



Analytic Bias Correction for Maximum Likelihood Estimators When the Bias Function is Non-Constant

Ryan T. Godwin

Department of Economics, University of Manitoba
Winnipeg, MB, Canada R3T 5V5

&

David E. Giles

Department of Economics, University of Victoria
Victoria, B.C., Canada V8W 2Y2

July 2017

Summary

Recently, many papers have obtained analytic expressions for the biases of various maximum likelihood estimators, despite their lack of closed-form solution. These bias expressions have provided an attractive alternative to the bootstrap. Unless the bias function is “flat,” however, the expressions are being evaluated at the wrong point(s). We propose an “improved” analytic bias adjusted estimator, in which the bias expression is evaluated at a more appropriate point (at the bias adjusted estimator itself). Simulations illustrate that the improved analytic bias adjusted estimator can eliminate significantly more bias than the simple estimator which has been well established in the literature.

Keywords: bias reduction; maximum likelihood; nonlinear bias function

JEL Classifications: C13; C15

Author Contact:

Ryan T. Godwin, Department of Economics, University of Manitoba, Winnipeg, MB, Canada R3T 5V5;
E-mail: Ryan.Godwin@umanitoba.ca; Voice: (204) 272-1534; FAX: (204) 474-7081

1. Introduction

Let $\hat{\theta}$ denote the MLE for a parameter vector, θ . Provided that $E[\hat{\theta}]$ exists, we can write

$$E[\hat{\theta}] = \theta + B_{\hat{\theta}}(\theta),$$

where $B_{\hat{\theta}}(\theta)$ is the bias function of $\hat{\theta}$, which depends on the true value of θ . Typically, $B_{\hat{\theta}}(\theta)$ is difficult to determine since $\hat{\theta}$ often does not have closed form, so that in order to obtain a bias corrected estimator $B_{\hat{\theta}}(\theta)$ must be either estimated or approximated.

A number of recent papers have successfully obtained analytic expressions for $B_{\hat{\theta}}(\theta)$ which are valid to $O(n^{-1})$, based on the work of Cox and Snell (1968), and Cordeiro and Klein (1994). Some examples include: Cordeiro and McCullagh (1991), Cordeiro *et al.* (1996), Cribari-Neto and Vasconcellos (2002), Cordeiro and Vasconcellos (1997), Silva *et al.* (2008), Giles (2012), Giles *et al.* (2013, 2016), Schwartz *et al.* (2013), Schwartz and Giles (2016), and Godwin (2016). A useful overview of this topic is provided by Cordeiro and Cribari-Neto (2014). This literature has established the analytic Cox-Snell/Cordeiro-Klein method as a viable way of reducing bias, and as an attractive competitor to bias correction *via* the bootstrap. However, this literature has ignored an important refinement in the bias correction technique; namely the appropriate point at which to evaluate $B_{\hat{\theta}}(\theta)$. The aim of this paper is to propose that the analytically obtained bias expression should be evaluated at the bias-corrected MLE itself, and to illustrate that this proposal is justified. The basic analytic bias correction technique is now outlined.

Let $l(\theta)$ be the log-likelihood function based on a sample of n observations, and θ be a p -dimensional parameter vector. It is assumed that $l(\theta)$ is regular with respect to all

derivatives up to and including the third order. The joint cumulants of $l(\theta)$ are defined as follows:

$$k_{ij} = E \left(\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right) ; i, j = 1, 2, \dots, p \quad (1)$$

$$k_{ijl} = E \left(\frac{\partial^3 l}{\partial \theta_i \partial \theta_j \partial \theta_l} \right) ; i, j, l = 1, 2, \dots, p \quad (2)$$

$$k_{ij,l} = E \left(\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \frac{\partial l}{\partial \theta_l} \right) ; i, j, l = 1, 2, \dots, p. \quad (3)$$

The derivatives of these cumulants are:

$$k_{ij}^{(l)} = \frac{\partial k_{ij}}{\partial \theta_l} ; i, j, l = 1, 2, \dots, p. \quad (4)$$

All of the expressions in (1) - (4) are assumed to be $O(n)$, and (1) is the $(i, j)^{th}$ element of Fisher's expected information matrix, $K = \{-k_{ij}\}$. Cox and Snell (1968) proved that when the sample data are independent, the bias of the s^{th} element of the MLE of θ is:

$$B_{\hat{\theta}_s}(\theta) = \sum_{i=1}^p \sum_{j=1}^p \sum_{l=1}^p k^{si} k^{jl} [0.5k_{ijl} + k_{ij,l}] + O(n^{-2}) ; s = 1, 2, \dots, p, \quad (5)$$

where k^{ij} is the $(i, j)^{th}$ element of the inverse of K . Our notation of the bias function, $B_{\hat{\theta}_s}(\theta)$, highlights the dependence on the unknown parameter vector, θ . Extending the work of Cox and Snell (1968), Cordeiro and Klein (1994) showed that even when the data are not necessarily independent, (5) can be written as:

$$B_{\hat{\theta}_s}(\theta) = \sum_{i=1}^p k^{si} \sum_{j=1}^p \sum_{l=1}^p [k_{ij}^{(l)} - 0.5k_{ijl}] k^{jl} + O(n^{-2}) ; s = 1, 2, \dots, p, \quad (6)$$

provided that all of the k terms are $O(n)$.

Note that the expression in (6) does not involve the terms as defined in (3), and is generally easier to evaluate than (5). Now, let $a_{ij}^{(l)} = k_{ij}^{(l)} - 0.5k_{ijl}$, for $i, j, l = 1, 2, \dots, p$; and define the following matrices:

$$A^{(l)} = \{a_{ij}^{(l)}\}; \quad i, j, l = 1, 2, \dots, p \quad (7)$$

$$A = [A^{(1)} | A^{(2)} | \dots \dots | A^{(p)}]. \quad (8)$$

Equations (7) and (8) allow (6) to be rewritten as:

$$B_{\hat{\theta}}(\theta) = K^{-1}A \text{vec}(K^{-1}) + O(n^{-2}). \quad (9)$$

Finally, we can define the ‘‘bias-corrected’’ MLE for θ as:

$$\tilde{\theta} = \hat{\theta} - \hat{K}^{-1}\hat{A} \text{vec}(\hat{K}^{-1}), \quad (10)$$

where $\hat{K} = (K)|_{\theta=\hat{\theta}}$, $\hat{A} = (A)|_{\theta=\hat{\theta}}$, $\text{vec}(\hat{K}^{-1})$ is the vector obtained by stacking the columns of the matrix \hat{K}^{-1} , and $\hat{\theta}$ is the MLE of θ . Note how (9) and (10) illustrate that the Cox and Snell/Cordeiro and Klein bias correction procedure can be implemented even when the MLE cannot be expressed in a closed form solution, since (9) and (10) are obtained solely through manipulation of the cumulants of the log-likelihood function, and do not require an explicit algebraic expression for $\hat{\theta}$. This adds greatly to the usefulness of this bias correction procedure.

In contrast to this ‘‘corrective’’ approach to dealing with estimation bias, Firth (1993), and other authors, have considered ‘‘preventive’’ methods based on an adjustment

to the score vector associated with the log-likelihood function. We do not consider this alternative approach in this paper.

2. Improved Analytic Bias Correction

The analytic bias expression, $B_{\hat{\theta}}(\theta)$ in (9), is in general a function of the true and unobservable parameter vector, θ . The simple analytic bias corrected MLE (10), may be re-written as:

$$\tilde{\theta} = \hat{\theta} - B_{\hat{\theta}}(\theta)|_{\theta=\hat{\theta}}, \quad (11)$$

where the new notation highlights that the analytically determined bias expression is evaluated at the MLE, $\hat{\theta}$, in order for it to be used in practice. However, unless the bias function is ‘flat’ or constant, the motivation for evaluating $B_{\hat{\theta}}(\theta)$ at the biased MLE $\hat{\theta}$ is unclear. If the bias function is non-constant, then ideally the analytic bias expression should be evaluated at the bias corrected MLE itself. MacKinnon and Smith (1998) discuss just such an estimator, which they term the non-linear bias correcting (NBC) estimator (but which is also valid for linear bias functions):

$$\check{\theta} = \hat{\theta} - B_{\hat{\theta}}(\theta)|_{\theta=\check{\theta}}. \quad (12)$$

In (12), the bias function, $B_{\hat{\theta}}(\theta)$, is evaluated at $\check{\theta}$ (which is unbiased), rather than evaluated at $\hat{\theta}$ (which is biased). Of course, the difficulty with $\check{\theta}$ is that there is likely no closed-form solution for this expression. However, standard numerical algorithms may be used to solve for $\check{\theta}$. For example, by rearranging (12):

$$\check{\theta} - \hat{\theta} + B_{\hat{\theta}}(\theta)|_{\theta=\check{\theta}} = 0,$$

and algorithms which find the roots of non-linear equations may be used to find $\check{\theta}$. One such option is the “nleqslv” package (Hasselman, 2015), in *R*. The usual algorithms used for locating roots in nonlinear equations can be imprecise¹. Hence, an alternative and potentially more attractive way to solve for $\check{\theta}$ is seen by rearranging (12) yet again:

$$\hat{\theta} = \check{\theta} + B_{\hat{\theta}}(\theta)|_{\theta=\check{\theta}}.$$

The estimator $\check{\theta}$ may then be obtained by substituting $[\theta + B_{\hat{\theta}}(\theta)]$ for θ in the log-likelihood function, and maximizing as usual.

The estimator in (12) has been applied in the bootstrap bias correction literature, but to our knowledge, (12) has not been applied in the literature where $B_{\hat{\theta}}(\theta)$ is obtained analytically. Hence, we propose $\check{\theta}$ as an improvement over $\tilde{\theta}$. A practical limitation to obtaining $\check{\theta}$ *via* the bootstrap is that $B_{\hat{\theta}}(\theta)$ must be estimated several times requiring considerable additional computation in the form of additional bootstrap loops. In contrast, the proposed improved analytical bias corrected MLE requires minimal additional computation. In addition to MacKinnon and Smith (1998), Cribari-Neto and Lima (2011) propose a related estimator. They obtain a sequence of bias-corrected heteroskedasticconsistent estimators for standard errors in the linear regression model. The estimator obtained by Cribari-Neto and Lima (2011) is similar to the one proposed here in that their estimator is obtained by iteratively bias correcting “in an accurate fashion.”

The biases of $\check{\theta}$ and $\tilde{\theta}$, when $B_{\hat{\theta}}(\theta)$ is non-linear, have been determined by MacKinnon and Smith (1998), and are:

¹ We are grateful to an anonymous referee for pointing this out.

$$B_{\ddot{\theta}}(\theta) = -\frac{B_{\hat{\theta}}''(\theta)}{2(1+B_{\hat{\theta}}'(\theta))}E(\hat{\theta} - \theta)^2 + O(n^{-3}) \quad (13)$$

and

$$B_{\tilde{\theta}}(\theta) = -B_{\hat{\theta}}'(\theta)B_{\hat{\theta}}(\theta) - \frac{1}{2}B_{\hat{\theta}}''(\theta)E(\hat{\theta} - \theta)^2 + o(n^{-2}) \quad (14)$$

where $B_{\hat{\theta}}'(\theta)$ and $B_{\hat{\theta}}''(\theta)$ are the first and second derivatives of $B_{\hat{\theta}}(\theta)$. Note that (13) and (14) depend on the true unobservable bias function $B_{\hat{\theta}}(\theta)$, and that when $B_{\hat{\theta}}(\theta)$ is approximated, such as by equation (9) which is valid to $O(n^{-1})$, additional bias terms at $O(n^{-2})$ will enter (13) and (14). By examining (13) and (14) MacKinnon and Smith (1998) show that: (i) when $B_{\hat{\theta}}(\theta)$ is non-linear both $\ddot{\theta}$ and $\tilde{\theta}$ are unbiased to $O(n^{-1})$, and that $\ddot{\theta}$ and $\tilde{\theta}$ differ only by terms that are $O(n^{-2})$; (ii) depending on whether $B_{\hat{\theta}}'$ is positive or negative, $\ddot{\theta}$ may be either more or less biased than $\tilde{\theta}$, through terms of order n^{-2} ; (iii) when $B_{\hat{\theta}}(\theta)$ is linear then the first term of (13) disappears and $\ddot{\theta}$ is less biased than $\tilde{\theta}$. Furthermore, the slope of the bias function, as well as the magnitude of the bias of $\hat{\theta}$ relative to its variance, determines whether bias correction results in smaller or larger mean-squared error; when $B_{\hat{\theta}}'$ is negative the variances of $\ddot{\theta}$ and $\tilde{\theta}$ will be larger than that of $\hat{\theta}$.

Finally, it is possible to eliminate the $O(n^{-2})$ bias terms in (13) by obtaining the first and second derivatives of (9), and evaluating them and $E(\hat{\theta} - \theta)^2$ at the bias-corrected estimator. This would eliminate some of the $O(n^{-2})$ bias in $\ddot{\theta}$, which arises when $B_{\hat{\theta}}(\theta)$ is non-linear. However, the motivation for doing so is dubious since an $O(n^{-2})$ bias

will still remain due to $B_{\hat{\theta}}(\theta)$ being approximated or estimated by methods that are valid only to $O(n^{-1})$.

3. A Simple Example

Consider the biased estimator of the error variance of a linear regression model with p fixed regressors, $\hat{\sigma}^2 = SSR/n$, where SSR is the sum-of-squared residuals. Applying the Cordeiro-Klein procedure to this estimator yields the familiar bias expression:

$$B_{\hat{\sigma}^2}(\sigma^2) = -\frac{p}{n}\sigma^2 + O(n^{-2}),$$

and in this case we know that the higher order bias terms ($O(n^{-2})$, *etc.*) are zero. Using the bias correction as established in the Cordeiro-Klein based literature (the estimators in (10) or (11)), the simple analytic bias corrected estimator (that is unbiased to $O(n^{-1})$) is:

$$\tilde{\sigma}^2 = \hat{\sigma}^2 + \frac{p}{n}\hat{\sigma}^2 = \frac{SSR(n+p)}{n^2}.$$

Using the improved bias correction proposed in this paper, (12), a better bias-corrected estimator is:

$$\ddot{\sigma}^2 = \hat{\sigma}^2 + \frac{p}{n}\ddot{\sigma}^2 = \frac{SSR}{n-p},$$

giving a familiar estimator. In $\tilde{\sigma}^2$ the $O(n^{-1})$ bias has been evaluated at a biased estimator, whereas in $\ddot{\sigma}^2$ the $O(n^{-1})$ bias has been evaluated at an estimator that is unbiased. In this simple example, a closed-form solution for $\ddot{\theta}$ ($\ddot{\sigma}^2$) is available, however, in most practical situations numerical methods will be required to solve for $\ddot{\theta}$.

4. Illustrative Examples

In this section, four illustrative examples are provided. For each example, four different MLEs are simulated for each parameter of the distribution, in experiments consisting of 10,000 replications for each sample size. The different estimators simulated are: (i) $\hat{\theta}$, the (un-corrected MLE); (ii) $\tilde{\theta}$, the simple analytic bias adjusted estimator obtained from (11); (iii) $\check{\theta}$, the parametric bootstrap bias adjusted estimator; and (iv) $\ddot{\theta}$, the improved bias adjusted estimator from (12), proposed in this paper.

The parametric bootstrap bias adjusted estimator is included, but is not a focus. The intent of this paper is to improve upon $\tilde{\theta}$, which has already been extensively compared to bootstrap bias-adjustment (for example, in those papers referenced in Section 1). The parametric bootstrap bias adjusted estimator is obtained as $\check{\theta} = 2\hat{\theta} - (1/N_B)[\sum_{j=1}^{N_B} \hat{\theta}_{(j)}]$, where $\hat{\theta}_{(j)}$ is the MLE of λ obtained from the j^{th} of the N_B bootstrap samples, and where each bootstrap sample N_B has been randomly generated as though $\hat{\theta}$ were the data generating process. We set $N_B = 1,000$ which implies 1 million replications for each row of the results reported below, with some rows requiring several days of computational time. The choice of $N_B = 1,000$ is quite standard, and increasing N_B is unlikely to alter the performance of the bootstrap estimator.

In each illustrative example, the percent bias and percent mean-squared error (MSE) of the various estimators (averaged over the 10,000 replications) are reported. In most instances, the proposed estimator $\ddot{\theta}$ is the least biased of the four considered. As the sample size is increased to say $n = 1,000$, all four estimators become similar, reflecting the well known asymptotic properties of the MLEs. The MSE of $\ddot{\theta}$ can be higher than that

of $\tilde{\theta}$ or even $\hat{\theta}$. This is likely due to the variance of $\tilde{\theta}$ increasing when the slope of the bias function is negative (as discussed in Section 2 and by MacKinnon and Smith, 1998).

The simulations were carried out in the R Statistical Environment (versions 3.3.1 and 3.2.1) on two different machines: one with a 4.70 GHz AMD FX-9590 eight-core processor and the other with a 4.00 GHz AMD FX-8350 eight-core processor. Random variates were generated via: the `rlogis` function in the base of R for Section 4.1; the `rpois` function in the VGAM package (Yee, 2017) for Section 4.2; random uniform variates and the cumulative density function for the Lomax distribution in Section 4.3; and the `rgpd` function in the evd package (Stephenson, 2002).

4.1: The Half-Logistic Distribution with Known Location Parameter

Under independent sampling from the half-logistic distribution, with uncensored data, when the location parameter is equal to zero, the log-likelihood function is:

$$l = n \ln(2) - n \ln(\sigma) + (n\bar{y}/\sigma) - 2 \sum_{i=1}^n \ln[1 + \exp(y_i/\sigma)].$$

The MLE of σ is found by numerically solving the following first-order condition:

$$\partial l / \partial \sigma = -(n/\sigma) - (n\bar{y}/\sigma^2) + (2/\sigma^2) \sum_{i=1}^n [y_i \exp(y_i/\sigma)] / [1 + \exp(y_i/\sigma)] = 0.$$

Despite the lack of closed-form solution for the MLE, $\hat{\sigma}$, using the Cox-Snell/Cordeiro-Klein procedure outlined in Section 1, Giles (2012) determines the following bias expression for $\hat{\sigma}$:

$$B_{\hat{\sigma}}(\sigma) = -0.052567665(\sigma/n) + O(n^{-2}).$$

Using (11), the simple analytic bias adjusted estimator is:

$$\tilde{\sigma} = \hat{\sigma} + 0.052567665(\hat{\sigma}/n) = \hat{\sigma}(n + 0.052567665)/n,$$

however, it is clear that $B_{\hat{\sigma}}(\sigma)$ is not constant but is indeed a linear function of σ and that $B_{\hat{\sigma}}(\sigma)$ has been evaluated at the wrong point; a biased estimator. Using (12), the improved bias adjusted estimator is:

$$\check{\sigma} = \hat{\sigma} + 0.052567665(\check{\sigma}/n) = \hat{\sigma}n/(n - 0.052567665).$$

The improved bias-adjusted estimator has closed form solution, however, this is due to the linearity of $B_{\hat{\sigma}}(\sigma)$ and is the only such case that the authors are aware of.

The three bias-corrected MLEs, $\tilde{\sigma}$ (simple analytic adjusted, $\check{\sigma}$ (bootstrap adjusted) and $\ddot{\sigma}$ (improved analytic adjusted), are quite similar. Monte Carlo results for the percent bias and percent MSE for the various MLEs are presented in Table 1. All bias corrected estimators reduce the bias of $\hat{\sigma}$ in most cases, but tend to increase the MSE. This finding is not surprising since the bias function, $B_{\hat{\sigma}}(\sigma)$, is downward-sloping (MacKinnon and Smith, 1998).

4.2: The Positive Poisson Distribution

Under independent sampling from the positive Poisson distribution, the log-likelihood function is

$$l = n(\bar{y} \ln \lambda - \ln(e^\lambda - 1)) - \sum_{i=1}^n \ln y_i! .$$

The MLE of λ is obtained by solving $\bar{y} = \lambda e^\lambda / (e^\lambda - 1)$, numerically. Through the procedure outlined in Section 1, Godwin (2016) obtains the Cox-Snell/Cordeiro-Klein analytic bias expression of the MLE, $\hat{\lambda}$, to be:

$$B_{\hat{\lambda}}(\lambda) = \frac{\lambda(e^{\lambda}-1)(2e^{\lambda}-\lambda e^{\lambda}-\lambda-2)}{2ne^{\lambda}(\lambda-e^{\lambda}+1)^2} + O(n^{-2}).$$

Table 2 presents simulation results over 10,000 Monte Carlo replications, for the %bias and %MSE of: the MLE $\hat{\lambda}$; the simple analytic bias adjusted MLE $\tilde{\lambda}$, obtained from (11); the bootstrap bias adjusted estimator $\check{\lambda}$; and the improved analytic bias adjusted MLE $\ddot{\lambda}$, obtained from (12). For this example, the estimator $\ddot{\lambda}$ in (12) is solved for by substituting $\lambda + B_{\hat{\lambda}}(\lambda)$ for λ in the log-likelihood function:

$$l = n \left(\bar{y} \ln[\lambda + B_{\hat{\lambda}}(\lambda)] - \ln \left(e^{[\lambda + B_{\hat{\lambda}}(\lambda)]} - 1 \right) \right) - \sum_{i=1}^n \ln y_i!$$

and maximizing as usual. Obtaining $\ddot{\lambda}$ by solving the roots of (12) provides near identical results. Both methods are illustrated via R code in the Appendix.

All bias corrected estimators reduce the bias but slightly increase the MSE. The reason why both $\tilde{\lambda}$ and $\check{\lambda}$ are nearly identical is because the $O(n^{-1})$ bias expression is essentially ‘flat’ over the range of values in which the $O(n^{-1})$ bias expression is being evaluated. A plot of the bias expression is shown in Figure 1. While the expression certainly does not look flat, the MLE $\hat{\lambda}$ contains so little bias that it is of little import where the bias expression is evaluated.

Note that when $\lambda = 0.75$, the bias function is downward-sloping, and according to Table 2, the bias adjustment has increased the %MSE. When the bias function is increasing, e.g., for $\lambda > 1.5$, bias-correction tends to reduce the %MSE. For example, for $n = 15$ and $\lambda = 2.5$, the %MSE of $\hat{\lambda}$, $\tilde{\lambda}$, and $\check{\lambda}$ is 3.212, 3.193 and 3.193, respectively (obtained from 10,000 Monte Carlo replications). Again, this finding is consistent with MacKinnon and Smith (1998).

4.3: The Lomax Distribution

Based on n independent observations, the log-likelihood function for the Lomax distribution is:

$$l = n \log(\alpha) - n \log(\lambda) - (1 + \alpha) \sum_{i=1}^n \log(1 + y_i/\lambda).$$

The $O(n^{-1})$ Cox-Snell/Cordeiro-Klein bias expressions for the MLEs $\hat{\lambda}$ and $\hat{\alpha}$, obtained by Giles *et al.* (2013), are:

$$B_{\hat{\lambda}} \begin{pmatrix} \lambda \\ \alpha \end{pmatrix} = K^{-1} A \text{vec}(K^{-1}), \quad (15)$$

where

$$K = n \begin{bmatrix} \alpha/[\lambda^2(\alpha + 2)] & -1/[\lambda(\alpha + 1)] \\ -1/[\lambda(\alpha + 1)] & 1/\alpha^2 \end{bmatrix},$$

and

$$A = n \begin{bmatrix} \frac{2\alpha}{\lambda^3(\alpha+2)(\alpha+3)} & \frac{-1}{\lambda^2(\alpha+1)(\alpha+2)} & \frac{\alpha}{\lambda^2(\alpha+2)^2} & \frac{-1}{\lambda(\alpha+1)^2} \\ \frac{-1}{\lambda^2(\alpha+1)(\alpha+2)} & 0 & \frac{-1}{\lambda(\alpha+1)^2} & \frac{1}{\alpha^3} \end{bmatrix}.$$

The simple analytic bias adjusted MLEs, $\tilde{\lambda}$ and $\tilde{\alpha}$, are obtained by evaluating (15) at the MLEs $\hat{\lambda}$ and $\hat{\alpha}$. The improved analytic bias adjusted MLEs, $\check{\lambda}$ and $\check{\alpha}$, are obtained by evaluating (15) at the bias corrected MLEs themselves ($\tilde{\lambda}$ and $\tilde{\alpha}$). The algorithms proposed in the Appendix is not complicated by the fact that there are two parameters, and works just as well as in the one-parameter cases already discussed.

Monte Carlo results for the %bias and %MSE of the various MLEs are presented in Table 3. The reduction in bias when using the improved analytic bias adjustment

compared to the simple analytic bias adjustment can be quite dramatic. For example, for $n = 80$, $\alpha = 2.1$, $\lambda = 1.0$, the MLEs $\hat{\lambda}$ and $\hat{\alpha}$ have %bias of 29.50 and 20.87, respectively. The simple analytic bias adjusted estimators over-correct the bias, so that the estimators $\tilde{\lambda}$ and $\tilde{\alpha}$ have %bias of -25.43 and -17.95, respectively. The improved analytic bias adjusted estimators, however, do not suffer from over-correction. The improved estimators $\check{\lambda}$ and $\check{\alpha}$ have %bias of -1.642 and -1.040, respectively. In all but four instances, the improved estimator has the lowest bias, however, the MSE is almost always greater than the simple analytic adjusted estimator, but is almost always smaller than the bootstrap bias adjusted estimator. The increase in MSE for the improved estimator is likely due to the downward slope of the bias function.

4.4: The Generalized Pareto Distribution

Assuming independent observations the log-likelihood for the generalized Pareto distribution (GPD) is:

$$l(\xi, \sigma) = -n \ln(\sigma) - (1 + 1/\xi) \sum_{i=1}^n \ln(1 + \xi y_i / \sigma).$$

The log-likelihood surface is fraught with abnormalities, and great care must be taken when solving for the MLEs, $\hat{\xi}$ and $\hat{\sigma}$. For the simulations, the log-likelihood was maximized using the method outlined by Grimshaw (1993), and by using *R* code generously supplied by that author. For this same reason, the improved estimator is solved for by finding the roots of equation (12), and not by modifying the log-likelihood function.

Provided that $-1/3 < \hat{\xi} < 1$, Giles *et al.* (2016) obtain the $O(n^{-1})$ bias expressions for the MLEs $\hat{\xi}$ and $\hat{\sigma}$ to be:

$$\text{Bias}(\hat{\xi}) = -(1 + \xi)(3 + \xi)/[n(1 + 3\xi)],$$

$$\text{Bias}(\hat{\sigma}) = \sigma(3 + 5\xi + 4\xi^2)/[n(1 + 3\xi)].$$

Table 4 provides some Monte Carlo results for the %bias and %MSE of the MLEs ($\hat{\xi}$ and $\hat{\sigma}$), the simple analytic bias adjusted MLEs ($\check{\xi}$ and $\check{\sigma}$), the bootstrap bias adjusted estimators ($\tilde{\xi}$ and $\tilde{\sigma}$), and the improved analytic bias adjusted MLEs ($\ddot{\xi}$ and $\ddot{\sigma}$). In almost all cases the improved estimator has smaller bias than the simple analytic adjusted estimator, while maintaining a similar MSE. The bootstrap bias adjusted estimator, however, often has the lowest bias of all, albeit at the cost of higher MSE.

5. Conclusions

Many recent papers have obtained analytic $O(n^{-1})$ bias expressions for MLEs that do not admit a closed form solution, via the Cox-Snell/Cordeiro-Klein method. These papers then evaluate the bias expression at the MLE, and subtract it from the MLE, in order to correct for bias. This paper argues that the bias expression should be evaluated at the bias corrected MLE itself, leading to an “improved” analytic bias corrected estimator. In most cases this implies that the improved analytic bias corrected estimator does not have a closed form solution, however, it can be solved for numerically while imposing negligible computational burden. Monte Carlo evidence suggests that the improved analytic bias corrected estimator can reduce considerably more bias compared to the simple analytic bias corrected estimator already established in the literature, when the MLE has relatively high bias to begin with.

Acknowledgment: We are grateful to two anonymous referees for their helpful and constructive comments on an earlier version of this paper.

References

Cordeiro, G. M., Cribari-Neto, F. (2014). *An Introduction to Bartlett Correction and Bias Reduction*. Heidelberg: Springer.

Cordeiro, G. M., Da Rocha, E. C., Da Rocha, J. G. C., Cribari-Neto, F. (1996). Bias-corrected maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation* 58:21-35.

Cordeiro, G. M., Klein, R. (1994). Bias correction in ARMA models. *Statistics and Probability Letters* 19:169-176.

Cordeiro, G. M., McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society: Series B* 53:629-643.

Cordeiro, G. M., Vasconcellos, K. L. P. (1997). Bias correction for a class of nonlinear regression models. *Statistics and Probability Letters* 35:155-164.

Cox, D. R., Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B* 30:248-275.

Cribari-Neto, F., Vasconcellos, K. L. P. (2002). Nearly unbiased maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation* 72:107-118.

Cribari-Neto, F., Lima, M.G.A. (2011). A sequence of improved standard errors under heteroskedasticity of unknown form. *Journal of Statistical Planning and Inference* 141: 3617-3627.

David, F. N., Johnson, N. L. (1952). The truncated Poisson. *Biometrics* 8:275-285.

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

Giles, D. E. (2012). Bias reduction for the maximum likelihood estimators of the parameters in the half-logistic distribution. *Communications in Statistics-Theory and Methods* 41:212-222.

Giles, D. E., Feng, H., Godwin, R. T. (2016). Bias – corrected maximum likelihood estimation of the parameters of the generalized Pareto distribution. *Communications in Statistics-Theory and Methods* 45:2465-2483.

Giles, D. E., Feng, H., Godwin, R. T. (2013). On the bias of the maximum likelihood estimator for the two-parameter Lomax distribution. *Communications in Statistics-Theory and Methods* 42:1934-1950.

Godwin, R. T. (2016). Bias reduction for the maximum likelihood estimator of the doubly-truncated Poisson distribution. *Communications in Statistics-Theory and Methods* 45:1887-1901.

Grimshaw, S. D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics* 35:185-191.

Hasselmann, B. (2015). nleqslv. R Package.

Henningsen, A., Toomet, O. (2011). maxLik: A package for maximum likelihood estimation in R. *Computational Statistics* 26:443–458.

MacKinnon, J. G., Smith, A. A. (1998). Approximate bias correction in econometrics. *Journal of Econometrics* 85:205–230.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Schwartz, J., Giles, D. E. (2016). Bias-reduced maximum likelihood estimation for the zero-inflated Poisson distribution. *Communications in Statistics – Theory and Methods* 45:465-478.

Schwartz, J., Godwin, R. T., Giles, D. E. (2013). Improved maximum-likelihood estimation of the shape parameter in the Nakagami distribution. *Journal of Statistical Computation and Simulation* 83:434-445.

Silva, M., Cribari-Neto, F., Frery, A. C. (2008). Improved likelihood inference for the roughness parameter of the GA0 distribution. *Environmetrics* 19:347–368.

Stephenson, A. G. (2002). evd: Extreme Value Distributions. *R News* 2:31–32.

Yee, T. W., (2017). VGAM: Vector Generalized Linear and Additive Models. R package version 1.0-3.

Appendix. Two Suggested Methods for Finding $\hat{\theta}$ in the Case of the Positive Poisson Distribution, Illustrated by R Code

```
library (VGAM)
library (maxLik)
library (nleqslv)

# The Cordeiro/Klein analytic bias expression (equation (9))
ckbias <- function(lam) {
  lam * (exp(lam)-1) * (2 * exp(lam) - lam * exp(lam) - lam - 2) /
  (2 * n * exp(lam) * (lam - exp(lam) + 1) ^ 2)
}

# The log-likelihood for the positive Poisson distribution
loglik <- function(lam) sum(y) * log(lam) - n * log(exp(lam) - 1)

# The modified log-likelihood, where lambda has been replaced with
# lambda + bias(lambda hat)
loglikmod <- function(lcup) {
  sum(y) * (log(lcup + ckbias(lcup)))
  - n * (log(exp(lcup + ckbias(lcup)) - 1))
}

# Generate a sample from the positive Poisson distribution
n <- 20
y <- rpospois(n,lambda=1)

# Estimate the MLE
lhat <- maxLik(loglik, start=1)$estimate

# Obtain the simple analytic bias-adjusted estimator (equation (11))
ltilde <- lhat - ckbias(lam=lhat)

# Obtain the improved analytic bias-adjusted estimator from equation
# (12) by solving for the roots
lcup1 <- nleqslv(lhat, function(lcup) lcup - lhat + ckbias(lam=lcup))$x

# Obtain the improved analytic bias-adjusted estimator from equation
# (12) by maximizing the modified log-likelihood
lcup2 <- maxLik(loglikmod, start=1)$estimate
```

Table 1. %Bias, %MSE for $\hat{\sigma}$, $\tilde{\sigma}$, $\check{\sigma}$, and $\ddot{\sigma}$, for $\sigma = 1$, for various sample sizes, and for 10,000 Monte Carlo replications.

n	%Bias $\hat{\sigma}$ [%MSE]	%Bias $\tilde{\sigma}$ [%MSE]	%Bias $\check{\sigma}$ [%MSE]	%Bias $\ddot{\sigma}$ [%MSE]
10	-0.81 [6.90]	-0.29 [6.97]	-0.26 [6.97]	-0.28 [6.96]
15	-0.12 [4.72]	0.23 [4.75]	0.24 [4.76]	0.23 [4.75]
20	-0.56 [3.47]	-0.30 [3.48]	-0.30 [3.48]	-0.30 [3.48]
100	-0.07 [0.71]	-0.01 [0.72]	-0.01 [0.72]	-0.01 [0.72]

Table 2. %Bias, %MSE for $\hat{\lambda}$, $\tilde{\lambda}$, $\check{\lambda}$, and $\ddot{\lambda}$, for $\lambda = 0.75$, for various sample sizes, and for 10,000 Monte Carlo replications.

n	%Bias $\hat{\lambda}$ [%MSE]	%Bias $\tilde{\lambda}$ [%MSE]	%Bias $\check{\lambda}$ [%MSE]	%Bias $\ddot{\lambda}$ [%MSE]
15	-1.25 [14.20]	0.02 [14.39]	-0.01 [14.41]	0.03 [14.39]
25	-1.17 [8.38]	-0.40 [8.44]	-0.40 [8.44]	-0.39 [8.44]
50	-0.63 [4.27]	-0.23 [4.28]	-0.23 [4.29]	-0.23 [4.28]
100	-0.34 [2.13]	-0.14 [2.13]	-0.15 [2.14]	-0.14 [2.13]
1000	-0.03 [0.21]	-0.01 [0.21]	-0.01 [0.21]	-0.01 [0.21]

Table 3. %Bias, and %MSE for the various estimators for the parameters of the Lomax distribution, for various sample sizes and parameter configurations, and for 10,000 Monte Carlo replications.

%Bias [%MSE]								
$\alpha = 1.1, \lambda = 1.0$								
n	$\hat{\alpha}$	$\tilde{\alpha}$	$\check{\alpha}$	$\ddot{\alpha}$	$\hat{\lambda}$	$\tilde{\lambda}$	$\check{\lambda}$	$\ddot{\lambda}$
35	19.60 [30.82]	-15.82 [6.67]	7.91 [44.44]	-1.66 [11.79]	33.83 [84.72]	-28.54 [20.03]	16.34 [132.59]	-2.56 [30.92]
50	15.46 [21.79]	-6.28 [5.68]	3.79 [26.84]	0.29 [10.31]	27.61 [61.84]	-11.24 [15.94]	8.55 [80.90]	1.20 [27.85]
100	7.70 [8.13]	-0.71 [4.86]	-0.64 [6.96]	0.53 [5.49]	13.61 [23.29]	-1.37 [13.30]	-0.99 [20.36]	1.07 [15.15]
150	4.90 [4.41]	-0.21 [3.28]	-0.70 [3.41]	0.27 [3.42]	7.99 [12.16]	-1.03 [8.85]	-1.86 [9.33]	-0.07 [9.24]
250	2.79 [2.18]	-0.09 [1.86]	-0.30 [1.83]	0.07 [1.88]	4.79 [6.27]	-0.28 [5.25]	-0.67 [5.19]	0.04 [5.32]
500	1.43 [0.99]	0.05 [0.92]	-0.00 [0.92]	0.09 [0.92]	2.57 [2.82]	0.14 [2.58]	0.05 [2.58]	0.21 [2.58]
1000	0.64 [0.46]	-0.03 [0.44]	-0.05 [0.44]	-0.02 [0.44]	1.11 [1.31]	-0.08 [1.26]	-0.11 [1.26]	-0.06 [1.26]
$\alpha = 2.1, \lambda = 1.0$								
n	$\hat{\alpha}$	$\tilde{\alpha}$	$\check{\alpha}$	$\ddot{\alpha}$	$\hat{\lambda}$	$\tilde{\lambda}$	$\check{\lambda}$	$\ddot{\lambda}$
80	20.87 [34.63]	-17.95 [9.48]	9.33 [51.83]	-0.36 [12.17]	29.50 [67.65]	-25.43 [18.86]	14.23 [104.13]	-0.20 [23.58]
100	17.32 [27.50]	-10.42 [5.61]	5.22 [37.73]	0.34 [11.12]	24.21 [52.20]	-14.88 [10.99]	7.74 [72.91]	0.50 [20.90]
150	10.94 [14.78]	-3.51 [5.10]	-0.22 [15.49]	0.21 [7.91]	15.44 [28.75]	-5.02 [9.88]	-0.08 [30.56]	0.39 [15.22]
250	6.45 [6.87]	-0.72 [4.15]	-1.10 [5.50]	0.33 [4.72]	9.08 [13.21]	-1.04 [7.92]	-1.55 [10.52]	0.50 [8.99]
500	2.84 [2.48]	-0.27 [2.05]	-0.54 [2.02]	-0.06 [2.09]	3.99 [4.84]	-0.39 [3.98]	-0.78 [3.92]	-0.08 [4.06]
750	1.90 [1.52]	-0.10 [1.34]	-0.22 [1.33]	-0.02 [1.35]	2.73 [2.95]	-0.09 [2.61]	-0.25 [2.58]	0.05 [2.63]
2000	0.68 [0.50]	-0.04 [0.48]	-0.06 [0.48]	-0.03 [0.48]	0.96 [0.97]	-0.05 [0.93]	-0.08 [0.93]	-0.03 [0.93]

Table 4. %Bias, and %MSE for the various estimators of the parameters of the GPD distribution, for various sample sizes and parameter configurations, and for 10,000 Monte Carlo replications.

%Bias [%MSE]								
$\xi = 0.1, \sigma = 1$								
n	$\hat{\xi}$	$\check{\xi}$	$\tilde{\xi}$	$\ddot{\xi}$	$\hat{\sigma}$	$\check{\sigma}$	$\tilde{\sigma}$	$\ddot{\sigma}$
50	-56.41 [354.51]	5.16 [275.85]	3.38 [296.17]	-4.71 [246.35]	6.09 [5.99]	-2.28 [3.63]	-0.82 [4.72]	-0.72 [3.05]
100	-26.76 [150.17]	4.25 [113.02]	1.29 [137.05]	0.27 [114.17]	2.69 [2.57]	-1.12 [1.69]	-0.37 [2.32]	-0.48 [1.79]
200	-12.98 [66.97]	1.45 [57.07]	0.64 [63.84]	0.40 [58.07]	1.37 [1.21]	-0.26 [0.98]	-0.08 [1.15]	-0.11 [1.02]
100	-2.42 [12.71]	0.25 [12.36]	0.23 [12.60]	0.21 [12.38]	0.30 [0.22]	0.02 [0.22]	0.02 [0.22]	0.01 [0.23]
0								
$\xi = -0.1, \sigma = 1$								
50	-64.75 [306.65]	-27.94 [330.71]	5.27 [231.46]	-14.58 [267.22]	6.63 [5.45]	2.56 [4.92]	-1.18 [3.96]	-2.56 [2.25]
100	-31.26 [119.66]	-3.79 [120.83]	1.95 [103.45]	4.49 [85.21]	3.00 [2.25]	0.07 [1.92]	-0.44 [1.96]	-1.74 [1.17]
200	-15.99 [51.26]	1.66 [47.15]	0.64 [46.98]	3.68 [37.22]	1.59 [1.04]	-0.28 [0.87]	-0.09 [0.97]	-0.49 [0.75]
100	-3.31 [8.88]	0.60 [8.24]	0.17 [8.68]	0.51 [8.27]	0.36 [0.19]	-0.03 [0.17]	0.02 [0.18]	-0.03 [0.18]
0								
$\xi = -0.2, \sigma = 1$								
50	-34.98 [73.66]	-29.81 [78.08]	3.64 [51.40]	-14.41 [71.44]	7.03 [5.25]	6.42 [5.00]	-1.56 [3.55]	-3.58 [2.22]
100	-17.18 [27.43]	-12.40 [30.08]	1.15 [22.90]	1.61 [23.22]	3.23 [2.11]	2.58 [2.06]	-0.48 [1.80]	-3.90 [0.95]
200	-9.00 [11.48]	-5.74 [12.82]	0.33 [10.23]	2.68 [10.67]	1.75 [0.97]	1.25 [0.97]	-0.10 [0.89]	-0.60 [0.88]
100	-2.02 [1.86]	-1.03 [2.05]	0.07 [1.79]	0.91 [1.58]	0.42 [0.17]	0.25 [0.17]	0.02 [0.16]	-0.17 [0.15]
0								

Figure 1. The Cox-Snell $O(n^{-1})$ bias expression ($n = 15$) for the MLE $\hat{\lambda}$, evaluated numerically.

