The Annals of Probability
1973, Vol. 1, No. 3, 497–503

ON AN L_p VERSION OF THE BERRY–ESSEEN THEOREM FOR INDEPENDENT AND m-DEPENDENT VARIABLES

BY R. V. ERICKSON

Michigan State University

We show that the L_1 norm of the difference between the standard normal distribution and the distribution of the standardized sum of n independent random variables is less than 72 R_n, where R_n is a sum of standardized "inside" third and "outside" second moments. We conjecture that 72 can be replaced by 36 or even less. We also prove a similar result for m-dependent random variables, but no constant is specified.

1. Introduction. Recently some use has been made in statistics of an L_1 version of the Berry–Esséen theorem which is a trivial consequence of a result of Bikyalis [1] if absolute moments of order $2 + \alpha > 2$ are assumed finite.

We consider the case of independent variables having only finite second moments and show that both the L_1 and L_∞ version can be derived simultaneously by the usual characteristic function techniques (see Feller [5]), the only difference being the use of the appropriate smoothing lemma. Ibragimov [7] has a different simple proof for the independent identically distributed case.

We also extend the results of Egorov [2] in the m-dependent case to include the L_p norms, $1 \leqq p \leqq \infty$.

2. Notation and results. Throughout we consider random variables $X_1, X_2, \cdots$ with $EX_k = 0$, $EX_k^2 = \sigma_k^2 < \infty$, $k = 1, 2, \cdots$. We let

$$S_n = \sum_1^n X_k\,, \qquad B_n = ES_n^2\,, \qquad s_n^2 = \sum_1^n \sigma_k^2\,, \qquad \Delta_n(x) = F_n(x) - \mathscr{N}(x)\,,$$

where F_n is the distribution of $S_n/B_n^{\frac{1}{2}}$ and $\mathscr{N}$ is the standard normal distribution.

Denote the L_p norm of Δ_n by

$$\Delta_{np} = ||\Delta_n||_p\,.$$

When the random variables are independent we truncate as in Feller [5]: fix n and for $k = 1, \cdots, n$ fix $-\infty \leqq -\tau_k < 0 < \tau_k' \leqq \infty$, put $A_k = (-\tau_k, \tau_k')$, $X_k' = X_k I_{A_k}(X_k)$, $X_k'' = X_k - X_k'$. Write

$$\beta_k' = E(X_k')^2\,, \qquad \beta_k'' = E(X_k'')^2\,, \qquad \gamma_k' = E|X_k'|^3\,,$$
$$b_n'' = \sum_1^n \beta_k''\,, \qquad c_n' = \sum_1^n \gamma_k'\,,$$

and

$$B = b_n''/s_n^2\,, \qquad \Gamma = c_n'/s_n^3\,, \qquad R = B + \Gamma\,.$$

THEOREM 1. *If $X_1, X_2, \cdots$ are independent there is an absolute constant K_p*

Received July 15, 1971; September 27, 1972.

AMS 1970 *subject classifications.* Primary 60F05; Secondary 60F99.

Key words and phrases. L_p Berry–Esséen, m-dependent, asymptotic normality and error bounds.

such that

$$\Delta_{np} \leqq K_p R ,$$

where

$$K_p = (K_1)^{1/p}(6)^{1-1/p}$$

and K_1 is some constant less than 72.

REMARK $K_1 = 72$ is much too large, and by much more tedious calculations we can show that $K_1 \leqq 36$. Even this is probably way off the ultimate constant if the independent, identically distributed case, with finite third moments, can be used as a guide. In that case Zolotarev [10] shows that $\lim_{n\to\infty} n^{\frac{1}{2}}\Delta_{n1} \leqq (\frac{1}{2})E|X_1|^3/\sigma_1^3$. We will say more about the calculation of K_1 in the proof of the theorem.

For fixed n, taking $\tau_k = \tau_k' = s_n$, we have a

COROLLARY. *If $X_1, X_2, \cdots$ are independent and $0 < \delta \leqq 1$ then*

$$\Delta_{np} \leqq K_p \sum_1^n E|X_k|^{2+\delta}/s_n^{2+\delta} .$$

Recall that $X_1, X_2, \cdots$ are *m-dependent* if $(X_1, \cdots, X_r)$ and $(X_s, \cdots, X_n)$ are independent for all integers $1 \leqq r < s \leqq n$ with $s - r > m \geqq 0$.

The L_∞ version of the following theorem appears in Egorov [2].

THEOREM 2. *If $X_1, X_2, \cdots$ are m-dependent and if* (i) $B_n \to \infty$, (ii) $s_n^{-2} = O(B_n)$ *and* (iii) $\sum_1^n E|X_k|^{2+\delta} = O(B_n)$ *for some δ, $0 < \delta \leqq 1$, then there is an absolute constant C_p such that*

$$\Delta_{np} \leqq C_p B_n^{-\delta_p}$$

where

$$\delta_p = \delta/p(2 + 4\delta) + (1 - 1/p)\delta/(2 + 3\delta) .$$

Since $||\cdot||_p^p \leqq ||\cdot||_1||\cdot||_\infty^{p-1}$, we need prove only the L_1 estimates, the L_∞ estimates being known.

3. Proof of Theorem 1.

Fix n throughout this section.

We state the smoothing lemma only for the case at hand: let X_k have characteristic function χ_k, set $u_k(t) = \chi_k(t/s_n)$, $v_k(t) = \exp\{-\sigma_k^2 t^2/2s_n^2\}$ and $w_n = u_1 \cdots u_n - v_1 \cdots v_n = \sum_1^n (u_k - v_k)\Pi_k$, $\Pi_k = (\Pi_1^{k-1} u_i)(\Pi_{k+1}^n v_j)$.

LEMMA. *For any $T > 0$*

$$\pi\Delta_{n\infty} \leqq \textstyle\int_{-T}^{T} |w_n(t)/t|\, dt + (24/T)(2\pi)^{-\frac{1}{2}} ,$$
$$\Delta_{n1} \leqq 8\pi/T + (\tfrac{1}{2} + 4/T^2)^{\frac{1}{2}}\varepsilon + \delta_1 + \delta_2$$

where

$$\varepsilon^2 = \textstyle\int_{-T}^{T} |w_n(t)/t|^2\, dt ,$$
$$\delta_1^2 = \textstyle\int_{-T}^{T} |w_n(t)|t^2|^2\, dt ,$$
$$\delta_2^2 = \textstyle\int_{-T}^{T} |w_n'(t)/t|^2\, dt$$

and $w_n' = (d/dt)w_n$.

The L_∞ part of this lemma, due essentially to Berry, is proved in Feller [4] page 538. The L_1 part, due to Esséen, is proved in [8] page 25 and, save for the $8\pi/T$, is a simple consequence of the material of Chapter XIX. 7 of Feller [4]. The $8\pi/T$ term rests on a minimal extrapolation lemma of Esséen ([1] page 13); we update the references cited in the proof of this lemma and indicate its level of difficulty by noting that it is based on the fact that

$$\beta(z) = \sum_0^\infty (-1)^n/(z+n) = 1/z - \log 2 - \sum_1^\infty (-1)^n z/n(z+n)$$

is meromorphic with principal parts $p_n(z) = (-1)^n z$ at poles $z_n = -n \leqq 0$ and that $\pi G(z) = (\beta(z) - \frac{1}{2}z)\sin \pi z$ is therefore entire and $G(z) + G(-z) = 1$ (see Hille [6] pages 219, 221 and 264).

We use the L_1-part of this smoothing lemma in exactly the same way as Feller [5] uses the L_∞ part.

We have chosen to write the proof of Theorem 1 in a way that makes obvious what may be varied in hopes to improving the constant K_1. We then indicate choices of variables which give $K_1 = 36$, and $K_1 = 72$ and mention how these were made.

Using Feller's version of the L_∞ Berry–Esséen Theorem over $I = [-a, a]$ and Chebyshev's inequality and symmetry of η over I^c and integrating we have

$$\Delta_{n1} \leqq 12Ra + 3/2a\,.$$

This is minimal when $a^2 = 1/8R$ so without loss of generality we may suppose that $K_1 R \leqq 24R/(8R)^{\frac{1}{2}}$ and thus

$$R \leqq 72/K_1^2 = \rho^3\,. \tag{1}$$

By the moment inequality $(\beta_k')^3 \leqq (\gamma_k')^2$ and hence (see Feller [5] (17))

$$\begin{aligned}(\sigma_k/s_n)^4 &\leqq [(\gamma_k')^{\frac{2}{3}} + \beta_k'']^2 s_n^{-4}\\ &\leqq \Gamma^{\frac{1}{3}}\gamma_k'/s_n^3 + (2\Gamma^{\frac{2}{3}} + B)\beta_k''/s_n^2\,.\end{aligned}$$

Fix R, substiute $B = R - \Gamma$, and allow Γ to vary under the restriction $0 \leqq \Gamma \leqq R$. We see that the maximum of $2\Gamma^{\frac{2}{3}} + R - \Gamma$ is attained at $\Gamma = R$, $B = 0$ if $R \leqq (\frac{4}{3})^3$. Hence

$$\sum_1^n (\sigma_k/s_n)^4 \leqq \rho\Gamma + 2\rho^2 B \tag{2}$$

if we assume (1) and take $K_1 \geqq 6$ (which implies $R \leqq (\frac{4}{3})^3$).

Assuming (1) and (2) we obtain

$$\sum_1^n |u_k(t) - v_k(t)| \leqq \Gamma|t|^3/6 + B|t|^2 + |t|^4(\rho\Gamma + 2\rho^2 B)/8 = \varphi(t)\,, \tag{3}$$

$$\sum_1^n |u_k'(t) - v_k'(t)| \leqq \Gamma|t|^2/2 + 2B|t| + |t|^3(\rho\Gamma + 2\rho^2 B)/2 = |t|\psi(t)\,, \tag{4}$$

$$\sum_1^{k-1} |u_j'(t)| + \sum_{k+1}^n |v_j'(t)| \leqq \Gamma|t|^2/2 + 2B|t| + |t| = |t|(\eta(t) + 1)\,, \tag{5}$$

all t, $k = 1, \cdots, n$. The first of these is immediate from (3.3) of Feller [5] (his final $\sum$ is a misprint). Equations (4) and (5) are easy consequences of

$$\begin{aligned}|\chi_k'(t) + t\sigma_k^2| &= |E(e^{itX_k} - 1 - itX_k)(iX_k' + iX_k'')|\\ &\leqq t^2\gamma_k'/2 + 2\beta_k''|t|\,, \qquad \text{all } t\end{aligned}$$

which implies

$$|u_k'(t) - v_k'(t)| \leqq \gamma_k'|t|^2/2s_n^3 + 2\beta_k''|t|/s_n^2 + \sigma_k^4|t|^3/2s_n^4$$

and

$$\max\{|u_k'(t)|, |v_k'(t)|\} \leqq \gamma_k'|t|^2/2s_n^3 + 2\beta_k''|t|/s_n^2 + \sigma_k^2|t|/s_n^2 .$$

To get bounded on Π_k and Π_{kj}, where Π_{kj} is defined as Π_k/u_j if $j < k$ and Π_k/v_j if $j > k$, we argue exactly as in Feller [5], (3.6) to (3.14), but we use the index set $A = \{k \mid 1 - \beta_k'T^2/2s_n^2 \geqq 0\}$. This shows that

(6) $$|\Pi_k(t)| \leqq \exp\{-(t^2/2)[1 - T\Gamma/2^{\frac{1}{2}} - 2B - 2/T^2]\}$$

and

(7) $$|\Pi_{kj}(t)| \leqq \exp\{-(t^2/2)[1 - T\Gamma/2^{\frac{1}{2}} - 2B - 4/T^2]\}$$

for all k, $j \neq k$ and $|t| < T$.

Now define T by

(8) $$\frac{1}{T} = r\Gamma + sB , \qquad 0 \leqq s \leqq r ,$$

so that $1/T \leqq rR \leqq r\rho^3$. Then the bracket in (6) is bounded below by

(9) $$p_r = 1 - 1/r2^{\frac{1}{2}} - 2r^2\rho^6$$

while that in (7) is bounded below by

(10) $$q_r = 1 - 1/r2^{\frac{1}{2}} - 4r^2\rho^6$$

if $sT/r2^{\frac{1}{2}} > 2$, which is the case if

(11) $$s^2/8r^4 \geqq \rho^6 .$$

Combining the above, if we assume (1) with $K_1 \geqq 6$ and (8) and (11) we obtain

(12) $$|w_n(t)| \leqq \exp(-p_r t^2/2)\varphi(t) \qquad \text{and}$$

(13) $$|w_n'(t)| \leqq \exp(-q_r t^2/2)|t|\varphi(t)(\eta(t) + 1) + \exp(-p_r t^2/^2)|t|\psi(t)$$

for $|t| \leqq T$.

To carry out the computations define

(14) $$c_k = \Gamma\left(\frac{k+1}{2}\right).$$

so that for $\alpha > 0$, $k = 0, 1, \cdots,$

$$2\int_0^\infty t^k \exp(-\alpha t^2)\,dt = c_k(\alpha^{-\frac{1}{2}})^{k+1} .$$

For any polynomial $P(t) = \sum_0^m a_j t^j$ define for $\alpha > 0$,

(15) $$P^\frown(\alpha) = \sum_0^m a_j c_j(\alpha^{-\frac{1}{2}})^{j+1} = 2\int_0^\infty P(t)\exp(-at^2)\,dt .$$

Defining the polynomials $\varepsilon(t) = [\varphi(t)/t]^2$ and $\delta_1(t) = [\varphi(t)/t^2]^2$ we have, in the notation of the smoothing lemma,

(16) $$\varepsilon^2 \leqq \varepsilon^\frown(p_r)$$
$$\delta_1^2 \leqq \delta_1^\frown(p_r)$$

and

$$\delta_2^2 \leqq [\psi^2]^\wedge(p_r) + 2[\varphi\psi(\eta + 1)]^\wedge(m_r) + [\varphi^2(\eta + 1)^2]^\wedge(q_r),$$

where $m_r = (p_r + q_r)/2$.

These computations are quite tedious for any given value of r and ρ. To find a really good approximation to the best value of K_1 would require a computer. We have preferred to make rough upper estimations of the above for a few well chosen values of r with a ρ corresponding to a K_1 we hoped to attain. We then used that value of r which gave the best rough results to calculate a much more precise estimate. In this more precise estimate all the transformations in the right-hand side of (16) were calculated in full and then dominated by expressions of the form $(a_i\Gamma + b_i B)^2$, save the term $[\varphi^2(\eta + 1)^2]^\wedge(q_r)$. This term seemed too difficult to calculate explicitly and was bounded by

$$Q\varphi^2(Q)[c_{10}Q^2\Gamma^2/4 + 2Qc_9B\Gamma + 4c_8B^2 + 4c_8B + Qc_9\Gamma] + [\varphi^2]^\wedge(q_r) \leqq (a\Gamma + bB)^2,$$

some $a, b > 0$, where $Q = q_r^{-\frac{1}{2}}$.

Notice that the estimate given by the above calculations and smoothing lemma can only be improved by a better choice of r and a better approximation of the $[\varphi^2(\eta + 1)^2]^\wedge$ term. We feel this will yield little gain on $K_1 = 36$, but we invite the interested reader to better 36 if he can.

We have done much calculating and have found that the choice $r = 1.1$, $s = .62$, $K_1 = 36$ will lead to $\Delta_{n1} \leqq 36R$. Rather than present these tedious and uninformative calculations here, we content ourselves with stating in theorem form only the result $K_1 \leqq 72$.

We point out that $K_1 = 72$ is easily proved using $r = 1.1$, $s = 3.7/8\pi$, $P = p_r^{-\frac{1}{2}}$, $Q = q_r^{-\frac{1}{2}}$ and $\varepsilon^2 \leqq c_6\varphi^2(P)/P$, $\delta_1^2 \leqq c_0\varphi^2(P)/P^3$ and $\delta_2^2 \leqq c_{10}Q[\varphi(Q)(\eta(Q) + 1) + \psi(Q)]^2$. This is inefficient because of the use of c_{10} throughout the estimate of δ_2. We leave this easy verification to the reader.

4. Proof of Theorem 2. Egorov's proof of the L_∞ version of this theorem is based on the following well-known, easily-proved

LEMMA. *Let X and Y be random variables (in general dependent) with distributions F and G, and let H be the distribution of $X + Y$. If $\mathscr{N}$ denotes the standard normal distribution, then for all $\varepsilon > 0$, x real*

$$F(x - \varepsilon) - \mathscr{N}(x) - P(|Y| > \varepsilon) \leqq H(x) - \mathscr{N}(x) \leqq F(x + \varepsilon) - \mathscr{N}(x) + P(|Y| > \varepsilon).$$

COROLLARY. *With assumptions and notation of the above lemma and $0 < \eta < 1$ we have*

(a) $$\|H - \mathscr{N}\|_\infty \leqq \|F - \mathscr{N}\|_\infty + \eta(2\pi)^{-\frac{1}{2}} + P(|Y| > \eta)$$

(b) $$\|H - \mathscr{N}\|_1 \leqq 2(1 - \eta^2)^{-1}\|F - \mathscr{N}\|_1 + 4\eta(1 - \eta^2)^{-1}(2\pi)^{-\frac{1}{2}} + 2\eta^{-1}\int_0^\infty P(|Y| > y)\,dy.$$

PROOF. (a) Clear if one estimates $||\mathscr{N}(\cdot + \eta) - \mathscr{N}(\cdot)||_\infty$.

(b) The lemma implies that

$$\begin{aligned}|H(x) - \mathscr{N}(x)| \leqq |F(x+\varepsilon) - \mathscr{N}(x+\varepsilon)| + |\mathscr{N}(x+\varepsilon) - \mathscr{N}(x)| \\ + |F(x-\varepsilon) - \mathscr{N}(x-\varepsilon)| + |\mathscr{N}(x-\varepsilon) - \mathscr{N}(x)| \\ + P(|Y| > |\varepsilon|)\end{aligned}$$

for $\varepsilon \neq 0$. Set $\varepsilon = \eta x$ and integrate. (b) follows from the fact that

$$(2\pi)^{\frac{1}{2}} \int_0^\infty |\mathscr{N}(x) - \mathscr{N}(x \pm \eta x)|\, dx = \eta(1 \pm \eta)^{-1} .$$

Egorov now uses a technique originated by Bernstein: decompose $S_n/B_n^{\frac{1}{2}} = X + Y$, where X is a standardized sum of independent random variables to which our corollary to Theorem 1 may be applied. This gives an estimate for $||F - \mathscr{N}||_p$, $p = 1, \infty$ in the notation of (a) and (b) above. This decomposition must also be arranged so that Chebyshev's inequality yields a nice estimate of $P(|Y| > y)$. If K_n is any sequence of reals increasing to infinity, Egorov shows how to decompose $S_n/B_n^{\frac{1}{2}} = X + Y$ so that, under the assumptions of the theorem and notation of the above lemma,

$$||F - \mathscr{N}||_p \leqq C(K_n/B_n)^{\delta/2}$$

and

$$[*] \qquad P(|Y| > y) \leqq C/K_n y^2 , \qquad y > 0 ,$$

C an absolute, $p = \infty$. Our corollary to Theorem 1 shows this holds also for $p = 1$. Using [*] over the interval $[K_n^{-\frac{1}{4}}, \infty)$ and the bound 1 over $[0, K_n^{-\frac{1}{4}})$, substitution in (b) above gives

$$\Delta_{n1} \leqq C\{(K_n/B_n)^{\delta/2}(1 - \eta^2)^{-1} + \eta(1 - \eta^2)^{-1} + \eta^{-1}K_n^{-\frac{1}{4}}\} .$$

Now take $K_n = (s_n^2)^{2\delta/(2\delta+1)}$, $\eta = (s_n^2)^{-\delta/4\delta+2}$. The theorem follows since $B_n \leqq (1 + m)s_n^2$ and $s_n^2 = O(B_n)$ by assumption.

REMARK. L_p Chebyshev–Cramér expansions are attainable using an appropriate version of the smoothing lemma above. We present these elsewhere, if they are not already known.

Acknowledgment. We wish to thank Professor James Hannan for pointing out the need for an L_1 version of the Berry–Esséen Theorem and for several interesting discussions.

REFERENCES

[1] BIKYALIS, A. (1966). Estimates for the remainder term in the central limit theorem. *Litovsk. Matem. Sb.* **6** 321–346. (In Russian).

[2] EGOROV, V. A. (1970). Some limit theorems for *m*-dependent random variables. *Litovsk. Matem. Sb.* **10** 51–59. (In Russian).

[3] ESSÉEN, C. G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law. *Acta Math.* **77** 1–125.

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

[5] FELLER, W (1968). On the Berry–Esséen theorem. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **10** 261–268.
[6] HILLE, E. (1959). *Analytic Function Theory* **1**. Ginn, Boston.
[7] IBRAGIMOV, I. A. (1966). On the accuracy of Gaussian approximation to the distribution functions of sums of independent variables. *Theor. Probability Appl.* **11** 559–579.
[8] IBRAGIMOV, I. A. and LINNIK, YU. V. (1965). *Independent and Stationary Connected Random Variables*, Izd-vo "Nauka," Moscow. (In Russian).
[9] PETROV, V. V. (1970). On the central limit theorem for *m*-dependent variables. *Selected Transl. Math. Statist. and Prob.* **9** 83–88.
[10] ZOLOTAREV, V. M. (1964). On asymptotically best constants in refinements of mean limit theorems. *Theor. Probability Appl.* **9** 268–276.

DEPARTMENT OF STATISTICS
MICHIGAN STATE UNIVERSITY
WELLS HALL
EAST LANSING, MICHIGAN 48823