of, respectively,

$$\begin{aligned} b_n &= \frac{n}{r_n} m(3, a_n), \\ \theta_n &= \frac{n}{r_n} (G^+(a_n) - G^-(a_n)), \\ \mu_n &= \frac{n}{r_n} \tilde{m}(2, a_n), \\ c_n^2 &= \frac{n}{r_n} m(4, a_n). \end{aligned} \tag{5.73}$$

However, the form (5.72)–(5.73) also represents $\Pi_0$ as in Corollary 5.64(i), i.e., even along a subsequence where $c_n^2 \to 0 = c^2$.

The quantities appearing in (5.73) can be consistently estimated. Defining the obvious estimators, we will show that

$$\hat{b}_n - b_n \equiv \frac{n}{r_n} m_n(3, \hat{a}_n) - \frac{n}{r_n} m(3, a_n) \to_p 0, \tag{5.74i}$$

$$\begin{aligned} \hat{\theta}_n - \theta_n \equiv{}& \frac{n}{r_n} (G_n^+(\hat{a}_n) - G_n^-(\hat{a}_n)) \\ &- \frac{n}{r_n} (G^+(a_n) - G^-(a_n)) \to_p 0, \end{aligned} \tag{5.74ii}$$

$$\hat{\mu}_n - \mu_n \equiv \frac{n}{r_n} \tilde{m}_n(2, \hat{a}_n) - \frac{n}{r_n} \tilde{m}(2, a_n) \to_p 0, \tag{5.74iii}$$

$$\hat{c}_n^2 - c_n^2 \equiv \frac{n}{r_n} m_n(4, \hat{a}_n) - \frac{n}{r_n} m(4, a_n) \to_p 0. \tag{5.74iv}$$

But first a consistent estimate of $a^2(X)$ must be produced. The obvious choice is $\hat{A}_n^2 = 1 - n\hat{\gamma}_n^2/(r_n\hat{a}_n^2) = 1 - \hat{\gamma}_n^2/M_n(2, \hat{a}_n)$, for then $\hat{A}_n^2$ is the sample variance for the sample $\{(|X_j| \wedge \hat{a}_n)\mathrm{sgn}(X_j)/(\sqrt{r_n}\,\hat{a}_n): j \le n\}$. Now if $EX^2 < \infty$, then $((\hat{a}_n - a_n))/a_n\sqrt{r_n} \to_p 0$, so with probability tending to 1,

$$M_n\left(2, a_n\left(1 - \frac{1}{\sqrt{r_n}}\right)\right) \le M_n(2, \hat{a}_n) \le M_n\left(2, a_n\left(1 + \frac{1}{\sqrt{r_n}}\right)\right). \tag{5.75}$$

It is easy to check that $M_n(2, a_n(1 \pm 1/\sqrt{r_n})) - M(2, a_n(1 \pm 1/\sqrt{r_n})) \to_p 0$. Since $M(2, a_n(1 \pm 1/\sqrt{r_n})) \to EX^2 < \infty$, (5.75) forces $M_n(2, \hat{a}_n) \to_p EX^2$. But $E|X| < \infty$ as well, so $\hat{\gamma}_n \to_p EX$. It follows that $\hat{A}_n^2 \to_p \mathrm{Var}(X)/EX^2 = a^2(X)$, when $EX^2 < \infty$.

If, however, $EX^2 = \infty$, then $a^2(X) = 1$ and it must be proven that $\hat{\gamma}_n^2/M_n(2, \hat{a}_n) \to_p 0$. Observe that by (4.36) [which implies (4.9)] and Theorem 5.1,

$$\begin{aligned}\frac{\hat{\gamma}_n^2}{M_n(2, \hat{a}_n)} &= \frac{\hat{\gamma}_n^2/\hat{a}_n^2}{m_n(2, \hat{a}_n)} = \frac{n}{r_n}\left(\frac{\hat{\gamma}_n}{\hat{a}_n}\right)^2 \\ &= \frac{n}{r_n}\left(\frac{n(\hat{\gamma}_n - \gamma_n)}{\hat{a}_n\sqrt{r_n}}\,\frac{\sqrt{r_n}}{n} + \frac{\gamma_n}{a_n}\,\frac{a_n}{\hat{a}_n}\right)^2 \\ &= \frac{n}{r_n}\left(O_p\left(\frac{\sqrt{r_n}}{n}\right) + \frac{\gamma_n}{a_n}(1 + O_p(1))\right)^2 \\ &= O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{r_n}}\right)\frac{\gamma_n}{a_n} + \frac{n}{r_n}\left(\frac{\gamma_n}{a_n}\right)^2(1 + O_p(1)).\end{aligned} \tag{5.76}$$

But, since $EX^2 = \infty$, we have $(E(|X| \wedge t))^2 = o(E(X^2 \wedge t^2))$ as $t \to \infty$, so that

$$\gamma_n^2 \le (E|X| \wedge a_n)^2 = o(EX^2 \wedge a_n^2) = o(a_n^2 m(2, a_n)) = o\left(a_n^2 \frac{r_n}{n}\right)$$

and therefore (5.76) forces $\hat{\gamma}_n^2/M_n(2, \hat{a}_n) \to_p 0$ and $\hat{A}_n^2 \to_p a^2(X)$, as desired.

Turning to (5.74), we will verify (ii) here; the proofs for the other cases are similar and even easier. It will be clear that due to continuity considerations à la (2.5) and analogues, (i) and (iv) depend only on (4.9), although (ii) and (iii) appear to require the full strength of (4.36).

Under (4.36), a diagonalization argument shows there exists $0 < \rho_n \to \infty$ such that $\rho_n/\sqrt{r_n} \to 0$ and

$$\frac{n}{r_n}\left(G\left(a_n\left(1 \pm \frac{\rho_n}{\sqrt{r_n}}\right)\right) - G(a_n)\right) \to 0. \tag{5.77}$$

Since $\{((\hat{a}_n - a_n)/a_n)\sqrt{r_n}\}$ is tight, with probability tending to 1,

$$(5.78)\qquad G_n\left(a_n\left(1+\frac{\rho_n}{\sqrt{r_n}}\right)\right) \le G_n(\hat{a}_n) \le G_n\left(a_n\left(1-\frac{\rho_n}{\sqrt{r_n}}\right)\right).$$

Now

$$\operatorname{Var}\frac{n}{r_n}G_n\left(a_n\left(1\pm\frac{\rho_n}{\sqrt{r_n}}\right)\right) \le \frac{n}{r_n^2}G\left(a_n\left(1\pm\frac{\rho_n}{\sqrt{r_n}}\right)\right) = O\left(\frac{1}{r_n}\right),$$

so that (5.77) guarantees

$$(5.79)\qquad \frac{n}{r_n}\left(G_n\left(a_n\left(1\pm\frac{\rho_n}{\sqrt{r_n}}\right)\right) - G(a_n)\right) \to_p 0.$$

The argument in (5.61) shows that (5.79) guarantees

$$\frac{n}{r_n}\left|G_n^+\left(a_n\left(1\pm\frac{\rho_n}{\sqrt{r_n}}\right)\right)-G^+(a_n)\right| + \frac{n}{r_n}\left|G_n^-\left(a_n\left(1\pm\frac{\rho_n}{\sqrt{r_n}}\right)\right)-G^-(a_n)\right| \to_p 0,$$

i.e., $\hat{\theta}_n - \theta_n \to_p 0$.

We note that the impossibility of singular covariances $\Pi_0$ in Corollary 5.64 is equivalent to (4.28), since we are proceeding along the full sequence, so that by Remark 5.40, degenerate limits in Theorem 5.1 occur if and only if $X$ is in the domain of partial attraction of the normal.

Putting [appropriate to (5.72) and (5.74)]

$$(5.80)\qquad \hat{\Pi}_n = \begin{pmatrix} 1 & \hat{\theta}_n/\hat{\mu}_n \\ 0 & 1/\hat{\mu}_n \end{pmatrix}\begin{pmatrix} \hat{A}_n^2 & \hat{b}_n \\ \hat{b}_n & \hat{c}_n \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \hat{\theta}_n/\hat{\mu}_n & 1/\hat{\mu}_n \end{pmatrix}$$

and taking $\hat{\Pi}_n^{-1/2} = I$ if $\hat{\Pi}_n$ is singular, we have the following analogue to the classical results.

5.81. THEOREM. *Let $n_k = k$ and assume* (4.36). *Then*

$$(5.82)\qquad \mathscr{L}\left(\left(\hat{A}_n^2 + \frac{2\hat{\theta}_n\hat{b}_n}{\hat{\mu}_n} + \frac{\hat{\theta}_n^2\hat{c}_n^2}{\hat{\mu}_n^2}\right)^{-1/2}\frac{n}{\hat{a}_n\sqrt{r_n}}(\hat{\gamma}_n - \gamma_n)\right) \to N(0,1).$$

*If in addition* (4.28) *holds, then*

$$(5.83)\qquad \mathscr{L}\left(\hat{\Pi}_n^{-1/2}\left(\frac{n(\hat{\gamma}_n-\gamma_n)}{\hat{a}_n\sqrt{r_n}}, \frac{\hat{a}_n - a_n}{\hat{a}_n}\sqrt{r_n}\right)^t\right) \to N(0,I)$$

*and, in particular, $\hat{\Pi}_n$ is nonsingular with probability tending to* 1.

5.84. REMARK. The subsequential analogues of (5.82) and (5.83) are available even without (4.28) if (4.36) holds only on the given subsequence and if on no further subsequence do (4.5) and (5.5) both hold.

## 6. Applications and examples.

6.1. *Domains of attraction.* Fix a sequence $\{r_n\}$ as in (4.1). The various empirical results of Section 5 are considerably simplified for $F \in \mathrm{DA}(\alpha)$. In particular, *full sequential joint asymptotic normality for the self-normalized quantities is available throughout the entire class* $\mathrm{DA}(\alpha)$, $0 < \alpha \le 2$ [compare Davis and Resnick (1984)].

6.2. THEOREM. *Suppose* $F \in \mathrm{DA}(\alpha)$, *where* $0 < \alpha \le 2$. *Then*

$$\mathscr{L}\left(\frac{n}{\hat{a}_n\sqrt{r_n}}(\hat{\gamma}_n - \gamma_n), \frac{\hat{a}_n - a_n}{\hat{a}_n}\sqrt{r_n}\right) \to \mathscr{N}(0, \Pi_0),$$

*where*

$$\Pi_0 = \begin{pmatrix} 1 & \dfrac{(2p-1)(2-\alpha)}{\alpha} \\ 0 & \dfrac{2}{\alpha} \end{pmatrix} \begin{pmatrix} 1 & \dfrac{3(2-\alpha)(2p-1)}{2(3-\alpha)} \\ \dfrac{3(2-\alpha)(2p-1)}{2(3-\alpha)} & \dfrac{4-2\alpha}{4-\alpha} \end{pmatrix}$$

$$\times \begin{pmatrix} 1 & 0 \\ \dfrac{(2p-1)(2-\alpha)}{\alpha} & \dfrac{2}{\alpha} \end{pmatrix}. \tag{6.3}$$

*Moreover,*

$$\mathscr{L}\left(\left(\hat{A}_n^2 + \frac{2\hat{\theta}_n\hat{b}_n}{\hat{\mu}_n} + \frac{\hat{\theta}_n^2\hat{c}_n^2}{\hat{\mu}_n^2}\right)^{-1/2} \frac{n}{\hat{a}_n\sqrt{r_n}}(\hat{\gamma}_n - \gamma_n)\right) \to \mathscr{N}(0,1), \tag{6.4}$$

*where* $\hat{A}_n^2 = 1 - n\hat{\gamma}_n^2/(r_n\hat{a}_n^2)$ *and* $\hat{b}_n$, $\hat{c}_n$, $\hat{\mu}_n$ *and* $\hat{\theta}_n$ *are given by* (5.74). *If* $\alpha = 2$, *then*

$$\frac{\hat{a}_n - a_n}{\hat{a}_n}\sqrt{r_n} \to 0 \quad \textit{in probability}. \tag{6.5}$$

*Finally, if* $\alpha \ne 2$,

$$\mathscr{L}\left(\hat{\Pi}_n^{-1/2}\left(\frac{n}{\hat{a}_n\sqrt{r_n}}(\hat{\gamma}_n - \gamma_n), \frac{\hat{a}_n - a_n}{\hat{a}_n}\sqrt{r_n}\right)\right) \to \mathscr{N}(0, I), \tag{6.6}$$

*where* $\hat{\Pi}_n$ *is given by* (5.80) *and* $\hat{\Pi}_n^{-1/2}$ *is defined to be* $I$ *when* $\hat{\Pi}_n$ *is singular.*

6.7. REMARKS. The proof of Theorem 6.2 is an application of Theorems 5.1 and 5.81 involving relatively routine computation.

Theorem 6.2 motivates a new approach to the statistical problem of estimating an exponent of regular variation in the tail of a distribution. Development of this approach, together with a fuller discussion of the proof and statistical implications of a generalization of Theorem 6.2 may be found in Hahn, Kuelbs and Weiner (1989b).

6.8. *Remarks on the Feller class.* Many of the results of Section 5 improve and simplify when $X$ is in the Feller class, i.e., (4.27) holds. In Corollary 4.29 it was shown that (4.9) holds along the full sequence $\{n_k = k\}$. Thus subsequential results can be extended to the full sequence. Also, bounds on the quantities appearing in the various results can be given via (4.27), for example, in (5.63), $\xi \geq 1/(2(1 - \Lambda))$. Moreover, Pruitt's 1983 result shows that under (4.27), $X$ cannot be in the domain of partial attraction of an ordinary Poisson law. Hence, Remark 5.40 shows that the only limit laws of the sequence in (5.2) of a degenerate form in (5.4) will be

$$N\left(0, \begin{pmatrix} a^2 & 0 \\ 0 & 0 \end{pmatrix}\right),$$

where $a^2 = a^2(X)$. In particular when $X$ is in the Feller class, *a necessary and sufficient condition that with $n_k = k$, every limit from* (5.2) *have support exactly $\mathbf{R}^2$, is that $X$ be outside the domain of partial attraction of the normal*, i.e., *that* (4.28) *hold*. When (4.28) fails, there will be degenerate limits and they will be concentrated on the line $\{(x, 0)\colon x \in \mathbf{R}\}$.

6.9. *Examples of joint asymptotic normality.* Combining Corollary 5.59 with the preceding remarks on the Feller class enables us to present a relatively large class of examples generating joint asymptotic normality (with nonsingular limiting covariance) for the quantities $\hat{\gamma}_n$ and $\hat{a}_n$. In particular, the invariance principle, Theorem 5.81 applies. Now in Theorem 6.2 it was seen that $X \in DA(\alpha)$ $(0 < \alpha < 2)$ generates exactly this desired behavior, but here we show that regular variation/balance in the tails $G, G^{\pm}$ can be replaced by mild assumptions concerning the smoothness of $G$ and $\tilde{m}(2, \cdot\,)$. Suppose the distribution $H$ of the random variable $|X|$ has Lebesgue density $f$, such that for some $T > 0$, either $f$ is continuous on $(T, \infty)$ or the function $g(t) = tf(t)$ is nonincreasing on $(T, \infty)$. Suppose, further, that

$$0 < \liminf_{\xi \to \infty} \frac{\xi^3 f(\xi)}{\int_0^\xi 2s^2 f(s)\, ds} \leq \limsup_{\xi \to \infty} \frac{\xi^3 f(\xi)}{\int_0^\xi 2s^2 f(s)\, ds} < 1. \tag{6.10}$$

[In particular, $\tilde{m}(2, \cdot\,)$ varies "dominatedly" at $\infty$, since we will see that $\tilde{m}(2, \cdot\,)$ is monotone; cf. Seneta (1976).] We claim that

(i) $X$ is in the Feller class,
(ii) $X \notin \mathrm{DPA}(N(0,1))$,
(iii) (4.36) holds.

The desired nonsingular joint asymptotic normality will then follow from the aforementioned results. If $f$ is continuous on $(T, \infty)$, an application of the extended (or Cauchy) mean value theorem as in the proof of l'Hospital's rule leads to

$$\begin{aligned}\liminf_{t\to\infty}\frac{G(t)}{m(2,t)} &\geq \liminf_{\xi\to\infty}\frac{f(\xi)}{2\tilde{m}(2,\xi)/\xi}\\ &= \liminf_{\xi\to\infty}\frac{\xi^3 f(\xi)}{2\int_0^\xi s^2 f(s)\,ds} > 0,\end{aligned}\tag{6.11}$$

recalling (2.4) and $\tilde{m}(2,x) = x^{-2}\int_0^x s^2 f(s)\,ds$. But monotonicity of $tf(t)$ makes it easy to remove the assumption of continuity in (6.11). [Indeed, this monotonicity assumption could be replaced by any condition validating (6.11).] Claim (ii) follows and claim (i) is obtained in the same manner. For (iii), we will first show that for some $T' \geq T$, $\tilde{m}(2,\cdot)$ is nonincreasing on $(T',\infty)$. From $G + \tilde{m}(2,\cdot) = m(2,\cdot)$ it will then follow that for $b > a > T'$, $0 \leq G(a) - G(b) \leq m(2,a) - m(2,b)$. Then (4.36) will follow (for $x \leq 0$) from (i) and

$$\begin{aligned}&\frac{n}{r_n}\left\{G\left(a_n\left(1+\frac{x}{\sqrt{r_n}}\right)\right) - G(a_n)\right\}\\ &\leq \frac{n}{r_n}\left\{m\left(2, a_n\left(1+\frac{x}{\sqrt{r_n}}\right)\right) - m(2,a_n)\right\}\\ &= \frac{n}{r_n}m(2,a_n)\left\{\frac{m\left(2,a_n\left(1+x/\sqrt{r_n}\right)\right)}{m(2,a_n)} - 1\right\} \to 0,\end{aligned}\tag{6.12}$$

by (2.5) and (2.7), with a similar argument applying when $x > 0$.

To see that $\tilde{m}(2,\cdot)$ is eventually nonincreasing, use the quotient rule of differentiation to write (for large $t$ and off a countable set)

$$\begin{aligned}\frac{d}{dt}\tilde{m}(2,t) &= t^{-4}\left\{t^4 f(t) - 2t\int_0^t s^2 f(s)\,ds\right\}\\ &= 2t^{-3}\int_0^t s^2 f(s)\,ds\left\{\frac{t^3 f(t)}{\int_0^t 2sf(s)\,ds} - 1\right\}\\ &< 0,\end{aligned}\tag{6.13}$$

by (6.10).

It is interesting to note that the essential feature of these examples, namely, that $\tilde{m}(2,\cdot)$ is nonincreasing on $(T',\infty)$, actually forces $F$ to have a Lebesgue density on $(-\infty,-T'] \cup [T',\infty)$. To see this, note as above that $\tilde{m}(2,\cdot)$ nonincreasing forces $0 \leq G(a) - G(b) \leq m(2,a) - m(2,b)$ for $b > a > T'$, so that on $(T',\infty)$, the measure induced by $G$ is dominated by the measure induced by $m(2,\cdot)$. The latter is absolutely continuous with respect to Lebesgue measure by (2.4). Finally, the measures induced by $G^{\pm}$ are dominated by that induced by $G$, and thus $G^{\pm}$ are absolutely continuous functions. Hence, $F$ is absolutely continuous. Examples of probability densities $f$ as required are easy to generate: Take any random variable $Y$ in the Feller class but outside the DPA($N(0,1)$) and put $f(t) = ct^{-1}P(|Y| > t)h(t)$, where $h$ is slowly varying at $\infty$, $h$ is eventually nonincreasing, $\int_0^1 (h(t)/t)\,dt < \infty$ and $c$ is the suitable normalizing constant. [Note that $P(|Y| > t)$ varies dominatedly.] Then (6.10) is immediate.

Finally, we present a computationally convenient and illuminating example where even the simple consistent estimation of scale (i.e., $\hat{a}_n/a_n \to 1$ in probability) can fail. Naturally, our choice is a distribution with slowly varying tail.

6.14. *Example of inconsistent scale estimation.* Since the behavior of $\hat{a}_n$ as governed by Theorem 4.7 depends only on the distribution of $|X|$, it is enough to describe the latter. Let $Y = |X|$ satisfy, for each Borel set $A \subset \mathbf{R}$,

$$(6.15)\qquad P(Y \in A) = \frac{1}{2}I_A(0) + \int_{(e,\infty)\cap A} \{(\log x)^{-2} - (\log x)^{-3}\}\frac{dx}{x}.$$

Thus,

$$(6.16)\qquad G(t) = P(Y > t) = \begin{cases} \dfrac{1}{\log x} - \dfrac{1}{2(\log x)^2}, & \text{if } x \ge e,\\[2ex] \dfrac{1}{2}, & \text{if } 0 \le x \le e,\end{cases}$$

so that for $t \ge e$,

$$(6.17)\qquad m(2,t) = t^{-2}\int_0^t 2xG(x)\,dx = \frac{1}{\log t} - \frac{e^2}{2t^2}.$$

Fix $r_n \to \infty$ such that $r_n/n \to 0$ and fix $\rho_n \to \infty$ such that $\rho_n/\sqrt{r_n} \to 0$. Computations show that, relevant to (4.8)–(4.9), one has

$$(6.18)\qquad \begin{aligned} &\frac{n}{\sqrt{r_n}}\left\{m\left(2, a_n\left(1 - \frac{\rho_n}{\sqrt{r_n}}\right)\right) - m(2,a_n)\right\}\\ &\qquad = \frac{r_n\rho_n}{n}(1 + o(1)) - \frac{e^2}{2}\frac{n\rho_n}{r_n a_n^2}(1 + o(1)). \end{aligned}$$

But the size of $a_n$ can be (grossly) estimated from

$$(6.19)\qquad 1 = \frac{n}{r_n}m(2,a_n) = \frac{1}{\log a_n}(1 - o(1))\frac{n}{r_n}$$

to satisfy $a_n = \exp((n/r_n)(1 + o(1)))$. Thus in (6.18), clearly (whatever $\rho_n$ is)

$$(6.20)\qquad \frac{n}{\sqrt{r_n}}\left\{m\left(2, a_n\left(1 - \frac{\rho_n}{\sqrt{r_n}}\right)\right) - m(2,a_n)\right\} \sim \frac{r_n}{n}\rho_n = o\left(\frac{r_n^{3/2}}{n}\right).$$

Therefore if $r_n = O(n^{2/3})$, the left member of (6.20) tends to zero no matter how $\rho_n$ is chosen, and thus by Theorem 4.7, $\hat{a}_n/a_n \not\to 1$. But if $r_n/n^{2/3} \to \infty$, it may be possible to choose $\rho_n$ so that $r_n \rho_n / n \to \infty$ and hence $P(|\hat{a}_n / a_n 1| \geq \rho_n / \sqrt{r_n}) \to 0$ (and thus $\hat{a}_n / a_n \to 1$ in probability). However, under no circumstances can $\{\sqrt{r_n}(\hat{a}_n / a_n - 1)\}$ be tight, because in (6.20), it is always possible to find $\rho_n \to \infty$ such that $r_n \rho_n / n \to 0$, since $r_n / n \to 0$. Hence tightness fails due to Theorem 4.7.

**Acknowledgment.** The authors would like to thank the referees for their thorough reading of the manuscript of this article and for their many useful suggestions which have improved the paper.

## REFERENCES

CSÖRGŐ, S., HAEUSLER, E. and MASON, D. M. (1988a). The asymptotic distribution of trimmed sums. *Ann. Probab.* **16** 672–699.

CSÖRGŐ, S., HAEUSLER, E. and MASON, D. M. (1988b). A probabilistic approach to the asymptotic distribution of sums of independent, identically distributed random variables. *Adv. in Appl. Math.* **9** 259–333.

CSÖRGŐ, S., HORVÁTH, L. and MASON, D. M. (1986). What portion of the sample makes a partial sum asymptotically stable or normal? *Probab. Theory Related Fields* **72** 1–16.

DAVIS, R. and RESNICK, S. (1984). Tail estimates motivated by extreme value theory. *Ann. Statist.* **12** 1467–1487.

DOEBLIN, W. (1940). Sur l'ensemble des puissances d'une loi de probabilite. *Studia Math.* **9** 71–96.

FELLER, W. (1967). On regular variation and local limit theorems. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **2** 373–388. Univ. California Press.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

GNEDENKO, B. V. and KOLMOGOROV, A. N. (1954). *Limit Theorems for Sums of Independent Random Variables*. Addison-Wesley, Reading, Mass.

GRIFFIN, P. S. and PRUITT, W. E. (1987). The central limit problem for trimmed sums. *Math. Proc. Cambridge Philos. Soc.* **102** 329–349.

GRIFFIN, P. S. and PRUITT, W. E. (1989). Asymptotic normality and subsequential limits of trimmed sums. *Ann. Probab.* **17** 1186–1219.

HAHN, M. G. and KUELBS, J. (1988). Universal asymptotic normality for conditionally trimmed sums. *Statist. Probab. Lett.* **7** 9–15.

HAHN, M. G., KUELBS, J. and WEINER, D. C. (1989a). The asymptotic distribution of magnitude-winsorized sums via self-normalization. *J. Theoret. Probab.* To appear.

HAHN, M. G., KUELBS, J. and WEINER, D. C. (1989b). On estimating an exponent of regular variation for tail distributions. In *Sums, Trimmed Sums and Extremes* (M. G. Hahn, D. M. Mason and D. C. Weiner, eds). Birkhäuser, Boston. To appear.

HAHN, M. G., KUELBS, J. and WEINER, D. C. (1989c). A more robust approach to weak convergence for partial sums via trimming and self-normalization. In *Sums, Trimmed Sums and Extremes* (M. G. Hahn, D. M. Mason and D. C. Weiner, eds.). Birkhäuser, Boston. To appear.

KESTELMAN, H. (1937). *Modern Theories of Integration*. Oxford Univ. Press, London.

KUELBS, J. and LEDOUX, M. (1987). Extreme values for vector valued random variables and a Gaussian central limit theorem. *Probab. Theory Related Fields* **74** 341–355.

LÉVY, P. (1937). *Théorie de l'Addition des Variables Aléatoires*. Gauthier-Villars, Paris.

LOGAN, B., MALLOWS, C., RICE, S. and SHEPP, L. (1973). Limit distributions of self-normalized sums. *Ann. Probab.* **1** 788–809.

MALLER, R. (1982). Asymptotic normality of lightly trimmed means—a converse. *Math. Proc. Cambridge Philos. Soc.* **92** 535–545.

MALLER, R. (1988). Asymptotic normality of trimmed means in higher dimensions. *Ann. Probab.* **16** 1608–1622.

MORI, T. (1984). On the limit distributions of lightly trimmed sums. *Math. Proc. Cambridge Philos. Soc.* **96** 507–516.

PRUITT, W. (1983). The class of limit laws for stochastically compact normed sums. *Ann. Probab.* **11** 962–969.

PRUITT, W. (1985). Sums of independent random variables with the extreme terms excluded. Unpublished.

SENETA, E. (1976). *Regularly Varying Functions. Lecture Notes in Math.* **508**. Springer, Berlin.

STIGLER, S. M. (1973). The asymptotic distribution of the trimmed mean. *Ann. Statist.* **1** 472–477.

TAYLOR, A. (1985). *General Theory of Functions and Integration*. Dover, New York.

WEINER, D. C. (1986). Center, scale and asymptotic normality for censored sums of independent random variables. In *Sums, Trimmed Sums and Extremes* (M. G. Hahn, D. M. Mason and D. C. Weiner, eds.). Birkhäuser, Boston. To appear.

MARJORIE G. HAHN
DEPARTMENT OF MATHEMATICS
TUFTS UNIVERSITY
MEDFORD, MASSACHUSETTS 02155

JIM KUELBS
DEPARTMENT OF MATHEMATICS
213 VAN VLECK HALL
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706

DANIEL C. WEINER
DEPARTMENT OF MATHEMATICS
BOSTON UNIVERSITY
BOSTON, MASSACHUSETTS 02215