Open Journal of Statistics, 2012, 2, 98-105
http://dx.doi.org/10.4236/ojs.2012.21010 Published Online January 2012 (http://www.SciRP.org/journal/ojs)

On the Spectrum of Asymptotic Expansions for an Asymptotic Normal Sequence

Min Tsao
Department of Mathematics and Statistics, University of Victoria, Victoria, Canada
Email: tsao@math.uvic.ca

Received September 20, 2011; revised October 18, 2011; accepted October 29, 2011

ABSTRACT

We present a family of formal expansions for the density function of a general one-dimensional asymptotic normal sequence X_n. Members of the family are indexed by a parameter τ with an interval domain which we refer to as the spectrum of the family. The spectrum provides a unified view of known expansions for the density of X_n. It also provides a means to explore for new expansions. We discuss such applications of the spectrum through that of a sample mean and a standardized mean. We also discuss a related expansion for the cumulative distribution function of X_n.

Keywords: Asymptotic Expansion; Asymptotic Normal Sequence; Edgeworth Expansion; Saddlepoint Expansion; Saddlepoints Expansion; Hermite Polynomials

1. Introduction

Historically, formal expansions (*i.e.*, non-rigorous expansions) for distributions of random variables have played an important role in the development of asymptotic theories in statistics. The most well-known example is the Edgeworth expansion for the density of a standardized mean which was first derived in 1905 as a formal expansion for the density [1]. The method used by Edgeworth in his derivation made use of Charlier differential series and a standard normal density as a developing function [2]. This method did not address the validity of the expansion, but the expansion was proven valid by Cramér 23 years later in [3,4]. See also [5]. Another well-known example is that of the formal expansion by Wallace [2]. Its validity was given 20 years later in [6]. Indeed, formal expansions often serve as the first step in exploring valid expansions for random variables. Furthermore, they may be valuable approximation tools even in the absence of a rigorous treatment of their validity. There are many useful formal approximations that are numerically very accurate. For more recent examples, see [7,8].

Nevertheless, in spite of the usefulness of formal expansion there does not seem to exist a systematic approach for deriving such expansions in the literature. To obtain an Edgeworth type of expansion for an asymptotically normal sequence X_n, the common approach is to use the moments of X_n (or in the absence of the exact moments, the approximate moments obtained through the delta method) and substitute them into the Edgeworth expansion formula for the standardized mean. To obtain a saddlepoint type of expansion, one often follows Daniels's derivation [9] of the expansion for a sample mean by writing the cumulant generating function (or an approximation to the cumulant generating function) as a product of the sample size n and another function, say $h(x,n)$. Then applying the method of steepest descent to an inversion formula as if $h(x,n)$ is independent of the asymptotic factor n. See, for example, [10,11]. Such methods work on the particular cases in question but offer little insight into how a formal expansion should be sought in general.

The main purpose of this paper is to introduce a family of formal expansions for a general asymptotic normal sequence X_n. Members of the family are indexed by a parameter τ with an interval domain which we call the *spectrum* of the expansions. The spectrum has the following applications. 1) It provides a means to study the whole family of formal expansions and search for good and valid expansions. 2) It provides a way to view known expansions from a unified standpoint, thereby linking seemingly unrelated expansions under a unified framework. For the case of a standardized mean, for example, the Edgeworth expansion and the saddlepoints expansion [12] are actually members of the same family, although they are based on different asymptotic sequences and are substantially different in structure. 3) Existing expansions are mostly power sequence expansions in that individual terms of the expansions are of the form $a_k n^{-k/2}$

or $a_k n^{-k}$. The spectrum contains "non-standard" expansions which are not power sequence expansions. This allows one to explore new expansions which are not power sequence expansions. In cases where X_n is neither the mean nor the standardized mean of iid observations, such "non-standard" expansions may be more natural expansions than those based on the power sequences $\{n^{-k/2}\}$ or $\{n^{-k}\}$.

The rest of this paper is organized as follows: in Section 2, we derive the family of formal expansions. In Section 3, we discuss the validity of the family for the cases of the sample mean and the standardized mean. For the latter case, the family led to a set of valid new expansions for the density function. In Section 4, we consider a related formal expansion for the distribution function. Concluding remarks in Section 5 will include further notes on previous work which have motivated this paper.

2. The Family of Formal Expansions

Suppose X_n $(n = 1, 2, \cdots)$ has a density function $f_n(x)$ and a moment generating function $M_n(t)$. Assume that as n approaches infinity the interval in which $M_n(t)$ exists approaches a non-vanishing open interval (l, u) where l, u are constants and $l < u$. We will derive a formal expansion for $f_n(x)$ at each point $\tau \in (l, u)$ and thus we call (l, u) the *spectrum* of the formal expansions for $f_n(x)$. To derive the formal expansion at a point τ, we need the following inversion formula

$$f_n(x) = \frac{1}{2\pi i}\int_{\tau - i\infty}^{\tau + i\infty} \mathrm{e}^{K_n(T) - Tx}\,\mathrm{d}T, \tag{1}$$

where $K_n(T) = \log M_n(T)$, $T = \tau + iy$ and $\tau \in (l, u)$. A key result that will be used in the derivation is the following lemma which establishes a new defining relation for Hermite polynomials.

Lemma 1: *Let $\phi(v)$ be the density function of the standard normal distribution and $H_k(x)$ be the Hermite polynomial of degree k. Then*

$$\int_{-\infty}^{\infty}\phi(v)(iv - x)^k\,\mathrm{d}v = (-1)^k H_k(x), \tag{2}$$

for $k = 0, 1, 2, \cdots$

Proof of Lemma 1: See Appendix.

The derivation of the formal expansion at τ involves the following two steps. Step 1: obtaining a formal series representation of the density function, and Step 2: rearranging terms in the formal series representation according to their asymptotic orders. Step 1 is achieved by first replacing the exponent of the integrand in (1) with its Taylor series expansion, then isolating the quadratic term of the Taylor series and performing a term-by-term integration. Note that in this step no attempt will be made to isolate the asymptotic factor n from the exponent since we do not presume that the expansion of interest is based on a power sequence of n. Also, with the aid of Lemma 1, Step 1 is independent of τ and gives a unified series representation for all τ values in the spectrum as we will see in the proof of Theorem 1 below.

Theorem 1: *For any $\tau \in (l, u)$ where $K_n''(\tau) > 0$, $f_n(x)$ has the following formal series representation* (3) where $\phi_{(n,\tau)}(x)$ is the density function of the normal distribution with mean $\mu_n = K_n'(\tau)$ and variance $\sigma_n^2 = K_n''(\tau)$, and $\lambda_{(n,r)}(\tau) = K_n^{(r)}(\tau) \big/ \left[K_n''(\tau)\right]^{r/2}$.

Proof of Theorem 1: For convenience of presentation, we first consider the special case of $\tau = 0$. By setting τ in (1) to 0, on the path of integration near the origin we have

$$\begin{aligned} K_n(T) - Tx &= \left(K_n'(0) - x\right)iy - \frac{1}{2}K_n''(0)y^2 \\ &\quad + \frac{1}{3!}K_n^{(3)}(0)(iy)^3 + \frac{1}{4!}K_n^{(4)}(0)(iy)^4 + \cdots \end{aligned} \tag{4}$$

Since $K_n'(0) = \mu_n$ and $K_n''(0) = \sigma_n^2$, where μ_n and σ_n^2 are the mean and variance of X_n, (4) may be written as

$$\begin{aligned} K_n(T) - Tx &= (\mu_n - x)iy - \frac{1}{2}\sigma_n^2 y^2 \\ &\quad + \frac{1}{3!}K_n^{(3)}(0)(iy)^3 + \frac{1}{4!}K_n^{(4)}(0)(iy)^4 + \cdots \\ &= -\frac{\sigma_n^2}{2}\left[y - \frac{(\mu_n - x)i}{\sigma_n^2}\right]^2 - \frac{(\mu_n - x)^2}{2\sigma_n^2} \\ &\quad + \frac{1}{3!}K_n^{(3)}(0)(iy)^3 + \cdots. \end{aligned} \tag{5}$$

Letting $c = (\mu_n - x)/\sigma_n^2$, Equation (1) may be formally rewritten as

$$\begin{aligned} f_n(x) &= \frac{1}{2\pi}\mathrm{e}^{-(\mu_n - x)^2/2\sigma_n^2} \\ &\quad \times \int_{-\infty}^{\infty}\mathrm{e}^{-\frac{\sigma_n^2}{2}(y - ic)^2 + \frac{1}{3!}K_n^{(3)}(0)(iy)^3 + \frac{1}{4!}K_n^{(4)}(0)(iy)^4 + \cdots}\,\mathrm{d}y. \end{aligned} \tag{6}$$

$$\begin{aligned} f_n(x) = \mathrm{e}^{K_n(\tau) - \tau x}\phi_{(n,\tau)}(x)\Bigg\{1 &- \frac{\lambda_{(n,3)}(\tau)}{3!}H_3\left(\frac{\mu_n - x}{\sigma_n}\right) \\ &+ \frac{\lambda_{(n,4)}(\tau)}{4!}H_4\left(\frac{\mu_n - x}{\sigma_n}\right) + \cdots + \frac{\lambda_{(n,3)}^2(\tau)}{2!(3!)^2}H_6\left(\frac{\mu_n - x}{\sigma_n}\right) + \cdots\Bigg\}, \end{aligned} \tag{3}$$

Letting $v=\sigma_n(y-ic)$ and $\lambda_{(n,j)}(0)=K_n^{(j)}(0)/\sigma_n^j$ for $j\geq 3$, we may write (6) as

$$\begin{aligned} f_n(x) &= \frac{1}{2\pi\sigma_n} e^{-(\mu_n-x)^2/2\sigma_n^2} \int_{-\infty-ic\sigma_n}^{\infty-ic\sigma_n} e^{-\frac{1}{2}v^2+\frac{1}{3!}\lambda_{(n,3)}\left[i(v+ic\sigma_n)\right]^3+\frac{1}{4!}\lambda_{(n,4)}\left[i(v+ic\sigma)_n\right]^4+\cdots} dv \\ &= \phi_n(x)\int_{-\infty-ic\sigma_n}^{\infty-ic\sigma_n}\phi(v) e^{\frac{1}{3!}\lambda_{(n,3)}\left[i(v+ic\sigma_n)\right]^3+\frac{1}{4!}\lambda_{(n,4)}\left[i(v+ic\sigma_n)\right]^4+\cdots} dv, \end{aligned} \tag{7}$$

where ϕ_n is the density of $N(\mu_n,\sigma_n^2)$, and for brevity we have written $\lambda_{(n,j)}(0)$ as $\lambda_{(n,j)}$. Expanding the function $\exp(\cdot)$ in the integrand, we get

$$f_n(x)=\phi_n(x)\int_{-\infty-ic\sigma_n}^{\infty-ic\sigma_n}\phi(v)\left\{1+\frac{1}{3!}\lambda_{(n,3)}\left[i(v+ic\sigma_n)\right]^3+\frac{1}{4!}\lambda_{(n,4)}\left[i(v+ic\sigma_n)\right]^4+\cdots+\frac{1}{2!(3!)^2}\lambda_{(n,3)}^2\left[i(v+ic\sigma_n)\right]^6+\cdots\right\}dv \tag{8}$$

We now perform the term-by-term integration for the right-hand side of (8). This is easily carried out using Lemma 1 by noting that $\phi(v)v^r$ is an entire function. Thus the contour of integration in (8) may be deformed from $Im(v)=-c\sigma_n$ to $Im(v)=0$. This and Lemma 1 lead to

$$\begin{aligned} &\int_{-\infty-ic\sigma_n}^{\infty-ic\sigma_n}\phi(v)\left[i(v+ic\sigma_n)\right]^k dv \\ &= \int_{-\infty}^{\infty}\phi(v)(iv-x)^k dv = (-1)^k H_k(x), \end{aligned} \tag{9}$$

for $k=0,1,\cdots$. It follows that

$$f_n(x)=\phi_n(x)\left\{1+\frac{1}{3!}\lambda_{(n,3)}(-1)^3 H_3(c\sigma_n)+\frac{1}{4!}\lambda_{(n,4)}(-1)^4 H_4(c\sigma_n)+\cdots+\frac{1}{2!(3!)^2}\lambda_{(n,3)}^2(-1)^6 H_6(c\sigma_n)+\cdots\right\}, \tag{10}$$

or equivalently

$$f_n(x)=\phi_n(x)\left\{1-\frac{1}{3!}\lambda_{(n,3)}H_3\left(\frac{\mu_n-x}{\sigma_n}\right)+\frac{1}{4!}\lambda_{(n,4)}H_4\left(\frac{\mu_n-x}{\sigma_n}\right)+\cdots+\frac{1}{2!(3!)^2}\lambda_{(n,3)}^2 H_6\left(\frac{\mu_n-x}{\sigma_n}\right)+\cdots\right\}, \tag{11}$$

which is the formal series representation (3) at $\tau=0$.

For a general $\tau\in(l,u)$, $K_n(\tau)-\tau x$ may not be zero and (4) becomes

$$\begin{aligned} K_n(T)-Tx &= K_n(\tau)-\tau x+\left(K_n'(\tau)-x\right)iy \\ &\quad -\frac{1}{2}K_n''(\tau)y^2+\frac{1}{3!}K_n^{(3)}(\tau)(iy)^3+\cdots \end{aligned} \tag{12}$$

Let $\phi_{(n,\tau)}(x)$ be the density function of the normal distribution with mean $\mu_n=K_n'(\tau)$ and variance $\sigma_n^2=K_n''(\tau)$ and write $\lambda_{(n,r)}(\tau)=K_n^{(r)}(\tau)/\left[K_n''(\tau)\right]^{r/2}$. By replacing (4) with (12) and then following the same steps for the case of $\tau=0$ shown above, we obtain (3). □

The series representation (3) takes on a simpler form (11) for the special case of $\tau=0$ because the $\exp\{K_n(\tau)-\tau x\}$ term in (3) is not in (11). Another special case where (3) has a simpler form is the case where τ is the saddlepoint $\tau=\hat{T}_n$ satisfying $K_n'(\hat{T}_n)-x=0$. Define $g_n(x)$, the *generalized saddlepoint approximation* for $f_n(x)$, as

$$g_n(x)=\sqrt{\frac{1}{2\pi K_{n''}(\hat{T}_n)}}e^{K_n(\hat{T}_n)-\hat{T}_n x}. \tag{13}$$

Setting τ to $\tau=\hat{T}_n$ in (3) and noting that $(\mu_n-x)/\sigma_n=\left(K_n'(\tau)-x\right)/K_n''(\tau)=0$, we obtain the series representation of $f_n(x)$ at the saddlepoint:

$$\begin{aligned} f_n(x)=g_n(x)\Big\{&1+\frac{1}{8}\lambda_{(n,4)}(\hat{T}_n)+\frac{1}{2\times 4!}\lambda_{(n,6)}(\hat{T}_n) \\ &+\cdots-\frac{5}{24}\lambda_{(n,3)}^2(\hat{T}_n)+\cdots\Big\}. \end{aligned} \tag{14}$$

Note that here $\tau=\hat{T}_n$ depends on n and thus is not a constant in the spectrum as n changes.

We now discuss Step 2. Series (3), (11) or (14) are not particularly useful from an asymptotic expansion point of view in that they, like Charlier differential series, do not use information concerning the asymptotic properties of the distribution of X_n. They are not asymptotic expansions for X_n. When X_n is asymptotically normal, the sequence $\{\lambda_{(n,r)}(\cdot)\}$ may be an asymptotic sequence and may thus be used to transform these series into formal asymptotic expansions. To do so, we need to rearrange terms in the curly brackets in (3), (11) and (14) in ascending order according to the rates at which the $\lambda_{(n,r)}(\cdot)$'s approach zero. Corollary 1 below gives the rearranged series at the saddlepoint (14).

Corollary 1: *Suppose* $\lambda_{(n,i+1)}(\hat{T}_n)=o\left(\lambda_{(n,i)}(\hat{T}_n)\right)$ *as* n *approaches infinity for* $i\geq 2$ *and in particular* $\lambda_{(n,4)}(\hat{T}_n)=O\left(\lambda_{(n,3)}^2(\hat{T}_n)\right)$. *Then we have formally*

$$f_n(x) = g_n(x)\left\{1 + \left[\frac{1}{8}\lambda_{(n,4)}\left(\hat{T}_n\right) - \frac{5}{24}\lambda^2_{(n,3)}\left(\hat{T}_n\right)\right] + o\left(\lambda_{(n,4)}\left(\hat{T}_n\right)\right)\right\}. \quad (15)$$

We refer to (15) this as the *generalized saddlepoint expansion* for $f_n(x)$ based on the asymptotic sequence $\left\{\lambda_{(n,i)}\left(\hat{T}_n\right)\right\}_{i=2}^{\infty}$. Note that conditions in Corollary 1 are satisfied by a large class of statistics X_n, including the sample mean.

To transform the general series (3) into a formal expansion, we also need to consider the Hermite polynomials that appear in (3). If the absolute value of their common argument, $\frac{(\mu_n - x)}{\sigma_n}$, goes to infinity when n goes to infinity, then since H_k is a polynomial of order k. The reciprocals of these polynomials $1/H_k$ $(k = 3, 4, \cdots)$ will form an asymptotic sequence with respect to n. Thus (3) contains ratios of terms in two asymptotic sequences $\left\{\lambda_{(n,i)}(\tau)\right\}_{i=2}^{\infty}$ and $\left\{1/H_k\right\}_{k=3}^{\infty}$, and its asymptotic properties become complicated. To avoid this complication, we assume that $(\mu_n - x)/\sigma_n$ is bounded. With this assumption, the relative rate at which terms in the curly bracket of (3), such as $\lambda_{(n,3)}(\tau)H_3$ and $\lambda_{(n,4)}(\tau)H_4$, approach zero is determined by that of the $\lambda_{(n,i)}(\tau)$'s. Rearranging terms in (3), we have

Corollary 2: *Suppose* $\lambda_{(n,i+1)}(\tau) = o\left(\lambda_{(n,i)}(\tau)\right)$ *for* $i \geq 2$ *and* $(\mu_n - x)/\sigma_n$ *is bounded as* n *approaches infinity. Then we have formally* (16) where $\mu_n = K'_n(\tau)$ and $\sigma_n = K''_n(\tau)$.

In particular, at $\tau = 0$ formal expansion (16) becomes (17).

We refer to (16) as the *general expansion* for $f_n(x)$. and (17) as the *generalized Edgeworth expansion* because the latter is the expansion at the origin but unlike the Edgeworth expansion which is based on the power sequence $\left\{n^{-j/2/}\right\}_{j=1}^{\infty}$, (17) is based on a general asymptotic sequence $\left\{\lambda_{(n,i)}\right\}_{i=2}^{\infty}$.

Note that conditions on the relative order of the $\lambda_{(n,r)}(\cdot)$'s and the boundedness of $\frac{(\mu_n - x)}{\sigma_n}$ in the corollaries are easily verified once X_n is given. When some of these conditions are not met, terms in the series representations need to be arranged accordingly. The resulting formal expansions may be different from those obtained above but the leading term should still be $\mathrm{e}^{K_n(\tau)-\tau x}\phi_{(n,\tau)}(x)$.

3. The Spectrum of the Sample Mean and Standardized Mean

To demonstrate the use of the spectrum, we now examine the spectrum for the important cases of sample mean and standardized mean. We show that the known expansions such as the saddlepoint, Edgeworth and saddlepoints expansions, can all be located through the spectrum. Moreover, we examine the validity of other expansions in the spectrum.

3.1. Expansions for the Density of the Sample Mean

Let $X_n = \overline{X}_n$ be the average of n independent copies of a random variable X. How does the generalized saddlepoint expansion relate to the saddlepoint expansion for $\overline{X}_n$ given by Daniels [9]? Let $K(T)$ be the cumulant generating function of W, then $K_n(T) = nK(T/n)$. Let $\hat{T}$ be the solution of $K'\left(\hat{T}\right) = x$, then $\hat{T}_n = n\hat{T}$. Furthermore, $K_n\left(\hat{T}_n\right) = nK\left(\hat{T}\right)$ and $K''_n\left(\hat{T}_n\right) = K''\left(\hat{T}\right)/n$. Thus the generalized saddlepoint approximation (13) is the same as Daniels's saddlepoint approximation,

$$g_n\left(\overline{x}\right) = \left[n/2\pi K''\left(\hat{T}\right)\right]^{\frac{1}{2}} \exp\left\{n\left[K\left(\hat{T}\right) - \hat{T}\overline{x}\right]\right\}.$$

To examine the asymptotic property of other terms of the generalized saddlepoint expansion (15), we first note that $K_n^{(r)}\left(\hat{T}_n\right) = K^{(r)}\left(\hat{T}\right)/n^{r-1}$ for any $r \in N$. Hence $\lambda_{(n,r)}\left(\hat{T}_n\right) = K^{(r)}\hat{T}\Big/\left[n^{r/2-1}\left(K''\left(\hat{T}\right)\right)^{r/2}\right]$ for $r \geq 2$. Denote $K^{(r)}\left(\hat{T}\right)\Big/\left(K''\left(\hat{T}\right)\right)^{r/2}$ by λ_r. It is not difficult to show that

$$\frac{1}{8}\lambda_{(n,4)}\left(\hat{T}_n\right) - \frac{5}{24}\lambda^2_{(n,3)}\left(\hat{T}_n\right) = \frac{1}{n}\left(\frac{1}{8}\lambda_4(T) - \frac{5}{24}\lambda_3^2(T)\right)$$

which is a_1 in the saddlepoint expansion in [9]. Further

$$f_n(x) = \mathrm{e}^{K_n(\tau)-\tau x}\phi_{(n,\tau)}(x)\left\{1 - \frac{1}{6}\lambda_{(n,3)}(\tau)H_3\left(\frac{\mu_n - x}{\sigma_n}\right) + o\left(\lambda_{(n,3)}(\tau)\right)\right\}, \quad (16)$$

$$f_n(x) = \phi_n(x)\left\{1 + \frac{1}{6}\lambda_{(n,3)}\left[\left(\frac{x - \mu_n}{\sigma_n}\right)^3 - 3\left(\frac{x - \mu_n}{\sigma_n}\right)\right] + o\left(\lambda_{(n,3)}\right)\right\} \quad (17)$$

terms in expansion (15) may be constructed for this particular case and it can be shown that they are equal to the corresponding terms in Daniels's saddlepoint expansion. Thus the generalized saddlepoint expansion (15) is Daniels's saddlepoint expansion.

It may also be easily verified using the same arguments demonstrated above that the general expansion (16) coincides with the expansion Daniels derived through the Edgeworth expansion at τ in [9]. See (4.3) in Section 4 in Daniels (1954). We will refer to this (4.3) as D(4.3). Daniels [9] stated that the family of expansions given by D(4.3) are asymptotic expansions for $f_n(\bar{x})$. This, however, is not accurate. The reason is that the Edgeworth expansion for a standardized variable $Z_u = \left(\bar{U} - E(U)\right)\big/\left(\sigma_u/\sqrt{n}\right)$ may not be used to obtain an asymptotic expansion for the density of $\bar{U}$ at anywhere except for $E(U)$. The distribution of the random variable U described before D(4.3) has mean $E(U) = K'(\tau) - \bar{x}$, but D(4.3) was derived through the Edgeworth expansion for Z_u at $\bar{U} = 0$ or $Z_u = -\left[K'(\tau) - \bar{x}\right]\left[n/K''(\tau)\right]^{1/2}$. Thus when $K'(\tau) - \bar{x} \neq 0$, expansions given by D(4.3) are not valid. More specifically, the coefficient A_1 in D(4.3), for example, is in general $O\left(n^{3/2}\right)$. Thus the second term in D(4.3), $A_1/\sqrt{n}$, is in general $O(n)$. Hence D(4.3) cannot even be an asymptotic expansion in a formal sense. This illustrates the necessity of the condition that $(\mu_n - x)/\sigma_n$ be bounded, which we have used in arriveing at (16).

3.2. Expansions for the Density of the Standardized Mean

It is not difficult to verify that for this case the generalized Edgeworth expansion (17) coincides with the Edgeworth expansion. Furthermore, the saddlepoints approximation for the density of a standardized mean given by Routledge and Tsao [12] is actually the generalized saddlepoint expansion (15). We now focus on the validity of a set of new expansions within the family. These correspond to members of the family at other points of the spectrum. By (16), these have the expression

$$f_n(x) = \mathrm{e}^{K_n(\tau) - \tau x}\phi_{(n,\tau)}(x)\left\{1 + O\left(\lambda_{(n,3)}(\tau)\right)\right\}. \quad (18)$$

Although in this case the $O\left(\lambda_{(n,3)}(\tau)\right)$ term in (18) and the $O\left(\lambda_{(n,4)}(\tau)\right)$ term in (16) may be easily further expanded, verification of the validity of expansions with more terms than that in (18) is more involved and will not be considered here. We only consider (18) for which the validity of the family can be established. The following equation will be used implicitly for showing the validity:

$$\left[1 + O\left(\frac{1}{n^p}\right)\right]^q = 1 + O\left(\frac{1}{n^p}\right), \quad (19)$$

where $p, q \in R$ and $p > 0$.

Let $K_n(T)$ be the cumulant generating function of the standardized mean. Then its derivatives have the following expansions: (i) $K_n(T) = T^2/2 + O\left(n^{-1/2}\right)$, (ii) $K_n'(T) = T + O\left(n^{-1/2}\right)$, (iii) $K_n''(T) = 1 + O\left(n^{-1/2}\right)$, and iv) $K_n^{(r)}(T) = O(1/n^{r/2-1})$ for $r \geq 3$. Denote the leading term of the expansion in (18) by $l_{(n,\tau)}(x)$. We have

$$l_{(n,\tau)}(x) = \frac{1}{\sqrt{2\pi K_n''(\tau)}}\exp\left\{K_n(\tau) - \tau x - \frac{\left(x - K_n'(\tau)\right)^2}{2K_n''(\tau)}\right\}.$$

Equations (i), (ii), (iii) and (19) imply that

$$\begin{aligned} l_{(n,\tau)}(x) &= \frac{\exp\left\{\frac{1}{2}\tau^2 + O\left(n^{-1/2}\right) - \tau x - \frac{\left[x - \tau + O\left(n^{-1/2}\right)\right]^2}{2\left[1 + O\left(n^{-1/2}\right)\right]}\right\}}{\sqrt{2\pi\left[1 + O\left(n^{-1/2}\right)\right]}} \\ &= \frac{1 + O\left(n^{-1/2}\right)}{\sqrt{2\pi}}\exp\left\{\frac{-x^2/2 + O\left(n^{-1/2}\right)}{1 + O\left(n^{-1/2}\right)}\right\} \\ &= \phi(x)\mathrm{e}^{O\left(n^{-1/2}\right)}\left[1 + O\left(n^{-1/2}\right)\right] \\ &= \phi(x)\left[1 + O\left(n^{-1/2}\right)\right]. \end{aligned} \quad (20)$$

Also, (iii) and (iv) imply that $\lambda_{(n,3)}(\tau) = O\left(n^{-1/2}\right)$. Thus (20) may be written as

$$l_{(n,\tau)}(x) = \phi(x)\left\{1 + O\left(\lambda_{(n,3)}(\tau)\right)\right\}.$$

By the Edgeworth expansion, $f_n(x) = \phi(x)\left[1 + O\left(n^{-1/2}\right)\right]$. Thus

$$\begin{aligned} \frac{f_n(x)}{l_{(n,\tau)}(x)} &= \frac{\phi(x)\left[1 + O\left(n^{-1/2}\right)\right]}{\phi(x)\left[1 + O\left(n^{-1/2}\right)\right]} \\ &= 1 + O\left(n^{-1/2/}\right) \ or \ 1 + O\left(\lambda_{(n,3)}(\tau)\right). \end{aligned} \quad (22)$$

This proves the validity of (18). We have compared the numerical accuracy of $l_{(n,\tau)}(x)$ to the normal approximation $\phi(x)$ for small and moderately large sample sizes through a number of examples. Not surprisingly, $l_{(n,\tau)}(x)$ is substantially more accurate than $\phi(x)$ when τ is close to the saddlepoint. They are about the same when τ is near zero.

To summarize, all known expansions for the above two special cases have been located in their spectrums. For the sample mean, the generalized saddlepoint expansion is the only member which is a valid asymptotic expansion. For the standardized mean, new valid expansions have been found.

4. Expansions for the Distribution Function

The formal expansions for density functions may be integrated to obtain expansions for the corresponding distribution function, $F_n(x)$. Consider the case where $\mu_n = 0$ and $\sigma_n = 1$. By formally integrating (17), we obtain the *generalized Edgeworth expansion*

$$F_n(x) = \Phi(x) + \frac{1}{6}\lambda_{(n,3)}\left(1-x^2\right)\phi(x) + o\left(\lambda_{(n,3)}\right). \quad (23)$$

It may be easily verified that (23) is the same as the Edgeworth expansion for the distribution function when X_n is the standardized mean. We now consider another example where (23) is valid. Let U_n be a U-statistic of degree 2,

$$U_n = \frac{2}{n(n-1)}\sum_{1\le i<j\le n} h\left(X_i, X_j\right),$$

where the X_i's are iid and h is a symmetric function of two variables with $E\left[h\left(X_1,X_2\right)\right]=0$ and $E\left[h^2\left(X_1,X_2\right)\right]<\infty$. Let σ_n be the standard deviation of U_n and $F_n(x)$ be the distribution function of U_n/σ_n, then under certain conditions [13] and [14] showed that

$$F_n(x) = \Phi(x) + \frac{1}{6}\frac{\omega_3}{\sqrt{n}}\left(1-x^2\right)\phi(x) + O\left(n^{-1/2}\right), \quad (25)$$

where $\omega_3/\sqrt{n}$ is an approximation with error $o(1/n)$ to the third cumulant of U_n/σ_n, $\lambda_{(n,3)}$. With $X_n = U_n/\sigma_n$, (25) then implies that (23) is indeed valid. Furthermore, it can be shown that the fourth cumulant of U_n/σ_n, $\lambda_{(n,4)}$, satisfies $\lambda_{(n,4)} = O(1/n)$. The right-hand side of (25) and thus that of (23) can be further expanded. The expansion in (25) is simpler than that in (23) in that it is defined in terms of a simpler asymptotic sequence $\left\{1/n^{k/2}\right\}$ while (23) is defined in terms of $\left\{\lambda_{(n,r)}\right\}$ $(r \ge 3)$) which may be difficult to compute. From the present point of view, however, $\left\{\lambda_{(n,r)}\right\}$ is a more natural asymptotic sequence upon which to base asymptotic expansions. Presently, it is not clear which one is more accurate for small and moderate sample sizes.

Steps similar to Steps 1 and 2 in Section 2 may be devised to derive formal expansions for the distribution function directly using the inversion formula,

$$Q_n(x) = \frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} e^{K_n(T)-Tx}\frac{dT}{T}, \quad (26)$$

where $Q_n(x)$ is the tail probability and $\tau > 0$. This process, however, is more complicated due to the extra term $1/T$ in the integrand and it leads to different expansions depending on whether or not $1/T$ is expanded. We will not discuss such expansions here.

5. Concluding Remarks

For the cases of the sample mean and standardized mean, the spectrum has provided a new perspective on asymptotic expansions for density functions. It revealed that the saddlepoint expansion is the only valid expansion in the spectrum for the sample mean. It led to new expansions and provided a unified standpoint for viewing known expansions for the standardized mean. It also led to valid expansions outside the iid setting. These suggest that the spectrum is a valuable tool in finding expansions for density functions.

The derivation in Section 2 does not explicitly use the condition that X_n is asymptotically normal. Without this condition, however, the sequence $\left\{\lambda_{(n,r)}(\tau)\right\}$ may not be an asymptotic sequence and this condition has been used implicitly in the corollaries. Our derivation also shows that to obtain a saddlepoint type of expansion it is not necessary to isolate the asymptotic factor n by expressing the cumulant generating function of X_n as $K_n(nT) = nR_n(T)$. Instead, one can use K_n directly to obtain an expansion. Although the former approach will lead to the same saddlepoint approximation as the latter, it will obscure the underlying asymptotic sequence of the expansion and consequently that of the asymptotic order of the saddlepoint approximation. Indeed, the fact that the cumulant generating function can be written as n times a function not dependent on n is only a coincidence in the iid case. It has made it possible to establish the validity of the saddlepoint expansion through the method of steepest descent for this case. But it is not essential for deriving a formal expansion in general.

Turning now to some historical notes and remarks on previous work which have motivated this work. The Charlier difference series and the Gram-Charlier series of type A are mathematically elegant formal techniques which have contributed to the discovery of the Edgeworth expansion. However, they were not specifically aimed at approximating distributions from an asymptotic point of view and were unable to make use of the information that X_n is asymptotically normal beyond choosing the normal density function as the developing function. When the focus is on obtaining accurate approximations for the distributions of X_n rather than obtaining the speed at which the sequence approaches normality, other developing functions may be more suitable. In the present paper, we have found the leading term of the general expansion (16) to be very useful for this purpose.

Although in the extended version of Poincaré's definition of an asymptotic expansion,

$$f(z) = \sum_{n=1}^{N} a_n\phi_n(z) + o\left(\phi_N(z)\right),$$

the asymptotic sequence $\{\phi_n\}$ needs not to be a power sequence, important developments in the theory of asymptotic analysis are mostly concerned with power series expansions. The developments in asymptotic expansions in statistics reflect that of the theory of asymptotic analysis. Our use of the sequence $\{\lambda_{(n,r)}(\tau)\}$ was inspired by [15,16] which have used non-power sequences to characterize the Edgeworth expansion and the saddlepoint expansion. Indeed, with an appropriate standardizetion the cumulant generating function of an asymptotically normal sequence approaches a second order polynomial. If the limiting normal distribution is not a degenerate distribution, then the sequence $\{\lambda_{(n,r)}(\tau)\}$ may be an asymptotic sequence which can be used to construct asymptotic expansions.

6. Acknowledgements

I would like to thank a referee for helpful comments which have led to improvements in this paper.

REFERENCES

[1] F. Y. Edgeworth, "The Law of Error," *Cambridge Philosophical Society Transactions*, Vol. 20, 1905, pp. 33-66.

[2] D. L. Wallace, "Asymptotic Approximations to Distributions," *Annals of Mathematical Statistics*, Vol. 29, No. 3, 1958, pp. 635-654. doi:10.1214/aoms/1177706528

[3] H. Cramér, "On the Composition of Elementary Errors," *Skand Aktuarietidskr*, Vol. 11, 1928, pp. 13-74.

[4] H. Cramér, "Random Variables and Probability Distributions," Cambridge University Press, Cambridge, 1937.

[5] W. Feller, "An Introduction to Probability Theory and Its Applications," Wiley, New York, 1966.

[6] R. N. Bhattacharya and J. K. Ghosh, "On the Validity of the Formal Edgeworth Expansion," *Annals of Statistics*, Vol. 6, No. 2, 1978, pp. 434-451. doi:10.1214/aos/1176344134

[7] R. Strawderman, G. Casella and M. Wells, "Practical Small-Sample Asymptotics for Regression Problems," *Journal of the American Statistical Association*, Vol. 91, No. 434, 1996, pp. 643-654. doi:10.2307/2291660

[8] C. A. Field, "Tail Areas of Linear Combinations of Chi-Squares and Non-Central Chi-Squares," *Journal of Statistical Computation and Simulation*, Vol. 45, No. 3-4, 1996, pp. 243-248. doi:10.1080/00949659308811484

[9] H. E. Daniels, "Saddlepoint Approximations in Statistics," *Annals of Mathematical Statistics*, Vol. 25, No. 4, 1954, pp. 631-649. doi:10.1214/aoms/1177728652

[10] M. S. Srivastava and W. K. Yau, "Tail Probability Approximations of General Statistics," Technical Report No. 88-38, Center for Multivariate Analysis, University of Pittsburgh, Pittsburgh, 1988.

[11] G. S. Easton and E. Ronchetti, "General Saddlepoint Approximation with Application to L Statistics," *Journal of the American Statistical Association,* Vol. 81, No. 394, 1986, pp. 420-429. doi:10.2307/2289231

[12] R. D. Routledge and M. Tsao, "Uniform Validity of Saddlepoint Expansion on Compact Set," *Canadian Journal of Statistics*, Vol. 23, No. 4, 1995, pp. 425-431. doi:10.2307/3315386

[13] H. Callaert, P. Janssen and N. Veraverbeke, "An Edgeworth Expansion for U-Statistics," *Annals of Statistics*, Vol. 8, No. 2, 1980, pp. 299-312. doi:10.1214/aos/1176344955

[14] P. J. Bickel, F. Götze and W. R. van Zwet, "The Edgeworth Expansion for U-Statistic of Degree Two," *Annals of Statistics*, Vol. 14, No. 4, 1986, pp. 1463-1484. doi:10.1214/aos/1176350170

[15] I. M. Skovgaard, "On Multivariate Edgeworth Expansions," *International Statistical Review*, Vol. 54, No. 2, 1986, pp. 169-186. doi:10.2307/1403142

[16] J. L. Jensen, "Saddlepoint Approximations," Clarendon Press, Oxford, 1995.

[17] M. G. Kendall and A. Stuart, "Advanced Theory of Statistics," 3rd Edition, High Wycombe, London, 1969.

Appendix. Proof of Lemma 1

We need the following identities which may be found in [17]:

$$\exp\left\{tx - \frac{1}{2}t^2\right\} = \sum_{k=0}^{\infty} \frac{t^k}{k!} H_k(x), \tag{27}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} H_k(x) = kH_{k-1}(x), \tag{28}$$

where $H_k(x)$ is the Hermite polynomial of degree k and by convention $H_0(x) = 1$.

By setting x to zero in (27) we obtain

$$\begin{aligned}\exp\left\{\frac{1}{2}(it)^2\right\} &= \sum_{k=0}^{\infty} \frac{t^k}{k!} H_k(0) \\ &= \sum_{k=0}^{\infty} \frac{(-i^2 t)^k}{k!} H_k(0) \\ &= \sum_{k=0}^{\infty} \frac{(it)^k}{k!} (-i)^k H_k(0).\end{aligned} \tag{29}$$

The left-hand side of (29) is the moment generating function of the standard normal distribution evaluated at it. It follows that

$$H_k(0) = (-i)^{-k} m_k, \tag{30}$$

for $k = 0, 1, \cdots$, where m_k is the k th moment of the standard normal distribution. Since $m_k = 0$ when k is odd, (30) may be written as

$$H_k(0) = (i)^k m_k. \tag{31}$$

To prove (2), we show that if P_k satisfies

$$\int_{-\infty}^{\infty} \phi(v)(iv - x)^k \,\mathrm{d}v = (-1)^k P_k(x), \tag{32}$$

for $k = 0, 1, \cdots$, then $P_k(x) = H_k(x)$. We first note that

$$P_k(0) = (-1)^k \int_{-\infty}^{\infty} \phi(v)(iv)^k = (-1)^k (i)^k m_k. \tag{33}$$

Again, since $m_k = 0$ when k is odd, (33) may be written as

$$P_k(0) = (i)^k m_k. \tag{34}$$

Furthermore, by differentiating (32) with respect to x we obtain

$$\begin{aligned}(-1)^k \frac{\mathrm{d}P_k(x)}{\mathrm{d}x} &= -k\int_{-\infty}^{\infty} \phi(v)(iv - x)^{k-1} \,\mathrm{d}v \\ &= k(-1)^k P_{k-1}(x),\end{aligned}$$

for $k = 1, 2, \cdots$. Thus

$$\frac{\mathrm{d}}{\mathrm{d}x} P_k(x) = kP_{k-1}(x), \tag{35}$$

for $k = 1, 2, \cdots$. It follows that $H_k(x)$ and $P_k(x)$ are the solutions of the same differential Equation (28) or (35). Furthermore, $H_0(x) = P_0(x) = 1$ and $H_k(0) = P_k(0) = (i)^k m_k$, by induction $P_k(x) = H_k(x)$ for $k = 0, 1, \cdots$. □

