Brazilian Journal of Probability and Statistics (2001), 15, pp. 85–105.

SKEWNESS AND KURTOSIS FOR MAXIMUM LIKELIHOOD ESTIMATOR IN ONE-PARAMETER EXPONENTIAL FAMILY MODELS

Francisco José de A. Cysneiros[1], **Sylvio José P. dos Santos**[1] and **Gauss M. Cordeiro**[2]

[1] *Departamento de Estatística, CCEN/UFPE, Cidade Universitária, 50749-540, Recife - PE, Brazil. E-mails: cysneiros@de.ufpe.br and sylvio@de.ufpe.br*

[2] *Departamento de Estatística, UFBA, Av. Adhemar de Barros, s/n, Ondina, 40170-110, Salvador - BA, Brazil. E-mail: gauss@ufba.br*

Summary

In this paper we derive approximate formulae for the skewness and kurtosis of the maximum likelihood estimator in the one-parameter exponential family. The key idea underlying these formulae is that they indicate when the normal approximation usually employed with maximum likelihood estimators can be misleading in small samples. We apply our main result to a number of special distributions of this family. We also use a graphical analysis to examine how the skewness and kurtosis vary with the true value of the parameter in some special cases.

Key Words: Asymptotic expansion; exponential family; kurtosis; maximum likelihood estimator; skewness.

1 Introduction

The assumption of symmetry plays a crucial role in many statistical procedures. The notion of skewness of a distribution is related to a symmetry property. The most commonly used measure of skewness is the standardized third cumulant. In fact, the classical tests of symmetry make use of the standardized third sample cumulant measure and a departure from the normal value of zero then indicates skewness. Intuitively, we think of a distribution as being skewed if it systematically deviates from symmetrical form. Kurtosis is a measure of a type of departure from normality. The kurtosis is given by the standardized fourth cumulant which equals zero for any normal distributions. Often "peaked" (as compared with normal) distributions have positive kurtosis, and flat-topped-ones have negative kurtosis.

Any distribution expressed in standardized form has zero mean and unit variance. The standardized distributions can be readily compared in regard to form, its departure from symmetry (skewness) and other qualities, though not of course in regard to mean and variance. Commonly used indices of the shape of a distribution are the moment ratios, namely the indices of skewness γ_1 and kurtosis γ_2 defined by $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ and $\gamma_2 = \kappa_4/\kappa_2^2$, respectively, where κ_r is the rth cumulant of the distribution. The indices γ_1 and γ_2 are widely used as measures of departure from normality since $\gamma_1 = \gamma_2 = 0$ for the normal distribution. These univariate measures are constructed such that they are invariant under change of scale and origin, γ_1 is a function of κ_3, the lowest cumulant measuring symmetry and γ_2 is a function of κ_4, the lowest cumulant measuring "peakedness". When $\gamma_1 > 0$ ($\gamma_1 < 0$) the distribution is positively (negatively) skewed and will have a longer (shorter) right tail and a shorter (longer) left tail. Clearly, if the distribution is symmetrical, γ_1 vanishes and therefore its value will give some indication of the extent of departure from symmetry. According to van Zwet (1964), if we transform Y to $\varphi(Y)$ then we increase right-skewness if φ is convex, while we decrease right-skewness if φ is concave. Therefore, $\varphi(Y)$ will have a greater or smaller skewness coefficient than Y if φ is convex or concave. Distributions for which $\gamma_2 = 0$ are called mesokurtic. The distributions for which $\gamma_2 > 0$ are called leptokurtic and those for which $\gamma_2 < 0$ are called platykurtic. The leptokurtic distribution has a sharper peak at the mode and more extended tails, whereas the platykurtic distribution is characterized by a flatter top and more abrupt terminals than the normal curve. It is impossible to have $\gamma_1^2 > 2 + \gamma_2$ and thus $\gamma_2 \geq -2$ always. We note that γ_2 can be interpreted as a nonnormality adjustment for the variance of $(Y - E(Y))^2$ since $Var\{(Y - E(Y))^2\} = \kappa_4 + 2\kappa_2^2 = 2\kappa_2^2(1 + \gamma_2/2)$. The moment ratios γ_1 and γ_2 have the same values for any linear function $a + bY$ with $b > 0$. When $b < 0$, the absolute values are not altered, but ratios of odd order cumulants have their signs reversed.

Consider a general uniparametric model $f(y;\theta)$ indexed by an unknown scalar parameter $\theta \in \Theta$, where Θ is an open set of $\mathbb{R}$. Let y be the data vector of n observations which are assumed independent identically distributed with log likelihood for a single observation defined by $l(\theta; y) = \log f(y;\theta)$. We know that the maximum likelihood estimator (MLE) $\hat\theta$ of an unknown scalar parameter θ in regular problems is asymptotically distributed as $N(\theta, (n\kappa_{\theta,\theta}^{-1}))$ with an error of order $O(n^{-1/2})$, where $\kappa_{\theta,\theta} = E\{-\frac{d^2 l(\theta;y)}{d\theta^2}\}$ is the expected information for θ. Our main purpose here is to obtain simple asymptotic formulae for the indices $\gamma_1(\hat\theta)$ and $\gamma_2(\hat\theta)$ of the distribution of the MLE $\hat\theta$ in one-parameter exponential family models up to orders

$n^{-1/2}$ and n^{-1}, respectively. The values of the indices $\gamma_1(\hat\theta)$ and $\gamma_2(\hat\theta)$ can be used as measures of nonnormality of the distribution of $\hat\theta$ since they vanish when $\hat\theta$ is normally distributed. These include many important and commonly used distributions and only require knowledge of simple functions and their first few derivatives with respect to θ. It is possible to use the values of $\gamma_1(\hat\theta)$ and $\gamma_2(\hat\theta)$ to examine for what exponential models the distribution of $\hat\theta$ is closer to the normal distribution. The dependence of the finite-sample distribution of the MLE $\hat\theta$ on the sample size and on the value of θ is assessed by numerical and graphical inspection for some of the distributions considered.

The plan of the paper is as follows. In Section 2 we give simple asymptotic formulae for the standardized cumulants $\gamma_1(\hat\theta)$ and $\gamma_2(\hat\theta)$ in one-parameter exponential family models. We also discuss the effects that these standardized cumulants have on coverage of confidence intervals for θ and review two different higher-order refinements to obtain highly accurate small-sample confidence intervals for this parameter. However, the method employed here is both simpler and more direct than higher-order methods which give near-exact results. In Section 3 we present a number of special cases thus showing that our main result covers a wide range of important distributions. A graphical analysis that shows the dependence of the skewness $\gamma_1(\hat\theta)$ and kurtosis $\gamma_2(\hat\theta)$ on θ is performed in Section 4. Concluding comments are in Section 5.

2 Basic formula

Consider a general uniparametric model indexed by an unknown scalar parameter θ. Suppose there are n independent and identically distributed observations $y_1, \ldots, y_n$ and define the log likelihood function for a single observation by $l(\theta) = l(\theta; y)$. We assume that $l(\theta)$ satisfies the regularity conditions stated in Rao (1973, p.364) and Serfling (1980, p.144). The derivatives of the log likelihood $l(\theta)$ are denoted by $U_\theta = dl(\theta)/d\theta, U_{\theta\theta} = d^2l(\theta)/d\theta^2$, etc., and we use the following notation for their cumulants (Lawley, 1956): $\kappa_{\theta\theta} = E(U_{\theta\theta})$, $\kappa_{\theta\theta\theta} = E(U_{\theta\theta\theta})$, $\kappa_{\theta,\theta} = E(U_\theta^2)$, $\kappa_{\theta,\theta\theta} = E(U_\theta U_{\theta\theta})$, $\kappa_{\theta\theta,\theta\theta} = E(U_{\theta\theta}^2) - \kappa_{\theta\theta}^2$, $\kappa_{\theta\theta\theta\theta} = E(U_{\theta\theta\theta\theta})$, $\kappa_{\theta,\theta,\theta\theta} = E(U_\theta^2 U_{\theta\theta}) - \kappa_{\theta,\theta}\kappa_{\theta\theta}$, $\kappa_{\theta,\theta,\theta,\theta} = E(U_\theta^4) - 3\kappa_{\theta,\theta}^2$ and $\kappa_{\theta,\theta\theta\theta} = E(U_\theta U_{\theta\theta\theta})$.

We denote the derivatives of the cumulants with superscripts as $\kappa_{\theta\theta}^{(\theta)} = d\kappa_{\theta\theta}/d\theta$, etc. All $\kappa's$ refer to a single observation and are of order $O(1)$. Under these regularity conditions the asymptotic distribution of the MLE $\hat\theta$ is normal $N(\theta, n^{-1}\kappa_{\theta,\theta}^{-1})$ with an error of order $O(n^{-1/2})$. The $\kappa's$ satisfy certain Bartlett identities such as $\kappa_{\theta,\theta} = -\kappa_{\theta\theta}$, $\kappa_{\theta,\theta,\theta} = -\kappa_{\theta\theta\theta} - 3\kappa_{\theta,\theta\theta} = 2\kappa_{\theta\theta\theta} - 3\kappa_{\theta\theta}^{(\theta)}$, $\kappa_{\theta,\theta\theta} = \kappa_{\theta\theta}^{(\theta)} - \kappa_{\theta\theta\theta}$, $\kappa_{\theta,\theta,\theta,\theta} = -\kappa_{\theta\theta\theta\theta} - 4\kappa_{\theta,\theta\theta\theta} - 6\kappa_{\theta,\theta,\theta\theta} - 3\kappa_{\theta\theta,\theta\theta} =$

$-3\kappa_{\theta\theta\theta\theta}+8\kappa_{\theta\theta\theta}^{(\theta)}-6\kappa_{\theta\theta}^{(\theta\theta)}+3\kappa_{\theta\theta,\theta\theta}$, $\kappa_{\theta,\theta,\theta\theta}=\kappa_{\theta\theta\theta\theta}-2\kappa_{\theta\theta\theta}^{(\theta)}+\kappa_{\theta\theta}^{(\theta\theta)}-\kappa_{\theta\theta,\theta\theta}$, etc. (see Lawley, 1956).

Let $\mu_r(\hat{\theta}) = E\{(\hat{\theta}-\theta)^r\}$ for $r = 3$ and 4 be the third and fourth central moments of $\hat{\theta}$, respectively. Using the general formulae for $\mu_3(\hat{\theta})$ and $\mu_4(\hat{\theta})$ given by Shenton and Bowman (1963, equations (17c)-(17d); 1977) and some Bartlett identities, which usually simplify the computation of the $\kappa' s$, we can express the third and fourth central moments of $\hat{\theta}$ to orders n^{-2} and n^{-3}, respectively, by

$$\mu_3(\hat{\theta}) = \frac{3\kappa_{\theta\theta}^{(\theta)} - \kappa_{\theta\theta\theta}}{n^2\kappa_{\theta,\theta}^3} \tag{2.1}$$

and

$$\mu_4(\hat{\theta}) = \frac{3}{n^2\kappa_{\theta,\theta}^2} - \frac{9\kappa_{\theta,\theta,\theta,\theta} + 36\kappa_{\theta,\theta,\theta\theta} + 10\kappa_{\theta,\theta\theta\theta}}{n^3\kappa_{\theta,\theta}^4} + \frac{54\kappa_{\theta\theta}^{(\theta)^2} - 3\kappa_{\theta\theta\theta}^2 - 24\kappa_{\theta\theta\theta}\kappa_{\theta\theta}^{(\theta)}}{n^3\kappa_{\theta,\theta}^5}. \tag{2.2}$$

Equations (2.1) and (2.2) are in agreement with the formulae in Peers and Iqbal (1985) who give the asymptotic expansions of the second, third and fourth cumulants of the MLE for a general $p-$dimensional model, $p > 1$.

The third cumulant of $\hat{\theta}$ is simply $\kappa_3(\hat{\theta}) = \mu_3(\hat{\theta})$ whereas the fourth cumulant of $\hat{\theta}$ to order $O(n^{-3})$ is given by the last two terms in equation (2.2) since $\kappa_4(\hat{\theta}) = \mu_4(\hat{\theta}) - 3Var(\hat{\theta})^2$. Our aim here is to give formulae for the skewness $\kappa_3(\hat{\theta})$ and kurtosis $\kappa_4(\hat{\theta})$ of $\hat{\theta}$ in the one-parameter exponential model which are algebraically more appealing for applications than the general formulae (2.1) and (2.2). Unlike these formulae, our results can be readily used by applied researchers since they only require trivial operations on suitably defined functions and their derivatives.

Let $Y_1, \ldots, Y_n$ be a set of n independent and identically distributed random variables with probability or density function in the one-parameter exponential family, that is,

$$f(y;\theta) = \frac{1}{\zeta(\theta)} exp\{-\alpha(\theta)d(y) + v(y)\}, \tag{2.3}$$

where θ is a scalar parameter, $\alpha(.), \zeta(.), d(.)$ and $v(.)$ are known functions and $\theta \in \Theta$, Θ being an open set of $\mathbb{R}$. We also assume that the support of $f(y;\theta)$ does not depend on the unknown parameter θ and that $\alpha(.)$ and $\zeta(.)$ have continuous first four derivatives with respect to θ, and that ζ

is positive valued. Further, we require that $d\alpha(\theta)/d\theta$ and $d\beta(\theta)/d\theta$ are different from zero for all values of $\theta \in \Theta$, where $\beta(\theta) = E\{-d(y)\}$ is given by $\beta(\theta) = \frac{d\zeta(\theta)}{d\theta}(\zeta(\theta)\frac{d\alpha(\theta)}{d\theta})^{-1}$. From now on we omit the dependence of $\alpha(\theta)$, $\zeta(\theta)$ and $\beta(\theta)$ on θ with primes denoting derivatives with respect to the unknown parameter θ.

Many commonly used distributions in applied research are special cases of (2.3). Also, this family of distributions enjoys important mathematical properties; see Bickel and Doksum (1977) and Barndorff-Nielsen (1978). As is well known exponential family models allow for a unified treatment of several important distributions and have a number of interesting statistical properties for estimation, testing and inference problems.

Let $y_1, \ldots, y_n$ be the data set of n observations from (2.3). The maximum likelihood estimator (MLE) $\hat{\theta}$ of θ comes from $n^{-1}\sum_{i=1}^{n} d(y_i) = -\beta(\hat{\theta})$ if the solution to this equation belongs to Θ. For several important distributions in (2.3) the MLE $\hat{\theta}$ cannot be expressed as an explicit function of the data. To circumvent this problem we usually use iterative techniques to derive approximate solutions to the exact MLE.

Bias correction for the MLE $\hat{\theta}$ is discussed by Firth (1993). Ferrari et al. (1996) obtained the bias and variance of the MLE $\hat{\theta}$ in the exponential family (2.3) up to order n^{-2}. Here, "to order n^{-p}" means that terms of order smaller than n^{-p} are neglected. The expressions for the bias and variance of $\hat{\theta}$ only require knowledge of $\alpha(\theta)$ and $\zeta(\theta)$ and their first five derivatives with respect to θ. Cribari-Neto et al. (1998) give closed-form expressions for the second and third order biases of $\hat{\theta}$ for a number of distributions in (2.3). Cordeiro et al. (1999) proposed a pivotal quantity which is a function of $\hat{\theta}$ and whose distribution is standard normal up to order $n^{-3/2}$. The proposed pivot takes the form of a polynomial transformation of the standardized MLE of at most third degree. In this section we give asymptotic formulae for the standardized skewness and kurtosis of the distribution of the MLE $\hat{\theta}$ in the one-parameter exponential family (2.3) up to orders $n^{-1/2}$ and n^{-1}, respectively. In the next section we use these formulae to obtain approximate closed-form expressions for the skewness and kurtosis of $\hat{\theta}$ for a number of important distributions in this family.

Now, the log likelihood $l(\theta) = l(\theta; y)$ for the model (2.3) is $l(\theta) = -\alpha(\theta)d(y) - \log\zeta(\theta) + v(y)$. From the first four derivatives of $l(\theta)$ with respect to θ and using the Bartlett identities we obtain: $\kappa_{\theta,\theta} = \alpha'\beta'$, $\kappa_{\theta\theta\theta} = -2\alpha''\beta' - \alpha'\beta''$, $\kappa_{\theta\theta}^{(\theta)} = -\alpha''\beta' - \alpha'\beta''$, $\kappa_{\theta,\theta\theta} = \alpha''\beta'$, $\kappa_{\theta,\theta,\theta} = \alpha'\beta'' - \alpha''\beta'$, $\kappa_{\theta,\theta\theta\theta} = \alpha'''\beta'$, $\kappa_{\theta,\theta,\theta\theta} = \alpha''\beta'' - \alpha''^2\beta'/\alpha'$ and $\kappa_{\theta,\theta,\theta,\theta} = -\alpha'''\beta' - 3\alpha''\beta'' + \alpha'\beta''' + 3\alpha''^2\beta'/\alpha'$. Finally, replacing these cumulants in the expressions

(2.1) and (2.2), we get after some algebra

$$\kappa_3(\hat{\theta}) = \frac{-2\alpha'\beta'' - \alpha''\beta'}{n^2(\alpha'\beta')^3} \tag{2.4}$$

and

$$\begin{aligned}\kappa_4(\hat{\theta}) &= \frac{27(\alpha'\beta'')^2 - 6(\alpha''\beta')^2 + 24\alpha''\alpha'\beta''\beta'}{n^3(\alpha'\beta')^5} \\ &\quad - \frac{\alpha'''\alpha'\beta' + 9\alpha''\alpha'\beta'' + 9\alpha'^2\beta''' - 9\alpha''^2\beta'}{n^3\alpha'^5\beta'^4}.\end{aligned} \tag{2.5}$$

More convenient quantities than $\kappa_3(\hat{\theta})$ and $\kappa_4(\hat{\theta})$ for certain purposes are the standardized cumulants $\gamma_1(\hat{\theta}) = \kappa_3(\hat{\theta})/Var(\hat{\theta})^{3/2}$ and $\gamma_2(\hat{\theta}) = \kappa_4(\hat{\theta})/Var(\hat{\theta})^2$, where $Var(\hat{\theta}) = \frac{1}{n\alpha'\beta'}$ is the asymptotic variance of $\hat{\theta}$. From (2.4) and (2.5) we then obtain the standardized cumulants $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ to orders $n^{-1/2}$ and n^{-1}, respectively, as

$$\gamma_1(\hat{\theta}) = \frac{-2\alpha'\beta'' - \alpha''\beta'}{\sqrt{n}(\alpha'\beta'\sqrt{\alpha'\beta'})} \tag{2.6}$$

and

$$\gamma_2(\hat{\theta}) = \frac{27(\alpha'\beta'')^2 + 3(\alpha''\beta')^2 + 15\alpha''\alpha'\beta''\beta' - \alpha'\beta'(\alpha'''\beta' + 9\alpha'\beta''')}{n(\alpha'\beta')^3}. \tag{2.7}$$

Some features of the formulae (2.4)–(2.7) are noteworthy. First, the asymptotic expressions for the skewness and kurtosis of the distribution of the MLE $\hat{\theta}$ are now very easy to compute for any exponential family model. They depend on the model through the functions $\alpha(\theta)$ and $\beta(\theta)$ and their first three derivatives with respect to θ. Second, although the calculation of $\mu_3(\hat{\theta})$, $\mu_4(\hat{\theta})$, $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ is straightforward for any distribution in (2.3), it is rather difficulty to explain the general structures of their formulae. The main difficulty in interpreting these formulae is that their individual terms are not invariant under reparameterization and therefore they have no geometric interpretation which is independent of the coordinate system chosen. Third, by entering equations (2.6) and (2.7) into a computer algebra system such as MATHEMATICA (Wolfram, 1996) or MAPLE (Abell and Braselton, 1994), one can obtain the standardized cumulants $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ for several models with minimal effort (see Section 3). Further, a simple application of our method can be performed easily by hand using directly equations (2.6) and (2.7), although application of near-exact higher-order methods usually require computer algebra. Fourth, when $\alpha(\theta) = \theta$, which corresponds to the natural exponential family, $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ reduce to

$$\gamma_1(\hat{\theta}) = \frac{-2\beta''}{\beta'\sqrt{n\beta'}} \tag{2.8}$$

and

$$\gamma_2(\hat{\theta}) = \frac{27\beta''^2 - 9\beta'\beta'''}{n\beta'^3}, \tag{2.9}$$

where $\beta(\theta) = d\log\zeta(\theta)/d\theta$. Finally, if the distribution of $\hat{\theta}$ is approximately normal, both coefficients $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ should be close to zero and departure from this value is evidence to the contrary.

Often one is interested in the estimation of a function of θ, say $\tau = g(\theta)$, which does not depend on n. One can reparameterize the model in terms of τ and then use the results (2.6)–(2.7) to obtain the skewness $\gamma_1(\hat{\tau})$ and kurtosis $\gamma_2(\hat{\tau})$ of the distribution of the MLE $\hat{\tau} = g(\hat{\theta})$. Alternatively, we can express $\gamma_1(\hat{\tau})$ and $\gamma_2(\hat{\tau})$ as functions of the standardized cumulants $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ of $\hat{\theta}$. After some algebra, we obtain

$$\gamma_1(\hat{\tau}) = \gamma_1(\hat{\theta}) + \frac{3\tau''}{\sqrt{n\alpha'\beta'}}$$

and

$$\gamma_2(\hat{\tau}) = \gamma_2(\hat{\theta}) + \frac{-42\tau''(\alpha'\beta'') - 18\tau''(\alpha''\beta') + 25\tau''^2(\alpha'\beta') + 10\tau'''\tau'(\alpha'\beta')}{n(\alpha'\beta')^2}.$$

Formulae (2.6) and (2.7) are useful in telling when it is safe to base inference on the asymptotic normal distribution of the MLE $\hat{\theta}$ and to compute first-order confidence intervals for θ from $\hat{\theta} \pm 1.96(n\kappa_{\theta,\theta})^{-1/2}$. In real-life situations, if the values of the skewness and kurtosis indices $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ are large, the first-order normal approximation for $\hat{\theta}$ will be probably poor in samples of small to moderate size, and therefore may deliver inaccurate inference. In these cases, we have to work with second-order distributional refinements for the cumulative distribution function of the MLE $\hat{\theta}$. Two different methods of adjustments to improve the asymptotic standard normal approximation to the distribution of $\hat{\theta}$ are considered below. Despite their usefulness, the improved intervals based on these methods entail more algebra than those intervals based on the asymptotic normal $N(0, \kappa_{\theta,\theta}^{-1})$ distribution of $\sqrt{n}(\hat{\theta} - \theta)$.

2.1 The Edgeworth expansion

The second-order asymptotic distribution theory associated with Edgeworth expansion for the distribution function of $\hat{\theta}$ is summarized by Hill and Davis

(1968). Provided some regularity conditions hold, they showed that the distribution function of the pivot statistic $S = \sqrt{n}(\hat{\theta} - \theta)\kappa_{\theta,\theta}^{1/2}$ is given to order $O(n^{-1})$ by

$$\begin{aligned} P(S \leq z; y) &= \Phi(z) - \phi(z) \left[\{6\eta_1 + \eta_3 h_2(z)\} /6 \right. \\ &\quad \left. + \{360\eta_2 h_1(z) + 30\eta_4 h_3(z) + \eta_6 h_5(z)\} /24 \right], \end{aligned} \tag{2.10}$$

where $\Phi(.)$ and $\phi(.)$ are the cumulative distribution function and the density function of the standard normal variate, respectively, and $h_j(z)$ is the jth degree Hermite polynomial. The coefficients in (2.10) are given by $\eta_1 = (n\kappa_{\theta,\theta})^{1/2} B_1(\hat{\theta})$, $\eta_2 = n\kappa_{\theta,\theta}\{V_2(\hat{\theta}) + B_1(\hat{\theta})^2\}$, $\eta_3 = \gamma_1(\hat{\theta})$, $\eta_4 = \gamma_2(\hat{\theta}) + 4(n\kappa_{\theta,\theta})^{1/2} B_1(\hat{\theta})\gamma_1(\hat{\theta})$ and $\eta_6 = 10\gamma_1(\hat{\theta})^2$, where $B_1(\hat{\theta})$ and $V_2(\hat{\theta})$ are the n^{-1} and n^{-2} terms in the bias and variance of $\hat{\theta}$, respectively. Expressions for $B_1(\hat{\theta})$ and $V_2(\hat{\theta})$ under the one-parameter exponential model (2.3) are given by Ferrari et al. (1996). The coefficients η_1 and η_3 are of order $O(n^{-1/2})$ whereas η_2, η_4 and η_6 are of order $O(n^{-1})$.

Confidence intervals for θ computed from (2.10), which take into account the formulae of $B_1(\hat{\theta}), V_2(\hat{\theta}), \gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$, can be viewed as meaningful improvements over the usual confidence intervals $\{\theta; \mid S \mid \leq z\}$, where z is the appropriate upper point of a standard normal distribution. Equation (2.10) shows clearly that the bias term $B_1(\hat{\theta})$ and skewness $\gamma_1(\hat{\theta})$ affect both the corrected terms of orders $O(n^{-1/2})$ and $O(n^{-1})$ in the distribution function of S. However, the kurtosis $\gamma_2(\hat{\theta})$ only affects the correction term of order $O(n^{-1})$. It is then clear from the degrees of the polynomials in (2.10) that the inference based on the asymptotic normal distribution of $\hat{\theta}$ can be innacurate in the tails of the distribution of S, when $\mid z \mid$ is not very small, if any of the basic quantities $B_1(\hat{\theta}), \gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$, which correspond to terms involving z^3 and z^5, is large.

2.2 The modified direct likelihood

We now outline the modified direct likelihood method that yields approximate interval probabilities of one-dimensional distribution functions. Under a one-dimensional exponential family model, the cumulative distribution function of $\hat{\theta}$ can be approximated by (Barndorff-Nielsen, 1990; Fraser, 1990; Barndorff-Nielsen and Cox, 1994, Chapter 6)

$$P(\hat{\theta} \leq \theta; y) = \Phi(r) + \phi(r)\left(\frac{1}{r} - \frac{1}{u}\right) + O(n^{-3/2}), \tag{2.11}$$

where r is the modified direct likelihood given by

$$r = sgn(\hat{\theta} - \theta) \left[2 \left\{ l(\hat{\theta}; y) - l(\theta; y) \right\} \right]^{1/2}.$$

Here $l(\theta; y)$ is the log likelihood function and the likelihood quantity u is easily calculated from

$$u = \left\{ \frac{\partial l(\theta; y)}{\partial y} \Big|_{\hat{\theta}} - \frac{\partial l(\theta; y)}{\partial y} \right\} k(\hat{\theta})^{-1} j(\hat{\theta})^{1/2},$$

where $j(\theta) = -\frac{\partial^2 l(\theta;y)}{\partial \theta^2}$ is the observed information and $k(\theta) = \frac{\partial^2 l(\theta;y)}{\partial \theta \partial y}$.

The derivation of (2.11) using saddlepoint techniques is reviewed in Barndorff-Nielsen and Cox (1994, Chapter 6). Note that r is the signed square root of the log likelihood ratio statistic, and u is the standardized MLE. Here, u is the Wald statistic only if the exponential family is a natural one, i.e. if $\alpha(\theta) = \theta$ and $d(y) = y!$. To first order both r and u have standard normal distributions and the correction term involving $\phi(.)$ in (2.11) provides the improvement from $O(n^{-1/2})$ to $O(n^{-3/2})$. Approximation (2.11) can be unstable near $\theta = \hat{\theta}$ and some care must be taken if the distribution function is to be computed over its entire domain. Confidence intervals calculated from (2.11) are highly accurate in possibly very small samples and they typically have much better coverage than those based on the asymptotic distribution of $\hat{\theta}$. For model (2.3) we obtain $r = sgn(\hat{\theta} - \theta) \left[\left\{ \alpha(\theta) - \alpha(\hat{\theta}) \right\} \sum_{i=1}^{n} d(y_i) + n \log \left\{ \frac{\xi(\theta)}{\xi(\hat{\theta})} \right\} \right]$, $u = \{\alpha(\theta) - \alpha(\hat{\theta})\} \sum_{i=1}^{n} d'(y_i) k(\hat{\theta})^{-1} j(\hat{\theta})^{1/2}$, $k(\theta) = -\alpha'(\theta) \sum_{i=1}^{n} d'(y_i)$, and $j(\theta) = n\alpha'(\theta)\beta'(\theta)$, where $d'(y) = \frac{\partial d(y)}{\partial y}$.

3 Special cases

In this section, we use equations (2.6) and (2.7) to obtain approximate expressions for the skewness $\gamma_1(\hat{\theta})$ and kurtosis $\gamma_2(\hat{\theta})$ of the distribution of the MLE $\hat{\theta}$ for a number of important distributions that belong to the one-parameter exponential family (2.3). Twelve special distributions are considered and we give closed-form expressions for the standardized third and fourth cumulants of the MLE $\hat{\theta}$. The MAPLE code used to perform the algebraic calculations is available on a web page (www.de.ufpe.br/∼cysneiros/) so it can be download if anyone wants it. Distributions (i) through (iii) involve discrete random variables whereas continuous random variables are considered in cases (iv) through (xii). The special cases listed below have a wide range of practical applications in various fields such as engineering, biology, medicine, economics, among others (Johnson, Kotz and Balakrishnan, 1994, 1995; Johnson, Kotz and Kemp, 1992). We considered below some distributions whose values of $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ do not depend on θ and some distributions for which these values are very complicated functions of θ. The formulae derived which yield constants are compared in a chart given in Figure 1. For those distributions whose values of $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ are complicated functions of θ, we show graphically the dependence of these indices on θ in Section 4.

(i) Truncated Poisson ($\theta > 0$, $y = 1, 2, \ldots$): $\alpha(\theta) = -\log\theta$, $\zeta(\theta) = e^{\theta}(1 - e^{-\theta})$, $d(y) = y$, $v(y) = -\log(y!)$, $\hat{\theta}$ is obtained as solution of the equation $1/\{\hat{\theta}(1 - e^{-\hat{\theta}})\} = 1/\overline{y}$,

$$\begin{aligned}
\gamma_1(\hat{\theta}) &= \frac{(-1 + e^{-\theta})\{(2\theta^2 - 3\theta + 2)e^{-\theta} + (2\theta^2 + 3\theta - 1)e^{-2\theta} - 1\}}{\sqrt{-n\,(e^{-\theta}\,\theta - 1 + e^{-\theta})^3\theta\,(-1 + e^{-\theta})^2}}, \\
\gamma_2(\hat{\theta}) &= -\{1 + e^{-\theta}(-4 + 28\theta - 42\theta^2 + 9\theta^3) + e^{-2\theta}(6 - 84\theta + 121\theta^2 \\
&\quad -39\theta^3 + 18\theta^4) + e^{-3\theta}(-4 + 84\theta - 116\theta^2 - 27\theta^3 + 18\theta^4) \\
&\quad + e^{-4\theta}(1 - 28\theta + 37\theta^2 + 57\theta^3 + 18\theta^4)\}\Big/\{\theta n(e^{-\theta}\theta - 1 + e^{-\theta})^3\}.
\end{aligned}$$

(ii) Logarithmic series ($0 < \theta < 1$, $y = 1, 2, \ldots$): $\alpha(\theta) = -\log\theta$, $\zeta(\theta) = -\log(1-\theta)$, $d(y) = y$, $v(y) = -\log y$, $\hat{\theta}$ is obtained as solution of the equation $\overline{y} = -\hat{\theta}/\{(1 - \hat{\theta})\log(1 - \hat{\theta})\}$,

$$\begin{aligned}
\gamma_1(\hat{\theta}) &= -\{\log(1-\theta)\,(-1+\theta)\,[3\,\theta^2\,\log(1-\theta) + 5\,\theta\,\log^2(1-\theta) \\
&\quad + 4\,\theta^2 + 3\,\theta\,\log(1-\theta) - \log^2(1-\theta)]\} \\
&\quad \Big/\{-n\,[\theta + \log^3(1-\theta)]\theta\,(-1+\theta)^2\,\log^2(1-\theta)\}^{1/2}, \\
\gamma_2(\hat{\theta}) &= -\{\log^4(1-\theta) + 232\,\log^2(1-\theta)\,\theta^3 + 116\,\log^3(1-\theta)\,\theta^2 \\
&\quad + 78\,\theta^3\,\log(1-\theta) - 5\,\log^2(1-\theta)\,\theta^2 - 28\,\theta\,\log^3(1-\theta) \\
&\quad + 85\,\theta^2\,\log^4(1-\theta) + 84\,\theta^4\,\log(1-\theta) - 32\,\log^4(1-\theta)\,\theta \\
&\quad + 25\,\theta^4\,\log^2(1-\theta) + 83\,\theta^3\,\log^3(1-\theta) + 54\,\theta^4\} \\
&\quad \Big/\{\theta\,n\,[\theta + \log(1-\theta)]^3\}.
\end{aligned}$$

(iii) Zeta ($\theta > 0$, $y = 1, 2, \ldots$): $\alpha(\theta) = \theta + 1$, $\zeta(\theta) = Zeta(\theta + 1)$, $d(y) = \log y$, $v(y) = 0$, where $\zeta(\theta)$ is the Riemann zeta function, i.e., $\zeta(\theta) = \sum_{i=1}^{\infty} i^{-(\theta+1)}$ (see, e.g., Patterson, 1988) and $\mathrm{g} = \mathrm{g}(\theta) = d\log Zeta(\theta + 1)/d\theta$, $\hat{\theta}$ is obtained as solution of the equation $\mathrm{g}(\hat{\theta}) = -n^{-1}\Sigma_{i=1}^{n}\log y_i$,

$$\gamma_1(\hat{\theta}) = \frac{-2\mathrm{g}''}{\sqrt{n\mathrm{g}'^3}}, \quad \gamma_2(\hat{\theta}) = \frac{-9(\mathrm{g}'''\ \mathrm{g}' - 3\,\mathrm{g}''^2)}{n\,\mathrm{g}'^3}.$$

(iv) Gamma ($k>0$, $\theta>0$, $y>0$):

(a) k known: $\alpha(\theta)=\theta^{-1}$, $\zeta(\theta)=\theta^k$, $d(y)=ky$, $v(y)=(k-1)\log y-\log\{\Gamma(k)\}$, where $\Gamma(\cdot)$ is the gamma function , $\hat\theta=\overline{y}$,

$$\gamma_1(\hat\theta)=\frac{2}{\sqrt{kn}},\quad \gamma_2(\hat\theta)=\frac{6}{n\,k}.$$

(b) θ known: $\alpha(k)=1-k$, $\zeta(k)=\theta^{-k}\Gamma(k)$, $d(y)=\log y$, $v(y)=-\theta y$, $\hat k$ is obtained as solution of the equation $\psi(\hat k)=n^{-1}\log\left(\theta^n/\prod_{i=1}^n y_i\right)$, where $\psi(\cdot)$ is the digamma function,

$$\gamma_1(\hat k)=\frac{-2\psi''(k)}{\sqrt{n\,\psi'^3(k)}},\quad \gamma_2(\hat k)=\frac{9\{3\,\psi''^2(k)-\psi'''(k)\,\psi'(k)\}}{n\,\psi'^3(k)}.$$

(v) Rayleigh ($\theta>0$, $y>0$): $\alpha(\theta)=\theta^{-2}$, $\zeta(\theta)=\theta^2$, $d(y)=y^2$, $v(y)=\log(2y)$,

$\hat\theta=\left(n^{-1}\sum_{i=1}^n y_i^2\right)^{1/2}$,

$$\gamma_1(\hat\theta)=\frac{1}{2\sqrt{n}},\quad \gamma_2(\hat\theta)=-\frac{3}{4n}.$$

(vi) Extreme value ($-\infty<\theta<\infty$, $\phi>0$, ϕ known, $-\infty<y<\infty$): $\alpha(\theta)=\exp(\theta/\phi)$, $\zeta(\theta)=\phi\exp(-\theta/\phi)$, $d(y)=\exp(-y/\phi)$, $v(y)=-y/\theta$, $\hat\theta=-\phi\log\left\{n^{-1}\sum_{i=1}^n\exp(-y_i/\phi)\right\}$,

$$\gamma_1(\hat\theta)=\frac{1}{\sqrt{n}},\quad \gamma_2(\hat\theta)=\frac{5}{n}.$$

(vii) Lognormal ($\theta>0$, $-\infty<\mu<\infty$, μ known, $y>0$): $\alpha(\theta)=\theta^{-2}$, $\zeta(\theta)=\theta$, $d(y)=(\log y-\mu)^2/2$, $v(y)=-\log y-\{\log(2\pi)\}/2$, $\hat\theta=\left[n^{-1}\sum_{i=1}^n(\log y_i-\mu)^2\right]^{1/2}$,

$$\gamma_1(\hat\theta)=\frac{1}{\sqrt{2n}},\quad \gamma_2(\hat\theta)=\frac{-3}{2n}.$$

(viii) Normal ($\theta>0$, $-\infty<\mu<\infty$, $-\infty<y<\infty$):

(a) μ known: $\alpha(\theta)=(2\theta)^{-1}$, $\zeta(\theta)=\theta^{1/2}$, $d(y)=(y-\mu)^2$, $v(y)=-\{\log(2\pi)\}/2$, $\hat\theta=n^{-1}\sum_{i=1}^n(y_i-\mu)^2$,

$$\gamma_1(\hat\theta)=\frac{2\sqrt{2}}{\sqrt{n}},\quad \gamma_2(\hat\theta)=\frac{12}{n}.$$

(b) θ known: $\alpha(\mu) = -\mu/\theta$, $\zeta(\mu) = \exp\{\mu^2/(2\theta)\}$, $d(y) = y$, $v(y) = -\{y^2 + \log(2\pi\theta)\}/2$, $\hat{\mu} = \overline{y}$,

$$\gamma_1(\hat{\mu}) = 0, \quad \gamma_2(\hat{\mu}) = 0.$$

(ix) Inverse Gaussian ($\theta > 0$, $\mu > 0$, $y > 0$):

(a) μ known: $\alpha(\theta) = \theta^{-1}$, $\zeta(\theta) = \theta^{1/2}$, $d(y) = (y-\mu)^2/(2\mu^2 y)$, $v(y) = -\{\log(2\pi\, y^3)\}/2$, $\hat{\theta} = n\left[\sum_{i=1}^n \left\{(y_i-\mu)^2 \,/\, (y_i\mu^2)\right\}\right]$,

$$\gamma_1(\hat{\theta}) = \frac{2\sqrt{2}}{\sqrt{n}}, \quad \gamma_2(\hat{\theta}) = \frac{12}{n}.$$

(b) θ known: $\alpha(\mu) = \theta(2\mu^2)^{-1}$, $\zeta(\mu) = \exp(-\theta/\mu)$, $d(y) = y$, $v(y) = -\theta(2y)^{-1} + [\log\{\theta/(2\pi\, y^3)\}]/2$, $\hat{\mu} = \overline{y}$,

$$\gamma_1(\hat{\mu}) = \frac{3\sqrt{\mu}}{\sqrt{n\theta}}, \quad \gamma_2(\hat{\mu}) = \frac{15\mu}{n\,\theta}.$$

(x) McCullagh ($\theta > -1/2$, $-1 \leq \mu \leq 1$, μ known, $0 < y < 1$): $\alpha(\theta) = -\theta$, $\zeta(\theta) = 4^{-\theta}B(\theta{+}1/2,\, 1/2)$, $d(y) = \log[y(1-y)/\{(1+\mu)^2 - 4\mu\, y\}]$, $v(y) = -[\log\{y(1-y)\}]/2$, where $B(\cdot,\cdot)$ is the beta function (see McCullagh, 1989), $\hat{\theta}$ is obtained as solution of the equation $\psi(\hat{\theta}+1/2) - \psi(\hat{\theta}+1) = \log 4 - n^{-1}\sum_{i=1}^n \log\left[\{y_i(1-y_i)\}/\left\{(1+\mu)^2 - 4\mu y_i\right\}\right]$,

$$\begin{aligned}
\gamma_1(\hat{\theta}) &= \frac{2\{\psi''(1+\theta) - \psi''(\theta+1/2)\}}{\sqrt{\{\psi'(\theta+1/2) - \psi'(1+\theta)\}^3\, n}},\\
\gamma_2(\hat{\theta}) &= 9\{-\psi'''(1+\theta)\,\psi'(\theta+1/2) + \psi'''(1+\theta)\,\psi'(1+\theta)\\
&\quad +\psi'''(\theta+1/2)\psi'(\theta+1/2) - \psi'''(\theta+1/2)\psi'(1+\theta)\\
&\quad -3\psi''^2(1+\theta) + 6\psi''(1+\theta)\,\psi''(\theta+1/2)\\
&\quad -3\psi''^2(\theta+1/2)\}\Big/ n\,\{\psi'(1+\theta) - \psi'(\theta+1/2)\}^3.
\end{aligned}$$

(xi) von Mises ($\theta > 0$, $0 < \mu < 2\pi$, μ known, $0 < y < 2\pi$): $\alpha(\theta) = -\theta$, $\zeta(\theta) = 2\pi\, I_0(\theta)$, $d(y) = \cos(y-\mu)$, $v(y) = 0$, where $I_v(\cdot)$ is the modified Bessel function of first kind and vth order, and $\mathrm{r} = \mathrm{r}(\theta) = I_0'(\theta)/I_0(\theta)$, $\hat{\theta} = \mathrm{r}^{-1}\left\{n^{-1}\sum_{i=1}^n \cos(y_i-\mu)\right\}$,

$$\gamma_1(\hat{\theta}) = \frac{-2\mathrm{r}''}{\sqrt{n\,\mathrm{r}'^3}}, \quad \gamma_2(\hat{\theta}) = \frac{-9(\mathrm{r}'''\,\mathrm{r}' - 3\,\mathrm{r}''^2)}{n\,\mathrm{r}'^3}.$$

(xii) Log gamma ($\theta > 0$, $-\infty < \mu < \infty, -\infty < y < \infty$):

a) μ known : $\alpha(\theta) = -\theta, \zeta(\theta) = \theta^{-1}\Gamma(\theta), d(y) = y - \mu - \exp(y - \mu), v(y) = 0$, where $\Gamma(.)$ is the gamma function, $\hat{\theta}$ is obtained as solution of the equation $\psi(\hat{\theta}) - \hat{\theta}^{-1} - \mu = \overline{y} - n^{-1}\sum_{i=1}^{n} e^{y_i - \mu}$,

$$\begin{aligned}
\gamma_1(\hat{\theta}) &= \frac{2\{2 - \psi''(\theta)\,\theta^3\}}{\sqrt{n\,\{1 + \psi'(\theta)\,\theta^2\}^3}}, \\
\gamma_2(\hat{\theta}) &= \frac{9\{-\psi'''(\theta)\,\theta^4 - \psi'''(\theta)\,\theta^6\,\psi'(\theta) + 6 - 6\,\psi'(\theta)\,\theta^2}{n\,\{1 + \psi'(\theta)\,\theta^2\}^3} \\
&\quad \frac{-12\,\psi''(\theta)\,\theta^3 + 3\,\psi''^2(\theta)\,\theta^6\}}{n\,\{1 + \psi'(\theta)\,\theta^2\}^3}.
\end{aligned}$$

b) θ known : $\alpha(\mu) = \exp(-\mu), \zeta(\mu) = \exp(\theta\mu), d(y) = \theta\exp(y), v(y) = \theta y + \log\theta - \log\{\Gamma(\theta)\}$, $\hat{\mu} = \log\left(n^{-1}\sum_{i=1}^{n} e^{y_i}\right)$,

$$\gamma_1(\hat{\mu}) = -\frac{1}{\sqrt{n\theta}}, \quad \gamma_2(\hat{\mu}) = \frac{5}{n\,\theta}.$$

For some of the special cases considered here, the asymptotic formulae for the skewness $\gamma_1(\hat{\theta})$ and kurtosis $\gamma_2(\hat{\theta})$ of the standardized distribution of $\hat{\theta}$ are very simple and for some of them the formulae do not even depend on θ. For these cases, it is easier to verify the degree of departure from normality of the standardized distribution of $\hat{\theta}$. In some cases, however, the asymptotic formulae for $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ are very complicated functions of θ and can vary substantially depending on the true value of the parameter θ. We give in Table 1 the values of $\sqrt{n}\gamma_1(\hat{\theta})$ and $n\gamma_2(\hat{\theta})$, when $\theta = 1, 2, 5, 10$ and 15, for the following distributions: truncated Poisson, von Mises, log gamma and gamma (θ here represents k in case iv-b). For the von Mises, log gamma and gamma distributions, there is the greatest departure from normality for both n and θ small. In Table 2 the values of $\sqrt{n}\gamma_1(\hat{\theta})$ and $n\gamma_2(\hat{\theta})$ are given when $\theta = 0.1, 0.2, 0.5, 0.7, 0.9$ for the logarithmic series and zeta distributions and when $\theta = -0.2, -0.1, 0, 0.1, 0.2$ for the McCullagh distribution. For the logarithmic series distribution, $\gamma_1(\hat{\theta})$ ($\gamma_2(\hat{\theta})$) changes from positive (negative) to negative (positive) values when θ varies from 0 to 1. For the logarithmic series, zeta and McCullagh distributions, the kurtosis of $\hat{\theta}$ is significantly large.

Table 1
Skewness ($\sqrt{n}\gamma_1$) and kurtosis ($n\gamma_2$) of MLEs for some distributions

	truncated Poisson		von Mises		log gamma		gamma	
θ	$\sqrt{n}\gamma_1$	$n\gamma_2$	$\sqrt{n}\gamma_1$	$n\gamma_2$	$\sqrt{n}\gamma_1$	$n\gamma_2$	$\sqrt{n}\gamma_1$	$n\gamma_2$
1	0.6683	−0.9283	2.1268	31.2961	2.0477	12.2297	2.2791	13.4613
2	0.3407	0.5223	4.1126	76.4142	1.5452	6.3530	1.5605	5.7494
5	0.3558	0.6409	5.5847	113.8179	0.9700	2.2618	0.9372	1.9914
10	0.3140	0.1233	5.6833	108.5828	0.6678	1.0289	0.6478	0.9465
15	0.2582	0.0671	5.6664	106.8515	0.5376	0.6588	0.5249	0.6204

Table 2
Skewness ($\sqrt{n}\gamma_1$) and kurtosis ($n\gamma_2$) of MLEs for some distributions

	logarithmic series		zeta			McCullagh	
θ	$\sqrt{n}\gamma_1$	$n\gamma_2$	$\sqrt{n}\gamma_1$	$n\gamma_2$	θ	$\sqrt{n}\gamma_1$	$n\gamma_2$
0.1	2.5734	−22.7753	4.0105	54.3750	−0.2	4.5171	71.5131
0.2	0.4213	−21.3393	4.0394	55.4020	−0.1	4.6894	77.3013
0.5	−2.6839	14.7915	4.2094	61.4948	0	4.8347	82.1760
0.7	−4.2951	54.0575	4.3761	67.6370	0.1	4.9557	86.2129
0.9	−6.3359	127.0926	5.4173	101.1454	0.2	5.0557	89.5333

It is possible to check by direct calculation that the formulae for $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ are correct for some distributions which have closed-form $MLEs$, namely: gamma (case a), extreme value, normal (cases viii-a and b) and inverse Gaussian (cases ix-a and b). The normal $N(\mu, \theta)$ distribution with known variance θ (case viii-b) is the only case for which the moment ratios vanish, since the MLE $\hat{\mu}$ has an exact normal distribution. In a few cases, the distribution of a certain multiple of the MLE $\hat{\theta}$ has the same distribution proposed for the data. It is then possible to obtain by direct calculation the moment ratios of $\hat{\theta}$ from the corresponding moment ratios of the data. In this situation we have the following case: inverse Gaussian with known precision parameter θ (case ix-b) for which the MLE $\hat{\mu}$ of the mean μ has an inverse Gaussian $IG(\mu, n\theta)$ distribution with parameters μ and $n\theta$. The invariance property also holds for other distributions, not included in the above examples, such as: binomial $B(m, \theta)$ for which $nm\hat{\theta}$ has a binomial $B(nm, \theta)$ distribution and Poisson $P(\theta)$ where $n\hat{\theta}$ has a Poisson $P(n\theta)$ distribution.

For the gamma with known index k and unknown mean (case iv-a), extreme value, normal with known mean μ and unknown variance (case viii-a) and inverse Gaussian with known mean and unknown dispersion parameter (case ix-a) distributions, the $MLE\ \hat{\theta}$ is proportional to a chi-squared random variable and we can easily obtain $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ from the corresponding moment ratios $\gamma_1(\chi_r^2) = \frac{2\sqrt{2}}{\sqrt{r}}$ and $\gamma_2(\chi_r^2) = \frac{12}{\sqrt{r}}$ of a χ_r^2 distribution.

The division of the $(\sqrt{n}\gamma_1, n\gamma_2)$ plane among the various distributions for which $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ do not depend on the value of θ is shown in Figure 1. Good agreement with normal asymptotic theory happens when the point (γ_1, γ_2) is close to the origin. Departure of $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ from the normal value of zero is an indication of nonnormality in the distribution of the $MLE\ \hat{\theta}$.

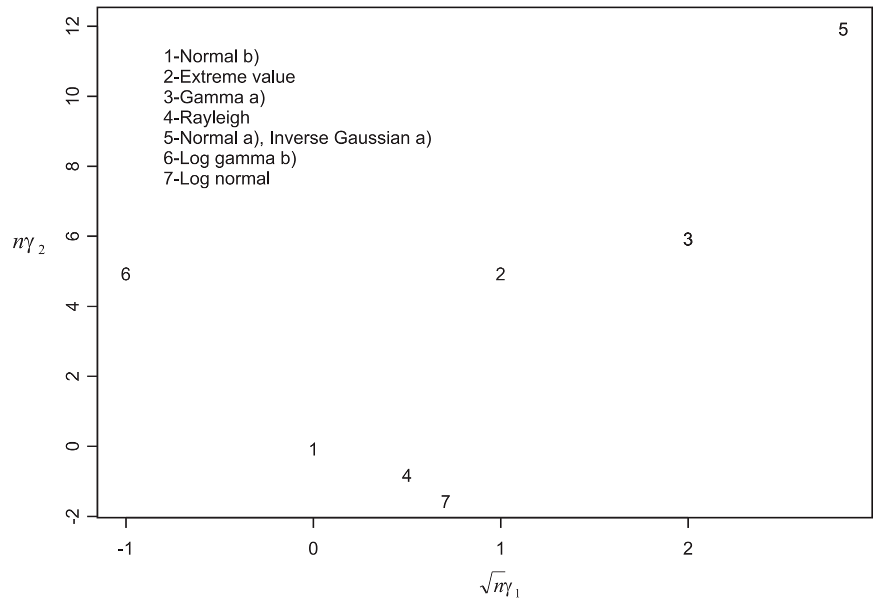

Figure 1
A chart relating the distribution of $\hat{\theta}$ *to the values of* $\sqrt{n}\gamma_1(\hat{\theta})$ *and* $n\gamma_2(\hat{\theta})$

4 Graphical analysis

It is clear from the results in Section 3 that for several distributions $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ are not constant but functions of θ. We now discuss the dependence of these indices on θ by plotting $\sqrt{n}\gamma_1(\hat{\theta})$ and $n\gamma_2(\hat{\theta})$ versus θ for the following distributions: truncated Poisson, logarithmic series, von Mises, McCullagh and zeta. These plots are given in Figures 2 through 6, respectively, to shed some light on how different values of θ affect the behavior of the standardized cumulants of $\hat{\theta}$ for these distributions and hence the normal approximation for $\hat{\theta}$. Figure 2 shows that $\sqrt{n}\gamma_1(\hat{\theta})$ and $n\gamma_2(\hat{\theta})$ become very small for large values of θ for the truncated Poisson distribution in agreement with the fact that the normal approximation for $\hat{\theta}$ is better when θ is large. The distribution of $\hat{\theta}$ is always positively skewed and is platykurtic for (approximately) $0.2567 < \theta < 1.5123$. $\sqrt{n}\gamma_1(\hat{\theta})$ and $n\gamma_2(\hat{\theta})$ increase quite fast when $\theta \to 0$ showing that the normal approximation for the distribution of $\hat{\theta}$ breaks down for very small θ. In the case of the logarithmic series distribution (Figure 3), the absolute skewness and kurtosis of $\hat{\theta}$ become very large for values of θ close to zero or one and, as might be expected, very marked departures from normality would occur in these cases for most sample sizes. In particular, $\gamma_1(\hat{\theta})$ vanishes for θ around 0.2293 and $\gamma_2(\hat{\theta})$ vanishes for values of θ around 0.0407 and 0.3994. The distribution of $\hat{\theta}$ is positively (negatively) skewed for $\theta < 0.2293$ ($\theta > 0.2293$) and platykurtic for $0.0407 < \theta < 0.3994$.

The case of the von Mises distribution (Figure 4) shows a rather interesting behavior. Both curves of $\sqrt{n}\gamma_1(\hat{\theta})$ and $n\gamma_2(\hat{\theta})$ increase quite fast for small θ and after reaching peaks at approximately 4.4586 and 3.7466, respectively, decrease continuously approaching asymptotic values around 5.67 and 106.85 as θ increases. This is shown via computational methods. For $\theta > 15$, $n\gamma_2(\hat{\theta})$ has an instable behavior. The distribution of $\hat{\theta}$ is always positively skewed and leptokurtic. Figure 5 displays $\sqrt{n}\gamma_1(\hat{\theta})$ and $n\gamma_2(\hat{\theta})$ versus θ for the McCullagh distribution. Both moment ratios $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ are positive and increase from 4.002 and 54.067 to 5.657 and 108, respectively, as θ increases from $-1/2$. In fact, $\lim_{\theta\to\infty}\gamma_1(\hat{\theta}) = \frac{4\sqrt{2}}{\sqrt{n}}$ and $\lim_{\theta\to\infty}\gamma_2(\hat{\theta}) = 108/n$ (see Section 5) showing that both moment ratios approach asymptotic levels as θ increases to ∞. Finally, consider the zeta distribution (Figure 6). For small values of θ, it can be shown (see also Section 5) that $\lim_{\theta\to 0}\gamma_1(\hat{\theta}) = \frac{4}{\sqrt{n}}$ and $\lim_{\theta\to 0}\gamma_2(\hat{\theta}) = \frac{54}{n}$. However, the moment ratios $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ diverge to ∞ as θ becomes large. Figures 5 and 6 show that the normal approximation for the distribution of $\hat{\theta}$ deteriorates when θ increases for the McCullagh and zeta distributions. In fact, because of the large kurtosis for any θ, the distribution of $\hat{\theta}$ is far from the normal distribution for these cases except if n is very large.

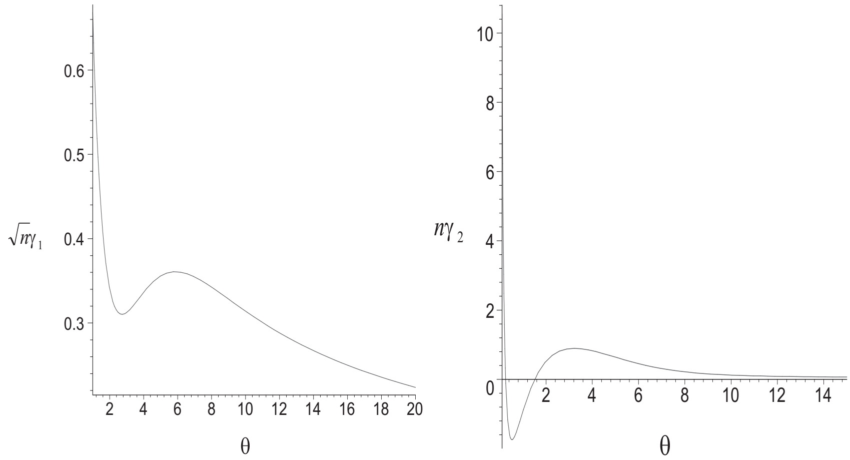

Figure 2
Truncated Poisson Distribution

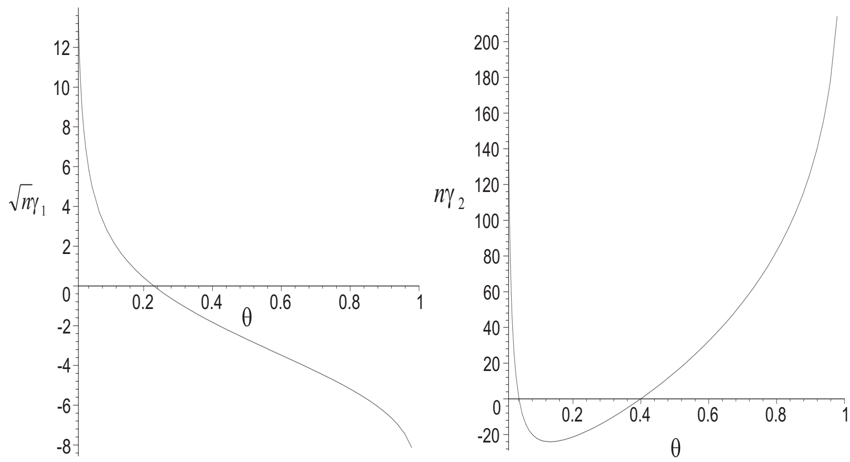

Figure 3
Logarithmic Series Distribution

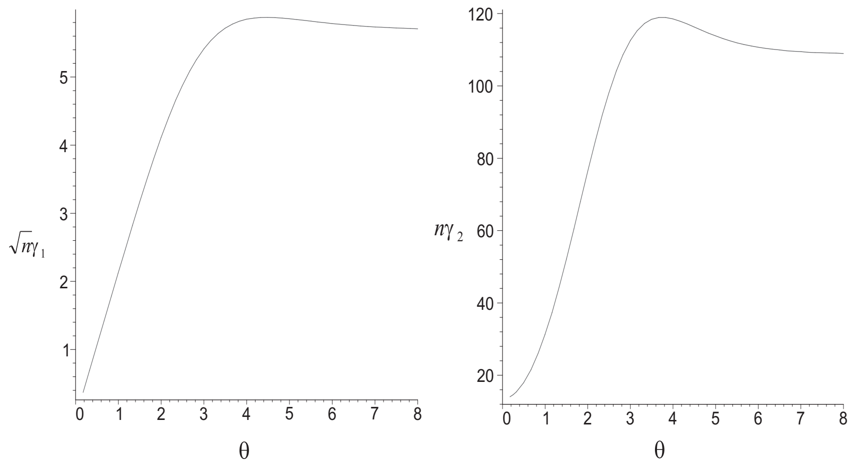

Figure 4
von Mises Distribution

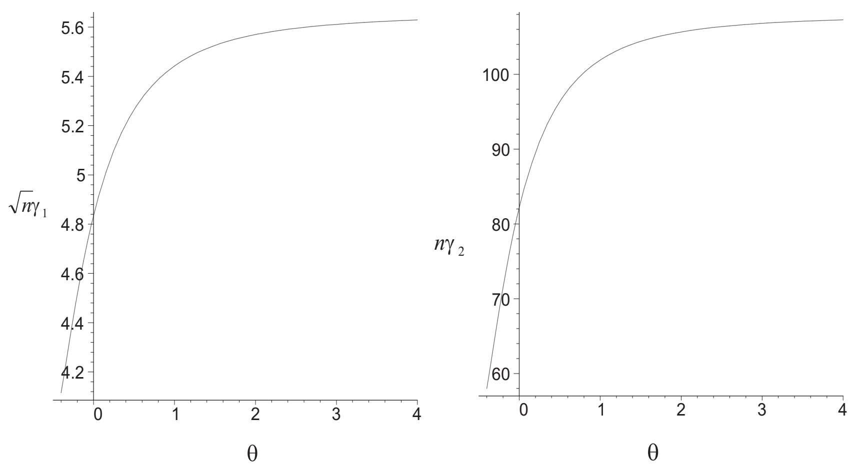

Figure 5
McCullagh Distribution

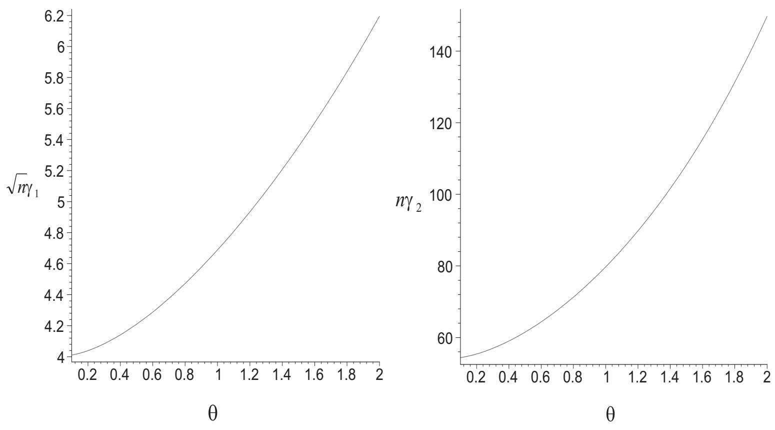

Figure 6
Zeta Distribution

5 Concluding remarks

We derive simple formulae for the asymptotic skewness $\gamma_1(\hat{\theta})$ and kurtosis $\gamma_2(\hat{\theta})$ of the maximum likelihood estimator $\hat{\theta}$ in the one-parameter exponential model. We apply the formulae to a number of distributions, thus giving closed-form expressions for these moment ratios for the purposes of computing first-order confidence intervals for θ from the normal distribution of $\hat{\theta}$. We also review two second-order methods to obtain improved confidence intervals for θ when the normal approximation for the distribution of $\hat{\theta}$ is probably bad. A graphical analysis that shows the dependence of these moment ratios on θ is performed for some distributions. This analysis is useful to examine to which intervals of the parameter space correspond smaller values of $\mid \gamma_1(\hat{\theta}) \mid$ and $\mid \gamma_2(\hat{\theta}) \mid$ in order to guarantee approximate normality for the distribution of $\hat{\theta}$. The distribution of the MLE for all of the given distributions (other than the normal) will have some degree of departure from normality. For some distributions, it may be possible to specify intervals of θ for which the distribution of $\hat{\theta}$ is more likely to departure from normality.

The knowledge of $\gamma_1(\hat{\theta})$ and $\gamma_2(\hat{\theta})$ can be used to reduce skewness and kurtosis of the distribution of $\hat{\theta}$ to insignificance by making the sample size n large enough. For example, if these moment ratios are less than $1/10$, then one might consider that approximate normality of $\hat{\theta}$ has been achieved. From this point of view, it is very simple to recommend the minimum value of n if both moment ratios do not depend on θ. If this is not the case, the critical sample size needed to make the asymptotic skewness and kurtosis less than 1/10 can be evaluated if we have any prior idea of the neighbourhood where the true parameter lies. Two illustrative examples are now given. A critical sample size of 143 would be needed to control these moment ratios for the von Mises distribution when $\theta < 0.20$. Far larger samples may be required whenever θ exceeds 0.20. The smallest sample size needed for the truncated Poisson distribution when $\theta > 0.2$ is 719, although we require far smaller sample size when θ becomes large.

Acknowledgments

We thank two anonymous referees for useful suggestions that improved the paper. We also gratefully acknowledge financial support from the Brazilian agency CNPq.

(Received December, 2001. Accepted December, 2001.)

References

Abell, M. L. and Braselton, J. P. (1994). *The Maple V Handbook.* New York: AP Professional.

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory.* New York: Wiley.

Barndorff-Nielsen, O. E. (1990). Approximate interval probabilities. *J. R. Statist. Soc. B*, **52**, 485–496.

Barndoff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics.* London: Chapman and Hall.

Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics.* Oakland: Holden Day.

Cordeiro, G. M., Ferrari, S. L. P., Botter, D. A, and Cribari-Neto, F. (1999). Modified maximum likelihood estimation in one-parameter exponential family models. *Comm. Statist. Theor. Meth.*, **28**, 157–178.

Cribari-Neto, F., Botter, D. A., Cordeiro, G. M. and Ferrari, S. L. P. (1998). Bias reduction in one-parameter exponential family models. *Comm. Statist. Simul. Comput.*, **27**, 761–768.

Ferrari, S. L. P., Botter, D. A., Cordeiro, G. M. and Cribari-Neto, F. (1996). Second and third-order bias reduction for one-parameter family models. *Statist. Prob. Lett.*, **30**, 339–345.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27–38.

Fraser, D. A. S. (1990). Tail probabilities from observed likelihoods. *Biometrika*, **77**, 65–76.

Hill, G. W. and Davis, A. W. (1968). Generalized asymptotic expansions of Cornish-Fisher type. *Ann. Math. Statist.*, **39**, 1264–1273.

Johnson, N. L, Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions.* v.1. New York: Wiley.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions.* v.2. New York: Wiley.

Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Univariate Discrete Distributions.* New York: Wiley.

Lawley, D. N. (1956). A general method for approximating to the distribution of the likelihood ratio criteria. *Biometrika*, **71**, 233–244.

McCullagh, P. (1989). Some statistical properties of a family of continuous univariate distributions. *J. Am. Statist. Assoc.*, **84**, 125–141.

Patterson, S. J. (1988). *An Introduction to the Theory of the Riemann Zeta-Function.* New York: Cambridge University Press.

Peres, H. W. and Iqbal, M. (1985). Asymptotic expansions for confidence limits in the presence of nuisance parameters with applications. *J. R. Statist. Soc. B*, **47**, 547–554.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications.* New York: Wiley.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

Shenton, L. R. and Bowman, K. O. (1963). Higher moments of a maximum likelihood estimate. *J. R. Statist. Soc. B*, **25**, 305–317.

Shenton, L. R. and Bowman, K. O. (1977). *Maximum Likelihood Estimation in Small Samples.* London: Griffin.

Van Zwet, W. R. (1964). *Convex Transformations of Random Variables.* Amsterdam: Mathematisch Centrum.

Wolfram, S. (1996). *The Mathematica Book.* 3.ed. New York: Addison Wesley.