# Second and Third Order Bias Reduction for One–Parameter Family Models

Silvia L.P. Ferrari

*Dept. de Estatística, Univ. de São Paulo, Caixa Postal 66281, Ag. Cidade de São Paulo, São Paulo SP, 05389-970, Brazil*

Denise A. Botter

*Dept. de Estatística, Univ. de São Paulo, Caixa Postal 66281, Ag. Cidade de São Paulo, São Paulo SP, 05389-970, Brazil*

Gauss M. Cordeiro

*Dept. de Estatística, Universidade Federal de Pernambuco, Cidade Universitária, Recife PE, 50740-540, Brazil*

Francisco Cribari–Neto

*Department of Economics, Southern Illinois University, Carbondale IL, 62901-4515, USA*

**Abstract.** In this paper we derive second and third order bias-corrected maximum likelihood estimates in general uniparametric models. We compare the corrected estimates and the usual maximum likelihood estimate in terms of their mean squared errors. We also obtain closed-form expressions for bias-corrected estimates in one-parameter exponential family models. Our results cover many important and commonly used distributions.

*Some key words*: Asymptotic expansion; bias correction; exponential family; maximum likelihood estimate.

## 1. Introduction

Consider the probability or density function

$$\pi(y;\theta) = \exp\{t(y;\theta)\}, \tag{1}$$

where $\theta$ is a scalar parameter, and let $\hat{\theta}$ denote the maximum likelihood estimate of $\theta$. Expansions for the bias and variance of $\hat{\theta}$ to order $n^{-2}$, where $n$ is the sample size, can be written as

$$B(\theta) = \frac{B_1(\theta)}{n} + \frac{B_2(\theta)}{n^2} \quad \text{and} \quad V(\theta) = \frac{V_1(\theta)}{n} + \frac{V_2(\theta)}{n^2},$$

respectively. We define three bias-corrected estimates, namely:

$$\hat{\theta}_1 = \hat{\theta} - \frac{B_1(\hat{\theta})}{n}, \tag{2}$$

$$\hat{\theta}_2 = \hat{\theta} - \frac{B_1(\hat{\theta})}{n} - \frac{B_2(\hat{\theta})}{n^2} \tag{3}$$

and

$$\tilde{\theta}_2 = \hat{\theta} - \frac{B_1(\hat{\theta})}{n} - \frac{B_2^*(\hat{\theta})}{n^2}, \tag{4}$$

where

$$B_2^*(\theta) = B_2(\theta) - B_1(\theta)B_1'(\theta) - \frac{1}{2}B_1''(\theta)V_1(\theta),$$

primes denoting derivatives with respect to $\theta$. These three modified estimates are bias-free to order $n^{-1}$, but only $\tilde{\theta}_2$ has no bias to order $n^{-2}$. Although $\hat{\theta}_2$ seems to be a third order bias-corrected estimate, its third order bias is in general nonnull, as we shall prove later. It will also be proved that the modified estimates in (2)-(4) have the same variance and mean squared error to order $n^{-2}$. Closed-form formulas for $B_1(\theta)$, $B_2(\theta)$, $B_2^*(\theta)$, $V_1(\theta)$ and $V_2(\theta)$ for general uniparametric models are given in the next section together with a comparison of the variances and mean squared errors of the maximum likelihood estimate and its bias-corrected versions. Section 3 presents closed-form expressions for the bias and variance coefficients for one-parameter exponential family models. It also gives bias-corrected estimates for natural exponential family models.

## 2. Derivation of the second and third order biases

Let $y_1, \ldots, y_n$ be a set of $n$ independent and identically distributed random variables with regular density (or probability) function given by (1). The asymptotic expansions for the bias and variance of the maximum likelihood estimate of $\theta$ up to order $n^{-2}$ were written in the previous section as $B(\theta) = n^{-1}B_1(\theta) + n^{-2}B_2(\theta)$ and $V(\theta) = n^{-1}V_1(\theta) + n^{-2}V_2(\theta)$, respectively. In order to obtain closed-form expressions for the coefficients in these expansions, we need to introduce some notation for cumulants of log-likelihood derivatives with respect to $\theta$ for a single observation. Let

$$\begin{aligned}
&\kappa_{\theta\theta} = \mathrm{E}\{t''(y;\theta)\}, \quad \kappa_{\theta,\theta} = -\kappa_{\theta\theta} = \mathrm{E}[\{t'(y;\theta)\}^2], \quad \kappa_{\theta\theta\theta} = \mathrm{E}\{t'''(y;\theta)\},\\
&\kappa_{\theta\theta\theta\theta} = \mathrm{E}\{t^{iv}(y;\theta)\}, \quad \kappa_{\theta\theta\theta\theta\theta} = \mathrm{E}\{t^{v}(y;\theta)\}, \quad \kappa_{\theta\theta,\theta} = \mathrm{E}\{t''(y;\theta)t'(y;\theta)\},\\
&\kappa_{\theta\theta,\theta\theta} = \mathrm{E}[\{t''(y;\theta)\}^2] - \kappa_{\theta\theta}^2, \quad \kappa_{\theta,\theta\theta\theta} = \mathrm{E}\{t'(y;\theta)t'''(y;\theta)\},\\
&\kappa_{\theta,\theta\theta\theta\theta} = \mathrm{E}\{t'(y;\theta)t^{iv}(y;\theta)\}, \quad \kappa_{\theta,\theta\theta,\theta\theta} = \mathrm{E}[t'(y;\theta)\{t''(y;\theta)\}^2] - 2\kappa_{\theta,\theta\theta}\kappa_{\theta\theta},\\
&\kappa_{\theta\theta,\theta\theta\theta} = \mathrm{E}\{t''(y;\theta)t'''(y;\theta)\} - \kappa_{\theta\theta}\kappa_{\theta\theta\theta}, \quad \kappa_{\theta\theta}^{(\theta)} = \frac{\mathrm{d}\kappa_{\theta\theta}}{\mathrm{d}\theta}, \quad \kappa_{\theta\theta}^{(\theta\theta)} = \frac{\mathrm{d}^2\kappa_{\theta\theta}}{\mathrm{d}\theta^2}, \quad \text{etc.}
\end{aligned}$$

Using the results in Shenton and Bowman (1977, pp.44-47) together with the notation for cumulants introduced above and the regularity equations (Lawley, 1956) $\kappa_{\theta,\theta} = -\kappa_{\theta\theta}$, $\kappa_{\theta,\theta,\theta} =$

$2\kappa_{\theta\theta\theta} - 3\kappa_{\theta\theta}^{(\theta)}$, $\kappa_{\theta,\theta\theta} = \kappa_{\theta\theta}^{(\theta)} - \kappa_{\theta\theta\theta}$, $\kappa_{\theta,\theta,\theta\theta} = \kappa_{\theta\theta\theta\theta} - 2\kappa_{\theta\theta\theta}^{(\theta)} + \kappa_{\theta\theta}^{(\theta\theta)} - \kappa_{\theta\theta,\theta\theta}$, we get, after some algebra,

$$
\begin{aligned}
B_1(\theta) &= -\frac{1}{2\kappa_{\theta\theta}^2}(\kappa_{\theta\theta\theta} - 2\kappa_{\theta\theta}^{(\theta)}),\\
B_2(\theta) &= -\frac{1}{8\kappa_{\theta\theta}^3}(\kappa_{\theta\theta\theta\theta\theta} + 12\kappa_{\theta\theta,\theta\theta\theta} + 4\kappa_{\theta,\theta\theta\theta\theta} + 8\kappa_{\theta,\theta\theta,\theta\theta} + 4\kappa_{\theta,\theta,\theta\theta\theta}) + \frac{1}{12\kappa_{\theta\theta}^4}(13\kappa_{\theta\theta\theta\theta}\kappa_{\theta\theta\theta}\\
&\quad - 18\kappa_{\theta\theta,\theta\theta}\kappa_{\theta\theta\theta} - 18\kappa_{\theta\theta\theta\theta}\kappa_{\theta\theta}^{(\theta)} + 36\kappa_{\theta\theta,\theta\theta}\kappa_{\theta\theta}^{(\theta)} - 36\kappa_{\theta\theta\theta}\kappa_{\theta\theta\theta}^{(\theta)} + 36\kappa_{\theta\theta}^{(\theta)}\kappa_{\theta\theta\theta}^{(\theta)} + 18\kappa_{\theta\theta\theta}\kappa_{\theta\theta}^{(\theta\theta)})\\
&\quad - \frac{1}{8\kappa_{\theta\theta}^5}\{11\kappa_{\theta\theta\theta}^3 - 48\kappa_{\theta\theta\theta}^2\kappa_{\theta\theta}^{(\theta)} + 48\kappa_{\theta\theta\theta}(\kappa_{\theta\theta}^{(\theta)})^2\},\\
V_1(\theta) &= -\frac{1}{\kappa_{\theta\theta}},\\
V_2(\theta) &= -\frac{1}{\kappa_{\theta\theta}^3}(\kappa_{\theta\theta,\theta\theta} - \kappa_{\theta\theta\theta}^{(\theta)} + 2\kappa_{\theta\theta}^{(\theta\theta)}) + \frac{1}{2\kappa_{\theta\theta}^4}(10\kappa_{\theta\theta}^{(\theta)^2} - \kappa_{\theta\theta\theta}^2 - 4\kappa_{\theta\theta}^{(\theta)}\kappa_{\theta\theta\theta}).
\end{aligned}
$$

Bias-corrected estimates can be obtained from (2)-(4) and the equations above.

In what follows, we shall use some expansions to obtain the bias, variance and mean squared error of the estimates $\hat{\theta}_1$, $\hat{\theta}_2$ and $\tilde{\theta}_2$. Let $H(\cdot)$ be a function defined in the parameter space that can be expanded in a Taylor series. Expanding $H(\hat{\theta})$ around $\theta$ we obtain

$$H(\hat{\theta}) = H(\theta) + H'(\theta)(\hat{\theta} - \theta) + \frac{1}{2}H''(\theta)(\hat{\theta} - \theta)^2 + \dots \qquad (5)$$

Taking $H(\theta) = \theta - B_1(\theta)/n$ and using (5) we get, after some algebra, that $\mathrm{E}(\hat{\theta}_1 - \theta) = B_2^*(\theta)/n^2 + O(n^{-3})$. Similarly, taking $H(\theta) = \theta - B_1(\theta)/n - B_2(\theta)/n^2$, we obtain $\mathrm{E}(\hat{\theta}_2 - \theta) = \{B_2^*(\theta) - B_2(\theta)\}/n^2 + O(n^{-3})$. Thus, $\hat{\theta}_2$ is, like $\hat{\theta}_1$, an unbiased estimate to order $n^{-1}$. Next, letting $H(\theta) = \theta - B_1(\theta)/n - B_2^*(\theta)/n^2$, we obtain $\mathrm{E}(\tilde{\theta}_2 - \theta) = O(n^{-3})$. It then follows that only $\tilde{\theta}_2$ is unbiased to order $n^{-2}$. Although one would expect $\hat{\theta}_2$ to be a third order bias-corrected estimate of $\theta$, it was shown above that the correction term $B_1(\hat{\theta})/n + B_2(\hat{\theta})/n^2$ does not make the $n^{-2}$ term in the bias expansion vanish. This leads us to consider the correction term $B_1(\hat{\theta})/n + B_2^*(\hat{\theta})/n^2$, which, as shown, delivers a third order bias-corrected estimate.

By making use of similar developments, it is possible to show that, to order $n^{-2}$, the mean squared errors (MSE) of the three corrected estimates are the same and coincide with their respective variances. They are given by

$$\mathrm{MSE}(\hat{\theta}_1) = \mathrm{MSE}(\hat{\theta}_2) = \mathrm{MSE}(\tilde{\theta}_2) = \frac{V_1(\theta)}{n} + \frac{1}{n^2}\{V_2(\theta) - 2V_1(\theta)B_1'(\theta)\} + O(n^{-3}).$$

Therefore, from the three bias-corrected estimates, the estimate $\tilde{\theta}_2$ is to be preferred since they all have the same variance to order $n^{-2}$ and $\tilde{\theta}_2$ is the only one that is bias-free to this order.

Finally, it remains to compare $\hat{\theta}$ and $\tilde{\theta}_2$. The mean squared error of $\hat{\theta}$ is given by

$$\mathrm{MSE}(\hat{\theta}) = \frac{V_1(\theta)}{n} + \frac{1}{n^2}\{V_2(\theta) + B_1(\theta)^2\} + O(n^{-3}),$$

and hence

$$\mathrm{MSE}(\hat{\theta}) - \mathrm{MSE}(\tilde{\theta}_2) = \frac{1}{n^2}\{B_1(\theta)^2 + 2V_1(\theta)B_1'(\theta)\} + O(n^{-3}).$$

The choice between $\hat{\theta}$ and $\tilde{\theta}_2$ is to be based on the sign of $\Delta = \Delta(\theta) = B_1(\theta)^2 + 2V_1(\theta)B_1'(\theta)$. For example, when $B_1(\theta) = c$, a constant, we have that, to order $n^{-2}$, $\mathrm{MSE}(\hat{\theta}) \geq \mathrm{MSE}(\tilde{\theta}_2)$.

Oftentimes one is interested in the estimation of a function of $\theta$, say $\tau = G(\theta)$, which does not depend on $n$. One can reparameterize the model in terms of $\tau$ and then use the results above to obtain $\hat{\tau}_1$, $\hat{\tau}_2$ and $\tilde{\tau}_2$, that is, bias-corrected maximum likelihood estimates for $\tau$.

## 3. One-parameter exponential family

Let $y_1, \ldots, y_n$ be a set of $n$ independent and identically distributed random variables, and assume an exponential family form for $\pi(\cdot;\cdot)$: $\pi(y;\theta)$ is as defined in (1) with $t(y;\theta) = -\log\zeta(\theta) - \alpha(\theta)d(y) + v(y)$, that is,

$$\pi(y;\theta) = \frac{1}{\zeta(\theta)}\exp\{-\alpha(\theta)\, d(y) + v(y)\}, \tag{6}$$

where $\theta$ is a scalar parameter and $\zeta = \zeta(\cdot)$, $\alpha = \alpha(\cdot)$, $d(\cdot)$ and $v(\cdot)$ are known functions, and it is assumed that the support of $\pi(y;\theta)$ does not depend upon $\theta$. It is also assumed that $\alpha$ and $\zeta$ have continuous first five derivatives with respect to $\theta$, and that $\zeta$ is positive valued, $\mathrm{d}\alpha(\theta)/\mathrm{d}\theta$ and $\mathrm{d}\beta(\theta)/\mathrm{d}\theta$ being different from zero for all values of $\theta$ in the parameter space, where $\beta = \beta(\theta)$ is defined as $\beta = (\mathrm{d}\zeta(\theta)/\mathrm{d}\theta)(\zeta(\theta)\mathrm{d}\alpha(\theta)/\mathrm{d}\theta)^{-1}$. Many commonly used distributions in applied research are special cases of the family of distributions in (6). Also, this family of distributions enjoys important mathematical properties; see Bickel and Doksum (1977). We then have that $t'(y;\theta) = -\alpha'\{\beta + d(y)\}$, and since $\mathrm{E}\{t'(y;\theta)\} = 0$ it follows that $\mu = \mathrm{E}\{d(y)\} = -\beta$. The maximum likelihood estimate $\hat{\theta}$ comes from $n^{-1}\sum d(y_i) = \beta(\hat{\theta})$. This equation may not be easy to solve. It might even require the use of numerical methods.

It can also be shown that $t''(y;\theta) = -\alpha''\{\beta + d(y)\} - \alpha'\beta'$, $t'''(y;\theta) = -\alpha'''\{\beta + d(y)\} - 2\alpha''\beta' - \alpha'\beta''$, $t^{iv}(y;\theta) = -\alpha^{iv}\{\beta + d(y)\} - 3(\alpha'''\beta' + \alpha''\beta'') - \alpha'\beta'''$, and $t^{v}(y;\theta) = -\alpha^{v}\{\beta + d(y)\} - 4\alpha^{iv}\beta' - \alpha''\beta^{iv} - \alpha'''\beta''' - 3(\alpha''\beta''' + \alpha'''\beta'')$. Using the results above, we can show, after some algebra, that

$$B_1(\theta) = -\frac{\beta''}{2\alpha'\beta'^2}, \tag{7}$$

$$B_2(\theta) = \frac{-12\beta'\alpha''\beta''^2 - 33\alpha'\beta''^3 + 4\beta'^2\alpha''\beta''' + 26\alpha'\beta'\beta''\beta''' - 3\alpha'\beta'^2\beta^{iv}}{24\alpha'^3\beta'^5}, \tag{8}$$

$$B_2^*(\theta) = \frac{1}{24\alpha'^4\beta'^5}(12\beta'^2\alpha''^2\beta'' + 18\alpha'\beta'\alpha''\beta''^2 + 15\alpha'^2\beta''^3 - 6\alpha'\beta'^2\beta''\alpha''' - 8\alpha'\beta'^2\alpha''\beta''' - 16\alpha'^2\beta'\beta''\beta''' + 3\alpha'^2\beta'^2\beta^{iv}), \tag{9}$$

$$V_1(\theta) = \frac{1}{\alpha'\beta'}, \tag{10}$$

$$V_2(\theta) = \frac{2\beta'\alpha''\beta'' + 5\alpha'\beta'' - 2\alpha'\beta'\beta'''}{2\alpha'^3\beta'^4}. \tag{11}$$

It is noteworthy that the expressions for $B_1(\theta)$, $B_2(\theta)$, $B_2^*(\theta)$, $V_1(\theta)$ and $V_2(\theta)$ given in (7)-(11) only require knowledge of $\alpha$, $\beta$ and their first few derivatives with respect to $\theta$. Unlike the general expressions given in the previous section, these expressions do not require the evaluation of moments or cumulants. Formulas (7)-(9) can be implemented in a computer algebra system such as Maple (Abell and Braselton, 1994) or Mathematica (Wolfram, 1991) to obtain bias-corrected estimates with minimal effort. They are capable of generating both simple and complex expressions for different special cases. Also, it is possible to use the expressions above to examine under what conditions the maximum likelihood estimate of $\theta$ is unbiased to a certain order of magnitude. For example, equations (7) and (8) imply that the second and third order biases vanish when $\beta'' = 0$.

It is possible to check equations (7)-(9) from first principles for some special distributions. The simplest special case is the normal distribution with mean $\mu$ and variance $\theta$ for which $B_1(\theta)$ and $B_2(\theta)$ vanish. Here, $\hat{\theta} = \sum(y_i - \mu)^2/n \sim \theta\chi_n^2/n$ and $\hat{\mu} = \sum y_i/n$ are clearly unbiased estimates. It is also easy to verify that the same happens for the following distributions: binomial, Poisson, exponential, Laplace, truncated extreme value and inverse Gaussian with known scale parameter $\theta$. For the inverse Gaussian distribution with known mean $\mu > 0$ and scale parameter $\theta > 0$, for which $\alpha(\theta) = \theta$, $\zeta(\theta) = \theta^{-1/2}$, $d(y) = (y-\mu)^2/(2\mu^2 y)$ and $v(y) = -\{\log(2\pi y^3)\}/2$, $\hat{\theta} = n\mu^2\{\sum(y_i - \mu)^2/y_i\}^{-1} \sim n\theta/\chi_n^2$. By Taylor series expansion we can prove that $\mathrm{E}(\hat{\theta} - \theta) = 2\theta/n + 4\theta/n^2 + O(n^{-3})$, in agreement with the corresponding expressions for $B_1(\theta)$ and $B_2(\theta)$ that are obtained using (7) and (8). For the gamma distribution with known index $k > 0$ and scale parameter $\theta > 0$, for which $\alpha(\theta) = \theta$, $\zeta(\theta) = \theta^{-k}$, $d(y) = y$ and $v(y) = (k-1)\log(y) - \log\{\Gamma(k)\}$, $\Gamma(\cdot)$ being the gamma function, the distribution of $\hat{\theta}$ is $2nk\theta/\chi_{2nk}^2$. Thus, by direct expansion $\mathrm{E}(\hat{\theta} - \theta) = \theta/(nk) + \theta/(nk)^2 + O(n^{-3})$, as can be obtained from (7) and (8).

For the Pareto distribution with $k > 0$ known $(y > k)$ and scale parameter $\theta > 0$, where $\alpha(\theta) = \theta + 1$, $\zeta(\theta) = (\theta k^\theta)^{-1}$, $d(y) = \log(y)$ and $v(y) = 0$, $2n\theta/\hat{\theta} \sim \chi_{2n}^2$, we have that $\mathrm{E}(\hat{\theta} - \theta) = \theta/n + \theta/n^2 + O(n^{-3})$, leading to the corresponding expressions for $B_1(\theta)$ and $B_2(\theta)$. The maximum likelihood estimate $\hat{\theta}$ of $\theta \in I\!R$ for the extreme value distribution, $\phi > 0$ known, for which $\alpha(\theta) = \exp\{\theta/\phi\}$, $\zeta(\theta) = \phi\exp\{-\theta/\phi\}$, $d(y) = \exp\{-y/\phi\}$ and $v(y) = -y/\phi$, reduces to $\hat{\theta} = -\phi\log(n^{-1}\sum e^{-y_i/\phi})$ and we can also show that $\mathrm{E}(\hat{\theta} - \theta) = \phi/(2n) + \phi/(12n^2) + O(n^{-3})$. This result again agrees with the expressions of $B_1(\theta)$ and $B_2(\theta)$ obtained from (7) and (8). Finally, it is of interest to note that for the lognormal distribution with known location parameter $\mu > 0$ and scale parameter $\theta > 0$, where $\alpha(\theta) = \theta^{-2}$, $\zeta(\theta) = \theta$, $d(y) = \{\log(y) - \mu\}^2/2$ and $v(y) = -\log(y) + \{\log(2\pi)\}/2$, $\hat{\theta}^2 \sim \theta^2\chi_n^2/n$. The resulting expansion $\mathrm{E}(\hat{\theta} - \theta) = -\theta/(4n) + \theta/(32n^2)$ to order $n^{-2}$ follows after some algebra, which agrees with the results from (7) and (8).

Next, we obtain the bias-corrected maximum likelihood estimates of the canonical parameter $\alpha = \alpha(\theta)$ in uniparametric natural exponential family models. The one-parameter exponential family (6) parameterized in terms of the canonical parameter $\alpha = \alpha(\theta)$ has the form

$$\pi(y;\alpha) = \frac{1}{\delta(\alpha)}\exp\{-\alpha d(y) + v(y)\}, \tag{12}$$

where $-d(y)$ is the canonical statistic and $\delta(\alpha)$ is its cumulant generating function. Similar conditions to the ones stated for the probability or density function in (6) are assumed to hold.

The $r$th cumulant of $-d(y)$ is $\kappa_r = \mathrm{d}^r\delta(\alpha)/\mathrm{d}\alpha^r$, for $r = 1, 2, \ldots$ Letting $\beta(\alpha) = \delta'(\alpha)/\delta(\alpha)$ we write the $(r+1)$th cumulant of $-d(y)$ as $\kappa_{r+1} = \mathrm{d}^r\beta/\mathrm{d}\alpha^r$, for $r = 1, 2, \ldots$ From now on primes will denote derivatives with respect to $\alpha$. The mean and variance of $-d(y)$ are $\kappa_1 = \beta$ and $\kappa_2 = \mathrm{d}\beta/\mathrm{d}\alpha = \beta'$. The variance function $\beta'$ uniquely characterizes the distribution in (12) and does not depend upon the particular parameterization used.

Using the notation introduced in Section 1, let $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\tilde{\alpha}_2$ be the bias corrected maximum likelihood estimates defined by equations (2), (3) and (4), respectively. The $n^{-1}$ and $n^{-2}$ biases of the maximum likelihood estimate $\hat{\alpha}$ of $\alpha$ are obtained from equations (7) and (8) as $B_1(\alpha) = -\beta''/(2\beta'^2)$ and $B_2(\alpha) = (-33\beta''^3 + 26\beta'\beta''\beta''' - 3\beta'^2\beta^{iv})/(24\beta'^5)$. Also, formulas (10) and (11) yield, to order $n^{-2}$, $\mathrm{var}(\hat{\alpha}) = V_1(\alpha)/n + V_2(\alpha)/n^2$ with $V_1(\alpha) = 1/\beta'$ and $V_2(\alpha) = (5\beta''^2 - 2\beta'\beta''')/(2\beta'^4)$. Let $\gamma_1^2 = \beta''^2/\beta'^3$ and $\gamma_2 = \beta'''/\beta'^2$ be the third and fourth standardized cumulants of $-d(y)$. From the equations for the mean squared errors given in Section 2 we find, after some algebra, that to order $n^{-2}$

$$\mathrm{MSE}(\hat{\alpha}_1) = \mathrm{MSE}(\hat{\alpha}_2) = \mathrm{MSE}(\tilde{\alpha}_2) = \frac{1}{n\beta'} + \frac{\gamma_1^2}{2n^2\beta'}$$

and

$$\mathrm{MSE}(\hat{\alpha}) - \mathrm{MSE}(\tilde{\alpha}_2) = \frac{9\gamma_1^2 - 4\gamma_2}{4n^2\beta'}.$$

Since $\beta' > 0$, it is clear that $\mathrm{MSE}(\hat{\alpha}) < \mathrm{MSE}(\tilde{\alpha}_2)$ when $9\gamma_1^2 - 4\gamma_2 < 0$ and $\mathrm{MSE}(\hat{\alpha}) > \mathrm{MSE}(\tilde{\alpha}_2)$ when $9\gamma_1^2 - 4\gamma_2 > 0$. The values of $\gamma_1$ and $\gamma_2$ must satisfy the inequalities $\gamma_1^2 + \gamma_2 + 2 \geq 0$ and $\gamma_2 - \gamma_1^2 \geq 1$. Then, the sign of $9\gamma_1^2 - 4\gamma_2$ determines the best estimate of $\alpha$.

We shall now derive simple expressions for the biases $B_1(\alpha)$, $B_2(\alpha)$ and $B_2^*(\alpha)$ for a special family of variance functions: the power variance function defined by $\beta' = \beta^p/c$, where $c > 0$ and we assume that $\beta > 0$ for $p \neq 0$. Here, $p = 0, 1, 2, 3$ for the normal, Poisson, gamma and inverse Gaussian distributions, respectively. Other values of $p$ define a number of distributions which have been classified by Jørgensen (1987). Cases with power $0 < p < 1$ do not correspond to distributions in (12).

For the power variance function we get $B_1(\alpha) = -p/(2\beta)$, $B_2(\alpha) = p(p-6)(p+1)\beta^{p-3}/(24c)$ and $B_2^*(\alpha) = p(p-3)(p-2)\beta^{p-3}/(24c)$. Note that $B_1(\alpha) > 0$ for $p < 0$ and $B_1(\alpha) < 0$ for $p \geq 1$. Also, $B_2(\alpha)$ is positive for $-1 < p < 0$ or $p > 6$ and negative for $p < -1$ or $1 < p < 6$. Clearly, $p = 0$ yields $B_1(\alpha) = B_2(\alpha) = 0$, as expected, and $p = -1$ and $p = 6$ lead to $B_2(\alpha) = 0$. $B_2^*(\alpha)$ vanishes when $p = 0, 2, 3$, being positive for $1 < p < 2$ or $p > 3$ and negative for $p < 0$ or $2 < p < 3$. A check of the expressions for $B_1(\alpha)$ and $B_2(\alpha)$ is provided for the Poisson and gamma distributions. For the Poisson distribution ($p = 1$) $\alpha = \log(\theta)$, $c = 1$ and $d(y) = -y$ (since $\beta = \theta > 0$). By Taylor series expansion we can find $\mathrm{E}(\hat{\alpha} - \alpha) = -1/(2n\theta) - 5/(12n^2\theta^2) + O(n^{-3})$, which agrees with the formulas of $B_1(\alpha)$ and $B_2(\alpha)$ for $p = 1$. For the gamma distribution with known scale parameter $k$ ($p = 2$) $\alpha = -\theta$, $c = k$ and $d(y) = -y$ (since $\beta = k/\theta > 0$). Direct expansion yields $\mathrm{E}(\hat{\alpha} - \alpha) = \alpha/(nk) + \alpha/(n^2k^2) + O(n^{-2})$, in agreement with the values of $B_1(\alpha)$ and $B_2(\alpha)$ for $p = 2$.

For the power variance function, the $n^{-2}$ term in the variance of $\hat{\alpha}$ reduces to $V_2(\alpha) = p(p+2)/(2\beta^2)$, which vanishes for $p = 0$ and $p = -2$. The mean squared errors can be expressed in the forms

$$\text{MSE}(\hat{\alpha}_1) = \text{MSE}(\hat{\alpha}_2) = \text{MSE}(\tilde{\alpha}_2) = \frac{c}{n\beta^p} + \frac{p^2}{2n^2\beta^2} + O(n^{-3})$$

and

$$\text{MSE}(\hat{\alpha}) = \frac{c}{n\beta^p} + \frac{p(3p+4)}{4n^2\beta^2} + O(n^{-3}).$$

Therefore,

$$\text{MSE}(\hat{\alpha}) - \text{MSE}(\tilde{\alpha}_2) = \frac{p(p+4)}{4n^2\beta^2} + O(n^{-3}).$$

This equation shows that, to order $n^{-2}$, $\text{MSE}(\hat{\alpha}) > \text{MSE}(\tilde{\alpha}_2)$ for $p < -4$ or $p > 1$, and $\text{MSE}(\hat{\alpha}) < \text{MSE}(\tilde{\alpha}_2)$ for $-4 < p < 0$. For $p = -4$ or $p = 0$, $\text{MSE}(\hat{\alpha}) = \text{MSE}(\tilde{\alpha}_2)$.

Bias-corrected maximum likelihood estimates for other families of variance functions can also be obtained using our results. For example, one can use the results in this paper to obtain the second and third order biases of the maximum likelihood estimate for quadratic (Morris, 1982), cubic (Letac and Mora, 1990) and Babel variance functions.

## Acknowledgements

The partial financial support of CNPq/Brazil is gratefully acknowledged.

## References

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

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

Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society B* **49**, 127-162.

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

Letac, G. & Mora, M. (1990). Natural real exponential families with cubic variance function. *Annals of Statistics* **18**, 1-37.

Morris, C.N. (1982). Natural exponential families with quadratic variance functions. *Annals of Statistics* **10**, 65-80.

Shenton, L.R. & Bowman, K. (1977). *Maximum Likelihood Estimation in Small Samples*. New York: MacMillan.

Wolfram, S. (1991). *Mathematica: A System for Doing Mathematics by Computer*, 2nd ed. New York: Addison-Wesley.