A Note on Bias of Closed-Form Estimators for the Gamma Distribution Derived from Likelihood Equations

Francisco Louzada, Pedro L. Ramos, Eduardo Ramos

Institute for Mathematical Sicences and Computing

Universidade de São Paulo, São Carlos, Brazil

April 20, 2021

Abstract

We discuss here an alternative approach for decreasing the bias of the closed-form estimators for the gamma distribution recently proposed by Ye and Chen [4]. We show that, the new estimator has also closed-form expression, is positive and can be computed for $n > 2$. Moreover, the corrective approach returns better estimates when compared with the former ones.

Keywords: bias correction, improve estimators, two-parameter gamma distribution.

1 Introduction

We read with interest the recent paper by Ye and Chen [4], in which they provide a new simple method to derive closed-form estimators for the parameters of the gamma distribution. Although the authors propose bias corrected estimators, the bias corrected estimator for the shape parameter continues to have a systematic bias in presence of small samples. Particularly, for the ones with less than 20 elements. So, bias correction for this parameter needs to be further improved. Here, we propose two improved bias corrected estimators which outperform the one proposed by the authors under such circumstances.

Let X be a non negative random variable with a gamma probability density function (p.d.f) given by

$$f(t|\alpha,\beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} t^{\alpha-1}\exp\left(-\frac{t}{\beta}\right),$$

where $\alpha > 0$ and $\beta > 0$ are the shape and scale parameters and $\Gamma(\alpha) = \int_0^{\infty} e^{-x}x^{\alpha-1}dx$ is the gamma function. The hybrid maximum likelihood estimators are given by

$$\hat{\alpha}_{YC} = \frac{n\sum_{i=1}^{n} t_i}{\left(n\sum_{i=1}^{n} t_i \log\left(t_i\right) - \sum_{i=1}^{n} t_i \sum_{i=1}^{n}\log\left(t_i\right)\right)} \tag{1}$$

and

$$\hat{\beta}_{YC} = \frac{1}{n^2}\left(n\sum_{i=1}^{n} t_i \log\left(t_i\right) - \sum_{i=1}^{n} t_i \sum_{i=1}^{n}\log\left(t_i\right)\right). \tag{2}$$

Firstly we derive a closed expression for the bias correction using Cox and Snell [2] approach as presented in Section 2. Secondly, still in this section, we present a useful closed form approximation for the first bias corrected estimator that does not depend on transcendental functions and can be calculated without computational cost. It is proved that new estimator is always positive for $n > 2$. In Section 3, an simulation study is conducted to show that our approach outperforms Ye and Chen's bias corrected estimator in terms of mean square error (MSE) and mean relative error (MRE), proving accurate estimates for the parameters even for very small samples.

2 Bias corrected estimators

Bias corrections for (1) and (2) were presented by Ye and Chen [4]. They modified maximum likelihood estimators for α and β, hereafter BC_1 estimators, are given by

$$\hat{\alpha}_{\mathrm{BC}_1} = \frac{(n-1)}{(n+2)} \frac{n\sum_{i=1}^{n} t_i}{\left(n\sum_{i=1}^{n} t_i \log(t_i) - \sum_{i=1}^{n} t_i \sum_{i=1}^{n} \log(t_i)\right)} \tag{3}$$

and

$$\hat{\beta}_{\mathrm{BC}_1} = \frac{1}{n(n-1)}\left(n\sum_{i=1}^{n} t_i \log(t_i) - \sum_{i=1}^{n} t_i \sum_{i=1}^{n} \log(t_i)\right). \tag{4}$$

Although $\hat{\beta}_{\mathrm{BC}_1}$ is an unbiased estimator for β, the $\hat{\alpha}_{\mathrm{BC}_1}$ has a systematic bias as we shall see in the next section. Therefore, our effort is to provide improved bias corrected estimators for α.

Cox and Snell [2] presented elegant expressions to derive the bias for the parameters of parametric models. However, to derive the bias correction for (1) we need to calculate the bias of the generalized gamma distribution parameter and use the delta method. Unfortunately, these bias are very complex to be calculated. On the order hand, Ye and Chen [4] showed that both, the MLE and the closed form estimator of α return similar results. Therefore, we consider the bias correction presented by Cox and Snell [2] for the MLE of α given by

$$Bias(\hat{\alpha}) = \frac{\hat{\alpha}\psi^{(1)}(\hat{\alpha}) - \hat{\alpha}^2\psi^{(2)}(\hat{\alpha}) - 2}{2n\left(\hat{\alpha}\psi^{(1)}(\hat{\alpha}) - 1\right)^2} + O(n^{-2}). \tag{5}$$

It follows that

$$\hat{\alpha}_{\mathrm{BC}_2} = \hat{\alpha}_{\mathrm{YC}} - Bias(\hat{\alpha}_{\mathrm{YC}}), \tag{6}$$

hereafter, BC_2 estimator. This expression is easily obtained by using the orthogonal reparametrization of the gamma distribution and considering the same steps as described by Schwartz et al. [3]. It is important to say that the solution of this estimator involves the computation of transcendental functions, increasing considerably the computational time of being obtained.

In the following, we describe a different approach considering the useful approximation of (5). The function $r(x) = \frac{x\psi^{(1)}(x) - x^2\psi^{(2)}(x) - 2}{2\left(x\psi^{(1)}(x)-1\right)^2}$ is approximated via a rational function. The first thing to be noticed is that $r(x)$ has $3x - \frac{2}{3}$ as asymptote, as $x \to \infty^+$. Formally, defining $r(x) - 3x$ and using the asymptotic relations

$$\lim_{x\to 0^+} \frac{\psi'(x)}{\frac{1}{x^2}} = 1, \psi'(x) = \frac{1}{x} + \frac{1}{2x^2} + o\left(\frac{1}{x^2}\right), \lim_{x\to 0^+} \frac{\psi''(x)}{\frac{1}{x^3}} = -2, \psi'(x) = -\frac{1}{x^2} - \frac{1}{x^3} + o\left(\frac{1}{x^3}\right),$$

we obtained

$$\lim_{x\to 0^+} r(x) - 3x = 0 \quad \text{and} \quad \lim_{x\to\infty^+} r(x) - 3x = -\frac{2}{3}.$$

In order to adjust this difference to have zero as lateral limits, we can sum up $\frac{2x}{3(1+x)}$ in $r(x) - 3x$, obtaining

$$\lim_{x\to 0^+} r(x) - 3x + \frac{2x}{3(1+x)} = 0 \quad \text{and} \quad \lim_{x\to\infty^+} r(x) - 3x + \frac{2x}{3(1+x)} = 0. \tag{7}$$

Therefore, $3x - \frac{2x}{3(1+x)}$ can be considered a good candidate for an approximation of $r(x)$. Now, let $g_0(x) = r(x) - 3x + \frac{2x}{3(1+x)}$ represents the error of approximating $r(x)$ by $3x - \frac{2x}{3(1+x)}$. In order to globally estimate this error function, $g_0(x)$, we considered the re-parametrization $x(y) = \frac{y}{1-y}$, therefore obtaining the graph of $g_0(x(y))$ as in Figure 1.

We noticed in Figure 1 that the function $g_0(x(y))$ closely resembles a quadratic function with roots at $x = 0$, $x = 1$ and maximum point at $(y, g_0(x(y))) = (\frac{1}{2}, -\frac{1}{5})$, i.e., it closely resembles the quadratic function $-\frac{4}{5}y(1-y)$. Therefore $3x(y) - \frac{2x(y)}{3(1+x(y))} + \frac{4}{5}y(1-y)$ should lead to an even better approximation for $r(x(y))$. Defining $g_1(x(y)) = g_0(x(y)) + \frac{4}{5}y(1-y)$, the global error $\max_{y\in(0,1)} |g_1(x(y))|$ of this new approximation is expected to be very small, as we can see in Figure 1.

The following proposition formalizes this fact.

Proposition 2.1. *For* $g_1(x) = r(x) - \left(3x - \frac{2x}{3(1+x)} - \frac{4x}{5(1+x)^2}\right)$, *it follows that* $\max_{x\in(0,\infty)} |g_1(x)| < 5 \cdot 10^{-2}$.

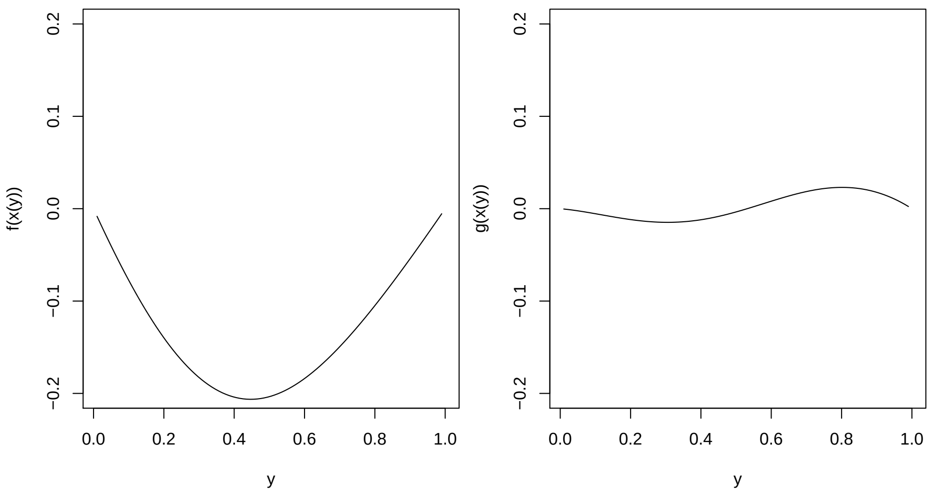

Figure 1: Left Panel: Graph of $g_0(x(y))$. Right Panel: Graph of $g_1(x(y))$.

Proof. Following Chen [1], let $p_1(n,x) = \frac{(n-1)!}{x} + \frac{n!}{2x^2} + \frac{(n+2)!}{12x^3} - \frac{(n+4)!}{720x^5}$ and $p_2(n,x) = \frac{(n-1)!}{x} + \frac{n!}{2x^2} + \frac{(n+2)!}{12x^3}$, then

$$p_1(n,x) < (-1)^{n+1}\psi^{(n)}(x) < p_2(n,x). \tag{8}$$

Notice that $p_2(n,x) - p_1(n,x) = \frac{(n+4)!}{720x^5}$ and therefore these inequalities provide good approximations for $\psi^{(n)}(x)$ for x sufficiently large, but not so good for small x. To overcome this problem and provide a good global approximation for $\psi^{(n)}(z)$ we used the recurrence relation, $\psi^{(n)}(x) = \psi^{(n)}(x+2) + (-1)^{n+1}\left(\frac{n!}{x^{n+1}} + \frac{n!}{(x+1)^{n+1}}\right)$, which is a direct consequence of the classical recurrence relation $\psi^{(n)}(x) = \psi^{(n)}(x+1) + (-1)^{n+1}\frac{n!}{x^{n+1}}$. This recurrence in combination with (8), and defining $q_1(n,x) = \frac{n!}{x^{n+1}} + \frac{n!}{(x+1)^{n+1}} + p_1(n,x+2)$, and $q_2(n,x) = \frac{n!}{x^{n+1}} + \frac{n!}{(x+1)^{n+1}} + p_2(n,x+2)$, provides

$$q_1(n,x) < (-1)^{n+1}\psi^{(n)}(x) < q_2(n,x). \tag{9}$$

Now, the difference between the two extremes $q_2(n,x)$ and $q_1(n,x)$ in the above inequality is equal to $\frac{(n+4)!}{720(x+2)^5}$, which for $n = 1,2$ and $x > 0$ is lower than $6 \cdot 10^{-3}$ and is lower than $4 \cdot 10^{-2}$ for $n = 3$ and $x > 0$. Therefore, the inequalities (9) above provides a good global approximation to $\psi^n(x)$ for $n = 1,2,3$ and $x > 0$.

Now, to prove the result, we extend $h(y) = g_1(x(y))$ from $(0,1)$ to $\bar{h} : [0,1] \to \mathbb{R}$ by defing $\bar{h}(0) = \bar{h}(1) = 0$, $\bar{h}(y) = h(y)$ for $y \in (0,1)$, and $\bar{h}(y)$ is still continuous due to (7). Using (9) for $n = 1, 2$, we can conclude formally that $\left|\bar{h}(y)\right| < 3 \cdot 10^{-2}$ for $y_k = \dfrac{k}{2^4}$, $k = 0, \cdots, 2^4$. We then used (9) for $n = 1, 2, 3$ to conclude that $\left|\frac{d\bar{h}(y)}{dy}\right| < \frac{1}{2}$ for all $y \in [0,1]$. Therefore by the mean value theorem, it follows that $\left|\bar{h}(y)\right| \leq \left|\bar{h}(y_k)\right| + \sup_{z\in[0,1]} \left|\frac{d\bar{h}(z)}{dy}\right| (y - y_k) < 3 \cdot 10^{-2} + 2^{-6} < 5 \cdot 10^{-2}$ for all $y \in [0,1]$, where y_k is chosen to be the closest real to y in the set $\left\{0, \frac{1}{2^4}, \cdots, \frac{2^4-1}{2^k}, 1\right\}$. Since $g_1(x(y)) = \bar{h}(y)$ for all $y \in (0,1)$, the result follows. $\square$

Therefore, the closed form estimator for α is given by

$$\hat{\alpha}_{\mathrm{BC}_3} = \hat{\alpha}_{\mathrm{YC}} - \frac{1}{n}\left(3\hat{\alpha}_{\mathrm{YC}} - \frac{2}{3}\left(\frac{\hat{\alpha}_{\mathrm{YC}}}{1+\hat{\alpha}_{\mathrm{YC}}}\right) - \frac{4}{5}\frac{\hat{\alpha}_{\mathrm{YC}}}{(1+\hat{\alpha}_{\mathrm{YC}})^2}\right), \tag{10}$$

hereafter, BC_3 estimator.

Notice that, by Proposition 2.1 it follows that $|\hat{\alpha}_{\mathrm{BC}_3} - \hat{\alpha}_{\mathrm{BC}_2}| < \frac{5.10^{-5}}{n} + o(n^{-2})$, and since $\hat{\alpha}_{\mathrm{BC}_2}$ is assymptically unbiased, it follows that $\hat{\alpha}_{\mathrm{BC}_3}$ also has this property.

Theorem 2.2. *Let $n > 2$ and let $\boldsymbol{t} = (t_1, \ldots, t_n)$ be an arbitrary sample such that t_1, t_2, $\cdots$, t_n are positive and not all equal, then $\hat{\alpha}_{\mathrm{BC}_3} > 0$.*

Proof. Without loss of generality let us suppose an index were chosen such that $t_n \geq t_{n-1} \geq \cdots \geq t_1 > 0$.

Now, with the hypothesis above, we shall first show that $\hat{\alpha}_{\mathrm{YC}} > 0$. In fact we have that

$$\hat{\alpha}_{\mathrm{YC}} = \frac{\frac{1}{n}\sum_{i=1}^n t_i}{\sum_{k=1}^n \left(t_k - \frac{1}{n}\sum_{i=1}^n t_i\right)\log(t_k)}.$$

On the other hand, we have that

$$
\begin{aligned}
&\sum_{h=1}^{n-1} h\left(\frac{1}{n}\sum_{i=1}^{n} t_i - \frac{1}{h}\sum_{i=1}^{h} t_i\right)\log(t_{h+1}/t_h) = \\
&\sum_{h=1}^{n-1}\left(\frac{h}{n}\sum_{i=1}^{n} t_i - \sum_{i=1}^{h} t_i\right)\log(t_{h+1}) - \sum_{h=1}^{n-1}\left(\frac{k}{n}\sum_{i=1}^{n} t_i - \sum_{i=1}^{k} t_i\right)\log(t_k) = \\
&\sum_{k=1}^{n}\left(\frac{(k-1)}{n}\sum_{i=1}^{n} t_i - \sum_{i=1}^{k-1} t_i\right)\log(t_k) - \sum_{k=1}^{n}\left(\frac{k}{n}\sum_{i=1}^{n} t_i - \sum_{i=1}^{k} t_i\right)\log(t_k) = \\
&\sum_{k=1}^{n}\left(\frac{(k-1)}{n}\sum_{i=1}^{n} t_i - \sum_{i=1}^{k-1} t_i - \frac{k}{n}\sum_{i=1}^{n} t_i + \sum_{i=1}^{k} t_i\right)\log(t_k) = \\
&\sum_{k=1}^{n}\left(t_k - \frac{1}{n}\sum_{i=1}^{n} t_i\right)\log(t_k),
\end{aligned}
$$

where in the second expression we used the index reparametrization $h = k-1$ for the first summand. Therefore it follows the identity

$$
\hat{\alpha}_{\text{YC}} = \frac{\frac{1}{n}\sum_{i=1}^{n} t_i}{\sum_{h=1}^{n-1} h\left(\frac{1}{n}\sum_{i=1}^{n} t_i - \frac{1}{h}\sum_{i=1}^{h} t_i\right)\log(t_{h+1}/t_h)}. \tag{11}
$$

Now, since all t_i are positive, it is easy to see that $\frac{1}{n}\sum_{i=1}^{n} t_i - \frac{1}{h}\sum_{i=1}^{h} t_i > 0$, for every $1 \leq h \leq n-1$. Moreover $\log(t_{h+1}/t_h) \geq 0$, for $0 \leq h \leq n-1$, due to the monotonicity of the t_h, and since $t_1, \cdots, t_n$ are not all equal, we have that $\log(t_{j+1}/t_j) > 0$, for at least one $0 \leq j \leq n$. Combining these results, it follows that $\sum_{h=1}^{n-1} h\left(\frac{1}{n}\sum_{i=1}^{n} t_i - \frac{1}{h}\sum_{i=1}^{h} t_i\right)\log(t_{h+1}/t_h) > 0$, which together with the fact that $\frac{1}{n}\sum_{i=1}^{n} t_i > 0$ and the identity (11), implies that $\hat{\alpha}_{\text{YC}} > 0$

We now prove that $\beta(x) = x - \frac{1}{n}\left(3x - \frac{2}{3}\left(\frac{x}{1+x}\right) - \frac{4}{5}\frac{x}{(1+x)^2}\right) > 0$ for all $x > 0$ and $n > 2$, which, via the reparametrization $x(y) = \frac{y}{(1-y)}$, is equivalent to proving that $\delta(y) = \beta(x(y)) = \frac{y}{(1-y)} - \frac{1}{n}\left(\frac{3y}{(1-y)} - \frac{2y}{3} - \frac{4}{5}y(1-y)\right) > 0$, for $y \in (0,1)$. Notice that $\delta(y)$ can be defined in all of $[0,1)$ and is C^1 in this domain with $\delta(y) = 0$. Now

$$
\frac{d\delta(y)}{dy} = \frac{15n - 24y^3 + 70y^2 - 68y - 23}{15n(1-y)^2}.
$$

We now show that $15n > 24y^3 - 70y^2 + 68y + 23$ for $n > 3$. To see this, since $0 < y < 1$ it, follows that $y^3 \leq y^2$ and therefore, $24y^3 - 70y^2 + 68y + 23 < -46y^2 + 68y + 23$ and elementary properties of quadratic functions shows that the max value of $-46y^2 + 68y + 23$ is given by $\frac{1107}{23} < \frac{1150}{23} = 50$ and therefore $24y^3 - 70y^2 + 68y + 23 < 50 < 15n$, for $n > 3$.

Hence, it follows that $\dfrac{d\delta(y)}{dy} > 0$, for $y \in (0, 1)$ and $n > 3$, and therefore, by the mean value theorem, $\delta(y) > \delta(0) = 0$, for $y \in (0, 1)$ and $n > 3$.

On the other hand, for $n = 3$, we have that

$$\delta(y) = \frac{y}{(1-y)} - \frac{y}{(1-y)} + \frac{1}{3}\left(\frac{2y}{3} + \frac{4}{5}y(1-y)\right) = \frac{2y(11-6y)}{45},$$

which is clearly positive for $y \in (0, 1)$.

Therefore, it is proved that $\beta(x) > 0$ for all $x > 0$ and $n > 2$ and the proof is completed.

$\square$

3 Simulation and Discussion

A simulation is performed in order to compare the performance of the estimators BC_1, BC_2 and BC_3. Following Ye and Chen [4], we consider $\beta = 1$. The mean relative error (MRE) and the root-mean-square error (RMSE) are considered as comparative measures based on 50,000,000 simulated samples. Therefore, we expect the most efficient estimators to yield MREs closer to one with smaller RMSEs. Figure 2 shows the MREs and MREs for different values of α and $n = 4$. Figure 3 shows the same results for $n = (4, 5, \ldots, 20)$, $\alpha = 0.5$ and $\alpha = 4$.

We observe that both BC_2 and BC_3 provide much less bias than BC_1, with preference to BC3, which is straightforwardly obtained. Also the proposed estimator performs better in terms of RMSE when compare with BC_1. Therefore, the BC_3 returned improved estimates for α in the three possible scenarios: $0 < \alpha < 1$, $\alpha = 1$ and $\alpha > 1$, even for small samples

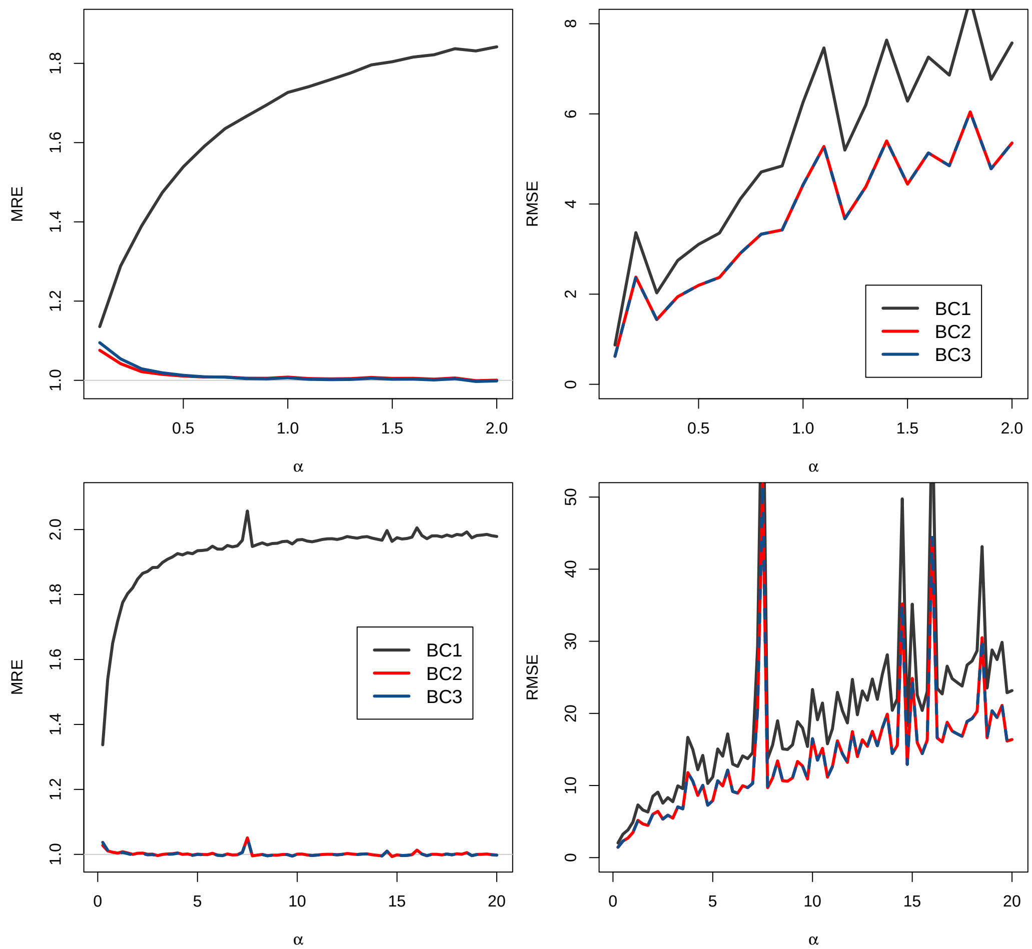

Figure 2: MREs and RMSE. Upper panel: for different values of $\alpha = (0.1, 0.2, \ldots, 2)$ for sample size of 4 elements. Lower panel: for different values of $\alpha = (0.25, 0.50, \ldots, 20)$ for sample size of 4 elements.

and should be used to perform inference for this important distribution. An interesting aspect of our approach is that such approximations can be extended to bias correction related to other distributions.

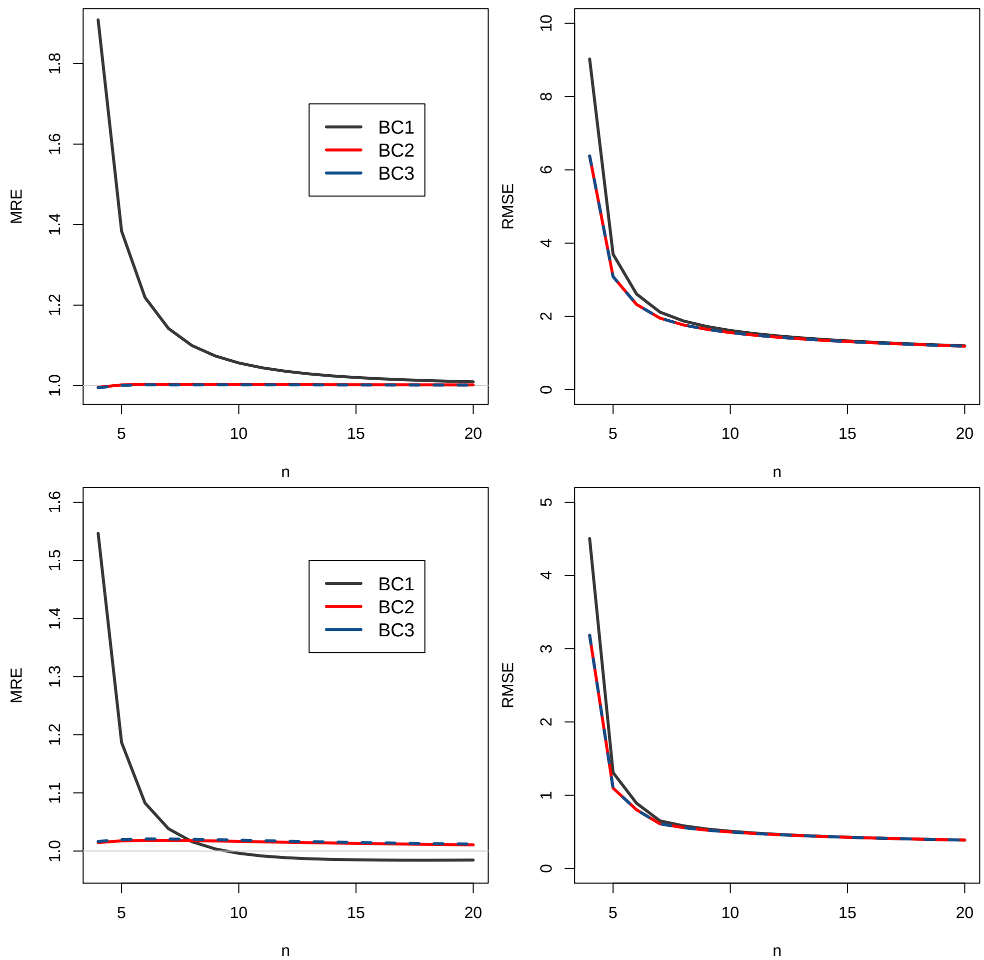

Figure 3: MREs. Upper panel: for samples sizes of $4, 5, 6, \ldots, 20$ elements considering $\alpha = 4$. Lower panel: for samples sizes of $4, 5, 6, \ldots, 20$ elements considering $\alpha = 0.5$.

Acknowledgement

The authors are very grateful to the Editor and the reviewers for their helpful and useful comments that improved the manuscript. The authors' researchers are partially supported by the Brazilian Institutions: CNPq, CAPES and FAPESP.

References

[1] Chen, C. (2005). Inequalities for the polygamma functions with application. *Scientia Magna 1*(2), 91–95.

[2] Cox, D. R. and E. J. Snell (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, 248–275.

[3] Schwartz, J., R. T. Godwin, and D. E. Giles (2013). Improved maximum-likelihood estimation of the shape parameter in the nakagami distribution. *Journal of Statistical Computation and Simulation 83*(3), 434–445.

[4] Ye, Z.-S. and N. Chen (2017). Closed-form estimators for the gamma distribution derived from likelihood equations. *The American Statistician 71*(2), 177–181.