Ann. Inst. Statist. Math.
Vol. 54, No. 4, 743–747 (2002)


# SADDLEPOINT APPROXIMATION FOR THE DISTRIBUTION FUNCTION NEAR THE MEAN

BO YANG[1] AND JOHN E. KOLASSA[2*]

[1]*Department of Biostatistics, Shering–Plough Research Institute, 2000 Galloping Hill Road, Kenilworth, NJ 07033-0530, U.S.A.*
[2]*Department of Biostatistics, University of Rochester, Box 630, Rochester, NY 14642, U.S.A.*

(Received July 3, 2001; revised March 19, 2002)

**Abstract.** We repair numerical difficulties in applying saddlepoint tail probability approximations when the ordinate at which the approximation is evaluated is near the mean of the distribution approximated. These modifications apply to double saddlepoint approximations to conditional distributions as well.

*Key words and phrases*: Saddlepoint approximation, singularity.

## 1. Introduction

Huzurbazar and Huzurbazar (1999) describe the use of saddlepoint tail probability approximations for the convolution of distributions, but note that some of the approximations that behave very well in the tails of the distribution perform poorly for ordinates near the mean of the distribution approximated. We repair numerical difficulties in applying double saddlepoint tail probability approximations when the ordinate at which the approximation is evaluated is near the mean of the conditional distribution. This correction is necessary in order to calculate significance tests. The underlying test will then be inverted to form confidence regions.

## 2. Saddlepoint approximations

Saddlepoint approximations (McCullagh (1987), Chapter 5; Reid (1988)) are asymptotic approximations to densities and distribution functions of means $\boldsymbol{T}$ of independent and identically distributed random vectors $\boldsymbol{Y}_j$. These approximations are derived from the cumulant generating function of each of the addends,

$$\mathcal{K}(\boldsymbol{\beta}) = \log(E[\exp(\boldsymbol{\beta}^T \boldsymbol{Y}_j)]). \tag{2.1}$$

When the cumulant generating function $\mathcal{K}$ is defined for vectors of complex numbers whose real parts are sufficiently close to zero, saddlepoint approximations for the density and cumulative distribution function may be defined by approximately inverting the relationship (2.1). These approximations are expressed in terms of derivatives of the cumulant generating function evaluated at the real vector $\hat{\boldsymbol{\beta}}$ determined by the saddlepoint equation

$$\mathcal{K}'(\hat{\boldsymbol{\beta}}) = \boldsymbol{t}, \tag{2.2}$$

*Now at Department of Statistics, Rutgers University, Piscataway, NJ 08854-8019, U.S.A.

where $\mathcal{K}'$ is the vector of first derivatives of $\mathcal{K}$, and $\boldsymbol{t}$ is the point where the density or cumulative distribution function is to be evaluated. The vector $\hat{\boldsymbol{\beta}}$ is called the multivariate saddlepoint.

Skovgaard (1987) applies saddlepoint techniques to the problem of approximating tail probabilities for conditional distributions. Approximating $F(t^1 \mid \boldsymbol{t}_{-1}) = P[T^1 \geq t^1 \mid \boldsymbol{T}_{-1} = \boldsymbol{t}_{-1}]$, for $\boldsymbol{T}_{-1} = (T^2, \ldots, T^d)$ and $\boldsymbol{t}_{-1} = (t^2, \ldots, t^d)$, involves the multivariate saddlepoint solving (2.2), both for the full distribution of $\boldsymbol{T}$ and for the distribution of the shorter random vector $\boldsymbol{T}_{-1}$. Let $\hat{\boldsymbol{\beta}}$ solve (2.2), and let $\hat{\boldsymbol{\gamma}}$ be the vector of length $d$ such that

$$\mathcal{K}^j(\hat{\boldsymbol{\gamma}}) = t^j \quad \text{for} \quad j \neq 1, \ \hat{\gamma}_1 = 0, \tag{2.3}$$

where $\mathcal{K}^j$ denotes the derivative of $\mathcal{K}$ with respect to component $j$ of its argument. The conditional tail probability approximation of interest is

$$1 - \tilde{F}(t^1 \mid \boldsymbol{t}_{-1}) = 1 - \Phi(\sqrt{n}\hat{w}) + \phi(\sqrt{n}\hat{w}) \left[\frac{1}{\sqrt{n}\hat{z}} - \frac{1}{\sqrt{n}\hat{w}}\right], \tag{2.4}$$

where $\hat{z} = \hat{\beta}_1\sqrt{\det[\mathcal{K}''(\hat{\boldsymbol{\beta}})]/\det[\mathcal{K}''_{-1}(\hat{\boldsymbol{\gamma}})]}$, $\hat{w} = \operatorname{sgn}(\hat{\beta}_1)\sqrt{2[\hat{\boldsymbol{\beta}}^T\boldsymbol{t} - \mathcal{K}(\hat{\boldsymbol{\beta}})] - 2[\hat{\boldsymbol{\gamma}}^T\boldsymbol{t} - \mathcal{K}(\hat{\boldsymbol{\gamma}})]}$, $\mathcal{K}''_{-1}$ is the $(d-1) \times (d-1)$ submatrix of the matrix of second derivatives of $\mathcal{K}$, corresponding to all components of $\boldsymbol{\beta}$ and $\boldsymbol{T}$ except the first, and $\Phi$ and $\phi$ are the normal distribution function and density respectively.

Also of interest are inversion techniques for lattice distributions. Skovgaard (1987) derives a counterpart of (2.4) in the lattice case, in which $\hat{\beta}_1$ is replaced by $2\sinh(\frac{1}{2}\hat{\beta}_1)$, and in which $t^1$ is corrected for continuity when calculating $\hat{\boldsymbol{\beta}}$. That is, if possible values for $T^1$ are one unit apart, then $\hat{\boldsymbol{\beta}}$ solves $\mathcal{K}'(\hat{\boldsymbol{\beta}}) = \tilde{\boldsymbol{t}}$ where $\tilde{t}^j = t^j$ if $j \neq 1$ and $\tilde{t}^1 = t^1 - \frac{1}{2}$.

## 3. Linear approximation near the mean

The difficulty in applying (2.4) for certain values of $t^1$ near the conditional mean is that $w(\boldsymbol{t})$ and $z(\boldsymbol{t})$ both take on the value 0, and hence $1/w - 1/z$ can not be evaluated. With all components of $\boldsymbol{t}_{-1}$ held fixed, $w(\boldsymbol{t})$ is a smooth invertible function of $t^1$. Hence $1/z - 1/w$ is a function of $w$, and since $z = 0$ if and only if $w = 0$, and $dz/dw \neq 0$, then $1/z - 1/w$ may be approximated as a linear function in $w$, with error of size $O(w)$. In order to determine this approximation, we require derivatives of the saddlepoint $\hat{\boldsymbol{\beta}}$ with respect to $t^1$. Let $\hat{\beta}^j_k = (d/dt^j)\hat{\beta}_k$. Then $\kappa^{ij}\hat{\beta}^l_j = \delta^{li}$, for $\delta$ the array that is one if all indices are equal and zero otherwise, and $\hat{\beta}^l_i = \kappa_{ij}\delta^{jl}$. Here $\kappa$ with superscripts denotes partial derivatives of $\mathcal{K}$ with respect to the corresponding components of $\boldsymbol{\beta}$; for example, $\kappa^{12} = \frac{d^2}{d\beta_1 d\beta_2}\mathcal{K}(\boldsymbol{\beta})$. The symbol $\kappa$ with two subscripts denotes the generic element of the inverse of the matrix of second derivatives of $\mathcal{K}$. We employ the Einstein summation notation, whereby an index repeated as a subscript and as a superscript implies summation over that index. Also, $\kappa^{ijk}\hat{\beta}^l_j\hat{\beta}^m_k + \kappa^{ij}\hat{\beta}^{lm}_j = 0$, and so $\hat{\beta}^{lm}_i = -\kappa_{it}\kappa^{tjk}\hat{\beta}^l_j\hat{\beta}^m_k$, and

$$\begin{aligned}
\hat{\beta}^{lmn}_i &= \kappa_{it}\kappa^{tjk}\kappa_{jw}\kappa^{wvs}\hat{\beta}^l_v\hat{\beta}^m_s\hat{\beta}^n_k - \kappa_{it}\kappa^{tjku}\hat{\beta}^l_j\hat{\beta}^m_k\hat{\beta}^n_u - \kappa_{it}\kappa^{tjk}\hat{\beta}^l_j\hat{\beta}^{mn}_k - \kappa_{it}\kappa^{tjk}\hat{\beta}^{ln}_j\hat{\beta}^m_k \\
&= \kappa_{it}\kappa^{tjk}\kappa_{jr}\kappa^{rvw}\hat{\beta}^l_k\hat{\beta}^n_v\hat{\beta}^m_w + \kappa_{it}\kappa^{tjk}\kappa_{jr}\kappa^{rvw}\hat{\beta}^l_w\hat{\beta}^n_k\hat{\beta}^m_v + \kappa_{it}\kappa^{tjk}\kappa_{jr}\kappa^{rvw}\hat{\beta}^l_v\hat{\beta}^m_w\hat{\beta}^n_k \\
&\quad -\kappa_{it}\kappa^{tjku}\hat{\beta}^l_j\hat{\beta}^m_k\hat{\beta}^n_u.
\end{aligned}$$

The functions $z$ and $w$ are functions of quantities whose derivatives with respect to $t^1$ are easy to calculate; $w(\boldsymbol{t}) = \sqrt{2h(t^1)}$ for $h(t^1) = -\hat{\kappa} + \hat{\beta}_j\hat{\kappa}^j + \tilde{\kappa} - \hat{\gamma}_j\tilde{\kappa}^j$, and

$z(\boldsymbol{t}) = \hat{\beta}_1(t^1)\exp(g(t^1)/2)$, where $g(t^1) = \log(\det[\mathcal{K}''(\hat{\beta})]/\det[\mathcal{K}''_{-1}(\hat{\gamma})])$. Then

$$h'(t^1) = -\hat{\kappa}^j\hat{\beta}^1_j + \hat{\beta}_j\hat{\kappa}^{jk}\hat{\beta}^1_k + \hat{\beta}^1_j\hat{\kappa}^j = \hat{\beta}_j\hat{\kappa}^{jk}\hat{\beta}^1_k = \hat{\beta}_1,$$
$$h''(t^1) = \hat{\beta}^1_1, h'''(t^1) = \hat{\beta}^{11}_1, h''''(t^1) = \hat{\beta}^{111}_1,$$

and $g'(t^1) = \kappa_{ij}\kappa^{ijk}\hat{\beta}^1_k$, and

$$\begin{aligned} g''(t^1) &= \kappa_{ij}\kappa^{ijk}\hat{\beta}^{11}_k + \kappa_{ij}\kappa^{ijkl}\hat{\beta}^1_k\hat{\beta}^1_l - \kappa_{is}\kappa^{stl}\kappa_{tj}\kappa^{ijk}\hat{\beta}^1_k\hat{\beta}^1_l \\ &= -\kappa_{ij}\kappa^{ijk}\kappa_{kl}\kappa^{lmn}\hat{\beta}^1_m\hat{\beta}^1_n + \kappa_{ij}\kappa^{ijkl}\hat{\beta}^1_k\hat{\beta}^1_l - \kappa_{ij}\kappa^{jml}\kappa_{mn}\kappa^{ink}\hat{\beta}^1_k\hat{\beta}^1_l. \end{aligned}$$

Expanding $1/w - 1/z$ in $w$ about 0,

$$\begin{aligned} (3.1)\quad \frac{1}{w(\boldsymbol{t})} - \frac{1}{z(\boldsymbol{t})} &= \left(\frac{1}{2}g'(0)h''(0)^{-1/2} + \frac{1}{3}h^{(3)}(0)h''(0)^{-3/2}\right) \\ &\quad + (-3g'(0)^2h''(0)^{-1} + 6g''(0)h''(0)^{-1} - 6g'(0)h''(0)^{-2}h^{(3)}(0) \\ &\qquad -5h^{(3)}(0)^2h''(0)^{-3} + 3h''(0)^{-2}h^{(4)}(0)) \\ &\quad \cdot w(\boldsymbol{t})/24 + O(w(\boldsymbol{t})^2), \end{aligned}$$

The first term in (3.1) was given by Skovgaard (1987); the term of order $O(w(\boldsymbol{t}))$ is new.

When $\boldsymbol{T}$ is scalar, and hence no effective conditioning is occurring, then $h''(t^1) = \kappa_{11}$, $h'''(t^1) = -(\kappa_{11})^3\kappa^{111}$, $h''''(t^1) = 3(\kappa_{11})^5(\kappa^{111})^2 - (\kappa_{11})^4\kappa^{1111}$, $g'(t^1) = (\kappa_{11})^2\kappa^{111}$, and $g''(t^1) = -2(\kappa_{11})^4(\kappa^{111})^2 + (\kappa_{11})^3\kappa^{1111}$, and

$$(3.2)\qquad 1/w - 1/z = \rho_3/6 + [(3\rho_4 - 5\rho_3^2)/24]w + O(w^2),$$

for $\rho_3 = (\kappa_{11})^{3/2}\kappa^{111}$ and $\rho_4 = (\kappa_{11})^2\kappa^{1111}$.

The quantity $w^2$ is twice the log of the likelihood ratio statistic, and, considering $w$ as a random variable, has a distribution that is well-approximated by a $\chi^2$ distribution with one degree of freedom. Approximations (2.4) and (3.2) imply that for $q$ near zero,

$$(3.3)\qquad P[w^2 \le q] \approx 2\Phi(\sqrt{q}) - 1 + a\sqrt{q}\phi(\sqrt{q}),$$

for $a = (3\rho_4 - 5\rho_3^2)/12$. The right hand side of (3.3) is the first order Taylor expansion of $1 - 2\Phi(\sqrt{q(1-a)})$ as $a$ approaches zero. Quantities $a$ chosen to make a twice the log of a likelihood ratio statistic divided by $1 - a$ more closely approximated by a $\chi^2$ distribution with the appropriate degrees of freedom are examples of Bartlett's correction (Kolassa (1997), p. 148ff). We are grateful to a referee for calling our attention to this interpretation of the quantity we calculate as part of our corrrection.

The expressions (3.1) and (3.2) are valid in the lattice case as well, since $\sinh(0) = 0$, $\sinh'(0) = 1$, and $\sinh''(0) = 0$.

We suggest using (3.1) when $w$ is small enough that direct evaluation of $1/w - 1/z$ causes numerical difficulties. Specifically, we suggest choosing $\epsilon > 0$ such that if $|w| \ge \epsilon$, then (2.4) is used without modification, and if $|w| < \epsilon$ then (3.1) is used. The choice of $\epsilon$ depends primarily on the accuracy with which the saddlepoint equations (2.2) and (2.3) are solved. If $\epsilon$ is chosen too small, then $1/w - 1/z$ will be evaluated in a region where it is numerically unstable. If $\epsilon$ is chosen too large, then some of the accuracy of (2.4) will be sacrificed, and furthermore, the approximation will have a jump as $|w|$ crosses $\epsilon$, potentially adding instability into algorithms that depend on inverting (2.4) to calculate confidence intervals. One might expect that the order $O(w(\mathbf{t})^2)$ error in (3.1) might require some smoothing of the transition between use of $1/w - 1/z$ and (3.1), but in the many examples that we investigated, (3.1) proved accurate enough that no such smoothing was required. We routinely require that numeric solutions to (2.2) and (2.3) make each component of the two sides agree to $10^{-11}$, and choose $\epsilon = 10^{-3}$. This tolerance was used in the calculations that produced the figures below.

## 4. Two examples

Let $F$ be the cumulative distribution function resulting from the convolution of an exponential variable with parameter $\lambda$, and hence mean $1/\lambda$, and a gamma variable with parameters $\alpha$ and $\beta$, and hence mean $\alpha/\beta$. Figure 1 contains an approximation to

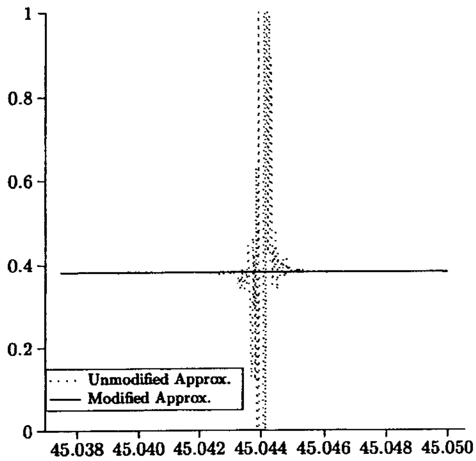


Fig. 1. Comparison of modified and unmodified saddlepoint approximations to the CDF.

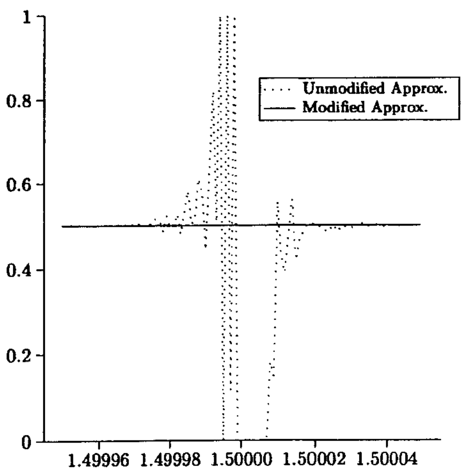


Fig. 2. Comparison of modified and unmodified saddlepoint approximations to the CDF.

$1 - F(x)$. This example was discussed by Huzurbazar and Huzurbazar (1999), and is simple enough to enable use of (3.2). Of particular interest were the values $\lambda = 0.0348$, $\alpha = 3.490$, and $\beta = 0.314$. Pictured are the saddlepoint approximations with and without the refinement presented in the last section.

The refinements of the previous section are perhaps even more useful when $\boldsymbol{T}$ is of length at least two, and (3.2) is unavailable. In this case the saddlepoint equations involve simultaneous nonlinear equations, and one may have to settle for a less accurate solution. Let $X_1$ and $X_2$ be independent variables, each with cumulative distribution function $F$. Figure 2 contains an approximation to a distribution of $X_1$ conditional on $X_1 + X_2$.

## Acknowledgements

This research was supported in part by grant NSF DMS 0092659. The authors are grateful for discussions with faculty at the University of Rochester, and particularly with Dr. Yang's thesis committee members, Dr. Dragalin, Prof. Hall, and Prof. Cranston. We are also grateful for very helpful comments from referees.

## References

Huzurbazar, S., and Huzurbazar, A. V. (1999). Survival and hazard functions for progressive diseases using saddlepoint approximations, *Biometrics*, **55**, 198–203.

Kolassa, J. E. (1997). *Series Approximation Methods in Statistics*, 2nd ed., Springer, New York.

McCullagh, P. (1987). *Tensor Methods in Statistics*, Chapman and Hall, London.

Reid, N. (1988). Saddlepoint methods and statistical inference, *Statist. Sci.*, **3**, 213–238.

Skovgaard, I. M. (1987). Saddlepoint expansions for conditional distributions, *J. Appl. Probab.*, **24**, 875–887.