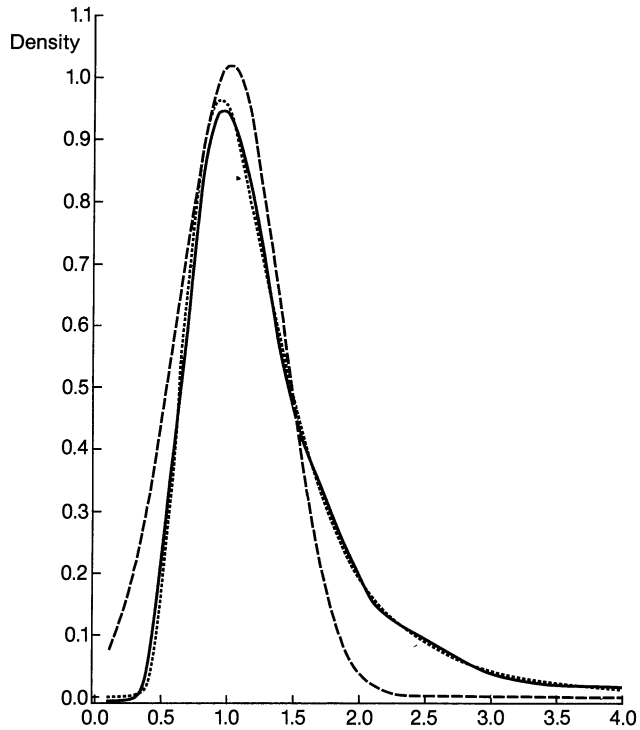

FIG. 1. *Saddlepoint approximation (dotted line) to the density of the maximum likelihood estimator of the shape parameter of a gamma distribution, based on a sample of size* 10. *The "exact" density (solid line) was estimated from* 10,000 *simulations. The normal approximation (dashed line) is shown for comparison.*

To construct a confidence interval for ν it is necessary either to carry out repeated numerical integration or to expand (13) and invert it algebraically. For the general one-parameter case, the details of the expansion are carried out in Barndorff-Nielsen (1985b) and McCullagh (1984a). Unfortunately, the relatively simple approximations to tail probabilities discussed in Daniels (1987) (cf. also Section 6.3) cannot be directly applied to this example because $\hat{\theta}$ is not a one-to-one function of a sample average. It may be possible to adapt the conditional probability tail approximation of Skovgaard (1988b) to this example.

Outside the exponential family setting, the maximum likelihood estimator will not be a one-to-one function of the minimal sufficient statistic, so even if we contemplated using the righthand side of (13) we would not be able to write $L(\theta; x)$ for example, as a function only of θ and $\hat{\theta}$. However, (13) does continue to provide an approximation to a conditional density of $\hat{\theta}$, as is illustrated in the next example.

Example 2. Location-scale family. Suppose $f_X(x; \theta)$ is an arbitrary continuous density on R^1, with θ as a two-dimensional location-scale parameter (μ, σ), so that for an independent, identically distributed sample,

$$f(x_1, \cdots, x_n; \mu, \sigma) = \prod_{i=1}^{n} \sigma^{-1} f_X\{(x_i - \mu)/\sigma\}.$$

Without further assumptions about f, the minimal sufficient statistic is the order statistic $(x_{(1)}, \cdots, x_{(n)})$. It can be separated into two components, $\hat{\theta}$, the maximum likelihood estimate of θ, and $a = (a_1, \cdots, a_n)$, where $a_i = (x_{(i)} - \hat{\mu})/\hat{\sigma}$. The vector a has $n - 2$ independent components, and is ancillary; i.e., its distribution does not depend on θ. The conditional distribution of $\hat{\theta}$, given a, is

$$f_{\hat{\Theta}|A}(\hat{\theta} \mid a; \theta) = c_0(a)\hat{\sigma}^{n-2}\Pi f_X(\hat{\sigma}a_i + \hat{\mu}; \mu, \sigma)$$

which can be re-expressed as

$$(14) \qquad f(\hat{\theta} \mid a; \theta) = c(a) \mid j(\hat{\theta}) \mid^{1/2}\{L(\theta)/L(\hat{\theta})\},$$

using the fact that $\mid j(\hat{\mu}, \hat{\sigma}) \mid = \hat{\sigma}^{-4}d(a)$, where $d(\cdot)$ depends on the derivatives of $\log f$.

Note the similarity of (14) to approximation (13), and also that (14) is the exact conditional density of $\hat{\theta}$, given the maximal ancillary a. Fisher (1934) derived (14) and argued that inference for θ should be based on this conditional distribution; see also Cox and Hinkley (1974, page 115). Different versions of formula (14) have been derived by several authors, including Pitman (1938), Fraser (1968), Efron and Hinkley (1978), and Barndorff-Nielsen (1980, 1983). Barndorff-Nielsen (1983) emphasized the similarity of (13) and (14), and showed further that (14) provides an expression for the conditional density of the maximum likelihood estimate in any transformation model, i.e., any model generated by a group.

That the same formula provides either a highly accurate approximation or an exact expression for the distribution of the maximum likelihood estimator in full exponential families or transformation families is rather surprising. Exponential families and transformation families are usually considered to be quite different statistical objects, but this suggests that there may be a close connection between them. McCullagh (1987, Chapter 8) has investigated to what extent an arbitrary family of densities can be made to "look like" an exponential family, by conditioning on some approximately distribution-free statistic. Also relevant is Mitchell (1988) in which the geometry of a subclass of transformation models, the elliptic families, is studied. This geometry has some striking similarities to the geometry of exponential families outlined in Amari (1982; 1985, Chapter 2) and Efron (1978).

What about densities that are not members of exponential or transformation families? Remarkably, the same formula continues to provide an approximation to the conditional distribution of the maximum likelihood estimate, conditioned on an approximately ancillary statistic a. Approximately ancillary is taken to mean that the distribution of a depends on θ only in terms of $O(n^{-1})$ or higher, for θ