that case the function f_n represents the joint density, g_n the marginal density. The point x_n represents the center of the marginal distribution g_n and $y_n = y_n(x)$ is the conditional mode of the distribution of y given x. Of course this approximate marginalization does not require f_n to be a posterior distribution. Phillips (1983) uses this approach to approximate the marginal sampling distribution of various econometric estimators.

The asymptotic properties of the saddlepoint approximation to sampling distributions and the Laplace approximation to marginal densities are very similar. Both yield approximations that have errors uniformly of order $O(n^{-1})$ on fixed neighborhoods of x_n. Both benefit from numerical renormalization in the sense that the absolute errors of the approximations are of order $O(n^{-3/2})$ in $n^{-1/2}$-neighborhoods of x_n. Another interpretation of this result is that the shapes of the densities $g_n(\cdot)$ are approximated to order $O(n^{-3/2})$ by both methods.

The approximation of posterior expectations by Laplace's method is somewhat different. A single number is to be approximated rather than a function. Direct application of Laplace's method yields the maximum likelihood estimate or the posterior mode as an approximation to the posterior mean. The error of this approximation is of order $O(n^{-1})$. More accurate approximations with an error of order $O(n^{-2})$ can be obtained by using higher order terms, as described by Lindley (1980), or by using different centers for the expansions of numerator and denominator integrals, as described in Tierney and Kadane (1986) and Tierney, Kass and Kadane (1987).

The approximate predictive densities discussed in Leonard (1982), Tierney and Kadane (1986) and Davison (1986) fall somewhere between marginal density and moment approximations. Because a predictive density is a density, its approximation would appear to be more closely related to the approximation of marginal and sampling densities. On the other hand, the predictive density at a particular point can be expressed as a posterior expectation. The result of applying second order expectation approximation, as in (4.1) of Tierney and Kadane (1986), is an approximation to the predictive density with an error of order $O(n^{-2})$. The order of this error term will generally not be improved by numerical renormalization. As a result I feel that these approximate predictive densities are more closely related to approximate expectations than to approximate marginal densities.

I hope that these comments have added to the discussion in Section 6 of Professor Reid's excellent paper.

ADDITIONAL REFERENCES

LEONARD, T. (1982). Comment on "A simple predictive density function" by M. Lejeune and G. D. Faulkenberry. *J. Amer. Statist. Assoc.* **77** 657–658.

LINDLEY, D. V. (1980). Approximate Bayesian methods. In *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 223–237. University Press, Valencia.

PHILLIPS, P. C. B. (1983). Marginal densities of instrumental variables estimators in the general single equation case. Cowles Foundation Paper No. 582.

Comment

Robert E. Kass

The world of asymptotics is beautiful and mysterious. Witness Stirling's approximation, and recall the first time you needed to use it. What explains the odd yet simple formula, you may have asked, and more, How is it that with one correction term it already achieves 99.95% accuracy in approximating factorials as small as 2? Marvel at Figure 1. But recognize, each time we consider a sample of size n to be part of an infinite sequence of observations, we are faced with an irony: limits do not depend on the first n values, yet they are able to inform us about the behavior of the sample. Our finite world seems tied to asymptopia, but how?

Second-order asymptotic results continue to produce this feeling of awe and amazement in those who aren't yet familiar with them. Nancy Reid's review not only tells the saddlepoint story, it also nicely demonstrates the similarity of method in applications to maximum likelihood and conditional inference, robust estimation and Bayesian analysis. My comment consists of (i) a brief description of the relationship between Laplace's method and the saddlepoint

Robert E. Kass is Associate Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.

approximation, (ii) a concern about accuracy of approximations in data analysis and (iii) a remark that current methodology is capable of treating special cases, but does not yet provide satisfactory general solutions to the basic problem of making inferences about one parameter in the presence of others.

In his 1954 paper, Daniels provided two derivations of the saddlepoint approximation, and Reid describes in some detail the second, based on an Edgeworth expansion of the "conjugate" density. I rather like the first derivation. It begins with an inversion of the characteristic function which, of course, is the way we get the Edgeworth series in the first place; the method of steepest descents is then applied to the integral given by Reid in the paragraph preceding equation (4). I would like to elaborate just a little on what Reid says in that paragraph, in order to explain the connection with Laplace's method.

Suppose f is the analytic function in the exponent, here $f(z) = K(z) - z\bar{x}$, let z_0 be a saddlepoint, and let P be the path through z_0 along which the imaginary part of f is constant. Consider the integral of $\exp[nf(z)]$ along P. Because the imaginary part of f is constant on P, the exponential of this constant may be brought outside the integral, leaving a real integrand. (In the present application of steepest descents, the saddlepoint z_0 is real and $f(z_0)$ is real so the constant being taken outside the integral is 1.) It is easily shown that this real integrand has a maximum along P at z_0, and that P is a path of steepest descent. The actual path of the integral is distorted to coincide with P near z_0 so that the value of the integral will come mainly from contributions occurring near z_0. (By the way, z_0 is chosen to be a saddlepoint rather than a maximum because analytic functions have no maxima in open regions of analyticity—all they have is maxima along paths.) Once it has been assured that the integral is dominated by contributions occurring near z_0, a quadratic expansion of f is substituted for f itself. This leaves an integrand that looks like a normal density (although with a complex argument) and it is easily integrated, yielding (1).

The replacement of an integrand by a normal-looking approximation, meanwhile, is the basis of Laplace's method. Let H be a real function on Φ in R^k, and let $\hat{\phi}$ be its maximum. Expanding H as a quadratic about $\hat{\phi}$, we get

$$\int \exp(nH(\phi))\, d\phi \doteq (2\pi n)^{-k/2} \mid H''(\hat{\phi}) \mid^{1/2} \exp[nH(\hat{\phi})]$$

which has a form comparable with that in equation (1). From these brief descriptions of Laplace's method and steepest descents, loosely speaking, we may consider the saddlepoint approximation to be an application of Laplace's method in the complex plane.

Terminology varies. Some authors refer to Laplace's method as being applied in the complex domain; some distinguish steepest descents from the saddlepoint approximation, emphasizing the latter's local nature and its use with paths other than steepest descent, allowing the quadratic approximation to f rather than f itself to have constant imaginary part along the chosen local path. Copson (1967) attributes the saddlepoint approximation to Riemann who, he says, used "essentially Laplace's method" (roughly as outlined above). See Stigler (1986) for commentary on Laplace's treatment.

In any case, the Laplace and saddlepoint approximations are closely related. (Each may be used to derive Stirling's approximation, which Copson attributes to Laplace rather than Stirling, the formula derived by Stirling apparently being different than the one most commonly used.) As a statistical consequence, the Laplace approximation to a marginal density has properties analogous to those of the saddlepoint approximation. Thus, the good accuracy apparent in Figure 1 of Tierney and Kadane (1986) is like that shown in Figure 1 here.

What are we to conclude from the approximations illustrated in each Figure 1? As far as explanations of their accuracy are concerned, I am inclined to take the uniform order $O(1/n)$ error at face value. The surprise elicited by the results, I think, is due in part to the choice of examples. When we pick cases in which the normal approximation is poor but not horrible, uniform second-order approximations do a good job of mopping up the remainder. Examination of the stochastic behavior of the remainder in common examples would probably be informative. On the other hand, even if we accept the asymptotics as being more than heuristic devices for deriving potentially useful approximations, in practical work one still needs to know whether good results are likely to be obtained in a given example. Thus, in each new problem, a data analyst must face the question of accuracy in finite samples.

It is possible to get bounds on the error from the proofs of the saddlepoint and Laplace approximations. In the case of Stirling's formula, this is helpful. Usually, however, the bounds are not only likely to be inaccurate, they are also difficult to use in nontrivial statistical examples because they depend on maximum magnitudes of higher derivatives. Another possibility is to develop diagnostics based on higher order asymptotics, although these are also likely to be complicated and would still rely on limiting behavior.

The other available approach is to use nonasymptotic numerical methods, such as Gaussian quadrature and Monte Carlo. In the saddlepoint case, one would think that the Fast Fourier Transform might be useful

as, perhaps, might be a complex version of Gaussian quadrature along the path of steepest descent. These methods would require sufficiently accurate computation of the cumulant generating function that its second derivative matrix could be obtained. This might present further difficulties. As we consider checking the accuracy of Laplace's method in Bayesian analysis or the saddlepoint method in conditional or robust inference, an interesting statistical point emerges. Although most nontrivial problems, in either approach, will present nontrivial integrals to be handled numerically, the Bayesian integrals are far easier to evaluate. This is because they involve only the parameter space, and not the sample space. In practice, one can envision a data analytical process in which experience plays a role, with one or a small number of integrals being checked by nonasymptotic numerical methods as the need arises; when satisfied with accuracy, the data analyst could rely on asymptotic approximations for much of the analysis (Kass, Tierney and Kadane, 1988). To proceed along similar non-Bayesian lines is more difficult, especially when conditional inference procedures are employed.

I come now to my opinion about the implications for statistical practice. Much progress has been made in the use of saddlepoint and Laplace approximations, and I am at least in partial agreement with Reid's conclusion that the impact of saddlepoint approximations on statistical practice will soon be felt. Most practical problems of parametric inference, it seems to me, fall in the category of "inference in the presence of nuisance parameters," with the acknowledgment that various functions $g(\theta)$, which might be components of a parameter vector, are of simultaneous interest. That is, in practice, we may wish to make inferences about each of many such functions $g(\theta)$ in turn, keeping in mind that joint inferences are being made, yet requiring separate statements for ease of contemplation. Currently, the asymptotic normality of MLE's and the asymptotic χ^2 distribution of the likelihood ratio test are the workhorses of statistical practice using parametric families. This is not only because these first-order results are often sufficiently accurate to be of use, but, in addition, they offer simple and easily comprehended answers to the questions being asked. Which alternative solutions to this basic problem can and should be used to analyze data?

As we all know, statistical practice is largely determined by widely available statistical software. Thus, to have an impact on practice in the near future, new procedures must be in a form that can be easily adapted and adopted by major software distributors. It would not be hard to get alternative p-values for modified likelihood ratio statistics included in computer output and, if we were lucky, modified confidence intervals and more accurate assessments of coverage and posterior probabilities of the usual Normal-based intervals might be included, too. But can current asymptotic theory supply these?

From the non-Bayesian point of view, it appears that second-order theory is now able to provide improved solutions to particular problems based on multivariate normal and generalized linear models. Once we move away from these special cases, however, dealing with nuisance parameters becomes more difficult, and the worries described at the end of Section 6.2 seem quite serious. Furthermore, as indicated above, nonasymptotic numerical evaluation of the characteristics of the procedures is also very demanding. Meanwhile, for solving these general problems, implementation of both asymptotic and nonasymptotic methods is more advanced for Bayesian technology than for its non-Bayesian counterpart. The practical difficulty here, however, is that an additional multivariate distribution (the prior) is required. There will be progress with the Bayesian approach as flexible families of multivariate distributions are further developed, used and understood; the recent advances in implementation should add impetus to this effort.

Thus, I believe there will, fairly soon, be a trickle of data analyses using asymptotic methods based on saddlepoint and Laplace approximations. Widespread use in general applications, however, must wait for further methodological advance on the old and difficult problem of handling nuisance parameters. For the present, Nancy Reid has done a great service in providing a concise, informative review of this very active area of statistical research.

ACKNOWLEDGMENT

This research was supported by National Science Foundation Grant DMS-87-05646. This comment was prepared while the author was a visiting scholar in the Department of Statistics, Harvard University.

ADDITIONAL REFERENCES

COPSON, E. T. (1967). *Asymptotic Expansions.* Cambridge Univ. Press, Cambridge.

STIGLER, S. M. (1986). Laplace's 1774 memoir on inverse probability. *Statist. Sci.* **1** 359–378.