ASYMPTOTIC EXPANSIONS ASSOCIATED WITH THE nth POWER OF A DENSITY[1]

By R. A. Johnson

University of Wisconsin

1. Introduction and summary. Let $\{X_n\}_{n=1}^{\infty}$ be the sequence of random variables associated with the sequence of densities of the form

$$C_n k(x) f^n(x) \qquad n = 1, 2, \cdots, \tag{1.1}$$

where k is a positive function and f has a unique mode m at which it is sufficiently smooth. It is known that $n^{\frac{1}{2}}(X_n - m)b$ converges in law to the standard normal distribution when b is a suitably chosen scaling constant (see Laplace (1847), pages 400–403, or von Mises (1964), page 269). In Section 2, it is shown that the cumulative distribution function F_n possesses an asymptotic expansion in powers of $n^{-\frac{1}{2}}$ where each coefficient is the product of a polynomial and the standard normal density. The polynomials have coefficients which are expressible in terms of k and $\log f$, together with their derivatives evaluated at m.

Section 3 shows that the normalizing transformation $\eta_n(\xi) = \Phi^{-1}(F_n(\xi))$, where Φ is the standard normal cdf, also has an asymptotic expansion in powers of $n^{-\frac{1}{2}}$ and Section 4 makes the same conclusion for the percentiles of F_n. The coefficients in each of these expansions are polynomials. Similar theorems are given by Bol'shev (1959), (1963), Dorogovcev (1962), Peiser (1949), and Wasow (1956).

Examples of these expansions, namely the t-distribution and central order statistics, are given.

1.1. *General assumptions.* Consideration of the random variables $b(X_n - m)$ where $b^2 = -f''(m)/f(m)$ shows that it is possible, without loss of generality, to specialize to the case where $m = 0$, $f(0) = 1$, $f'(0) = 0$, and $f''(0) = -1$. The general assumptions are stated for this case.

Assumption (i). $f(x)$ and $k(x)$, considered as functions of a complex variable, are analytic for $|x| \leqq \delta_1$ where δ_1 is a positive constant.

Assumption (ii). $f(x)$ has an absolute maximum at $x = 0$ and $f(x) \leqq \rho_1 < 1$ for all real x with $|x| \geqq \delta_1$.

We further assume that $k(0) \neq 0$ and also that $f(x) \neq 0$ whenever $|x| \leqq \delta_1$. These assumptions are chosen for simplicity and certainly are not the most general under which Laplace's approximation provides an asymptotic expansion.

1.2. *Notation.* Let F_n denote the cdf of $n^{\frac{1}{2}} X_n$ where X_n has a density satisfying the general assumptions. Denote by Φ and φ the standard normal cdf and density

Received 10 January 1967.

[1] Research sponsored in part by the National Science Foundation under Grant No. GP 3816. Part of this work appears in the author's Ph.D. dissertation, University of Minnesota.

respectively. For any $0 < \alpha < 1$, let ξ_n denote the upper αth percentile of F_n. It follows from the general assumptions that ξ_n is uniquely determined if n is sufficiently large. When considering percentiles, we shall assume that this is the case.

The notation $\sim$ is used rather than $=$ for relating a function to its asymptotic series since the series may or may not converge.

2. An asymptotic expansion for F_n. The procedure used to obtain an expansion of F_n is to consider $F_n(\xi)$ as being the ratio

$$(2.1)\qquad n^{\frac{1}{2}} \int_{-\infty}^{\xi n^{-\frac{1}{2}}} k(x)f^n(x)\,dx / n^{\frac{1}{2}} \int_{-\infty}^{\infty} k(x)f^n(x)\,dx$$

to first find asymptotic expansions for the numerator and denominator separately and then to form the quotient expansion.

The first part of this development, that given in Section 2, Lemma 2.1 through Theorem 2.1, is essentially the verification of Laplace's approximation (see Laplace (1847), Book One, Section 22–27) given by de Bruijn (1961), Chapter 4, and consequently proofs are omitted.

2.1. *Expansion of* $\int kf^n$. Under the general assumptions, it is possible to determine a $\rho > 0$ and $\delta > 0$ with $\delta < \min(1, \delta_1/2)$, so that not only are the above assumptions satisfied, but in addition, we have $\log f(x) \leqq -x^2/4$ for $-\delta \leqq x \leqq \delta$. This last inequality enables us to establish the following lemmas.

LEMMA 2.1. *For every positive integer* M,

$$\{\int_{-\infty}^{-\delta} + \int_{\delta}^{\infty}\} kf^n = O(n^{-M}) \qquad (n > 1).$$

LEMMA 2.2.

$$\{\int_{-\delta}^{-n^{-\frac{1}{3}}} + \int_{n^{-\frac{1}{3}}}^{\delta}\} kf^n = O(exp\,(-n^{\frac{1}{3}}/4)) \qquad (n > \delta^{-3}).$$

On the remaining interval, $\log f$ has the expansion $\sum a_s x^s$ which converges for $|x| \leqq 2\delta$. The principal branch of the log has been selected so that substitution for $f(0)$, $f'(0)$ and $f''(0)$ gives $a_0 = a_1 = 0$ and $a_2 = -\frac{1}{2}$. Define a function ψ by $\psi(z) = \sum_{s=3}^{\infty} a_s z^{s-3}$ where

$$(2.2)\qquad a_s = d^s \log f(x)/dx^s \,|_{x=0}/s! \quad \text{for } s = 3, 4, \cdots.$$

For $n > \delta^{-3}$ and $|x| \leqq n^{-\frac{1}{3}}$, we write

$$(2.3)\qquad k(x)f^n(x) = \exp(-nx^2/2)[k(x) \exp(nx^3 \textstyle\sum_{s=3}^{\infty} a_s x^{s-3})]$$

and regard the first factor on the right hand side of (2.3) as being the main factor and the second as being a particular evaluation of the function $P(w, z) = k(z) \exp[w\psi(z)]$ which is analytic in the region $\{|w| \leqq 2, |z| \leqq 2\delta\}$. Therefore

$$(2.4)\qquad P(w, z) = \sum_{l,m=0}^{\infty} c_{lm} w^l z^m$$

where

$$(2.5)\qquad l!\,m!\,c_{lm} = \partial^{l+m} P(w, z)/\partial^l w \partial^m z \,|_{w=0,z=0} \quad \text{each} \quad l, m = 0, 1, 2, \cdots,$$

and the series converges absolutely and uniformly in the region $\{|w| \leqq 1, |z| \leqq \delta\}$ (see Fuks (1963), pages 39–40, or Markushevich (1965), pages 101–105). Denote the truncated series $\sum_{l+m \leqq N} c_{lm} w^l z^m$ by $P_N(w, z)$. The integrand kf^n is approximated by $\exp(-nx^2/2) P_N(nx^3, x)$.

LEMMA 2.3. *For every positive integer* N, *there exists a constant* A_1 *such that*

$$\int_{-n^{-\frac{1}{3}}}^{n^{-\frac{1}{3}}} e^{-nx^2/2} |P(nx^3, x) - P_N(nx^3, x)| \, dx \leqq A_1 n^{-\frac{1}{2}N-1} \quad \text{for} \quad n > \delta^{-3}.$$

We now state a lemma which justifies the integration of the approximation over the whole real line.

LEMMA 2.4. *For each fixed positive integer* N,

$$\{\int_{-\infty}^{n^{-\frac{1}{3}}} + \int_{n^{-\frac{1}{3}}}^{\infty}\} |P_N| \, e^{-nx^2/2} \, dx = O(n^N \exp(-n^{\frac{1}{3}}/4) \qquad (n > 1).$$

From the above, we obtain the asymptotic series for the normalization constant.

THEOREM 2.1.

$$\int_{-\infty}^{\infty} k(x) f^n(x) \, dx \sim \sum_{j=0}^{\infty} \beta_j n^{-\frac{1}{2}(j+1)} \qquad (n > 1)$$

where

$$(2.6) \qquad \beta_j = 0, \quad \text{for } j \text{ odd}$$
$$= 2^{\frac{1}{2}(j+1)} \sum_{r=0}^{j} c_{r,j-r} 2^r \Gamma(r + j/2 + \tfrac{1}{2}), \quad \text{for } j \text{ even.}$$

The c_{lm}*'s are given by* (2.5).

Throughout this section, it is assumed that n goes to infinity through the integers, so that the expressions hold for $n \geqq 1$ if the bounding constants are suitably modified.

Upon integration of $P_N \exp(-nx^2/2)$ over $(-\infty, \xi n^{-\frac{1}{2}})$, we obtain the following result.

THEOREM 2.2.

$$\int_{\infty}^{\xi n^{-\frac{1}{2}}} kf^n \sim \sum_{j=0}^{\infty} \alpha_j(\xi) n^{-\frac{1}{2}(j+1)} \qquad (n \geqq 1)$$

where

$$(2.7) \quad \alpha_j(\xi) = \sum_{r=0}^{j} c_{r,j-r} \int_{-\infty}^{\xi} y^{2r+j} e^{-y^2/2} \, dy \quad \text{for each} \quad j = 0, 1, 2, \cdots,$$

and the c_{lm}*'s are given by* (2.5). *Here the error in using any finite sum is uniform in* ξ.

PROOF. For fixed but arbitrary integer N, consider the error in using the approximation when ξ is arbitrary but fixed. The error is bounded by

$$\{\int_{-\infty}^{-n^{-\frac{1}{3}}} + \int_{n^{-\frac{1}{3}}}^{\infty}\} [k(x) f^n(x) + |P_N| \, e^{-nx^2/2}] \, dx + \int_{-n^{-\frac{1}{3}}}^{n^{-\frac{1}{3}}} |k(x) f^n(x) - P_N e^{-nx^2/2}| \, dx$$

which, according to Lemmas 2.1–2.4, is $O(n^{-\frac{1}{2}(N+2)})$. Integrating the approximation and combining terms having the same power of $n^{-\frac{1}{2}}$, we establish the result. The change of variable $y = n^{\frac{1}{2}x}$ has been made so as to obtain (2.7).

We now obtain the desired asymptotic expansion of F_n.

THEOREM 2.3. *Under the general assumptions*, $F_n(\xi)$ *admits the expansion*

$$(2.8) \qquad F_n(\xi) \sim \Phi(\xi) + \sum_{j=1}^{\infty} \gamma_j(\xi) n^{-j/2} \qquad (n \geqq 1)$$

uniformly in ξ. *Each* γ_j *is the product of a polynomial and* φ.

PROOF. The assumption $k(0) \neq 0$ implies that $\alpha_0(\xi) = (2\pi)^{\frac{1}{2}}k(0)\Phi(\xi)$ and $\beta_0 = (2\pi)^{\frac{1}{2}}k(0)$ are different from zero. It follows from argument given by Erdelyi (1956), Chapter 1, that the asymptotic expansion for a quotient such as (2.1) exists, and the coefficients may be obtained by formal substitution. The uniform property is clear if we first find the expansion

$$(n^{\frac{1}{2}} \int_{-\infty}^{\infty} kf^n)^{-1} \sim \sum \beta_j' n^{-j/2}$$

and then multiply and collect terms, recalling that the numerator expansion is uniform for all ξ and noting from (2.7) that each $\alpha_j(\xi)$ is bounded.

The sequence $\{\gamma_j(\xi)\}_{j=0}^{\infty}$, obtained by formal division of the two asymptotic series, satisfies

$$\alpha_j(\xi) = \sum_{r=0}^{j} \gamma_r(\xi)\beta_{j-r} \quad \text{each} \quad j = 0, 1, 2, \cdots, \tag{2.9}$$

where β_j and $\alpha_j(\xi)$ are given by (2.6) and (2.7). In Section 2.2, it will be shown that $\gamma_0 = \Phi$ and that γ_1 is of the form asserted. We proceed by induction.From (2.9), it is sufficient to prove that $\alpha_j(\xi) - \beta_j\gamma_0(\xi)$ does not have a term involving Φ. If j is odd, this result is clear from the expressions (2.6) and (2.7) for β_j and $\alpha_j(\xi)$. When j is even, it suffices to show that

$$\int_{-\infty}^{\xi} y^{2r+j} e^{-y^2/2}\, dy - \Phi(\xi) 2^{r+j/2+\frac{1}{2}}\Gamma(r + j/2 + \tfrac{1}{2})$$

does not involve Φ. Repeated integration by parts confirms this result.

2.2. *Calculation of the first few terms.* Focusing our attention on the integrals $\int_{-\infty}^{\xi} y^s e^{-y^2/2}\, dy$ which enter the γ_j through the α_j's, we see that there is some arbitrariness as to their expression in terms of known functions. Upon repeated integration by parts, we express the integrals in terms of $\Phi(\xi)$ and polynomials multiplied by φ.

By straightforward calculation from (2.9), we find the first four terms of the expansion:

$$\begin{aligned}
\gamma_0(\xi) &= \Phi(\xi); \qquad \gamma_1(\xi) = -\varphi(\xi)c_{00}^{-1}[c_{10}\xi^2 + 2) + c_{01}];\\
\gamma_2(\xi) &= -\varphi(\xi)c_{00}^{-1}[c_{20}\xi^5 + (5c_{20} + c_{11})\xi^3 + (15c_{20} + 3c_{11} + c_{02})\xi];\\
\gamma_3(\xi) &= -\varphi(\xi)c_{00}^{-1}\{c_{30}\xi^8 + (8c_{30} + c_{21})\xi^6 + (48c_{30} + 6c_{21} + c_{12})\xi^4\\
&\quad + [192c_{30} + 24c_{21} + 4c_{12} + c_{30} - c_{10}c_{00}^{-1}(c_{02} + 3c_{11} + 15c_{20})]\xi^2\\
&\quad + [384c_{30} + 48c_{21} + 8c_{12} + 2c_{03} - c_{00}^{-1}(2c_{10} + c_{01})\\
&\qquad \cdot (c_{02} + 3c_{11} + 15c_{20})]\}.
\end{aligned} \tag{2.10}$$

The coefficients which enter the $\gamma_j(\xi)$ may be expressed as functions of the derivatives of k and $\log f$. The coefficients c_{lm} are defined by (2.5) and by the definition of P, they can be seen to be functions of k and ψ together with their derivatives. Since $\psi(z) = \sum a_s z^{s-3}$ where a_s is given by (2.2), we may ultimately express each c_{lm} as a function of k and $\log f$ together with their derivatives. Table 2.1 gives these expressions for the first few c_{lm}.

The entry in row l column m is c_{lm}. For example, $c_{02} = k''(0)/2$. In the table, $k'' = k''(0)$ etc.

TABLE 2.1
Coefficients c_{lm} in terms of derivatives of $k(\cdot)$ and $\log f(\cdot)$

	m			
l	0	1	2	3
0	k	k'	$k''/2$	$k'''/6$
1	ka_3	$ka_4 + k'a_3$	$ka_5 + k'a_4 + k''a_3/2$	
2	$ka_3^2/2$	$ka_3a_4 + k'a_3^2/2$		
3	$ka_3^3/6$			

3. Normalizing transformation. Let $\eta_n(\xi) = \Phi^{-1}(F_n(\xi))$. That is, $\eta_n(\xi)$ is the solution of the equation

$$\Phi(\eta) = F_n(\xi). \tag{3.1}$$

From knowledge of the expansion for F_n as given in Theorem 2.3 and the results of Section 2.1, it is possible to conclude from the proof given by Wasow (1956), page 255, that $\eta_n(\xi)$ possesses an asymptotic expansion which is uniform in every finite ξ interval.

THEOREM 3.1. *Under the general assumptions, equation* (3.1) *possesses a solution* $\eta_n(\xi)$ *which admits an asymptotic expansion of the form*

$$\eta_n(\xi) \sim \xi + \textstyle\sum_{j=1}^{\infty} \omega_j(\xi)n^{-j/2} \qquad (n \to \infty) \tag{3.2}$$

uniformly in every finite ξ interval. The functions ω_j are polynomials.

The terms are obtained by formal substitution of (2.8) into the series of Φ^{-1}. In particular, we have

$$\omega_1 = \gamma_1/\varphi \quad \text{and} \quad \omega_2 = (\gamma_2/\varphi) + \xi(\gamma_1^2/2\varphi^2). \tag{3.3}$$

The coefficients may be expressed in terms of the c_{lm} through the expressions (2.10).

4. Expansion of the percentiles of F_n. For any $0 < \alpha < 1$, let η and $\xi_n(\eta)$ denote the upper αth percentiles of Φ and F_n respectively. That is $\Phi(\eta) = 1 - \alpha$ and $\Phi(\eta) = F_n(\xi_n(\eta))$. Recall that the basic assumptions imply that $F_n'(\xi_n(\eta)) > 0$ for all sufficiently large n and η in any bounded interval. Therefore $\xi_n(\eta)$ is uniquely determined and $\lim_{n\to\infty} \xi_n(\eta) = \eta$. The expansion for $\xi_n(\eta)$ is obtained by inverting (3.2). The argument given by Wasow (1956), pages 256–257, established the following theorem. For an alternative approach, see Johnson (1966).

THEOREM 4.1. *Under the general assumptions, $\xi_n(\eta)$ admits an asymptotic expansion of the form*

$$\xi_n(\eta) \sim \eta + \textstyle\sum_{j=1}^{\infty} \tau_j(\eta)n^{-j/2} \qquad (n \to \infty), \tag{4.1}$$

uniformly in every finite η interval. The functions τ_j are polynomials.

The coefficients may be obtained by formally substituting the expansion (4.1) into (3.2), setting the coefficients of each power of $n^{-\frac{1}{2}}$ equal to zero and solving.

In particular, we have the following expressions for τ_1 and τ_2 in terms of the c_{lm} which are defined in (2.5).

$$\tau_1 = c_{00}^{-1}(c_{10}\eta^2 + 2c_{10} + c_{01}) \tag{4.2}$$

and

$$\begin{aligned}\tau_2 = {} & (5c_{20}c_{00}^{-1} + c_{11}c_{00}^{-1} - c_{01}c_{10}c_{00}^{-2})\eta^3 \\ & + (2c_{10}^2c_{00}^{-2} - c_{01}^2c_{00}^{-2}/2 + 15c_{20}c_{00}^{-1} + 3c_{11}c_{00}^{-1} + c_{02}c_{00}^{-1})\eta.\end{aligned} \tag{4.3}$$

To obtain the expression (4.3), it is necessary to employ the relationship between c_{10} and c_{20} as given in Table 2.1 to show that the coefficient of η^5 is zero.

We remark that $\xi_n^N(\eta) = \eta + \sum_{j=1}^N \tau_j(\eta)n^{-j/2}$ is an approximate upper αth percentile for F_n when η is the upper αth percentile of Φ. This follows directly upon expanding F_n about $\xi_n(\eta)$ and then noting that F_n' is bounded in some neighborhood of η for all sufficiently large n.

THEOREM 4.2. *Under the basic assumptions with $\xi_n^N(\eta)$ defined above,*

$$F_n(\xi_n^N(\eta_\alpha)) = 1 - \alpha + O(n^{-\frac{1}{2}(N+1)}) \qquad (n \to \infty) \quad \textit{for each} \quad N = 1, 2, \cdots.$$

5. Examples. In this section, we shall consider two examples. Others may be found in Buehler (1965).

Let Z_n be the $(\lambda n + 1)$st order statistic from a sample of size $(\lambda + \mu)n + 1$ where λ and μ denote positive integers. Let G and g be the population cdf and pdf respectively. Assume that (1) there exists a z_0 such that $G(z_0) = \lambda/(\lambda + \mu)$ with $g(z_0) > 0$ and (2) g is analytic in some neighborhood of z_0. Then $X_n = (Z_n - z_0)g(z_0)b_2^{\frac{1}{2}}$, where $b_2 = (\lambda + \mu)^2(\lambda^{-1} + \mu^{-1})$, has a density which satisfies the general assumptions. Calculating the first few derivatives of $\log f$ and k and then the coefficients c_{lm}, we obtain from (2.10) the following terms of the expansion for F_n:

$$\begin{aligned}\gamma_1(\xi) &= \varphi(\xi)[(g'/2g^2b_2^{\frac{1}{2}} - b_3/3b_2^{\frac{3}{2}})\xi^2 - 2b_3/3b_2^{\frac{3}{2}}], \\ \gamma_2(\xi) &= \varphi(\xi)[\xi^5(12g^2g'b_2b_3 - 4g^4b_3^2 - 9g'^2b_2^2)/72g^4b_2^3 \\ &\quad + (g''/6g^3b_2 - 5b_3^2/18b_2^3 + b_4/4b_2^2)\xi^3 \\ &\quad + (3b_4/4b_2^2 - 5b_3^2/6b_2^3)\xi]\end{aligned} \tag{5.1}$$

where $b_j = (\lambda + \mu)^j[\lambda^{1-j} - (-\mu)^{1-j}]$ for $j = 2, 3, 4, \cdots$. It is understood that g and all its derivatives are all evaluated at z_0.

The normalizing transformation is found from Section 3 to have

$$(g'/2g^2b_2^{\frac{1}{2}} - b_3/3b_2^{\frac{3}{2}})\xi^2 - 2b_3/3b_2^{\frac{3}{2}}$$

as a first correction term ω_1.

The percentile has the first two correction terms:

$$\begin{aligned}\tau_1 &= (-g'/2g^2b_2^{\frac{1}{2}} + b_3/3b_2^{\frac{3}{2}})\eta^2 + 2b_3/3b_2^{\frac{3}{2}}, \\ \tau_2 &= (-g''/6g^3b_2 + 5b_3^2/18b_2^3 - b_4/4b_2^2 + g'^2/2g^4b_2 - g'b_3/3g^2b_2^2)\eta^3 \\ &\quad + (-3b_4/4b_2^2 + 19b_3^2/18b_2^3 - 2g'b_3/3g^2b_2^2)\eta.\end{aligned} \tag{5.2}$$

As a second example, let $k(x) = (1 + x^2)^{-\frac{1}{2}} = f(x)$. The basic assumptions are clearly satisfied and $n^{\frac{1}{2}}X_n$ has the t-distribution with n degrees of freedom. Using Table 2.1, we obtain expressions for the c_{lm} from the expansion of $\log f(x)$. The expansion becomes

$$F_n(\xi) = \Phi(\xi) - [\varphi(\xi)(\xi^3 + \xi)/4]n^{-1} + O(n^{-\frac{3}{2}}) \qquad (n \to \infty)$$

which agrees with Fisher (1925), who also gives more terms. Pieser (1949) has shown the error is of the correct order. Both Fisher and Pieser, as well as others, seem unaware of the relation to the work of Laplace.

Hotelling and Frankel (1938), page 89, give the normalizing transformation, and Pieser (1949) gives the expansion for the percentiles.

6. Acknowledgment. I wish to express my gratitude to Professor R. J. Buehler for his guidance and advice throughout the preparation of this paper.

REFERENCES

BOL'SHEV, L. (1959). On transformations of random variables. *Theor. Prob. Appl.* **4** 129–141.

BOL'SHEV, L. (1963). Asymptotically Pearson transformations. *Theor. Prob. Appl.* **8** 121–146.

BUEHLER, R. J. (1965). The limit of the nth power of a density. *Ann. Math. Statist.* **36** 1878–1882.

DE BRUIJN, N. G. (1961). *Asymptotic Methods in Analysis*. Interscience, New York.

DOROGOVCEV, A. (1962). Confidence intervals associated with the evaluation of parameters. *Select. Transl. Math. Statist. Prob.* **3** 321–326. Amer. Math. Soc., Providence.

ERDELYI, A. (1956). *Asymptotic Expansions*. Dover, New York.

FISHER, R. (1925). Expansion of Student's integral in powers of n^{-1}. *Metron* **5** 109–111.

FUKS, B. (1963). *Analytic Functions of Several Complex Variables*. Translations of Mathematical Monographs 8. Amer. Math. Soc., Providence.

HOTELLING, H. and FRANKEL, L. (1938). The transformation of statistics to simplify their distribution. *Ann. Math. Statist.* **8** 87–96.

JOHNSON, R. A. (1966). Asymptotic expansions associated with the nth power of a density. Technical Report No. 80, Department of Statistics, University of Minnesota.

LAPLACE, P. S. DE (1847). *Oeuvers*, **7**. Imprimerie Royale, Paris.

MARKUSHEVICH, A. (1965). *Theory of Functions of a Complex Variable*. **2**. Prentice-Hall, Englewood Cliffs.

PEISER, A. (1949). Correction to asymptotic formulas for significance levels of certain distributions. *Ann. Math. Statist.* **20** 128–129.

VON MISES, R. (1964). (edited by HILDA GEIRINGER) *Mathematical Theory of Probability and Statistics*. Academic Press, New York.

WASOW, W. (1956). On the asymptotic transformation of certain distributions into the normal distribution. *Proc. Symposia Appl. Math.* **6** 251–259. McGraw-Hill, New York.