Statistica Sinica **24** (2014), 1597-1611
doi:http://dx.doi.org/10.5705/ss.2012.333

SADDLEPOINT APPROXIMATION OF NONLINEAR MOMENTS

Xiaoping Shi, Xiang-Sheng Wang and Nancy Reid

St. Francis Xavier University, Southeast Missouri State University and University of Toronto

Abstract: A saddlepoint approximation to a nonlinear function of a sum of independent and identically distributed random variables is provided, using Watson's lemma applied to a double integral. The accuracy of the expansion depends on the smoothness of the nonlinear function. The result is applied to variance approximation, inverse moments, convergence rate, likelihood ratio statistics, and harmonic mean, and compared with results in the literature.

Key words and phrases: Asymptotic expansion, nonlinear moment, saddlepoint approximation, Watson's lemma.

1. Introduction

While the basic ideas of saddlepoint approximation date back to Laplace (1774), the approximation was introduced to statistics by Daniels (1954, 1987), who derived saddlepoint expansions for the density and distribution function of the sample mean. Barndorff-Nielsen and Cox (1989, 1994) give a detailed treatment and a number of examples relevant for statistical theory and higher order approximation for high order inference, and Kolassa (1994) provides a careful derivation of several expansions based on saddlepoint techniques. Jensen (1995) and Butler (2007) provide book length treatments. Butler and Wood (2004) discuss approximating the moment generating function of truncated random variables, Zhao, Cheng, and Yang (2011) introduce approximation of the moments for the stop-loss premium in insurance and Broda and Paolella (2012) give a recent review.

These applications typically consider a single integral, but we are interested in the moments of nonlinear functions of a sum that are determined by a double integral. Although a Taylor series expansion can be used to approximate the nonlinear moment, it is more difficult to control the order of the error terms when the sum in unnormalized. In this note, we derive a saddlepoint expansion for this double integral by first applying Daniels' method to approximate the probability density function, and then using Watson's lemma. Since Daniels' method itself uses Watson's lemma, this is a double application of the technique.

2. Main Result

Let $Z_i, i = 1, \ldots, n$ be a sequence of independently identical distributed (iid) random variables with probability density function $p(x) = p_1(x)$ for $x \in (x_-, x_+)$. The moment generating function is

$$M(T) = e^{K(T)} = \int_{x_-}^{x_+} e^{Tx} p(x) dx, \tag{2.1}$$

which we assume converges for real T in some nonvanishing interval containing the origin. As in Daniels (1954), let $-c_1 < T < c_2$ be the largest such interval, where $0 \leq c_1 \leq \infty$ and $0 \leq c_2 \leq \infty$ but $c_1 + c_2 > 0$. By Lebesgue's Dominated Convergence Theorem, the moment generating function is analytically defined in the strip $\{T \in \mathbb{C} : -c_1 < \Re(T) < c_2\}$.

The probability density function $p_n(x)$ for the sum $X_n = \sum_{i=1}^{n} Z_i$ is determined by the inverse Fourier integral

$$p_n(x) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} e^{nK(T)-Tx} dT, \tag{2.2}$$

where $-c_1 < \Re(T) < c_2$ on the path of integration. For a given function $f(\cdot)$,

$$E\{f(X_n)\} = \int_{nx_-}^{nx_+} f(x) p_n(x) dx = \frac{1}{2\pi i} \int_{nx_-}^{nx_+} \int_{-i\infty}^{i\infty} f(x) e^{nK(T)-Tx} dT dx, \tag{2.3}$$

where we assume the expectation exists. To find an asymptotic expansion for $E\{f(X_n)\}$ for large n, we make use of Watson's lemma (see Jeffreys and Jeffreys (1950)).

Lemma 1 (Watson's Lemma)**.** *Let $\psi(w)$ be analytic in a neighbourhood of $w = 0$ and bounded for real $w \in [-A, B]$ with positive A and B. For any $m \in \mathbb{N}$ we have as $n \to \infty$,*

$$\left(\frac{n}{2\pi}\right)^{1/2} \int_{-A}^{B} e^{-nw^2/2} \psi(w) dw = \sum_{k=0}^{m} \frac{\psi^{(2k)}(0)}{(2n)^k k!} + O(n^{-m-1}). \tag{2.4}$$

Remark 1. Following convention, (2.4) can be expressed as

$$\left(\frac{n}{2\pi}\right)^{1/2} \int_{-A}^{B} e^{-nw^2/2} \psi(w) dw \sim \sum_{k=0}^{\infty} \frac{\psi^{(2k)}(0)}{(2n)^k k!};$$

more precisely,

$$G(n) \sim \sum_{k=0}^{m} \frac{a_k}{n^k},$$

if $G(n) - \sum_{k=0}^{m} a_k/n^k = O(n^{-m-1})$, and if this holds for any $m \in \mathbb{N}$, we write $G(n) \sim \sum_{k=0}^{\infty} a_k/n^k$.

We make a change of variable $x = nK'(y)$ to transform the double integral in (2.3) into

$$E\{f(X_n)\} = \frac{n}{2\pi i}\int_{y_-}^{y_+}\int_{-i\infty}^{i\infty} f\{nK'(y)\}K''(y)e^{n\{K(T)-TK'(y)\}}dTdy, \tag{2.5}$$

where $y_- < 0$ is the root of $K'(y_-) = x_-$, and $y_+ > 0$ is the root of $K'(y_+) = x_+$. Note that

$$K'(y) = \frac{M'(y)}{M(y)} = \frac{\int_{x_-}^{x_+} xe^{xy}p(x)dx}{\int_{x_-}^{x_+} e^{xy}p(x)dx} > x_-$$

for any $y \in \mathbb{R}$, which implies $y_- = -\infty$. To illustrate the transformation from x_- and x_+ to y_- and y_+, we take two distributions as examples. For the binomial distribution with expectation denoted by θ, we have $x_- = 0$ and $x_+ = 1$, $M(y) = 1 - \theta + \theta e^y$, $K(y) = \ln(1 - \theta + \theta e^y)$, $K'(y) = 1/[1 + e^{-y}(1 - \theta)/\theta]$, so $y_- = -\infty$ and $y_+ = \infty$. For the normal distribution with mean μ and variance σ^2, we have $x_- = -\infty$ and $x_+ = \infty$, $M(y) = \exp(\mu y + \sigma^2 y/2)$, $K(y) = \mu y + \sigma^2 y/2$ and $K'(y) = \mu + \sigma y$, so $y_- = -\infty$ and $y_+ = \infty$.

Let $u = u(T, y)$ be the solution in T to the equation

$$K(T) - TK'(y) = K(y) - yK'(y) + \frac{u^2}{2}. \tag{2.6}$$

It follows that the first integral in (2.5) becomes

$$\int_{-i\infty}^{i\infty} e^{n\{K(T)-TK'(y)\}}dT = \int_{u(-i\infty,y)}^{u(i\infty,y)} e^{n\{K(y)-yK'(y)\}+nu^2/2}\phi(u, y)du, \tag{2.7}$$

where

$$\phi(u, y) = \frac{\partial T}{\partial u}(u, y). \tag{2.8}$$

Following Daniels (1954), we transform the integral contour of u in (2.7) into the steepest-descent curve $(-iA, iB)$ through the saddle point $u = 0$, where $A > 0$ and $B > 0$. An application of Watson's lemma yields

$$\frac{n}{2\pi i}\int_{-i\infty}^{i\infty} e^{n\{K(T)-TK'(y)\}}dT \sim \left(\frac{n}{2\pi}\right)^{1/2} e^{n\{K(y)-yK'(y)\}}\sum_{k=0}^{\infty}\frac{(-1)^k\phi^{(2k)}(0, y)}{(2n)^k k!}, \tag{2.9}$$

where the derivatives of ϕ are partial derivatives with respect to u. Substituting this into (2.5) gives

$$E\{f(X_n)\} \sim \left(\frac{n}{2\pi}\right)^{1/2}\int_{y_-}^{y_+} f\{nK'(y)\}K''(y)\widetilde{\phi}(y)e^{n\{K(y)-yK'(y)\}}dy, \tag{2.10}$$

where

$$\widetilde{\phi}(y) \sim \sum_{k=0}^{\infty} \frac{(-1)^k \phi^{(2k)}(0,y)}{(2n)^k k!}. \tag{2.11}$$

We make the second transformation

$$K(y) - yK'(y) = -\frac{w^2}{2}, \tag{2.12}$$

from which we have

$$E\{f(X_n)\} \sim \left(\frac{n}{2\pi}\right)^{1/2} \int_{y_-}^{y_+} e^{-nw^2/2} \psi(w) dw, \tag{2.13}$$

where

$$\psi(w) = f\{nK'(y)\} K''(y) \widetilde{\phi}(y) \frac{dy}{dw}. \tag{2.14}$$

A further application of Watson's lemma yields the following.

Theorem 1. *Assume that*

(A1) $\kappa_r := K^{(r)}(0)$ *exists for any* $0 \le r \le 2m$.

(A2) $f \in C^{2m}(nx_-, nx_+)$ *and* $f_r := n^r f^{(r)}(n\kappa_1) = O(|f_0|)$ *for any* $0 \le r \le 2m$.

Then we have the m-th order approximation

$$E\{f(X_n)\} = \sum_{r=0}^{m} \frac{\psi^{(2r)}(0)}{(2n)^r r!} + O(|f_0| n^{-m-1}). \tag{2.15}$$

Remark 2. If the condition $f_r = O(|f_0|)$ is violated, as for example in the case of an exponential function $f(\cdot)$, we can modify the definition of w in (2.12) according to the expression of $f(x)$ and then apply Watson's lemma. For example, if $f(x)$ is replaced by $f(x)e^{\beta x}$ with f satisfying (A2) and $-c_1 < \beta < c_2$ (where c_1 and c_2 are given in the first paragraph in Section 2) and the moment exists, we need to change (2.3), (2.5), and (2.10) into

$$\begin{aligned}
E\{f(X_n)e^{\beta X_n}\} &= \frac{1}{2\pi i} \int_{nx_-}^{nx_+} \int_{-i\infty}^{i\infty} f(x) e^{nK(T) - Tx + \beta x} dT dx \\
&= \frac{n}{2\pi i} \int_{y_-}^{y_+} \int_{-i\infty}^{i\infty} f\{nK'(y)\} K''(y) e^{n\{K(T) - TK'(y) + \beta K'(y)\}} dT dy \\
&\sim \left(\frac{n}{2\pi}\right)^{1/2} \int_{y_-}^{y_+} f\{nK'(y)\} K''(y) \widetilde{\phi}(y) e^{n\{K(y) - yK'(y) + \beta K'(y)\}} dy.
\end{aligned}$$

Consequently, the transformation (2.12) becomes

$$K(y) - yK'(y) + \beta K'(y) - K(\beta) = -\frac{w^2}{2}.$$

We still have (2.15) with $\psi(w)$ in (2.14) replaced by

$$\psi(w) = e^{nK(\beta)} f\{nK'(y)\}K''(y)\widetilde{\phi}(y)\frac{dy}{dw}.$$

3. Approximations

To illustrate how our theorem can be applied to obtain an approximation up to a given order, we provide details for obtaining an explicit approximation from our main result; the higher-order terms can be obtained in a similar manner. Our goal is to obtain the coefficients c_i in the asymptotic expansion

$$E\{f(X_n)\} = c_0 + \frac{c_1}{n} + \frac{c_2}{n^2} + \frac{c_3}{n^3} + \frac{c_4}{n^4} + O(|c_0|n^{-5}).$$

First, from (2.6) we obtain

$$\begin{aligned} u = {} & \sqrt{K''(y)}(T-y) + \frac{K^{(3)}(y)}{6\sqrt{K''(y)}}(T-y)^2 + \frac{3K^{(4)}(y)K''(y) - \{K^{(3)}(y)\}^2}{72\{K''(y)\}^{3/2}}(T-y)^3 \\ & + \frac{5\{K^{(3)}(y)\}^3 - 15K''(y)K^{(3)}(y)K^{(4)}(y) + 18\{K''(y)\}^2K^{(5)}(y)}{2160\{K''(y)\}^{5/2}}(T-y)^4 \\ & + u_5(y)(T-y)^5 + O(|T-y|^6), \end{aligned} \tag{3.1}$$

where

$$\begin{aligned} 51840\{K''(y)\}^{7/2}u_5(y) = {} & 72\{K''(y)\}^3K^{(6)}(y) - 72\{K''(y)\}^2K^{(3)}(y)K^{(5)}(y) \\ & - 45\{K''(y)\}^2\{K^{(4)}(y)\}^2 + 90K''(y)\{K^{(3)}(y)\}^2K^{(4)}(y) \\ & - 25\{K^{(3)}(y)\}^4. \end{aligned}$$

Inverting this yields

$$\begin{aligned} T(u,y) = {} & y + \frac{u}{\sqrt{K''(y)}} - \frac{K^{(3)}(y)}{6\{K''(y)\}^2}u^2 + \frac{5\{K^{(3)}(y)\}^2 - 3K^{(4)}(y)K''(y)}{72\{K''(y)\}^3\sqrt{K''(y)}}u^3 \\ & + \frac{45K''(y)K^{(3)}(y)K^{(4)}(y) - 40\{K^{(3)}(y)\}^3 - 9\{K''(y)\}^2K^{(5)}(y)}{1080\{K''(y)\}^5}u^4 \\ & + T_5(y)u^5 + O(|u|^6), \end{aligned} \tag{3.2}$$

where

$$\begin{aligned} 17280T_5(y)\{K''(y)\}^{13/2} = {} & 385\{K^{(3)}(y)\}^4 - 630K''(y)\{K^{(3)}(y)\}^2K^{(4)}(y) \\ & + 105\{K''(y)\}^2\{K^{(4)}(y)\}^2 + 168\{K''(y)\}^2K^{(3)}(y)K^{(5)}(y) \\ & - 24\{K''(y)\}^3K^{(6)}(y). \end{aligned}$$

By applying this to (2.8) and (2.11), we obtain

$$\widetilde{\phi}(y) = \frac{1}{\sqrt{K''(y)}} - \frac{1}{2n}\frac{5\{K^{(3)}(y)\}^2 - 3K^{(4)}(y)K''(y)}{12\{K''(y)\}^{7/2}} + \frac{\widetilde{\phi}_2(y)}{n^2} + O(n^{-3}), \tag{3.3}$$

where

$$
\begin{aligned}
1152\widetilde{\phi}(y)\{K''(y)\}^{13/2} = {} & 385\{K^{(3)}(y)\}^4 - 630K''(y)\{K^{(3)}(y)\}^2K^{(4)}(y) \\
& +105\{K''(y)\}^2\{K^{(4)}(y)\}^2 + 168\{K''(y)\}^2K^{(3)}(y)K^{(5)}(y) \\
& -24\{K''(y)\}^3K^{(6)}(y).
\end{aligned}
$$

On the other hand, denoting $K^{(r)}(0)$ by κ_r for convenience, we obtain from the second transformation (2.12) that

$$
\begin{aligned}
w = {} & \sqrt{\kappa_2}y + \frac{\kappa_3}{3\sqrt{\kappa_2}}y^2 + \frac{9\kappa_4\kappa_2 - 4\kappa_3^2}{72\kappa_2^{3/2}}y^3 + \frac{20\kappa_3^3 - 45\kappa_2\kappa_3\kappa_4 + 36\kappa_2^2\kappa_5}{1080\kappa_2^{5/2}}y^4 \\
& +w_5y^5 + O(|y|^6),
\end{aligned} \tag{3.4}
$$

where

$$
51840w_5\kappa_2^{7/2} = 360\kappa_2^3\kappa_6 - 576\kappa_2^2\kappa_3\kappa_5 - 405\kappa_2^2\kappa_4^2 + 1080\kappa_2\kappa_3^2\kappa_4 - 400\kappa_3^4.
$$

Inverting this asymptotic expansion yields

$$
\begin{aligned}
y = {} & \frac{w}{\sqrt{\kappa_2}} - \frac{\kappa_3}{3\kappa_2^2}w^2 + \frac{20\kappa_3^2 - 9\kappa_4\kappa_2}{72\kappa_2^{7/2}}w^3 + \frac{135\kappa_2\kappa_3\kappa_4 - 160\kappa_3^3 - 18\kappa_2^2\kappa_5}{540\kappa_2^5}w^4 \\
& +y_5w^5 + O(|w|^6),
\end{aligned} \tag{3.5}
$$

where

$$
17280y_5\kappa_2^{13/2} = 6160\kappa_3^4 - 7560\kappa_2\kappa_3^3\kappa_4 + 945\kappa_2^2\kappa_4^2 + 1344\kappa_2^2\kappa_3\kappa_5 - 120\kappa_2^3\kappa_6.
$$

Therefore, it follows from (3.5) and (A2) that

$$
\begin{aligned}
f\{nK'(y)\} = {} & f_0 + f_1\kappa_2y + (f_2\kappa_2^2 + f_1\kappa_3)y^2/2 + (f_3\kappa_2^3 + 3f_2\kappa_2\kappa_3 + f_1\kappa_4)y^3/6 \\
& +(f_4\kappa_2^4 + 6f_3\kappa_2^2\kappa_3 + 3f_2\kappa_3^2 + 4f_2\kappa_2\kappa_4 + f_1\kappa_5)y^4/24 + O(|f_0||y|^5) \\
= {} & f_0 + f_1\sqrt{\kappa_2}w + (\frac{f_2\kappa_2}{2} + \frac{f_1\kappa_3}{6\kappa_2})w^2 \\
& +\frac{12f_3\kappa_2^4 + 12f_2\kappa_2^2\kappa_3 - 4f_1\kappa_3^2 + 3f_1\kappa_2\kappa_4}{72\kappa_2^{5/2}}w^3 \\
& +\frac{A_4}{1080\kappa_2^4}w^4 + O(|f_0||w|^5),
\end{aligned} \tag{3.6}
$$

where

$$
\begin{aligned}
A_4 = {} & 45f_4\kappa_2^6 + 90f_3\kappa_2^4\kappa_3 - 45f_2\kappa_2^2\kappa_3^2 + 40f_1\kappa_3^3 + 45f_2\kappa_2^3\kappa_4 \\
& -45f_1\kappa_2\kappa_3\kappa_4 + 9f_1\kappa_2^2\kappa_5.
\end{aligned}
$$

Moreover, we obtain from (3.5) that

$$\begin{aligned}
K''(y) &= \kappa_2 + \kappa_3 y + \kappa_4 y^2/2 + \kappa_5 y^3/6 + \kappa_6 y^4/24 + O(|y|^5) \\
&= \kappa_2 + \frac{\kappa_3}{\sqrt{\kappa_2}} w + \frac{3\kappa_2\kappa_4 - 2\kappa_3^2}{6\kappa_2^2} w^2 + \frac{20\kappa_3^3 - 33\kappa_2\kappa_3\kappa_4 + 12\kappa_2^2\kappa_5}{72\kappa_2^{7/2}} w^3 \\
&\quad + \frac{45\kappa_2^3\kappa_6 - 216\kappa_2^2\kappa_3\kappa_5 - 135\kappa_2^2\kappa_4^2 + 630\kappa_2\kappa_3^3\kappa_4 - 320\kappa_3^4}{1080\kappa_2^5} w^4 + O(|w|^5), \quad (3.7) \\
\frac{dy}{dw} &= \frac{1}{\sqrt{\kappa_2}} - \frac{2\kappa_3}{3\kappa_2^2} w + \frac{20\kappa_3^2 - 9\kappa_4\kappa_2}{24\kappa_2^{7/2}} w^3 + \frac{135\kappa_2\kappa_3\kappa_4 - 160\kappa_3^3 - 18\kappa_2^2\kappa_5}{135\kappa_2^5} w^4 \\
&\quad + 5y_5 w^5 + O(|w|^6). \quad (3.8)
\end{aligned}$$

Multiplying (3.3), (3.6), (3.7), and (3.8), and taking into account (2.14), gives

$$\begin{aligned}
\psi(0) &= f_0 + \frac{f_0(3\kappa_2\kappa_4 - 5\kappa_3^2)}{24\kappa_2^3 n} \\
&\quad + \frac{f_0(385\kappa_3^4 - 630\kappa_2\kappa_3^2\kappa_4 + 105\kappa_2^2\kappa_4^2 + 168\kappa_2^2\kappa_3\kappa_5 - 24\kappa_2^3\kappa_6)}{1152\kappa_2^6 n^2} \\
&\quad + O(|f_0| n^{-3}), \quad (3.9) \\
\psi''(0) &= \frac{12 f_2\kappa_2^4 + 5f_0\kappa_3^2 - 3f_0\kappa_2\kappa_4}{12\kappa_2^3} + \frac{\psi_1^2}{n} + O(|f_0| n^{-2}), \quad (3.10)
\end{aligned}$$

with

$$\begin{aligned}
288\psi_1^2\kappa_2^6 &= 36 f_0\kappa_2^3\kappa_6 - 300 f_0\kappa_2^2\kappa_3\kappa_5 + 72 f_2\kappa_2^4\kappa_5 - 201 f_0\kappa_2^2\kappa_4^2 + 1338 f_0\kappa_2\kappa_3^2\kappa_4 \\
&\quad - 382 f_1\kappa_2^3\kappa_3\kappa_4 + 36 f_2\kappa_2^5\kappa_4 - 925 f_0\kappa_3^4 + 360 f_1\kappa_2^2\kappa_3^3 - 60 f_2\kappa_2^4\kappa_3^2,
\end{aligned}$$

and

$$\psi^{(4)}(0) = \psi_0^4 + O(|f_0| n^{-1}), \quad (3.11)$$

with

$$\begin{aligned}
144\psi_0^4\kappa_2^6 &= 144 f_4\kappa_2^8 + 192 f_3\kappa_2^6\kappa_3 + 120 f_2\kappa_2^4\kappa_3^2 - 720 f_1\kappa_2^2\kappa_3^3 + 1465 f_0\kappa_3^4 \\
&\quad - 72 f_2\kappa_2^5\kappa_4 + 768 f_1\kappa_2^3\kappa_3\kappa_4 - 2046 f_0\kappa_2\kappa_3^2\kappa_4 + 297 f_0\kappa_2^2\kappa_4^2 \\
&\quad - 144 f_1\kappa_2^4\kappa_5 + 432 f_0\kappa_2^2\kappa_3\kappa_5 - 48 f_0\kappa_2^3\kappa_6.
\end{aligned}$$

Applying (3.9), (3.10), and (3.11) to (2.15) yields

$$E\{f(X_n)\} = f_0 + \frac{f_2\kappa_2}{2n} + \frac{3f_4\kappa_2^2 + 4f_3\kappa_3}{24n^2} + O(|f_0| n^{-3}), \quad (3.12)$$

as $n \to \infty$. Recall the notations $\kappa_r = K^{(r)}(0)$ and $f_r = n^r f^{(r)}(n\kappa_1)$. A similar approach can give an approximation up to any order; the terms to $O(n^{-4})$ are, with the details omitted,

$$
\begin{aligned}
E\{f(X_n)\} = f_0 &+ \frac{f_2\kappa_2}{2n} + \frac{3f_4\kappa_2^2+4f_3\kappa_3}{24n^2} + \frac{f_6\kappa_2^3+4f_5\kappa_2\kappa_3+2f_4\kappa_4}{48n^3} \\
&+ \frac{15f_8\kappa_2^4+120f_7\kappa_2^2\kappa_3+80f_6\kappa_3^2+120f_6\kappa_2\kappa_4+48f_5\kappa_5}{5760n^4} \\
&+ O(|f_0|n^{-5}). \qquad (3.13)
\end{aligned}
$$

Remark 3. By using Taylor series approximation, we have

$$E\{f(X_n)\} = f_0 + \frac{f_2\kappa_2}{2n} + \frac{f_3E(X_n-n\kappa_1)^3}{6n^3} + \frac{f_4E(X_n-n\kappa_1)^4}{24n^4} + \cdots ,$$

where the order of successive terms is governed by the order of the central moments of X_n, and it is not straightforward to determine the order of the remainder term when the Taylor series expansion is truncated; for example, see Rosén (1970) for bounds on even moments. The saddlepoint expansion, being built on the cumulant generating function, provides an easier method for obtaining an approximation up to a desired order.

4. Examples

Example 1. *Approximate variance compared to the delta method.* Assume the Z_i are nonnegative. Using the delta method, Garcia and Palacios (2001, Theorem 1) proved

$$\sqrt{n}\Big\{\Big(1+\sum_{i=1}^{n}\frac{Z_i}{n}\Big)^{-\alpha} - (1+\kappa_1)^{-\alpha}\Big\} \to_d N\left\{0, \alpha^2(1+\kappa_1)^{-2(\alpha+1)}\kappa_2\right\}. \qquad (4.1)$$

Letting $f(x) = (1+x/n)^{-\alpha}$ with $x \geq 0$, which satisfies the conditions of Theorem 1, we have

$$E\Big\{\Big(1+\sum_{i=1}^{n}\frac{Z_i}{n}\Big)^{-\alpha}\Big\} = (1+\kappa_1)^{-\alpha}\Big\{1+\frac{c_1(\alpha)}{n}+\frac{c_2(\alpha)}{n}^2+O(\frac{1}{n^3})\Big\}, \qquad (4.2)$$

where $c_1(\alpha) = -\alpha(-\alpha-1)\kappa_2/\{2(1+\kappa_1)^2\}$ and

$$c_2(\alpha) = \frac{1}{24}\left\{\frac{3(-\alpha)(-\alpha-1)(-\alpha-2)(-\alpha-3)\kappa_2^2}{(1+\kappa_1)^4} + \frac{4(-\alpha)(-\alpha-1)(-\alpha-2)\kappa_3}{(1+\kappa_1)^3}\right\}.$$

Further, we have

$$
\begin{aligned}
&\mathrm{Var}\Big\{\sqrt{n}\Big(1+\sum_{i=1}^{n}\frac{Z_i}{n}\Big)^{-\alpha}\Big\} \\
&\quad = n\{Ef^2(X_n) - E^2f(X_n)\} \\
&\quad = n(1+\kappa_1)^{-2\alpha}\Big[1+\frac{c_1(2\alpha)}{n}+\frac{c_2(2\alpha)}{n^2}-\{1+\frac{c_1(\alpha)}{n}+\frac{c_2(\alpha)}{n^2}\}^2+O(\frac{1}{n^3})\Big] \\
&\quad = (1+\kappa_1)^{-2\alpha}\Big[\{c_1(2\alpha)-2c_1(\alpha)\}+\frac{\{c_2(2\alpha)-c_1^2(\alpha)-2c_2(\alpha)\}}{n}+O(\frac{1}{n^2})\Big]. \qquad (4.3)
\end{aligned}
$$

Note that $(1+\kappa_1)^{-2\alpha}\{c_1(2\alpha)-2c_1(\alpha)\} = \alpha^2(1+\kappa_1)^{-2(\alpha+1)}\kappa_2$. Thus the leading term in (4.3) agrees with the variance approximation in (4.1).

Example 2. *Inverse moment.* Let X_n follow a binomial distribution with mean $n\theta$. A formula was given in Chao and Strawderman (1972) for the inverse moment $E(a+X_n)^{-\alpha}$, for positive integers a and α. In particular for $a=1$ and $\alpha=1$, they obtained

$$E\left(\frac{1}{1+X_n}\right) = \frac{1-(1-\theta)^{n+1}}{(n+1)\theta}. \tag{4.4}$$

The corresponding leading term approximation is

$$E\left(\frac{1}{1+X_n}\right) \doteq \frac{1}{1+n\theta}, \tag{4.5}$$

see Garcia and Palacios (2001) and Shi, Wu, and Liu (2010). An upper bound given by Pittenger (1990) is

$$\frac{1}{1+n\theta/(1-\theta)} + \frac{1}{(1-\theta+n\theta)(1+(1+\theta(1-\theta))/(n\theta))}. \tag{4.6}$$

Since $f(x) = 1/(1+x)$ satisfies the conditions of Theorem 1, we have the fourth order approximation (3.13), where $\kappa_r = \{\log(1-\theta+\theta e^T)\}^{(r)}(0)$ and $f_r = n^r f^{(r)}(n\theta)$.

We compare the exact result (4.4), the leading term approximation (4.5), the fourth order approximation (3.13), and the upper bound (4.6) for different values of n and θ in Figure 1. It shows that the fourth order approximation is better than the leading term approximation and is closest to the exact result. To obtain more accurate approximation requires either larger values of n or more terms in the saddlepoint expansion. The upper bound is not as accurate as the leading term approximation.

The inverse moment approximation for other values of a and α can be useful in theoretical analysis, as closed-form expressions are difficult to obtain. Applications of (4.4) for non-integer a and α include Stein estimation and post-stratification, evaluation of risks of estimators or powers of tests, reliability and life testing, insurance and financial mathematics, complex systems and other problems; see Wu, Shi, and Miao (2009). Recent applications include Au and Zhang (2011) and Jordan (2011).

Example 3. *Convergence rate.* In this example, assuming that $f(x)$ satisfies the conditions of Theorem 1, we are interested in the convergence rate of $f(X_n)$ to $f(n\kappa_1)$. Using Markov's inequality, we have for arbitrary $\varepsilon, r > 0$,

$$P\{|f(X_n)-f(n\kappa_1)| \geq \varepsilon\} \leq \frac{E|f(X_n)-f(n\kappa_1)|^r}{\varepsilon^r}. \tag{4.7}$$

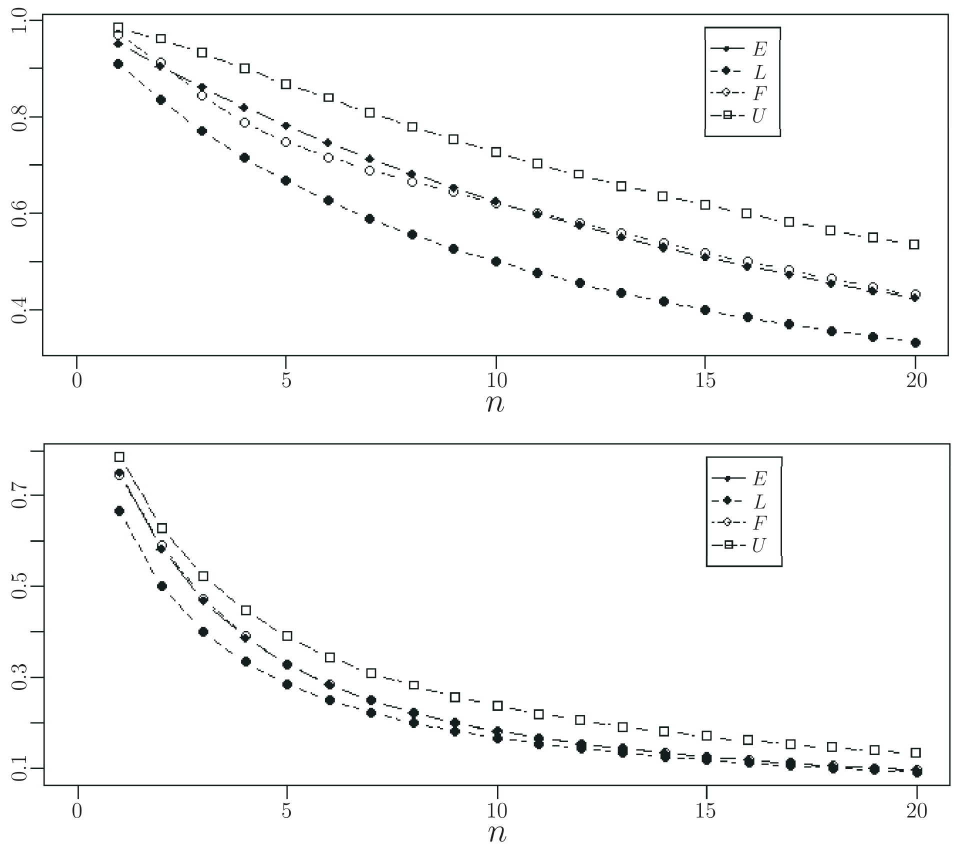

Figure 1. Comparison of inverse moment approximations for $n = 1, \ldots, 20$ and $\theta = 0.1$ (top) and $\theta = 0.5$ (bottom). The exact result (E) is obtained by Chao and Strawderman (1972), the leading term approximation (L) is given by Garcia and Palacios (2001) and Shi, Wu, and Liu (2010), the fourth order approximation (F) is shown in (3.13), and the upper bound (U) is obtained by Pittenger (1990).

To obtain the convergence rate, we need to find an accurate bound on $E|f(X_n) - f(n\kappa_1)|^r$. In view of the general inequality

$$E|f(X_n) - f(n\kappa_1)|^s \leq [E|f(X_n) - f(n\kappa_1)|^r]^{s/r}$$

for $s \leq r$, it suffices to consider the case when r is an even integer. For simplicity, we assume $r = 4$; the other cases can be studied similarly. We intend to show $E|f(X_n) - f(n\kappa_1)|^4 = O(|f(n\kappa_1)|^4/n^2)$. This formula reduces to the well known bound if $f(x) = x$. Now

$$\begin{aligned} E|f(X_n) - f(n\kappa_1)|^4 = {} & E\{f^4(X_n)\} - 4E\{f^3(X_n)\}f(n\kappa_1) + 6E\{f^2(X_n)\}f^2(n\kappa_1) \\ & - 4E\{f(X_n)\}f^3(n\kappa_1) + f^4(n\kappa_1). \end{aligned} \tag{4.8}$$

By Theorem 1, we have

$$\begin{aligned}
E\{f(X_n)\} &= f(n\kappa_1) + f''(n\kappa_1)n\kappa_2/2 + O(|f(n\kappa_1)|n^{-2}),\\
E\{f^2(X_n)\} &= f^2(n\kappa_1) + [\{f'(n\kappa_1)\}^2 + f(n\kappa_1)f''(n\kappa_1)]n\kappa_2 + O(|f^2(n\kappa_1)|n^{-2}),\\
E\{f^3(X_n)\} &= f^3(n\kappa_1) + [6f(n\kappa_1)\{f'(n\kappa_1)\}^2 + 3f^2(n\kappa_1)f''(n\kappa_1)]n\kappa_2/2\\
&\quad + O(|f^3(n\kappa_1)|n^{-2}),\\
E\{f^4(X_n)\} &= f^4(n\kappa_1) + [6f^2(n\kappa_1)\{f'(n\kappa_1)\}^2 + 2f^3(n\kappa_1)f''(n\kappa_1)]n\kappa_2\\
&\quad + O(|f^4(n\kappa_1)|n^{-2}).
\end{aligned}$$

Substituting these into (4.8) gives $E|f(X_n) - f(n\kappa_1)|^4 = O(|f(n\kappa_1)|^4/n^2)$ and hence

$$P(|f(X_n) - f(n\kappa_1))| \geq \varepsilon) \leq \frac{C|f(n\kappa_1)|^4}{n^2\varepsilon^4}. \tag{4.9}$$

Example 4. *Likelihood ratio statistic.* Consider a single Poisson process of rate ρ with the null hypothesis that $\rho = \rho_0$. If n points are observed in the total time of observation t, the log-likelihood function is

$$\ell(\rho) = n\log(\rho t) - \rho t, \tag{4.10}$$

and the likelihood ratio statistic for the hypothesis is

$$w(\rho_0) = 2[n\log\left\{\frac{n}{\rho_0 t}\right\} - (n - \rho_0 t)]. \tag{4.11}$$

Suppose n is preassigned, so that $2\rho_0 t$ follows a χ^2_{2n} distribution. Barndorff-Nielsen and Cox (1994, Example 7.4) have shown that

$$E_0\{w(\rho_0)\} = 1 + \frac{1}{6n} + O(\frac{1}{n^2}). \tag{4.12}$$

Since $\rho_0 t$ follows a gamma distribution with shape n and rate 1, we can write

$$E_0\{w(\rho_0)\} = 2n\log n - 2nE(\log X_n), \tag{4.13}$$

where $X_n = \sum_{i=1}^n Z_i$, and Z_i follows an exponential distribution with rate 1. Then the conditions of Theorem 1 are satisfied for $f(x) = \log x, x > 0$. Substituting (3.13) into (4.13) gives

$$E_0\{w(\rho_0)\} = 1 + \frac{1}{6n} - \frac{1}{60n^3} + O(n^{-4}), \tag{4.14}$$

where the first order term $1/(6n)$ is consistent with (4.12), there is no second order term n^{-2}, and for the last term we use the fact that $|f_r| = O(1)$ with $r > 0$.

Example 5. *Harmonic mean.* The harmonic mean is used to measure the price ratio in finance, and the program execution rate in computer engineering. Some statistical applications are given in Pakes (1999, Section 3). Jones (2003) describes the role of the harmonic mean in assessing results of multi-center clinical trials, and we use the model in that paper. Let $\xi_1, \ldots, \xi_n$ be iid positive Poisson random variables with parameter λ and

$$P(\xi_1 = k) = \frac{1}{1-e^{-\lambda}} \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{for} \quad k = 1, 2, \ldots$$

Set $H_n = n/X_n$ to be the harmonic mean of random variables $\xi_1, \ldots, \xi_n$, where $X_n = \sum_{i=1}^n Z_i$ and $Z_i = 1/\xi_i$. Assuming $\lambda \to \infty$, Jones (2003) derived the approximations $\kappa_1 = E(1/\xi_1) = 1/\lambda + 1/\lambda^2 + 2/\lambda^3 + O(1/\lambda^4)$ and $E(1/\xi_1^2) = 1/\lambda^2 + 3/\lambda^3 + O(1/\lambda^4)$, which imply that $\kappa_2 = 1/\lambda^3 + O(1/\lambda^4)$ and $\kappa_2/\kappa_1 = 1 + O(1/\lambda)$.

Letting $f(x) = n/x$, the first order approximation is

$$E(H_n) = E\{f(X_n)\} \doteq \frac{1}{\kappa_1} + \frac{\kappa_2}{(n\kappa_1^3)} = \lambda - 1 + \frac{1}{n} + O(\frac{1}{\lambda}),$$

by (3.13), which is consistent with Jones (2003, Section 3.3). In addition, we are able to obtain higher-order approximations using (3.12). First, making use of the identity

$$\frac{1}{k(k+m)!} = \sum_{j=m+1}^{\infty} \frac{(j-1)!}{(k+j)!m!},$$

we can extend the asymptotic formulas in Jones (2003) as follows:

$$E_1 := E\left(\frac{1}{\xi_1}\right) \sim \sum_{j=1}^{\infty} \frac{(j-1)!}{\lambda^j};$$

$$E_2 := E\left(\frac{1}{\xi_1^2}\right) \sim \sum_{j_2=2}^{\infty} \sum_{j_1=1}^{j_2-1} \frac{(j_2-1)!}{j_1 \lambda^{j_2}};$$

$$E_3 := E\left(\frac{1}{\xi_1^3}\right) \sim \sum_{j_3=3}^{\infty} \sum_{j_2=2}^{j_3-1} \sum_{j_1=1}^{j_2-1} \frac{(j_3-1)!}{j_1 j_2 \lambda^{j_3}};$$

$$E_m := E\left(\frac{1}{\xi_1^m}\right) \sim \sum_{1 \le j_1 < j_2 < \cdots < j_m < \infty} \frac{(j_m-1)!}{j_1 \cdots j_{m-1} \lambda^{j_m}}.$$

In particular, we obtain

$$E_1 := E\left(\frac{1}{\xi_1}\right) \sim \frac{1}{\lambda} + \frac{1}{\lambda^2} + \frac{2}{\lambda^3} + \frac{6}{\lambda^4} + \cdots ;$$
$$E_2 := E\left(\frac{1}{\xi_1^2}\right) \sim \frac{1}{\lambda^2} + \frac{3}{\lambda^3} + \frac{11}{\lambda^4} + \frac{50}{\lambda^5} + \cdots ;$$
$$E_3 := E\left(\frac{1}{\xi_1^3}\right) \sim \frac{1}{\lambda^3} + \frac{6}{\lambda^4} + \frac{35}{\lambda^5} + \frac{225}{\lambda^6} + \cdots .$$

Using the identities $\kappa_1 = E_1$, $\kappa_2 + \kappa_1^2 = E_2$ and $\kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3 = E_3$, it follows that

$$\kappa_1 \sim \frac{1}{\lambda} + \frac{1}{\lambda^2} + \frac{2}{\lambda^3} + \frac{6}{\lambda^4} + \cdots ;$$
$$\kappa_2 \sim \frac{1}{\lambda^3} + \frac{6}{\lambda^4} + \frac{34}{\lambda^5} + \cdots ;$$
$$\kappa_3 \sim \frac{5}{\lambda^5} + \frac{68}{\lambda^6} + \cdots .$$

With $f(x) = n/x$ and $f_r := n^r f^{(r)}(n\kappa_1)$, we have

$$f_0 = \frac{1}{\kappa_1},\ f_1 = \frac{-1}{\kappa_1^2},\ f_2 = \frac{2}{\kappa_1^3},\ f_3 = \frac{-6}{\kappa_1^4},\ f_4 = \frac{24}{\kappa_1^5},$$

and from (3.12),

$$\begin{aligned} E(H_n) = E\{f(X_n)\} &\sim \frac{1}{\kappa_1} + \frac{\kappa_2}{n\kappa_1^3} + \frac{3\kappa_2^2 - \kappa_1\kappa_3}{n^2\kappa_1^5} \\ &\sim \left(\lambda - 1 - \frac{1}{\lambda} - \frac{3}{\lambda^2}\right) + \frac{1}{n}\left(1 + \frac{3}{\lambda} + \frac{16}{\lambda^2}\right) + \frac{1}{n^2}\left(\frac{-2}{\lambda} + \frac{-27}{\lambda^2}\right). \end{aligned}$$

The above approximation neglects terms that are $O(n^{-3}) + O(\lambda^{-3})$.

5. Discussion

Using the saddlepoint approximation and Watson's lemma, we provided a general framework to derive an asymptotic expansion to nonlinear moments of a sum of independent, identically distributed random variables. We have illustrated this on several examples, and compared our result to existing approximations in literature. Our method can be applied to more general nonlinear moments.

We conclude with a short remark on some potential applications of our results. Rice (2008) uses Taylor series approximations to nonlinear moments that arise in the study of evolutionary theory. For example, the ratio of individual fitness to mean population fitness is important in studying the effects of selection.

In Shi, Wu, and Reid (2014) we extended the methods of this paper to approximate the moments of $Z_i/\sum_{i=1}^{n} Z_i$, which could in turn be used to approximate this fitness ratio.

Reviewers of an earlier version asked about simulating the moments by Monte Carlo methods, and this is certainly possible and potentially accurate enough for numerical work, but is not as useful for theoretical analysis of the type provided in Example 3. Analytical approximations are much faster, which may not be an issue for a single computation, but if the computation is embedded in a larger iteration and needs to be carried out repeatedly then the time savings may become important.

It would be of interest, although challenging, to relax the assumption of identical distributions, or of independence.

Acknowledgement

We are grateful to the reviewers for helpful comments on the first version. This work was done when XSW was visiting York University as a postdoctoral fellow of Professor Jianhong Wu. XSW would like to express his gratitude to Professor Wu for his generous support. The research of NR and XS is partially supported by the Natural Sciences and Engineering Research Council of Canada.

References

Au, S.-K. and Zhang, F.-L. (2011). On assessing the posterior mode shape uncertainty in ambient modal identification. *Probabilistic Engineering Mechanics* **26**, 427-434.

Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics.* Chapman and Hall, London.

Broda, S. A. and Paolella, M. S. (2012). Saddlepoint approximations: A review and some new applications. *Handbook of Computational Statistics* **3**, 953-983.

Butler, R. W. (2007). *Saddlepoint Approximations with Applications.* Cambridge University Press, Cambridge.

Butler, R. W. and Wood, A. T. A. (2004). Saddlepoint approximation for moment generating functions of truncated random variables. *Ann. Statist.* **32**, 2712-2730.

Chao, M. T. and Strawderman, W. E. (1972). Negative moments of positive random variables. *J. Amer. Statist. Assoc.* **67**, 429-431.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631-650.

Daniels, H. E. (1987). Tail probability approximations. *Internat. Statist. Rev.* **55**, 37-48.

Garcia, N. L. and Palacios, J. L. (2001). On inverse moments of nonnegative random variables. *Statist. Probab. Lett.* **53**, 235-239.

Jeffreys, H. and Jeffreys, B. S. (1950). *Methods of Mathematical Physics.* Cambridge University Press, Cambridge.

Jensen, J. L. (1995). *Saddlepoint Approximations.* Oxford University Press, Oxford.

Jones, C. M. (2003). Approximating negative and harmonic mean moments for the Poisson distribution. *Math. Comm.* **8**, 157-172.

Jordan, J. (2011). Randomised reproducing graphs. *Electr. J. Probab.* **57**, 1549-1562.

Kolassa, J. E. (1994). *Series Approximation Methods in Statistics.* Springer-Verlag, New York.

Laplace, P. S. (1774). Memoir on the probability of the causes of events. Mémoires de Mathématique et de Physique, Tome Sixième. (English translation by S. M. Stigler 1986. *Statist. Sci.* **1**, 364-378).

Pakes, A. G. (1999). On the convergence of moments of geometric and harmonic means. *Statist. Neerlandica* **53**, 96-110.

Pittenger, A. O. (1990). Sharp mean-variance bounds for Jensen-type inequalities. *Statist. Probab. Lett.* **10**, 91-94.

Rice, S. H. (2008). A stochastic version of the Price equation reveals the interplay of deterministic and stochastic processes in evolution. *BMC Evolutionary Biology* **8**, 262.

Rosén, B. (1970). On bounds on the central moments of even order of a sum of independent random variables. *Ann. Math. Statist.* **41**, 1074-1077.

Shi, X., Wu, Y. and Liu, Y. (2010). A note on asymptotic approximations of inverse moments of nonnegative random variables. *Statist. Probab. Lett.* **80**, 1260-1264.

Shi, X., Wu, Y. and Reid, N. (2014). Approximation to the moments of ratios of cumulative sums. *Canad. J. Statist.* **42**, 325-336.

Wu, T.-J., Shi, X. and Miao, B. (2009). Asymptotic approximation of inverse moments of nonnegative random variables. *Statist. Probab. Lett.* **79**, 1366-1371.

Zhao, K., Cheng, X. and Yang, J. (2011). Saddlepoint approximation for moments of random variables. *Front. Math. China* **6**, 1265-1284.

Department of Mathematics, Statistics and Computer Science, St. Francis Xavier University, Antigonish, Nova Scotia, B2G 2W5, Canada.

E-mail: shermanshixp@gmail.com

Department of Mathematics, Southeast Missouri State University, Cape Girardeau, Missouri 63701, USA.

E-mail: xswang@semo.edu

Department of Statistical Sciences, University of Toronto, Toronto, Ontario, M5S 3G3, Canada.

E-mail: reid@utstat.utoronto.ca

(Received November 2012; accepted November 2013)