EMPIRICAL SADDLEPOINT CONVERGENCE

Andrey Feuerverger

Presented by D.A.S. Fraser, F.R.S.C.

ABSTRACT. The uniform consistency, moment structure, and weak convergence to normality of the empirical moment generating function and empirical cumulant generating function and also of the arbitrary derivatives of these processes is established and used to investigate the properties of the saddlepoint approximation in the case that the required cumulant generating function is obtained empirically.

1. INTRODUCTION. If $X_1, X_2, \cdots, X_n$ are iid with density $f(x)$, moment generating function $M(t) = \int e^{tx} f(x)\,dx$ assumed finite in an interval I about the origin, and cumulant generating function $K(t) = \log M(t)$, then the saddlepoint approximation (see Daniels, 1954, 1980; Reid, 1988) for the density of $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is given by

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

where $t = t(x)$ is the unique real root of $K'(t) = x$. Our object is to study the consequence of replacing $K(t)$ in (1) by its sample version $K_n(t) = \log M_n(t)$ where $M_n(t) = \frac{1}{n}\sum_{i=1}^{n} e^{tX_i}$. The importance of this modification stems from the fact that the analytic form of $M(t)$ often is not tractable; a similar situation arises also when $f(x)$ itself is not available, but where a sample may be obtained. Another important application is given in Davison and Hinkley (1988) where saddlepoint approximations are applied in the context of the bootstrap and other resampling schemes.

2. MAIN RESULTS. In order to study the consequences of replacing $K(t)$ in (1) by $K_n(t)$ it is necessary to understand the sampling properties of the transforms $M_n(t)$ and $K_n(t)$. Letting D^ℓ denote differentiation applied ℓ times we have:

(A) $$\sup_{a \le t \le b} |D^\ell M_n(t) - D^\ell M(t)| \to 0 \text{ a.s.}, \quad \ell = 0, 1, 2, \cdots;$$

(B) $$\sup_{a \le t \le b} |D^\ell K_n(t) - D^\ell K(t)| \to 0 \text{ a.s.}, \quad \ell = 0, 1, 2, \cdots;$$

(C) $$E D^\ell M_n(t) = D^\ell M(t), \quad \ell = 0, 1, 2, \cdots;$$

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

for $s, t, s+t \in I$, and integers $\alpha \ge 0$, $\beta \ge 0$; and

(E) If $Y_n(t) = \sqrt{n}\,(M_n(t) - M(t))$ and $Z_n(t) = \sqrt{n}\,(K_n(t) - K(t))$ then $D^\ell Y_n(t)$ and $D^\ell Z_n(t)$ converge weakly, in the space of continuous functions on $[a, b]$ under the supremum norm, for $\ell = 0, 1, 2, \cdots$ to zero mean Gaussian processes having covariance structures respectively given by (D) and by

$$\text{asymp cov}\,(D^\alpha Z_n(t), D^\beta Z_n(t)) = D_s^\alpha D_t^\beta \left[\frac{M(s+t)}{M(s)M(t)} - 1 \right]$$

where $\alpha \ge 0$, $\beta \ge 0$ are integers and subscripts on D denote variables of partial differentiation.

In (A) and (B) we require $a, b \in I$, and for $\ell > 0$ we require a, b to be interior. In (E) we require $[a, b] \subseteq I/2$, and for $\ell > 0$ we require a,b to be interior. The proof of (A) and (B) involves the strong law of large numbers and convexity, while a proof for (E) may be based on Theorem 12.3 of Billingsley (1968) and Taylor expansion arguments. Related results in a different context are given in Ghosh (1987). The proofs of (A) - (E) will be given elsewhere.

Now let $\hat{f}_n(x)$ denote the *empirical saddlepoint approximation* given by expression (1) except with K_n replacing K and $\hat{t}$ replacing t throughout, where $\hat{t} = \hat{t}(x)$ is defined by $K_n'(\hat{t}) = x$; let $g_n(x)$ denote the saddlepoint approximation for the normalized variable $\sqrt{n}\cdot(\overline{X} - \mu)$ where $\mu = EX$; and let $\hat{g}_n(x)$ denote the correspondingly normalized empirical saddlepoint approximation, but centered now at $\overline{X}$ instead of μ. Then we have

$$\frac{\hat{g}_n(x)}{g_n(x)} = 1 + O_P(\frac{1}{\sqrt{n}}) \qquad (2)$$

where the error term is best possible in powers of n, and uniform over finite intervals.

To see this note that

$$g_n(x) = n^{-1/2} f_n(t(\mu + n^{-1/2}x)) \quad \text{and} \quad \hat{g}_n(x) = n^{-1/2} \hat{f}_n(\hat{t}(\overline{X} + n^{-1/2}x))$$

so that

$$\frac{\hat{g}_n(x)}{g_n(x)} = \left(\frac{K''(t)}{K_n''(\hat{t})}\right)^{1/2} \exp\left[n\{(K_n(\hat{t}) - K(t) - \overline{X}\hat{t} + \mu t) - (\hat{t} - t)\cdot\frac{x}{\sqrt{n}}\}\right] \qquad (3)$$

where now $K'(t) = \mu + n^{-1/2}x$ and $K_n'(\hat{t}) = \overline{X} + n^{-1/2}x$. The fractional term on the right in (3) is $1 + O_P(n^{-1/2})$, and since $K(t) = \mu t + \frac{1}{2}\sigma^2 t^2 + O(n^{-3/2})$ and $K_n(\hat{t}) = \overline{X}\hat{t} + \frac{1}{2}S^2\hat{t}^2 + O_P(n^{-3/2})$ where σ^2, S^2 are the variances of the actual and empirical distributions, the exponent in (3) equals

$$n\left[\frac{1}{2}(S^2\hat{t}^2 - \sigma^2 t^2) - (\hat{t} - t)\frac{x}{\sqrt{n}}\right] + O_P(\frac{1}{\sqrt{n}}). \qquad (4)$$

But since $\mu + n^{-1/2}x = K'(t) = \mu + \sigma^2 t + O(n^{-1})$ and $\overline{X} + n^{-1/2}x = K_n'(\hat{t}) = \overline{X} + S^2\hat{t} + O_P(n^{-1})$ then $t = O(n^{-1/2})$ and $\hat{t} = O_P(n^{-1/2})$ and therefore we find, in turn, $S^2\hat{t} - \sigma^2 t = O_P(n^{-1})$, $\hat{t} - t \equiv \sigma^{-2}[(S^2\hat{t} - \sigma^2 t) + (\sigma^2 - S^2)\hat{t}] = O_P(n^{-1})$ and

$S^2\hat{t}^2 - \sigma^2 t^2 = O_P(n^{-3/2})$. Consequently the exponent term (4) is $O_P(n^{-1/2})$. The normalization is required for (2) to hold, while studentizing does not improve the order of convergence. For the nonnormalized case we state the result

$$\frac{\hat{f}_{m,n}(x)}{f_n(x)} = 1 + O_P\left(\frac{n}{\sqrt{m}}\right) \tag{5}$$

where $\hat{f}_{m,n}(x)$ dnotes the saddlepoint approximation $f_n(x)$ of (1) but based now on the sample cumulant function $K_m(t)$ from a sample of size m. The error term is uniform over any interval of x-values corresponding to an interval of t-values interior to the domain on which M(t) is finite. The case $n = 1$ in (5) in conjunction with higher terms in the saddlepoint approximation leads to some interesting new possibilities for nonparametric density estimation. A fuller analysis (Feuerverger, 1988) will be given elsewhere.

Essentially similar analyses may be carried out for empirical versions of the tail area saddlepoint approximation (Lugannani and Rice, 1980; Daniels, 1987), for Edgeworth expansions (Feller, 1971, Theorem 2, page 535; Barndorff-Nielsen and Cox, 1979), and also for quantities such as the Chernoff index $\inf_z e^{-z(t+\mu)}M(z)$ for large deviation probabilities (Serfling, 1980, chapter 10).

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 N. Reid for valuable discussions on the saddlepoint, and to R. Tibshirani for valuable discussions on bootstrap methods.

REFERENCES

Barndorff-Nielsen, O. and Cox, D.R. (1979). Edgeworth and saddle-point approximations with statistical applications. J. Royal Statist. Soc., B41, 279-312.

Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

Csorgo, S. (1980). The empirical moment generating function. In Colloquia Mathematica Societatis Janos Bolyai, 32. Nonparametric Statistical Inference. Budapest. (I. Vincze, ed.) North-Holland.

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

Daniels, H.E. (1980). Exact Saddlepoint approximations. Biometrika, 67, 59-63.

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

Davison, A.C. and Hinkley, D.V. (1988). Saddlepoint approximations in resampling methods. Biometrika, to appear.

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

Feuerverger, A. (1988). On the empirical saddlepoint. Manuscript. 41pp.

Ghosh, S. (1987). Some Tests of Normality Using Methods Based on Transforms. Doctoral dissertation, University of Toronto.

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

Reid, N. (1988). Saddlepoint methods and statistical inference. Statistical Science, to appear.

Serfling, R.J. (1980). Approximation Theorems in Mathematical Statistics. Wiley, New York.

Andrey Feuerverger
Department of Statistics
University of Toronto
Toronto, Ontario
Canada M5S 1A1

Received February 28, 1988