Biometrika (1989), **76**, 3, *pp.* 457–64
Printed in Great Britain

On the empirical saddlepoint approximation

BY ANDREY FEUERVERGER
Department of Statistics, University of Toronto, Toronto, Canada M5S 1A1

SUMMARY

The properties of the saddlepoint approximation are investigated when the required cumulant generating function is obtained empirically. Properties of the empirical moment generating function and empirical cumulant generating function and derivatives of these processes which are needed for this study are derived first, in particular their uniform consistency, moment structure, and weak convergence to normality are established. A numerical investigation exploring use of the empirical saddlepoint approximation as a tool in density estimation is discussed briefly.

Some key words: Convergence; Density estimation; Empirical transform; Saddlepoint approximation; Stable law.

1. INTRODUCTION

Let $X_1, \ldots, X_n$ be identically and independently distributed random variables having probability density function $f(x)$ and moment generating function

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

defined, i.e. finite, in an interval I about the origin, and write

$$K(t) = \log M(t)$$

for the corresponding cumulant generating function. Then the saddlepoint approximation for the density of $\bar{X} = (X_1 + \ldots + X_n)/n$ is

$$f_n(x) = \left\{\frac{n}{2\pi K''(t)}\right\}^{\frac{1}{2}} \exp\,[n\{K(t) - tx\}], \qquad (1\cdot 1)$$

where $t = t(x)$ is the unique real root of the equation $K'(t) = x$. A saddlepoint approximation for the tail area $\mathrm{pr}\,(\bar{X} > x)$ may be expressed, when $x \neq \mu \equiv E(X)$, as

$$1 - F_n(x) \sim 1 - \Phi(y) + \phi(y)(1/z - 1/y), \qquad (1\cdot 2)$$

where

$$y = 2n\{tx - K(t)\}^{\frac{1}{2}}, \quad z = t\{nK''(t)\}^{\frac{1}{2}},$$

and where Φ and ϕ are the distribution function and density of the standard normal distribution. The approximations (1·1) and (1·2) are due to Daniels (1954) and Lugannani & Rice (1980) respectively. Daniels (1980, 1987) provides helpful brief expositions and Reid (1988) gives a recent general review.

Our object here is to study the consequences of replacing $K(t)$ in (1·1) and (1·2) by its sample version

$$K_n(t) = \log M_n(t), \tag{1·3}$$

where

$$M_n(t) = \frac{1}{n} \sum_{i=1}^{n} e^{tX_i}. \tag{1·4}$$

The potential interest of such modifications derives from a number of distinct sources. For example, in all but the most straightforward cases, the exact analytic form of the transforms $M(t)$ and $K(t)$ will not be tractable. On the other hand, these transforms may be estimated empirically as (1·4) and (1·3) when sampling from $f(x)$ is possible, and the behaviour of the resulting empirical saddlepoint approximation is then of interest. A similar situation arises also in cases where $f(x)$ itself is not available, but where a sample may be obtained. Other potentially important applications are given by Davison & Hinkley (1988) where saddlepoint approximations are applied in the context of the bootstrap and other resampling schemes. Of course, these applications provide only a part of the overall reason for studying such empirical approximations, since an overriding feature of this problem is that the empirical saddlepoint in fact constitutes an especially interesting application within the purview of statistical transform methods.

In § 2 we detail an investigation of those properties of the relevant empirical transforms that have statistical bearing, in particular their uniform consistency, moment structure and asymptotic normality. Related results in a different context are given by S. Ghosh in a Toronto doctorial dissertation. Our main results are given in § 3, where we study the properties of the empirical saddlepoint approximation. Finally, we comment briefly on a numerical investigation into the possible use of the empirical saddlepoint approximation as a tool in density estimation.

2. Properties of the empirical transforms

To study the consequences of replacing $K(t)$ by $K_n(t)$ in expressions such as (1·1) it is necessary to understand the properties of the transforms $M_n(t)$ and $K_n(t)$. By the law of large numbers we have that $M_n(t) \to M(t)$ almost surely at any fixed t. It follows that also $K_n(t) \to K(t)$ almost surely by continuity of the logarithm. This convergence is in fact uniform and extends also to the derivatives of these processes. The notation D^γ represents differentiation applied γ times.

THEOREM 2·1. *Let $-\infty < a \leqslant b < \infty$ both lie in the interval I on which $M(t)$ is finite. Then almost surely*

$$\sup_{a \leqslant t \leqslant b} |M_n(t) - M(t)| \to 0, \tag{2·1}$$

$$\sup_{a \leqslant t \leqslant b} |K_n(t) - K(t)| \to 0. \tag{2·2}$$

If further a, b are both interior points of I then the derivative processes satisfy almost surely

$$\sup_{a \leqslant t \leqslant b} |D^\gamma M_n(t) - D^\gamma M(t)| \to 0, \tag{2·3}$$

$$\sup_{a \leqslant t \leqslant b} |D^\gamma K_n(t) - D^\gamma K(t)| \to 0, \tag{2·4}$$

for all γ.

Proof. First note that $M_n(t)$ and $M(t)$ are convex, being the average and integral of the functions e^{tx} which are convex. By the strong law of large numbers, $M_n(t) \to M(t)$ almost surely for any t and hence for all t in any countable collection $\{t_i\}$. Then (2·1) follows because, for convex functions, convergence on a dense subset implies uniform convergence on compact subsets (Roberts & Varberg, 1973, § 13). The result (2·2) follows from (2·1).

To prove (2·3) note that $D^\gamma M(t) = \int x^\gamma e^{tx}\, dF(x)$ and that $X^\gamma e^{tX}$ has finite expectation whenever t is in the interior of I, and also that $x^\gamma e^{tx}$ is convex for γ even. For γ odd, the convexity fails but the argument may be applied separately to the components from $(-\infty, 0)$ and $(0, \infty)$ in the definitions of $M_n(t)$ and $M(t)$. Finally, (2·4) follows because $D^\gamma K(t)$ has the form $P(t)/\{M(t)\}^{2^\gamma}$, where P is a polynomial function in the variables $D^q M(t)$, for $q = 0, 1, \ldots, \gamma$. □

The result (2·1) for $M_n(t)$ is given by Csorgö (1980) but with an incomplete proof.

Note that $M_n(t)$ and its derivatives are unbiased estimators for $M(t)$ and its corresponding derivatives, that is

$$E\{D^\gamma M_n(t)\} = D^\gamma M(t) \quad (\gamma = 0, 1, \ldots).$$

Concerning the covariance structure of these quantities we have the following simple result.

LEMMA 2·2. *If s, t, $s+t \in I$, then for any integers $\alpha \geq 0$, $\beta \geq 0$ we have*

$$n \operatorname{cov}\{D^\alpha M_n(s), D^\beta M_n(t)\} = D^{\alpha+\beta} M(s+t) - \{D^\alpha M(s)\}\{D^\beta M(t)\} \tag{2·5}$$

independently of n.

On the half interval $I/2$ the variance of $M_n(t)$ exists and the multivariate central limit theorem in effect implies that, for $t_1, \ldots, t_k \in I/2$, $\{M_n(t_1), \ldots, M_n(t_k)\}$ is asymptotically normal with mean $\{M(t_1), \ldots, M(t_k)\}$ and covariance matrix having (i, j)th entry $M(t_i + t_j) - M(t_i)M(t_j)$. Csorgö (1980) proved weak convergence for the normalized $M_n(t)$ process. In fact more generally, that result may be extended to cover also the associated derivative processes.

THEOREM 2·3. *The sequence of processes*

$$Y_n(t) = n^{\frac{1}{2}}\{M_n(t) - M(t)\}$$

in the space of continuous functions under the supremum norm, converges weakly to a Gaussian process having zero mean and covariance structure identical to $Y_n(t)$, on any finite, closed interval $[a, b] \subseteq I/2$. Further, if $[a, b]$ lies in the interior of $I/2$, then for $\gamma = 1, 2, \ldots$ the derivative process $D^\gamma Y_n(t)$ converges weakly to a zero mean Gaussian process having the covariance structure given by (2·5) *with $\alpha = \beta = \gamma$.*

Proof. By (2·5) we have at once that

$$\begin{aligned} E\{D^\gamma Y_n(s) - D^\gamma Y_n(t)\}^2 = {} & \{D^{2\gamma} M(2s) + D^{2\gamma} M(2t) - 2D^{2\gamma} M(s+t)\} \\ & - \{D^\gamma M(s) - D^\gamma M(t)\}^2. \end{aligned} \tag{2·6}$$

Then, using an elementary Taylor expansion argument based on the twice differentiability of $D^{2\gamma} M(.)$ and $D^\gamma M(.)$, we find that (2·6) is bounded by $C(s-t)^2$ so that (Billingsley, 1968, Th. 12·3) tightness, and hence the claimed weak convergences, follow. □

The moment structure of $K_n(t)$ is more involved than that of $M_n(t)$. Existence questions can be studied by means of the inequality

$$t\bar{X} \leqslant K_n(t) \leqslant \max_{1\leqslant i\leqslant n} (tX_i),$$

whose left part involves Jensen's inequality. Because the sample mean $\bar{X}$ and the extremal order statistics possess the same number of moments as an individual X_i, it follows that $K_n(t)$ does also, and hence in particular that all moments of $K_n(t)$ will exist for $t \in I$ when $M(.)$ is finite in an interval about the origin. In any case, we will require only the asymptotic mean and variance of $K_n(t)$ as given in Theorem 2·4 below.

Using standard arguments, it follows from the asymptotic normality of $M_n(t)$ and differentiability of the logarithm that $K_n(t)$ is also asymptotically normal. In fact the weak convergence can be shown to also be induced using only elementary Taylor-expansion based arguments. The analogous results apply also to the derivative processes of any order and a notationally simple representation for their general covariance structure can be obtained.

THEOREM 2·4. *The processes*

$$Z_n(t) = n^{\frac{1}{2}}\{K_n(t) - K(t)\}$$

converge weakly on any finite interval $[a, b] \subseteq I/2$ *to a zero mean Gaussian process having the covariance function*

$$R(s, t) = \frac{M(s+t)}{M(s)M(t)} - 1.$$

The corresponding result holds also for each of the derivative processes

$$D^{\gamma}Z_n(t) = n^{\frac{1}{2}}\{D^{\gamma}K_n(t) - D^{\gamma}K(t)\} \quad (\gamma = 1, 2, \ldots)$$

but with $[a, b]$ *further restricted to lie in the interior of* $I/2$. *These processes converge weakly to zero mean Gaussian processes with covariance structures identical to the asymptotic covariance functions for* $D^{\alpha}Z_n(s)$ *with* $D^{\beta}Z_n(t)$ *obtained from the representation*

$$\operatorname{acov}\{D^{\alpha}Z_n(s), D^{\beta}Z_n(t)\} = D_s^{\alpha} D_t^{\beta} R(s, t), \tag{2·7}$$

which holds for any nonnegative integers α, β. *Here the subscripts on D indicate the variable of partial differentiation.*

The algebraic details regarding (2·7) are tedious; the proof of the theorem is otherwise fairly straightforward.

In view of Theorems 2·3 and 2·4 we may refer to $I/2$ as the zone of normal convergence for the various empirical processes which we have defined. On the part of I outside $I/2$ the situation is substantially more complicated. For some essential background on this, see, for example, § 5, Ch. 17 of Feller (1971), especially Theorems 1, 2 and 3. In particular, pointwise asymptotic normality and the usual $n^{\frac{1}{2}}$ convergence rate associated with the Central Limit Theorem no longer apply. For such values of t, a suitably renormalized version of $M_n(t)$ will possess a nondegenerate limiting distribution if and only if e^{tX} satisfies the necessary and sufficient conditions for belonging to some domain of attraction, and that limiting distribution will be a stable law having index $\alpha = c/t$, where c is that boundary point of I having the same sign as t. But whether or not a limiting distribution actually exists, we will have $n^{\delta}\{M_n(t)-M(t)\}\to 0$ almost surely for every $\delta<\delta_0$, and diverging for every $\delta>\delta_0$, where $\delta_0=\alpha^{-1}-1$. From this it follows that almost surely

$$K_n(t)-K(t)=\log\frac{M_n(t)}{M(t)}=\log\{1+o(n^{-\delta})\}=o(n^{-\delta})$$

for all $\delta<\delta_0$. Using similar arguments, together with the method used in the proof of Theorem 2·1, we are led to the following result.

THEOREM 2·5. *For t in I but outside $I/2$, and nonnegative integers γ, we have*

$$D^{\gamma}\{M_n(t)-M(t)\}=o(n^{-\delta}),\quad D^{\gamma}\{K_n(t)-K(t)\}=o(n^{-\delta})$$

almost surely for every $\delta<\delta_0=\alpha^{-1}-1$. The error term may be taken as being uniform over any subinterval of I that includes points outside $I/2$, provided that δ_0 is its minimum value on that interval.

In interpreting the statement of this result, the value of δ_0 should be taken to be $\frac{1}{2}$ on $I/2$. Note also that the intervals of uniform convergence in the present case may include the endpoints of I even in the cases $\gamma>0$.

3. THE EMPIRICAL SADDLEPOINT APPROXIMATION

Let $\hat{f}_n(x)$ denote the empirical saddlepoint approximation, namely $\hat{f}_n(x)$, identical to $f_n(x)$ of (1·1) except with K_n replacing K and $\hat{t}$ replacing t where $\hat{t}$ is defined through $K_n'(\hat{t})=x$. Then it is quite readily established that $\hat{f}_n(x)$ is unsatisfactory as an estimator for $f_n(x)$. To see this heuristically, consider the ratio

$$\frac{\hat{f}_n(x)}{f_n(x)}=\left\{\frac{K''(t)}{K_n''(\hat{t})}\right\}^{\frac{1}{2}}\exp[n\{K_n(\hat{t})-\hat{t}K_n'(\hat{t})\}-n\{K(t)-tK'(t)\}]. \qquad (3\cdot 1)$$

By (2·4) the fractional term on the right in (3·1) tends to unity; in fact it does so at rate $O_p(n^{-\frac{1}{2}})$ in the zone $t\in I/2$ of normal convergence of the empirical quantities. Thus the requirement that $\hat{f}_n(x)/f_n(x)\to 1$ is equivalent to the requirement that the exponent term in (3·1) tend to zero. However the quantities in the curly brackets of the exponent of (3·1) are a sample and a population quantity and their difference can be expected to be $O_p(n^{-\frac{1}{2}})$ in the zone of normal convergence. Hence overall (3·1) will be $O_p(n^{\frac{1}{2}})$ in $I/2$; at $t=0$, it may be shown using a more detailed argument to be $O_p(1)$. Consequently $\hat{f}_n(x)$ is unsatisfactory as an estimator for $f_n(x)$.

On the other hand if instead of K_n above we use K_m based on a sample size m substantially larger than n, then the $\hat{f}_n$ so obtained may be a good estimator for $f_n(x)$. However, before we can state our next result, we must attend to a technical difficulty associated with the fact that (3·1) is defined only in the intersection of the regions where each of $\hat{f}_n(x)$ and $f_n(x)$ is defined. Now $\hat{f}_n(x)$ is defined on the region of x-values where $K_n'(t)=x$ has a solution, and this may easily be shown to consist of the interval $(X^{(1)}, X^{(n)})$ from the smallest to largest order statistics. For $f_n(x)$ however the situation is more involved since $f_n(x)$ is defined for all x for which $K'(t)=x$ has a solution and thus for all $x\in K'(I)$, and this region need not correspond, as one would wish for the purposes of the arguments below, to the support range, i.e. to the convex support, of $f(x)$. In order to avoid this situation we shall make use of the following.

Assumption 1. As t tends to the upper and lower boundary points of I, $K'(t)$ tends respectively to the upper and lower points in the support range of $f(x)$.

This assumption is discussed in detail by Daniels (1954, § 6); in the context of exponential families, the assumption is known as steepness. See, for example, Barndorff-Nielsen (1978, Th. 9.2). Of course in cases where Assumption 1 does not hold, the stated results would remain true provided that x is additionally restricted to lie within the range of values assumed by $K'(t)$.

THEOREM 3·1. *Suppose the conditions of Assumption 1 are met, and let $\hat{f}_{m,n}(x)$ denote the saddlepoint approximation $f_n(x)$ of* (1·1) *but based now on the sample cumulant function $K_m(t)$ from a sample of size m. Then*

$$\frac{\hat{f}_{m,n}(x)}{f_n(x)} = 1 + O_p(m^{-\frac{1}{2}}n), \tag{3·2}$$

where the error term is uniform over any interval of x-values corresponding to an interval of t-values interior to the domain $I/2$ of normal convergence. On subintervals of I containing points outside $I/2$, the convergence result (3·2) *still holds uniformly but the error term must be taken as $o_p(m^{-\delta}n)$ for $\delta < \delta_0$, where δ_0 is the smallest value of $\alpha^{-1}-1$ occurring on that part of the interval outside $I/2$. Here, as before, $\alpha = c/t$, where c is the boundary point of I having the same sign as t.*

The analytical details of the proof are omitted; they can be obtained from the author.

One useful way to think of the results so far is to note that basically $K_n(t)$ is a good estimator of $K(t)$ by virtue of Theorem 2·1, while $n^{\frac{1}{2}}K_n(t)$, as an estimator for $n^{\frac{1}{2}}K(t)$, is 'on the boundary' because the difference $n^{\frac{1}{2}}\{K_n(t)-K(t)\}$ is $O_p(1)$ under the conditions of Theorem 2·4. The fact that $nK_n(t)$ is a poor estimator for $nK(t)$ is then just a consequence of this last assertion for we will then have $n\{K_n(t)-K(t)\} = O_p(n^{\frac{1}{2}})$.

It is possible to compare mean-corrected versions of the true and empirical saddlepoint approximations by replacing $K(t)$ and $K_n(t)$ by $K(t)-\mu t$ and $K_n(t)-\bar{X}t$ respectively, before applying the foregoing analyses. We omit the details but note that the resulting ratio $\hat{f}_n(x)/f_n(x)$ of centred terms will then be $O_p(n^{-\frac{1}{2}})$ at $t=0$, and not $O_p(1)$ as before. For other values of t in $I/2$ the ratio is $O_p(n^{\frac{1}{2}})$ as before, but if $t = O(n^{-\frac{1}{2}})$, which essentially corresponds to normalization of the densities being compared, then the uncentered ratio becomes $O_p(1)$ while the centred ratio becomes $O_p(n^{-\frac{1}{2}})$.

The normalized case is important in connexion with certain asymptotic bootstrap results (Bickel & Freedman, 1981, Th. 2·1); the following result is thus of interest.

THEOREM 3·2. *Let $g_n(x)$ denote the saddlepoint approximation for the normalized variable $n^{\frac{1}{2}}(\bar{X}-\mu)$, where $\mu = E(X)$ and let $\hat{g}_n(x)$ denote the correspondingly normalized empirical saddlepoint approximation, but centred now at $\bar{X}$ instead of μ. Then*

$$\frac{\hat{g}_n(x)}{g_n(x)} = 1 + O_p(n^{-\frac{1}{2}}). \tag{3·3}$$

Proof. To obtain (3·3) note that

$$g_n(x) = n^{-\frac{1}{2}}f_n\{t(\mu + n^{-\frac{1}{2}}x)\}, \quad \hat{g}_n(x) = n^{-\frac{1}{2}}\hat{f}_n\{\hat{t}(\bar{X} + n^{-\frac{1}{2}}x)\}$$

so that

$$\frac{\hat{g}_n(x)}{g_n(x)} = \left\{\frac{K''(t)}{K_n''(\hat{t})}\right\}^{\frac{1}{2}} \exp\left[n\{K_n(\hat{t}) - K(t) - \bar{X}\hat{t} + \mu t\} - n(\hat{t}-t)\frac{x}{n^{\frac{1}{2}}}\right], \tag{3·4}$$

where now $K'(t) = \mu + n^{-\frac{1}{2}}x$ and $K_n'(\hat{t}) = \bar{X} + n^{-\frac{1}{2}}x$. The fractional term on the right in (3·4) is now $1 + O_p(n^{-\frac{1}{2}})$, and since

$$K(t) = \mu t + \tfrac{1}{2}\sigma^2 t^2 + O(n^{-3/2}), \quad K_n(\hat{t}) = \bar{X}\hat{t} + \tfrac{1}{2}S^2\hat{t}^2 + O_p(n^{-3/2}),$$

where σ^2, S^2 are the variances of the actual and empirical distributions, the exponent in (3·4) equals

$$n\{\tfrac{1}{2}(S^2\hat{t}^2-\sigma^2t^2)-n^{-\frac{1}{2}}(\hat{t}-t)x\}+O_p(n^{-\frac{1}{2}}). \tag{3·5}$$

But since

$$\mu+n^{-\frac{1}{2}}x=K'(t)=\mu+\sigma^2t+O(n^{-1}),\quad \bar{X}+n^{-\frac{1}{2}}x=K'_n(\hat{t})=\bar{X}+S^2\hat{t}+O_p(n^{-1})$$

then $t=O(n^{-\frac{1}{2}})$ and $\hat{t}=O_p(n^{-\frac{1}{2}})$ and therefore we find, in turn,

$$S^2\hat{t}-\sigma^2t=O_p(n^{-1}),\quad \hat{t}-t=\sigma^{-2}\{(S^2\hat{t}-\sigma^2t)+(\sigma^2-S^2)\hat{t}\}=O_p(n^{-1}),$$
$$S^2\hat{t}^2-\sigma^2t^2=\{S(\hat{t}-t)+(S-\sigma)t\}(S\hat{t}+\sigma t)=O_p(n^{-3/2}).$$

Consequently the exponent term (3·5) is $O_p(n^{-\frac{1}{2}})$. □

Assumption 1 is not required for Theorem 3·2 as may be seen from simple arguments involving the fact that $t=O(n^{-\frac{1}{2}})$ uniformly on finite x-intervals.

Essentially similar analyses may be carried out for empirical versions of the tail area approximation (1·2), for Edgeworth expansions, and quantities such as the Chernoff index $\inf_z e^{-z(t+\mu)}M(z)$ for large deviation probabilities (Serfling, 1980, Ch. 10).

Finally, because the saddlepoint approximation $f_n(x)$ tends often to be extremely close to the distribution of $\bar{X}$ even for very small values of n, it seems natural to consider using empirical versions of the $n=1$ saddlepoint approximation $\hat{f}_{m,1}(x)$, as a non-parametric density estimator for $f(x)$. Since

$$\hat{f}_{m,1}(x)-f(x)=\{\hat{f}_{m,1}(x)-f_1(x)\}+\{f_1(x)-f(x)\} \tag{3·6}$$

the error of this density estimate consists of two additive components, the first representing essentially sampling variability, and the second mainly bias, it being the error of the $n=1$ saddlepoint approximation, i.e. the error involved in approximating the density $f(x)$ by inversion of $K(t)$ using Laplace's method of approximation for evaluating the inversion integral.

A numerical study of the two components of error in (3·6) showed that $\hat{f}_{m,1}$ was quite generally a reliable estimator of $f_1(x)$ while, as could be expected, $f_1(x)$ was often, but not always, very close to $f(x)$. Presumably this bias might be reduced through prior adjustment of the sample or by the inclusion of higher terms in the saddlepoint approximation but this matter requires further investigation. Further details may be found in a technical report available on request from the author.

Acknowledgements

This work was supported in part by a grant from the National Sciences and Engineering Research Council of Canada. The author is indebted to A. C. Davison and D. V. Hinkley for a preprint of their paper, to L. Devroye, R. Lockhart, N. Reid and R. Tibshirani for valuable discussions, to L. Wasserman for computing assistance, and to the referees for their valuable suggestions.

References

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory.* Chichester: Wiley.

Bickel, P. J. & Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* 9, 1196–217.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* New York: Wiley.

CSORGÖ, S. (1980). The empirical moment generating function. In *Colloquia Mathematica Societatis Janos Bolyai,* 32, *Nonparametric Statistical Inference, Budapest,* Ed. I. Vincze, pp. 139-50. Amsterdam: North-Holland.

DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631-50.

DANIELS, H. E. (1980). Exact saddlepoint approximations. *Biometrika* **67**, 59-63.

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

DAVISON, A. C. & HINKLEY, D. V. (1988). Saddlepoint approximations in resampling methods. *Biometrika* **75**, 417-31.

FELLER, W. (1971). *An Introduction to Probability Theory and its Applications,* **2**, 2nd ed. New York: Wiley.

LUGANNANI, R. & RICE, S. (1980). Saddlepoint approximation for the distribution of a sum of independent random variables. *Adv. Appl. Prob.* **12**, 475-90.

REID, N. (1988). Saddlepoint methods and statistical inference (with discussion). *Statist. Sci.* **3**, 213-27.

ROBERTS, A. W. & VARBERG, D. E. (1973). *Convex Functions.* New York: Academic Press.

SERFLING, R. J. (1980). *Approximation Theorems in Mathematical Statistics.* New York: Wiley.

[*Received February* 1988. *Revised January* 1989]