The Annals of Statistics
1975, Vol. 3, No. 6, 1349–1353

A CENTRAL LIMIT THEOREM UNDER CONTIGUOUS ALTERNATIVES

BY K. BEHNEN AND G. NEUHAUS

University of Freiburg

In the statistical literature Le Cam's third lemma (cf. Hájek and Šidák (1967), page 208) is extensively used in order to get asymptotic normality of a statistic S_n under contiguous alternatives from asymptotic normality of S_n under the nullhypothesis. Since Le Cam's lemma utilizes the joint asymptotic normality of S_n and log-likelihood-ratio $\log L_n$, which is a sufficient but in general not a necessary condition for contiguity, it is not possible to get asymptotic normality of S_n for all contiguous alternatives from this lemma. On the other hand one is interested in the limiting distribution of S_n under all contiguous alternatives in order to get general power and efficiency results for the respective tests. In this paper we utilize a truncation method in order to prove asymptotic normality under all contiguous alternatives from asymptotic normality under the nullhypothesis for sums of independent random variables which are interesting in rank test theory, since they often are asymptotically equivalent to certain rank statistics under the nullhypothesis, and thus under contiguous alternatives, too.

For each $n \in \mathbb{N}$ let random variables Z_{ni}, $i = 1, \cdots, n$, be given which are independent under the nullhypothesis $H_0 : \mathfrak{L}(Z_{ni}) = P_{ni}$, $i = 1, \cdots, n$, as well as under the alternative $K : \mathfrak{L}(Z_{ni}) = Q_{ni}$, $i = 1, \cdots, n$, where the P_{ni} and Q_{ni} are probability measures on the real line such that the sequence $\{Q_n\}$ of product measures $Q_n = \times_{i=1}^n Q_{ni}$ is contiguous (in the sense of Hájek and Šidák (1967), page 202) to the sequence $\{P_n\}$ of product measures $P_n = \times_{i=1}^n P_{ni}$. (For simplicity reasons we omit $i = 1, \cdots, n, j = 1, \cdots, n$, and $n \to \infty$ in the sequel.) Moreover, we assume without loss of generality

$$\int z \, dP_{ni}(z) = 0 , \qquad \int z^2 \, dP_{ni}(z) = 1 .$$

Then we want to consider statistics of the form

$$S_n = \sum_i \sigma_{ni} Z_{ni} , \qquad \sigma_{ni} \geqq 0 , \qquad \sum_i \sigma_{ni}^2 = 1 , \tag{1}$$

and pose the question under what conditions asymptotic normality of S_n under H_0 implies asymptotic normality of S_n under K. Statistics of the form (1) play a central role in the theory of rank tests, since they often are asymptotically equivalent to certain rank statistics under H_0 and thus under K, too. The limiting distribution of S_n under K is needed for power examinations of the respective tests. Up to now the tool for deriving the limiting distributions under contiguous alternatives was Le Cam's third lemma (cf. Hájek and Šidák (1967),

Received October 1974; revised May 1975.

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

Key words and phrases. Asymptotic normality, contiguity, Le Cam's third lemma, central limit theorem, alternatives, rank tests.

page 208) which states that weak convergence of the pairs $(S_n, \log L_n)$ to a normal distribution $\mathfrak{N}(\mu_1, \mu_2, \tau_1^2, \tau_2^2, \tau_{12})$ with $\mu_2 = -\frac{1}{2}\tau_2^2$ under the nullhypothesis H_0 implies asymptotic normality of S_n to $\mathfrak{N}(\mu_1 + \tau_{12}, \tau_1^2)$ under the alternatives K, where L_n is the likelihood ratio dQ_n/dP_n. Clearly, a necessary condition for applying Le Cam's third lemma is the weak convergence of $\log L_n$ to a normal distribution $\mathfrak{N}(-\frac{1}{2}\tau_2^2, \tau_2^2)$ under H_0. This is a sufficient but in general not a necessary condition for contiguity of $\{Q_n\}$ to $\{P_n\}$. In cases where $\{Q_n\}$ is contiguous to $\{P_n\}$ but the above mentioned asymptotic normality of $\log L_n$ does not hold the described approach breaks down. Below we shall give an example where $\log L_n$ has nonnormal asymptotic distribution. In the sequel we shall describe a truncation technique for proving asymptotic normality of S_n which works for general contiguous alternatives. Moreover, our result is simpler to apply, even in cases where Le Cam's third lemma is applicable, since the relatively difficult proof of joint asymptotic normality of $(S_n, \log L_n)$ under P_n is superfluous. First in Behnen (1971), (1972) this truncation technique was used in special situations. Since the proof is omitted there, we shall formulate it here in a more general setting.

THEOREM. *Let the sequence $\{Q_n\}$ be contiguous to the sequence $\{P_n\}$ and let the constants σ_{ni} and the random variables Z_{ni} in $S_n = \sum_i \sigma_{ni} Z_{ni}$ fulfill the following conditions*

(i) $\max_i \sigma_{ni} \to 0$, $\sum_i \sigma_{ni}^2 = 1$, $\sigma_{ni} \geqq 0$, $i = 1, \cdots, n$,

(ii) $\max_i \int_{\{|z|>M_n\}} z^2\, dP_{ni}(z) \to 0$ *for all sequences $\{M_n\}$ of positive numbers tending to infinity.*

Then S_n under P_n converges in distribution to a standard normal distribution:

$$\mathfrak{L}(S_n \mid P_n) \to \mathfrak{N}(0, 1)\,, \tag{2}$$

and there exists a sequence $\{a_n\}$ of real numbers such that under Q_n

$$\mathfrak{L}(S_n - a_n \mid Q_n) \to \mathfrak{N}(0, 1) \tag{3}$$

holds true.

REMARK 1. From the assumptions (i) and (ii) we can show that, for each n and i there exist measurable, real-valued, bounded functions h_{ni} such that

$$\int h_{ni}\, dP_{ni} = 0\,, \tag{4}$$

$$\max_i \int (h_{ni}(z) - z)^2\, dP_{ni}(z) \to 0\,, \tag{5}$$

$$\max_i \sigma_{ni} \max_j \sup_z h_{nj}^2(z) \to 0\,. \tag{6}$$

Take for example

$$h_{ni} = h'_{ni} - \int h'_{ni}\, dP_{ni}\,,$$

$$h'_{ni}(z) = z \quad \text{if } |z| < (\max_j \sigma_{nj})^{-\frac{1}{4}}\,,$$
$$= 0 \quad \text{elsewhere.}$$

The proof of the theorem shows that the sequence $\{a_n\}$ may have the form

$$a_n = \sum_i \sigma_{ni} \int h_{ni}\, dQ_{ni}\,. \tag{7}$$

REMARK 2. Any sequence $\{a_n\}$ of the theorem is bounded, since convergence of a subsequence $\{a_{n'}\}$ to infinity would imply either

$$P_{n'}(S_{n'} < a_{n'}) \to 1\,, \qquad Q_{n'}(S_{n'} < a_{n'}) \to \tfrac{1}{2}\,,$$

or

$$P_{n'}(S_{n'} < a_{n'}) \to 0\,, \qquad Q_{n'}(S_{n'} < a_{n'}) \to \tfrac{1}{2}\,,$$

which contradicts, in any case, the contiguity of $\{Q_n\}$ to $\{P_n\}$.

Before proving the theorem we give a simple example, where Le Cam's third lemma is not applicable whereas our theorem applies. It should be mentioned that it is easy to construct trivial examples by mixing different contiguous sequences such that $\log L_n$ has not a limiting distribution.

EXAMPLE. For each $n \in \mathbb{N}$ and $i = 1, \cdots, n$ let P_{ni} and Q_{ni} be normal distributions, $P_{ni} = \mathfrak{N}(0, 1)$, $Q_{ni} = \mathfrak{N}(0, \tau_i^2)$, with $\tau_i^2 = \exp(2^{-i}) > 1$. Then we have

$$-\log L_n(z_1, \cdots, z_n) = \left(\frac{1}{2} - \frac{1}{2^{n+1}}\right) + \tfrac{1}{2} \sum_{i=1}^n z_i^2(\tau_i^{-2} - 1)\,.$$

Therefore the monotone convergence theorem implies the existence of a limiting distribution of L_n under P_n with expectation 1, and Le Cam's first lemma (cf. Hájek and Šidák (1967), page 203) implies the contiguity of $\{Q_n\}$ to $\{P_n\}$. Since condition (ii) is obviously fulfilled, the theorem is applicable whereas Le Cam's third lemma cannot work since the limiting distribution of $\log L_n$ is apparently a nonnormal distribution.

PROOF. First we prove an auxiliary result on contiguity which may be interesting in its own right.

PROPOSITION. *For each $n \in \mathbb{N}$ let P_n and Q_n be probability measures on some measurable space $(\Omega_n, \mathfrak{A}_n)$. Then contiguity of $\{Q_n\}$ to $\{P_n\}$ implies*

$$\limsup_{n\to\infty} \|P_n - Q_n\| < 1\,. \tag{8}$$

Moreover, in the special situation of product measures $P_n = \times_{i=1}^n P_{ni}$ and $Q = \times_{i=1}^n Q_{ni}$, the inequality (8) *implies*

$$\limsup_{n\to\infty} \sum_{i=1}^n \|P_{ni} - Q_{ni}\|^2 < \infty\,. \tag{9}$$

Here the distance $\|P - Q\|$ of P and Q defined on $(\Omega, \mathfrak{A})$ *is given by* $\|P - Q\| = \sup\{|P(A) - Q(A)| : A \in \mathfrak{A}\}$.

PROOF OF THE PROPOSITION. Suppose that (8) is not true. Then, with the notation $\mu_n = P_n + Q_n$, $p_n = dP_n/d\mu_n$, $q_n = dQ_n/d\mu_n$, $B_n = \{q_n > p_n\}$, there exists a subsequence $\{n'\}$ of $\mathbb{N}$ such that

$$Q_{n'}(B_{n'}) - P_{n'}(B_{n'}) = \|P_{n'} - Q_{n'}\| \to 1\,.$$

This entails $P_{n'}(B_{n'}) \to 0$ and $Q_{n'}(B_{n'}) \to 1$ which contradicts the contiguity assumption. Hence (8) holds.

For proving (9) we use the additional notation $\mu_{ni} = P_{ni} + Q_{ni}$, $p_{ni} = dP_{ni}/d\mu_{ni}$, $q_{ni} = dQ_{ni}/d\mu_{ni}$, out, s.c. and $\mu_n = \times_i \mu_{ni}$. Then we get following chain of inequalities.

$$\begin{aligned}1 - \|P_n - Q_n\| &= 1 - \tfrac{1}{2} \int |p_n - q_n|\, d\mu_n = \int (p_n \wedge q_n)\, d\mu_n \leqq \int (p_n q_n)^{\frac{1}{2}}\, d\mu_n \\ &= \prod_i \int (p_{ni} q_{ni})^{\frac{1}{2}}\, d\mu_{ni} \leqq \prod_i \left(\int p_{ni} \wedge q_{ni})\, d\mu_{ni} \int (p_{ni} \vee q_{ni})\, d\mu_{ni}\right)^{\frac{1}{2}} \\ &= \prod_i ((1 - \|P_{ni} - Q_{ni}\|)(1 + \|P_{ni} - Q_{ni}\|))^{\frac{1}{2}} \\ &= \prod_i (1 - \|P_{ni} - Q_{ni}\|^2)^{\frac{1}{2}} \leqq \exp(-(\tfrac{1}{2}) \textstyle\sum_i \|P_{ni} - Q_{ni}\|^2)\,.\end{aligned}$$

Thus it is obvious that (8) implies (9).

PROOF OF THE THEOREM. Given $\varepsilon > 0$, let us define $M_n = \varepsilon/\max_i \sigma_{ni}$. Then the right side of the obvious (cf. assumption (i)) inequality

$$\sum_i \sigma_{ni}^2 \int_{\{\sigma_{ni}|z| > \varepsilon\}} z^2\, dP_{ni}(z) \leqq \max_i \int_{\{|z| > M_n\}} z^2\, dP_{ni}(z)$$

tends to zero according to (i) and (ii). Therefore, Lindeberg's condition for $\{S_n\}$ is fulfilled, and (2) is proved. In order to prove (3) let h_{ni} be functions which fulfill the conditions (4), (5), and (6) stated in Remark 1, and define

$$\bar{S}_n = \sum_i \sigma_{ni} h_{ni}(Z_{ni})\,.$$

Then, from (4), (5), and (i), we have

$$\int (\bar{S}_n - S_n)^2\, dP_n = \sum_i \sigma_{ni}^2 \int (h_{ni} - z)^2\, dP_{ni} \leqq \max_i \int (h_{ni} - z)^2\, dP_{ni} \to 0\,.$$

This, together with the contiguity assumption, implies

$$\bar{S}_n - S_n \to 0 \qquad \text{in probability under} \quad \{Q_n\}\,.$$

Thus, it suffices to prove (3) with $\bar{S}_n$ instead of S_n, which can be done by Lindeberg's theorem: In a first step we get

$$\rho_n^2 = \sum_i a_{ni}^2 \to 0 \qquad \text{if} \quad a_{ni} = \sigma_{ni} \int h_{ni}\, dQ_{ni}\,, \tag{10}$$

because of (4), (i), (6), and (9), which imply

$$\begin{aligned}\sum_i a_{ni}^2 &= \sum_i \sigma_{ni}^2 \left(\int h_{ni}\, dQ_{ni}\right)^2 = \sum_i \sigma_{ni}^2 \left(\int h_{ni}(q_{ni} - p_{ni})\, d\mu_{ni}\right)^2 \\ &\leqq \max_i \sigma_{ni}^2 \max_j \sup_z h_{nj}^2(z) \sum_i 4\|Q_{ni} - P_{ni}\|^2 = o(1)\,.\end{aligned}$$

In a second step we get

$$\tau_n^2 = \sum_i \tau_{ni}^2 \to 1 \qquad \text{if} \quad \tau_{ni}^2 = \int (\sigma_{ni} h_{ni} - a_{ni})^2\, dQ_{ni}\,, \tag{11}$$

because of (10), (5), (6), (i), and (9), which imply

$$\begin{aligned}|1 - \tau_n^2| &= |\textstyle\sum_i (\sigma_{ni}^2 - \tau_{ni}^2)| \leqq \sum_i \sigma_{ni}^2 \left|\int z^2\, dP_{ni}(z) - \int h_{ni}^2\, dQ_{ni}\right| + \rho_n^2 \\ &\leqq \max_i \left|\int (z^2 - h_{ni}^2)\, dP_{ni}\right| + \textstyle\sum_i \sigma_{ni}^2 \left|\int h_{ni}^2 (p_{ni} - q_{ni})\, d\mu_{ni}\right| + o(1) \\ &\leqq o(1) + \left(\textstyle\sum_j \sigma_{nj}^4 \sum_i \left(\int h_{ni}^2 |p_{ni} - q_{ni}|\, d\mu_{ni}\right)^2\right)^{\frac{1}{2}} \\ &\leqq o(1) + \max_i \sigma_{ni} \max_j \sup_z h_{nj}^2(z) \left(\textstyle\sum_i \sigma_{ni}^2 \sum_j 4\|P_{nj} - Q_{nj}\|^2\right)^{\frac{1}{2}} = o(1)\,.\end{aligned}$$

Now, in a last step, we show that Lindeberg's condition holds true for $\{\bar{S}_n - \sum_i a_{ni}\}$. Let $\varepsilon > 0$ be given. Then, if n is sufficiently large, we have

$$\frac{1}{\tau_n^2} \sum_i \int_{\{z : |\sigma_{ni} h_{ni}(z) - a_{ni}| > \varepsilon \tau_n\}} (\sigma_{ni} h_{ni} - a_{ni})^2 \, dQ_{ni} = 0 ,$$

because of (i), (6), (10), (11), and

$$|\sigma_{ni} h_{ni}(z) - a_{ni}| \leqq \max_i \sigma_{ni} \max_j \sup_z |h_{nj}(z)| + \max_i |a_{ni}| .$$

REMARK 3. The truncation technique of this proof can be applied, with suitable modifications, in many situations where sums of independent random variables are considered under the nullhypothesis as well as under contiguous alternatives. An example where sums of independent Hilbert-space random variables are treated in this way is given in Neuhaus (1973).

REMARK 4. At the Oberwolfach meeting in November, 1974, Professor L. Le Cam pointed out that the asymptotic normality of S_n under Q_n can be proved by the general version of Le Cam's third lemma (cf. Le Cam (1960)) and the representation of infinitely divisible distributions. But our proof is more elementary and contains the form of the centering constants a_n, which is useful for asymptotic power investigations.

REFERENCES

[1] BEHNEN, K. (1971). Asymptotic optimality and ARE of certain rank-order tests under contiguity. *Ann. Math. Statist.* **42** 325–329.

[2] BEHNEN, K. (1972). A characterization of certain rank-order tests with bounds for the asymptotic relative efficiency. *Ann. Math. Statist.* **43** 1839–1851.

[3] HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

[4] LE CAM, L. (1960). Locally asymptotically normal families of distributions. Univ. of Calif. Publ. Statist. **3** 37–98.

[5] NEUHAUS, G. (1974). Asymptotic properties of the Cramér-von Mises statistic when parameters are estimated. *Proc. Prague Symp. on Asymptotic Statist.* Sept. 3–6, 1973 (J. Hájek, ed.). Universita Karlova Praha. **2** 257–297

STUDIENBEREICH MATHEMATIK
UNIVERSITÄT BREMEN
28 BREMEN 33
POSTFACH
GERMANY

MATHEMATISCHES INSTITUT
UNIVERSITÄT GIESSEN
63 GIESSEN
ARNDTSTRASSE 2
GERMANY