

THE MULTIVARIATE CENTRAL LIMIT THEOREM FOR REGULAR CONVEX SETS

BY T. K. MATTHES¹

University of Oregon

Let X_1, X_2, \dots be i.i.d. random vectors in R^k . Let P_n denote the probability measure induced by the normalized sum and let Q_n denote the multivariate Edgeworth signed measure with terms through $n^{-\frac{1}{2}}$. If C is a member of a class of convex bodies whose boundaries are sufficiently smooth and possess positive Gaussian curvatures, and X_1 has fourth moments, it is shown that $P_n(C) - Q_n(C) = O(n^{-k/(k+1)})$ where the bound is uniform. If, moreover, X_1 has a nonlattice distribution, the difference is $o(n^{-k/(k+1)})$.

1. Introduction and main results. Let $X_i \in R^k$, $i = 1, 2, \dots$; $k > 1$ be a sequence of independent identically distributed random vectors which, for convenience, are normalized to have zero means and covariance matrix identity. Let P_n denote the probability measure on the Borel subsets of R^k induced by $Z_n = n^{-\frac{1}{2}} \sum_{i=1}^n X_i$. There is a continuing interest in the rate of convergence of P_n to the normal probability measure Φ .

Beginning with the distribution function F_n of Z_n , Bergstrom (1945) showed $\sup_x |F_n(x) - \Phi(x)| < cn^{-\frac{1}{2}}$ where c depends on k and the third moments. Assuming $E(|X_1|^{3+\delta}) < \infty$, Bhattacharya (1968) strengthened this to $\sup\{|P_n(C) - \Phi(C)| : C \in \mathcal{C}\} < cn^{-\frac{1}{2}}$, where \mathcal{C} is the class of convex subsets of R^k . Bhattacharya (1971) studied the weak convergence of P_n and has several theorems for convex sets. Suppose X_1 has a strongly nonlattice distribution; i.e., its characteristic function satisfies Cramér's condition: $\limsup_{|t| \rightarrow \infty} |\varphi(t)| < 1$. Then he showed $\sup\{|P_n(C) - Q_{ns}(C)| : C \in \mathcal{C}\} = o(n^{-(s-2)/2})$ if $E(|X_1|^s) < \infty$ for $s \geq 3$, where Q_{ns} denotes the multivariate Edgeworth "approximate measure" with terms up to $n^{-(s-2)/2}$.

At the other extreme R. R. Rao (1961) considers the lattice case. He has an expansion of $P_n(B)$ for any Borel set, with error $O(n^{-(s-2)/2})$. The usefulness of Rao's expansion would appear to be limited by its complexity. However, Yarnold (1972) applies Rao's expansion in case B is convex, and in particular, applies it to the distribution of the chi square goodness of fit test statistic. The approximation with error $O(n^{-1})$ leads to the interesting and difficult lattice point problem of finding the number of lattice points in an ellipsoid.

Inasmuch as these Edgeworth type expansions quickly become unmanageable

Received January 11, 1974; revised August 27, 1974.

¹ Ted Matthes died February 10, 1975, after this paper had been accepted for publication. His friends and colleagues wish to express their deepest sorrow at his death.

AMS 1970 subject classifications. Primary 62E20, 60F05; Secondary 60B15.

Key words and phrases. Multivariate central limit theorem, Edgeworth expansion, rates of convergence, probability of convex sets, nonlattice distributions.

as more terms are included, one is motivated to seek the weakest conditions under which P_n is approximated by Q_{n3} . In this regard Esseen (1945) showed that $|P_n(C) - \Phi(C)| \leq c\beta_4^{\frac{1}{2}}n^{-k/(k+1)}$ for all centered spheres, where β_4 denotes the sum of the 4th moments of the components of X_1 . Consequently, not much is gained for large k by adding the term of apparent order $n^{-\frac{1}{2}}$ that arises in the lattice case, at least for such spheres.

Our theorems below extend Esseen's results uniformly to a class of smooth positively curved convex bodies, with error still $O(n^{-k/(k+1)})$, and $o(n^{-k/(k+1)})$ if the distribution is nonlattice.

Suppose the boundary of C is continuously twice differentiable. Let $x_0 \in \partial C$ and introduce a local Euclidean coordinate system in Π , the tangent plane at x_0 . Consider the distance of $x \in \partial C$ to its projection in Π as a function of the projection point. In the Taylor series approximation to the distance the eigenvalues $\kappa_1, \kappa_2, \dots, \kappa_{k-1}$ of the matrix associated with the quadratic terms are the principal curvatures at x_0 . The Gaussian curvature $K(x_0)$ is defined as the product of the κ_i 's.

Suppose in addition C is bounded and $K(x) > 0$, $x \in \partial C$. Then there is a one to one continuous correspondence between $x \in \partial C$ and unit outer normals t , given by $x = \text{grad } H(t)$, $|t| = 1$. Here H is the support function of C : $H(t) = \sup_{x \in C} (x \cdot t)$. If H is twice differentiable, the matrix of its second order partial derivatives at $|t| = 1$ has $k - 1$ nonzero eigenvalues, R_1, R_2, \dots, R_{k-1} which are the principal radii of curvature at the corresponding $x \in \partial C$. Hence, one may compute $K(x)$ in terms of the support function. Only elementary facts concerning the differential geometry of convex bodies are assumed here. A standard reference is [4].

Define $\mathcal{C}(M)$ to be the class of convex bodies $C \subset R^k$ with the following properties:

- (i) $0 < K(x) < M$, $x \in \partial C$ and
- (ii) the support function of C has partial derivatives up to order $[(k + 7)/2]$

which are absolutely bounded by M on $|t| = 1$.

It follows that the principal radii of curvatures are uniformly bounded away from zero and infinity over the class. In particular, the Gaussian curvatures have the same property.

Our principal results are the following theorems. Throughout $Q_n(C) = Q_{n3}(C)$ and $\gamma = k/(k + 1)$.

THEOREM 1. *There exists a constant $c(k, M)$ depending only on k and M such that for all normalized distributions*

$$(1.1) \quad \sup \{|P_n(C) - Q_n(C)| : C \in \mathcal{C}(M)\} < c(k, M)\beta_4^{\frac{1}{2}}n^{-\gamma}, \quad n = 1, 2, \dots$$

The distribution of X_1 is said to be nonlattice if $|\varphi(t)| < 1$, $t \neq 0$; that is, if each nonzero linear combination of coordinates has a one-dimensional nonlattice distribution.

THEOREM 2. Suppose $\beta_4 < \infty$ and X_1 has a normalized nonlattice distribution. Then

$$(1.2) \quad \lim_{n \rightarrow \infty} n^r \sup\{|P_n(C) - Q_n(C)| : C \in \mathcal{C}(M)\} = 0.$$

It is desirable to escape the restriction that the breadths of the convex set are bounded, as is entailed for $C \in \mathcal{C}(M)$, owing to the bounded radii of curvatures. Define the subclass $\mathcal{C}(M, \rho) = \{C \in \mathcal{C}(M) : C \text{ contains a sphere about the origin of radius } \rho\}$. For $C \in \mathcal{C}(M, \rho)$ consider the dilation $\lambda C = \{\lambda x : x \in C\}$.

THEOREM 3. There exists a constant $c(k, M, \rho)$ such that

$$(1.3) \quad |P_n(\lambda C) - Q_n(\lambda C)| \leq c(k, M, \rho) \beta_4^{\frac{3}{4}r} n^{-r}, \quad 0 < \lambda < \infty; \quad n = 1, 2, \dots$$

whenever $C \in \mathcal{C}(M, \rho)$ and X_1 is normalized.

The proofs of these theorems are given in Sections 2–4, in each case preceded by preparatory lemmas. Lemma 6 in Section 4 has independent interest inasmuch as the difference in (1.3) is more precisely bounded as a function of λ and $\theta = \beta_4^{\frac{3}{4}} n^{-\frac{1}{4}}$.

2. Proof of Theorem 1. The proof follows broadly along lines developed by Esseen in the multivariate case. Let φ_n denote the characteristic function of P_n and put

$$(2.1) \quad \varphi_n^*(t) = \{1 + E(it \cdot X_1)^3 / (6n^{\frac{1}{4}})\} \exp(-|t|^2/2).$$

The inverse Fourier transform of φ_n^* is dQ_n/dx . Denote the Fourier transform of the indicator function $I_{[x \in C]}$ by \hat{I}_C ; the uniform distribution on $\{x \in R^k : |x| \leq \varepsilon\}$ by U_ε ; and the characteristic function of U_1 by u . Then U_ε has characteristic function $u(\varepsilon t)$. Here and henceforth $|x|$ is Euclidean norm, $(t \cdot x)$ is inner product and $\|x\|$ is sup norm. Integration is over R^k unless indicated. In this section c will denote any constant depending on k but not β_4 or $\mathcal{C}(M)$.

Inasmuch as $\varphi_n \cdot \hat{I}_C$ may not be integrable, I_C is smoothed by convolution. Let C^ε denote a Euclidean ε neighborhood of C , $\varepsilon > 0$ and where possible, $C^{-\varepsilon}$ is that set such that $(C^{-\varepsilon})^\varepsilon = C$. Choosing $\varepsilon_0 > 0$ strictly less than the minimum (positive) principal radii of curvature that applies to $\mathcal{C}(M)$, $C^{-\varepsilon}$ is defined provided $\varepsilon \leq \varepsilon_0$. Define the convolutions

$$(2.2) \quad G_\varepsilon(x) = I_{C^\varepsilon} * U_{|\varepsilon|}(x), \quad |\varepsilon| < \varepsilon_0.$$

LEMMA 1. There exists $c(k)$ such that

$$|P_n(C) - Q_n(C)| \leq \max |\int G_{\pm\varepsilon} d(P_n - Q_n)| + c(1 + \beta_4^{\frac{3}{4}} n^{-\frac{1}{4}}) \varepsilon$$

whenever $C \in \mathcal{C}(M)$; $0 < \varepsilon < \varepsilon_0$; and all n, β_4 . (The maximum refers to the choice of sign.)

PROOF. G_ε has range $[0, 1]$ and if $\varepsilon > 0$, $G_\varepsilon(x) = 1$, $x \in C$; $= 0$, $x \notin C^{2\varepsilon}$. Hence

$$(2.3) \quad \int G_\varepsilon d(P_n - Q_n) \geq P_n(C) - Q_n(C) - \int_{C^{2\varepsilon}-C} |dQ_n|.$$

Now,

$$(2.4) \quad \left| \frac{dQ_n(x)}{dx} \right| \leq c(1 + \beta_4^{\frac{1}{2}} n^{-\frac{1}{2}}) \exp(-|x|^2/3),$$

and

$$\int_{C^{2\varepsilon-C}} d\Phi \leq c\varepsilon,$$

as Rao and Bhattacharya have shown (cf. [3]). This leads to an upper bound for $P_n(C) - Q_n(C)$. A similar argument with $G_{-\varepsilon}$ yields a lower bound and proves the lemma. \square

Herz (1962) has studied the precise asymptotic behavior of the Fourier transform of a convex body with smooth positively curved boundary. Under the conditions we are assuming, an immediate consequence of his theorem is

$$(2.5) \quad |\hat{I}_C(t)| \leq K^{-\frac{1}{2}} |t|^{-(k+1)/2} + O(|t|^{-(k+3)/2}), \quad |t| \rightarrow \infty,$$

where K is the minimum Gaussian curvature over ∂C and the error term depends on the bound on the derivatives of $H(t)$ to order $[(k+7)/2]$.

LEMMA 2. *There exist constants B_1, B_2 depending on k and M such that*

$$(2.6) \quad |\hat{I}_C \in (t)| < B(t) \equiv \min\{B_1 |t|^{-(k+1)/2}, B_2\}, \quad |\varepsilon| < \varepsilon_0$$

uniformly for $C \in \mathcal{C}(M)$.

PROOF. It is required that Herz's theorem holds uniformly. Now, the support function of C^ε is $H(t) + \varepsilon|t|$. These clearly have uniformly bounded derivatives on $|t| = 1$. Moreover, in view of the way ε_0 was chosen, the principal radii of curvatures for C^ε , $-\varepsilon_0 < \varepsilon < 0$ remain bounded below; the Gaussian curvatures are bounded both ways. It is more difficult to establish that under these conditions the error term in (2.5) is also uniform for $C \in \mathcal{C}(M)$, $|\varepsilon| < \varepsilon_0$ (cf. Lemma 4 of [8]). Finally, $|\hat{I}_C \varepsilon|$ is always bounded by the measure of C^ε , which in turn is bounded in terms of the maximum possible radii of curvature. The lemma follows. \square

When C is a sphere \hat{I}_C is related to a Bessel function. Esseen obtained the bounds (2.5) in that case from asymptotic properties of Bessel functions. Denote the cube

$$R_T = \{(t_1, \dots, t_k) : 0 \leq t_j \leq T, \text{ each } j\}$$

and let $R_T + a$ be its translation by a .

LEMMA 3. *Assume $\beta_4 < \infty$ and X_1 is normalized. For $T > 0$,*

$$(2.7) \quad e(T, \varphi) \equiv \sup_n \sup_a n^{k/2} \int_{R_T+a} |\varphi|^n dt < \infty;$$

$$(2.8) \quad \text{for } T < c\beta_4^{-\frac{1}{2}}, \quad e(T, \varphi) < c,$$

independent of φ .

PROOF. Esseen ([7]) page 100) bounds the integrals in question in (2.8) if instead R_T is replaced by any sphere of radius essentially $\beta_4^{-\frac{1}{2}}$. Since any R_T

may be covered by finitely many such spheres and $R_r + a$ is covered by their translations, the finiteness in (2.7) follows. \square

Recall that $\|t\|$ is sup norm. The following lemma is also true with Euclidean norm. Define the function

$$(2.9) \quad u_0(t) = \min\{|t|^{-(k+1)/2}, 1\}$$

LEMMA 4. For each $\varepsilon > 0$, $T > 0$,

$$(2.10) \quad \int_{\|t\| > n^{1/2}T} B(t)u_0(\varepsilon t)|\varphi_n(t)| dt < cB_1e(T, \varphi)(n^{1/2}T)^{-k}\varepsilon^{-(k-1)/2},$$

where B_1 and $e(T, \varphi)$ are given in (2.6), (2.7).

PROOF. Making a change of variables, the integral in (2.10) becomes

$$(2.11) \quad I = n^{k/2} \int_{\|t\| > T} B(n^{1/2}t)u_0(n^{1/2}\varepsilon t)|\varphi^n| dt.$$

Let $\{t: \|t\| > T\} = \bigcup_{m=1}^{\infty} A_m$ be a covering of the domain of integration by cubes of side T which are disjoint except perhaps at boundary points and let $t_m \in A_m$ maximize $[n^{1/2}t]^{-(k+1)/2}u_0(n^{1/2}\varepsilon t)$ over A_m . From Lemma 3,

$$(2.12) \quad \int_{A_m} |\varphi|^n dt \leq e(T, \varphi)n^{-k/2}.$$

One obtains from (2.6) and (2.12)

$$(2.13) \quad \begin{aligned} I &\leq B_1e(T, \varphi) \sum_{m=1}^{\infty} (n^{1/2}|t_m|)^{-(k+1)/2}u_0(n^{1/2}\varepsilon t_m) \\ &= B_1e(T, \varphi)(n^{1/2}T)^{-k}\varepsilon^{-(k-1)/2} \sum_{m=1}^{\infty} W(s_m)\Delta A, \end{aligned}$$

where we have put $s_m = n^{1/2}\varepsilon t_m$, $W(s) = |s|^{-(k+1)/2}u_0(s)$ and $\Delta A = (n^{1/2}\varepsilon T)^k$. One easily verifies that on any of the cubes $K = n^{1/2}\varepsilon A_m$, $\max_K W(s) \leq c \min_K W(s)$. The sum in (2.13) may then be bounded by c times the integral of W , which is convergent for $k > 1$. \square

PROOF OF THEOREM 1. The function G_ε defined in (2.2) has Fourier transform

$$g_\varepsilon(t) = \hat{I}_G \varepsilon(t)u(\varepsilon t),$$

and by Lemma 2

$$(2.14) \quad |g_{\pm\varepsilon}(t)| \leq cB(t)u_0(\varepsilon t), \quad 0 < \varepsilon \leq \varepsilon_0.$$

Since g_ε is integrable it is immediate that

$$(2.15) \quad \int G_\varepsilon d(P_n - Q_n) = (2\pi)^{-k} \int g_\varepsilon(-t)(\varphi_n - \varphi_n^*) dt.$$

Put $T_0 = k^{-1}\beta_4^{-1}$. If $\|t\| < n^{1/2}T_0$, it follows (e.g., Esseen [7] page 99)

$$(2.16) \quad |\varphi_n(t) - \varphi_n^*(t)| \leq c\beta_4^{1/2}n^{-1}(|t|^4 + |t|^6) \exp(-|t|^2/3).$$

The integral on the right side of (2.15) is now bounded by means of (2.14), as the sum $I_1 + I_2 + I_3$ where

$$(2.17) \quad I_1 = \int_{\|t\| > n^{1/2}T_0} B(t)u_0(\varepsilon t)|\varphi_n - \varphi_n^*| dt \leq cB_2\beta_4^{1/2}n^{-1},$$

by (2.6), (2.16);

$$(2.18) \quad \begin{aligned} I_2 &= \int_{||t|| > n^{\frac{1}{2}} T_0} B(t) u_0(\varepsilon t) |\varphi_n^*| dt \\ &\leq c B_2 (1 + \beta_4^{\frac{2}{3}} n^{-\frac{1}{2}}) O((n^{\frac{1}{2}} T_0)^{-\infty}); \end{aligned}$$

$$(2.19) \quad \begin{aligned} I_3 &= \int_{||t|| > n^{\frac{1}{2}} T_0} B(t) u_0(\varepsilon t) |\varphi_n| dt \\ &\leq c B_1 (\beta_4^{\frac{2}{3}} n^{-\frac{1}{2}})^k \varepsilon^{-(k-1)/2}, \end{aligned}$$

from Lemmas 3 and 4. Incorporating these bounds with Lemma 1 and putting $\theta = \beta_4^{\frac{2}{3}} n^{-\frac{1}{2}}$, we obtain for $\varepsilon \leq \varepsilon_0$

$$(2.20) \quad |P_n(C) - Q_n(C)| < c \{A_0(\theta) + B_2 \theta^2 + B_2(1 + \theta) O(\theta^\infty)\},$$

where

$$A_0(\theta) = \min_{0 < \varepsilon \leq \varepsilon_0} \{B_1 \theta^k \varepsilon^{-(k-1)/2} + (1 + \theta) \varepsilon\}.$$

It is straightforward to show

$$(2.21) \quad A_0(\theta) \leq c \max \{B_1^{2/(k+1)} \theta^{2\gamma}, B_1 \theta^k \varepsilon_0^{-(k-1)/2}\}.$$

Now, $k > 1$ implies $1 < 2\gamma < 2$. Together (2.20) and (2.21) imply

$$(2.22) \quad |P_n(C) - Q_n(C)| < c B_1^{2/(k+1)} \theta^{2\gamma}, \quad \theta < 1.$$

However, the trivial bound $|P_n(C) - Q_n(C)| < c(1 + \theta)$ ensues from (2.4) so that for a possibly larger choice of c , (2.22) holds for all θ . The proof of theorem is completed by noting the uniformity of the bound in $C \in \mathcal{C}(M)$, n and β_4 . \square

The following makes more explicit how the constant in Theorem 1 relates to $\mathcal{C}(M)$. Let $K_0 = \inf\{K(x) : x \in \partial C, C \in \mathcal{C}(M)\}$.

COROLLARY. *There exists $c(k)$ such that*

$$\sup\{|P_n(C) - Q_n(C)| : C \in \mathcal{C}(M)\} \leq c(k) K_0^{-1/(k+1)} \beta_4^{\frac{2}{3}} n^{-\gamma},$$

whenever $n\beta_4^{-\frac{2}{3}} \rightarrow \infty$ and X_1 is normalized.

PROOF. The condition means that $n^{\frac{1}{2}} T_0 \rightarrow \infty$ in Lemma 4. Therefore, one needs the bound $B(t)$ only for $|t| \rightarrow \infty$. In this situation the effect of the error term in (2.5) becomes negligible and one may take $B_1 = K_0^{-\frac{1}{2}}$ in (2.6). The corollary follows from (2.22). \square

3. The nonlattice case. In this section we are dealing with a given nonlattice distribution; that is, $|\varphi(t)| < 1$ for all $t \neq 0$. This definition of nonlattice distribution appears to be due to Stone (1965). It should be pointed out that when $k > 1$ there exist distributions which are neither nonlattice in this sense, nor lattice.

In contrast to (2.7) define

$$(3.0) \quad e(T) = \limsup_{n \rightarrow \infty} \sup_{a \in R^k} n^{k/2} \int_{R_T + a} |\varphi|^n dt.$$

The principal technical result needed is

LEMMA 5. If X_1 has a normalized nonlattice distribution and $\beta_1 < \infty$, then

$$(3.1) \quad e(T) = o(T^k), \quad T \rightarrow \infty.$$

PROOF. There are three cases to consider.

Case i. X_1 has a continuous component. By considering $X_1 - X_2$ we may suppose without loss of generality that $\varphi = \lambda_1 \varphi_1 + \lambda_2 \varphi_2$ where $\lambda_1 > 0$, $\varphi \geq 0$ is the characteristic function of a continuous distribution, and $\lambda_1 + \lambda_2 = 1$, $|\varphi| \leq 1$. Then for any $\delta > 0$,

$$(3.2) \quad T^{-k} \int_{R_T+a} \varphi_1 dt < \delta, \quad T \rightarrow \infty,$$

uniformly in a .

Partition $R_T + a$ into cubes of side 1, say, and let k_T be the number of these cubes K for which $\max_{t \in K} \varphi_1(t) > 1 - \varepsilon$. If $\varphi_1(t_0) > 1 - \varepsilon$ for some $t_0 \in K$ then $\varphi_1(t) \geq (1 - \varepsilon - c|t - t_0|)^+$ where c bounds $|\text{grad } \varphi_1|$. In view of (3.2) a Chebyshev argument shows that $k_T = o(T^k)$, $T \rightarrow \infty$.

Bounding the integral of $|\varphi|^n$ on such cubes as in Lemma 3 and observing that on the remaining cubes $|\varphi| \leq 1 - \lambda_1 \varepsilon < 1$, one obtains

$$(3.3) \quad \int_{R_T+a} |\varphi|^n dt \leq k_T e(1, \varphi) n^{-k/2} + (1 - \lambda_1 \varepsilon)^n T^k,$$

uniformly in a . Multiplying by $n^{k/2}$ and letting $n \rightarrow \infty$ yields (3.1) in this case.

In the remaining cases X_1 is discrete nonlattice and, by symmetrizing, we may suppose its distribution is symmetric and has an atom at zero. Inasmuch as any nonsingular linear transformation preserves the nonlattice character, it suffices to suppose in addition that X_1 has atoms at $x^{(j)} = (x_1^{(j)}, \dots, x_k^{(j)})$, $j = 1, 2, \dots$, where

$$(3.4) \quad x_i^{(j)} = \delta_{ij}, \quad 1 \leq i \leq k, \quad 0 \leq j \leq k.$$

Such a reduction is possible because X_1 cannot be degenerate in a subspace. We then have

$$(3.5) \quad \varphi(t) = \sum_{j \geq 0} p_j \cos(x^{(j)} \cdot t); \quad p_0 > 0.$$

Case ii. All coordinates $x_i^{(j)}$ are rational. Put $\lambda_m = \sum_{j=0}^m p_j$ and define the characteristic functions

$$(3.6) \quad \varphi_m(t) = \sum_{j=0}^m p_j \cos(x^{(j)} \cdot t) / \lambda_m, \quad m = 1, 2, \dots;$$

(not to be confused with φ_n in Section 1).

For each m let $t_m \neq 0$ be a point with smallest sup norm for which $\varphi_m(t_m) = 1$. ($|\varphi_m(t)| = 1$ only when $\varphi_m(t) = 1$ since $p_0 > 0$.) It is impossible that $\{t_m\}_{m=1}^\infty$ has a limit point. For were there a limit point $t_* \neq 0$, then $\varphi(t_*) = 1$ since $\varphi_m \rightarrow \varphi$ uniformly, contradicting the nonlattice assumption. Nor could $t_* = 0$ be a limit point since then appropriate multiples of t_m would have a nonzero limit point.

For arbitrary $T > 0$ choose m such that $\|t_m\| > T + 2$. Let $\mathcal{L} = \{t: \varphi_m(t) = 1\}$

and

$$(3.7) \quad N(\mathcal{L}) = \{u \in R^k : \exists t \in \mathcal{L}, \|t - u\| < 1\}.$$

Because $\varphi_m(t_0 + t) = \varphi_m(t)$ whenever $t_0 \in \mathcal{L}$, it is clear that for this choice of m any set $R_T + a$ contains at most one element of \mathcal{L} . Moreover, in Case ii φ_m is periodic in each variable, so that $\sup\{|\varphi_m(t)| : t \notin N(\mathcal{L})\} = 1 - \varepsilon < 1$. In the same way as (3.3) was obtained

$$(3.8) \quad \int_{R_T+a} |\varphi|^n dt \leq 1 \cdot e(1, \varphi) n^{-k/2} + (1 - \lambda_m \varepsilon)^n T^k,$$

and again (3.1) follows.

Case iii. At least one coordinate, say $x_1^{(k+1)}$, is irrational.

For arbitrary $\varepsilon > 0$ cover R^k by cubes of side ε and to each cube let correspond that vertex with maximum coordinates. Then

$$(3.9) \quad \max_{1 \leq j \leq k+1} |x^{(j)} \cdot t - x^{(j)} \cdot t_0| \leq d\varepsilon$$

whenever $\|t - t_0\| < \varepsilon$, where $d = k^{\frac{1}{2}} \max_{1 \leq j \leq k+1} |x^{(j)}|$. For $x \in R$ let $\{x\} = x - [x]$ denote the fractional part. If K is one of the above cubes with designated vertex t_0 , an easy calculation shows there exists $\eta > 0$ such that

$$(3.10) \quad (d+1)\varepsilon < \{x^{(j)} \cdot t_0/2\pi\} < 1 - (d+1)\varepsilon, \quad \text{some } j = 1, 2, \dots, k+1$$

implies

$$(3.11) \quad \max_{t \in K} |\varphi(t)| < 1 - \eta.$$

Define the $k+1$ linear forms $L_j(m) = (\varepsilon/2\pi)x^{(j)} \cdot m$, $j = 1, 2, \dots, k+1$; $m \in Z^k$. Equivalent to (3.10) when the vertex is given by $t_0 = m\varepsilon$ is

$$(3.10)' \quad (d+1)\varepsilon < \{L_j(m)\} < 1 - (d+1)\varepsilon, \quad \text{some } j = 1, 2, \dots, k+1.$$

The following theorem of Weyl ([5] page 64) pertains to the asymptotic distribution of linear forms modulo one.

THEOREM. *Let $L_1(m), L_2(m), \dots$ be a finite set of linear forms in $m = (m_1, m_2, \dots, m_k) \in Z^k$ such that for integral u_j , $\sum_j u_j L_j(m)$ has all integer coefficients of m_1, m_2, \dots, m_k only when $u_1 = u_2 = \dots = 0$. Then the fractional parts $\{L_j\}$ are asymptotically uniformly and independently distributed in $(0, 1)$. This means that the $\{L_j\}$'s regarded as random variables defined on $\{1, 2, \dots, M\}^k$ with uniform probability measure, have a distribution which converges weakly to that stipulated as $M \rightarrow \infty$.*

A uniform version of Weyl's theorem is needed for the problem at hand. For $N = (N_1, N_2, \dots, N_k) \in Z^k$ put

$$S_j = \{n \in Z : N_j \leq n < N_j + M\}$$

and assign a uniform probability measure to $\times S_j$. With $\nu_{N,M}$ denoting the distribution induced by the $\{L_j\}$'s it can be established that as $M \rightarrow \infty$, the weak

convergence of $\nu_{N,M}$ is uniform in the "initial vector" N . (Essentially, the bounds on page 70 of [5] do not depend on N .) For a more recent discussion of these uniform distribution theorems one is referred to Chapter VIII of [6].

In Case iii $x_1^{(k+1)}$ is irrational and the condition in Weyl's theorem is easily verified provided only ε is chosen so that the numbers 1, $x_1^{(k+1)}$, and $2\pi/\varepsilon$ are linearly independent over the rationals. This we assume.

Suppose T and a are of the form $T = M\varepsilon$, $a = N\varepsilon \in R^k$. For any cube $K \subset R_T + a$ with designated vertex $t_0 = m_0\varepsilon$, either none of the inequalities in (3.10') obtains, in which case (3.11) applies, or else

$$m_0 \in B_{N,M} = \{(m_1, m_2, \dots, m_k) : \forall j = 1, 2, \dots, k+1, N_j \leq m_j < N_j + M \text{ and either } \{L_j(m)\} \leq (d+1)\varepsilon \text{ or } \geq 1 - (d+1)\varepsilon\}.$$

This set has cardinality $M^k \nu_{N,M}(B_{N,M})$. As before,

$$(3.12) \quad T^{-k} \int_{R_T+a} |\varphi|^n dt \leq \nu_{N,M}(B_{N,M}) e(\varepsilon, \varphi) (n^{\frac{1}{2}}\varepsilon)^{-k} + (1-\eta)^n.$$

Letting $n \rightarrow \infty$ and appealing to the above mentioned uniformity in Weyl's theorem,

$$(3.13) \quad \limsup_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_a T^{-k} n^{k/2} \int_{R_T+a} |\varphi|^n dt \leq [2(d+1)\varepsilon]^{k+1} \varepsilon^{-k} e(\varepsilon, \varphi).$$

Since ε is essentially arbitrary, the left side of (3.13) is zero and Case iii is concluded. \square

PROOF OF THEOREM 2. The proof of Theorem 1 goes through with modification. The distribution of X_1 is given so that β_4 is fixed. First, observe that for each $T > 0$ there exists an integer n_T such that

$$(3.14) \quad I'_3 = \int_{||t|| > n^{\frac{1}{2}}T} B(t) u_0(\varepsilon t) |\varphi_n| dt < c B_1 e(T) (n^{\frac{1}{2}}\varepsilon)^{-k} \varepsilon^{-(k-1)/2}, \quad n > n_T.$$

This is because in the proof of Lemma 4, and with regard to the definition of $e(T)$ in (3.0), the bound (2.12) may be replaced by $2e(T)$ as $n \rightarrow \infty$. The conclusion of Lemma 4 holds with this replacement.

The integral on the right side of (2.15) is bounded by dividing its region of integration into the regions $||t|| < n^{\frac{1}{2}}T_0$, $n^{\frac{1}{2}}T_0 \leq ||t|| < n^{\frac{1}{2}}T$ and $||t|| \geq n^{\frac{1}{2}}T_0$ where T_0 is the same as before. There results

$$(3.15) \quad |\int G_\varepsilon d(P_n - Q_n)| \leq I_1 + I_2 + I'_3 + I_4,$$

where I_1 and I_2 are still given by (2.17) and (2.18), I'_3 is given in (3.14) and

$$I_4 = \int_{n^{\frac{1}{2}}T_0 < ||t|| < n^{\frac{1}{2}}T} B(t) u_0(\varepsilon t) |\varphi_n| dt \leq B_2 n^{k/2} \int_{T_0 \leq ||t|| \leq T} |\varphi|^n dt.$$

Owing to the nonlattice hypothesis, for each T , $I_4 = O(n^{-\infty})$. Now replace ε by $\varepsilon_n = \varepsilon n^{-\gamma}$ in (3.15) and in Lemma 1. Incorporating the resulting bounds, we obtain

$$(3.16) \quad \sup\{|P_n(C) - Q_n(C)| : C \in \mathcal{C}(M)\} \leq c(k, M) e(T) T^{-k} \varepsilon^{-(k-1)/2} n^{-\gamma} + c \varepsilon n^{-\gamma} + O(n^{-1}).$$

Multiplying by n^r and letting in turn $n \rightarrow \infty$, $T \rightarrow \infty$ and $\varepsilon \rightarrow 0$, the right side of (3.16) becomes zero on account of Lemma 5. \square

4. Generalization to a class of dilations. Recall the definitions of $\mathcal{C}(M, \rho)$ and λC given in the introduction. For notational brevity we shall put $\theta = \beta_4^3 n^{-1}$, $\xi = \lambda^k \exp\{-\rho^2 \lambda^2/2\}$ and $l(\theta) = \max\{|\log \theta|, 1\}$, $\theta > 0$. In this section c will denote any constant depending at most on the class $\mathcal{C}(M, \rho)$; i.e., on k , M and ρ . There is no loss of generality in supposing ε_0 , defined in Section 2, satisfies $\varepsilon_0 < \min\{\rho/4, 1\}$.

LEMMA 6. *There exists $c(k, M, \rho)$ such that for all $C \in \mathcal{C}(M, \rho)$ and all normalized distributions,*

$$|P_n(\lambda C) - Q_n(\lambda C)| \leq c(k, M, \rho)G(\theta, \lambda), \quad 0 < \lambda < \infty; \theta < 1; n = 1, 2, \dots$$

where G is defined as follows; for $k \geq 3$:

$$(4.1) \quad G(\theta, \lambda) = \zeta^{(k-1)/(k+1)} \theta^{2r}, \quad \theta^{2/(k+1)} \lambda^{(k+1)/2} < \xi; \quad 1 < \lambda < \theta^{2-k}$$

$$(4.2) \quad = \theta^2 l^{(k-1/2)}(\theta), \quad e^{-2\rho\lambda\theta^2} \lambda^{(k+1)/2} < \xi < \theta^{2/(k-1)} \lambda^{(k+1)/2}; \quad 1 < \lambda < \theta^{2-k}$$

$$(4.3) \quad = \frac{\theta^2 l^k(\theta)}{l^{(k+1/2)}(\theta^2/\xi)}, \quad e^{-2\varepsilon_0\rho\lambda^2\theta^2}/\lambda < \xi < e^{-2\rho\lambda\theta^2} \lambda^{(k+1)/2}; \quad 1 < \lambda < \theta^{2-k}$$

$$(4.4) \quad = \theta^2/\lambda, \quad \xi < e^{-2\varepsilon_0\rho\lambda^2\theta^2}/\lambda; \quad 1 < \lambda < \theta^{2-k}$$

$$(4.5) \quad = \theta^k, \quad \lambda > \theta^{2-k}$$

$$(4.6) \quad = \max\{\lambda^{(k-1)r}\theta^{2r}, \theta^k\}, \quad \lambda < 1;$$

for $k = 2$ (4.1)–(4.4) are replaced by:

$$(4.1') \quad G(\theta, \lambda) = \xi^{1/3}\theta^3, \quad \theta^2\lambda^3 < \xi; \quad 1 < \lambda$$

$$(4.3') \quad \frac{\theta^2 l(\theta)}{l^{1/2}(\theta^2\lambda^3/\xi)}, \quad e^{-2\varepsilon_0\rho\lambda^2\theta^2} < \xi < \theta^2\lambda^3; \quad 1 < \lambda$$

$$(4.4') \quad = \theta^2, \quad \xi < e^{-2\varepsilon_0\rho\lambda^2\theta^2}; \quad 1 < \lambda.$$

REMARK. These various conditions on ξ could be given explicitly in terms of λ since $x^\alpha \exp\{-x^2\} = ca$, $a \rightarrow 0$ is solved by $x^2 = |\log a| + \alpha \log |\log a|$. We have not done so because Theorem 3 follows immediately from the lemma in the form given.

PROOF. The proof will be only outlined. One verifies $\hat{I}_{\lambda C}(t) = \lambda^k \hat{I}_C(\lambda t)$ so that

$$(4.7) \quad |\hat{I}_{\lambda(C^\varepsilon)}(t)| \leq B_\lambda(t) \equiv \lambda^k B(\lambda t) \\ \leq c \min\{\lambda^{(k-1)/2} |t|^{-(k+1)/2}, \lambda^k\}, \quad |\varepsilon| \leq \varepsilon_0.$$

Define the smoothings

$$G_{\varepsilon, \lambda} = I_{\lambda(C^\varepsilon)} * U_{|\varepsilon|}, \quad |\varepsilon| \leq \varepsilon_0.$$

From (2.4), $|dQ_n/dx| \leq c(1 + \theta|x|^3) \exp\{-|x|^2/2\}$. By considering the support function representation of a convex set one sees $(\lambda C)^{\lambda^\varepsilon} = \lambda(C^\varepsilon)$ so that after a

change of variable

$$(4.8) \quad J = \int_{(\lambda C)^{2\lambda\varepsilon - (\lambda C)^{-2\lambda\varepsilon}}} |dQ_n| \leq c(1 + \theta\lambda^3) \exp\{2\varepsilon\rho\lambda^2\}\xi\varepsilon.$$

Passing to the Fourier transform,

$$(4.9) \quad |P_n(\lambda C) - Q_n(\lambda C)| \leq c \int B_\lambda(t) u_0(\varepsilon\lambda t) |\varphi_n - \varphi_n^*| dt + J.$$

Now apply Lemma 4 to the integral on the right here to obtain

$$(4.10) \quad \int_{|t| > \theta^{-1}} B_\lambda(t) u_0(\varepsilon\lambda t) |\varphi_n| dt \leq c\theta^k \varepsilon^{-(k-1)/2}.$$

Because $|\varphi_n^*(t)| \leq c \exp\{-|t|^2/3\}$, $\theta < 1$, inequality (4.10) applies as well to φ_n^* .

The region of integration in (4.9) must be subdivided:

$$|t| \leq \lambda^{-1}, \lambda^{-1} < |t| \leq (\varepsilon\lambda)^{-1} \quad \text{and} \quad (\varepsilon\lambda)^{-1} < |t| \leq \theta^{-1}.$$

Bounding $|\varphi_n - \varphi_n^*|$, B_λ and u_0 appropriately in each subregion; integrating first over the spherical surface; and putting $dN = c \exp\{-r^2/3\} dr$ we have

$$(4.11) \quad \begin{aligned} & \int_{|t| < \theta^{-1}} B_\lambda(t) u_0(\varepsilon\lambda t) |\varphi_n - \varphi_n^*| dt \\ & \leq \theta^2 \lambda^k \int_0^{\lambda^{-1}} r^{k+3} dN + \theta^2 \lambda^{(k-1)/2} \int_{\lambda^{-1}}^{(\varepsilon\lambda)^{-1}} r^{(k+5)/2} dN \\ & \quad + (\theta^2/\lambda) \varepsilon^{-(k+1)/2} \int_{(\varepsilon\lambda)^{-1}}^{\theta^{-1}} r^2 dN. \end{aligned}$$

Suppose for the time being $\lambda > 1$. As $\varepsilon\lambda \rightarrow 0$ the middle term is of order $\theta^2 \lambda^{(k-1)/2}$ while the other two terms are of smaller order; as $\varepsilon\lambda \rightarrow \infty$, the third term is of order $\theta^2 \lambda^{-1} \varepsilon^{-(k+1)/2}$ while the other terms are smaller. It now follows from (4.9)–(4.11) and these remarks that for $\lambda > 1$

$$(4.12) \quad |P_n(\lambda C) - Q_n(\lambda C)| \leq cF(\theta, \lambda, \varepsilon), \quad \varepsilon \leq \varepsilon_0,$$

where

$$(4.13) \quad \begin{aligned} F(\theta, \lambda, \varepsilon) &= \theta^2 \lambda^{(k-1)/2} + \theta^k \varepsilon^{-(k-1)/2} \\ &\quad + (1 + \theta\lambda^3) \exp\{2\varepsilon\rho\lambda^2\}\xi\varepsilon, \quad \varepsilon < \varepsilon_1 \equiv \lambda^{-1} \end{aligned}$$

$$(4.14) \quad = \theta^2 \lambda^{-1} \varepsilon^{-(k+1)/2} + \theta^k \varepsilon^{-(k-1)/2} + \exp\{2\varepsilon\rho\lambda^2\}\xi\varepsilon, \quad \varepsilon > \varepsilon_1.$$

Now we seek the order of magnitude of $\min_{0 < \varepsilon < \varepsilon_0} F$. Only the case $k \geq 3$ is discussed. Suppose, first of all, $\xi > \exp\{-2\rho\lambda\}\theta^2 \lambda^{(k+1)/2}$, $\lambda > 1$. This forces $\lambda \leq c\lambda^k(\theta)$. Hence the term $\theta\lambda^3$ is bounded and may be ignored. Suppose as well $\lambda > \varepsilon_0^{-1}$ so that $\varepsilon_1 < \varepsilon_0$. One is led to minimize $\tilde{F}(\varepsilon) = \theta^k \varepsilon^{-(k-1)/2} + \exp\{2\varepsilon\rho\lambda^2\}\xi\varepsilon$, $0 < \varepsilon \leq \varepsilon_1$. By supposition the second term exceeds the first at the endpoint ε_1 so that up to a constant factor, the minimum is obtained by equating the terms. The solution ε_* is given by

$$\varepsilon_* \lambda^2 = c\omega[(\lambda^{k+1}\theta^k/\xi)^{2/(k+1)}],$$

where ω denotes the inverse of the map $x \mapsto xe^x$, $x > 0$. If the condition in (4.1) holds then, since $k \geq 3$, $\lambda^{k+1}\theta^k/\xi < c$ and since $\omega(a) \sim a$, $a \rightarrow 0$, $\tilde{F}(\varepsilon_*) \leq c\xi^{2/(k-1)/(k+1)}\theta^{2r}$. Under the same condition $\tilde{F}(\varepsilon_*)$ dominates the first term in (4.13) and is the bound given in this case. But under condition (4.2) the contrary

holds. Though unnecessary for the validity of (4.1) and (4.2), it can be shown that the effective minimum of F , $\varepsilon \leq \varepsilon_0$ occurs for $\varepsilon \leq \varepsilon_1$.

The condition (4.3) is treated in a similar way except that the minimum is sought in $(\varepsilon_1, \varepsilon_0)$. Cases (4.4) and (4.5) lead to endpoint minima and the bounds given are the larger of the first two terms in (4.14) up to a factor.

Finally, if $\lambda < 1$ then $\xi \leq c\lambda^k$ and reconsideration of (4.11) shows that (4.12) holds with F now defined as

$$F = \theta^2 \lambda^k + \theta^k \varepsilon^{-(k-1)/2} + \lambda^k \varepsilon, \quad \varepsilon \leq \varepsilon_0.$$

Minimizing this gives rise to the bound in (4.6). This completes our discussion of the proof. \square

REMARK. Inasmuch as $Q_n(R^k) = 0$, it is easily seen that $|Q_n(\lambda C)| = |Q_n(\lambda C)'| \leq c\xi/\lambda^2$. It follows that $P_n(\lambda C) \leq cG(\theta, \lambda)$ whenever $\xi/\lambda^2 \leq \theta^2 \lambda^{(k-1)/2}$.

PROOF OF THEOREM 3. It is sufficient to show that $G(\theta, \lambda) \leq c\theta^{2\gamma}$ for $0 < \lambda < \infty$, $\theta \leq 1$ where G is the function in Lemma 6. This is because there always is the bound $|P_n(\lambda C) - Q_n(\lambda C)| \leq c(1 + \theta) \leq c\theta^{2\gamma}$, $\theta \rightarrow \infty$. Now, ξ is bounded and $2\gamma < 2$ so that all cases but (4.3) are immediate. But condition (4.3) implies

$$l(\theta^2/\xi) \geq 2\rho\lambda \geq cl(\theta). \quad \square$$

Von Bahr (1967) has extended the result of Esseen for spheres. Let S_λ denote a centered sphere of radius λ . Assuming $\beta_k < \infty$ and with the supposition of a normalized distribution, he obtains the result

$$(4.15) \quad |P_n(S_\lambda) - \Phi(S_\lambda)| \leq c(\beta_k, k) \{ (1 + \lambda^{k+2}) \exp\{-\delta\lambda^2\} n^{-\gamma} \\ + n^{-1}(\log n)^{(k-1)/4} \}, \quad \lambda \leq (2.5 \log n)^{\frac{1}{2}}$$

where $\delta = \frac{1}{8}$ if $k = 2$ and $\delta = \frac{1}{2}(k-1)/(k+1)$, $k \geq 3$.

His bound is equivalent to our bound in case (4.2) but is worse in cases (4.1) and (4.3). In cases (4.1), (4.1'), for example, we have

$$G(\theta, \lambda) = \lambda^{(k-1)/(k+1) \cdot k} \exp\{-\delta\lambda^2\} \theta^{2\gamma}, \quad k \geq 3 \\ = \lambda^{\frac{1}{2}} \exp\{-\lambda^2/6\} \theta^{\frac{1}{2}}, \quad k = 2.$$

Von Bahr also states, with the same hypothesis

$$|P_n(S_\lambda) - \Phi(S_\lambda)| \leq c(\beta_k, k) d(n) \lambda^{-4} n^{-1}, \quad \lambda \geq (2.5 \log n)^{\frac{1}{2}},$$

where $d(n)$ is a sequence tending to zero. This appears to be an improvement over our bounds by the factors $d(n)\lambda^{-3}$ when $k \geq 3$ and case (4.4) obtains, and $d(n)\lambda^{-4}$ when $k = 2$. The contrast with the previous discussion is surprising. Apparently the second term in (4.11) does not arise in Von Bahr's estimation. He does not provide details of the contribution corresponding to (4.11) so that the discrepancy cannot be resolved.

5. Conclusions. In the introduction reference was made to Rao's expansion in the lattice case of $P_n(B)$ for any Borel set B . An interesting consequence of

Theorem 1 is that when $B \in \mathcal{C}(M)$ and $\beta_4 < \infty$, the additional term of apparent order $n^{-\frac{1}{2}}$ that arises in the lattice case is in fact of order n^{-r} . This was noted by Yarnold in his application of Rao's expansion to the chi-square statistic.

Even for centered spheres Theorem 3 is a slight strengthening of Esseen's theorem inasmuch as for a normalized distribution, $\beta_4 \geq 1$ and $\beta_4^{\frac{3}{2}r} < \beta_4^{\frac{3}{2}}$ except when $\beta_4 = 1$.

Theorem 1 will be true under the following weaker hypothesis: $\beta_s < \infty$ for some $s > 3 + (k - 1)/(k + 1)$. As Bhattacharya has shown, an expansion of $\varphi_n - \varphi_n^*$ analogous to that in (2.16) still is available where the constant appearing is replaced by $(\beta_s^{3/s} n^{-\frac{1}{2}})^{2r}$.

It would be interesting to generalize to sufficiently regular unbounded convex sets, or bounded ones which are not dilations of sets in $\mathcal{C}(M, \rho)$, and even to nonconvex sets. The difficulty that immediately arises is that the intersection of an unbounded convex set possessing a smooth boundary with a large centered sphere violates our smoothness assumptions. Another consideration for a bounded though large set is that one would prefer not to smooth I_G by convolution. Rather, the smoothing could be more at boundary points far away from the origin. For a smooth though nonconvex set, $K(x) = 0$ at some boundary points and the proofs given clearly break down. However, if $K(x) = 0$ only on a null set $\hat{I}_B(t)$ may still behave properly as $|t| \rightarrow \infty$ in which case the proofs would go through with minor modification.

REFERENCES

- [1] BERGSTROM, H. (1945). On the central limit theorem in the space R_k , $k > 1$. *Skand. Aktuarietidskr.* **28** 106-127.
- [2] BHATTACHARYA, R. N. (1968). Berry-Esséen bounds for the multi-dimensional central limit theorem. *Bull. Amer. Math. Soc.* **75** 285-287.
- [3] BHATTACHARYA, R. N. (1971). Rates of weak convergence and asymptotic expansions for classical central limit theorems. *Ann. Math. Statist.* **42** 241-259.
- [4] BONNESON, T. and FENCHEL, W. (1948). *Theorie der Konvexen Korper*. Chelsea, New York.
- [5] CASSELS, J. W. S. (1957). *An Introduction to Diophantine Approximation*. Cambridge.
- [6] CHANDRASEKHARAN, K. (1968). *Introduction to Analytic Number Theory*. Springer-Verlag, New York.
- [7] ESSEEN, C. G. (1945). Fourier analysis of distribution functions. *Acta. Math.* **77** 1-125.
- [8] HERZ, C. S. (1962). Fourier transforms related to convex sets. *Ann. of Math.* **75** 81-92.
- [9] RAO, R. R. (1961). On the central limit theorem in R_k . *Bull. Amer. Math. Soc.* **67** 359-361.
- [10] STONE, C. (1965). A local limit theorem for nonlattice multi-dimensional distribution functions. *Ann. Math. Statist.* **36** 546-551.
- [11] YARNOLD, J. K. (1972). Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set. *Ann. Math. Statist.* **43** 1566-1580.
- [12] VON BAHR, B. (1967). On the central limit theorem in R_k . *Ark. Mat.* **7** 61-69.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OREGON
EUGENE, OREGON 97403