Saddlepoint Approximations for Generalized Linear Models: A Gentle Introduction

By G.U.H. SEEBER[1]

Leopold-Franzens-Universität Innsbruck, Austria

SUMMARY

Saddle point approximations to the density of a sum of i.i.d. random variables are introduced via exponential tilting and an Edgeworth expansion in the conjugate family of distributions. Applications to exponential families and generalized linear models are reviewed.

Keywords: Higher order asymptotics; Edgeworth expansion; Saddlepoint approximation; Conditional inference; Exponential tilting; Exponential families

1 Introduction

It is almost fourty years that saddlepoint approximations to densities of sums of random variables were introduced by Daniels (1954), but it's only recently that this method for constructing large sample approximations finds greater attention among statisticians. Techniques relying on saddlepoint expansions are reported to be of high accuracy, even when sample sizes are (very) small, making them potentially useful to applied statisticians. Since the discussion paper by Barndorff-Nielsen and Cox (1979) much literature has appeared mainly contributing to higher order asymptotic theory. Davison (1988) developed approximations to the conditional densities and distributions of sufficient statistics in generalized linear models with canonical link functions, thus emphasizing the relevance of the methods for applied work. The saddlepoint approximation to the density of the maximum likelihood estimator in curved exponential families admitting non-canonical link functions is given in Hougaard (1985). Very recently Pierce and Peters (1992) reviewed saddle point methods for exponential families. They emphasized situations envolving discrete data, particularly contingency tables, and presented some real data examples. However, it seems that, at this time, research on mathematical issues as well as practical experience is still needed to fully understand and appreciate the method and its usefulness for the purposes of data analysis.

In this paper I attempt to give a non-rigorous introduction to the subject. Readers interested in formal details or an extensive treatment are referred to the books by Barndorff-Nielsen and Cox (1989) or Field and Ronchetti (1990). Reid (1988, 1991) provides very concise overviews on the derivation of saddlepoint techniques and their application to statistical inference.

2 Exponential Tilting and Saddlepoint Approximations

Let $Y_1, \ldots, Y_n$ be i.i.d. random variables with mean μ and variance σ^2, and $S_n = Y_1 + \cdots + Y_n$ their sum, suitably standardized as $Z_n = (S_n - n\mu)/\sqrt{n\sigma^2}$. We are interested in approximating the density function of S_n or Z_n, respectively. Derivation of both Edgeworth and saddlepoint expansion relies on the cumulant generating function K of the Y_i-s, i.e. the logarithm of their moment generating function $M(t) = \mathrm{E}[\exp(tY_i)]$. r-th derivatives at 0 of the cumulant generating function $K(t) = \ln(M(t))$ are called cumulants and will be denoted by κ_r. Cumulants are related to moments, in particular $\kappa_1 = \mu$ and $\kappa_2 = \sigma^2$.

[1] *Address for correspondence*: Institut für Statistik, Universität Innsbruck, Innrain 52, A–6020 Innsbruck, Austria. e-mail: Gilg.Seeber@uibk.ac.at

Formally, Edgeworth series are obtained by expanding the cumulant generating function K in powers of $1/\sqrt{n}$ and inverting the corresponding moment generating function. This results in an approximate density

$$f_{Z_n}(z) \approx \varphi(z) \cdot \left[1 + \frac{1}{\sqrt{n}} \frac{\varrho_3}{6} H_3(z) + \frac{1}{n} \left(\frac{\varrho_4}{24} H_4(z) + \frac{\varrho_3^2}{72} H_6(z) \right) \right], \tag{1}$$

where φ denotes the standard normal density, $\varrho_r = \kappa_r / \sqrt{\kappa_2}$ are the standardized cumulants. $H_3(z) = z^3 - 3z$, $H_4(z) = z^4 - 6z^2 + 3$, and $H_6(z) = z^6 - 15z^4 + 45z^2 - 15$ are Hermite polynomials of respective degrees. The $1/\sqrt{n}$ term adjusts for skewness, while the $1/n$ term is an adjustment for both skewness and kurtosis.

As a simple example consider i.i.d. random variables Y_1, Y_2, Y_3 distributed as exponential with parameter $\lambda = \sqrt{3}$. The expected value and the variance of their sum equals $\sqrt{3}$ and 1, respectively. The normal approximation to the density of S_3 deviates considerably from the exact gamma density. The Edgeworth expansion

$$f_{S_3}(s) \approx \varphi(s - \sqrt{3}) \cdot \left[\frac{107}{108} + \frac{1}{9}(s - \sqrt{3})^2 + \frac{1}{9}(s - \sqrt{3})^3 - \frac{7}{108}(s - \sqrt{3})^4 + \frac{1}{162}(s - \sqrt{3})^6 \right],$$

certainly means an improvement, but still seems to be less than satisfactory.

In general, Edgeworth expansions are reported to be reliable in the center of the distribution for moderate sample sizes. Due to the polynomial factor in (1) the accuracy can be worse in tail areas, where the approximate density may even become negative. This limits the usefulness of the method, particularly if one is interested in calculating tail probabilities. One approach to overcome some of the difficulties in finding an approximation to $f_{S_n}(s)$ is to associate, for each s, a distribution, the density of which is accurately approximated at its mean value. Let $f(y)$ be the density of the Y_i-s, then we associate with f an exponential family defined by

$$f(y|\xi) \quad = \quad \exp(\xi y - K(\xi)) \cdot f(y), \tag{2}$$

of which $f(y) = f(y|0)$ is an element. In the theory of large deviations this embedding in a conjugate family of distributions is known as exponential tilting. Applying this to the density of S_n gives

$$f_{S_n}(s|\xi) \quad = \quad \exp(\xi s - nK(\xi)) \cdot f_{S_n}(s), \tag{3}$$

or

$$f_{S_n}(s) = f_{S_n}(s|0) = \exp(-\xi s + nK(\xi)) \cdot f_{S_n}(s|\xi). \tag{4}$$

Now, an approximation to $f_{S_n}(s)$ may be obtained from an approximation to $f_{S_n}(s|\xi)$ at any value of ξ. A suitable choice $\hat{\xi}$ is to put s in the 'centre' of the resulting distribution by defining $\hat{\xi} = \hat{\xi}(s)$ to be its expected value, i.e. by solving

$$K'(\hat{\xi}) = \frac{s}{n}. \tag{5}$$

$\hat{\xi}$ is called the saddlepoint. Via an Edgeworth expansion and after some algebra we obtain

$$f_{S_n}(s) \approx \frac{1}{\sqrt{2\pi K''(\hat{\xi})}} \cdot \exp(-\hat{\xi}s + nK(\hat{\xi})), \tag{6}$$

which is called the saddlepoint approximation. The function on the right hand side of (6) does not necessarily integrate to 1, so that it may need renormalization by multiplication with a constant factor.

If $Y_1, \ldots, Y_n$ are exponentially distributed with parameter λ, then $K(t) = \ln(\lambda/(\lambda - t))$ and $\hat{\xi} = \lambda - n/s$. Applying (6) results in

$$f_{S_n}(s) = \frac{1}{\sqrt{2\pi}} n^{1-n} \exp(n) \cdot y^{n-1} \lambda^n \exp(-y\lambda),$$

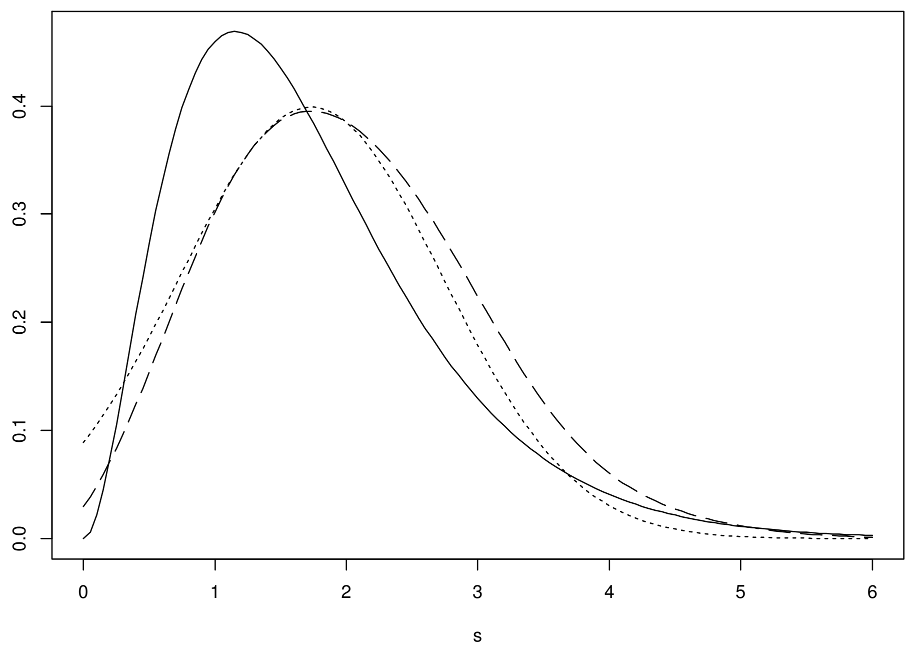

Figure 1: Exact density and saddlepoint approximation (solid line), normal (dotted line) and Edgeworth approximation (dashed line) to an exponential distribution with parameter $\lambda = \sqrt{3}$.

which, after renormalization, is the exact gamma distribution of S_n. The saddlepoint approximation is known to be exact for the normal, gamma, and inverse Gaussian distributions (being the only univariate distributions, for which this is true). Figure 1 exhibits approximate and exact densities for the example given above.

There is a different derivation of the saddlepoint approximation using techniques from complex analysis: By the Fourier inversion theorem (see, e.g., Grimmett and Stirzaker (1982, p. 106)) we have

$$f_{S_n}(s) = \frac{1}{2\pi} \cdot \int_{-\infty}^{\infty} \exp(-its) \cdot M(it)^n \, dt, \tag{7}$$

where $i = \sqrt{-1}$. The idea is to choose the path of integration through a point in the complex plane that, in its neighbourhood, allows accurate approximation of the integral in (7). This point is seen to be the solution of (5), which is (under some regularity conditions, of course) unique, real and a saddlepoint of the complex function $K(v + iw) - (v + iw)s$. This provides both new insights and justifies application of the method for sums of discrete random variables, as long as their probability functions can be regarded continuous functions of a real variable.

If, for instance, Y is distributed as binomial with parameters n and π, then $f_B(y|n, \pi) = \Gamma(n+1)/(\Gamma(y+1)\Gamma(n - y + 1)) \cdot \pi^y (1 - \pi)^{(n-y)}$ is a continuous function on $(0, n)$. The cumulant generating function is $K(t) = \ln(1 - \pi + \pi \exp(t))$, hence giving a saddlepoint

$$\hat{\xi} = \ln \left[\frac{y}{n - y} \Big/ \frac{\pi}{1 - \pi} \right].$$

Note that there does not exist a saddlepoint for $y = 0, n$. Using (6) we obtain

$$f_B(y|n, \pi) \approx \frac{1}{\sqrt{2\pi}} \frac{n^{n+1}}{y^{y+1/2}(n - y)^{n-y+1/2}} \cdot \pi^y (1 - \pi)^{n-y},$$

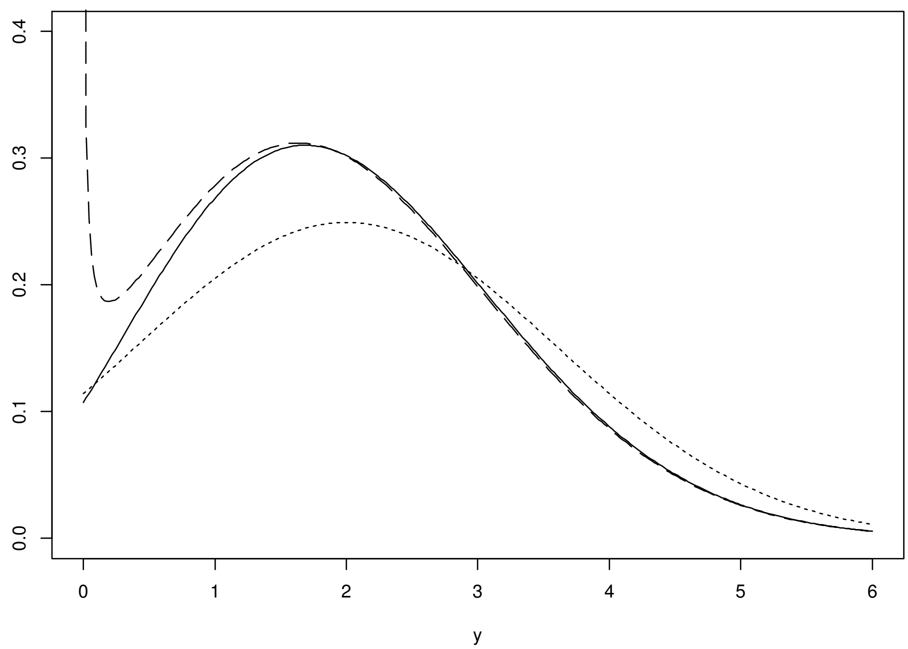

Figure 2: Exact probability function (solid line), saddlepoint (dashed line) and normal approximation (dotted line) to a binomial distribution with parameters $n = 10$ and $\pi = 0.2$.

which differs from an approximation obtained by Stirling's formula by a factor of $n/\sqrt{y(n-y)}$.

To obtain tail probabilities the saddlepoint approximation can be integrated, though this is not an easy task. Daniels (1987) provides

$$\begin{aligned} F_{S_n}(s) &\approx \Phi(r) + \varphi(r) \cdot \left(\frac{1}{r} - \frac{1}{v}\right), \\ r &= \operatorname{sign}(\hat{\xi})\sqrt{2n(-K(\hat{\xi}) + \hat{\xi}y/n)}, \\ v &= \hat{\xi}\sqrt{nK''(\hat{\xi})}, \end{aligned} \tag{8}$$

which is known as the Lugannani and Rice (1980) formula. Φ denotes the standard normal distribution function. (8) is regarded as a highly accurate approximation, except in a neighbourhood of $r = 0$, see Reid (1991) and the references therein.

3 Statistical Applications

In the context of statistical modelling, particularly if a likelihood approach is adopted, the distribution of the maximum likelihood estimator is of special importance. Consider the case of n not necessarily scalar, i.i.d. random variables Y_i having an exponential family density

$$f_Y(y|\vartheta) = \exp(\vartheta' t(y) - b(\vartheta) - c(y)),$$

where $t(Y) = (t_1(Y), \ldots, t_k(Y))'$ is the minimal sufficient statistic and ϑ the canonical parameter. Let $S = t(Y_1) + \cdots + t(Y_n)$, then $f_T(s|\vartheta) = \exp(\vartheta' s - nb(\vartheta) - h(s))$, say, i.e. S is distributed according to an exponential family law. Typically, the function $h(s)$ is difficult to obtain. However, applying the saddlepoint approximation (6) to $\exp(-h(s))$ leads to an expression, which does not involve $h(s)$,

$$f_S(s) \approx (2\pi)^{-k/2} \sqrt{|nb''(\hat{\vartheta})|}\, \exp\left[(\vartheta - \hat{\vartheta})' t - n(b(\vartheta) - b(\hat{\vartheta}))\right], \tag{9}$$

which allows derivation of a formula for the approximate density of $\hat{\vartheta} = \hat{\vartheta}(s)$,

$$f_\Theta(\hat{\vartheta}|\vartheta) \approx c\sqrt{|j(\hat{\vartheta})|}\, \frac{L_t(\vartheta)}{L_t(\hat{\vartheta})}, \tag{10}$$

where L_t is the likelihood function of $Y_1, \ldots, Y_n$ in terms of t and j denotes the Jacobian $\partial^2/(\partial\vartheta'\partial\vartheta)\, b(\vartheta)|_{\vartheta=\hat{\vartheta}}$ of the transformation that maps t to $\hat{\vartheta}$, i.e. the information matrix evaluated at the maximum likelihood estimate. c is a suitably chosen renormalizing constant. (10) is sometimes called the Barndorff-Nielsen approximation, which also provides a valid approximation for models outside the exponential family.

Introducing a not necessarily linear 'regression' function $\eta(\beta)$ that maps regression parameters β to expected values of the Y_i-s, and, a fortiori, to canonical parameters ϑ, thus restricting ϑ to be element of a smooth subset of the canonical parameter space, defines a curved exponential family model. This is not an exponential family model, if, e.g., it is specified by a non-canonical link function in a GLM context. Hougaard (1985) derives a formula for the saddlepoint approximation for the resulting density function, which I do not reproduce here. If the regression function is defined via a canonical link, formula (9) may be applied.

For a one parameter exponential family model the distribution function of the maximum likelihood estimator for the canonical parameter is easily calculated using the Lugannani and Rice formula, which in this case has exactly the same structure as (8), but

$$\begin{aligned} r &= \operatorname{sign}(\hat{\vartheta} - \vartheta)\sqrt{2 \cdot [\ln(L(\hat{\vartheta})/L(\vartheta))]}, \\ v &= (\hat{\vartheta} - \vartheta) \cdot \sqrt{j(\hat{\vartheta})} \end{aligned}$$

are the signed square root of the likelihood ratio statistic and the standardized maximum likelihood estimate, respectively. For non-canonical parameters or other models see Fraser (1990) and the references cited there.

In the multiparameter case, if there is a parameter ϑ of interest and a nuisance parameter ν and the data can be split into components S_1 and S_2, where S_2 is sufficient for ν, then inference on ϑ can be based on the conditional likelihood $L_{s_1|s_2}(\vartheta)$. A double saddlepoint approximation to its logarithm, obtained by computing the ratio of saddlepoint approximations to the joint density of (S_1, S_2) and the marginal density of S_2, is given by

$$\ln(L_{s_1|s_2}(\vartheta)) \approx \ln[L_{(s_1,s_2)}(\vartheta, \hat{\nu}(\vartheta))] + \frac{1}{2}\ln\left|-\frac{\partial^2}{\partial\nu'\partial\nu}\ln[L_{(s_1,s_2)}(\vartheta,\nu)|_{\vartheta,\hat{\nu}(\vartheta)}]\right|, \tag{11}$$

where $\hat{\nu}(\vartheta)$ maximizes the likelihood function for given ϑ. $\ln(L_{(s_1,s_2)}(\vartheta, \hat{\nu}(\vartheta))$ is the profile (log-) likelihood for the parameter of interest, (11) therefore suggest one of several ways to modify the profile likelihood as to achieve more desireable distributional properties. Davison (1988) shows that approximate conditional inference for regression and dispersion parameters in generalized linear models involving the normal, gamma and inverse Gaussian distributions based on (11) may be performed using standard statistical packages such as GLIM.

Fraser (1990) suggest that (11) may be used in conjunction with the Lugannani and Rice formula for the distribution of the maximum likelihood estimate, but see also Pierce and Peters (1992).

Acknowledgements

This work was done while visiting the Department of Statistics, University of Florida, partially funded by a grant provided by the Austrian *Fond zur Förderung der Wissenschaftlichen Forschung.* I would like to thank my colleagues at UF for introducing me to the subject.

References

Barndorff-Nielsen, O.E. and Cox, D.R. (1979) Edgeworth and saddlepoint approximations with statistical applications (with discussion). *Journal of the Royal Statistical Society* B **41**, 279–312.

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

Daniels, H.E. (1954) Saddlepoint approximations in statistics. *Annals of Mathematical Statistics* **25**, 631–650.

Daniels, H.E. (1987) Tail probability approximations. *International Statistical Review* **55**, 37–48.

Davison, A.C. (1988) Approximate conditional inference in generalized linear models. *Journal of the Royal Statistical Society* B **50**, 445–461.

Field, C. and Ronchetti, E. (1990) *Small Sample Asymptotics.* Hayward, CA: Institute of Mathematical Statistics.

Fraser, D.A.S. (1990) Statistical inference: likelihood to significance. *Journal of the American Statistical Association* **86**, 258–265.

Fraser, D.A.S., Reid, N. and Wong, A. (1991) Exponential linear models: a two-pass procedure for saddlepoint approximation. *Journal of the Royal Statistical Society* B **53**, 483–492.

Grimmett, G.R. and Stirzaker, D.R. (1982) *Probability and Random Processes.* Oxford University Press.

Hougaard, P. (1985) Saddlepoint approximations for curved exponential families. *Statistics & Probability Letters* **3**, 161–166.

Lugannani, R., and Rice, S. (1980) Saddle point approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability* **12**, 475–490.

Pierce, D.A. and Peters, D. (1992) Practical use of higher-order asymptotics for multiparameter exponential families (with discussion). *Journal of the Royal Statistical Society* B **54**, to appear.

Reid, N. (1988) Saddlepoint methods and statistical inference (with discussion). *Statistical Science* **3**, 213–238.

Reid, N. (1991) Approximations and asymptotics. In *Statistical Theory and Modelling* (eds. D.V. Hinkley, N. Reid and E.J. Snell), pp. 287–305. London: Chapman and Hall.