Comment

Philip Hougaard

I welcome this paper, because there is need to spread the knowledge of the saddlepoint approximation. The approximation is very useful, both in theory and in practice, but it is still not used as much as it deserves. I would like to comment on the tail area approximation and on approximations to ratios. Furthermore, I will present the approximation to the noncentral gamma distribution. This is on one hand an example to show the usefulness and accuracy of the approximation and on the other hand there is need to publish as many specific formulas as possible, because many potential users do not make these derivations on their own. Specific formulas will be helpful for increasing the use of the saddlepoint approximation.

This paper has shown that a number of quantities, sums, estimates, score statistics and likelihood ratio statistics can be approximated in similar ways. Even though the approximations are very good, they can be improved in various ways, by renormalization or by the first terms in the remainder. Interestingly, the longest section is that on further developments. Therefore the paper is filled with ideas and this makes it difficult to do a simple and fast derivation, but it is most inspiring and I have learned a lot from the examples I have studied, including the one reported below. I think we still need more experience with these many approximations, but I also see possibilities for further work. Can the approximation (28) due to Lugannani and Rice be generalized to more complicated models, both curved models and multivariate models? More work is needed on treatment of nuisance parameters, both concerning estimates and test statistics. Also we need substitutes for the moment generating function when it is too complicated.

Concerning approximation of the tail area discussed in Section 6.3, it is typically difficult to explicitly integrate the density approximation. However, in the one-parameter nonlinear regression it is possible. In fact, it is much simpler to derive an approximation for the distribution function rather than the density, because this requires no use of moment generating functions or conjugated families. In these models the set of observations y for which $\hat{\theta}(y) < \theta$ is approximately a half space, given by the derivative $\dot{\eta}$ of the mean value vector function η calculated at θ. Locally around θ the only possible error in this approximation is caused by a set of observations on one side of the mean value space, and further away from it than the radius of curvature. On a global basis it is as usual more difficult to express whether the approximation is good. The advantage of this approach is that the probability of the half space can be calculated exactly by using the mean $\eta_0 = \eta(\theta_0)$ under the true value θ_0. This is different from the standard normal approximation that is based on the linear tangent space of the mean value curve at either θ_0 or $\hat{\theta}$. We find the approximation, first noticed by Pazman (1984),

$$P(\hat{\theta} < \theta) \approx \Phi\{\sigma^{-1}(\dot{\eta}'\dot{\eta})^{-1/2}\dot{\eta}'(\eta - \eta_0)\}$$

where σ^2 is the variance on an observation. The derivative gives the saddlepoint approximation for the density.

The saddlepoint approach is also very useful for studying ratios, say $R = X/Y$, if we know the joint moment generating function of (X, Y). This gives an approximation different from that based on the moment generating function of R. Even though this was noticed already by Daniels (1954, Section 9), it is apparently not well known. If X and Y are bivariate normal we get the approximation suggested by Hinkley (1969). If X and Y are independent and gamma distributed R follows an F distribution. The saddlepoint approach yields a density proportional to the true one. The factor is explained by the Stirling approximation applied to all of the three gamma functions in the normalizing constant. Even if X and Y are dependent, being the diagonal of a Wishart distributed matrix, the approximation is proportional to the true density. This example corresponds to the ratio of variances in a bivariate normal distribution. This was also realized by Daniels (1954).

Finally, I would like to consider an example, which to my knowledge has not been studied before, the noncentral gamma distribution. Cox and Reid (1987) studied other approximations for this distribution. The density is only given as an infinite sum

$$f(x) = e^{-\lambda-\theta x} \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} \frac{\theta^{r+\gamma}}{\Gamma(r+\gamma)} x^{r+\gamma-1},$$

with three parameters θ, γ and λ. The latter is the noncentrality parameter; if $\lambda = 0$, the distribution is the familiar gamma of shape γ and inverse scale θ. The moment generating function is simply

$$M(\phi) = \left(\frac{\theta}{\theta-\phi}\right)^{\gamma} \exp\left\{\lambda\left(\frac{\theta}{\theta-\phi} - 1\right)\right\}.$$

Philip Hougaard is Biostatistician, Novo Research Institute, DK-2880 Bagsvaerd, Denmark.

Interestingly, the equation (2) for ϕ is a quadratic equation, which can be solved explicitly. The solution can be described as $\hat{\phi} = -\theta + \varepsilon^{-1}$, where for $\lambda = 0$, $\varepsilon = -x/\gamma$ and for $\lambda > 0$,

$$\varepsilon = \{-\gamma + (\gamma^2 + 4\lambda\theta x)^{1/2}\}/(2\lambda\theta).$$

It is convenient to express the density by means of ε rather than $\hat{\phi}$ and this gives

$$f(x) = (2\pi)^{-1/2}\{2\theta\lambda\varepsilon^3 + \gamma\varepsilon^2\}^{-1/2}(\theta\varepsilon)^{\gamma} \cdot \exp(-\theta x + x/\varepsilon - \lambda + \lambda\theta\varepsilon).$$

We have excluded n from the formula because the repetition parameter is already included in the parameters. As described by Reid (equation (8)), this formula can be improved by renormalization, at least for $\lambda = 0$. In that case the approximation should be multiplied by

$$c(\gamma, 0) = \gamma^{\gamma-1/2}(2\pi)^{1/2}e^{-\gamma}/\Gamma(\gamma)$$

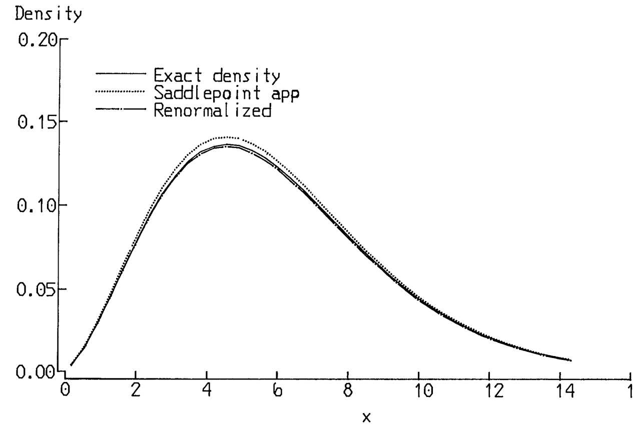
in order to get the true density. This factor is independent of θ. Interestingly, it is helpful to act as if it applies also to non-zero values of λ. This gives for small to moderate values of λ, approximations of an impressive quality. Figure 1 shows an example, $\gamma = 2$, $\theta = 1$, $\lambda = 4$, which has a clear noncentrality. The mean and variance are 6 and 10, respectively. A central gamma with the same mean and γ would have a variance of 18. It is difficult to see any differences on Figure 1 and therefore Figure 2 shows the ratio of the approximate density to the true density. The approximate density named renormalized is $c(\gamma, 0)f(x)$ normalized as if $\lambda = 0$, and stays within 2% of the true density over the range studied. With smaller λ this approximation is even closer to the exact.

In practice we are more interested in fractiles of these distributions and therefore we need approximations to tail areas. The approximation (28) originally due to Lugannani and Rice (1980) is simply calculated and surprisingly accurate in this case. We find

$$y = \operatorname{sgn}(\theta - \varepsilon^{-1})[2\{x(\theta - \varepsilon^{-1}) - (-\lambda + \lambda\alpha y + \gamma \log \alpha y)\}]^{1/2},$$

$$z = (\theta - \varepsilon^{-1})(2\theta\lambda\varepsilon^3 + \gamma\varepsilon^2)^{1/2}.$$

In the example mentioned above, the true 0.95 fractile is 11.888. The approximate fractile is 11.893, which has a true level of 0.95012. This is a typical example. In fact it is very difficult to find cases where the approximate fractile is markedly wrong. The worst cases are the central case ($\lambda = 0$) with few repetitions (small γ) and in particular at the lower to mediate range. For example, in the rather extreme case of



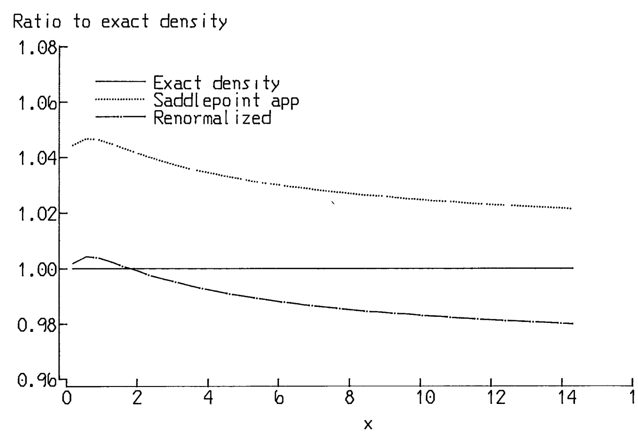


FIG. 2. *Ratio of approximate to exact density of noncentral gamma,* $\gamma = 2$, $\theta = 1$ *and* $\lambda = 4$.

$\gamma = \frac{1}{2}$, corresponding to a single degree of freedom and $\lambda = 0$ the approximate 1% fractile has a true level of 0.88% and the exact probability of being below the approximate median is 48.9%. We cannot demand that an approximation should perform better than this.

That brings me around to my conclusion. The approximations studied by Nancy Reid in this paper have in most cases a terrific accuracy and therefore the only possible objection to them is computational. To be provocative, if the saddlepoint approximation is simpler than the exact distribution, we rarely need the exact distribution.

ADDITIONAL REFERENCES

COX, D. R. and REID, N. (1987). Approximations to noncentral distributions. *Canad. J. Statist.* **15** 105–114.

HINKLEY, D. V. (1969). On the ratio of two correlated normal random variables. *Biometrika* **56** 635–639.

PAZMAN, A. (1984). Probability distribution of the multivariate nonlinear least squares estimates. *Kybernetika* **20** 209–230.