SADDLEPOINT APPROXIMATIONS IN THE CASE OF INTRACTABLE CUMULANT GENERATING FUNCTIONS

John E. Kolassa
IBM T. J. Watson Research Center

Abstract

Saddlepoint Approximations have long been used to approximate densities and distribution functions of random variables with known cumulant generating function defined on an open interval about the origin. This approximation has very desirable asymptotic properties when approximating densities and tail probabilities for sums of random variables, and also often performs remarkably well for small sample sizes, including samples of one.

Calculating the saddlepoint approximation requires calculating the Legendre transform of the log of the cumulant generating function. In some cases this cumulant generating function may be unavailable; in other cases the Legendre transform is difficult to calculate analytically. This paper discusses modifications to the saddlepoint approximation necessary when the cumulant generating function is replaced by a similar but more tractable function whose Legendre transform can be given explicitly. Calculations for the logistic distribution are presented to illustrate the case of a known but intractable cumulant generating function, and an example involving an overdispersed binomial model is presented to illustrate the case of an unavailable cumulant generating function. An application to a random effects logistic linear model is discussed.

1. The Problem. Consider the following problem: $X_1, X_2, X_3, \ldots, X_n, \ldots$ are independent and identically distributed random variables. Assume that their common distribution is continuous, with probability density function $f_1(x)$ and cumulant generating function $K(t) = \log E[e^{tX}]$, defined on some convex set $I \subset \boldsymbol{R}$. Without loss of generality assume further that $E[X_i] = 0$. I wish to approximate the density function $f_n(s)$ and distribution function $F_n(s)$ of

$$S = \frac{X_1 + \ldots + X_n}{\sqrt{n}}$$

analytically, and am faced with three options:

a. Convolute the density n times.
b. Use the classical Edgeworth series approximation to $f_n(x)$.

c. Use a saddle point approximation to $f_n(x)$.

Option (a) is often too difficult. The Edgeworth series of option (b) often behaves poorly in the tails. In order to exercise option (c), one must solve the saddlepoint equation

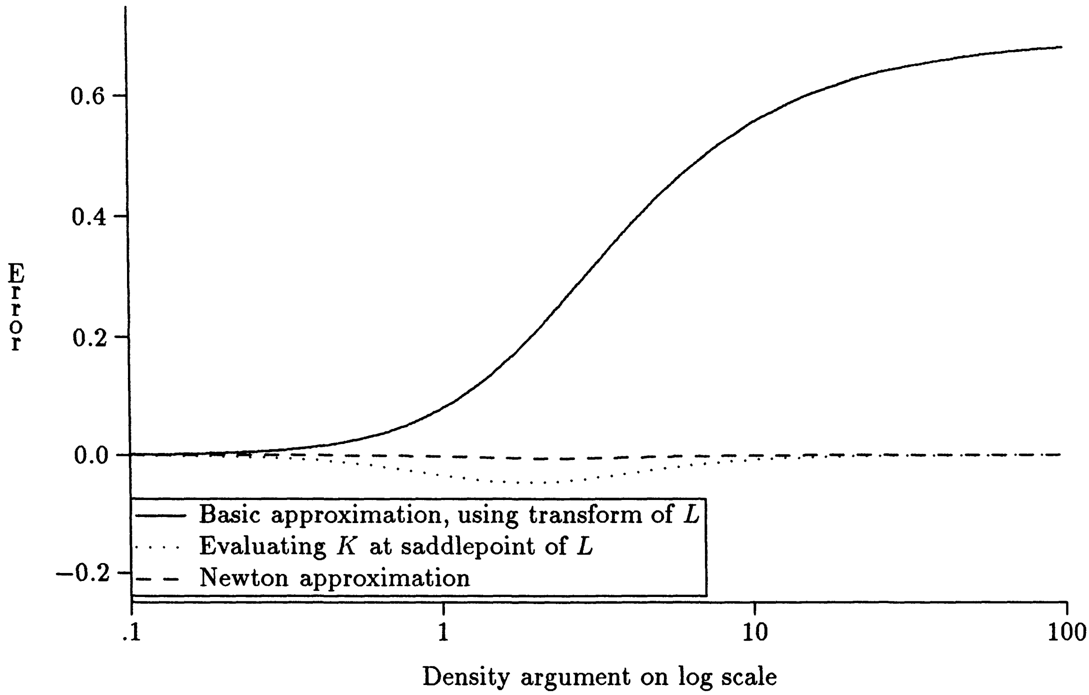
$$\sqrt{n}K'(t_e/\sqrt{n}) = s \tag{1}$$

The solution t_e is known as the saddlepoint. The value t_e is that element of the domain of K for which the tilted density, described in section 3, has its mean exactly at s. Often this is easy to solve analytically, but in some statistical calculations this is very difficult (Table 1). In these situations I propose finding a similar but computationally simpler approximate cumulant generating function L, performing the computations with the saddlepoint of L rather than that of K, and substituting it into the derivatives of K as specified in the standard saddlepoint approximation. Below I discuss the modifications to this process needed to retain good asymptotic properties. Denote the difference between these cumulant generating functions $K(t) - L(t)$ by $a(t)$. Suppose furthermore that

i. 0 is in the interior of I, the domain of K.
ii. L is analytic on $I \times i\boldsymbol{R}$
iii. $L(0) = 0$ and $L'(0) = 0$.
iv. a' is bounded on I.
v. $K(t) - tL'(t) \leq 0$ for all $t \geq 0$.

Table 1: *Examples of Legendre-Fenchel Calculations*

Distribution	Density	Domain	Cumulant Generating Function	Domain	Saddlepoint
Normal	$\frac{\exp\left[-\frac{1}{2}x^2\right]}{\sqrt{2\pi}}$	R	$\frac{1}{2}t^2$	R	x
Uniform	1	$\left[-\frac{1}{2},\frac{1}{2}\right]$	$\log\left[\frac{\sinh\left(\frac{1}{2}t\right)}{t/2}\right]$	R	intractable
Exponential	$\exp[-x]$	$[0, \infty)$	$-\log[1 - t]$	$(-\infty, 1)$	$(x - 1)/x$
Double Exponential	$\frac{1}{2}\exp[\lvert x\rvert]$	R	$-\log[1 - t^2]$	$(-1, 1)$	$(\sqrt{1+x^2} - 1)/x$
Logistic	$\frac{\exp[-x]}{(1+\exp[-x])^2}$	R	$\log\left[\frac{\pi t}{\sin(\pi t)}\right]$	$(-1, 1)$	intractable



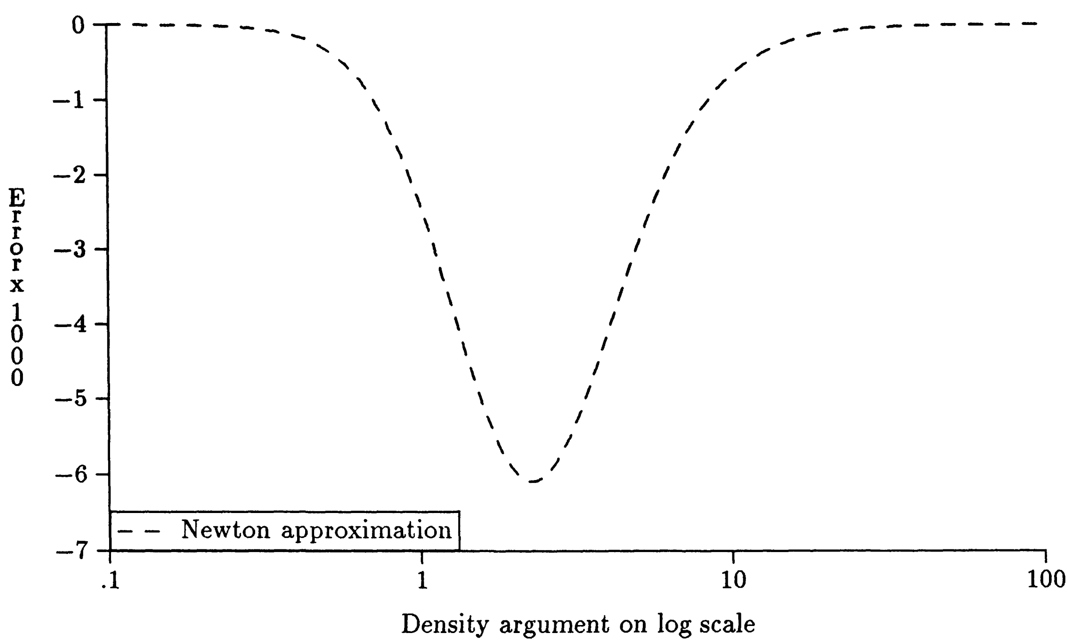


Fig. 1: The Legendre Transform for a CGF $K(t)$ and an approximating $L(t)$

In particular, the cumulant generating functions for the logistic and double exponential distributions satisfy these requirements. Note that although the saddlepoint equation $\sqrt{n}K'(t_e/\sqrt{n}) = s$ need not have a solution for all s [Daniels (1954)], condition (i) implies that a solution always exists for large enough n.

One might try to substitute the approximate saddlepoint t_a for the exact saddlepoint t_e in the expression for the saddlepoint approximation to f_n. (Figure 1). In fact, using knowledge about a, one can improve on this naive approximation.

2. The Edgeworth Series. Let the random variables X_i and S_n be defined as in Section 1. The cumulant generating function of S_n is $K_n(t) = \sqrt{n}K(t/\sqrt{n})$. Note that these functions are defined for at least t pure imaginary. In the present case, however, we assume that K exists on $I + i\mathbf{R}$ for some open interval I containing 0. One may easily show that K is infinitely differentiable at every point in this domain. Define the cumulant of order r of X_i to be $\kappa_r = K^{(r)}(0)$. Then the cumulant of order r of S_n is $\kappa_r^n = \kappa_r n^{1-r/2}$.

The density $f_n(s)$ and the cumulative distribution function $F_n(s)$ can, for large n, be well-approximated for moderate values of s, by an Edgeworth series. This series has the form

$$e_m(s;\kappa^n) = \phi(s)\left\{\sum_{k=2}^{m}\sum_{j=0}^{3(k-2)}\gamma_{j,k}h_j(s)\right\},$$

where the standard Hermite polynomials $h_r(s)$ are given by $(-1)^r\phi^{(r)}(s)/\phi(s)$. When $m = 2$, this reduces to the usual normal approximation, and $\gamma_{0,2} = 1$ and $h_0 \equiv 1$. The function ϕ is the standard normal density, and differentiation is with respect to s. The coefficients $\gamma_{j,k}$ are sums of products of the first k cumulants. McCullagh (1987, Chapter 5) gives a thorough description of this series. Feller (1971, page 535) shows that under Cramer's condition,

$$\lim_{|t|\to\infty}\sup|\exp[K(it)]| < 1,$$

that

$$f_n(t) = e_r(t;\kappa^n) + o\left(n^{-\frac{r-2}{2}}\right),$$

holds uniformly in t. Cramer's condition is always satisfied in cases where the random variables X_i have a density.

The Edgeworth series for the density can easily be integrated analytically to give an approximation $E_r(t;\kappa^n)$ to the cumulative distribution function, valid to the same order.

3. The Expansion for the Density. Denote the density of S by $f_n(s)$. Embed the density $f_n(s)$ in the exponential family $\{f_n(s;t) = \exp[ts - nK(t/\sqrt{n})]$ $f_n(s)$: $t \in I_n\}$, where I is the domain of the cumulant generating function $K(t)$ of X_1, and $I_n = \sqrt{n}I$ is the domain of the cumulant generating function $K_n(t) = nK(t/\sqrt{n})$ of S_n. Note that $f_n(s;t)$ has the cumulant generating function

$$\hat{t} \to n[K((\hat{t}+t)/\sqrt{n}) - K(t/\sqrt{n})].$$

Standard saddlepoint techniques proceed by choosing t_e to satisfy (1); here assume that a solution to this equation is not available, but that a suitable approximation L to K exists, with properties as described in section 1, and such that one can solve

$L'(t_a/\sqrt{n}) = s/\sqrt{n}$. In what follows, t_a is treated as an approximation to t_e, and is substituted into K in place of t_e. Corrections for the difference between t_a and t_e then enter the approximation through the function N of t_a described below.

By construction,

$$\begin{aligned} f_n(s) &= \exp[nK(t_a/\sqrt{n}) - t_a s] f_n(s;t_a) \\ &= \exp[nK(t_a/\sqrt{n}) - t_a s] f_n(\mu_a + \sigma_a N(t_a);t_a) \end{aligned}$$

where

$$\mu_a = \sqrt{n}K'(t_a/\sqrt{n}),$$

$$\sigma_a = \sqrt{K''(t_a\ \sqrt{n})}, \tag{2}$$

and

$$N(t_a) = (s - \sqrt{n}K'(t^a/\sqrt{n}))/\sigma_a = \sqrt{n}(L'(t_a/\sqrt{n}) - K'(t_a/\sqrt{n}))/\sigma_a$$

Approximate the density $\sigma_a f_n(\mu_a + \sigma_a N; t_a)$ as an Edgeworth series

$$e_m(N;\beta^n) = \phi(N)\left\{\sum_{k=2}^{m}\sum_{j=0}^{3(k-2)}\gamma_{j,k}h_j(N)\right\},$$

where

$$\beta_r^{n,t_a} = n^{1-r/2}K^{(r)}(t_a/\sqrt{n})/[K^{(2)}(t_a/\sqrt{n})]^{n/2}$$

for $2 \le r \le m$, and $\beta_n^{n,t_a} = 0$. Let β^{n,t_a} denote the collection of these. Denote the error in this approximation by $R_n^m(s)$. The β^{n,t_a} are the cumulants of a variable with the density $f_n(s;t_a)$, after normalizing to unit variance. The coefficients γ are the appropriate coefficients for an Edgeworth series using cumulants β^{n,t_a}. Then

$$f_n(s) = \exp[nK(t_a/\sqrt{n}) - t_a s]\frac{1}{\sigma_a}[e_m(N;\beta^n) + R_n^m(s)] \tag{3}$$

The error term is of order $O(n^{(1-m)/2})$. The size of this term depends on the cumulants β^{n,t_a}, which depend in turn on s. Daniels (1954) describes conditions on the density of the X_i which in turn imply that β_r^{n,t_a} are of order $O(n^{1-m/2})$ uniformly in t_a. The appendix of this paper provides a proof that in these cases, the above bound on the error term is uniform.

In the case where the saddlepoint equation (1) is solvable, then $K = L$, $t_a = t_e$, and N is identically 0. Then the expression (3) simplifies to the usual saddlepoint density, involving the evaluation of an Edgeworth series at 0. Note that in an Edgeworth series for a density, all odd powers of $1/\sqrt{n}$ are multiplied by odd-ordered Hermite polynomials, which are 0 when evaluated at 0. Hence, an Edgeworth series evaluated at 0 has only terms in integer powers of $1/n$. The leading term of this approximation,

$$[2\pi K^{(2)}(t_e/\sqrt{n})]^{-1/2}\exp[nK(t_e/\sqrt{n}) - t_e s], \tag{4}$$

is accurate to order $O(1/n)$ rather than $O(1/\sqrt{n})$. Note also that this first-order approximation is always positive. Daniels (1954) provides an early account of these developments.

4. The Expansion for the Tail Probabilities. I now desire an approximation to tail probabilities $Q_n(s) = P[S_n \ge s]$ for the random variable S_n described earlier,and proceed as follows:

Denote the density of S by $f_n(s)$. Again embed the density $f_n(s)$ in the exponential family $\{f_n(s;t) = \exp[ts - nK(t/\sqrt{n})]f_n(s) : t \in I_n\}$, and chose t_a such that $\sqrt{n}L'(t_a/\sqrt{n}) = s$. Let $z_a = t_a\sqrt{K''(t_a\sqrt{n})}$. Then,

$$Q_n(s) = \int_s^\infty f_n(y)\,dy$$

$$= \exp[nK(t_a/\sqrt{n}) - t_a\sqrt{n}K'(t_a/\sqrt{n})]\int_s^\infty \exp[-t_a(y - \sqrt{n}K'(t_a/\sqrt{n}))]$$

$f_n(y;t_a)dy$

$$= \exp[nK(t_a/\sqrt{n}) - t_a\sqrt{n}K'(t_a/\sqrt{n})]\int_{-N(t_a)}^\infty \exp[-z_a\xi]\,\sigma_a f_n(\mu_a + \sigma_a\xi;t_a)\,d\xi$$

where

$$\xi = (y - \sqrt{n}K'(t_a/\sqrt{n}))/\sigma_a,$$

σ_a, μ_a, and N are defined as in (2). Approximate the density $\sigma_a f_n(\mu_a + \sigma_a\xi;t_a)$ as an Edgeworth series

$$e_m(\xi;\beta^n) = \phi(\xi)\left\{\sum_{k=2}^{m}\sum_{j=0}^{3(k-2)}\gamma_{j,k}h_j(\xi)\right\},$$

where the β^{n,t_a} are as before. Then

$$Q(s) = \exp[nK(t_a/\sqrt{n}) - t_a\sqrt{n}K'(t_a/\sqrt{n})]\int_{-N(t_a)}^\infty \exp[-z_a\xi]e_m(\xi;\beta^n)d\xi + R_n^m(s)$$

where the remainder term $R_n^m(s)$ will be discussed later,

$$= \exp[nK(t_a/\sqrt{n}) - t_a\sqrt{n}K'(t_a/\sqrt{n})]\sum_{k=2}^{m}\sum_{j=0}^{3(k-2)}\gamma_{j,k}I_j(z_a, -N(t_a)) + R_n^m(s), \quad (5)$$

where $I_j(a, b) = \int_b^\infty \exp[-a\xi]\,\phi(\xi)\,h_j(\xi)\,d\xi$. Integration by parts shows that

$$I_0(a,b) = \exp\left[\frac{1}{2}a^2\right][1 - \Phi(a+b)]$$

$$I_j(a,b) = aI_{j-1}(a,b) - \exp[-ab]\,\phi(b)\,h_{j-1}(\xi),$$

whence by induction we find that

$$I_j(a,b) = \exp\left[\frac{1}{2}a^2\right]\{(-a)^j[1 - \Phi(a+b)] + \phi(a+b)\sum_{l=0}^{j-1}(-a)^l h_{j-1-l}(b)\}.$$

Hence we can in principle evaluate this series to any order desired; however, calculations beyond the first order approximation are very messy. The second appendix contains a proof that the error $R_n^m(s)$ is of order $O\left(n^{\frac{1-m}{2}}\right)$ uniformly in s. The approximation in the case when $m = 2$, accurate to order $O(1/\sqrt{n})$, is

$$\exp[nK(t_a/\sqrt{n}) - t_a\sqrt{n}K'(t_a/\sqrt{n}) + \frac{1}{2}(t_a)^2K''(t_a/\sqrt{n})\}][1 - \Phi(z_a - N(t_a))]. \tag{6}$$

Here $z_a = t_a\sqrt{K''(t_a\ \sqrt{n})}$. This is the main result of this paper.

Note that if $m = 2$, the Berry-Esseen theorem [Feller (1971), p. 542] implies that the result (5) does not depend on X_i having a density, but only on the existence of a finite third moment, which in turn is guaranteed by condition i of section 1. Hence the result (6) holds whether or not the X_i have a density.

Note also that the difference between $nK(t_a/\sqrt{n}) - t_a\sqrt{n}K'(t_a/\sqrt{n}) + \frac{1}{2}(t_a)^2K''({}^tt_a/\sqrt{n})$ and this quantity with K replaced by L is $na(t_a/\sqrt{n}) - t_a\sqrt{n}a'(t_a) + \frac{1}{2}(t_a)^2a''(t_a) = (t_a/\sqrt{n})^3a^{(3)}(\theta)/6$, for some $\theta \in [0, t_a)$, through a judicious use of Taylor's theorem. If the third derivatives of a is uniformly bounded, then, this difference is of order $O(1/\sqrt{n})$. The same argument shows that substituting L'' for K'' in the definitions of N and z_a introduces an error of order $O(1/\sqrt{n})$. Hence (6) can be changed to

$$\exp[nL(t_a/\sqrt{n}) - t_a s + \frac{1}{2}(t_a)^2L''(t_a/\sqrt{n})\ [1 - \Phi(z_a - N(t_a))] \tag{7}$$

while remaining accurate to $O(1/\sqrt{n})$ uniformly in s, and once again the requirement that the X_i have a density can be dropped.

Note that in the case where the saddlepoint equation (1) is solvable, and hence $K = L$, $t_a = t_e$, and N is identically 0, expression (6) agrees with the usual saddlepoint approximation given by Robinson (1982),

$$\exp[nK(t_e/\sqrt{n}) - t_e s + \frac{1}{2}z_m^2]\left\{[1 - \Phi(z_m)]\left(1 - \frac{\beta_3^{n,t_e}z_m^3}{6}\right) + \phi(z_m)\frac{\beta_3^{n,t_e}}{6}(z_m^2 - 1)\right\}, \tag{8}$$

to order $O(1/\sqrt{n})$. Note that the coefficient β_3^{n,t_e} has a factor of $1/\sqrt{n}$ included. Daniels (1987) discusses alternate derivations for this expansion, as well as other expansions for tail probabilities. These in turn might inspire similar variations on the expansion presented here.

5. The Choice of L. For the logistic examples considered in this paper, the cumulant generating functions are defined on an open interval (τ_1, τ_2), with $\tau_1 < 0 < \tau_2$, and there exist r_1 and r_2 such that $(\tau_i - t) K'(t) \to r_i$ as $t \to \tau_i$, for $i = 1$ and 2. Then the approximate cumulant generating function $L(t) = -\Sigma_{i=1}^{2} [r_i \log(1 - t/\tau_i) + (r_i t)/\tau_i]$ will satisfy the requirements i-v of section 1, and the resulting approximate saddlepoint t_a is the solution of the quadratic equation in t:

$$\frac{r_1}{\tau_1 - t} + \frac{r_2}{\tau_2 - t} - \frac{r_1}{\tau_1} - \frac{r_2}{\tau_2} = s.$$

Finding approximate cumulant generating functions more closely fitting the actual functions appears to be very difficult.

6. An Example: The Logistic Distribution. Consider for example the logistic distribution, whose density is $\frac{\exp[-x]}{(1 + \exp[-x])^2}$ and whose cumulant generating function is

$$K(t) = \log\left(\frac{\pi t}{\sin(\pi t)}\right) \text{ for } t \in (-1,1),$$

and set $n = 1$, so that $S = X_1$. The approximating cumulant generating function defined by the algorithm described in section 5 is

$$L(t) = -\log(1 - t^2) \text{ for } t \in (-1,1),$$

the cumulant generating function of the double exponential, with density $\frac{1}{2}\exp[-|x|]$, defined over the same interval. The saddlepoint of L is easily found to be $(\sqrt{1 + x^2} - 1\ /x)$. Straight-forward calculations verify the conditions (2) are fulfilled; the fifth condition is checked by examining coefficients of the series expansion for $K(t)$ - $tL'(t)$ and noting that they are all negative. The function L was chosen so that the difference $a(t)$ between $K(t)$ and $L(t)$ is bounded, and has a bounded first derivative.

I plotted the differences between the logarithms of various approximating densities and the logarithm of the true density (Figure 2). The approximations considered are the Edgeworth series, the Saddle Point series in which the approximating saddlepoint t_a satisfying $L'(t_a) = x$ is used naively instead of t_e satisfying $K'(t_e) = x$, the quasi-saddlepoint approximation (6), and the true saddle point series. The Edgeworth series performs best for small values of the ordinate, but for large values this approximation fails miserably. The other three approximations all perform similarly. Maximal fluctuation in the differences from the logarithm of the true density are about the same, implying that under an optimal standardization, these approximations would have similar maximal errors.

Part of the discrepancy between the approximation and the exact saddlepoint approximation can be explained by the difference in variance standardizations used. The true saddle point approximation uses the second derivative of the cumulant generating function evaluated at the true saddle point, while the approximation uses the this derivative evaluated at the saddle point of the simpler cumulant generating function. The ratio of these standardizations is far from unity when the quasi-saddlepoint approximation and the exact saddle point approximation are far apart.

The quasi-saddlepoint approximation may be improved by using a one-term Taylor series approximation in $t_e - t_a$ to $K''(t_e)$. Note that by Taylor's theorem, $K'_n(t_e) \approx L'_n(t_a) + a'_n(t_a) + (t_e - t_a)K''_n(t_a)$, and hence $t_e - t_a \approx -a'_n(t_a)/K''_n(t_a)$. Hence $K''_n(t_a) \approx K''_n(t_a) + (t_e - t_a)K'''_n(t_a) \approx K''_n(t_a) - a'_n(t_a)K'''_n(t_a)/K''_n(t_a)$.

Replacing $\sqrt{K''_n(t_a)}$ in the definition of σ_a by this expression results in a variance standardization much closer to that of the exact saddlepoint, and a resulting approximation that works better than the exact saddlepoint approximation (Figure 3).

Note that the choice of approximating cumulant generating function was crucial here. For example, if L is modified by multiplying by the ratio of the variances corresponding to K and L, in order to match variances, the resulting difference is not well-behaved, and indeed the quasi-saddlepoint approximation works poorly (Figure 4).

7. An Application. Convolution of the normal and logistic distributions has important applications in random effects logistic models where the random effect is normally distributed. Consider the following logistic model for random variables X_i taking on the values 0 or 1. The probability of realizing a 1 depends on the random effect ε_i:

$$\text{logit } P(X_i = 1 | \varepsilon_i) = \mu + \varepsilon_i$$

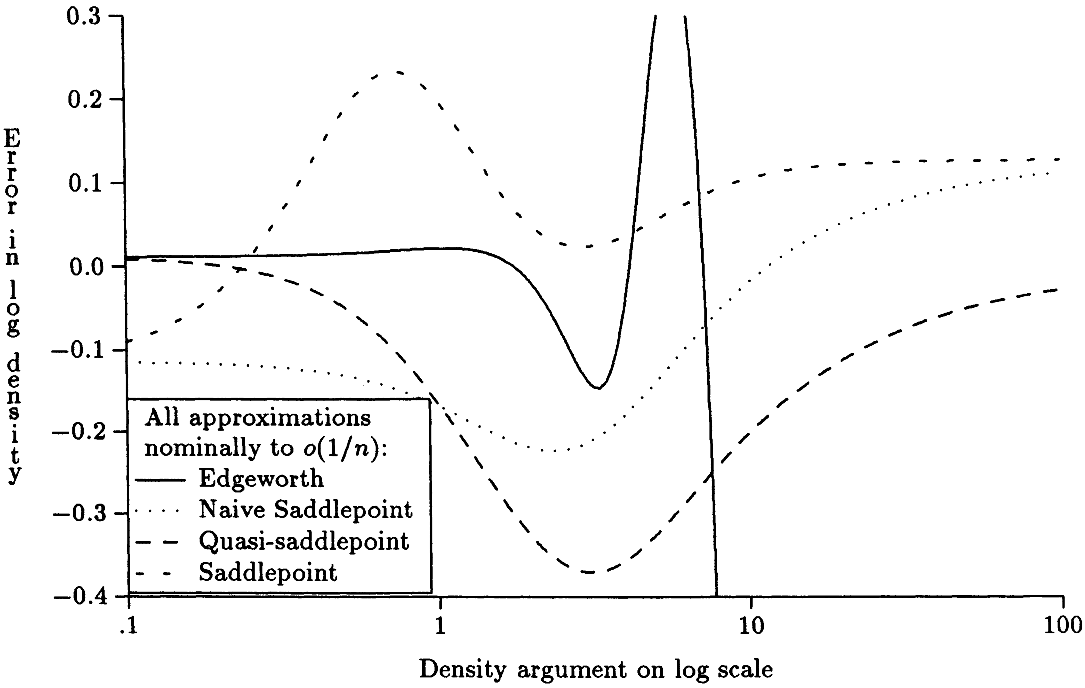


Fig. 2: Error in Unstandardized Approximations to log of Logistic Density

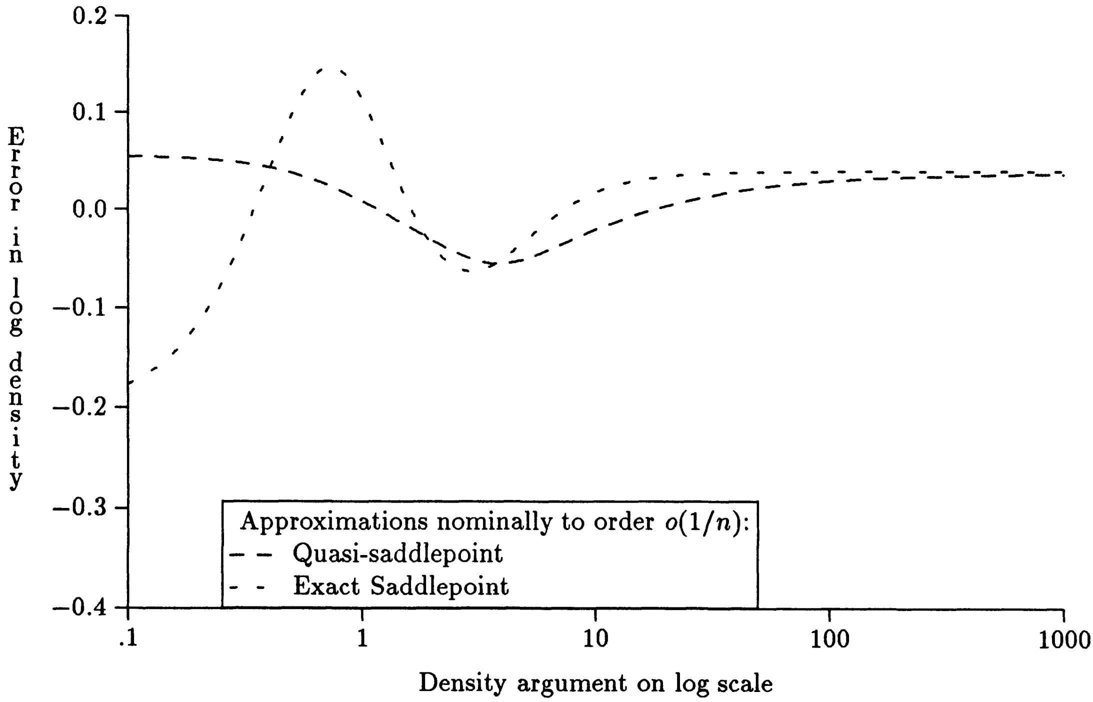


Fig. 3: Error in Approximation with Improved Variance Standardization

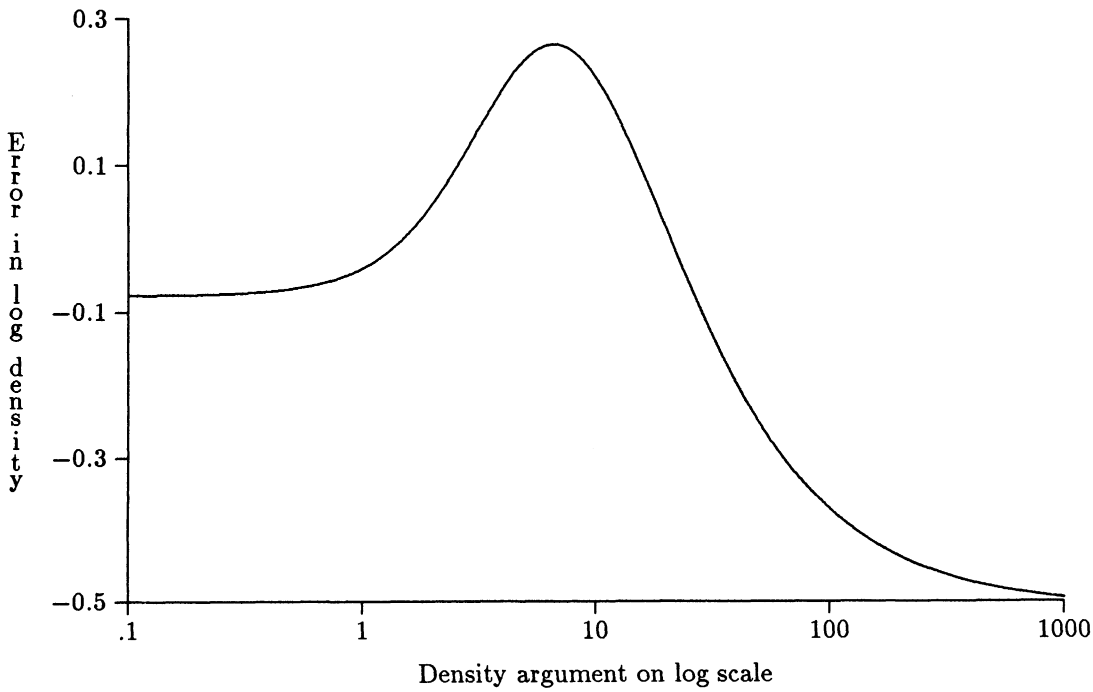


Fig. 4: Performance of Quasi-saddlepoint Approximation for Poorly Chosen L

where ε_i are normals with zero mean and variance σ^2. I wish to derive approximations to the unconditional means and variances of the $P(X_i = 1|\varepsilon_i)$'s.

Note that

$$E(P(X_i = 1|\varepsilon_i)) = \int_{-\infty}^{\infty} P(X_i = 1|\varepsilon_i)\frac{1}{\sigma}\phi\left(\frac{\varepsilon_i}{\sigma}\right)d\varepsilon_i$$

$$= \int_{-\infty}^{\infty} G_1(\mu-\varepsilon)\frac{1}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)d\varepsilon$$

where $G_1(x)$ is the cumulative distribution function $e^x/(1+e^x)$.

$$= G_2(\mu)$$

where $G_2(x)$ is the cumulative distribution function of the sum of a logistic random variable and an independent $N(0, \sigma^2)$ random variable.

Hence we see that $E(P(X_i = 1|\varepsilon_i))$ can be approximated by (6), where

$$K(t) = \log(\pi t \csc(\pi t)) + \frac{1}{2}t^2\sigma^2$$

$$L(t) = -\log(1-t^2)$$

$$a(t) = \log(\pi t(1-t^2)\csc(\pi t)) + \frac{1}{2}t^2\sigma^2$$

$$t_a = \frac{\sqrt{1+x^2}-1}{x}$$

Similarly,

$$E(P(X_i = 1|\varepsilon_i)^2) = \int_{-\infty}^{\infty} P(X_i = 1|\varepsilon_i)^2\frac{1}{\sigma}\phi\left(\frac{\varepsilon_i}{\sigma}\right)d\varepsilon_i$$

$$= \int_{-\infty}^{\infty} G_1(\mu-\varepsilon)^2\frac{1}{\sigma}\varphi\left(\frac{\varepsilon}{\sigma}\right)d\varepsilon.$$

Note that G_1^2 is an increasing differentiable function taking values in [0, 1], such that

$$\lim_{z\to-\infty} G_1^2(z) = 0$$

and

$$\lim_{z \to \infty} G_1^2(z) = 1$$

and hence is a distribution function. Differentiation shows that its density is

$$(2e^{2x}) / (1 + e^x).$$

Integration shows that its cumulant generating function is $K(t) = \log(t(1+t)\pi\csc(\pi t))$ defined on (-2, 1). It also has first moment 1. Hence

$$E(P_i^2) = G_3(\mu - 1).$$

where G_3 is the cumulative distribution function corresponding to the cumulant generating function $K(t) = \log(t(1+t)\pi\csc(\pi t)) - t + \sigma^2 t^2/2$ defined on (-2, 1). The methods of section 5 lead to the approximation $L(t) = -\log((1-t)(2+t)) - t/2 + \log(2)$, also defined on the same interval, and satisfying the conditions (2). Furthermore, the saddlepoint equation arising from L is easy to solve. Hence we see that $E(P_i^2)$ can be approximated by (6), where K and L are as just described, and

$$a(t) = \log(\pi t(1-t^2)(2+t)\csc(\pi t)) + \frac{1}{2}t^2\sigma^2 + \frac{1}{2}t - \log(2)$$

$$t_a = \frac{\sqrt{1 + \frac{9}{4}\left(x + \frac{1}{2}\right)^2} - 1}{x + \frac{1}{2}} - \frac{1}{2}.$$

These approximations appear at first glance to work well (Figures 5 and 6) but note approximation (6) performs poorly when evaluated at values of the ordinate less than the mean of the distribution. These approximations can easily be modified to calculate the lower tail when the ordinate is less than the mean. Taking care to calculate the correct tail, and subtracting from 1 if necessary, these approximations appear to work well over all, but for moderate σ, do not perform well enough to approximate the variance (Figure 7). Furthermore, the resulting discrepancy is more a result of the inability of the quasi-saddlepoint expression to approximate the true saddlepoint than the error inherent in the saddlepoint approximation.

8. A Second Example: An Overdispersed Binomial Distribution. Consider random variables X_i distributed as Bernoulli trials each with success probability p_i. I aim to calculate tail probabilities for the standardized sum S_n of

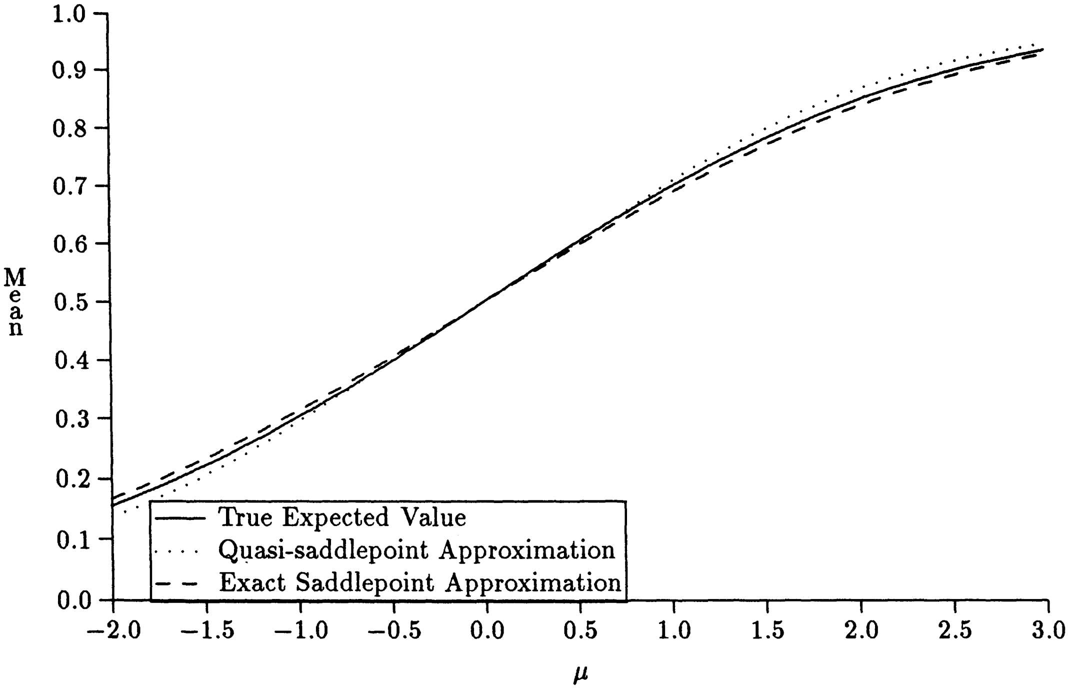


Fig. 5: Approximation to the First Moment of the Resulting Parameter

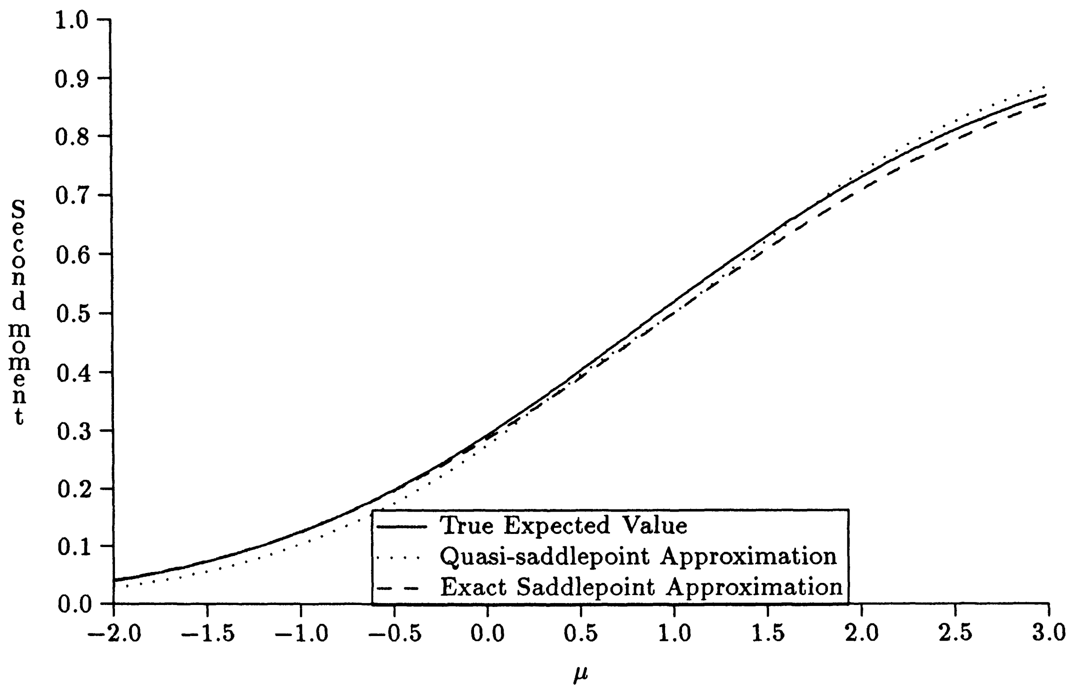


Fig. 6: Approximations to the Second Moment of the Resulting Parameter

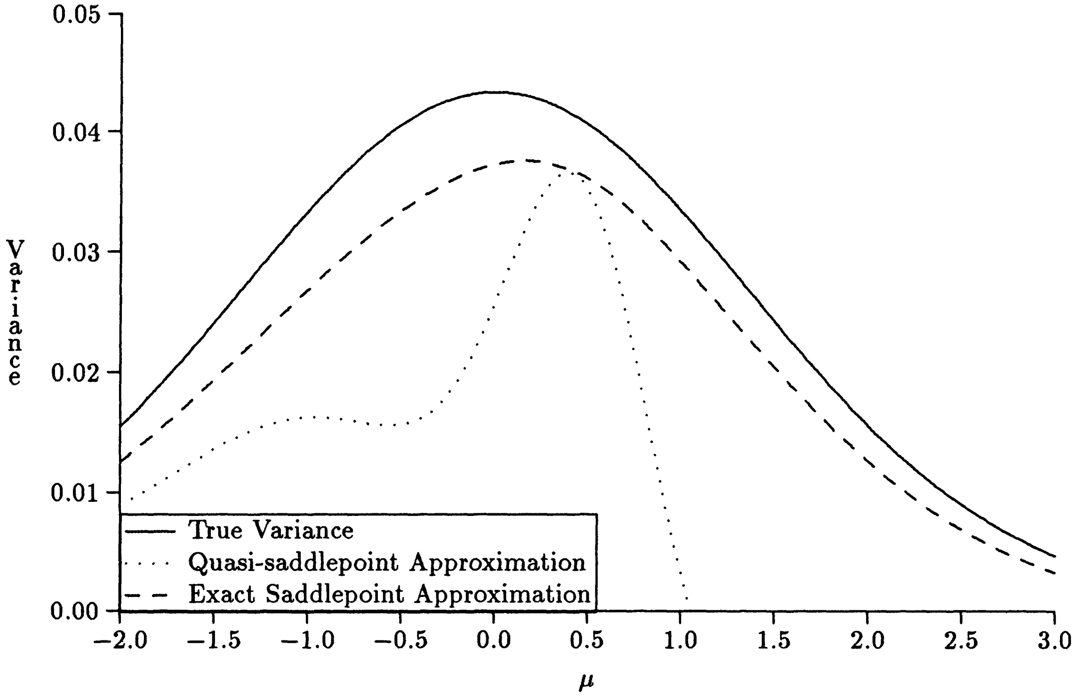


Fig. 7: Approximations to the Variance of the Resulting Parameter

the centered variables $X_i - p_i$. Were p_i all the same, exact tail probability calculations would be trivial. Were they distributed with some postulated distribution, techniques described for the previous example could be used to calculate unconditional success probabilities and also tail probabilities. Under more lax assumptions about their distribution, one might use a normal approximation to calculate unconditional tail probabilities. Unfortunately, the distribution of S_n has much lighter tails than the normal, and we might hope for better tail behavior. This section investigates what information about the p_i is necessary to apply the quasi-saddlepoint techniques to approximate tail probabilities.

Choose $p \in (0,1)$. We will approximate the exact cumulant generating function of S_n, $K_n(t) = \Sigma_{i=1}^{n} [\log(p_i(\exp(t\ /\sqrt{n}) - 1) + 1) - p_i t/\sqrt{n}]$, by the cumulant generating function that arises when all of the p_i are set equal to p: $L_n(t) = n\log(p(\exp(t\ /\sqrt{n}) - 1) + 1 - \sqrt{n}pt$. Solving the saddlepoint equation for L_n, $L_n(t_a) = s$, yields $t_a = -\sqrt{n}\text{logit}(p + s/\sqrt{n}) - \text{log}it(p)\ ]$. Note that $p + s/\sqrt{n}$ is the average of indicator variables, and hence always in [0, 1]. The difference between K_n and L_n is given by

$$a_n(t) = \sum_{i=1}^{n} \log\left[1 - \frac{(p_i - p)(\exp(t\ \sqrt{n}) - 1)}{1 + p(\exp(t\ \sqrt{n}) - 1) - (p_i - p)t\ \sqrt{n}}\right].$$

Hence

$$a'_n(t_a) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{p_i - p}{[1 + p_i(\exp(t_a\ \sqrt{n}) - 1)]\,[1 + p(\exp(t_a\ \sqrt{n}) - 1)]} - (p_i - p).$$

Lengthy but unenlightening calculations show that the third derivative of a_n is bounded over the possible choices of the p_i, and hence the error incurred by substituting L for K in (6) is of order $O(1/\sqrt{n})$ uniformly over choices of the p_i. This choice of L_n satisfies the requirements i-iv of section 1 regardless of the relation between the p_i and p. The quality of this approximation improves as p moves to represent a more central value for the p_i.

If we now assume that the p_i are distributed with mean p and some variance σ^2, the Central Limit Theorem implies that $a''_n(0)$ is asymptotically $\sigma^2 + O_p(1/\sqrt{n})$. Hence $a'_n(t_a)$ is asymptotically $t_a\sigma^2 + O_p(1/\sqrt{n})$. Also, $L''_n(t_a) = p(1-p)^2 + O(1/\sqrt{n})$. The resulting conditional tail probability approximation as a function of the random variables p_i is then

$$\exp[nL(t_a/\sqrt{n}) - t_a s + \frac{1}{2}(t_a)^2 L''(t_a \sqrt{n})]$$

$$[1 - \Phi\left(t_a\left(\sqrt{p}(1-p) - \frac{\sigma^2}{(1-p)\sqrt{p}}\right)\right)] + O_p(1/\sqrt{n}).$$

Hence this is also the unconditional probability to order $O(1/\sqrt{n})$.

Appendix 1. Order of the Error term in the Density Approximation. I aim to show that the error term in (3) is of the order stated. It suffices to prove the following theorem:

Theorem. Let $\{p_t(u)\}$ be a family of functions indexed by a parameter t, with a power series representation

$$p_t(u) = \sum_{k=3}^{m} n^{\frac{2-k}{2}} \frac{\beta_{t,k}}{k!} u^k + n^{\frac{1-m}{2}} r_{m,t}(u) \tag{9}$$

where the coefficients $\beta_{t,k}$ are bounded t in for all k, and a function $r_{m,t}(u)$ such that $|r_{m,t}(u)| \le C_m |u|^m$, then the inverse Fourier transform of $\exp(u^2/2 + p_t)$ evaluated at α is

$$\sum_{k=0}^{m-2} \sum_{j=0}^{\infty} n^{-\frac{k}{2}} \gamma_{jk} h_j(\alpha) + R_t(n), \tag{10}$$

where $\gamma = \Gamma(\beta)$ and

$$R_t(n) = o\left(n^{-\frac{m+1}{2}}\right)$$

uniformly in t. Dependence of γ is suppressed.

Proof: To see this, observe that

$$\exp(p(u)) = \left[\prod_{k=3}^{m} \exp\left[\frac{\beta_{t,k}}{k!} u^k n^{\frac{2-k}{2}}\right]\right] \exp\left[n^{\frac{1-m}{2}} r_{m,t}(u)\right]$$

$$= \left[\prod_{k=3}^{m} \sum_{r \le \frac{m}{k-2}} \frac{1}{r!}\left(\frac{\beta_{t,k}}{k!}\right)^r n^{\frac{(2-k)r}{2}} u^{rk} + R_{t,m,k}(u)\right]\left(1 + n^{\frac{1-m}{2}} R_{m,t}(u)\right) \tag{11}$$

where

$$R_{t,m,k}(u) = \frac{w_{k,m,n}^{\left[\frac{m}{k-2}\right]+1}}{\left(\left[\frac{m}{k-2}\right]+1\right)!}$$

for $w_{k,m,n}$ between 0 and $n^{\frac{2-k}{2}}\frac{\beta_{t,k}}{k!}u^k$, by Taylor's theorem.

Retaining all terms of order (2 - m)/2 and smaller in n in (11), and including all others in the remainder, gives

$$\exp(p(u)) = \sum_{k=0}^{m-2}\sum_{j=0}^{\infty} n^{-\frac{k}{2}}\gamma_{jk}u^j + R^*_{m,t}(u), \tag{12}$$

where $R^*_{m,t}(u)$ is of the form $u^m n^{\frac{1-m}{2}}$ times sums of products of the β's, negative powers of n, and positive powers of u, plus the same kind of expression times $r_{m,t}(u)$. It's absolute value can be bounded by $n^{\frac{1-m}{2}}$ times a polynomial $q_m(|u|)$ in $|u|$ whose coefficients are independent of n and t. Taking the inverse Fourier transform of (12) at α yields (10) plus a term bounded by $n^{\frac{1-m}{2}}$ times the normal density ϕ evaluated at α times a polynomial in α whose coefficients are independent of t and n.

Appendix 2. Order of the Error term in the Distribution Function Approximation I now calculate the order of the error $R_n^m(s)$ in (5). Denote the cumulative distribution function associated with $\sigma_a f_n(\mu_a + \sigma_a\xi; t_a)$ by D_n. Let

$$g(t_a) = \exp[nK(t_a/\sqrt{n}) - t_a\sqrt{n}K'(t_a/\sqrt{n})].$$

Then, using integration by parts,

$$|R_n^m(s)| = g(t_a)\left|\int_{-\alpha(t_a)}^{\infty} \exp[-z_a\xi](f_n(\mu_a+\sigma_a\xi;t_a) - e_m(\xi;\beta^n))\,d\xi\right|$$

$$= g(t_a)\left|(D_n(\xi) - E_m(\xi;\beta^n))\exp[z_a\alpha(t_a)] + \right.$$

$$\left.\int_{-\alpha(t_a)}^{\infty} \exp[-z_a\xi]\,z_a(D_n(\xi) - E_m(\xi;\beta^n))\,d\xi\right|$$

$$\le 2\left(\sup\left| \left(D_n(\xi) - E_m(\xi;\beta^n)\right)\right|\right)$$

$$\times \exp\left[nK(t_a/\sqrt{n}) - t_a\sqrt{n}K'(t_a/\sqrt{n}) + t_a\sqrt{n}a'(t_a/\sqrt{n})\right]$$

$$= 2\left(\sup\left| \left(D_n(\xi) - E_m(\xi;\beta^n)\right)\right|\right)\exp\left[nK(t_a/\sqrt{n}) - t_a\sqrt{n}L'(t_a/\sqrt{n})\right]$$

The Berry-Esseen theorem tells us that $\sup\left|(D_n(\xi) - E_m)(\xi;\beta^{n,t_a}))\right|$ is of order $o\left(n^{\frac{2-m}{2}}\right)$. Hence the error is always the order of the first term omitted, and our error bounds can be sharpened to $O\left(n^{\frac{1-m}{2}}\right)$. Furthermore, the Berry-Esseen bound involves the quantities β, which are uniformly bounded in t_a. By assumption,

$$\exp\left[nK(t_a/\sqrt{n}) - \sqrt{n}t_aL'(t_a/\sqrt{n})\right]$$

is uniformly bounded, and the error bounds given above hold uniformly in t_a.

References

[1] Blackwell, D., and Hodges, J.L. (1959). The probability in the extreme tail of a convolution. *Ann. Math. Statist.* **30** 1113-1120.

[2] Daniels, H.E. (1954). Saddlepoint Approximations in Statistics. *Ann. Math. Statist.* **25** 614-649.

[3] Daniels, H.E. (1987). Tail probability approximations. *Intern. Statist. Rev.* **55** 37-46.

[4] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, II. Wiley.

[5] Lugannani, R. and Rice, S.O. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. Appl. Prob.* **12** 475-490.

[6] McCullagh, P. (1987). *Tensor Methods in Statistics*, Chapman and Hall.

[7] Reid, N. (1988). Saddlepoint methods. *Statist. Sci.* **2** 213-238.

[8] Robinson, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *J. Roy. Statist. Soc. Ser. B* **44** 91-101.