ASYMPTOTIC EXPANSIONS IN GLOBAL CENTRAL LIMIT THEOREMS

By RALPH PALMER AGNEW¹

Cornell University

1. Introduction. Let ξ_1 , ξ_2 , \cdots be independent random variables having the same d.f. (distribution function) F(x). We suppose that

(1.1)
$$\int_{-\infty}^{\infty} x \, dF(x) = 0, \qquad \int_{-\infty}^{\infty} x^2 \, dF(x) = 1$$

so that F(x) has mean 0 and standard deviation 1. Let $F_n(x)$ denote the d.f. of the normalized sum

$$(1.2) (\xi_1 + \xi_2 + \cdots + \xi_n)/n^{\frac{1}{2}}.$$

A special case of the central limit theorem then asserts that, for each individual x in the interval $-\infty < x < \infty$,

$$\lim_{n \to \infty} F_n(x) = \Phi(x)$$

where $\Phi(x)$ is the Gaussian d.f. defined by

(1.22)
$$\Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-u^{2}/2} du.$$

It is our purpose to study the behavior as $n \to \infty$ of the constants C_n defined by

$$(1.3) C_n = \int_{-\infty}^{\infty} |F_n(x) - \Phi(x)|^2 dx.$$

For each p > 0, let constants $C_n(p)$ be defined by

$$(1.31) C_n(p) = \int_{-\infty}^{\infty} |F_n(x) - \Phi(x)|^p dx$$

when these integrals exist, that is, are finite. It is known from [1] and [2] that the hypotheses (1.1) imply that if $p > \frac{1}{2}$ then the constants $C_n(p)$ exist and $\lim_{n\to\infty} C_n(p) = 0$. Beyond this, not very much is known about the constants $C_n(p)$. The moments α_k and the absolute moments β_k of F(x) are defined by

(1.4)
$$\alpha_k = \int_{-\infty}^{\infty} x^k dF(x), \qquad \beta_k = \int_{-\infty}^{\infty} |x|^k dF(x)$$

Received March 11, 1957.

¹ The research of this author was supported in part by the United States Air Force under Contract No. AF18(600)-685 monitored by the Office of Scientific Research.

721

when these integrals exist. If β_3 exists, then an inequality of Esseen ([5], p. 78) shows that there is a constant $K(\beta_3)$, depending only upon β_3 , such that

(1.5)
$$|F_n(x) - \Phi(x)| \le \frac{K(\beta_3)}{n^{\frac{1}{2}}} \frac{\log (2 + |x|)}{1 + |x|^3}.$$

This implies that if β_3 exists, then the constants $C_n(p)$ exist when $p > \frac{1}{3}$ and $C_n(p) = O(n^{-p/2})$. In particular, if β_3 exists then $C_n = O(n^{-1})$. There is a sense in which this result cannot be improved because it is shown in [2] that if F(x) is the symmetric binomial d.f. satisfying (1.1), then

(1.6)
$$C_n = \frac{1}{n} \frac{1}{6\pi^{\frac{1}{2}}} + O\left(\frac{1}{n^2}\right).$$

The only other case in which the constants C_n have been appraised is that for which F(x) is the d.f. of a random variable ξ uniformly distributed over $-a \le x \le a$; in this case (1.1) implies that $a = 3^{\frac{1}{2}}$ and it is shown in [2] that

(1.61)
$$C_n = \frac{1}{n^2} \frac{3}{1280\pi^{\frac{1}{2}}} + O\left(\frac{1}{n^3}\right).$$

One of our main purposes is to give conditions under which there exist constants D_1 , D_2 , D_3 , \cdots such that the expansion

(1.7)
$$C_n = \frac{D_1}{n} + \frac{D_2}{n^2} + \cdots + \frac{D_k}{n^k} + O\left(\frac{1}{n^{k+1}}\right)$$

is valid for each $k=1, 2, 3, \cdots$ and to give explicit expressions for D_1 and D_2 . Such results are given at the ends of sections 4 and 6. Binomial distributions are treated in section 7, and the symmetric binomial d.f. is treated more extensively in section 9.

2. Formulas for the constants C_n . Information about the constants C_n is obtainable by use of the c.f. (characteristic function) $\phi(t)$ of F(x) which is defined by

(2.01)
$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x).$$

It is shown in [2] that

(2.02)
$$\int_{-\infty}^{\infty} |F_n(x) - \Phi(x)|^2 dx = \frac{1}{n^{\frac{1}{2}}} \frac{1}{\pi} \int_{0}^{\infty} |[\phi(t)]^n - e^{-nt^2/2}|^2 \frac{dt}{t^2}.$$

Hence

(2.1)
$$C_n = \frac{1}{n^{\frac{1}{2}}} \frac{1}{\pi} \int_0^\infty | \left[\phi(t) \right]^n - e^{-nt^2/2} |^2 \frac{dt}{t^2}.$$

The hypotheses (1.1) imply that, at least when k = 0, 1, 2,

(2.2)
$$\phi^{(k)}(t) = \int_{-\infty}^{\infty} (ix)^k e^{itx} dF(x)$$

and hence that $\phi(0) = 1$, $\phi'(0) = 0$, and $\phi''(0) = -1$. This implies that we can choose a positive constant T such that

$$|\phi(t)| < 1 \qquad (0 < t \le T).$$

Let constants δ_1 , δ_2 , \cdots be defined by

(2.22)
$$\delta_n = \log n / n^{\frac{1}{2}} \qquad (n = 1, 2, 3, \cdots).$$

For values of n so great that $0 < \delta_n < T$, we split C_n into the sum of four terms by putting

$$(2.3) C_n = C_n^{(1)} + C_n^{(2)} + C_n^{(3)} + C_n^{(4)}$$

where

$$(2.4) C_n^{(1)} = \frac{1}{n^{\frac{1}{2}}} \frac{1}{\pi} \int_{t_-}^{\infty} \left\{ e^{-nt^2/2} - \left[\phi(t) \right]^n - \left[\overline{\phi(t)} \right]^n \right\} e^{-nt^2/2} \frac{dt}{t^2},$$

(2.5)
$$C_{n}^{(2)} = \frac{1}{n^{\frac{1}{2}}} \frac{1}{\tau} \int_{x}^{\tau} |\phi(t)|^{2n} \frac{dt}{t^{2}},$$

(2.6)
$$C_n^{(3)} = \frac{1}{n^{\frac{1}{2}}} \frac{1}{\pi} \int_0^{\delta_n} |\left[\phi(t)\right]^n - e^{-nt^2/2}|^2 \frac{dt}{t^2},$$

(2.7)
$$C_{n}^{(4)} = \frac{1}{n^{\frac{1}{2}}} \frac{1}{\pi} \int_{T}^{\infty} |\phi(t)|^{2n} \frac{dt}{t^{2}}.$$

Estimation of $C_n^{(1)}$ and $C_n^{(2)}$ offers no difficulty; as we shall see,

$$(2.81) C_n^{(1)} = o(n^{-\omega}), C_n^{(2)} = o(n^{-\omega})$$

where $o(n^{-\omega})$ denotes a quantity which is $o(n^{-k})$ for each fixed positive constant k. Since $|e^{-nt^2/2}| \le 1$ and $|\phi(t)| \le 1$, it follows from (2.4) that

$$|C_{n}^{(1)}| \leq \frac{1}{n^{\frac{3}{4}}} \frac{3}{\pi} \int_{\delta_{n}}^{\infty} t^{-3} e^{-nt^{2}/2} (nt) dt$$

$$< \frac{1}{n^{\frac{3}{4}}} \frac{3}{\pi} \int_{\delta_{n}}^{\infty} e^{-nt^{2}/2} (nt) dt = \frac{1}{n^{\frac{3}{4}}} \frac{3}{\pi} \frac{1}{\delta_{n}^{3}} e^{-n\delta_{n}^{2}/2} = o(n^{-\omega}).$$

To estimate $C_n^{(2)}$, let $\psi(t) = |\phi(t)|^2$. Then $\psi(0) = 1, \psi'(0) = 0$, and $\psi''(0) = -2$. This and the fact that $\psi(t) < 1$ when $0 < t \le T$ imply that, for each sufficiently great n,

(2.83)
$$\max_{\delta_n \le t \le T} |\psi(t)| < 1 - \delta_n^2/2.$$

Hence, when n is sufficiently great,

$$(2.84) |C_{n}^{(2)}| < \frac{1}{n^{\frac{1}{2}}} \frac{1}{\pi} \int_{\delta_{n}}^{T} (1 - \delta_{n}^{2}/2)^{n} \delta_{n}^{-2} dt$$

$$< \frac{1}{n^{\frac{1}{2}}} \frac{T}{\pi \delta_{n}^{2}} \left(1 - \frac{\delta_{n}}{2} \right)^{n} = \frac{n^{\frac{1}{2}} T}{n (\log n)^{2}} \left[1 - \frac{(\log n)^{2}}{2n} \right]^{n} = o(n^{-\omega}).$$

This proves (2.81).

The problem of estimating C_n is therefore reduced to the problem of estimating $C_n^{(3)}$ and $C_n^{(4)}$. Instead of (2.3), we henceforth use the formula

$$(2.9) C_n = o(n^{-\omega}) + C_n^{(3)} + C_n^{(4)}.$$

3. The Constants $C_n^{(3)}$. In this section we appraise the constants $C_n^{(3)}$ in terms of the Thiele [7] semi-invariants γ_k of F(x). We suppose that, for some integer m for which $m \geq 3$, the moments α_m and β_m exist. Then, as $t \to 0$,

(3.1)
$$\phi(t) = 1 - \frac{t^2}{2} + \sum_{k=2}^{m} \frac{(it)^k}{k!} \alpha_k + o(t^m)$$

and using the ordinary expansion of $\log (1 + x)$ in powers of x gives

(3.2)
$$\log \phi(t) = -\frac{t^2}{2} + \sum_{k=3}^{m} \frac{(it)^k}{k!} \gamma_k + o(t^m)$$

where

(3.21)
$$\gamma_3 = \alpha_3, \quad \gamma_4 = \alpha_4 - 3, \quad \gamma_5 = \alpha_5 - 10\alpha_3,$$

$$\gamma_6 = \alpha_6 - 15\alpha_4 - 10\alpha_3^2 + 30, \cdots.$$

The constants γ_3 , γ_4 , \cdots are the Thiele semi-invariants of F(x) which are treated in the books of Cramer [3], [4] and Gnedenko and Kolmogoroff [6] and which have simplified forms here because $\alpha_0 = 1$, $\alpha_1 = 0$, and $\alpha_2 = 1$. From (3.2) we obtain, for each fixed n,

$$[\phi(t)]^n = e^{-nt^2/2}e^w$$

where w is the function of n and t defined by

(3.31)
$$w = n \left[\sum_{k=3}^{m} \frac{(it)^k}{k!} \gamma_k + o(t^m) \right].$$

From (2.6) and (3.3) we find that

(3.4)
$$C_{n}^{(3)} = \frac{1}{n^{\frac{1}{2}}} \frac{1}{\pi} \int_{0}^{\delta_{n}} e^{-nt^{2}} |e^{w} - 1|^{2} \frac{dt}{t^{2}}.$$

Supposing henceforth that $0 < t < \delta_n$, we see from (2.22) that

$$0 < nt^3 < (\log n)^3/n^{\frac{1}{2}}$$

and hence from (3.31) that w=o(1) as $n\to\infty$. Therefore we can use the formula

(3.41)
$$e^{w} - 1 = \sum_{k=1}^{m} \frac{w^{k}}{k!} + O(w^{m+1})$$

and (3.31) to obtain a formula giving $e^w - 1$ as the sum of a finite number of terms involving n, t, and γ_3 , γ_4 , \cdots , γ_m . When this finite sum is written down, it is found that

(3.42)
$$\left[\frac{1}{t} |e^{w} - 1| \right]^{2} = |u + iv|^{2} = u^{2} + v^{2}$$

where

(3.43)
$$u = \frac{\gamma_4}{24} n t^3 - \frac{\gamma_3^2}{72} n^2 t^5 + \cdots,$$

$$(3.44) v = -\frac{\gamma_3}{6} nt^2 + \frac{\gamma_5}{120} nt^4 - \frac{\gamma_3 \gamma_4}{144} n^2 t^6 + \frac{\gamma_3^3}{1296} n^3 t^8 + \cdots$$

In (3.43), (3.44), and formulas which follow, the final dots represent finite sums of terms which turn out to give contributions to $C_n^{(3)}$ which are of lower orders of magnitude than the contributions of the terms which precede the dots. From (3.42), (3.43), and (3.44) we obtain

(3.45)
$$\left[\frac{1}{t} \left| e^{w} - 1 \right| \right]^{2} = \frac{\gamma_{3}^{2}}{36} n^{2} t^{4} + \frac{5\gamma_{4}^{2} - 8\gamma_{3}\gamma_{5}}{2880} n^{2} t^{6} + \frac{\gamma_{3}^{2} \gamma_{4}}{864} n^{3} t^{8} - \frac{\gamma_{3}^{4}}{15552} n^{4} t^{10} + \cdots$$

From (3.4) and (3.45) we see that $C_n^{(3)}$ is a linear combination, with coefficients depending upon γ_3 , γ_4 , \cdots , of integrals of the form

(3.5)
$$J_n = \frac{1}{n^{\frac{1}{2}}} \int_0^{\delta_n} e^{-nk^2} n^p t^{2q} dt$$

where p and q are positive integers. Putting $t = n^{-\frac{1}{2}}u$ in (3.5) and using (2.22) gives

(3.51)
$$J_n = \frac{1}{n^{1+q-p}} \int_0^{\log n} e^{-u^2} u^{2q} du.$$

But, when n is sufficiently great,

(3.52)
$$\int_{\log n}^{\infty} e^{-u^2} u^{2q} \, du = \int_{\log n}^{\infty} (u^{2q-1} e^{-u^2/2}) e^{-u^2/2} u \, du$$
$$< \int_{\log n}^{\infty} e^{-u^2/2} u \, du = e^{-(\log n)^2/2} = o(n^{-\omega}).$$

Hence

(3.53)
$$J_n = o(n^{-\omega}) + \frac{1}{n^{1+q-p}} \int_0^{\infty} e^{-u^2} u^{2q} du.$$

Using a standard formula for the integral in (3.53) gives

(3.54)
$$J_n = o(n^{-\omega}) + \frac{\pi^{\frac{1}{2}}}{n^{1+q-p}} \frac{(2q)!}{q! 2^{2q+1}}.$$

Use of (3.4), (3.45), (3.5), and (3.54) gives

(3.6)
$$C_n^{(3)} = \frac{A_1}{n} + \frac{A_2}{n^2} + \cdots + \frac{A_k}{n^k} + O\left(\frac{1}{n^{k+1}}\right)$$

where

$$(3.61) A_1 = \frac{1}{\pi^2} \frac{\gamma_3^2}{96},$$

(3.62)
$$A_2 = \frac{1}{\pi^{\frac{1}{2}}} \left[\frac{5\gamma_4^2}{3072} - \frac{\gamma_3}{384} + \frac{35\gamma_3^2}{9216} + \frac{35\gamma_3^4}{36864} \right],$$

and each of the constants A_1 , A_2 , A_3 , \cdots depends only upon a finite number of the semi-invariants γ_3 , γ_4 , \cdots . In case the given d.f. F(x) has finite moments of all positive integer orders, the integer m of this section can be chosen as great as we wish and (3.6) is then valid for each $k = 1, 2, 3, \cdots$. The expressions for A_1 and A_2 given in (3.61) and (3.62) are particularly simple in the important case in which F(x) is symmetric because in this case $\gamma_3 = 0$. The complexity of the expression for A_k increases very rapidly as k increases.

4. The constants $C_n^{(4)}$; case $\limsup |\phi(t)| < 1$. In this section, we suppose that F(x) is a d.f. having a c.f. $\phi(t)$ for which

$$\lim_{t\to\infty}\sup|\phi(t)|<1$$

and show that in this case

$$C_n^{(4)} = o(n^{-\omega})$$

where, as above, $o(n^{-\omega})$ denotes a quantity which is $o(n^{-k})$ for each positive constant k.

It is known ([3], p. 26) that if a c.f. $\phi(t)$ satisfies the hypothesis (4.1), then $|\phi(t)| < 1$ when t > 0. Since each c.f. is everywhere continuous, the hypothesis (4.1) therefore implies that if T > 0, then there is a constant θ such that $0 < \theta < 1$ and $|\phi(t)| \le \theta$ when $t \ge T$. The definition (2.7) of $C_n^{(4)}$ therefore implies that

(4.3)
$$C_n^{(4)} \le \int_T^{\infty} \theta^{2n} t^{-2} dt = T^{-1} \theta^{2n}$$

and the desired conclusion (4.2) follows.

Thus in case F(x) has a nonvanishing absolutely continuous component [3, page 17 and page 25] and in other cases where (4.1) holds, we have

$$(4.4) C_n = o(n^{-\omega}) + C_n^{(3)}$$

and the results of section 3 suffice for the estimation of C_n . In particular, if (4.1) holds and F(x) has finite moments of all positive integer orders, then (1.7) is valid when the constants D_1 , D_2 , \cdots are the constants A_1 , A_2 , \cdots in (3.6).

5. The uniform distribution. Let F(x) be the d.f. of a random variable ξ uniformly distributed over $-a \le x \le a$ so that F(x) = 0 when $x \le a$, F(x) = (x+a)/2a when $-a \le x \le a$, and F(x) = 1 when $x \ge a$. This d.f. has mean 0, and we assume that $a = 3^{\frac{1}{2}}$ so that the standard deviation is 1. In this case (4.4) holds. The moments α_1 , α_2 , \cdots defined by (1.4) are

(5.1)
$$\alpha_k = \int_{-a}^a (x^k/2a) \ dx$$

so that $\alpha_k = 0$ when k is odd and $\alpha_k = a^k/(k+1)$ when k is even. Using (3.21) gives $\gamma_3 = 0$ and $\gamma_4 = -6/5$. Using (4.4) and (3.6) then gives the result (1.61).

6. The constants $C_n^{(4)}$; case $|\phi(t)|$ periodic. It is well known that if there is a positive value of t for which $|\phi(t)| = 1$, then F(x) must be a lattice distribution and $|\phi(t)|$ must be periodic; and, conversely, if F(x) is a lattice distribution, then $|\phi(t)|$ must be periodic. Throughout this section, we suppose that F(x) is a d.f. for which (1.1) holds and $|\phi(t)|$ is periodic. Then $|\phi(t)|$ has a least positive period which we call 2T.

To estimate $C_n^{(4)}$ we start with a method employed in [2] for the case in which F(x) is the symmetric binomial d.f. From (2.7) we obtain

(6.1)
$$C_n^{(4)} = \frac{1}{n^{\frac{1}{2}}} \frac{1}{\pi} \sum_{k=1}^{\infty} \int_{2kT-T}^{2kT+T} |\phi(t)|^{2n} \frac{dt}{t^2}.$$

Since $|\phi(t)|$ has period 2T, this implies that

(6.2)
$$C_n^{(4)} = \frac{1}{n^{\frac{1}{2}}} \frac{1}{\pi} \int_{-T}^{T} |\phi(t)|^{2n} S(t) dt$$

where

$$S(t) = \sum_{k=1}^{\infty} (2kT + t)^{-2}.$$

Since $|\phi(-t)| = |\phi(t)|$, this implies that

(6.21)
$$C_n^{(4)} = \frac{1}{n^{\frac{1}{2}}} \frac{1}{\pi} \int_0^T |\phi(t)|^{2n} S_1(t) dt$$

where $S_1(t) = S(t) + S(-t)$ and hence

(6.22)
$$S_1(t) = \sum_{k=1}^{\infty} [(2kT+t)^{-2} + (2kT-t)^{-2}].$$

The function $S_1(t)$ is, as a function of a complex variable t, analytic except for simple poles at the points $\pm 2T$, $\pm 4T$, $\pm 6T$, \cdots and, as we shall see, it is an elementary function. From (6.22) we obtain

$$S_1(t) = \frac{d}{dt} \sum_{k=1}^{\infty} \left[(2kT - t)^{-1} - (2kT + t)^{-1} \right] = \frac{d}{dt} \frac{t}{2T^2} \sum_{k=1}^{\infty} \frac{1}{k^2 - (t/2T)^2}.$$

Using the standard formula

(6.31)
$$\sum_{k=1}^{\infty} \frac{1}{k^2 - z^2} = \frac{1 - \pi z \cot \pi z}{2z^2},$$

which is valid when z is not an integer, gives

(6.32)
$$S_1(t) = \frac{d}{dt} \left[\frac{1}{t} - \frac{\pi}{2T} \cot \frac{\pi t}{2T} \right].$$

From this we obtain

(6.33)
$$S_1(t) = (\pi/2T)^2[\csc^2 x - x^{-2}]$$

where $x = \pi t/2T$. Differentiating the ordinary power series expansion of $(\cot x - x^{-1})$ gives a representation of the right side as a power series in x. Putting $x = \pi t/2T$ in this power series gives

(6.34)
$$S_1(t) = \left(\frac{\pi}{2T}\right)^2 \left[\frac{1}{3} + \frac{1}{15} \left(\frac{\pi t}{2T}\right)^2 + \frac{2}{189} \left(\frac{\pi t}{2T}\right)^4 + \frac{1}{675} \left(\frac{\pi t}{2T}\right)^6 + \cdots\right].$$

The numerical coefficients in (6.34) have simple expressions in terms of Bernoulli numbers, and the expansion is valid when |t| < 2T. Differentiating (6.22) gives

(6.35)
$$S_1'(t) = 4t \sum_{k=1}^{\infty} \frac{12k^2T^2 + t^2}{(4k^2T^2 - t^2)^3}.$$

This shows that $S_1'(t) > 0$ when 0 < t < T and hence that $S_1(t)$ is increasing when $0 \le t \le T$. With the aid of (6.34) and (6.33), we see that

(6.36)
$$\frac{1}{3} \left(\frac{\pi}{2T} \right)^2 = S_1(0) \le S_1(t) \le S_1(T) = \frac{\pi^2 - 4}{\pi^2} \left(\frac{\pi}{2T} \right)^2$$

when $0 \le t \le T$. This shows that we could delete the factor $S_1(t)$ from the integrand in (6.21) without changing the order of magnitude of $C_n^{(4)}$.

We now improve the formula (6.21) by showing that

(6.4)
$$C_n^{(4)} = o(n^{-\omega}) + \frac{1}{n^{\frac{1}{2}}} \frac{1}{\pi} \int_0^{\delta_n} |\phi(t)|^{2n} S_1(t) dt$$

where $\delta_n = (\log n)/n^{\frac{1}{2}}$ as in (2.22). For this purpose, let

(6.41)
$$E_n = \frac{1}{n^{\frac{1}{2}}} \frac{1}{\pi} \int_{\delta_n}^{\tau} |\phi(t)|^{2n} S_1(t) dt.$$

Letting $\psi(t) = |\phi(t)|^2$ and using (6.36) gives, for some constant M,

(6.42)
$$E_n < M \int_{\delta_n}^T |\psi(t)|^n dt.$$

Since $\psi(t)$ is continuous over $0 \le t \le T$, $\psi(0) = 1$, $\psi'(0) = 0$, $\psi''(0) = -2$, and $0 \le \psi(t) < 1$ when $0 < t \le T$, it follows that, when n is sufficiently great,

(6.43)
$$\max_{\delta_n \le t \le T} |\psi(t)| = \psi(\delta_n) < 1 - \delta_n^2/2.$$

This and (6.42) show that

$$(6.44) E_n < MT[1 - \delta_n^2/2]^n = o(n^{-\omega}).$$

Hence (6.21) and (6.41) imply (6.4) and (6.4) is proved.

Our estimate of $C_n^{(4)}$ will come from (6.4). As in Section 3 we suppose that, for some integer m for which $m \ge 3$, the moments α_m and β_m exist. The formulas (3.3) and (3.31) are then valid and, supposing henceforth that $0 < t < \delta_n$, we can use the formula

(6.5)
$$e^{w} = 1 + \sum_{k=1}^{m} \frac{w^{k}}{k!} + O(w^{m+1})$$

to obtain a formula giving e^w as the sum of a finite number of terms involving n, t, and γ_3 , γ_4 , \cdots , γ_m . When this finite sum is written down, it is found that

(6.51)
$$e^{w} = u + iv, \qquad |e^{w}|^{2} = u^{2} + v^{2}$$

where

(6.52)
$$u = 1 + \frac{\gamma_4}{24} n t^4 - \frac{\gamma_3^2}{72} n^2 t^6 + \cdots,$$

$$(6.53) v = -\frac{\gamma_3}{6} n t^3 + \frac{\gamma_5}{120} n t^5 - \frac{\gamma_3 \gamma_4}{144} n^2 t^7 + \frac{\gamma_3^3}{1296} n^3 t^9 + \cdots,$$

and remarks analogous to those following (3.44) are applicable. From (6.4), (3.3), (6.51), and (6.34), we find that

(6.54)
$$C_n^{(4)} = o(n^{-\omega}) + \frac{1}{n^{\frac{1}{2}}} \frac{1}{\pi} \int_0^{\delta_n} e^{-nt^2} G_n(t) dt$$

where

(6.55)
$$G_n(t) = \left(\frac{\pi}{2T}\right)^2 \left[\frac{1}{3} + \frac{\pi^2}{60T^2}t^2 + \frac{\gamma_4}{36}nt^4 + \cdots\right].$$

Using the values in (3.54) of the integrals in (3.5) then gives

(6.6)
$$C_n^{(4)} = \frac{B_1}{n} + \frac{B_2^2}{n^2} + \cdots + \frac{B_k}{n_k} + O\left(\frac{1}{n^{k+1}}\right)$$

where

(6.61)
$$B_1 = \frac{1}{6\pi^{\frac{1}{2}}} \left(\frac{\pi}{2T}\right)^2,$$

(6.62)
$$B_2 = \frac{1}{60\pi^{\frac{1}{2}}} \left(\frac{\pi}{2T}\right)^4 + \frac{\gamma_4}{96\pi^{\frac{1}{2}}} \left(\frac{\pi}{2T}\right)^2,$$

and each of the constants B_1 , B_2 , B_3 , \cdots depends only upon T and a finite number of the semi-invariants γ_3 , γ_4 , \cdots . Unlike the constants A_1 , A_2 , \cdots in (3.6), the constant B_1 in (6.6) can never be zero. In case the given d.f F(x) has finite moments of all positive integer orders, the integer m can be chosen as great as we wish and (6.6) is then valid for each $k = 1, 2, 3, \cdots$.

Our results show that if F(x) has finite moments of all positive orders, and if the c.f. $\phi(t)$ is such that $|\phi(t)|$ is periodic, then (1.7) is valid when $D_k = A_k + B_k$ and the constants A_k and B_k are the constants in (3.6) and (6.6).

7. The global version of the De Moivre theorem on binomial distributions. Let 0 and let <math>F(x) be the binomial d.f., associated with the probability p, which has mean 0 and standard deviation 1. To simplify our formulas, we define two constants h and β related to p and to each other by the formulas

$$(7.1) h = [p/(1-p)]^{\frac{1}{2}}, p = h^2/(1+h^2),$$

(7.11)
$$\beta = (h + h^{-1})^{-1} = [p(1 - p)]^{\frac{1}{2}}.$$

A random variable ξ governed by F(x) has the value $-h^{-1}$ with probability p and the value h with probability (1-p). Hence F(x)=0 when $x< h^{-1}$, F(x)=p when $-h^{-1} \leq x < h$, and F(x)=1 when $x \geq h$. A classic theorem of De Moivre states that, under these conditions, (1.21) holds. In case $p=\frac{1}{2}$ and F(x) is symmetric, the constant C_n in (1.3) has been estimated in [2] and the result is given in (1.6). We now treat the general case and shall show that there exists constants D_1 , D_2 , \cdots , depending only upon p, such that

$$(7.2) C_n = D_1 n^{-1} + D_2 n^{-2} + \cdots + D_k n^{-k} + O(n^{-k-1})$$

for each $k = 1, 2, 3, \cdots$. Moreover

(7.21)
$$D_1 = \frac{1}{24\pi^{\frac{1}{2}}} \frac{1 + \left(\frac{1}{2} - p\right)^2}{p(1 - p)}.$$

It is a straightforward but tedious task to extend our work to obtain explicit formulas for D_2 and D_3 .

The definition of F(x) implies that F(x) has finite moments of all positive integer orders and hence that (3.6) is valid. From (2.01) and the definition of F(x) we obtain

(7.3)
$$\phi(t) = pe^{-ih^{-1}t} + (1-p)e^{iht} = e^{-ih^{-1}t}[p + (1-p)e^{i\beta^{-1}t}].$$

While $\phi(t)$ is not necessarily periodic, we see that

(7.31)
$$|\phi(t)| = |[p + (1-p)\cos\beta^{-1}t] + i[(1-p)\sin\beta^{-1}t]|^{\frac{1}{2}}$$

$$= [p^2 + (1-p)^2 + 2p(1-p)\cos\beta^{-1}t]^{\frac{1}{2}}$$

and hence that $|\phi(t)|$ has least period $2\pi\beta$. Therefore (6.6) is valid with $2T = 2\pi\beta$ and hence $(\pi/2T) = (2\beta)^{-1}$. Using the notation of (3.6) and (6.6), we see that (7.2) is valid when $D_k = A_k + B_k$.

To obtain (7.21), we use the formula $D_1 = A_1 + B_1$ where A_1 and B_1 are given by (3.61) and (6.61). Since $\gamma_3 = \alpha_3$ and

(7.4)
$$\alpha_3 = (-h^{-1})^3 p + h^3 (1-p) = h - h^{-1}$$

(7.41)
$$\alpha_3^2 = (h^2 + h^{-2} - 2) = (1 - 2p)^2 / p(1 - p)$$

we see from (3.61) that

$$A_1 = (24\pi^{\frac{1}{2}})^{-1}(\frac{1}{2} - p)^2/p(1 - p).$$

Since

$$(7.5) \qquad (\pi/2T)^2 = 1/4\beta^2 = 1/4p(1-p)$$

we see from (6.61) that

$$(7.51) B_1 = (24\pi^{\frac{1}{2}})^{-1}/p(1-p).$$

From (7.42) and (7.51) we obtain (7.21).

When $p = \frac{1}{2}$, the d.f. F(x) of this section reduces to the symmetric binomial or Bernoulli d.f. which we shall treat further in Section 9. In this case $\phi(t) = \cos t$, $|\phi(t)|$ has period π , and $2T = \pi$. With the aid of (3.21) we obtain $\gamma_3 = 0$ and $\gamma_4 = -2$. Hence (2.9), (3.6), (3.61), (3.62), (6.6), (6.61), and (6.62) give

(7.6)
$$C_n = \frac{1}{6\pi^{\frac{1}{2}}} \frac{1}{n} + \frac{3}{1280\pi^{\frac{1}{2}}} \frac{1}{n^2} + O\left(\frac{1}{n^3}\right).$$

or

$$(7.61) C_n = .04903 \ 15973n^{-1} + .00132 \ 23193n^{-2} + O(n^{-3}).$$

8. An inequality for C_n . Throughout this section we suppose that F(x) is a d.f. having a finite third absolute moment β_3 . It is known ([5], p. 201) that there is an absolute constant E_1 such that

In the left member of (8.1) we have the distance between $F_n(x)$ and $\Phi(x)$ in the space of bounded measurable functions defined over $-\infty < x < \infty$. It is not unreasonable to conjecture that, for some constant E_2 , a valid companion inequality is obtained by replacing E_1 by E_2 and replacing the left member of (8.1) by the distance C_n^{\dagger} between $F_n(x)$ and $\Phi(x)$ in the Lebesgue space $L_2(-\infty, \infty)$. While $F_n(x)$ and $\Phi(x)$ themselves cannot belong to the space L_2 , we know that $F_n(x) - \Phi(x)$ belongs to the space L_2 whenever F(x) has finite second moments and hence whenever F(x) has finite third moments. Thus we conjecture that there is an absolute constant E_2 such that

(8.2)
$$C_n^{\frac{1}{2}} \leq E_2 \beta_3 n^{-\frac{1}{2}} \qquad (n = 1, 2, 3, \cdots).$$

To eliminate the fractional exponents, we write the conjecture (8.2) in the form

$$(8.3) C_n \le E_2^2 \beta_3^2 n^{-1}.$$

Evidence that the right member of (8.3) involves β_3 and n in the correct way is obtained by examining the manner in which C_n depends upon β_3 when F(x) is the binomial d.f. of Section 7. From (7.2) we obtain $C_n \sim D_1 n^{-1}$ where D_1 is defined by (7.21). From the definition of F(x) in section 7, we find that

(8.4)
$$\beta_3 = [p^2 + (1-p^2)][p(1-p)]^{-\frac{1}{2}}.$$

Squaring (8.4) and using the result in (7.21) gives

$$(8.5) D_1 = (24\pi^{\frac{1}{2}})^{-1}Q(p)\beta_3^2$$

where

$$Q(p) = \left[1 + \left(\frac{1}{2} - p\right)^2\right] \left[p^2 + \left(1 - p\right)^2\right]^{-2}.$$

In the range $0 where p must lie, we have <math>5/4 < Q(p) \le 4$. Moreover $Q(\frac{1}{2}) = 4$. Thus for the binomial d.f. of Section 7, we have

(8.6)
$$C_n \sim (24\pi^{\frac{1}{2}})^{-1}Q(p)\beta_3^2 n^{-1}$$

where $5/4 < Q(p) \le 4$.

While the conjecture involving (8.3) and (8.2) remains unproved, the above estimates show that if (8.3) and (8.2) are universally valid, then

$$(8.7) E_2^2 \ge (6\pi^{\frac{1}{2}})^{-1} = .09403 \ 15973$$

and

$$(8.71) E_2 \ge (6\pi^{\frac{1}{2}})^{-\frac{1}{2}} = .30664 57195$$

9. The symmetric binomial or Bernoulli d.f. Let F(x) be the symmetric binomial or Bernoulli d.f., this being the d.f. of Section 7 with $p = \frac{1}{2}$. This d.f. is commonly associated with problems in coin tossing. It is the purpose of this section to obtain precise information about the constants C_n defined by (1.3). We shall focus our attention upon the formula

$$(9.01) C_n = \frac{D_1}{n} + \frac{D_2}{n^2} + \frac{D_3}{n^3} + \frac{D_4}{n^4} + \frac{D_5}{n^5} + \frac{R_n}{n^6}$$

where D_1 , D_2 , \cdots are the constants in the asymptotic expansion of C_n and the numbers R_1 , R_2 , \cdots are determined by the formula (9.01) itself. Of course the theory of asymptotic expansions assures us that the result of neglecting R_n in the right member of (9.01) gives a good approximation to C_n when n is sufficiently great, but until the matter has been investigated we do not know whether the approximation will be good when n is 5 or 10 or 100.

In the first place, the numerical values of D_1 , D_2 , \cdots , D_5 can be calculated by the methods of Sections 3 and 6. The details of the calculations are quite lengthy and tedious even when full advantage is taken of the fact that the c.f. $\phi(t)$ is now the real periodic function cos t. The right side of (3.2) can be replaced by the known power series expansion of log cos t which is obtained by

integrating the expansion of $\tan t$. With the notation of (3.3), w is real and $|e^w - 1|^2$ can be replaced by $e^{2w} - 2e^w + 1$. Since $|\cos t|$ has period π , the constant 2T of Section 6 is π . It is found that each of the constants D_1 , D_2 , \cdots is a rational multiple of $\pi^{-\frac{1}{2}}$ and that

$$(9.02) D_1 = \frac{1}{6} \pi^{-\frac{1}{2}} = .09403 \ 15972 \ 57959$$

$$(9.03) D_2 = \frac{3}{1280} \pi^{-\frac{1}{2}} = .00132 \ 23193 \ 36440$$

$$(9.04) D_3 = \frac{-397}{258048} \pi^{-\frac{1}{2}} = -.00086 79907 01995$$

$$(9.05) D_4 = \frac{53461}{353 \quad 89440} \pi^{-\frac{1}{2}} = .00085 \quad 22920 \quad 77129$$

$$(9.06) D_5 = \frac{23}{20761} \frac{24491}{80480} \pi^{-\frac{1}{2}} = .00063 \quad 16664 \quad 76919$$

Only one of these five constants is negative, and the author has very little information about D_n when n > 5.

In order to obtain information about the numbers R_1 , R_2 , \cdots in (9.01) the values of C_1 , C_2 , \cdots , C_{10} in (9.1) were calculated by the method which is explained later in this section.

$$C_1 = .10244 \ 13576 \qquad \Gamma_1 = .09506 \ 98844$$

$$C_2 = .04706 \ 47193 \qquad \Gamma_2 = .04729 \ 68250$$

$$C_3 = .03147 \ 89023 \qquad \Gamma_3 = .03147 \ 05293$$

$$C_4 = .02358 \ 02730 \qquad \Gamma_4 = .02358 \ 07083$$

$$C_5 = .01885 \ 37826 \qquad \Gamma_5 = .01885 \ 37765$$

$$C_6 = .01570 \ 53613 \qquad \Gamma_6 = .01570 \ 53651$$

$$C_7 = .01345 \ 79250 \qquad \Gamma_7 = .01345 \ 79258$$

$$C_8 = .01177 \ 31393 \qquad \Gamma_8 = .01177 \ 31395$$

$$C_9 = .01046 \ 32283 \qquad \Gamma_9 = .01046 \ 32284$$

$$C_{10} = .00941 \ 56055 \qquad \Gamma_{10} = .00941 \ 56056$$

The exact value of C_1 is

$$(9.11) C_1 = (2/\pi e)^{\frac{1}{2}} + \psi(1) - \pi^{-\frac{1}{2}} - \frac{1}{2}$$

where $\psi(x)$ is the tabulated Gaussian function defined by (9.29) below, and

$$(9.12) C_1 = .10244 13576 27616$$

with uncertainty only in the last figure which should perhaps be 7. For our next step, the ten-decimal approximations given in (9.1) are scarcely adequate. With the aid of the values of C_1 , \cdots , C_{10} which were rounded to obtain the values in (9.1), it is possible to calculate the numbers R_1 , R_2 , \cdots , R_{10} in (9.01). It is found that the numbers R_5 , R_6 , R_7 , R_8 , R_9 , and R_{10} differ relatively little from -.0009 and some heuristic considerations suggest very strongly that R_n differs from -.0009 by less than .00018 when $n \ge 5$. This in turn suggests that the constant Γ_n defined by

(9.13)
$$\Gamma_n = \frac{D_1}{n} + \frac{D_2}{n^2} + \frac{D_3}{n^3} + \frac{D_4}{n^4} + \frac{D_5}{n^5} - \frac{.0009}{n^6}$$

must be a very good approximation to C_n at least when $n \geq 5$, and that

$$|\Gamma_n - C_n| < .00018 \, n^{-6} \qquad (n \ge 5).$$

The values of Γ_1 , Γ_2 , \cdots Γ_{10} calculated from (9.13) are given in (9.1), and it is easy to see how Γ_n compares with C_n when $1 \le n \le 8$. When n is 9 or 10, rounding errors obscure the relationships. After the above results were obtained, the values of Γ_{16} and C_{16} were obtained correct to 15 decimal places. The values are

$$\Gamma_{16} = .00588 \ 19417 \ 80443$$

$$(9.16) C_{16} = .00588 19417 81902,$$

and the agreement is neither better nor worse than was expected. It thus appears that C_n has an exceptionally useful asymptotic expansion and that, for example, use of (9.13) gives

$$(9.17) C_{100} = .00094 \ 04473 \ 45108$$

where the result is correct to the full 15 decimals. It would seem to be a formidable task to obtain even a crude approximation to C_{100} by direct computation of C_{100} .

We now proceed to obtain the formulas from which the numbers C_1 , C_2 , \cdots , C_{10} in (9.1) were calculated. Let $H_n = nC_n$ so that $C_n = H_n/n$. Since $\phi(t) = \cos t$, we find from (2.1) that

$$(9.2) H_n = n^{\frac{1}{2}} \frac{1}{\pi} \int_0^\infty |e^{-nt^2/2} - \cos^n t|^2 t^{-2} dt.$$

According to R. J. Walker, it is not desirable to undertake to calculate H_1 , H_2 , ... by direct application of a computing machine to the right member of (9.2); it is better to use the following way of expressing H_n as a finite sum of terms which are tabulated or easily calculated. From (9.2) we obtain

(9.21)
$$H_n = n^{\frac{1}{2}} \pi^{-1} [2R_n - P_n - Q_{2n}]$$

where

$$(9.22) P_n = \int_0^\infty \frac{1 - e^{-nt^2}}{t^2} dt, Q_n = \int_0^\infty \frac{1 - \cos^n t}{t^2} dt$$

(9.23)
$$R_n = \int_0^\infty \frac{1 - e^{-nt^2/2} \cos^n t}{t^2} dt.$$

We shall show that, for each $n = 1, 2, \dots$

$$(9.24) P_n = (n\pi)^{\frac{1}{2}},$$

$$(9.25) Q_{2n-1} = Q_{2n} = \frac{n\pi}{2^{2n}} \binom{2n}{n} = \frac{\pi}{2} \frac{3 \cdot 5 \cdot 7 \cdot \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdot \cdots (2n-2)},$$

and

$$(9.26) R_n = S_n + T_n$$

where

$$S_n = \frac{(2n\pi)^{\frac{1}{2}}}{2^{n+1}} \sum_{k=0}^n \binom{n}{k} e^{-(n-2k)^2/2n},$$

$$(9.28) T_n = \frac{\pi}{2^{n+1}} \sum_{k=0}^n \binom{n}{k} |n-2k| \psi(|n-2k| n^{-\frac{1}{2}}),$$

and $\psi(x)$ is the thoroughly tabulated Gaussian function

(9.29)
$$\psi(x) = (2\pi)^{-\frac{1}{2}} \int_{-x}^{x} e^{-t^{2}/2} dt.$$

In (9.27) and (9.28), $\sum_{k=0}^{n}$ can be replaced by $2\sum_{k\leq n/2}^{*}$ where the star on the \sum signifies that when n is even the term for which k=n/2 is to be divided by 2. The numbers H_n are calculated from (9.21) with the aid of (9.24), (9.25), and (9.26). We shall omit these calculations and hence it remains only for us to establish (9.24), (9.25), and (9.26).

Starting with (9.22) and using standard integral formulas gives

$$(9.3) P_n = \int_0^\infty dt \int_0^n e^{-xt^2} dx = \int_0^n dx \int_0^\infty e^{-xt^2} dt = \pi^{\frac{1}{2}} 2^{-1} \int_0^n x^{-\frac{1}{2}} dx = (n\pi)^{\frac{1}{2}}$$

and (9.24) is established.

We now establish (9.25) by a method which exhibits material we shall use to establish the more complicated formula (9.26). Using the Euler formula for $\cos t$ and the binomial formula we obtain, when t is real,

(9.4)
$$\cos^{n} t = 2^{-n} (e^{it} + e^{-it})^{n}$$

$$= \frac{1}{2^{n}} \sum_{k=0}^{n} \binom{n}{k} e^{i(n-2k)t} = \frac{1}{2^{n}} \sum_{k=0}^{n} \binom{n}{k} \cos(n-2k)t$$

and hence

(9.41)
$$1 - \cos^n t = \frac{1}{2^n} \sum_{k=0}^n \binom{n}{k} [1 - \cos(n-2k)t].$$

Putting this in the second of the formulas (9.22) and using a standard integral formula gives

(9.42)
$$Q_n = \frac{\pi}{2^{n+1}} \sum_{k=0}^n \binom{n}{k} |n-2k|.$$

While other proofs of (9.25) may be more elegant, we dismiss the matter with the remark that it is not difficult to use (9.42) to prove (9.25) by induction.

To establish (9.26) we suppose that n is a fixed positive integer, put

(9.5)
$$G(x) = \int_{0}^{\infty} \left[1 - e^{-xt^{2}/2} \cos^{n} t\right] t^{-2} dt,$$

and observe that $G(n) = R_n$ and $G(0) = Q_n$. Differentiating (9.5) gives

(9.51)
$$G'(x) = \frac{1}{2} \int_0^\infty e^{-xt^2/2} \cos^n t \, dt.$$

Use of (9.4) gives

(9.52)
$$G'(x) = \frac{1}{2^{n+1}} \sum_{k=0}^{n} \binom{n}{k} \int_{0}^{\infty} e^{-xt^{2}/2} \cos(n-2k)t \, dt$$

and use of a standard integral formula then gives

$$(9.53) G'(x) = \frac{(2\pi)^{\frac{1}{2}}}{2^{n+2}} \sum_{k=0}^{n} {n \choose k} x^{-\frac{1}{2}} e^{-(n-2k)^{\frac{n}{2}/2x}}.$$

Defining I(m) by the formula

(9.54)
$$I(m) = \int_0^n x^{-\frac{1}{2}} e^{-m^2/2x} dx,$$

we use (9.53) and the fact that $G(n) = R_n$ and $G(0) = Q_n$ to obtain

(9.6)
$$R_n = Q_n + \frac{(2\pi)^{\frac{1}{2}}}{2^{n+2}} \sum_{k=0}^n \binom{n}{k} I(n-2k).$$

Our next step is to obtain a better formula for I(m). Suppose first that $m \neq 0$. A change of the variable of integration in (9.54) gives

$$(9.61) I(m) = 2 | m | \int_{|m| \, n^{-1}}^{\infty} t^{-2} e^{-t^{2}/2} \, dt.$$

Using the well known formula

(9.62)
$$\int_{a}^{\infty} e^{-t^{2}/2} dt = a^{-1} e^{-a^{2}/2} - \int_{a}^{\infty} t^{-2} e^{-t^{2}/2} dt,$$

which is easily derived by intergration by parts, gives

$$(9.63) I(m) = 2n^{\frac{1}{2}}e^{-m^{2}/2n} - 2 \mid m \mid \int_{|m| n^{-\frac{1}{2}}}^{\infty} e^{-t^{2}/2} dt.$$

In case m=0, an easy evaluation of the right members of (9.54) and (9.63) shows that (9.63) is still valid. Substituting (9.63) in (9.6) and using (9.42) gives (9.26). This completes the derivations of the formulas used to obtain numerical values of C_1 , \cdots , C_{10} and C_{16} . The tables of the exponential and probability functions put out by the U.S. National Bureau of Standards were used.

While our work does not actually prove the result, it indicates very strongly that the sequence C_1 , C_2 , C_3 , \cdots converges monotonically to zero and hence that, in the mean square sense, each one of the distribution functions $F_1(x)$, $F_2(x)$, $F_3(x)$, \cdots is more nearly Gaussian than its predecessors. There was a time when the author rather expected that the sequence H_1 , H_2 , H_3 , \cdots defined by $H_n = nC_n$ would also be monotone, but it turns out that this is not so. In fact

(9.7)
$$H_1 = .10244 \ 136 \qquad H_3 = .09443 \ 671$$

$$H_2 = .09412 \ 944 \qquad H_4 = .09432 \ 109$$

As a check upon the value of H_2 and upon the relative values of H_1 , H_2 , and H_4 , the author, at that time, calculated H_1 , H_2 , and H_4 by a method completely independent of the calculations of this section and of the theories upon which they are based. By use of the distribution functions $F_1(x)$, $F_2(x)$, $F_4(x)$ and the formula (1.3) itself, the constants C_1 , C_2 , and C_4 were calculated by use of the Simpson parabolic formula for approximate evaluation of integrals. The resulting values of H_1 , H_2 , and H_4 were found to agree to 6 decimal places with the values in (9.7).

REFERENCES

- R. P. Agnew, "Global versions of the central limit theorem," Proc. Nat. Acad. Sci. (USA), Vol. 40 (1954), pp. 800-804.
- [2] R. P. Agnew, "Estimates for global central limit theorems," Ann. Math. Stat., Vol. 28 (1957), pp. 26-42.
- [3] HAROLD CRAMÉR, Random Variables and Probability Distributions, Cambridge Univ. Press, 1937, 121 pp.
- [4] HAROLD CRAMÉR, Mathematical Methods of Statistics, Princeton Univ. Press, 1946, 575 pp.
- [5] C. G. Esseen, "Fourier analysis of distribution functions; a mathematical study of the Laplace-Gaussian law," Acta Mathematica, Vol. 77 (1945), pp. 1-125.
- [6] B. V. GNEDENKO AND A. N. KOLMOGOROV, Limit Distributions for Sums of Independent Random Variables, (Translation from Russian by K. L. Chung) Cambridge Univ. Press, 1954, 264 pp.
- [7] T. N. THIELE, Theory of Observations, London, 1903, 143 pp.