The Annals of Probability
1988, Vol. 16, No. 1, 305–312

NORMAL CONVERGENCE BY HIGHER SEMIINVARIANTS WITH APPLICATIONS TO SUMS OF DEPENDENT RANDOM VARIABLES AND RANDOM GRAPHS

BY SVANTE JANSON

Uppsala University

If the means and variances of a sequence of random variables converge, and all semiinvariants (cumulants) of sufficiently high order tend to zero, then the variables converge in distribution to a normal distribution. Thus no information is needed on the remaining (finitely many) semiinvariants. This is applied to give a new criterion for asymptotic normality of sums of dependent variables. An example is included where this criterion is applied to the number of induced subgraphs of a particular type in a random graph.

1. Introduction and main results. Marcinkiewicz (1939, Théorème 2^{bis}) proved the following theorem:

Suppose that a random variable X has a characteristic function of the form $\exp(p(t))$ where p is a polynomial. Then $\deg(p(t)) \le 2$, i.e., X has a normal distribution.

(We regard a degenerate distribution as normal.)

An equivalent formulation using semiinvariants (see Section 2, in particular Lemma 2) is:

If the semiinvariants $\kappa_j(X)$ vanish for all sufficiently large j, then X has a normal distribution.

In Section 3 we will use this uniqueness result to prove the corresponding convergence result:

If the semiinvariants $\kappa_j(X)$ are close to zero for all sufficiently large j, then the distribution of X is close to a normal distribution.

More formally:

THEOREM 1. *Let $X_1, X_2, \dots$ be a sequence of random variables such that, as $n \to \infty$,*

$$\kappa_1(X_n) = EX_n \to \mu, \tag{1.1}$$

$$\kappa_2(X_n) = \operatorname{var}(X_n) \to \sigma^2, \tag{1.2}$$

$$\kappa_j(X_n) \to 0 \quad \textit{for every } j \ge m, \tag{1.3}$$

where $-\infty < \mu < \infty$, $\sigma^2 \ge 0$, and $m \ge 3$. Then

$$X_n \to_d N(\mu, \sigma^2) \quad \textit{as } n \to \infty. \tag{1.4}$$

Received November 1985; revised March 1986.
AMS 1980 *subject classifications.* Primary 60F05, 05C80.
Key words and phrases. Convergence in distribution, central limit theorem, semiinvariants, cumulants, method of moments, random graphs.

Furthermore, all moments of X_n converge to the corresponding moments of $N(\mu, \sigma^2)$.

Here (and in Lemma 1 below) it is not necessary that all variables X_n have moments of all orders, only that for every j, $E|X_n|^j < \infty$ (and thus $\kappa_j(X_n)$ is defined) for $n \geq n_j$.

The theorem may alternatively be formulated as an approximation result.

COROLLARY 1. *If $\varepsilon > 0$ and $m \geq 3$, then there exist $\delta > 0$ and $M < \infty$ such that $EX = 0$, $\operatorname{var} X = 1$, and $|\kappa_j(X)| < \delta$ for $m \leq j \leq M$ imply*

$$\sup_x |P(X \leq x) - \Phi(x)| < \varepsilon.$$

REMARK 1. Our methods give no information on the rate of convergence.

REMARK 2. A symmetrization and Cramér's decomposition theorem (Cramér (1937, Théorème 19)) show that it suffices in Marcinkiewicz's theorem that the semiinvariants vanish for all sufficiently large *even* j. Similarly, it is sufficient to consider *even* j in Theorem 1 and Corollary 1.

When $m = 3$, (1.1)–(1.3) say that all semiinvariants of X_n converge to the semiinvariants of $N(\mu, \sigma^2)$. This is obviously equivalent to the convergence of all moments, hence Theorem 1 is, for $m = 3$, only a reformulation of the method of moments. When $m > 3$, the situation is different. The conclusion of the theorem implies that, in fact, $\kappa_j(X_n) \to 0$ for every $j \geq 3$, and thus the theorem applies only to the same sequences $\{X_n\}$ as the method of moments. However, the point of the theorem is that it is not necessary to actually check that $\kappa_j(X_n) \to 0$ for small j, and that is sometimes a considerable simplification.

This is illustrated in Section 4, where Theorem 1 is used to prove the following central limit theorem for sums of dependent variables. Before stating the theorem, we introduce a convenient terminology.

DEFINITION. A graph Γ is a dependency graph for a family of random variables if the following two conditions are satisfied:

(i) There exists a one-to-one correspondence between the random variables and the vertices of the graph.

(ii) If V_1 and V_2 are two disjoint sets of vertices of Γ such that no edge of Γ has one endpoint in V_1 and the other in V_2, then the corresponding sets of random variables are independent.

Note that this definition does not define a unique dependency graph for every family of random variables. For example, we may add any edge to a dependency graph and obtain a new one (for the same family of random variables).

Recall that the maximal degree of a graph is the maximal number of edges incident to a single vertex.

THEOREM 2. *Suppose that, for each n, $\{X_{ni}\}_{i=1}^{N_n}$ is a family of bounded random variables; $|X_{ni}| \le A_n$ a.s. Suppose further that Γ_n is a dependency graph for this family and let M_n be the maximal degree of Γ_n (unless Γ_n has no edges at all, in which case we set $M_n = 1$).*

Let $S_n = \sum_1^{N_n} X_{ni}$ and $\sigma_n^2 = \operatorname{var}(S_n)$. If there exists an integer m such that

$$(N_n/M_n)^{1/m} M_n A_n/\sigma_n \to 0 \quad \text{as } n \to \infty, \tag{1.5}$$

then

$$(S_n - ES_n)/\sigma_n \to_d N(0,1) \quad \text{as } n \to \infty. \tag{1.6}$$

REMARK 3. By a standard truncation argument, it is possible to extend the theorem to unbounded variables. We may, e.g., replace the condition $|X_{ni}| \le A_n$ a.s. by $M_n \sum_i E(X_{ni}^2 I(|X_{ni}| > A_n))/\sigma_n^2 \to 0$ as $n \to \infty$ (for some sequence A_n satisfying (1.5)).

Since $N_n/M_n \ge 1$, the condition (1.5) gets weaker as m increases. It follows from the proof that necessarily $m \ge 3$.

The case $m = 3$ can be proved by the method of moments without recourse to Theorem 1 and Marcinkiewicz's theorem, but these results are an important ingredient of the proof when $m \ge 4$.

Section 5 furnishes an application to random graphs of the latter case ($m \ge 4$). Theorem 2 is used to show the asymptotic normality of a subgraph count statistic also in a degenerate case.

Many theorems on asymptotic normality for sums have been proved by the method of moments using, at least implicitly, dependency graphs and estimates of the type used in the proofs below (for $m = 3$). Some such theorems are in fact special cases of Theorem 2. For example, Noether (1970) studied the following situation, which includes many test statistics: (We change his notation.) Let $S_n = \sum_{i \in I_n, j \in J_n} X_{nij}$, where X_{nij} are uniformly bounded random variables such that we obtain a dependency graph for $\{X_{nij}\}_{i \in I_n, j \in J_n}$ by joining every pair of pairs (i, j) and (k, l) that contains a common subscript. Assume that $\#I_n \le CK_n$ and $\#J_n \le CK_n$ where $K_n \to \infty$ as $n \to \infty$.

Noether proved that S_n then is asymptotically normally distributed, provided $\operatorname{var} S_n \ge cK_n^3$ for some $c > 0$. Here

$$N_n = \#I_n \#J_n \le C_1 K_n^2, \qquad M_n \le C_2 K_n, \qquad A_n = C_3,$$

and thus

$$(N_n/M_n)^{1/m} M_n A_n/\sigma_n \le C_4 K_n^{1/m+1} (\operatorname{var} S_n)^{-1/2}.$$

Consequently, Noether's theorem follows from Theorem 2; in fact, the somewhat weaker condition $\operatorname{var} S_n \ge cK_n^{2+\varepsilon}$ for some $\varepsilon > 0$ is sufficient.

Chen (1978) has used a different method to prove asymptotic normality in a general situation under conditions related to our use of dependency graphs. His Theorem 4.2 contains, e.g., Noether's theorem, but neither the improved version thereof given above nor the example in Section 5.

2. Semiinvariants. For convenience, we give here the definition and some elementary properties of semiinvariants (cumulants); see, e.g., Cramér (1945). Assume that X is a random variable with $E|X|^j < \infty$ for some $j \geq 1$, i.e., $X \in L^j$. Then the characteristic function φ is j times continuously differentiable. Hence $\log \varphi(t)$ is a continuous, j times differentiable function in some open interval containing 0. We define the jth semiinvariant κ_j (or $\kappa_j(X)$ to be more precise) by

$$\kappa_j = (-i)^j \frac{d^j}{dt^j} \log \varphi(0). \tag{2.1}$$

($\kappa_j(X)$ is not defined if $X \notin L^j$.) For example, if $X \sim N(\mu, \sigma^2)$, then $\kappa_1 = \mu$, $\kappa_2 = \sigma^2$, and $\kappa_j = 0$, $j \geq 3$. Recall that the jth moment $\alpha_j(X) = EX^j = (-i)^j d^j/dt^j \varphi(0)$. It follows that κ_j is a polynomial function of $\alpha_1, \dots, \alpha_j$, and conversely, α_j is a polynomial function of $\kappa_1, \dots, \kappa_j$. In particular, $\kappa_1 = \alpha_1 = EX$ and $\kappa_2 = \alpha_2 - \alpha_1^2 = \operatorname{var}(X)$. We have the following semiinvariant version of the methods of moments.

LEMMA 1. *Suppose that X_n is a sequence of random variables such that for every $j \geq 1$, $\kappa_j(X_n) \to c_j$ as $n \to \infty$, where c_j are some real numbers. Then there exists a random variable X with $\kappa_j(X) = c_j$, $j \geq 1$, and, if this determines the distribution of X uniquely, then $X_n \to_d X$ and $\alpha_j(X_n) \to \alpha_j(X)$ for every j.*

PROOF. $\alpha_j(X_n) = p_j(\kappa_1(X_n), \dots, \kappa_j(X_n)) \to p_j(c_1, \dots, c_j)$ for some polynomial p_j. Hence $\{X_n\}$ is tight, and every convergent subsequence converges (in distribution) to a limit X with $\alpha_j(X) = p_j(c_1, \dots, c_j)$ for every j, and thus $\kappa_j(X) = c_j$. □

The sequence of semiinvariants determines the distribution uniquely iff the sequence of moments does so. The following lemma gives one criterion.

LEMMA 2. *Suppose that $\sum_1^\infty (|\kappa_j|/j!) r^j < \infty$ for some $r > 0$. Then the distribution of X is uniquely determined by $\{\kappa_j\}_1^\infty$, and the characteristic function*

$$\varphi(t) = \exp\left(\sum_1^\infty \frac{\kappa_j}{j!}(it)^j\right) \quad \textit{for } |t| < r.$$

PROOF. Let $F(t) = \exp(\sum_1^\infty \kappa_j/j!(it)^j)$. $F(t)$ is analytic for $|t| < r$. By definition, $\log F(t)$ and $\log \varphi(t)$ have the same derivatives of all orders at $t = 0$, whence $F(t)$ and $\varphi(t)$ have the same derivatives of all orders at $t = 0$. Hence $\sum \alpha_j/j!(it)^j$ converges for $|t| < r$, which implies that φ is analytic. Consequently, $\varphi(t) = F(t)$ for $|t| < r$, and φ is uniquely determined by $\{\kappa_j\}$. □

For future reference we observe two simple consequences of the definition (2.1). If X and Y are independent and a is a real number, then

$$\kappa_j(X + Y) = \kappa_j(X) + \kappa_j(Y), \tag{2.2}$$

$$\kappa_j(aX) = a^j \kappa_j(X). \tag{2.3}$$

3. Proof of Theorem 1. Suppose that (1.1)–(1.3) holds and define (for n large enough)

$$\varepsilon_n = \sup_{3 \le j < m} |\kappa_j(X_n)|^{1/j}. \tag{3.1}$$

If $m = 3$ or if $\varepsilon_n \to 0$ as $n \to \infty$, then $\kappa_j(X_n) \to 0$ as $n \to \infty$ for every $j \ge 3$, which together with (1.1) and (1.2) yield the conclusions by Lemmas 1 and 2.

We complete the proof by showing that the assumptions $m > 3$ and $\varepsilon_n \not\to 0$ lead to a contradiction. Restricting attention to a subsequence, we may assume that $\varepsilon_n > \delta$ for every n and some $\delta > 0$. Let $Y_n \sim N(0, (1 - (\delta/\varepsilon_n)^2)\mathrm{var}(X_n))$ be independent of X_n and set

$$Z_n = \frac{\delta}{\varepsilon_n}(X_n - EX_n) + Y_n. \tag{3.2}$$

Then, by (2.2) and (2.3),

$$\kappa_1(Z_n) = 0, \tag{3.3}$$

$$\kappa_2(Z_n) = \left(\frac{\delta}{\varepsilon_n}\right)^2 \kappa_2(X_n) + \kappa_2(Y_n) = \mathrm{var}(X_n) \to \sigma^2, \tag{3.4}$$

$$\kappa_j(Z_n) = \left(\frac{\delta}{\varepsilon_n}\right)^j \kappa_j(X_n) + \kappa_j(Y_n) = \left(\frac{\delta}{\varepsilon_n}\right)^j \kappa_j(X_n), \qquad j \ge 3. \tag{3.5}$$

Since $0 < \delta/\varepsilon_n < 1$, (3.5) and (1.3) give $\kappa_j(Z_n) \to 0$ as $n \to \infty$ for $j \ge m$. Furthermore, (3.5) and (3.1) give

$$\sup_{3 \le j < m} |\kappa_j(Z_n)|^{1/j} = \delta. \tag{3.6}$$

In particular, $|\kappa_j(Z_n)| \le \delta^j$, $3 \le j < m$. Hence we may select a subsequence of $\{Z_n\}$ such that, along this subsequence, $\kappa_j(Z_n) \to c_j$ for some real numbers c_j, $3 \le j < m$. Defining $c_1 = 0$, $c_2 = \sigma^2$, and $c_j = 0$, $j \ge m$, we see that $\kappa_j(Z_n) \to c_j$ for every $j \ge 1$ as $n \to \infty$ along the subsequence. By Lemma 1, there exists a random variable X with $\kappa_j(X) = c_j$, but this contradicts (the second version of) Marcinkiewicz's theorem given at the beginning of the introduction, since $c_j = 0$, $j \ge m$, but, by (3.6), $\sup_{3 \le j < m} |c_j|^{1/j} = \delta > 0$. □

PROOF OF COROLLARY 1. Suppose that for some ε and m no such δ and M exist. Then, choosing $\delta = 1/n$ and $M = n$, we may construct a sequence $\{X_n\}$ that contradicts Theorem 1. □

4. Sums of dependent variables. We introduce mixed semiinvariants; see, e.g., Leonov and Shiryaev (1959). If $X_1, \dots, X_j \in L^j$, then their joint characteristic function $\varphi_{X_1, \dots, X_j}(t_1, \dots, t_j)$ is j times continuously differentiable and we define

$$\kappa(X_1,\dots,X_j) = (-i)^j \frac{\partial^j}{\partial t_1 \cdots \partial t_j} \log \varphi_{X_1,\dots,X_j}(0,\dots,0). \tag{4.1}$$

It is easily seen that $\kappa(X_1,\dots,X_j)$ is a multilinear function of $X_1,\dots,X_j$, and that it equals $\kappa_j(X)$ when $X_1 = X_2 = \cdots = X_j = X$. Hence, if $S = \sum_1^N X_i$, then

$$\kappa_j(S) = \kappa(S,\dots,S) = \sum_{i_1=1}^{N} \cdots \sum_{i_j=1}^{N} \kappa(X_{i_1},\dots,X_{i_j}). \tag{4.2}$$

LEMMA 3. *Suppose that $\{X_1,\dots,X_j\}$ can be divided into two independent nonempty sets of random variables. Then $\kappa(X_1,\dots,X_j) = 0$.*

PROOF. Suppose, e.g., that $X_1,\dots,X_k$ are independent of $X_{k+1},\dots,X_j$, with $1 \le k < j$. Then

$$\begin{aligned} &\log \varphi_{X_1,\dots,X_j}(t_1,\dots,t_j) \\ &\qquad = \log \varphi_{X_1,\dots,X_k}(t_1,\dots,t_k) + \log \varphi_{X_{k+1},\dots,X_j}(t_{k+1},\dots,t_j) \end{aligned} \tag{4.3}$$

and it is clear that the mixed derivative in (4.1) vanishes. □

Furthermore, it follows from (4.1) that $\kappa(X_1,\dots,X_j)$ may be written as a linear combination of terms $\prod_k E \prod_{i \in I_k} X_i$, where $I_1,\dots,I_l$ is a partition of $\{1,\dots,j\}$. (See Leonov and Shiryaev (1959, (II.c)) for the exact formula.) This expansion and Hölder's inequality show that there exist some universal constants C_j such that

$$|\kappa(X_1,\dots,X_j)| \le C_j \|X_1\|_j \cdots \|X_j\|_j. \tag{4.4}$$

LEMMA 4. *Suppose that Γ is a dependency graph for $\{X_i\}_1^N$ and that M is the maximal degree of Γ. Suppose further that $|X_i| \le A$ a.s., $1 \le i \le N$. Then*

$$\left|\kappa_j\left(\sum_1^N X_i\right)\right| \le C_j' N(M+1)^{j-1} A^j, \qquad j \ge 1, \tag{4.5}$$

for some universal constants C_j'.

PROOF. We use the expansion (4.2). By Lemma 3, we only have to consider terms $\kappa(X_{i_1},\dots,X_{i_j})$ such that the corresponding j (not necessarily distinct) vertices of Γ form a connected subgraph. If $v_1,\dots,v_j$ are j such vertices, then it is possible to reorder them so that $\{v_1,v_2\},\{v_1,v_2,v_3\},\dots,\{v_1,\dots,v_j\}$ form connected subgraphs of Γ. We estimate the number of such sequences. v_1 may be chosen in N ways. Since either v_2 equals v_1 or is connected to v_1 by an edge, there are at most $M+1$ choices of v_2 for every v_1. Similarly, v_3 must equal v_1 or v_2 or be connected to one of them; hence there are at most $2(M+1)$ choices of v_3. Continuing in this way, we see that there are at most

$N(M+1)2(M+1)\cdots(j-1)(M+1)=(j-1)!N(M+1)^{j-1}$ such sequences. Taking the $j!$ possible orderings of each sequence into account, we see that there are at most $j!(j-1)!N(M+1)^{j-1}$ nonzero terms in the sum in (4.2). Because of (4.4) and the bound on X_i, each term is bounded (in absolute value) by $C_j A^j$, which gives (4.5) with $C_j' = j!(j-1)!C_j$. □

PROOF OF THEOREM 2. Replacing X_{ni} by $(X_{ni}-EX_{ni})/\sigma_n$ (and A_n by $2A_n/\sigma_n$), we may assume that $ES_n = 0$ and $\sigma_n = 1$. Then, since $N_n \ge M_n$, (1.5) implies that, for every $j \ge m$, $(N_n/M_n)^{1/j}M_n A_n \to 0$ and thus $N_n M_n^{j-1}A_n^j \to 0$ as $n \to \infty$. By Lemma 4, $\kappa_j(S_n) \to 0$ as $n \to \infty$ for $j \ge m$, and thus $S_n \to_d N(0,1)$ by Theorem 1. □

5. Random graphs. In this section we give an application of the results above to the theory of random graphs. We confine ourselves to a single example.

The random graph $G_{n,p}$ has n vertices $1, 2, \ldots, n$ and the $\binom{n}{2}$ possible edges occur independently with probability p. Define, for $1 \le i < j < k \le n$, $X_{ijk} = 1$ if exactly two of the edges $(i,j),(i,k),(j,k)$ occur in $G_{n,p}$, and $X_{ijk} = 0$ otherwise. Let $S_n = \sum_{i<j<k} X_{ijk}$. Thus S_n is the number of triads with exactly two edges in the random graph $G_{n,p}$.

Since $EX_{ijk} = 3p^2q$, with $q = 1-p$, $ES_n = \binom{n}{3}3p^2q$. Let $\sigma_n^2 = \operatorname{var} S_n$. It is easily seen that

$$\begin{aligned}\sigma_n^2 = \operatorname{var} S_n &= \binom{n}{3}\operatorname{var} X_{123} + \binom{n}{2}(n-2)(n-3)\operatorname{cov}(X_{123}, X_{124}) \\ &= \binom{n}{3}3p^2q(1-3p^2q) + 12\binom{n}{4}p^3q(3p-2)^2.\end{aligned} \tag{5.1}$$

Let p be fixed, $0 < p < 1$, and let $n \to \infty$. Then, by Nowicki (1985, Corollary 10),

$$n^{-2}(S_n - ES_n) \to_d N\left(0, \tfrac{1}{2}p^3q(3p-2)^2\right). \tag{5.2}$$

If $p \ne \frac{2}{3}$, this may by (5.1) be written

$$(S_n - ES_n)/\sigma_n \to_d N(0,1). \tag{5.3}$$

However, in the degenerate case $p = \frac{2}{3}$, (5.2) reduces to $n^{-2}(S_n - ES_n) \to_p 0$, while (5.1) gives $\sigma_n^2 = O(n^3)$. We will show that (5.3) nevertheless holds. Thus, by (5.1),

$$n^{-3/2}(S_n - ES_n) \to_d N\left(0, \tfrac{10}{243}\right) \quad \text{as } n \to \infty,\ p = \tfrac{2}{3}.$$

PROOF OF (5.3) FOR $p = \frac{2}{3}$. Construct a graph Γ_n as follows. The vertices of Γ_n are the $\binom{n}{3}$ triples (i,j,k), $1 \le i < j < k \le n$, and an edge joins two vertices (i,j,k) and (i',j',k') iff the (unordered) sets $\{i,j,k\}$ and $\{i',j',k'\}$ have two common elements. (For example, there is an edge between $(1,2,3)$ and $(1,3,5)$.) Γ_n is a dependency graph for the family $\{X_{ijk}\}_{1\le i<j<k\le n}$, and we apply Theorem 2 (with a slight change of notation). We have $N_n = \binom{n}{3} \le n^3$, $A_n = 1$, and $M_n = 3(n-3)$ (provided $n \ge 4$). Furthermore, $\sigma_n^2 = c\binom{n}{3}$ by (5.1), with

$c > 0$. Hence, for $n > 3$, $m > 0$, and some constant $C(m)$,

$$(N_n/M_n)^{1/m} M_n A_n/\sigma_n \leq C(m) n^{2/m} n n^{-3/2} = C(m) n^{2/m-1/2}.$$

Consequently, (1.5) holds with $m = 5$, and (5.3) follows by Theorem 2. □

Note that the variables X_{ijk}, in this degenerate case, are pairwise independent but not independent, e.g.,

$$EX_{123}X_{124}X_{134}X_{234} = \tfrac{16}{243} \neq (EX_{123})^4.$$

REFERENCES

CHEN, L. H. Y. (1978). Two central limit problems for dependent random variables. *Z. Wahrsch. verw. Gebiete* **43** 223–243.

CRAMÉR, H. (1937). *Random Variables and Probability Distributions.* Cambridge Univ. Press, Cambridge.

CRAMÉR, H. (1945). *Mathematical Methods of Statistics.* Almqvist & Wiksell, Uppsala.

LEONOV, V. P. and SHIRYAEV, A. N. (1959). On a method of calculation of semiinvariants. *Theory Probab. Appl.* **4** 319–329.

MARCINKIEWICZ, J. (1939). Sur une propriété de la loi de Gauss. *Math. Z.* **44** 612–618.

NOETHER, G. E. (1970). A central limit theorem with nonparametric applications. *Ann. Math. Statist.* **41** 1753–1755.

NOVICKI, K. (1985). Asymptotic normality of graph statistics. Technical Report 1985:7, Lund.

DEPARTMENT OF MATHEMATICS
THUNBERGSVÄGEN 3
S-752 38 UPPSALA
SWEDEN