of renormalization in the $k$-dimensional case, where the leading term of $R_n$ is given by (7).

A referee has suggested that the renormalization issue can be clarified by choosing the simpler renormalization of dividing by $f(\mu)$. Using $\tilde{f}(\cdot)$ to denote the saddlepoint approximation to $f(\cdot)$, we can write

$$f(\bar{x}) = \tilde{f}(\bar{x}) \left[ 1 + \frac{(\bar{x} - \mu)R_n(\bar{x})}{1 + R_n(0)/n} \right],$$

for some $\xi$ between $\bar{x}$ and $\mu$, by the mean value theorem. Because $R_n$ and $R_n'$ are uniformly bounded in a neighborhood of $\mu$, the renormalized approximation has a relative error of $O(n^{-1})$ for fixed $|\bar{x} - \mu|$ and $O(n^{-3/2})$ for $|\bar{x} - \mu| = O(n^{-1/2})$.

Many of the statistical applications of (1) derive from the fact that the saddlepoint $\phi = \hat{\phi}(\bar{x})$ is the maximum likelihood estimate of the “parameter” $\phi$ in the density (5) and that $n[K(\hat{\phi}) - \hat{\phi}^T \bar{x}]$ is the maximized “log likelihood ratio” for an independent, identically distributed sample from that density. The density is an artificial construct, however, and the true parameter of interest will be the parameter of the original density $f_X(\cdot)$, which has been suppressed in the present notation. An important exception is the case when $f_X(\cdot)$ is itself an exponential family density with canonical parameter $\theta$, in which case there is a simple relationship between $\hat{\theta}$ and $\hat{\phi}$, and between $n[K(\hat{\phi}) - \hat{\phi}^T \bar{x}]$ and the maximized log likelihood ratio statistic for $\theta$. This connection will be exploited in each of the next three sections.

3. MAXIMUM LIKELIHOOD ESTIMATION

Suppose that the density of $X$, takes the exponential family form

$$f_X(x; \theta) = \exp\{\theta^T t(x) - \psi(\theta) - d(x)\}$$

where $\theta$ is the canonical parameter and $t(x) = (t_1(x), \cdots, t_k(x))^T$ is the minimal sufficient statistic. The conjugate family is again of the form (10), with canonical parameter $\bar{\theta} + \phi$, and $K(\phi) = \psi(\bar{\theta} + \phi) - \psi(\theta)$ is the cumulant generating function for $f_X(x; \theta)$. The density of $T = \sum t(X_i)$ is given by

$$f_T(t; \theta) = \exp[\theta^T t - \psi(\theta) - h(t)],$$

and we now approximate $\exp[-h(t)]$ using the saddlepoint approximation. The saddlepoint equation is simply $n \psi'(\phi) = t$, so that the saddlepoint occurs at the maximum likelihood estimate $\hat{\theta}$, and (1) gives

$$f_T(t; \theta) = (2\pi)^{-k/2} |n \psi'(\hat{\theta})|^{-1/2} \exp\{\theta^T t - \psi(\theta) + n \psi'(\hat{\theta}) \cdot [1 + O(n^{-1})]\}.$$

As was first pointed out in Daniels (1958), equation (11) has a very simple likelihood formulation:

$$f_T(t; \theta) = (2\pi)^{-k/2} |n \psi'(\hat{\theta})|^{-1/2} \exp[\theta^T t - \psi(\theta) - \psi'(\hat{\theta}) \cdot [1 + O(n^{-1})]].$$

In (12) $L(\theta)$ is the joint likelihood for the sample $x_1, \cdots, x_n$ and $j(\theta) = -\bar{\theta} \log L(\theta)/\partial \theta \bar{\theta}$ is the observed Fisher information, in this case equal to $n \psi''(\theta)$. Both $L$ and $j$ should properly be written $L(\theta, t)$ and $j(\theta, t)$, to emphasize their dependence on the data. The transformation from $t$ to $\theta$ is one-to-one, with Jacobian $|j(\theta)|$, giving

$$f_0(\theta; \theta) = c |j(\theta)|^{-1/2} \exp[L(\theta)/L(\hat{\theta})][1 + O(n^{-3/2})].$$

We have replaced $2\pi^{k/2}$ by $c$ to indicate that an improvement via renormalization is incorporated into (13). In this case the renormalization does reduce the error to $O(n^{-3/2})$, because the region of integration can be truncated to $\sqrt{n} |\theta - \theta| < \epsilon$ and the error incurred is exponentially small (Durbin, 1980a, Section 2.3).

The righthand side of (13) is often called Barndorff-Nielsen’s approximation, as Barndorff-Nielsen has investigated extensively its application outside the exponential family. This will be discussed in detail below, but for the moment we simply point out that it is readily obtained from the saddlepoint approximation, in full exponential families. The argument outlined above is given in Barndorff-Nielsen (1983).

Example 1. Gamma density with unknown shape. We write $f_X(x) = (\nu/\mu)^x e^{-x/\mu} \Gamma^{-1}(\nu)$. The maximum likelihood estimate $\hat{\theta} = (\hat{\mu}, \hat{\nu})$ is given by $\hat{\mu} = t_0/n$ and $\hat{\nu} = (t_1/n - \log t_0/n)$, where $t_0 = \sum \log x_i$, $t_1 = \sum x_i$, and $\psi(\nu)$ is the digamma function $d \log \Gamma(\nu)/d\nu$. It is easily shown that (13) is

$$f(\hat{\mu}, \hat{\nu}; \nu, \nu) \equiv g_1(\hat{\mu}; \mu, \nu) g_2(\hat{\nu}; \nu)$$

where

$$g_1(\hat{\nu}; \mu, \nu) = (\nu/\mu)^{\nu/\mu} \cdot \exp(-n \nu \hat{\mu}/\mu)$$

and

$$g_2(\hat{\nu}; \nu) = \Gamma^n(\nu) \Gamma^{n}(\nu)[\psi'\hat{\nu} - 1]^{1/2} \cdot \exp[n|\hat{\nu} - \nu] \psi'\hat{\nu} - \nu \ln \psi]\]$$

showing that $(\hat{\mu}, \hat{\nu})$ are independent to the order considered, and that the approximation to the density of $\hat{\nu}$ is exact after renormalization. The renormalized version of $g_0(\hat{\nu}; \nu)$ is displayed in Figure 1, for $n = 10$, $\nu = 1$. Also shown are the “exact” density of $\hat{\nu}$ estimated from 10,000 simulations, and the approximating normal density with mean $\nu$ and variance $\nu^2$.

The gamma example is further discussed in Jensen (1986b), where an alternative approximation to the density of $\hat{\nu}$ and approximations to the similar test for $\mu$ are derived.