ASYMPTOTIC EXPANSIONS FOR A CLASS OF DISTRIBUTION FUNCTIONS

By K. C. Chanda[1]

Washington State University

1. Introduction and summary. Investigations have been made in the past by several people on the possibility of extending the content of the classical central limit Theorem when the basic random variables are no longer independent. Several interesting extensions have been made so far. Hoeffding and Robbins (1948) have established asymptotic normality for the distribution of mean of a sequence of m-dependent random variables, where m is a finite positive constant. The result has been proved by Diananda (1953) under more general conditions and has been extended to cover situations where the random variables $\{X_t\}$ $(t = 1, 2, 3, \cdots)$ are of the type $X_t - E(X_t) = \sum_{j=0}^{\infty} g_j Y_{t-j}$ where $\{Y_t\}$ $(t = 0, \pm 1, \cdots)$ is an m-dependent stationary process and

$$E(Y_t) = 0, \qquad \sum_{j=0}^{\infty} g_j < \infty.$$

Walker (1954) has established asymptotic normality for the distributions of serial correlations based on X_t of the above form. However, so far no attempt has been made to investigate whether the type of asymptotic expansions as discussed by Cramér (1937), Berry (1941) and Hsu (1945) for the distributions of means of independent random variables could also be extended to apply to situations where the random variables are not independent. Chanda (1962) has made a start in this direction, but the results are understandably incomplete. An attempt has been made in this paper to investigate this problem more systematically. The conclusion is that an extension is possible under conditions precisely similar to those under which Cramér, Berry and Hsu proved their results.

2. Asymptotic distribution of the mean of a sequence of m-dependent random variables. We assume that the process $\{X_t\}$ $(t = 1, 2, 3, \cdots)$, $E(X_t) = 0$ is either a stationary m-dependent process or is a linear process defined by $X_t = \sum_{j=0}^{\infty} g_j Y_{t-j}$ where $\{Y_t\}$ $(t = 0, \pm 1, \cdots)$ is a sequence of mutually independent random variables with a common distribution which is not purely discrete and $E(Y_t) = 0$, $E(Y_t^2) = 1$, and $\sum_{j=0}^{\infty} |g_j| < \infty$, $|\sum_{j=0}^{\infty} g_j| > 0$. Let the $(r_1, \cdots, r_n)$th order joint absolute moment of $X_1, \cdots, X_n$ exist for $\sum_{j=1}^{\infty} r_j \leqq r (r \geqq 3)$. Define $Z_n = \sum_{t=1}^{n} X_t/s_n$ where $s_n^2 = \text{Var}(\sum_{t=1}^{n} X_t)$. Let $F_n(x)$, $\phi_n(\alpha)$ denote respectively the distribution function (d.f.) and characteristic function (ch.f.) of Z_n and $F(x)$, $\phi(\alpha)$ be the corresponding quantities for the standard normal distribution. Our results can then be stated as follows.

Received June 26, 1962; revised May 20, 1963.

[1] Now at the Iowa State University.

THEOREM. *Let* $E\{\prod_{j=1}^{n}|X_j|^{r_j}\}$ *exist for* $\sum_{j=1}^{n} r_j \leqq r$ *where r can be any arbitrary integer* $\geqq 3$ *when* $\{X_t\}$ *is a linear process defined by* $X_t = \sum_{j=0}^{\infty} g_j Y_{t-j}$ *with* $E(Y_t) = 0$, $E(Y_t^2) = 1$ *and* $\sum_{j=0}^{\infty}|g_j| < \infty$, $|\sum_{j=0}^{\infty} g_j| > 0$ *and the distribution of* Y_t *is not purely discrete and* $r = 3$ *when* $\{X_t\}$ *is a general m-dependent stationary process with* $E(X_t) = 0$. *Then*

$$F_n(x) = F(x) + G_{n,r}(x) + R_{n,r}(x), \tag{2.1}$$

where $G_{n,r}(x)$ *is a linear combination of successive derivatives*

$$F^{(3)}(x), \cdots, F^{3(r-3)}(x)$$

with each coefficient of the form $n^{-\frac{1}{2}\nu}$ $(1 \leqq \nu \leqq r-3)$ *times a quantity depending on r and the multivariate moments of* $X_1, \cdots, X_n$ *but bounded for all n and*

$$|R_{n,r}(x)| \leqq M/n^{\frac{1}{2}(r-2)} \quad \text{for all} \quad x, \tag{2.2}$$

where M is a generic symbol denoting a finite positive constant. It can be noted, in passing, that $F(x) + G_{n,r}(x)$ is obtained by expansion of the cumulant function $\log \phi_n(\alpha)$ and term by term inversions.

Before proving this theorem we shall consider a few lemmas.

LEMMA 1. *Let* $\lambda_{\nu,n}$ *denote the* νth *order cumulant of* $s_n Z_n$. *Then* $\lambda_{\nu,n}$ *exist and* $|\lambda_{\nu,n}| \leqq Mn$ *for all* $\nu \leqq r$.

PROOF.

CASE I: X_t is an m-dependent stationary process. Evidently, the $(r_1, \cdots r_n)$th order cumulants $\kappa_{r_1,\cdots,r_n}$ of $X_1, \cdots, X_n$ exist for $\sum_{j=1}^{n} r_j \leqq r$. Let $2 \leqq \nu \leqq r$. Then $\lambda_{\nu,n} = \nu! \sum_1 (\kappa_{\nu_1,\cdots,\nu_n}/\prod_{j=1}^{n} \nu_j!)$ where $\sum_1$ denotes summation over all possible values of $\nu_j \geqq 0 (1 \leqq j \leqq n)$ such that $\sum_{j=1}^{n} \nu_j = \nu$. Note that $\kappa_{\nu_1,\cdots,\nu_n} = 0$ if at least one of the differences $j - j'(j > j')$, such that $\nu_{j'} > 0$, $\nu_j > 0$, $\nu_{j'+1} = \cdots = \nu_{j-1} = 0$, is greater than m. Let $\nu_{j_1}, \cdots, \nu_{j_q}$ $(j_1 < j_2 < \cdots < j_q)$ be the $q(q \leqq \nu)$ non-zero values of the ν_j corresponding to a non-zero term in $\sum_1$. From m-dependence we must have $1 \leqq j_{u+1} - j_u \leqq m$ $(1 \leqq u \leqq q-1)$, and it follows that the number of possible values for j_1, given $v_1 = j_2 - j_1, \cdots, v_{q-1} = j_q - j_{q-1}$ lies between $n - (q-1)$ and

$$n - (q-1)m.$$

To each of these corresponds the same joint cumulant, which by stationarity is a function of $v_1, \cdots, v_{q-1}$ for given values of $\nu_{j_1}, \cdots, \nu_{j_q}$. It follows from this that the total number of different non-zero cumulant terms occurring in $\sum_1$ remains finite as $n \to \infty$. As a matter of fact this number

$$\leqq \sum_{q=1}^{\nu} \binom{\nu-1}{q-1} m^{q-1}$$

Hence these cumulants have a finite upper bound and

$$|\lambda_{\nu,n}| \leqq M\nu! \sum_1 \left(1 \Big/ \prod_{j=1}^{n} \nu_j!\right) \leqq \nu! \sum_{q=1}^{\nu} M\{n - (q-1)\}m^{q-1} \sum_2 \left(1 \Big/ \prod_{j=1}^{q} \nu_j'!\right),$$

where $\sum_2$ denotes summation over all positive integral values of ν_j' subject to $\sum_{j=1}^{q} \nu_j' = \nu$. As a result we have

$$|\lambda_{\nu,n}| \leqq Mn. \tag{2.3}$$

Case II: $\{X_t\}$ is a linear process. Let $\xi(\alpha)$ denote the ch.f. of Y_t. Then the ch.f. of $s_n Z_n$ is given by

$$\prod_{k=1}^{n} \xi\left(\alpha \sum_{j=0}^{n-k} g_j\right) \prod_{k=0}^{\infty} \xi\left(\alpha \sum_{j=1+k}^{n+k} g_j\right), \tag{2.4}$$

so that

$$\begin{aligned} |\lambda_{\nu,n}| &\leqq M \left| \sum_{k=1}^{n} \left(\sum_{j=0}^{n-k} g_j \right)^{\nu} + \sum_{k=0}^{\infty} \left(\sum_{j=1+k}^{n+k} g_j \right)^{\nu} \right| \\ &\leqq M a^{\nu-2} \left\{ \sum_{k=1}^{n} \left(\sum_{j=0}^{n-k} g_j \right)^2 + \sum_{k=0}^{\infty} \left(\sum_{j=1+k}^{n+k} g_j \right)^2 \right\} \leqq M s_n^2, \end{aligned} \tag{2.5}$$

where $a = \sum_0^\infty |g_j|$. But $s_n^2/n \to \sum_{v=-\infty}^{\infty} \tau_v$, as $n \to \infty$, where

$$\tau_v = \sum_{j=0}^{\infty} g_j g_{j+v} \quad \text{and} \quad \sum_{-\infty}^{\infty} \tau_v = \left(\sum_{j=0}^{\infty} g_j\right)^2 > 0;$$

further $|\sum \tau_v| \leqq (\sum g_j)^2 < \infty$. Hence for sufficiently large n, $s_n^2/n \leqq M$ and

$$|\lambda_{\nu,n}| \leqq Mn. \tag{2.6}$$

Lemma 2. *Let $H(x)$ be any* d.f. *such that cumulants κ_r of $H(x)$ of order r and lower exist. Then if $\psi(\alpha)$ is the corresponding* ch.f. *we can write*

$$\log \psi(\alpha) = \sum_{\nu=0}^{r-1} (i\alpha)^{\nu} \kappa_\nu / \nu! + \theta_r(\alpha) \kappa_r \alpha^r, \tag{2.7}$$

where $|\theta_r(\alpha)| \leqq 1$ for all $|\alpha| \leqq \epsilon$ where ϵ is a suitably determined finite positive constant.

Proof. If we define $\theta_r(\alpha)$ by the relation above then it can easily be shown that $\lim_{\alpha \to 0} \theta_r(\alpha) = i^r/r!$ Since $\theta_r(\alpha)$ is a continuous function of α the result of the lemma follows.

Lemma 3. *Let $\phi(\alpha)\chi_{n,r}(i\alpha) = \int_{-\infty}^{\infty} \exp(i\alpha x)\, dG_{n,r}(x)$. Then if*

$$\delta_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x) - G_{n,r}(x)|,$$

for all sufficiently large values of $A(>0)$,

$$\begin{aligned} A\delta_n &\left\{ 3 \int_0^{A\delta_n} (1 - \cos x)/x^2\, dx - \pi \right\} \\ &\leqq MA \int_0^A |\phi_n(\alpha) - \phi(\alpha)\{1 + \chi_{n,r}(i\alpha)\}| \alpha^{-1}\, d\alpha. \end{aligned}$$

For proof see Hsu (1945). The result is stated in Equation (71) on page 18 of Hsu's paper.

We shall now prove our theorem.

CASE I: $\{X_t\}$ is an m-dependent stationary process. Let $\phi_n^*(\alpha)$ denote the ch.f. of $s_n Z_n$. Then from Lemma 2 and the fact that $\lambda_{1,n} = 0$, we have

$$\log \phi_n^*(\alpha) = -(\alpha^2/2)\lambda_{2,n} + \theta_3\lambda_{3,n}|\alpha|^3,$$

where for $|\alpha| \leqq \epsilon$, $|\theta_3| \leqq M$. Further since $\phi_n(\alpha) = \phi_n^*(\alpha/s_n)$ and $\lambda_{2,n} = s_n^2$, $\log \phi_n(\alpha) = -\alpha^2/2 + \theta_3\lambda_{3,n}|\alpha|^3/s_n^3$, where for $|\alpha| \leqq s_n\epsilon$, $|\theta_3| \leqq M$. We can therefore, write $\phi_n(\alpha) = \exp(-\frac{1}{2}\alpha^2)\{1 + \theta s \exp(|s|)\}$, where $|\theta| \leqq 1$ and $s = \theta_3\lambda_{3,n}|\alpha|^3/s_n^3$. Since $|\lambda_{3,n}| \leqq Mn$ by Lemma 1, and $\lim_{n\to\infty} s_n^2/n > 0$ so that $n/s_n^2 \leqq M$, $|s| \leqq M|\alpha^3|/s_n < \frac{1}{4}\alpha^2$ for sufficiently small values of $|\alpha|/s_n$. It follows, therefore, that there exists $\epsilon'(0 < \epsilon' \leqq \epsilon)$ such that for all $|\alpha| \leqq s_n\epsilon'$

$$\phi_n(\alpha) = \exp(-\alpha^2/2) + \theta|\alpha|^3/s_n \exp(-\alpha^2/4), \tag{2.9}$$

where $|\theta| \leqq M$. By Lemma 3, taking $\chi_{n,3}(i\alpha) = 0(G_{n,3}(x) = 0)$ $A = s_n\epsilon'$ we have

$$A\int_0^A |\phi_n(\alpha) - \phi(\alpha)|\alpha^{-1}\,d\alpha \leqq M. \tag{2.10}$$

Hence $A\delta_n \leqq M$ or

$$|R_{n,3}(x)| \leqq M/s_n \leqq M/n^{\frac{1}{2}}. \tag{2.11}$$

CASE II: $\{X_t\}$ is a linear process. Using the same symbols as in Case I, we have from Lemma 2

$$\log \phi_n^*(\alpha) = \sum_{\nu=0}^{r-1} (i\alpha)^\nu\lambda_{\nu,n}/(\nu! s_n^\nu) + \theta_r\lambda_{r,n}|\alpha|^r,$$

where for $|\alpha| \leqq \epsilon$, $|\theta_r| \leqq M$. Hence

$$\log \phi_n(\alpha) = \sum_{\nu=0}^{r-1} (i\alpha)^\nu\lambda_{\nu,n}/(\nu! s_n^\nu) + \theta_r\lambda_{r,n}|\alpha|^r/s_n^r.$$

Since $\lambda_{1,n} = 0$ and $\lambda_{2,n} = s_n^2$,

$$\phi_n(\alpha) = \exp(-\alpha^2/2)\left\{1 + \sum_{j=1}^{r-3} s^j/j! + \frac{\theta|s|^{r-2}\exp(|s|)}{(r-2)!}\right\},$$

where $s = \sum_{\nu=3}^{r-1} (i\alpha)^\nu\lambda_{\nu,n}/(\nu! s_n^\nu) + \theta_r\lambda_{r,n}|\alpha|^r/s_n^r$ and $|\theta| \leqq 1$ (here we have taken $r > 3$; when $r = 3$ the term $\sum_j s^j/j!$ is absent and the argument below simplifies accordingly). Also since $n/s_n^2 \leqq M$,

$$|s| \leqq M|\alpha|^3 n/s_n^3 \sum_{\nu=0}^{r-4} |\alpha|^\nu/(\nu! s_n^\nu) \leqq M|\alpha|^3/s_n \exp(|\alpha|/s_n) \leqq M|\alpha|^3/s_n \tag{2.12}$$

for $|\alpha| \leqq s_n\epsilon$. Hence

$$|s^j/j!| \leqq M|\alpha|^{3j}/s_n^j \exp(j|\alpha|/s_n), \qquad (1 \leqq j \leqq r-2). \tag{2.13}$$

If now we expand s^j and denote by P_j the polynomial part of $s^j/j!$ of degree

$r-3$ in s_n^{-1} and the remainder term by R_j we have from (2.13)

$$(2.14)\qquad \begin{aligned} |R_j| &\leqq M|\alpha|^{3j}/s_n^j \sum_{\nu=r-2-j}^{\infty} j^{\nu}|\alpha|^{\nu}/(\nu! s_n^{\nu}) \\ &\leqq M|\alpha|^{r-2+2j}/s_n^{r-2} \sum_{\nu=0}^{\infty} j^{\nu}|\alpha|^{\nu}/(\nu! s_n^{\nu}) \leqq M|\alpha|^{r-2+2j}/s_n^{r-2} \end{aligned}$$

for all $|\alpha| \leqq s_n\epsilon$. Further from (2.12), $-\alpha^2/2 + |s| \leqq -\alpha^2/2 + M|\alpha|^3/s_n \leqq -\alpha^2/4$ for some $\epsilon' \leqq \epsilon$. Hence for all $|\alpha| \leqq s_n\epsilon'$ we have

$$(2.15)\qquad \begin{aligned} \phi_n(\alpha) &= \exp(-\alpha^2/2)\{1 + \chi_{n,r}(i\alpha)\} \\ &\quad + \theta/s_n^{r-2}\{|\alpha|^r + |\alpha|^{r+1} + \cdots + |\alpha|^{3(r-2)}\} \exp(-\alpha^2/4), \end{aligned}$$

where $\chi_{n,r}(i\alpha) = \sum_{j=1}^{r-3} P_j$. Note that $\chi_{n,r}(i\alpha)$ is a polynomial in s_n^{-1} of degree $r-3$ with coefficients which are functions of n and α but, for given α, are bounded for all n. Evidently $\chi_{n,r}(i\alpha) = 0$ for $r = 3$. In Lemma 3, putting $A = (s_n\epsilon')^{r-2}$, we have

$$A \int_0^A |\phi_n(\alpha) - \phi(\alpha)\{1 + \chi_{n,r}(i\alpha)\}|\alpha^{-1}\, d\alpha = (s_n\epsilon')^{r-2}\left\{\int_0^{s_n\epsilon'} + \int_{s_n\epsilon'}^{(s_n\epsilon')^{r-2}}\right\} = J_1 + J_2 \quad \text{say.}$$

Then $J_1 \leqq M$ and

$$(2.16)\qquad J_2 \leqq (s_n\epsilon')^{r-2}\left[\int_{s_n\epsilon'}^{(s_n\epsilon')^{r-2}} |\phi_n(\alpha)|\alpha^{-1}\, d\alpha + \int_{s_n\epsilon'}^{(s_n\epsilon')^{r-2}} |\phi(\alpha)\{1 + \chi_{n,r}(i,\alpha)\}|\alpha^{-1}\, d\alpha\right].$$

The second term on the right hand side of (2.16) is evidently $\leqq M$. Now since $|\sum_0^\infty g_j| > 0$ there exists a positive integer N such that $|\sum_{j=0}^k g_j| > 0$ for all $k > N$. From (2.4), we have for $n > N + 1$, $|\phi_n^*(\alpha)| \leqq \prod_{k=0}^{n-1} |\xi(\alpha \sum_{j=0}^k g_j)| \leqq \prod_{k=N+1}^{n-1} |\xi(\alpha \sum_{j=0}^k g_j)|$. Again, since the distribution of Y_t is not purely discrete, the only solution to $\xi(\alpha) = 1$ is $\alpha = 0$. Since $|\xi(\alpha)| \leqq 1$ for all α, there exists a positive real number $\rho < 1$ such that $|\xi(\alpha)| \leqq \rho$ for all $|\alpha| \geqq \epsilon$ however small $\epsilon(>0)$ may be. Hence

$$(2.17)\qquad |\phi_n^*(\alpha)| \leqq M\rho^n$$

for all $|\alpha| \geqq \epsilon'$. Thus

$$\int_{s_n\epsilon'}^{(s_n\epsilon')^{r-2}} |\phi_n(\alpha)|\alpha^{-1}\, d\alpha = \int_{\epsilon'}^{s_n^{r-3}\epsilon'^{r-2}} |\phi_n^*(\alpha)|\alpha^{-1}\, d\alpha \leqq M\rho^n \log(s_n\epsilon').$$

Since $(s_n\epsilon')^{r-2}\rho^n \log(s_n\epsilon') \leqq M$ for all n, it follows that the left hand side of (2.16) is $\leqq M$. Hence $A\delta_n \leqq M$ and

$$(2.17)\qquad |R_{n,r}(x)| \leqq M/s_n^{r-2} \leqq M/n^{\frac{1}{2}(r-2)}$$

for all x.

REFERENCES

BERRY, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49** 122–136.

CHANDA, K. C. (1962). Asymptotic expansions for tests of goodness of fit for linear autoregressive schemes. (Submitted to *Biometrika.*)

CRAMÉR, H. (1937). *Random variables and probability distributions.* Cambridge Tracts in Mathematics, No. 36, Cambridge.

DIANANDA, P. H. (1953). Some probability limit theorems with statistical applications. *Proc. Cambridge Philos. Soc.* **49** 239–246.

HOEFFDING, W. and ROBBINS, H. (1948). The central limit theorem for dependent variables. *Duke. Math. J.* **15** 773–780.

HSU, P. L. (1945). The approximate distributions of the mean and variance of a sample of independent variables. *Ann. Math. Statist.* **16** 1–29.

WALKER, A. M. (1954). The asymptotic distribution of serial correlation coefficients for autoregressive processes with dependent residuals. *Proc. Cambridge Philos. Soc.* **50** 60–64.