The Annals of Probability
1988, Vol. 16, No. 3, 1265–1280

ON THE EFFECT OF RANDOM NORMING ON THE RATE OF CONVERGENCE IN THE CENTRAL LIMIT THEOREM

BY PETER HALL

Australian National University

It is shown that "studentizing," i.e., normalizing by the sample standard deviation rather than the population standard deviation, can improve the rate of convergence in the central limit theorem. This provides concise confirmation of one feature of the folklore that a studentized sum is in some sense more robust than a normed sum. The case of infinite population standard deviation is also examined.

1. Introduction. Our purpose in this paper is to show that "studentizing" —the operation of replacing a population standard deviation in a norming constant by its sample estimate—can improve the rate of convergence in the central limit theorem. Thus, we provide concise confirmation of one feature of the folklore that a studentized sum is in some sense more robust than a normed sum.

The reason is that large deviations of summands, which are responsible for slow rates of convergence when ordinary norming constants are used, "cancel" from the sample mean and sample standard deviation. Remark 3.7 in Section 3 will give an intuitive explanation of the mechanism behind this cancellation. Note that the large deviations have a particularly subtle influence, since under the condition that the distribution is attracted to the normal law, the largest summand is negligible in comparison with the sum itself.

Section 2 will briefly survey results on rates of convergence in the "ordinary" (i.e., nonstudentized) central limit theorem and will discuss their connection with our work. Section 3 will state our main results and summarize their consequences, and Section 4 will give proofs.

2. Background. At this point it is necessary to introduce a little notation. Let $X, X_1, X_2, \dots$ be independent and identically distributed nondegenerate random variables, and assume for the time being that $\sigma^2 \equiv \mathrm{var}(X) < \infty$. Put $\mu \equiv E(X)$,

$$\bar{X} \equiv n^{-1} \sum_{i=1}^{n} X_i, \qquad \hat{\sigma}^2 \equiv n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2,$$

$$S \equiv n^{1/2}\sigma^{-1}(\bar{X} - \mu) \quad \text{and} \quad T \equiv n^{1/2}\hat{\sigma}^{-1}(\bar{X} - \mu).$$

Of course if $E(|X|^3) < \infty$, then S converges to the standard normal distribution at the rate $O(n^{-1/2})$, as expressed by the Berry–Esseen theorem [e.g., Petrov

Received December 1986; revised November 1987.
AMS 1980 *subject classifications*. Primary 60F05; secondary 60G50.
Key words and phrases. Central limit theorem, rate of convergence, sample variance, studentizing.

(1975), Theorem 4, page 111]. We are interested in the case where $E(|X|^3) = \infty$; there rates of convergence are often described in terms of "characterizations." We pause here to introduce them.

Let $\mathscr{F}$ be the class of nonincreasing functions f: $[0, \infty) \to (0, \infty)$ such that for some $\varepsilon > 0$, $x^{1-\varepsilon}f(x^2)$ is eventually nondecreasing. Let $\mathscr{G}$ be the class of measurable functions g: $[0, \infty) \to (0, \infty)$ such that for some $\varepsilon > 0$, $x^{\varepsilon-1}g(x)$ is eventually nonincreasing. Examples include $f(x) \equiv x^{-\alpha/2}$ and $g(x) \equiv x^{\alpha}$ for $0 \le \alpha < 1$. Given $g \in \mathscr{G}$, let

$$G(x) \equiv \int_1^x u^{-1}g(u)\,du.$$

Put

$$\Delta_n(S) \equiv \sup_{-\infty<x<\infty} |P(S \le x) - \Phi(x)|,$$

$$(2.1)\qquad \Delta_n^* \equiv \inf_{c,\,d}\left\{\sup_{-\infty<x<\infty} |P(\bar{X} \le cx + d) - \Phi(x)|\right\},$$

where Φ denotes the standard normal distribution function. We next give two "order-of-magnitude" characterizations and one "summation" characterization of the rate of convergence in the central limit theorem.

(i) The following two conditions are equivalent:

$$\Delta_n(S) = O\{f(n)\}, \qquad E\{X^2 I(|X| > x)\} = O\{f(x^2)\}.$$

(ii) The following two conditions are equivalent:

$$\Delta_n^* = O\{f(n)\}, \qquad x^2P(|X| > x) = O\{f(x^2)\}.$$

(iii) The following three conditions are equivalent:

$$\sum n^{-1}g(n^{1/2})\Delta_n(S) < \infty, \qquad \sum n^{-1}G(n^{1/2})\Delta_n^* < \infty, \qquad E\{X^2G(|X|)\} < \infty.$$

Further results of this type and references to earlier work in the same vein may be found in Rozovskii (1978) and Hall (1980). There it is shown that (ii) and (iii) remain true if Δ_n^* is replaced by

$$\Delta_n^\dagger \equiv \sup_{-\infty<x<\infty} |P\{n^{1/2}\sigma_n^{-1}(\bar{X} - \mu) \le x\} - \Phi(x)|,$$

where $\sigma_n^2 \equiv E\{(X - \mu)^2 I(|X - \mu| \le n^{1/2})\}$ is the truncated population variance.

To appreciate the implications of these characterizations, consider the case where $f(x) \equiv 1/\log x$, $g(x) \equiv 1$ and

$$P(|X| > x) \sim \text{const}\, x^{-2}(\log x)^{-\alpha}$$

as $x \to \infty$. Then the ordinary rate of convergence in the central limit theorem is that of $\Delta_n(S)$ and equals $O(1/\log n)$ if and only if $\alpha \ge 2$; but Δ_n^* and $\Delta_n^\dagger$ are $O(1/\log n)$ under the weaker condition $\alpha > 1$. Furthermore, $\sum n^{-1}\Delta_n(S) < \infty$ if and only if $\alpha > 2$, whereas $\sum n^{-1}\Delta_n^* < \infty$ and $\sum n^{-1}\Delta_n^\dagger < \infty$ under the weaker condition $\alpha > 1$. When $1 < \alpha < 2$, and indeed for other values of α, the rate of convergence in the central limit theorem can be improved by norming with the truncated standard deviation σ_n rather than the ordinary standard deviation σ.

Now we examine what happens if we replace S by its studentized form T. Put

$$\Delta_n(T) \equiv \sup_{-\infty < x < \infty} |P(T \le x) - \Phi(x)|. \tag{2.2}$$

The following result is a corollary of Theorem 3.1 from the next section.

PROPOSITION 2.1. *If $f \in \mathscr{F}$ and $g \in \mathscr{G}$ and if the distribution of X satisfies condition $(*)$ in Section 3, then*

$$x^2 P(|X| > x) = O\{f(x^2)\} \quad \textit{implies} \quad \Delta_n(T) = O\{f(n)\} \tag{2.3}$$

and

$$E\{X^2 G(|X|)\} < \infty \quad \textit{implies} \quad \sum n^{-1} G(n^{1/2}) \Delta_n(T) < \infty. \tag{2.4}$$

That is, conditions which were necessary and sufficient for characterizing the *fastest* rate of convergence in the central limit theorem for S, using special norming constants, are sufficient to characterize the rate of convergence in the central limit theorem for the studentized mean T, using the sample standard deviation as the norming "constant." The preceding results show that the studentized mean converges to normality "at the fastest rate" to the extent to which that rate can be characterized by the usual order-of-magnitude or summation conditions. Interestingly, the optimal rate is actually *achieved* under conditions of symmetry (see Remark 3.6 in Section 3) but not in all other cases (see Remark 3.5).

3. Main results. We continue to use notation introduced in Section 2, except that we slightly extend the definition of T by allowing a general location constant, as follows. Let $\{m_n\}$ be any sequence of real numbers converging to the mean μ of X and redefine $T \equiv n^{1/2}(\bar{X} - m_n)/\hat{\sigma}$. Put

$$\delta_n \equiv nP(|X| > n^{1/2}) + n^{-1/2} E\{|X|^3 I(|X| \le n^{1/2})\},$$

$$\delta_{n1} \equiv nP(|X| > n^{1/2}) + |n^{-1/2} E\{X^3 I(|X| \le n^{1/2})\}| + n^{-1} E\{X^4 I(|X| \le n^{1/2})\},$$

$$\mu_n \equiv E\{X I(|X| \le n^{1/2})\}$$

and

$$p(y) \equiv P(X - y > 0 \mid |X - y|) \quad \text{(a random variable)}.$$

We assume that

$$(*) \qquad \text{for some } y \in (-\infty, \infty), \qquad E[p(y)\{1 - p(y)\}] > 0.$$

Condition $(*)$ holds for any distribution which has a nondegenerate absolutely continuous component [use the argument leading to Lemma 2.2 of Hall (1987)] or which has two or more atoms or which is symmetric about some point.

Our first theorem provides an upper bound to the rate of convergence of $\Delta_n(T)$, the latter defined in (2.2).

THEOREM 3.1. *Assume $E(X^2) < \infty$ and condition $(*)$ holds. Then*

$$\Delta_n(T) = O(\delta_n + n^{1/2}|m_n - \mu_n|) \tag{3.1}$$

as $n \to \infty$.

The following remarks discuss implications of Theorem 3.1 and generalizations.

REMARK 3.1. A slight modification of the proof of (3.1) allows us to show that if $n^{1/2}|m_n - \mu_n| \to 0$, then

$$n^{1/2}|m_n - \mu_n| = O\{\Delta_n(T) + \delta_n\}. \tag{3.2}$$

REMARK 3.2. If we take $m_n \equiv \mu = E(X)$, then

$$|m_n - \mu_n| = |E\{XI(|X| > n^{1/2})\}| \le E\{|X|I(|X| > n^{1/2})\},$$

in which case it follows from Theorem 3.1 that

$$\Delta_n(T) = O[n^{1/2}E\{|X|I(|X| > n^{1/2})\} + n^{-1/2}E\{|X|^3 I(|X| \le n^{1/2})\}]. \tag{3.3}$$

REMARK 3.3. Proposition 2.1 is a simple consequence of (3.3). To obtain result (2.3), integrate by parts in the formulas for both terms on the right-hand side of (3.3), thereby proving that those terms are $O\{f(n)\}$ if $x^2 P(|X| > x) = O\{f(x^2)\}$. To obtain (2.4), replace $\Delta_n(T)$ in the summation condition by the right-hand side of (3.3) and use an integral approximation to the sum.

REMARK 3.4. Suppose X is positive with probability 1. Again take $m_n \equiv \mu$. Then by (3.3),

$$\Delta_n(T) = O[n^{1/2}E\{XI(X > n^{1/2})\} + n^{-1/2}E\{X^3 I(X \le n^{1/2})\}] \tag{3.4}$$

and by (3.2),

$$\begin{aligned} &n^{1/2}E\{XI(X > n^{1/2})\} \\ &\quad = O[\Delta_n(T) + nP(X > n^{1/2}) + n^{-1/2}E\{X^3 I(X \le n^{1/2})\}]. \end{aligned} \tag{3.5}$$

The example discussed in Hall [(1982), page 138] shows that it is possible to have, along a subsequence $n = n(k) \to \infty$, both $nP(X > n^{1/2})$ and $n^{-1/2}E\{X^3 I(X \le n^{1/2})\}$ of smaller order than $n^{1/2}E\{XI(X > n^{1/2})\}$. It then follows from (3.4) and (3.5) that $\Delta_n(T)$ is of precise order $n^{1/2}E\{XI(X > n^{1/2})\}$ along that subsequence.

REMARK 3.5. The fastest rate of convergence in the central limit theorem for the sample mean is precisely δ_{n1} up to terms of order $n^{-1/2}$. That is to say, $(\Delta_n^* + n^{-1/2})/(\delta_{n1} + n^{-1/2})$ is bounded away from zero and infinity as $n \to \infty$. [See, for example, Hall (1982), Theorems 3.1 and 3.2, pages 87–90.] In the case of the example considered in Remark 3.4, this entails

$$\Delta_n^* = o\{\Delta_n(T)\}$$

as $n \to \infty$ through the sequence $\{n(k)\}$. It is therefore not necessarily true that $\Delta_n(T)$ converges to zero as rapidly as Δ_n^*.

Pathologies such as the one discussed in Remark 3.5 vanish if the underlying distribution is symmetric. Indeed, in the context of symmetry we may generalize our work to the case of infinite variance, as follows. Assume X is in the domain of attraction of the normal distribution, meaning that the function $H(x) \equiv E\{X^2 I(|X| \le x)\}$ is slowly varying at infinity [see, e.g., Ibragimov and Linnik (1971), page 83]. Let a_n be the largest solution of the equation $na_n^{-2}H(a_n) = 1$; then a_n is uniquely defined for sufficiently large n. Of course, $a_n \sim n^{1/2}\sigma$ if $E(X^2) = \sigma^2 < \infty$. Define Δ_n^* as in (2.1). If X is symmetric, then Δ_n^* is of precise order

$$\delta_{n2} \equiv nP(|X| > a_n) + na_n^{-4}E\{X^4 I(|X| \le a_n)\},$$

up to terms of order $n^{-1/2}$. That is to say, $(\Delta_n^* + n^{-1/2})/(\delta_{n2} + n^{-1/2})$ is bounded away from zero and infinity as $n \to \infty$. [See, for example, Hall (1982), Theorems 4.12 and 4.13, pages 206–207.] Our next result shows that in the case of symmetry, $\Delta_n(T)$ converges to zero at least as quickly as Δ_n^*.

THEOREM 3.2. *Assume X is symmetric and in the domain of attraction of the normal distribution. Take $m_n \equiv \mu = E(X)$. Then $\Delta_n(T) = O(\delta_{n2} + n^{-1/2})$ as $n \to \infty$.*

REMARK 3.6. In view of the comments made just prior to Theorem 3.2, we have

$$\limsup_{n \to \infty} \Delta_n(T)/\left(\Delta_n^* + n^{-1/2}\right) < \infty$$

under the conditions of Theorem 3.2. Remark 3.5 shows that this result can fail in cases of asymmetry.

REMARK 3.7. When the summands are symmetric, we may give an intuitive explanation of why studentizing enhances the rate of convergence. Notice that since $\mu = 0$, the event $n^{1/2}(\bar{X} - \mu)/\hat{\sigma} \le x$ is identical to $(\Sigma X_i)(\Sigma X_i^2)^{-1/2} \le x(1 + n^{-1}x^2)^{-1/2}$. We argue that the distribution of $(\Sigma X_i)(\Sigma X_i^2)^{-1/2}$ should be particularly close to the standard normal distribution, as follows. Condition on the absolute values $|X_1|, \dots, |X_n|$. Then ΣX_i has conditional mean zero and conditional variance $\hat{\sigma}^2 \equiv \Sigma X_i^2$. The *truncated* conditional variance is

$$\sum_{i=1}^n E\{X_i^2 I(|X_i| \le \hat{\sigma}) \mid |X_1|, \dots, |X_n|\} = \sum_{i=1}^n X_i^2 I\left(X_i^2 \le \hat{\sigma}^2\right) = \sum_{i=1}^n X_i^2 = \hat{\sigma}^2.$$

That is, the truncated conditional variance is identical to the ordinary conditional variance. Rates of convergence in the central limit theorem for general summands are improved by norming with the truncated variance rather than the untruncated variance; see Theorems 3.1 and 3.2 of Hall [(1982), pages 87–90]. On the present occasion these two variances are identical, at least as far as conditional distributions go, and so it stands to reason that $(\Sigma X_i)\hat{\sigma}^{-1} = (\Sigma X_i)(\Sigma X_i^2)^{-1/2}$ should exhibit fast rates of convergence.

REMARK 3.8. A conditional argument along lines of that given earlier is at the heart of our proofs. That technique has been used before by Hall (1987). It does not appear to extend to the asymmetric infinite variance case, largely because $p(y)\{1 - p(y)\}(X - y)^2$ can have finite mean even though $(X - y)^2$ has infinite mean.

Edgeworth expansions for distributions of general statistics, including the studentized mean, have been given by Chibisov (1972, 1973, 1980, 1981) and Bhattacharya and Ghosh (1978, 1986). See also Slavova (1986).

4. Proofs. The symbols $C, C_1, C_2, \ldots$ will denote positive generic constants and ϕ will be the standard normal density.

PROOF OF THEOREM 3.1. It suffices to prove a version of Theorem 3.1 in which the supremum is taken only over $|x| \le \log n$. To see why, observe that if $|X_i - \mu| \le n^{1/2}$ for $1 \le i \le n$, then

$$n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sigma_n^2 + n^{-1} \sum_{i=1}^{n} \left\{ (X_i - \mu)^2 I\left(|X_i - \mu| \le n^{1/2}\right) - \sigma_n^2 \right\} - \left(\bar{X} - \mu\right)^2,$$

where $\sigma_n^2 \equiv E\{(X - \mu)^2 I(|X - \mu| \le n^{1/2})\}$. Therefore for a positive constant ε and for sufficiently large n,

$$\begin{aligned} P(|T| > \log n) &\le nP(|X - \mu| > n^{1/2}) + P\left(n^{1/2}\sigma_n^{-1}|\bar{X} - m_n| > \varepsilon \log n\right) \\ &\quad + P\left[n^{-1} \left| \sum_{i=1}^{n} \left\{ (X_i - \mu)^2 I\left(|X_i - \mu| \le n^{1/2}\right) - \sigma_n^2 \right\} \right| > \varepsilon \right]. \end{aligned} \tag{4.1}$$

The second term on the right-hand side is dominated by

$$\begin{aligned} &2 \sup_{-\infty < x < \infty} \left| P\left\{ n^{1/2}\sigma_n^{-1}\left(\bar{X} - m_n\right) \le x \right\} - \Phi(x) \right| + 2\{1 - \Phi(\varepsilon \log n)\} \\ &\quad \le 2 \sup_{-\infty < x < \infty} \left| P\left\{ n^{1/2}\sigma_n^{-1}\left(\bar{X} - \mu_n\right) \le x \right\} - \Phi(x) \right| \\ &\qquad + 2 \sup_{-\infty < x < \infty} \left| \Phi\left\{ x + n^{1/2}\sigma_n^{-1}(\mu_n - m_n) \right\} - \Phi(x) \right| + 2\{1 - \Phi(\varepsilon \log n)\} \\ &\quad \le C\left(\delta_n + n^{1/2}|\mu_n - m_n|\right), \end{aligned}$$

using Theorem 3.9 of Hall [(1982), page 132] and noting that $1 - \Phi(\varepsilon \log n) = O(n^{-r})$ for all $\varepsilon > 0$ and all $r > 0$. The third term on the right-hand side of (4.1) is no greater than

$$(\varepsilon n)^{-2}E\left[\left|\sum_i\left\{(X_i-\mu)^2I(|X_i-\mu|\le n^{1/2})-\sigma_n^2\right\}\right|^2\right]$$
$$\le(\varepsilon n)^{-2}nE\left\{(X-\mu)^4I(|X-\mu|\le n^{1/2})\right\}\le C\delta_n.$$

Therefore $P(|T|>\log n)=O(\delta_n+n^{1/2}|\mu_n-m_n|)$, whence

$$\sup_{|x|>\log n}|P(T\le x)-\Phi(x)|\le P(|T|>\log n)+1-\Phi(\log n)$$
$$=O\left(\delta_n+n^{1/2}|\mu_n-m_n|\right).$$

In establishing this bound when the supremum is taken over $|x|\le\log n$, we may assume without loss of generality that condition (*) holds with $y=0$, although we are now not permitted to assume that X has zero mean. Let $\mathscr{F}$ be the σ-field generated by $|X_1|,\dots,|X_n|$ and put $p_i\equiv P(X_i>0\,|\,|X_i|)$,

$$Y_i\equiv X_i-E(X_i\,|\,|X_i|),$$
$$s^2\equiv\sum_{i=1}^n E\left(Y_i^2\,|\,|X_i|\right)=\sum_{i=1}^n 4p_i(1-p_i)X_i^2,\qquad V\equiv\sum_{i=1}^n Y_i,$$
$$\hat\delta_n\equiv s^{-2}\sum_{i=1}^n E\left\{Y_i^2I(|Y_i|>s)|\mathscr{F}\right\}+s^{-3}\sum_{i=1}^n E\left\{|Y_i|^3I(|Y_i|\le s)|\mathscr{F}\right\}.$$

Conditional on $\mathscr{F}$, the variables $Y_1,\dots,Y_n$ are independent with zero means and variances adding to s^2. It now follows from Theorems 2.2 and 2.3 of Hall [(1982), pages 25 and 44] that if $s^2>0$ and if the $\mathscr{F}$-measurable random variable $\hat\varepsilon_n$ satisfies

$$s^{-2}\sum_{i=1}^n E\left\{Y_i^2I(|Y_i|>\hat\varepsilon_n s)|\mathscr{F}\right\}\le\tfrac{1}{8},\tag{4.2}$$

then

$$\sup_{-\infty<x<\infty}\left|P(s^{-1}V\le x|\mathscr{F})-\Phi(x)\right|\le C\left[\hat\delta_n+\hat\varepsilon_n+s^{-4}\sum_{i=1}^n\left\{E\left(Y_i^2\,|\,|X_i|\right)\right\}^2\right],$$

where C is a universal constant. The condition $T\le x$ is equivalent to $s^{-1}V\le U_x$ for an $\mathscr{F}$-measurable random variable U_x whose value in most circumstances is defined in (4.8). Therefore if $\mathscr{E}_1$ is any event in $\mathscr{F}$ such that (4.2) holds and $s^2>0$ on $\mathscr{E}_1$, then

$$\sup_{|x|\le\log n}|P(T\le x)-\Phi(x)|$$
$$\le CE\left\{\left(\hat\delta_n+\hat\varepsilon_n+s^{-4}\sum_{i=1}^n X_i^4\right)I(\mathscr{E}_1)\right\}+2P(\tilde{\mathscr{E}}_1)+D,\tag{4.3}$$

where

$$D \equiv \sup_{|x| \le \log n} |E\{\Phi(U_x)\} - \Phi(x)|.$$

Given constants $c_2 > c_1 > 0$, define $\mathscr{E}_2 \equiv \{c_1 n \le s^2 \le c_2 n,\ c_1 n \le \Sigma X_i^2 \le c_2 n\}$. Lemmas 4.1–4.3 allow us to prove that if c_1 is sufficiently small and c_2 sufficiently large, then

$$P(\tilde{\mathscr{E}}_2) = O(\delta_n) \tag{4.4}$$

as $n \to \infty$. By way of notation, let $Z, Z_1, Z_2, \dots$ denote independent and identically distributed random variables. Given a random variable U, define

$$\delta_n(U) \equiv nP(|U| > n^{1/2}) + n^{-1/2}E\{|U|^3 I(|U| \le n^{1/2})\}.$$

LEMMA 4.1. *If Z is nondegenerate and $E(Z^2) < \infty$, then for sufficiently small $\varepsilon > 0$,*

$$P\left\{\sum_{i=1}^n Z_i^2 \le n(EZ)^2 + \varepsilon n\right\} = o(n^{-1})$$

as $n \to \infty$.

PROOF. Put $\zeta \equiv E(Z)$ and $W_i \equiv z_i - \zeta$. Let $\xi > 0$. Then

$$P\left(\sum_{i=1}^n Z_i^2 \le n\zeta^2 + \varepsilon n\right) = P\left(\sum_{i=1}^n W_i^2 + 2\zeta \sum_{i=1}^n W_\iota \le \varepsilon n\right)$$
$$\le P\left\{\sum_{i=1}^n W_i^2 \le (\varepsilon + \xi)n\right\} + P\left(2\zeta \sum_{i=1}^n W_i \le -\xi n\right).$$

Now, $P(|\Sigma W_i| > \lambda n) = o(n^{-1})$ for any $\lambda > 0$; see Theorem 28 of Petrov [(1975), page 286]. Choose $\eta > 0$ so small that $\pi \equiv P(W_1^2 > \eta) > 2\eta$, put $\varepsilon = \xi \equiv \eta^2/2$ and let the random variable B have the binomial (n, π) distribution. Then

$$P\left\{\sum_{i=1}^n W_i^2 \le (\varepsilon + \xi)n\right\} \le P\left\{\sum_{i=1}^n \eta I(W_i^2 > \eta) \le \eta^2 n\right\}$$
$$\le P(B - n\pi \le -\eta n) = O(n^{-r})$$

for all $r > 0$. The lemma is proved. □

LEMMA 4.2.

$$P\left[\left|\sum_{i=1}^n \{Z_i - EZI(|Z| \le n)\}\right| > n\right] \le nP(|Z| > n) + n^{-1}E\{Z^2 I(|Z| \le n)\}.$$

PROOF. The left-hand side is dominated by

$$nP(|Z| > n) + n^{-2}E\left[\sum_{i=1}^n \{Z_i I(|Z_i| \le n) - EZI(|Z| \le n)\}\right]^2$$
$$\le nP(|Z| > n) + n^{-1}E\{Z^2 I(|Z| \le n)\}.$$ □

LEMMA 4.3. *If for a positive constant* c, $|U| \le c|V|$ *with probability* 1, *then* $\delta_n(U) \le C\delta_n(V)$, *where* C *depends only on* c.

PROOF. Without loss of generality, $c > 1$. Then

$$\begin{aligned}\delta_n(U) &\le c^3\delta_n(V) + nP\big(c^{-1}n^{1/2} < |V| \le n^{1/2}\big)\\ &\quad + n^{-1/2}E\big\{|U|^3 I\big(|U| \le n^{1/2};\ |V| > n^{1/2}\big)\big\}\\ &\le 2c^3\delta_n(V),\end{aligned}$$

as required. □

To derive (4.4), apply Lemma 4.1 first with $Z \equiv X_1$ and then with $Z \equiv 2p_1^{1/2}(1-p_1)^{1/2}X_1$, apply Lemma 4.2 first with $Z \equiv X_1^2$ and then with $Z \equiv 4p_1(1-p_1)X_1^2$, and apply Lemma 4.3 with $U \equiv 2p_1^{1/2}(1-p_1)^{1/2}X_1$ and $V \equiv X_1$.

On the event $\mathscr{E}_2$ we have

$$s^2 = \sum_{i=1}^n 4p_i(1-p_i)X_i^2 \ge c_1 n \ge c_1 c_2^{-1}\sum_{i=1}^n X_i^2,$$

$$\begin{aligned}s^{-2}\sum_{i=1}^n E\big\{Y_i^2 I(|Y_i| > s)|\mathscr{F}\big\} &\le s^{-2}\sum_{i=1}^n 4X_i^2 I\big(4X_i^2 > c_1 n\big)\\ &\le 4c_1^{-1}c_2\sum_{i=1}^n X_i^2\left(\sum_{j=1}^n X_j^2\right)^{-1} I\big(4X_i^2 > c_1 n\big)\\ &\le 4c_1^{-1}c_2\sum_{i=1}^n I\big(4X_i^2 > c_1 n\big),\end{aligned}$$

$$\begin{aligned}s^{-3}\sum_{i=1}^n E\big\{|Y_i|^3 I(|Y_i| \le s)|\mathscr{F}\big\} &\le s^{-3}\sum_{i=1}^n E\big\{|Y_i|^3 I\big(4X_i^2 \le c_1 n\big)|\mathscr{F}\big\}\\ &\quad + s^{-3}\sum_{i=1}^n E\big\{|Y_i|^3 I\big(|Y_i| \le s, 4X_i^2 > c_1 n\big)|\mathscr{F}\big\}\\ &\le (c_1 n)^{-3/2}\sum_{i=1}^n 8|X_i|^3 I\big(4X_i^2 \le c_1 n\big)\\ &\quad + \sum_{i=1}^n I\big(4X_i^2 > c_1 n\big).\end{aligned}$$

Therefore on $\mathscr{E}_2$,

$$\hat{\delta}_n \le C(c_1, c_2)\left\{\sum_{i=1}^n I\big(|X_i| > n^{1/2}\big) + n^{-3/2}\sum_{i=1}^n |X_i|^3 I\big(|X_i| \le n^{1/2}\big)\right\}. \tag{4.5}$$

Take $\hat{\varepsilon}_n \equiv \lambda n^{-1/2}$ for a large but fixed positive constant λ and let $\mathscr{E}_3$ denote the event that relation (4.2) holds. On $\mathscr{E}_2$,

$$s^{-2}\sum_{i=1}^n E\big\{Y_i^2 I(|Y_i| > \hat{\varepsilon}_n s)|\mathscr{F}\big\} \le (c_1 n)^{-1}\sum_{i=1}^n 4X_i^2 I\big(4X_i^2 > \lambda^2 c_1\big).$$

Choose λ so large that $4c_1^{-1}E\{X^2I(4X^2 > \lambda^2 c_1)\} \le 1/16$. Then

$$P(\mathscr{E}_2 \cap \tilde{\mathscr{E}}_3) \le nP(|X| > n^{1/2}) + P\left\{4(c_1 n)^{-1} \sum_{i=1}^{n} X_i^2 I\left(\tfrac{1}{2}\lambda c_1^{1/2} < |X_i| \le n^{1/2}\right) > \tfrac{1}{8}\right\}$$

$$\begin{aligned} (4.6) \qquad &\le nP(|X| > n^{1/2}) + P\left[4(c_1 n)^{-1}\left|\sum_{i=1}^{n} \left\{X_i^2 I\left(\tfrac{1}{2}\lambda c_1^{1/2} < |X_i| \le n^{1/2}\right)\right.\right.\right. \\ &\qquad \left.\left.\left. - EX^2 I\left(\tfrac{1}{2}\lambda c_1^{1/2} < |X| \le n^{1/2}\right)\right\}\right| > \tfrac{1}{16}\right] \\ &\le C\delta_n, \end{aligned}$$

using Chebyshev's inequality to bound the last-written probability.

Let the event $\mathscr{E}_1$ appearing in (4.3) be $\mathscr{E}_2 \cap \mathscr{E}_3$. From (4.3)–(4.6), we deduce that

$$(4.7) \qquad \sup_{|x| \le \log n} |P(T \le x) - \Phi(x)| \le C\delta_n + D.$$

It remains to bound D. We shall assume that $(2p_1 - 1)|X_1|$ has a nondegenerate distribution, for otherwise it is very easy to bound D. Let ξ be a small positive constant and put

$$\begin{aligned} \mu^{(2)} &\equiv E\left\{X_1^2 I\left(|X_1| \le \xi n^{1/2}\right)\right\}, \\ \nu^{(2)} &\equiv E\left\{4p_1(1-p_1)X_1^2 I\left(|X_1| \le \xi n^{1/2}\right)\right\}, \\ u_1 &\equiv -(\nu^{(2)}n)^{-1/2} m_n x^2 (1 + n^{-1}x^2)^{-1}, \\ u_2 &\equiv (\nu^{(2)})^{-1/2}(1 + n^{-1}x^2)^{-1} x\left\{\left(\mu^{(2)} - m_n^2\right) + n^{-1}\mu^{(2)}x^2\right\}^{1/2}, \\ U_1 &\equiv (1 + n^{-1}x^2)\left\{\left(\mu^{(2)} - m_n^2\right) + n^{-1}\mu^{(2)}x^2\right\}^{-1} n^{-1} \\ (4.8) \qquad &\quad \times \sum_{i=1}^{n} \left\{X_i^2 I\left(|X_i| \le \xi n^{1/2}\right) - \mu^{(2)}\right\}, \\ U_2 &\equiv -(\nu^{(2)}n)^{-1/2} \sum_{i=1}^{n} \left\{(2p_i - 1)|X_i| I\left(|X_i| \le \xi n^{1/2}\right) - m_n\right\}, \\ U_3 &\equiv (\nu^{(2)}n)^{-1} \sum_{i=1}^{n} \left\{4p_i(1-p_i)X_i^2 I\left(|X_i| \le \xi n^{1/2}\right) - \nu^{(2)}\right\}, \\ U_x &\equiv \left\{u_1 + u_2(1 + U_1)^{1/2} + U_2\right\}(1 + U_3)^{-1/2}, \\ \mathscr{E}_4 &\equiv \left\{\sum_{i=1}^{n} X_i^2 \le n(EX)^2 + n\varepsilon \text{ and } |X_i| \le \xi n^{1/2} \text{ for } 1 \le i \le n\right\}. \end{aligned}$$

Arguing as in the Appendix of Hall (1987) we may show that on $\mathscr{E}_4$, $T \le x$ if and only if $s^{-1}V \le U_x$. Of course, $P(\tilde{\mathscr{E}}_4) = O(\delta_n)$; see Lemma 4.1. Therefore as far as proving Theorem 3.1 goes, we may suppose that the $\mathscr{F}$-measurable variable U_x (such that $s^{-1}V \le U_x$ is equivalent to $T \le x$) is given by (4.8); that is, we may ignore $\tilde{\mathscr{E}}_4$.

For this definition of U_x, we shall prove that

$$D = O\big(\delta_n + n^{1/2}|\mu_n - m_n|\big) \tag{4.9}$$

as $n \to \infty$. Theorem 3.1 follows on combining this estimate with (4.7).

Put

$$\mathscr{E}_5 \equiv \left\{ n^{-1}\left| \sum_{i=1}^{n} \left\{ X_i^2 I\big(|X_i| \le \xi n^{1/2}\big) - \mu^{(2)} \right\} \right| \le \eta, \right.$$
$$\left. n^{-1}\left| \sum_{i=1}^{n} \left\{ 4p_i(1-p_i) X_i^2 I\big(|X_i| \le \xi n^{1/2}\big) - \nu^{(2)} \right\} \right| \le \eta \right\}$$

for a small positive constant η. Let $U_{x1} \equiv u_1 + u_2(1+U_1)^{1/2} + U_2$. Then

$$\Phi(U_x) = \Phi(U_{x1}) + \left\{(1+U_3)^{-1/2} - 1\right\} U_{x1}\phi(U_{x1})$$
$$+ \tfrac{1}{2}\left\{(1+U_3)^{-1/2} - 1\right\}^2 U_{x1}^2 \phi'\Big(U_{x1}\Big[1 + \hat{\theta}_1\big\{(1+U_3)^{-1/2} - 1\big\}\Big]\Big),$$

where $0 \le \hat{\theta}_1 \le 1$. Therefore,

$$\Phi(U_x) = \Phi(U_{x1}) - \tfrac{1}{2}U_3 U_{x1}\phi(U_{x1}) + R_1, \tag{4.10}$$

where if η is sufficiently small, $|R_1| \le CU_3^2$ on $\mathscr{E}_5$. Put $U_{x2} \equiv u_1 + u_2 + U_2$. Then

$$\Phi(U_{x1}) = \Phi(U_{x2}) + \left\{(1+U_1)^{1/2} - 1\right\} u_2 \phi(U_{x2})$$
$$+ \tfrac{1}{2}\left\{(1+U_1)^{1/2} - 1\right\}^2 u_2^2 \phi'\Big(u_1 + U_2 + u_2\Big[1 + \hat{\theta}_2\big\{(1+U_1)^{1/2} - 1\big\}\Big]\Big),$$

where $0 \le \hat{\theta}_2 \le 1$. Therefore,

$$\Phi(U_{x1}) = \Phi(U_{x2}) + \tfrac{1}{2}U_1 u_2 \phi(U_{x2}) + R_2, \tag{4.11}$$

where if η is sufficiently small and $\mathscr{E}_5$ holds, $|R_2| \le CU_1^2 V$ and

$$V \equiv \big(1 + u_2^2\big) \sup_{1/2 \le t \le 3/2} \exp\left\{-\tfrac{1}{4}(u_1 + tu_2 + U_2)^2\right\}.$$

Likewise,

$$U_{x1}\phi(U_{x1}) = U_{x2}\phi(U_{x2}) + R_3, \tag{4.12}$$

where $|R_3| \le C|U_1|V$ on $\mathscr{E}_5$. Combining (4.10)–(4.12), we conclude that

$$\Phi(U_x) = \Phi(U_{x2}) + \tfrac{1}{2}U_1 u_2 \phi(U_{x2}) - \tfrac{1}{2}U_3 U_{x2}\phi(U_{x2}) + R_4, \tag{4.13}$$

where $|R_4| \le C(U_1^2 + U_3^2)(1 + V)$ on $\mathscr{E}_5$.

A slight modification of the argument used to prove Lemma 4.2 shows that $P(\tilde{\mathscr{E}}_5) = O(\delta_n)$. Therefore the desired result (4.9) will follow from (4.13) if we show that

$$(4.14)\qquad \sup_{|x|\le \log n} |E\{\Phi(U_{x2})\} - \Phi(x)| = O(\delta_n + n^{1/2}|\mu_n - m_n|),$$

$$(4.15)\qquad \sup_{|x|\le \log n} |E\{U_1 u_2 \phi(U_{x2}) I(\mathscr{E}_5)\}| = O(\delta_n),$$

$$(4.16)\qquad \sup_{|x|\le \log n} |E\{U_3 U_{x2} \phi(U_{x2}) I(\mathscr{E}_5)\}| = O(\delta_n),$$

$$(4.17)\qquad \sup_{|x|\le \log n} E\{(U_1^2 + U_3^2)(1+V)\} = O(\delta_n).$$

First we derive (4.14). Let $\lambda^{(1)}$ and $\lambda^{(2)}$ denote the mean and variance, respectively, of $Z_i \equiv (2p_i - 1)|X_i| I(|X_i| \le \xi n^{1/2})$ and put

$$W \equiv (\lambda^{(2)} n)^{1/2} \Sigma (Z_i - \lambda^{(1)}).$$

The argument leading to Theorem 3.9 of Hall [(1982), page 132] may be employed to prove that

$$\sup_{-\infty < z < \infty} |P(W \le z) - \Phi(z)| \le C\delta_{n1}.$$

Therefore if N denotes a standard normal random variable,

$$\Big|E\Big[\Phi\big\{U_{x2} + (n/\nu^{(2)})^{1/2}(m_n - \lambda^{(1)})\big\}\Big] - E\Big[\Phi\big\{u_1 + u_2 + (\lambda^{(2)}/\nu^{(2)})^{1/2} N\big\}\Big]\Big|$$
$$= \Big|\int_{-\infty}^{\infty} \Phi\big\{u_1 + u_2 + (\lambda^{(2)}/\nu^{(2)})^{1/2} z\big\}\, d\{P(W \le z) - \Phi(z)\}\Big| \le C\delta_{n1}$$

uniformly in x, the inequality following on integrating by parts. Noting that $\nu^{(2)} + \lambda^{(2)} = \mu^{(2)} - (\lambda^{(1)})^2$ we see that

$$E\Big[\Phi\big\{u_1 + u_2 + (\lambda^{(2)}/\nu^{(2)})^{1/2} N\big\}\Big]$$
$$= \Phi\Big[(u_1 + u_2)\{1 + (\lambda^{(2)}/\nu^{(2)})\}^{-1/2}\Big]$$
$$= \Phi\Big[\big\{(\mu^{(2)} - m_n^2)(\nu^{(2)} + \lambda^{(2)})^{-1}\big\}^{1/2} x\Big] + O(n^{-1/2})$$
$$= \Phi(x) + O(n^{-1/2} + |m_n - \lambda^{(1)}|)$$

uniformly in $|x| \le \log n$. Since $\lambda^{(1)} = E\{XI(|X| \le \xi n^{1/2})\}$, then

$$n^{1/2}|\lambda^{(1)} - \mu_n| \le nP(|X| > \xi n^{1/2}) \le C\delta_n.$$

The desired result (4.14) follows from these estimates.

Next we prove (4.15). Put $u_3 \equiv (1 + n^{-1}x^2)\{(\mu^{(2)} - m_n^2) + n^{-1}\mu^{(2)}x^2\}^{-1}$ and $U'_{x2} \equiv u_1 + u_2 + W'$, where

$$(4.18)\qquad W' \equiv -(\nu^{(2)} n)^{-1/2} \sum_{i=1}^{n} \big\{(2p_i - 1)|X_i| I(|X_i| \le \xi n^{1/2}) - m_n\big\}.$$

Notice that

$$v \equiv E\{U_1 u_2 \phi(U_{x2}) I(\mathscr{E}_5)\} = u_2 u_3 E\left[\{X_1^2 I(|X_1| \le \xi n^{1/2}) - \mu^{(2)}\}\phi(U_{x2}) I(\mathscr{E}_5)\right]$$

and $\phi(U_{x2}) = \phi(U'_{x2}) + R$, where

$$|R| \le C_1 n^{-1/2}\{1 + |X_1| I(|X_1| \le \xi n^{1/2})\} \sup_{|t| \le 1} \exp\{-\tfrac{1}{4}(U'_{x2} + C_1 t)^2\}.$$

It follows that

$$v = u_2 u_3 E\left[\{X_1^2 I(|X_1| \le \xi n^{1/2}) - \mu^{(2)}\}\phi(U'_{x2}) I(\mathscr{E}_5)\right] + r, \tag{4.19}$$

where

$$\begin{aligned} |r| &\le C_2 n^{-1/2} |u_2 u_3| E\left[\{1 + |X_1|^3 I(|X_1| \le \xi n^{1/2})\} \sup_{|t| \le 1} \exp\{-\tfrac{1}{4}(U'_{x2} + C_1 t)^2\}\right] \\ &\le C_3 \delta_n |u_2 u_3| E\left[\sup_{|t| \le 1} \exp\{-\tfrac{1}{4}(U'_{x2} + C_1 t)^2\}\right]. \end{aligned}$$

An elementary argument shows that the last-written expectation is dominated by $C_4(1 + |u_2|)^{-1}$ uniformly in $|x| \le \log n$, and so $|r| \le C_5 \delta_n$ uniformly in $|x| \le \log n$.

Put

$$\begin{aligned} \mathscr{E}'_5 \equiv \Bigg\{ & n^{-1}\left|\sum_{i=2}^n \{X_i^2 I(|X_i| \le \xi n^{1/2}) - \mu^{(2)}\}\right| \le \tfrac{1}{2}\eta, \\ & n^{-1}\left|\sum_{i=2}^n \{4p_i(1-p_i) X_i^2 I(|X_i| \le \xi n^{1/2}) - \nu^{(2)}\}\right| \le \tfrac{1}{2}\eta \Bigg\}. \end{aligned}$$

If ξ is sufficiently small (relative to η) and n sufficiently large, then $\mathscr{E}'_5 \cap \{|X_1| \le \xi n^{1/2}\} \subseteq \mathscr{E}_5$, in which case

$$\begin{aligned} &\left|E\left[\{X_1^2 I(|X_1| \le \xi n^{1/2}) - \mu^{(2)}\}\phi(U'_{x2}) I(\mathscr{E}_5)\right]\right| \\ &\quad = \left|E\left[\{X_1^2 I(|X_1| \le \xi n^{1/2}) - \mu^{(2)}\}\phi(U'_{x2}) I(\tilde{\mathscr{E}}_5)\right]\right| \\ &\quad \le E\{X_1^2 I(|X_1| \le \xi n^{1/2}) \phi(U'_{x2}) I(\tilde{\mathscr{E}}'_5)\} + |\mu^{(2)}| E\{\phi(U'_{x2}) I(\tilde{\mathscr{E}}_5)\} \\ &\quad \le C\left[E\{\phi(U'_{x2}) I(\tilde{\mathscr{E}}'_5)\} + E\{\phi(U'_{x2}) I(\tilde{\mathscr{E}}_5)\}\right], \end{aligned} \tag{4.20}$$

bearing in mind that X_1 is independent of U'_{x2} and of $\mathscr{E}'_5$. To bound the last two expectations, note for example that

$$\begin{aligned} E\{\phi(U'_{x2}) I(\tilde{\mathscr{E}}_5)\} \le 4\eta^{-2}\Bigg(& E\left[\phi(U'_{x2})\left|n^{-1}\sum_{i=2}^n \{X_i^2 I(|X_i| \le \xi n^{1/2}) - \mu^{(2)}\}\right|^2\right] \\ & + E\left[\phi(U'_{x2})\left|n^{-1}\sum_{i=2}^n \{4p_i(1-p_i) X_i^2 I(|X_i| \le \xi n^{1/2}) - \nu^{(2)}\}\right|^2\right]\Bigg). \end{aligned}$$

Techniques which we shall use later on to derive (4.17) may now be employed to prove that the right-hand side of (4.20) is dominated by $C\delta_{n1}$. Combining the results from (4.19) on, we conclude that $|v| \le C\delta_{n1}$, which is (4.15). The proof of (4.16) is similar to but simpler than the proof of (4.15).

Finally we establish (4.17). Trivially, $E(U_1^2 + U_3^2) = O(\delta_n)$ uniformly in $|x| \le \log n$. We shall sketch a proof of the fact that $E(U_1^2 V) = O(\delta_n)$ uniformly in $|x| \le \log n$. Similarly it may be shown that $E(U_3^2 V) = O(\delta_n)$.

Put $\psi(z) \equiv \sup_{1/2 \le t \le 3/2} \exp\{-\frac{1}{4}(u_1 + tu_2 + z)^2\}$ and $X_i^* \equiv X_i I(|X_i| \le \xi n^{1/2})$ and notice that

$$\begin{aligned} E(U_1^2 V) &= u_3^2(1 + u_2^2) \\ &\quad \times \Big[n^{-1} E\big\{ (X_1^{*2} - \mu^{(2)})^2 \psi(U_2) \big\} \\ &\qquad + (1 - n^{-1}) E\big\{ (X_1^{*2} - \mu^{(2)})(X_2^{*2} - \mu^{(2)}) \psi(U_2) \big\} \Big]. \end{aligned} \tag{4.21}$$

Now, $\psi(U_2) = \psi(W') + R$, where W' is defined at (4.18),

$$|R| \le C_1 n^{-1/2}(1 + |X_1^*| + |X_2^*|)\chi(W')$$

and

$$\chi(z) \equiv \sup_{1/2 \le t \le 3/2,\ |s| \le 1} \exp\big\{-\tfrac{1}{8}(u_1 + tu_2 + sC_1 + z)^2\big\}.$$

Furthermore, $E\{\psi(W')\} + E\{\chi(W')\} \le C_2(1 + u_2^2)^{-1}$ uniformly in $|x| \le \log n$. The result $E(U_1^2 V) = O(\delta_n)$ now follows from (4.21) on noting that X_1^* and X_2^* are independent of one another and of W'. □

PROOF OF THEOREM 3.2. When working under the assumption of symmetry, we do not separate suprema over $|x| \le \log n$ and $|x| > \log n$. Of course in the presence of symmetry, $E(X_1 | |X_1|) = 0$ and condition (*) holds with $y = 0$. Therefore in notation introduced during the proof of Theorem 3.1, $Y_i = X_i$ and $s^2 = \sum X_i^2$. Replace $\hat{\delta}_n$ by

$$\hat{\delta}_{n1} \equiv s^{-2} \sum_{i=1}^{n} E\{Y_i^2 I(|Y_i| > s) | \mathscr{F}\} + s^{-4} \sum_{i=1}^{n} E\{Y_i^4 I(|Y_i| \le s) | \mathscr{F}\}$$

and replace condition (4.2) by

$$s^{-2} \sum_{i=1}^{n} X_i^2 I(|X_i| > \hat{\varepsilon}_n s) \le \tfrac{1}{8}. \tag{4.22}$$

The constraint $T \le x$ is equivalent to $s^{-1} V \le (1 + n^{-1}x^2)^{-1/2} x$. Let $\mathscr{E}_1'$ be any event in $\mathscr{F}$ such that (4.22) holds and $s^2 > 0$ on $\mathscr{E}_1'$. The argument which formerly led to (4.3) now yields

$$\begin{aligned} &\sup_{-\infty < x < \infty} |P(T \le x) - \Phi(x)| \\ &\quad \le CE\left\{ \left(\hat{\delta}_{n1} + \hat{\varepsilon}_n + s^{-4} \sum_{i=1}^{n} X_i^4 \right) I(\mathscr{E}_1') \right\} + P(\tilde{\mathscr{E}}_1') + D', \end{aligned} \tag{4.23}$$

where

$$D' \equiv \sup_{-\infty < x < \infty} \left| \Phi\left\{ (1 + n^{-1}x^2)^{-1/2} x \right\} - \Phi(x) \right| \le C n^{-1/2}.$$

Define $\mathscr{E}_2' \equiv \{c_1 a_n^2 \le s^2 \le c_2 a_n^2\}$ for positive constants c_1 and c_2. Observe that

$$\begin{aligned} P&\left\{ \left| \sum_{i=1}^{n} \left(X_i^2 - n^{-1} a_n^2 \right) \right| > \varepsilon a_n^2 \right\} \\ &\le nP(|X| > a_n) + P\left[\left| \sum_{i=1}^{n} \left\{ X_i^2 I(|X_i| \le a_n) - n^{-1} a_n^2 \right\} \right| > \varepsilon a_n^2 \right] \\ &\le nP(|X| > a_n) + n\left(\varepsilon a_n^2 \right)^{-2} E\left\{ X^4 I(|X| \le a_n) \right\} \le C\delta_{n1}. \end{aligned}$$

Therefore if $0 < c_1 < 1 < c_2 < \infty$, then $P(\tilde{\mathscr{E}}_2') = O(\delta_{n1})$. On $\mathscr{E}_2$,

$$\begin{aligned} \hat{\delta}_{n1} &\le \sum_{i=1}^{n} X_i^2 \left(\sum_{j=1}^{n} X_j^2 \right)^{-1} I\left(X_i^2 > c_1 a_n^2 \right) + \left(c_1 a_n^2 \right)^{-2} \sum_{i=1}^{n} X_i^4 I\left(X_i^2 \le c_2 a_n^2 \right) \\ &\le C\left\{ \sum_{i=1}^{n} I\left(X_i^2 > c_1 a_n^2 \right) + a_n^{-4} \sum_{i=1}^{n} X_i^4 I\left(X_i^2 \le c_2 a_n^2 \right) \right\} \\ &\le \left(c_1^{-2} + c_2^2 \right) C\left\{ \sum_{i=1}^{n} I(|X_i| > a_n) + a_n^{-4} \sum_{i=1}^{n} X_i^4 I(|X_i| \le a_n) \right\} \end{aligned}$$

and

$$s^{-4} \sum_{i=1}^{n} X_i^4 \le \sum_{i=1}^{n} I(|X_i| > a_n) + \left(c_1 a_n^2 \right)^{-2} \sum_{i=1}^{n} X_i^4 I(|X_i| \le a_n).$$

Put $\hat{\varepsilon}_n \equiv \lambda a_n^{-1}$ for a large positive constant λ and let $\mathscr{E}_3'$ be the event on which (4.22) holds. The argument which formerly lead to (4.6) may be adapted to show that if λ is sufficiently large, $P(\mathscr{E}_2' \cap \tilde{\mathscr{E}}_3') = O(\delta_{n1})$. Taking $\mathscr{E}_1' \equiv \mathscr{E}_2' \cap \mathscr{E}_3'$ in (4.23) and noting the estimates derived in the previous paragraph, we conclude that the right-hand side of (4.23) equals $O(\delta_{n1} + n^{-1/2} + a_n^{-1})$. This proves Theorem 3.2. □

REFERENCES

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.

BHATTACHARYA, R. N. and GHOSH, J. K. (1986). On moment conditions for valid formal Edgeworth expansions. Preprint.

CHIBISOV, D. M. (1972). An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. *Theory Probab. Appl.* **17** 620–630.

CHIBISOV, D. M. (1973). An asymptotic expansion for the distribution of sums of a special form with application to minimum-contrast estimates. *Theory Probab. Appl.* **18** 453–466.

CHIBISOV, D. M. (1980). An asymptotic expansion for the distribution of a statistic admitting a stochastic expansion. I. *Theory Probab. Appl.* **25** 732–744.

CHIBISOV, D. M. (1981). Asymptotic expansion for the distribution of a statistic admitting a stochastic expansion. II. *Theory Probab. Appl.* **26** 1–12.

HALL, P. (1980). Characterizing the rate of convergence in the central limit theorem. *Ann. Probab.* **8** 1037–1048.

HALL, P. (1982). *Rates of Convergence in the Central Limit Theorem.* Pitman, London.

HALL, P. (1987). Edgeworth expansion for Student's t statistic under minimal moment conditions. *Ann. Probab.* **15** 920–931.

IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

PETROV, V. V. (1975). *Sums of Independent Random Variables.* Springer, Berlin.

ROZOVSKII, L. V. (1978). On the precision of an estimate of the remainder term in the central limit theorem. *Theory Probab. Appl.* **23** 712–730.

SLAVOVA, V. V. (1986). Berry–Esseen bound for Student's statistic. Report MS-R8608, Dept. Math. Statist., Centrum voor Wiskunde en Informatica, Amsterdam.

DEPARTMENT OF STATISTICS
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA, ACT 2601
AUSTRALIA