University of Victoria
Department of Economics

Econometrics Working Paper EWP0908

ISSN 1485-6441

Bias of the Maximum Likelihood Estimators of the Two-Parameter Gamma Distribution Revisited

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

&

Hui Feng
Department of Economics, Business & Mathematics
King's University College, University of Western Ontario
London, ON, Canada N6A 2N3

September, 2009

Abstract

We consider the quality of the maximum likelihood estimators for the parameters of the two-parameter gamma distribution in small samples. We show that the methodology suggested by Cox and Snell (1968) can be used very easily to bias-adjust these estimators. A simulation study shows that this analytic correction is frequently much more effective than bias-adjusting using the bootstrap – generally by an order of magnitude in percentage terms. The two bias-correction methods considered result in increased variability in small samples, and the original estimators and their bias-corrected counterparts all have similar percentage mean squared

Keywords Maximum likelihood estimator; bias reduction; gamma distribution

Mathematics Subject Classification 62F10; 62F40; 62N02; 62N05

Author Contact:
David E. Giles, Dept. of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, B.C., Canada V8W 2Y2; e-mail: dgiles@uvic.ca; Phone: (250) 721-8540; FAX: (250) 721-6214

1. Introduction

The gamma, or Pearson (1895) Type III, distribution has been used to model a wide range of data types in many disciplines, especially in the context of reliability modeling, life testing and fatigue testing. For example, Birnbaum and Saunders (1958) introduced the gamma distribution for modeling the life-length of certain materials, and the use of this distribution for various reliability problems is noted by both Herd (1959) and Drenick (1960). Gupta and Groll (1961) discuss acceptance sampling based on this distribution, and they derive the operating characteristic function, producer's risk, failure rates and minimum sample sizes for this problem.

Empirical applications of the gamma distribution arise in a diverse range of fields. For example, Stoney (1988) and Wein and Bajeva (2005) applied this distribution in analyses of human fingerprint data. Segal *et al*. (2000) used it for matching scores in the context of DNA fingerprint genotyping of tuberculosis, and Keaton (1995) adopted it for an inventory control problem. The gamma distribution has also been applied in a number of studies in the fields of signal processing (*e.g*., Brehm and Stammler, 1987, Martin, 2002, Jensen *et al*., 2005, and Kim and Stern, 2008), hydrology (*e.g*., Bobbé and Ashkar, 1991, Stedinger *et al*., 1993, Askoy, 2000, and Bhunya *et al*., 2007) and meteorology (*e.g*., Thom, 1958, 1959, and Simpson, 1972).[1]

The density function for the gamma distribution, with shape and scale parameters α and θ respectively, is:

$$f(x) = \frac{x^{\alpha-1} e^{-x/\theta}}{\Gamma(\alpha)\theta^{\alpha}} \quad ; \quad \alpha, \theta > 0 \; ; \quad x > 0 \, . \tag{1}$$

It is also common to parameterize (1) in terms of the so-called "rate parameter", $\lambda = 1/\theta$, reflecting the fact that if α is integer-valued, the gamma distribution collapses to the Erlang distribution which describes the waiting-time until the α^{th} arrival in a Poisson process with rate parameter λ.[2]

Various estimators of the parameters of the gamma distribution have been used in practice. Historically, and notwithstanding the associated computational issues of the day, the maximum likelihood estimator (MLE) was an early choice because of its optimal asymptotic properties (*e.g.*, Masuyama and Kuroiwa, 1951; Cohen, 1953; Raj, 1953; Chapman, 1956; Greenwood and Durand, 1960; Gupta and Groll, 1961; Harter and Moore, 1965). Fisher (1922) demonstrated the inefficiency of the method of moments (MOM) estimator, relative to

---

[1] As Yue *et al*. (2001, p.1) note, "A univariate gamma distribution is one of the most commonly adopted statistical distributions in hydrological frequency analysis."

[2] It is well known that the gamma distribution collapses to the exponential distribution with rate parameter λ when $\alpha = 1$; to the chi square distribution with υ degrees of freedom when $\alpha = \upsilon/2$ and $\theta = 2$; and for large α it can be approximated by a normal distribution with mean $\alpha\theta$ and variance $\alpha\theta^2$.

the MLE, for this problem.[3] Given its widespread use, often with samples of only modest size, the small-sample properties of this MLE are of considerable interest. Early Monte Carlo evidence regarding these properties was provided by Choi and Wett (1969). They considered sample sizes of n = 40, 120, 200; and shape parameter values[4] of α = 1, 2, 3, 5, 7. Computational constraints limited their simulation experiment to only 100 replications. They found that the MLEs of α and θ are both upward-biased, and they also provided measures regarding the variances of these MLEs. Bowman and Shenton (1982) derive analytic approximations based on asymptotic expansions, to $O(n^{-6})$, for the first four moments of the MLEs for the gamma distribution parameters.[5] Their methodology is discussed briefly in the next section. Among other things, they provided extensive tables of the approximate bias, variance, skewness and kurtosis measures for these MLEs. They confirmed the signs of the biases noted by Choi and Wett (1969) and formalized by Berman (1981).

In this paper we show that the methodology of Cox and Snell (1968, pp.251-252) provides a very simple way of deriving the expressions for the biases (to $O(n^{-1})$) of the MLEs for the parameters of the gamma distribution. These expressions are identical to those reported by Bowman and Shenton (1982, p.390), and stated by Masuyama and Kuroiwa (1951) and Greenwood and Durand (1960). We then go further by investigating the effectiveness of "bias-adjusting" the MLEs by subtracting the *estimated* $O(n^{-1})$ bias. This does not appear to have been considered for this particular estimation problem previously. We also compare this approach with the "obvious", but computationally more burdensome, alternative of using the bootstrapped bias for the purpose of bias reduction. Both approaches achieve dramatic reductions in bias, but the analytic approach generally out-performs that based on the bootstrap.

The remainder of the paper proceeds as follows. In the next section we briefly summarize some of the analytic methods used by Bowman and Shenton, and discuss the simple results due to Cox and Snell (1968) that we use to obtain analytical expressions for the biases of the MLEs of the parameters in the gamma distribution. Section 3 derives these biases and the corresponding bias-adjusted MLEs. In section 4 a Monte Carlo experiment is used to compare the performance of the latter estimators with that of bias-adjusted estimators that use the bootstrap to determine finite-sample bias. An empirical example using real data is presented in section 5 to illustrate the application, and potential merits, of the analytic bias adjustment. Section 6 concludes.

[3] Bowman and Shenton (1982, p.391) note that the $O(n^{-1})$ bias expressions discussed in the present paper can be derived from the first-order covariance terms given by Fisher (1922).
[4] They set θ = 1, but as we show in section 3.1, their results will be invariant to the value of this parameter.
[5] Their results drew on earlier work in Bowman and Shenton (1968, 1970) and Shenton and Bowman (1977).

2. Approximating the bias of maximum likelihood estimators

We are concerned with obtaining analytic approximations to the bias of an MLE, even when the likelihood equations do not admit a closed-form solution.[6] We use a sample of n observations to estimate the p elements of the parameter vector, ϕ, of a distribution whose density is assumed to be regular with respect to all derivatives up to and including the third order. Let $l(\phi)$ be the associated log-likelihood function, satisfying the usual regularity conditions, and let $\hat{\phi}$ be the MLE of ϕ.

Not surprisingly, considerable attention has been paid to the problem of obtaining analytic approximations to the moments of MLEs, beginning with the derivation of the $O(n^{-1})$ bias when $p = 1$ by Bartlett (1953a).[7] Haldane and Smith (1956), derived expressions for the first four cumulants to this same order of accuracy, and Shenton and Bowman (1963) obtained higher-order approximations for the first four moments of an MLE. Bartlett (1953b) and Haldane (1953) explored the bias of an MLE when $p = 2$, and Shenton and Wallington (1962) and Cox and Snell (1968) derived formulae for the $O(n^{-1})$ bias of an MLE in the multi-parameter case.

Shenton and Bowman (1977, Chap.2) distinguish three general methods for approximating the moments of MLEs in terms of expansions using inverse powers of n. The first (*e.g.*, Haldane and Smith, 1956) involves applying Lagrange's expansion to a Taylor series approximation to $l(\hat{\phi})$. This method does not lend itself well to the case where $p \geq 2$. The second is the method that Shenton and Bowman have adopted in the bulk of their own work. It involves an indirect Taylor series approach that involves viewing $\hat{\phi}$ as a function of the proportions of the sample observations falling into each of (say) c classes. This method was used by Shenton and Bowman (1963), and it extends readily to the multi-parameter case. It has the additional merits that it can be used to determine higher-order approximations to higher-order moments (but at the cost of a very large number of tedious evaluations).

The third approach is described by Shenton and Bowman (1977, p.35) as "….an adjusted order of magnitude process", and they attribute it to Cox and Snell (1968, pp.251-252). They also claim that it has "….little advantage over the Lagrange expansion" (p.40), except that it can be applied in the multi-parameter case. The latter case is mentioned only briefly by Shenton and Bowman (1977, p.77). The method of Cox and Snell is, in fact, straightforward to apply when $p \geq 2$, especially if only an $O(n^{-1})$ approximation is desired. To see this we define the following cumulants:

[6] Fundamentally different approaches could also be considered – *e.g.*, see Firth (1993).
[7] Earlier related results were obtained by Tukey (1949).

$$k_{ij} = E(\partial^2 l / \partial\phi_i \partial\phi_j) \quad ; \quad i, j = 1, 2, \ldots, p \tag{2}$$

$$k_{ijl} = E(\partial^3 l / \partial\phi_i \partial\phi_j \partial\phi_l) \; ; \quad i, j, l = 1, 2, \ldots, p \tag{3}$$

$$k_{ij,l} = E[(\partial^2 l / \partial\phi_i \partial\phi_j)(\partial l / \partial\phi_l)] \; ; \quad i, j, l = 1, 2, \ldots, p \, . \tag{4}$$

and

$$k_{ij}^{(l)} = \partial k_{ij} / \partial\phi_l \quad ; \quad i, j, l = 1, 2, \ldots, p. \tag{5}$$

The expressions in (2) – (5) are assumed to be $O(n)$. Cox and Snell showed that when the sample data are independent (but not necessarily identically distributed) the bias of the s^{th} element of the MLE of ϕ ($\hat{\phi}$) is:

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

where k^{ij} is the $(i, j)^{\text{th}}$ element of the inverse of the (expected) information matrix, $K = \{-k_{ij}\}$. More recently, Cordeiro and Klein (1994) observed that (6) holds if the data are non-independent, and they showed that it can be re-written in the more convenient form:[8]

$$Bias(\hat{\phi}_s) = \sum_{i=1}^{p} k^{si} \sum_{j=1}^{p}\sum_{l=1}^{p} [k_{ij}^{(l)} - 0.5 k_{ijl}] k^{jl} + O(n^{-2}) \; ; \quad s = 1, 2, \ldots, p. \tag{7}$$

Define: $a_{ij}^{(l)} = k_{ij}^{(l)} - (k_{ijl} / 2)$, for $i, j, l = 1, 2, \ldots, p$

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

$$A = [A^{(1)} \,|\, A^{(2)} \,|\, \ldots\ldots \,|\, A^{(p)}].$$

Cordeiro and Klein (1994) showed that the bias of $\hat{\theta}$ can be re-written as:

$$Bias(\hat{\phi}) = K^{-1} A \, vec(K^{-1}) + O(n^{-2}) \, , \tag{8}$$

and a "bias-corrected" MLE for θ is defined as:

$$\tilde{\phi} = \hat{\phi} - \hat{K}^{-1} \hat{A} \, vec(\hat{K}^{-1}) \, , \tag{9}$$

where $\hat{K} = (K)|_{\hat{\phi}}$ and $\hat{A} = (A)|_{\hat{\phi}}$. It can be shown that the bias of $\tilde{\phi}$ is $O(n^{-2})$. Clearly, (8) and (9) can be evaluated *even when the likelihood equation does not admit a closed-form analytic solution,* so that the MLE has to be obtained by numerical methods.

---

[8] The computational advantage of (7) is that it does not involve terms of the form in (4).

The Cox-Snell/Cordeiro-Klein methodology has recently been applied successfully to a range of problems by various authors, including Cordeiro and McCullagh (1991), Cordeiro *et al*. (1996), Cribari-Neto and Vasconcellos (2002), Giles (2009) and Giles and Feng (2009).

3. Biased-adjusted MLEs for the gamma distribution

In a series of studies, Bowman and Shenton (1968, 1969, 1982, 1988) and Shenton and Bowman (1970, 1972) considered the second type of asymptotic expansion (in inverse powers of n) noted in section 2 to derive analytic approximations to the first four moments of the MLEs for α and θ in (1). In addition, Shenton and Bowman (1969) and Bowman and Shenton (1970, 1982) considered asymptotic expansions in terms of descending powers of the parameter α, in order to obtain alternative approximations to the moments of the MLEs for the parameters of the gamma distribution. However, we do not pursue the latter type of expansion here as our concern is with small sample asymptotics.

In what follows, we consider two separate cases – one where the distribution is characterized by the shape and scale parameters (α and θ), and one where it is characterized by the shape and rate parameters (α and $\lambda = 1/\theta$). In each case, we show that the Cox-Snell/Cordeiro-Klein methodology quickly provides the biases of the MLEs to $O(n^{-1})$, and we construct the associated bias-adjusted estimators.

3.1 Shape and scale parameters

From (1), the log-likelihood function, based on a sample of n independent observations, is

$$l = (\alpha - 1)\sum_{i=1}^{n}\log(y_i) - (\sum_{i=1}^{n} y_i)/\theta - n[\log(\Gamma(\alpha)) + \alpha\log(\theta)]. \tag{10}$$

We then have:

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{n}\log(y_i) - n[\Psi(\alpha) + \log(\theta)] \tag{11}$$

$$\frac{\partial l}{\partial \theta} = [\sum_{i=1}^{n} y_i - n\alpha\theta]/\theta^2 \tag{12}$$

where $\Psi(\alpha)$ is the usual digamma function:

$$\Psi(\alpha) = d\log\Gamma(\alpha)/d\alpha = -\gamma - \sum_{k=1}^{\infty}\varsigma(k+3)(-(\alpha-1))^k ,$$

where $\gamma = \underset{n\to\infty}{Limit}[(\sum_{k=1}^{\infty}\frac{1}{k}) - \ln(n)] = 0.57721.....$ is the Euler-Mascheroni constant, and $\varsigma(s) = \sum_{n=1}^{\infty}(n^{-s})$ is the Riemann zeta function. In what follows we also need the trigamma and tetragamma functions, these being $\Psi_{(i)}(\alpha) = d^i\Psi(\alpha)/d\alpha^i$; i = 1, 2. These functions have various representations that are convenient for numerical evaluation.

From equations (11) and (12) it is clear that the maximum likelihood estimates of the parameters cannot be obtained analytically as solutions to the likelihood equations. Numerical methods must be used, and early discussions of some of the associated issues are provided by Gupta and Groll (1961) and by Choi and Wette (1969).

Noting that $E(y_i) = \alpha\theta$, the following results emerge immediately:

$$k_{11} = -n\Psi_{(1)}(\alpha)$$

$$k_{12} = -n/\theta$$

$$k_{22} = -n\alpha/\theta^2$$

$$k_{111} = k_{11}^{(1)} = -n\Psi_{(2)}(\alpha)$$

$$k_{112} = k_{121} = k_{211} = k_{11}^{(2)} = k_{12}^{(1)} = 0$$

$$k_{122} = k_{221} = k_{212} = k_{12}^{(2)} = -k_{22}^{(1)} = n/\theta^2$$

$$k_{222} = 4n\alpha/\theta^3$$

$$k_{22}^{(2)} = 2n\alpha/\theta^3 .$$

The information matrix is

$$K = n\begin{bmatrix} \Psi_{(1)}(\alpha) & 1/\theta \\ 1/\theta & \alpha/\theta^2 \end{bmatrix},$$

and

$$A = (n/2)\begin{bmatrix} -\Psi_{(2)}(\alpha) & 0 & 0 & 1/\theta^2 \\ 0 & -3/\theta^2 & 1/\theta^2 & 0 \end{bmatrix}.$$

So, to $O(n^{-1})$,

$$Bias(\hat{\alpha}) = \frac{[\alpha(\Psi_{(1)}(\alpha) - \alpha\Psi_{(2)}(\alpha)) - 2]}{2n[\alpha\Psi_{(1)}(\alpha) - 1]^2} \tag{13}$$

and

$$Bias(\hat{\theta}) = \frac{\theta[\alpha\Psi_{(2)}(\alpha) + \Psi_{(1)}(\alpha)]}{2n[\alpha\Psi_{(1)}(\alpha) - 1]^2} \ . \tag{14}$$

These bias expressions are identical to those in equation (4.23) of Bowman and Shenton (1982, p.390).[9] The bias of $\hat{\alpha}$, and the percentage biases of both $\hat{\alpha}$ and $\hat{\theta}$, are invariant to the value of θ (and hence λ). In addition, $\hat{\alpha}$ is upward-biased, and $\hat{\theta}$ is downward-biased, to $O(n^{-1})$, consistent with the analytic results of Shenton and Bowman (1977, pp.151-153) and Berman (1981), and the Monte Carlo results of Choi and Wette (1969). Bias-adjusted estimators are then obtained as

$$\tilde{\alpha} = \hat{\alpha} - \frac{[\hat{\alpha}(\Psi_{(1)}(\hat{\alpha}) - \hat{\alpha}\Psi_{(2)}(\hat{\alpha})) - 2]}{2n[\hat{\alpha}\Psi_{(1)}(\hat{\alpha}) - 1]^2} \tag{15}$$

and

$$\tilde{\theta} = \hat{\theta} - \frac{\hat{\theta}[\hat{\alpha}\Psi_{(2)}(\hat{\alpha}) + \Psi_{(1)}(\hat{\alpha})]}{2n[\hat{\alpha}\Psi_{(1)}(\hat{\alpha}) - 1]^2} \ . \tag{16}$$

Apparently, the sampling properties of such bias-adjusted estimators have not been evaluated previously, so the evidence that we present in section 4 extends Shenton and Bowman's results in this direction.

3.2 Shape and rate parameters

Re-parameterizing the gamma distribution in terms of the rate parameter, $\lambda = 1/\theta$, the log-likelihood function is

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

and it is easily verified that

$$k_{11} = -n\Psi_{(1)}(\alpha)$$

$$k_{12} = n/\lambda$$

[9] Also, see Masuyama and Kuroiwa (1951) and Greenwood and Durand (1960).

$$k_{22} = -n\alpha / \lambda^2$$

$$k_{111} = k_{11}^{(1)} = -n\Psi_{(2)}(\alpha)$$

$$k_{112} = k_{121} = k_{211} = k_{11}^{(2)} = k_{12}^{(1)} = 0$$

$$k_{122} = k_{221} = k_{212} = k_{22}^{(1)} = k_{12}^{(2)} - n / \lambda^2$$

$$k_{222} = k_{22}^{(2)} = 2n\alpha / \lambda^3$$

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

$$A = (n/2)\begin{bmatrix} -\Psi_{(2)}(\alpha) & 0 & 0 & -1/\lambda^2 \\ 0 & -1/\lambda^2 & -1/\lambda^2 & 2\alpha/\lambda^3 \end{bmatrix},$$

and so

$$Bias(\hat{\alpha}) = \frac{[\alpha(\Psi_{(1)}(\alpha) - \alpha\,\Psi_{(2)}(\alpha)) - 2]}{2n[\alpha\,\Psi_{(1)}(\alpha) - 1]^2} + O(n^{-2})$$

as in equation (13), and

$$Bias(\hat{\lambda}) = \frac{\lambda[2\alpha\,(\Psi_{(1)}(\alpha))^2 - 3\Psi_{(1)}(\alpha) - \alpha\Psi_{(2)}(\alpha)]}{2n[\alpha\,\Psi_{(1)}(\alpha) - 1]^2} + O(n^{-2})\ . \qquad (17)$$

A bias-adjusted estimator for λ is

$$\tilde{\lambda} = \hat{\lambda} - \frac{\hat{\lambda}[2\alpha\,(\Psi_{(1)}(\hat{\alpha}))^2 - 3\Psi_{(1)}(\hat{\alpha}) - \hat{\alpha}\Psi_{(2)}(\hat{\alpha})]}{2n[\hat{\alpha}\,\Psi_{(1)}(\hat{\alpha}) - 1]^2}\ . \qquad (18)$$

Clearly, the percentage bias of $\hat{\lambda}$ is invariant to the value of λ (and hence θ). It is readily shown that the numerator expression in (18) is positive for all $\alpha > 0$, so $\hat{\lambda}$ is unambiguously upward-biased, to $O(n^{-2})$, which is again consistent with the earlier results of Choi and Wette (1969) and Berman (1981).

4. Simulation results

The bias expressions in (13), (14) and (17) are valid only to $O(n^{-1})$.[10] The actual bias and mean squared error (MSE) of the maximum likelihood and bias-corrected maximum likelihood estimators are now compared in a Monte Carlo experiment. The maximum likelihood estimates were obtained using the Nelder-Mead algorithm in the *maxLik* package (Toomet and Henningsen, 2008) for the *R* statistical software environment (R, 2008). The *R* software also includes routines for generating gamma-distributed random variates, and for evaluating the digamma, trigamma, and tetragamma functions. An excellent recent discussion of the issues and difficulties associated with accurately computing these functions is provided by Ho *et al*. (2009).[11]

In addition to $\hat{\alpha}$, $\tilde{\alpha}$, $\hat{\theta}$, $\tilde{\theta}$, $\hat{\lambda}$ and $\tilde{\lambda}$, we have also considered the bootstrap-bias-corrected estimator. In the case of α, this is obtained as $\breve{\alpha} = 2\hat{\alpha} - (1/N_B)[\sum_{j=1}^{N_B} \hat{\alpha}_{(j)}]$, where $\hat{\alpha}_{(j)}$ is the MLE of α obtained from the j^{th} of the N_B bootstrap samples. Corresponding expressions apply for the estimators of the other two parameters. See Efron (1982, p.33). Bootstrap-bias-corrected estimators are also unbiased to $O(n^{-2})$, but it is known that that this generally comes at the expense of increased variance.

Each part of the experiment relating to the MLE's and Cox-Snell biased-corrected MLE's uses 100,000 Monte Carlo replications. In the case of the bootstrap-corrected estimators the number of Monte Carlo replications is 100,000, with 1,000 bootstrap samples *per* replication. The results that are reported in Tables 1 are *percentage* biases and MSE's. The latter are defined as $100 \times (\text{MSE} / \alpha^2)$, *etc*., and appear in square brackets below the corresponding percentage biases.

As the relative biases and mean squared errors are invariant to the value of the scale parameter, θ, and rate parameter, λ, these parameters are assigned the value unity in this experiment. Various values of the shape parameter, α, are considered, including ones that are consistent with some of the empirical studies discussed in section 1. The simulation results for $\hat{\alpha}$ and $\hat{\theta}$ in Table 1 agree very closely with the analytic results, given to $O(n^{-6})$, in Tables 17 and 21 of Bowman and Shenton (1982), and in particular the directions of all of the biases are as anticipated from section 3. The orders of magnitude of the percentage biases and MSEs are relatively invariant to the values we have considered for the scale parameter, α.

---

[10] It will be recalled, however, that Bowman and Shenton (1982) provide approximations to $O(n^{-6})$.

[11] Our numerical evaluations of the relative biases based on (13) and (14) exactly match those in the 'n^{-1}' columns of Tables 17 and 21 of Bowman and Shenton (1982), to the four decimal places reported by those authors.

We see that the MLEs, $\hat{\alpha}$, $\hat{\theta}$ and $\hat{\lambda}$ can exhibit considerable percentage bias when the sample size is small. As the sample size increases, the biases and mean squared errors of these MLEs fall, reflecting the consistency of the estimators. The analytic bias adjustment defined in (15) – (17) (resulting in the estimators $\tilde{\alpha}$, $\tilde{\theta}$ and $\tilde{\lambda}$) is extremely successful, often reducing the percentage bias by as much as two orders of magnitude in very small samples. There is at least one order of magnitude improvement in percentage bias even when n = 250. However, this comes at the expense of increased variance, as is seen from the fact that the MSEs of the original MLEs and these bias-adjusted MLEs are generally of the same order of magnitude, and in fact often very close in value. The consistency of these bias-adjusted estimators is clear from the (overall, but not necessarily monotonic) decline in their biases and MSEs as the sample size increases.

Bias-adjusting using the bootstrap also reduces the biases of the MLEs. Often an order of magnitude improvement can be obtained in percentage terms for small sample sizes, but again this comes at the expense of increased variability in the estimators. Generally, the original MLEs and both the analytically bias-adjusted and bootstrap bias-adjusted MLEs have similar percentage MSEs, so any choice between the two methodologies may be based on the resultant biases. The overall thrust of the results in Table 1 is that the analytical Cox-Snell/Cordeiro-Klein methodology out-performs the use of the bootstrap for bias reduction in almost all of the cases considered, and especially if the sample size is less than n = 100. The main exceptions arise in parts (a) and (b) of Table 1 when estimating the shape parameter (θ) if the scale parameter exceeds unity. Even in these cases, however, analytical bias adjustment performs extremely well, and may be preferred on the grounds of simplicity of application.

5. Illustrative example

Simpson (1972) fitted seven distributions, including the gamma distribution, to two small samples of data measuring the fourth root of rainfall (in acre-feet) from cumulus clouds in Florida. One sample relates to clouds that were "seeded" to induce rainfall, while the other is for a control group of unseeded clouds. In each case n = 26. Simpson found that the gamma distribution is the preferred model, on the basis of several criteria. In her parameterization of the gamma model, Simpson's parameters B and C correspond to our (α-1) and λ respectively. Using the same software and algorithm as in section 4, we have duplicated Simpson's MLE's and computed our bias-adjusted estimators. These results appear in the upper part of Table 2.

We know from section 3 that the MLE's of both the shape and rate parameters are upward biased, which accounts for the relative magnitudes of the MLE's and their bias-adjusted counterparts in Table 2. Adjusting the parameter estimates for the first-order bias has implications for the fitted gamma model, and in the lower part of Table 2 these are summarized *via* the changes in the mean, variance, mode, and coefficient of variation (CV) of the distribution. Simpson is especially interested in the latter descriptive statistic.

Simpson (1972, pp.310-311) reported that the seeding of the clouds reduces the CV associated with the fitted gamma distribution only slightly.[12] In fact there is a reduction of 4.34% as we move from the model estimated from the control data to the model estimated from the seeded data. These calculations are based on the (unadjusted) MLE's of the parameters in each case. Using the results for the bias-adjusted MLE's, as given in Table 2, the corresponding reduction in the CV as estimated to be 4.10%. Interestingly, Simpson (1972, p.311, fn.1) observes that the *sample* coefficient of variation is 0.43 for the "unseeded" sample and 0.37 for the "seeded" sample of data. From the lower part of Table 2 we see that, in this application, bias-adjusting the MLE's of the parameters of the gamma distribution results in a fitted model that captures certain key characteristics of the control data more accurately than is the case if no adjustment for bias is made. However, the converse is true for the model fitted to the "seeded" data.

6. Conclusions

The two-parameter gamma distribution is very widely used in many disciplines, and so there is considerable interest in the quality of the maximum likelihood estimators for its parameters in small samples. In this paper we have shown how the methodology suggested by Cox and Snell (1968) can be used very easily to construct a closed-form adjustment to these MLEs that corrects for the $O(n^{-1})$ bias. This analytic bias correction, wich has not previously been evaluated, is found to be much more effective than bias-adjusting using the bootstrap – generally by an order of magnitude in percentage terms. Not surprisingly, the two bias-correction methods compared in this paper result in increased variability in small samples, the net effect being that the original MLEs and their bias-corrected counterparts all have similar percentage mean squared errors. The use of the Cox-Snell bias correction is strongly recommended when estimating the parameters of the gamma distribution by maximum likelihood.

Acknowledgement

The second author acknowledges financial support for this research from King's College at the University of Western Ontario, We are also grateful to Ryan Godwin for several helpful discussions, and to Lief Bluck for arranging the necessary computational support.

---

[12] Three typesetting errors in Simpson (1972) paper should be noted. In equation (5) on p.310, the correct results (to the reported level of accuracy) are $\alpha = 6.52$, $V = 0.39$ and $\langle R \rangle = \alpha/\beta = 2.93$.

References

Askoy, H. (2000), "Use of gamma distribution in hydrological analysis", *Turkish Journal of Engineering and Environmental Science*, 24, 419-428.

Bartlett, M. S. (1953a), "Approximate confidence intervals", *Biometrika*, 40, 12-19.

Bartlett, M. S. (1953b), "Approximate confidence intervals II. More than one unknown parameter", *Biometrika*, 40, 306-317.

Berman, M. (1981), "The maximum likelihood estimators of the parameters of the gamma distribution are always positively biased", *Communications in Statistics*, *A*, 10, 693-697

Bhunya, P. K., R. Berndtsson, C. S. P. Ojha and S. K. Mishra (2007), "Suitability of gamma, chi-square, Weibull, and beta distributions as synthetic unit hydrographs", *Journal of Hydrology*, 334, 28-38.

Birnbaum, Z. W. and S. C. Saunders (1958), "A statistical model for life lengths of materials", *Journal of the American Statistical Association*, 53, 153-160.

Bobbé, B. and F. Ashkar, F. (1991), *The Gamma Family and Derived Distributions Applied in Hydrology*, Water Resources Publications, Littleton, CO.

Bowman, K. O. and L. R. Shenton (1968), "Properties of estimators for the gamma distribution", Report CTC-1, Union Carbide Corporation, Nuclear Division.

Bowman, K. O. and L. R. Shenton (1969), "Remarks on maximum likelihood estimators for the gamma distribution", *International Conference on Quality Control (Toyko) Proceedings*, 519-22.

Bowman, K. O. and L. R. Shenton (1970), "Small sample properties of estimators for the gamma distribution", Report CTC-28, Union Carbide Corporation, Nuclear Division.

Bowman, K. O. and L. R. Shenton (1982), "Properties of estimators for the gamma distribution", *Communications in Statistics*, *B*, 11, 377-519.

Bowman, K. O. and L. R. Shenton (1988), *Properties of Estimators for the Gamma Distribution*, Marcel Dekker, New York.

Brehm, H. and W. Stammler (1987), "Description and generation of spherically invariant speech-model signals", *Signal Processing*, 12, 119-141.

Chapman, D. G. (1956), "Estimating the parameters of a truncated gamma distribution", *Annals of Mathematical Statistics*, 27, 498-506.

Choi, S. C. and R. Wette (1969), "Maximum likelihood estimation of the parameters of the gamma distribution and their bias", *Technometrics*, 11, 683-690.

Cohen, A. C. (1950), "Estimating parameters of Pearson Type III populations from truncated samples", *Journal of the American Statistical Association*, 45, 411-423.

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

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

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

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

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

Drenick, R. F. (1960), "Mathematical aspects of the reliability problem", *Journal of the Society for Industrial and Applied Mathematics*, 8, 125-149.

Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia, PA.

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

Fisher, R. A. (1922), "On the mathematical foundations of theoretical statistics", *Philosophical Transactions of the Royal Society of London*, *A*, 222, 309-368.

Giles, D. E. A. (2009), "Bias reduction for the maximum likelihood estimator of the scale parameter in the half-logistic distribution", Econometrics Working Paper EWP0901, Department of Economics, University of Victoria.

Giles, D. E. A. and H. Feng (2009), "Bias - corrected maximum likelihood estimation of the parameters of the generalized Pareto distribution", Econometrics Working Paper EWP0902, Department of Economics, University of Victoria.

Greenwood, J. A. and D. Durand (1960), "Aids for fitting the gamma distribution by maximum likelihood", *Technometrics*, 2, 55-65.

Gupta, S. S. and P. A. Groll (1961), "Gamma distribution in acceptance sampling based on life tests", *Journal of the American Statistical Association*, 56, 942–970.

Haldane, J. B. S. (1953), "The estimation of two parameters from a sample", *Sankhyā*, 12, 313-320.

Haldane, J. B. S. and S. M. Smith (1956), "The sampling distribution of a maximum likelihood estimate", *Biometrika*, 43, 96-103.

Harter, H. L. and A. H. Moor (1965), "Maximum likelihood estimation of the parameters of gamma and Weibull populations from complete and from censored samples", *Technometrics*, 7, 639-643.

Herd, G. R. (1959), “Some statistical concepts and techniques for reliability analyses and prediction”, *Proceedings from Fifth National Symposium on Reliability Control*, Philadelphia PA, 126-136.

Ho, L-P., G-N. Wang and J-C. Chern (2009), “Development and computation of the psi function and its derivatives”, *Proceedings of the 2009 Hawaiian International Conference on Statistics, Mathematics and Related Fields*, Honolulu HI, 400-407.

Jensen, J., I. Batina, R. C. Hendriks and R. Heusdens (2005), “A study of the distribution of time-domain speech samples and discrete Fourier coefficients”, *Proceedings of SPS-DARTS 2005* (*The first annual IEEE BENELUX/DSP Valley Signal Processing Symposium*), 155-158.

Keaton, M. (1995), “Using the gamma distribution to model demand when lead time is random”, *Journal of Business Logistics*, 16, 107-131.

Kim, C. and R. M. Stern (2008), “Robust signal-to-noise ratio estimation based on waveform amplitude distribution analysis”, *Proceedings of the Interspeech 2008 Conference*, Brisbane, Australia, 2598-2601.

Martin, R. (2002), “Speech enhancement using mmse short time spectral estimation with gamma distributed speech priors”, in *Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing*, I, Orlando FL, 253–256.

Masuyama, M. and Y. Kuroiwa (1951), “Table for the likelihood solutions of gamma distribution and its medical applications”, *Reports of Statistical Applications, Union of Japanese Scientists and Engineers*, 7, 18-23.

Pearson, K. (1895), “Contributions to the mathematical theory of evolution. - II. Skew variation in homogeneous material”, *Philosophical Transactions of the Royal Society of London*, *A*, 186, 343-414.

R (2008), The R Project for Statistical Computing, http://www.r-project.org

Raj, D. (1953), “Estimating the parameters of Type III populations from truncated samples”, *Journal of the American Statistical Association*, 48, 336-349.

Segal, M. R., H. Salamon and P. M. Small (2000), “Comparing DNA fingerprints of infectious organisms”, *Statistical Science*, 15, 27-45.

Shenton, L. R. and K. Bowman (1963), “Higher moments of a maximum-likelihood estimate”, *Journal of the Royal Statistical Society, B*, 25, 305-317.

Shenton, L. R. and K. O. Bowman (1969), “Maximum likelihood estimator moments for the 2-parameter gamma distribution”, *Sankhyā*, *B*, 379-396.

Shenton, L. R. and K. O. Bowman (1970), “Remarks on Thom’s estimators for the gamma distribution”, *Monthly Weather Review*, 98, 154-161.

Shenton, L. R. and K. Bowman (1972), “Further remarks on maximum likelihood estimators for the gamma distribution”, *Technometrics*, 14, 725-733.

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

Shenton, L. R. and P. A. Wallington (1962), “The bias of moment estimators with an application to the negative binomial distribution”, *Biometrika*, 49, 193-204.

Simpson, J. (1972), “Use of the gamma distribution in single-cloud rainfall analysis”, *Monthly Weather Review*, 100, 309-312.

Stedinger, J. R., R. M. Vogel and E. Foufoula-Georgiou (1993), “Frequency analysis of extreme events”, in D. R. Maidment (ed.), *Handbook of Hydrology*, McGraw-Hill, New York, chapter 18.

Stoney, D. A. (1988), “Distribution of epidermal ridge minutiae”, *American Journal of Physical Anthropology*, 77, 367-376

Thom, H. C. S. (1957), “A statistical method of evaluating augmentation of precipitation by cloud seeding”, Final Report, U. S. Advisory Committee on Weather Control, Vol. 2, Washington, D. C., 5-25.

Thom, H. C. S. (1958), “A note on the gamma distribution”, *Monthly Weather Review*, 86, 117-122.

Toomet, O. and A. Henningsen (2008), maxLik: Maximum likelihood estimation, http://CRAN.R-project.org ; http://maxLik.org

Tukey, J. W. (1949), “Standard confidence points”, Memorandum Report No. 26, Statistical Research Group, Stanford University.

Wein, L. M. and M. Baveja (2005), “Using fingerprint image quality to improve the identification performance of the U.S. visitor and immigrant status indicator technology program”, *Proceedings of the National Academy of Sciences*, 102, 7772-7775.

Yue, S., T. B. M. J. Ouarda and B. Bobée (2001), “A review of bivariate gamma distributions for hydrological application”, *Journal of Hydrology*, 246, 1-18.

Table 1: Percentage bias [and percentage MSE] of the MLEs and bias-corrected MLEs

n	$\hat{\alpha}$	$\tilde{\alpha}$	$\breve{\alpha}$	$\hat{\theta}$	$\tilde{\theta}$	$\breve{\theta}$	$\hat{\lambda}$	$\tilde{\lambda}$	$\breve{\lambda}$
					(a): α = 0.5				
10	30.3087	0.0503	-16.9439	-8.3511	-0.9461	-0.0157	62.7715	0.0099	-47.5312
	[55.0804]	[23.7143]	[29.9336]	[33.0258]	[37.2629]	[37.9863]	[220.2250]	[81.5503]	[332.9895]
15	17.7896	0.0373	-3.6332	-5.4226	-0.3270	-0.1352	35.8608	-0.0249	-9.9794
	[21.9830]	[12.5042]	[12.4549]	[22.1354]	[24.0395]	[24.3807]	[75.5345]	[37.6372]	[41.4609]
25	9.7018	-0.0595	-0.8303	-3.2462	-0.1188	-0.0426	19.3239	0.0075	-2.0881
	[9.0589]	[6.4722]	[6.4897]	[13.3633]	[14.0513]	[13.9947]	[27.9439]	[18.0135]	[18.0833]
50	4.5579	-0.0357	-0.1056	-1.6526	-0.0630	-0.0125	8.9467	-0.0074	-0.3896
	[3.4794]	[2.9360]	[2.9404]	[6.6720]	[6.8424]	[6.8390]	[9.6422]	[7.6525]	[7.7030]
100	2.2000	-0.0316	-0.0379	-0.7999	0.0016	-0.0174	4.3164	-0.0026	-0.0011
	[1.5339]	[1.4088]	[1.4149]	[3.3656]	[3.4091]	[3.3863]	[4.0530]	[3.6001]	[2.6861]
250	0.8721	-0.0058	0.0070	-0.3705	-0.0485	0.0477	1.7449	0.0524	-0.0343
	[0.5697]	[0.5505]	[0.5521]	[1.3473]	[1.3539]	[1.3500]	[1.4590]	[1.3886]	[1.3703]
					(b): α = 1.0				
10	33.1554	0.1167	-21.0180	-9.3635	-1.1486	-0.4251	50.1401	0.1073	-31.5996
	[72.2336]	[29.7664]	[39.3795]	[24.7572]	[28.0324]	[28.2438]	[135.6396]	[53.4702]	[81.2329]
15	20.4645	0.0127	-4.6131	-6.0828	-0.3954	-0.3030	28.9366	-0.0886	-7.0535
	[27.0769]	[14.8398]	[15.0003]	[16.6527]	[18.1463]	[18.3550]	[49.6743]	[26.1654]	[26.7053]
25	11.1739	0.0029	-1.0784	-3.7252	-0.2206	-0.1569	15.8077	0.0037	-1.5510
	[10.9318]	[7.5679]	[7.5833]	[10.0178]	[10.5514]	[10.5860]	[19.2108]	[12.8536]	[13.0206]
50	5.2080	-0.0252	-0.1159	-1.8724	-0.0839	-0.0828	0.0569	0.0207	-0.1999
	[4.1068]	[3.4064]	[3.4412]	[5.0491]	[5.1835]	[5.2638]	[7.0100]	[5.6886]	[5.6997]
100	2.4938	-0.0428	-0.0779	-0.8443	0.0599	0.0103	3.5168	-0.0551	-0.0011
	[1.7757]	[1.6166]	[1.6247]	[2.5651]	[2.6011]	[2.5883]	[2.9996]	[2.7008]	[2.6861]
250	0.9648	-0.0318	-0.0050	-0.3199	0.0439	-0.30037	1.3562	-0.0458	-0.0146
	[0.6530]	[0.6290]	[0.6334]	[1.0182]	[1.0240]	[1.0117]	[1.0872]	[1.0425]	[1.0458]

Table 1: Percentage bias [and percentage MSE] of the MLEs and bias-corrected MLEs (continued)

n	$\hat{\alpha}$	$\tilde{\alpha}$	$\breve{\alpha}$	$\hat{\theta}$	$\tilde{\theta}$	$\breve{\theta}$	$\hat{\lambda}$	$\tilde{\lambda}$	$\breve{\lambda}$
					(c): α = 5.0				
10	40.5278	-0.3381	-21.3998	-9.7335	-0.7658	-1.2637	43.3039	-0.4092	-22.9995
	[92.7209]	[37.4035]	[47.9748]	[19.8566]	[22.8109]	[22.8425]	[103.0552]	[41.2993]	[54.8719]
15	23.6317	-0.2339	-5.5794	-6.4325	-0.2317	-5.9729	25.2575	-0.2729	-6.1081
	[34.3966]	[18.4480]	[18.9571]	[13.4550]	[14.8079]	[21.2819]	[38.3105]	[20.4317]	[21.0530]
25	12.8723	-0.1564	-1.2765	-3.7893	0.0381	-1.3759	13.7541	-0.1826	-1.3759
	[13.6023]	[9.2535]	[9.4863]	[8.2193]	[9.2535]	[10.5580]	[15.2001]	[10.3045]	[10.5580]
50	6.0525	-0.0529	-0.3949	-1.9175	0.0341	0.0991	6.4790	-0.0525	-0.4090
	[5.0751]	[4.1614]	[4.0927]	[4.1412]	[4.2679]	[4.2062]	[5.6620]	[4.6315]	[4.5541]
100	2.9164	-0.0138	0.0070	-0.9551	0.0305	-0.0729	3.1382	-0.0273	-0.0078
	[2.1855]	[1.9727]	[1.9823]	[2.0819]	[2.1138]	[2.0120]	[2.4334]	[2.1986]	[2.1881]
250	1.1155	-0.0463	0.0169	-0.3599	0.0368	-0.0246	1.2058	-0.0372	0.0171
	[0.7980]	[0.7669]	[0.7639]	[0.8293]	[0.8346]	[0.8352]	[0.8834]	[0.8481]	[0.8475]
					(d): α = 10.0				
10	41.2198	-0.4895	-18.4343	-9.7259	-0.7134	-1.0217	42.6328	-0.5039	-18.9922
	[91.9345]	[36.7292]	[51.7588]	[19.3936]	[22.3072]	[22.1556]	[97.3581]	[38.7994]	[54.747]
15	24.1024	-0.2806	-5.6652	-6.4683	-0.2421	-1.0217	24.9493	-0.2713	-0.4823
	[35.3445]	[18.9053]	[19.4916]	[13.1466]	[14.4746]	[22.1556]	[37.3506]	[19.9202	[14.4479]
25	13.1293	-0.1838	-1.2228	-3.7811	0.0624	-0.2527	13.5839	-0.1843	-1.3089
	[14.1745]	[9.6429]	[9.7256]	[8.0392]	[8.5380]	[8.3875]	[15.0086]	[10.1936]	[10.243]
50	6.1853	-0.0546	-0.2556	-1.9489	0.0096	0.0166	6.4123	-0.0414	-0.3387
	[5.2329]	[4.2859]	[4.3170]	[4.0407]	[4.1638]	[4.1584]	[5.5386]	[4.5305]	[4.5650]
100	2.9903	-0.0339	0.0363	-0.9669	0.0222	-0.0971	3.1049	-0.0227	-0.0491
	[2.2632]	[2.0454]	[2.0406]	[2.0353]	[2.0665]	[2.0556]	[2.3839]	[2.1523]	[2.1277]
250	1.1410	-0.0465	0.0085	-0.0366	0.0342	-0.0042	1.1950	-0.0331	-0.0387
	[0.8229]	[0.7906]	[0.7960]	[0.7918]	[0.8210]	[0.8207]	[0.8674]	[0.8328]	[0.8341]

Table 2: Estimates of the gamma model for Simpson's rainfall data

Seeded Clouds				Unseeded Clouds			
$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\lambda}$	$\tilde{\lambda}$	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\lambda}$	$\tilde{\lambda}$
7.104	6.309	1.831	1.616	6.523	5.795	2.224	1.962
(1.988)	[2.103]	(0.529)	[0.536]	(1.658)	[1.904]	(0.592)	[0.725]

MLE				Bias-Corrected MLE				MLE				Bias-Corrected MLE			
Mean	Var.	Mode	CV	Mean	Var.	Mode	CV	Mean	Var.	Mode	CV	Mean	Var.	Mode	CV
3.880	2.119	3.334	0.375	3.904	2.416	3.285	0.398	2.933	1.319	2.483	0.392	2.954	1.505	2.444	0.415

Asymptotic standard errors appear in parentheses, and bootstrapped standard errors (based on 1000 re-samples) appear in square brackets.