Z. Wahrscheinlichkeitstheorie verw. Geb. 10, 261–268 (1968)

On the Berry-Esseen Theorem *

WILLIAM FELLER

Received February 1, 1968

Summary. The paper presents a simple derivation of a generalized Berry-Esseen theorem not requiring moments.

1. Notations and Results

We consider a fixed number, n, of mutually independent random variables $X_1, \ldots, X_n$ with distributions $F_1, \ldots, F_n$. Let F denote the distribution of their sum $S = X_1 + \cdots + X_n$. We are interested in estimates for the discrepancy

$$\Delta = \sup |F(xs) - \mathfrak{N}(x)|; \tag{1.1}$$

here $\mathfrak{N}$ stands for the normal distribution with zero expectation and unit variance, while s is an appropriate positive norming factor. Upper bounds for Δ will be obtained in terms of truncated moments. As the derivation is independent of the mode of truncation, no restriction is placed on the latter. Thus, to each X_k we let correspond an interval $-\infty \leqq -\tau_k < 0 < \tau_k' \leqq \infty$ and put

$$\bar{X}_k = \begin{cases} X_k & \textit{if} \quad -\tau_k < X_k < \tau_k' \\ 0 & \textit{otherwise} \end{cases} \tag{1.2}$$

$$X_k = \bar{X}_k + X_k'. \tag{1.3}$$

Then, throughout this paper,

$$\alpha_k = E(\bar{X}_k), \qquad \beta_k = E(\bar{X}_k^2), \qquad \gamma_k = E(|\bar{X}_k|^3), \tag{1.4}$$

$$a = \alpha_1 + \cdots + \alpha_n, \qquad b = \beta_1 + \cdots + \beta_n, \qquad c = \gamma_1 + \cdots + \gamma_n. \tag{1.5}$$

We begin with a strengthening of the classical Berry-Esseen theorem. It is designed to cover variables without third moment, but the estimate (1.8) is sharper than the classical one even when third moments do in fact exist.

Theorem 1. *Suppose that* $E(X_k) = 0$ *and* $E(X_k^2) = \sigma_k^2 < \infty$ *for* $k = 1, \ldots, n$. *Put*

$$s^2 = \sigma_1^2 + \cdots + \sigma_n^2 \tag{1.6}$$

and

$$\beta_k' = E(X_k'^2), \qquad b' = \beta_1' + \cdots + \beta_n'. \tag{1.7}$$

Then

$$\Delta \leqq 6\left(\frac{c}{s^3} + \frac{b'}{s^2}\right). \tag{1.8}$$

Note. No attempt at sharp estimates was made. Instead, all calculations are based on simple rational approximations so as to render trivial the verification

* Research connected partly with a Project for research in probability at Princeton University, supported by the Army Research Office.

of each step. No impressive improvement of the constant 6 can be obtained by arguments based on the crude estimate of (2.2). For an improvement of this inequality see ZOLOTAREV [9]. (The direct calculations in [1] for identically distributed variables lead to the constant 33/4 instead of 6.)

From Theorem 1 it is easy to derive an analogous result for arbitrary random variables. To formulate it we introduce the quantities

$$\pi_k = P\{X_k' \neq 0\}, \qquad p = \pi_1 + \cdots + \pi_n \tag{1.9}$$

and

$$\lambda_k = \alpha_k^2/\pi_k. \tag{1.10}$$

If $\pi_k = 0$ we define $\lambda_k = 0$ when $\alpha_k = 0$, and $\lambda_k = \infty$ otherwise. (It will be seen that λ_k is a measure for the appropriateness of the given centering of X_k.)

Theorem 2. *If*

$$s^2 \geqq b + \sum \lambda_k \tag{1.11}$$

then

$$\Delta \leqq 6\left(\frac{c}{s^3} + \frac{s^2 - b}{s^2}\right) + p. \tag{1.12}$$

(In the limit as $s \to \infty$ the right side is $\geqq 6$, and so the theorem is true when $\lambda_k = \infty$ for some k.) Under the conditions of Theorem 1 one has $\alpha_k = -E(X_k')$ and hence by CHEBYSHEV's inequality

$$\lambda_k \leqq E(X_k'^2) = \sigma_k^2 - \beta_k. \tag{1.13}$$

Thus (1.11) is satisfied and $s^2 - b = b'$, so that (1.12) represents a variant of (1.8). The condition (1.11) is not essential because the norming constant s in (1.1) can always be replaced by $s' < s$ at the expense of increasing the bound Δ for the error by

$$\sup |\mathfrak{N}(x) - \mathfrak{N}(s' x/s)| \leqq \sqrt{2\pi e}\,\frac{s - s'}{s}.$$

It follows that Theorem 2 is generally applicable, but the unwieldy constants λ_k make it unattractive. It is therefore preferable to replace Theorem 2 by a weakened version in which the constants λ_k are eliminated. To achieve this we introduce condition (1.14) which assures that the centering of the variables X_k is not too absurd. In fact, condition (1.14) is trivially satisfied except when $E(X_k)$ is finite or equals $\pm\infty$ (in the obvious sense that only one tail diverges). In these exceptional cases the validity of (1.14) can be achieved by adding an appropriate centering constant to X_k. In theorem 1 the conditions (1.14) and (1.15) are satisfied with $t_k = t_k' = \infty$. As remarked before, (1.15) imposes no serious restriction.

Theorem 3. *Suppose that*

$$\int_{-t_k}^{\tau'_k} x\, F_k(dx) \leqq 0 \quad \textit{and} \quad \int_{-\tau_k}^{t'_k} x\, F_k(dx) \geqq 0 \tag{1.14}$$

for some $-\infty \leqq -t_k \leqq -\tau_k$ *and* $\tau_k' \leqq t_k' \leqq \infty$. *If*

$$s^2 \geqq \sum_{k=1}^{n} \int_{-t_k}^{t'_k} x^2 F_k(dx), \tag{1.15}$$

then (1.12) *holds.*

2. Discussion

It is clear in principle that the standard method of truncation enables one to apply the Berry-Esseen theorem to arbitrary variables, but the truncated variables are not properly centered and one requires error estimates in terms of quantities relating to the original variables. The case of identically distributed variables with finite variances and the classical norming was first considered by M. L. Katz [3] who gave a bound for the error in terms of $E(X^2 g(X))$, where g is an even unbounded monotone function. This result was generalized to variable distributions with variances by V. V. Petrov [6]. Now the main application of the Berry-Esseen theorem is to triangular arrays subject to the central limit theorem, and it is therefore desirable to estimate the error by quantities related to the Lindeberg conditions. This point of view is implicit in Yu. P. Studnev [7] who had previously in [8] derived an asymptotic form of the Berry-Esseen theorem depending on the integral of the function

$$L_n(x) = \frac{1}{s^2} \sum_{k=1}^{n} \int_{|t|>xs} t^2 F_k(dt) \tag{2.1}$$

occurring in the Lindeberg condition. Studnev's claim that his theorem implies the Katz-Petrov result is unfortunately based on an oversight since in [7] he has inadvertedly ommitted the term $o(s^{-1})$ which is essential in the asymptotic estimate of [8].

Lindeberg's condition requires that $L_n(x) \to 0$ as $n \to \infty$ for every fixed $x > 0$. In the notations of theorem 1 with truncation at $\pm s_n$ this would imply that $b' = o(s^2)$ and $c = o(s^3)$. Conversely, these two conditions together imply that $L_n(x) \to 0$, but obviously neither of these conditions alone suffices to ensure that $L_n(x) \to 0$. It follows trivially that both terms on the right side in (1.8) are indispensable. This simple argument answers a problem raised by Ibragimov and Osipov [2] who constructed a complicated example to show that an estimate of the form $\Delta < C \cdot L_n(x)$ (for a fixed x) is impossible even if the existence of moments of some order $2 + \alpha > 2$ is assumed. A similar remark applies to the estimate (1.12).

Several variants and generalizations of the Katz-Petrov theorem were derived in Osipov [4], and Osipov and Petrov [5], by increasingly complex method.

Our theorem 1 (with an unspecified constant on the right side) is actually contained in [5]. Theorems 2 and 3 formulate a general version of the Berry-Esseen theorem involving only quantities naturally tied to the central limit theorem.

It is shown in the present paper that the apparent great generality of these theorems is deceptive inasmuch as they can be derived as simple corollaries of theorem 1. This is done by a method of truncation leading to zero expectations and prescribed variances (Section 4). Another purpose of this paper is to show that theorem 1 can be derived in a simple manner using a new technique for bounding products of characteristic functions in the absence of third moments. It is hoped that this method may find wider applications.

The proof of theorem 1 is based on the well-known lemma (due essentially to BERRY) that if ϕ denotes the characteristic function of F then for any $T > 0$

$$\pi\Delta \leqq \int_{-T}^{T} \left|\phi\left(\frac{\xi}{s}\right) - \exp\left(-\frac{1}{2}\xi^2\right)\right| \cdot |\xi^{-1}|\, d\xi + \frac{24}{T\sqrt{2\pi}}. \tag{2.2}$$

(See, for example, [1], p. 512.) The integrand will be estimated using the obvious identity

$$u_1 \cdots u_n - v_1 \cdots v_n = \sum_{\nu=1}^{n} u_1 \cdots u_{\nu-1}(u_\nu - v_\nu)\, v_{\nu+1} \cdots v_n \tag{2.3}$$

with

$$u_k = \phi_k(\xi/s), \qquad v_k = \exp(-\tfrac{1}{2}\sigma_k^2 \xi^2/s^2), \tag{2.4}$$

where ϕ_k is the characteristic function of F_k.

3. Proof of Theorem 1

Denote the integral occurring in (2.2) by I. Our first aim is to show that

$$I \leqq \int_{-\infty}^{+\infty} e^{-\xi^2/8} \cdot \sum \left|\phi_k\left(\frac{\xi}{s}\right) - \exp\left(-\frac{\sigma_k^2\,\xi^2}{2s^2}\right)\right| \cdot |\xi|^{-1}\, d\xi. \tag{3.1}$$

From the familiar inequalities for the partial sums of the Taylor series for e^{ix} we get trivially

$$\left|\phi_k\left(\frac{\xi}{s}\right) - 1 + \frac{\sigma_k^2\,\xi^2}{2s^2}\right| \leqq \frac{\gamma_k\,|\xi|^3}{6s^3} + \frac{\beta_k'\,\xi^2}{s^2}. \tag{3.2}$$

Remembering that $0 \leqq e^{-x} - 1 + x < \frac{1}{2}x^2$ for all $x > 0$ we conclude that

$$\left|\phi_k\left(\frac{\xi}{s}\right) - \exp\left(\frac{-\sigma_k^2\,\xi^2}{2s^2}\right)\right| \leqq \frac{\gamma_k\,|\xi|^3}{6s^3} + \frac{\beta_k'\,\xi^2}{s^2} + \frac{\xi^4}{8s^4}\sum \sigma_k^4. \tag{3.3}$$

Assuming (3.1) to be true this implies

$$I \leqq \int_{-\infty}^{+\infty} e^{-\xi^2/8}\left[\frac{c\,\xi^2}{6s^3} + \frac{b'\,|\xi|}{s^2} + \frac{|\xi|^3}{8s^4}\sum \sigma_k^4\right] d\xi \leqq \frac{32}{9} + 8\,\frac{b'}{s^2} + 8\sum\left(\frac{\sigma_k}{s}\right)^4. \tag{3.4}$$

(In the first term we used the inequality $\sqrt{2\pi} < 8/3$.)

To justify (3.1) from (2.3) we need an upper bound for the u_k, and here the difficulty arises that these factors are not necessarily uniformly small in any appropriate interval of integration $(-T, T)$. It is therefore necessary to partition the subscripts into two classes. Leaving the parameter T for the time being indeterminate we say that *a subscript k belongs to the class A iff*

$$\frac{\beta_k^{1/2}\,T}{s} \leqq \frac{4}{3}. \tag{3.5}$$

(The right side may be replaced by any number that is $\leqq \sqrt{2}$ but close to $\sqrt{2}$.)

Clearly

$$\begin{aligned} \phi_k(\xi/s) - 1 + \frac{\beta_k\,\xi^2}{2s^2} &= E\left(e^{i\xi \bar{X}_k/s} - 1 - \frac{i\,\xi\,\bar{X}_k}{s} + \frac{\xi^2\,\bar{X}_k^2}{2s^2}\right) \\ &\quad + E\left(e^{i\xi X'_k/s} - 1 - \frac{i\,\xi\,X'_k}{s}\right) \end{aligned} \tag{3.6}$$

and hence

$$(3.7)\qquad |\phi_k(\xi/s)| \leqq \left|1 - \frac{\beta_k \xi^2}{2s^2}\right| + \frac{\gamma_k |\xi|^3}{6s^3} + \frac{\beta'_k \xi^2}{2s^2}.$$

When $k \in A$ and $|\xi| < T$ the first quantity within absolute values on the right is positive, and so

$$(3.8)\qquad |\phi_k(\xi/s)| \leqq \exp\left(-\frac{1}{2}\xi^2\left[\frac{\beta_k}{s^2} - \frac{\gamma_k T}{3s^3} - \frac{\beta'_k}{s^2}\right]\right).$$

Taking the product over all $k \in A$ we get (with obvious notations) the inequality

$$(3.9)\qquad \prod_{k \in A} |\phi_k(\xi/s)| \leqq \exp\left(-\frac{1}{2}\xi^2\left[\frac{b_A}{s^2} - \frac{c_A T}{3s^3} - \frac{b'_A}{s^2}\right]\right)$$

valid for $|\xi| < T$. We proceed to find a lower bound for the expression D within brackets.

From the moment inequality we conclude that for any subscript in the complement of A

$$(3.10)\qquad \gamma_k \geqq \beta_k^{3/2} \geqq \frac{4s}{3T}\beta_k.$$

Summing over $k \notin A$ we get

$$(3.11)\qquad c - c_A \geqq \frac{4s}{3T}(b - b_A).$$

Consider for the moment b_A and c_A as arbitrary non-negative parameters. A routine computation shows that among all such combinations satisfying (3.11) the one minimizing the quantity D within brackets in (3.9) corresponds to the smallest possible b_A, and so the minimum is achieved when

$$b_A = b - \frac{3T}{4s}c \quad \text{and} \quad c_A = 0.$$

Remembering that $b + b' = s^2$ we conclude that

$$(3.12)\qquad \prod_{k \in A} |\phi_k(\xi/s)| \leqq \exp\left(-\frac{1}{2}\xi^2\left[1 - \frac{3Tc}{4s^3} - \frac{2b'}{s^2}\right]\right).$$

We now return to the identity (2.3). For a subscript $k \in A$ the bound obtained in (3.8) for $|\phi_k(\xi/s)|$ exceeds $\exp\left(-\frac{\sigma_k^2 \xi^2}{2s^2}\right)$ and applies therefore to both u_k and v_k. For subscripts outside A we use the trite inequalities $|u_k| \leqq 1$ and $|v_k| \leqq 1$. When $\nu \notin A$ it follows that the factor of $u_\nu - v_\nu$ in (2.3) is in norm less than (3.12). When $\nu \in A$ the same is true provided the contribution of $|\phi_\nu(\xi/s)|$ is omitted. But a glance at (3.9) and (3.5) makes it clear that the omission of the ν-th term will increase the exponent by less than

$$(3.13)\qquad \frac{\beta_\nu \xi^2}{2s^2} \leqq \frac{8}{9} \cdot \frac{\xi^2}{T^2},$$

and so the right side of (2.3) is in absolute value less than

$$(3.14)\qquad \sum |u_k - v_k| \cdot \exp\left(-\frac{1}{2}\xi^2\left[1 - \frac{3Tc}{4s^3} - \frac{2b'}{s^2} - \frac{16}{9T^2}\right]\right).$$

This proves the assertion (3.1) *provided only that T is chosen so that*

$$(3.15)\qquad T \cdot \frac{c}{s^3} + \frac{8}{3}\frac{b'}{s^2} + \frac{64}{27T^2} \leqq 1.$$

With this proviso we have now the inequality (3.4) and we proceed to express the last sum on the right in terms of the quantities occurring in the theorem. From now on we suppose that

$$\frac{c}{s^3} + \frac{b'}{s^2} \leqq \frac{1}{6}, \tag{3.16}$$

for otherwise there is nothing to prove. By the moment inequality $\beta_k^3 \leqq \gamma_k^2$ and hence

$$\sigma_k^4 \leqq (\gamma_k^{2/3} + \beta_k')^2 \leqq c^{1/3}\gamma_k + (2c^{2/3} + b')\,\beta_k'. \tag{3.17}$$

Among all linear combinations of this form with coefficients subject to (3.16) the maximum is assumed for $c = s^3/6$ and $b' = 0$. Since $6^{1/3} > 9/5$ we conclude that

$$\sum\left(\frac{\sigma_k}{s}\right)^4 \leqq \frac{5}{9}\frac{c}{s^3} + \frac{50}{81}\frac{b'}{s^2}, \tag{3.18}$$

and hence finally from (3.4)

$$I < 8\frac{c}{s^3} + 13\frac{b'}{s^2}. \tag{3.19}$$

We introduce this estimate into the basic inequality (2.2) using the fact that $(2\pi)^{-1/2} < 2/5$. To prove theorem 1 it suffices to choose T such that

$$8\frac{c}{s^3} + 13\frac{b'}{s^2} + \frac{48}{5T} \leqq 6\pi\left(\frac{c}{s^3} + \frac{b'}{s^2}\right). \tag{3.20}$$

But $6\pi > 94/5$, and hence (3.20) holds with

$$\frac{1}{T} = \frac{9}{8}\frac{c}{s^3} + \frac{5}{8}\frac{b'}{s^2}. \tag{3.21}$$

With this definition the left side in (3.15) reduces to

$$\frac{8}{9} - \left(\frac{5T}{9} - \frac{8}{3}\right)\frac{b'}{s^2} + \frac{64}{27T^2}. \tag{3.22}$$

Now (3.16) implies that $T^{-1} < 3/16$; thus the coefficient of b' is positive and the last term in (3.22) is $< 1/12$. The quantity (3.22) is therefore < 1 as required.

4. Proof of Theorem 2

The theorem is trivially true when $\pi_k = 0$ but $\alpha_k \neq 0$ for some k. Otherwise we define for each k a point $x_k = -\alpha_k/\pi_k$ with the understanding that $x_k = 0$ when $\alpha_k = \pi_k = 0$. With λ_k defined in (1.10) we now define a number $0 \leqq \eta \leqq 1$ by

$$\sum \lambda_k = \eta(s^2 - b). \tag{4.1}$$

For definiteness assume first that $\eta \neq 0$.

Let $U_1, \ldots, U_n$ be random variables independent of each other and of the given variables X_k, and such that each U_k assumes the values 1 and 0 with probabilities η and $1 - \eta$, respectively. Put

$$*X_k = \begin{matrix} \bar{X}_k & \text{if} & X_k = \bar{X}_k \\ U_k x_k/\eta & \text{if} & X_k = X_k'. \end{matrix} \tag{4.2}$$

The distribution of $*X_k$ differs from the distribution of $\bar{X}_k$ in that it attributes a smaller weight to the origin and an extra weight of mass $\eta\pi_k$ to the point

x_k/η. Clearly

$$E(^*X_k) = E(\bar{X}_k) + x_k \pi_k = 0 \tag{4.3}$$
$$E(^*X_k^2) = E(\bar{X}_k^2) + \lambda_k/\eta = \beta_k + \lambda_k/\eta$$

so that the sum $^*S = {}^*X_1 + \cdots + {}^*X_n$ has variance s^2. This sum differs from S only if at least one among the variables $X_1', \ldots, X_n'$ is different from zero, and the probability for this event is $\leqq p$. Thus

$$\Delta \leqq {}^*\Delta + p \tag{4.4}$$

where $^*\Delta$ is the maximum discrepancy between the distribution of $^*S$ and the normal distribution. But theorem 1 applies to the variables $^*X_k$, and it remains to calculate the quantities $^*\gamma_k$ and $^*\beta_k'$ corresponding to the truncated moments γ_k and β_k'. For definiteness suppose $x_k > 0$. If $s \geqq \tau_k'$ we use truncation at τ_k' and see that $^*\beta_k' = \lambda_k/\eta$ or $^*\beta_k' = 0$ according as $x_k > \tau_k'\eta$ or $x_k < \tau_k'\eta$. In the first case $^*\gamma_k = \gamma_k$, in the second

$$^*\gamma_k = \gamma_k + x_k^3 \pi_k/\eta^2 \leqq \gamma_k + s\,\lambda_k/\eta\,.$$

If, on the other hand, $s \leqq \tau_k'$ we use truncation at s and conclude trivially that under any circumstances

$$\frac{^*\gamma_k}{s^3} + \frac{^*\beta_k'}{s^2} \leqq \frac{\gamma_k}{s^3} + \frac{\lambda_k}{\eta\, s^2}\,. \tag{4.5}$$

Summation over k now shows that $^*\Delta$ is indeed bounded by the first term on the right side in (1.12), and this concludes the proof when $\eta > 0$.

The contingency $\eta = 0$ arises if, and only if, $\lambda_k = 0$ for all k. This means that all $\bar{X}_k$ have zero expectation. The preceding argument applies, except that the U_k now assume the values ± 1 each with probability $\pi_k/2$, and $x_k^2 = (s^2 - b)/p$ for all k. The sum $^*S$ has again zero expectation and variance s^2.

5. Proof of Theorem 3

This is an immediate consequence of theorem 2. Suppose $\alpha_k < 0$. By the second condition in (1.14) and Schwarz' inequality

$$\alpha_k^2 \leqq \left(\int_{\tau'_k}^{t'_k} x\, F_k(dx)\right)^2 \leqq \pi_k \int_{\tau_k}^{t'_k} x^2\, F_k(dx)\,. \tag{5.1}$$

If $\alpha_k > 0$ the integration extends between $-t_k'$ and $-\tau_k'$. With the notation of the preceding section this implies that

$$\sum \lambda_k + b \leqq \sum \int_{-t_k}^{t'_k} x^2 F_k(dx) \leqq s^2 \tag{5.2}$$

and so the condition of theorem 2 is fulfilled.

References

1. Feller, W.: An introduction to probability theory and its applications, Vol. II. New York: J. Wiley 1966.
2. Ibragimov, I. A., and L. V. Osipov: On an estimate of the remainder in Lindeberg's theorem. Theor. Probab. Appl. **11**, 125—128 (1966).

3. KATZ, M. L.: Note on the Berry-Esseen theorem. Ann. math. Statistics **34**, 1107—1108 (1963).
4. OSIPOV, L. V.: Refinement of Lindeberg's Theorem. Theor. Probab. Appl. **11**, 299—302 (1966).
5. — and V. V. PETROV: On an estimate of the remainder in the central limit theorem. Theor. Probab. Appl. **12**, 281—286 (1967).
6. PETROV, V. V.: An estimate of the deviation of the distribution of a sum of independent random variables from the normal law. Doklady Akad. Nauk SSSR n. Ser. **160**, 1013 to 1015 (1965).
7. STUDNEV, YU. P.: A remark on the Katz-Petrov theorem. Theor. Probab. Appl. **10**, 682—684 (1965).
8. — On the role of Lindeberg's conditions. Dopovidi Akad. Nauk. Ukrain. RSR **3**, 239—242 (1958). Select. Transl. math. Statist. Probab. **4**, 303—306 (1963).
9. ZOLOTAREV, V. M.: An absolute estimate of the remainder in the central limit theorem. Theor. Probab. Appl. **11**, 95—105 (1966).

Prof. W. FELLER
Department of Mathematics
Princeton University, Princeton, N. J.
and
Rockefeller University, New York, N.Y.,
USA