Analysis of the effect of the sample size on the estimators of the parameters of a gamma distribution obtained by the Method of Moments and the Maximum Likelihood Estimation

Adrià Marín (1456429)
Albert Rodas (1418925)
Marc Seguí (1465613)

Abstract

A detailed analysis of the effect of the sample size n on the estimators of the parameters of a gamma distribution has been performed. Three methods have been compared to calculate such estimators, two involving the Method of Moments and one using the Maximum Likelihood Estimation. For a wide range of n, the Maximum Likelihood Estimation has been found to be the most reliable. However the bias of the estimators computed with the three different methods have been found to depend linearly on $1/n-1$, and the three have been showed to produce asymptotically unbiased estimators. The MSE of the estimators has also been computed and a linear dependence of $1/\text{MSE}$ with n has been showed for the three methods. Hence, it has been found that the estimators produced by the three different methods are consistent.

1 Introduction

1.1 Method of Moments

Consider an absolutely continuous random variable W with probability density function $f_W(w;\theta_1,\ldots,\theta_k)$ [1]. Suppose the parameters $\theta_1,\ldots\theta_k$ are unknown. Suppose the first k moments of the true distribution of W can be expressed as functions of $\theta_1,\ldots,\theta_k$, i.e.

$$\begin{cases} \mu_1 \equiv E[W] = g_1(\theta_1,\ldots\theta_k) \\ \mu_2 \equiv E[W^2] = g_2(\theta_1,\ldots\theta_k) \\ \quad\vdots \\ \mu_k \equiv E[W^k] = g_k(\theta_1,\ldots\theta_k) \end{cases}$$

Assume we can perform n measurements of W, thus obtaining a collection of n values $\{x_1,\ldots,x_n\}$ following the probability defined by $f_W(x;\theta_1,\ldots,\theta_k)$. Let $\hat{\mu_j} = \frac{1}{n}\sum_{i=1}^{n} x_i^j$ be the j-th sample moment.

By the Method of Moments, one can estimate the values of $\theta_1,\ldots\theta_k$, denoted by $\hat{\theta_1},\ldots\hat{\theta_k}$, by the solution (in case it exists) to

$$\begin{cases} \hat{\mu_1} = g_1(\hat{\theta_1},\ldots,\hat{\theta_k}) \\ \hat{\mu_2} = g_2(\hat{\theta_1},\ldots,\hat{\theta_k}) \\ \quad\vdots \\ \hat{\mu_k} = g_k(\hat{\theta_1},\ldots,\hat{\theta_k}) \end{cases} \tag{1}$$

1.2 Maximum Likelihood Estimation

Let W be an absolutely continuous random variable with probability density function $f_W(w;\theta_1,\ldots,\theta_k)$, where $\theta_1,\ldots,\theta_k$ are unknown [2]. Assume $\theta_i \in \Theta_i$, the parameter space, for all $1 \leq i \leq k$. Suppose a sample of size n, $\{x_1\ldots,x_n\}$ is drawn, and let us define the likelihood function $\mathcal{L}(\theta_1,\ldots,\theta_k|x_1,\ldots,x_n) = \prod_{i=1}^{n} f(x_i|\theta_1,\ldots,\theta_k)$.

By the Maximum Likelihood Estimation, one can estimate the real values of the unknown parameters $\theta_1,\ldots,\theta_k$, denoted by $\hat{\theta_1},\ldots\hat{\theta_k}$, as the solution to $(\hat{\theta_1},\ldots\hat{\theta_k}) = \underset{\theta_i\in\Theta_i}{\operatorname{argmax}}\ \mathcal{L}(\theta_1,\ldots,\theta_k|x_1,\ldots,x_n)$

In practice, it is often computationally preferred to use the log-likelihood function $\hat{l}(\theta_1,\ldots,\theta_k|x_1,\ldots,x_n) = \ln\mathcal{L}(\theta_1,\ldots,\theta_k|x_1,\ldots,x_n) = \sum_{i=1}^{n} \ln f(x_i|\theta_1,\ldots,\theta_k)$. Since the logarithm is a monotonic function, if the log-likelihood function is differentiable, it is necessary for $(\hat{\theta_1},\ldots\hat{\theta_k})$ to be a critical point of the likelihood function that

$$\begin{cases} \dfrac{\partial\hat{l}}{\partial\theta_1}(\hat{\theta_1},\ldots\hat{\theta_k}) = 0 \\ \dfrac{\partial\hat{l}}{\partial\theta_2}(\hat{\theta_1},\ldots\hat{\theta_k}) = 0 \\ \qquad\qquad\vdots \\ \dfrac{\partial\hat{l}}{\partial\theta_k}(\hat{\theta_1},\ldots\hat{\theta_k}) = 0 \end{cases} \tag{2}$$

1.3 Gamma distribution

Consider an absolutely continuous random variable with density function the gamma distribution

$$f(x;\alpha,\nu) = \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x}\mathbf{1}_{(0,+\infty)} \tag{3}$$

where $\mathbf{1}_{(0,+\infty)}$ is the indicator function on $(0,+\infty)$ and $\alpha,\nu > 0$ are, respectively, the rate and the shape [3].

One can straightforwardly compute the first and second order moments of such a variable X, by

$$\mu_1 \equiv E[X] = \int_0^\infty x \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x}\,\mathrm{d}x = \frac{\nu}{\alpha} \tag{4}$$

$$\mu_2 \equiv E[X^2] = \int_0^\infty x^2 \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x}\,\mathrm{d}x = \frac{\nu(\nu+1)}{\alpha^2} \tag{5}$$

Let $\hat{\mu_1}$, $\hat{\mu_2}$ be the first and second sample moments of a drawn sample following a gamma distribution as defined in 1.1. Then, 1 reads

$$\begin{cases} \hat{\mu_1} = \dfrac{\nu}{\alpha} \\ \hat{\mu_2} = \dfrac{\nu(\nu+1)}{\alpha^2} \end{cases} \tag{6}$$

whose solution is

$$\begin{cases} \hat{\alpha} = \dfrac{\hat{\mu_1}}{\hat{\mu_2} - \hat{\mu_1}^2} \\ \hat{\nu} = \dfrac{\hat{\mu_1^2}}{\hat{\mu_2} - \hat{\mu_1}^2} \end{cases} \tag{7}$$

Let now $\{x_1, \ldots, x_n\}$ be a drawn sample following a gamma distribution. The log-likelihood function defined in 1.2 is $\hat{l} = \sum_{i=1}^{n} \ln f(x_i; \alpha, \nu) = n(\nu \ln \alpha - \ln \Gamma(\nu)) + (\nu - 1) \ln (x_1 \cdot \ldots \cdot x_n) - \alpha(x_1 + \ldots + x_n)$ and 2 reads

$$\begin{cases} \dfrac{\partial \hat{l}}{\partial \alpha}(\hat{\alpha}, \hat{\nu}) = \dfrac{n\hat{\nu}}{\hat{\alpha}} - x_1 - \ldots - x_n = 0 \\ \dfrac{\partial \hat{l}}{\partial \nu}(\hat{\alpha}, \hat{\nu}) = n \ln \hat{\alpha} - n\Psi(\hat{\nu}) + \ln (x_1 \cdot \ldots \cdot x_n) = 0 \end{cases} \tag{8}$$

where $\Psi(z) = \frac{\mathrm{d}}{\mathrm{d}z} \ln \Gamma(z)$ denotes the digamma function. The system of Equations 8 can be solved numerically.

2 Methods

The results described in this paper have been obtained using the *rgamma* function of R, framed under the Gamma Distribution section of the stats package.

For all the combinations of $\alpha \in (0.2, 1)$ and $\nu \in (1, 2, 4)$, $N = 10^4$ samples of size n have been drawn, for different values of n. For each sample, the values of α and ν have been estimated using both the Method of Moments and the Maximum Likelihood Estimation as described in Section 1.3. Note that, for the Method of Moments, the experimental value $\hat{\mu_2}$ from Equation 6 has been calculated as defined in 1.1 and also such that $\hat{\mu_2} - \hat{\mu_1}^2$ is the corrected variance of the sample, being $\hat{\mu_1}$ as defined in 1.1, as done in p.37 of [4]. Whether this makes a difference in our estimators is discussed in the Section 3.

Let α and ν be the real values of the parameters used to drawn N samples of size n, thus obtaining estimators $\{\hat{\alpha_1}, \ldots, \hat{\alpha_N}\}$ and $\{\hat{\nu_1}, \ldots, \hat{\nu_N}\}$ using one of the methods described above. Then, the experimental bias of these estimators has been computed by

$$\mathrm{Bias}(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \hat{\theta}_i - \theta \tag{9}$$

Also, their mean squared error has been computed by

$$\mathrm{MSE}(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^{N} (\hat{\theta}_i - \theta)^2 \tag{10}$$

for $\theta \in \{\alpha, \nu\}$.

In order to discuss the asymptotic behaviour of the bias, a linear regression of the bias of each estimator against $\frac{1}{n-1}$ will be presented. From it, we can estimate the limit value of each estimator as n goes to infinity as the intercept of the corresponding linear regression. A hypothesis testing has been performed to show the compatibility of the intercept with 0, thus proving that we can assume our statistic to be asymptotically unbiased. The statistic used in the hypothesis testing has been $\frac{\hat{\beta}_0}{SE(\hat{\beta}_0)} \sim t_{n-2}$, where n is the number of points and $SE(\hat{\beta}_0)^2$ is an estimator for $\sigma(\hat{\beta}_0)^2$, being $\hat{\beta}_0$ the intercept calculated [5].

3 Results

From now on we state that $\hat{\alpha}_{\mathrm{MM}}$ ($\hat{\nu}_{\mathrm{MM}}$) denotes the estimator of α (resp. ν) calculated by the Method of Moments described in 1.1. $\hat{\alpha}_{\mathrm{MMcor}}$ ($\hat{\nu}_{\mathrm{MMcor}}$) denotes the estimator of α (resp. ν) calculated by the Method of Moments where $\hat{\mu_2}$ is such that $\hat{\mu_2} - \hat{\mu_1}^2$ is the corrected variance of the sample. $\hat{\alpha}_{\mathrm{MLE}}$ ($\hat{\nu}_{\mathrm{MLE}}$) denotes the estimator of α (resp. ν) calculated as the Maximum Likelihood Estimator, as described in Section 1.2.

3.1 Mean and standard deviation

The results shown in Table 1 represent the mean values of $\hat{\alpha}$ and $\hat{\nu}$ with its corrected standard deviation $S_{\hat{\alpha}}$ and $S_{\hat{\nu}}$ for the case where $\alpha = 1$ and $\nu = 2$. The results are representative of those obtained for different values of α and ν.

n	10	25	50	100	1000	5000
$\hat{\alpha}_{\mathrm{MM}}$	1.60 ± 1.07	1.20 ± 0.43	1.09 ± 0.28	1.05 ± 0.19	1.01 ± 0.06	1.00 ± 0.03
$\hat{\alpha}_{\mathrm{MMcor}}$	1.43 ± 0.96	1.15 ± 0.41	1.07 ± 0.28	1.04 ± 0.19	1.00 ± 0.06	1.00 ± 0.03
$\hat{\alpha}_{\mathrm{MLE}}$	1.46 ± 0.99	1.14 ± 0.37	1.06 ± 0.23	1.03 ± 0.16	1.00 ± 0.05	1.00 ± 0.02
$\hat{\nu}_{\mathrm{MM}}$	3.02 ± 1.81	2.35 ± 0.76	2.17 ± 0.51	2.09 ± 0.35	2.01 ± 0.11	2.00 ± 0.05
$\hat{\nu}_{\mathrm{MMcor}}$	2.72 ± 1.63	2.26 ± 0.73	2.13 ± 0.50	2.07 ± 0.34	2.01 ± 0.11	2.00 ± 0.05
$\hat{\nu}_{\mathrm{MLE}}$	2.77 ± 1.69	2.24 ± 0.65	2.11 ± 0.41	2.05 ± 0.28	2.00 ± 0.08	2.00 ± 0.04

Table 1: Mean and associated standard deviation for $\hat{\alpha}$ and $\hat{\nu}$ for the case where $\alpha = 1$ and $\nu = 2$. Different values of n were used.

As observed in Table 1 for small n the Method of Moments using the corrected variance of the sample can outperform the method that uses the Maximum Likelihood Estimator. As an example, if $n = 10$ the corrected standard deviation of $\hat{\alpha}_{\mathrm{MMcor}}$ is smaller than $\hat{\alpha}_{\mathrm{MLE}}$. The same happens with $\hat{\nu}$. However, for bigger values the Maximum Likelihood Estimator is better. One could think that these estimators are (asymptotically) unbiased as the values obtained are compatible with the real values of α and ν, but this will be later discussed.

Performing the same calculations, results for the estimators of α and ν have been obtained for the rest of the cases. With fixed α, $S_{\hat{\alpha}}$ decreases by increasing the value of ν. On the other hand, with fixed ν the values obtained for $S_{\hat{\nu}}$ varying α don't seem to follow any pattern. However, fluctuations of $S_{\hat{\alpha}}$ when tuning ν represent less than a 0.5% difference. The same occurs for $S_{\hat{\nu}}$ varying α.

3.2 Histogram results

Figures 1 and 2 show representative histograms of $\hat{\alpha}$ for different values of α, ν and n. Figures 3 and 4 show representative histograms of $\hat{\nu}$ for different values of α, ν and n. Note that histograms for the three methods described in Section 2 have been superimposed.

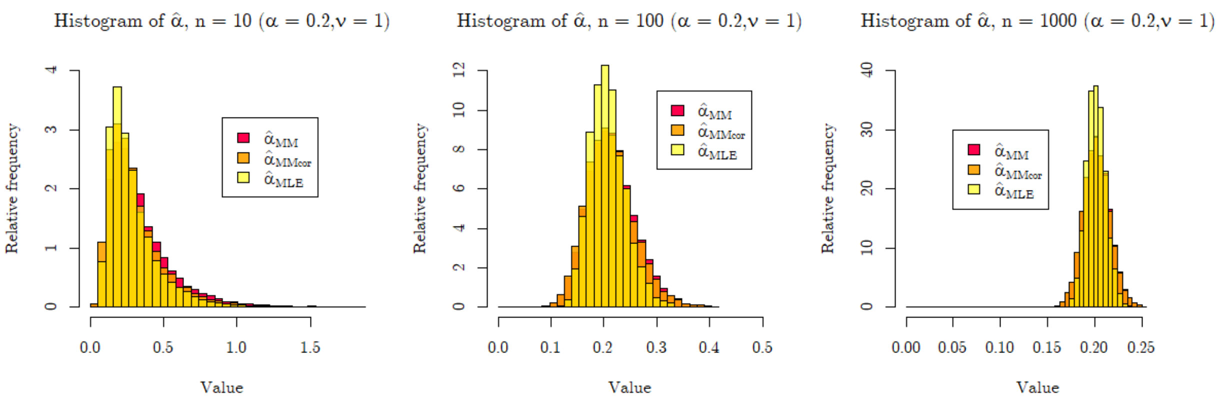

Figure 1: Histograms of $\hat{\alpha}$ for $n = 10$, 100 and 1000. In this case $\alpha = 0.2$ and $\nu = 1$.

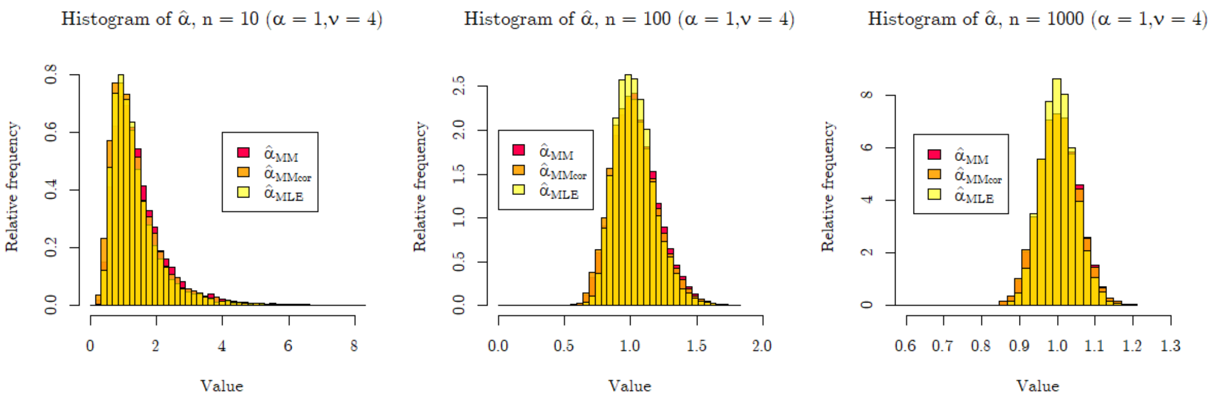

Figure 2: Histograms of $\hat{\alpha}$ for $n = 10$, 100 and 1000. In this case $\alpha = 1$ and $\nu = 4$.

As a qualitative analysis, note that, as one would expect, the histograms become sharper as n increases, and also less biased. In addition, as the Central Limit Theorem states, histograms have a bigger resemblance to a Gaussian function.

Also, for all cases, the Maximum Likelihood Estimation gives slightly sharper histograms, with substantially more values closer to the real value of α or ν. No clear qualitative results can be obtained on the difference between both approaches under the method of moments.

In a first analysis, no differences are seen as the real values of α and ν are tuned.

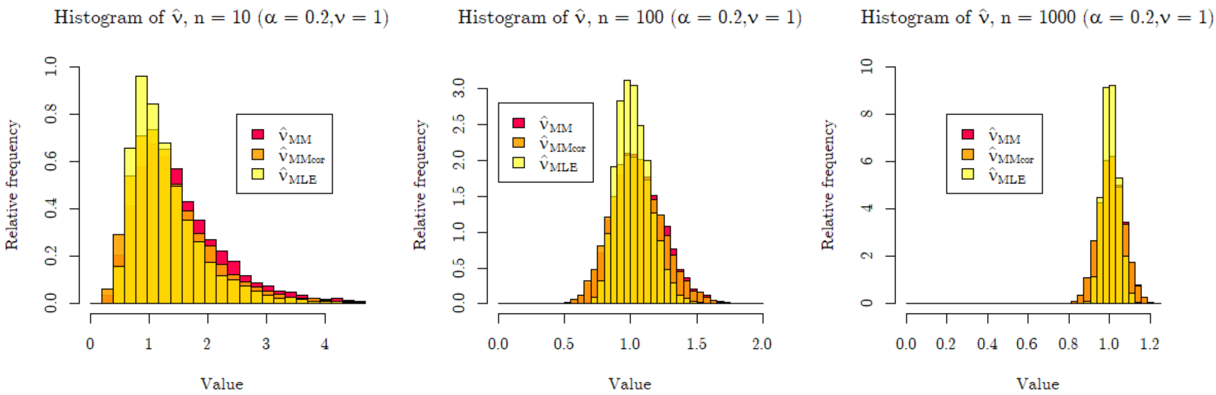

Figure 3: Histograms of $\hat{\nu}$ for $n = 10$, 100 and 1000. In this case $\alpha = 0.2$ and $\nu = 1$.

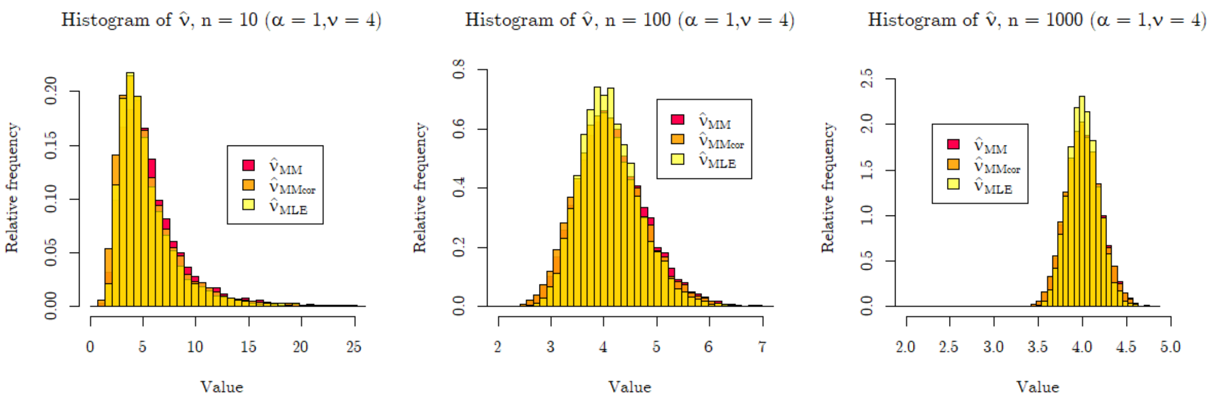

Figure 4: Histograms of $\hat{\nu}$ for $n = 10$, 100 and 1000. In this case $\alpha = 1$ and $\nu = 4$.

3.3 Calculation of the Bias

The biases of the estimators of α and ν have been calculated using Equation 9 for different values of α, ν and n.

It has been found that the biases decrease as n increase, for all pairs of parameters and methods.

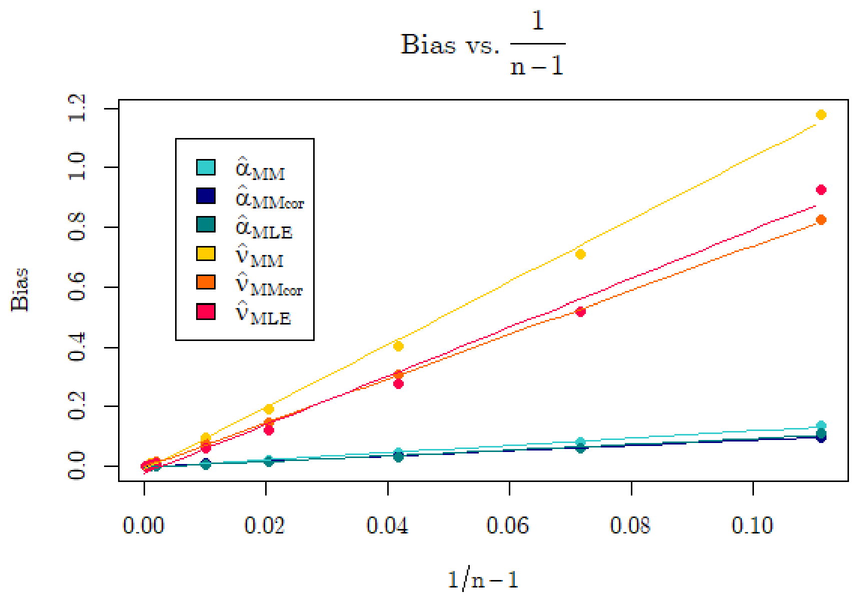

Figure 5: Bias versus $\frac{1}{n-1}$. $\alpha = 0.2$ and $\nu = 2$ were used in this case. The lines represent the linear dependence obtained from the linear regressions. The results for other values of α and ν are similar.

Figure 5 shows a linear dependence between the biases of the estimators and $\frac{1}{n-1}$ for the three methods, for a representative pair of α and ν. For all pairs of values of α and ν and methods, the R^2 values are greater than 0.995. As deduced in Section 3.2, the Maximum Likelihood Estimation produces a smaller bias in relation to the Method of Moments, and it can now be argued that the Method of Moments produces less biased estimators when used with the corrected variance rather than with the variance itself. The hypothesis testing described in 2 has been performed for the intercepts obtained for all methods and pairs of parameters. In all cases, with 5% statistical significance, there has been no statistical evidence of the intercept being different of zero. Hence, we claim this enables us to believe all estimators produced are asymptotically unbiased.

3.4 Calculation of the MSE

The mean standard error of the estimators of α and ν have been calculated using Equation 10 for different values of α, ν and n.

It has been found that the MSE of the studied estimators decreases as n increases, as expected.

Figure 6 shows a linear dependence between $\frac{1}{\text{MSE}}$ and n. For all pairs of values of α and ν and methods, the R^2 values are greater than 0.995. As assumed from the histograms in Section 3.1, the Maximum Likelihood Estimation produces smaller Mean Squared Errors, thus generating sharper histograms as seen in 3.1. It can be now argued as well that the Method of Moments using the corrected variance produces estimators with smaller mean squared error than the Method of Moments with the variance itself.

As discussed in the previous paragraph, it has been showed that MSE behaves like $\frac{1}{an+b}$ for all estimators and methods, where $a, b \in \mathbb{R}$ depend on the estimator and method. In all cases, this implies that the MSE converges to 0 as n goes to infinity.

4 Conclusions

Three methods have been used to estimate the real values of the parameters of a gamma distribution. The evolution of different features of the estimators has been analysed when varying the sample size n. For a

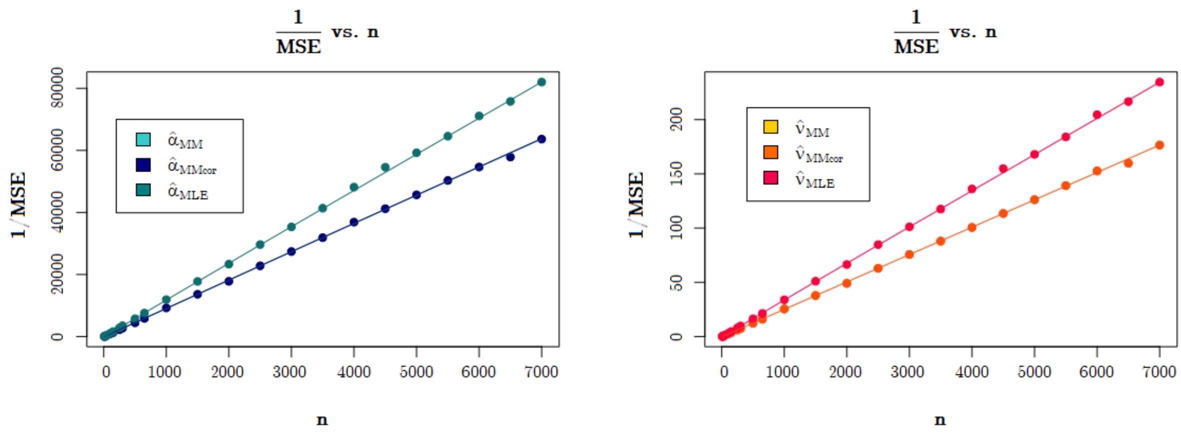

Figure 6: $\frac{1}{\text{MSE}}$ versus n. In this case $\alpha = 0.2$ and $\nu = 4$. The lines represent the linear dependence obtained from the linear regressions. Note that the linear regressions obtained with both versions of the Method of Moments are indistinguishable

given sample size, $N = 10^4$ estimators have been calculated for each method, thus producing a histogram per method per parameter. This histograms have been discussed and analysed.

Also, the bias and MSE of the estimators have been computed. It has been found, as one would expect, that the estimators become less biased as n increases, meaning that the peak of the histogram converges to the real value of the parameter for large n. It has been argued how the Maximum Likelihood Estimator is the most reliable of the three methods considered, as it produces smaller biases and MSE for a fixed n. This can be seen in the histograms, as they peak closer to the real value and they are sharper than those produced by the other two methods for the same n.

It has been showed that, for the three methods, the produced biases have a linear dependence with $\frac{1}{n-1}$. A hypothesis testing has been applied to show that, with 95% significance level, there is no statistical evidence to refuse that the intercepts of the linear regressions are zero, for all cases. Hence, it has been argued that all the estimators produced are asymptotically unbiased.

It has also been showed that there exists a linear dependence between $\frac{1}{\text{MSE}}$ and n for all the estimators produced. It has been claimed that this implies that all the estimators are MSE-consistent. Since all estimators are asymptotically unbiased, this implies that all estimators are consistent, given Chebyshev's inequality.

Under the Central Limit Theorem, all the histograms converge to a Gaussian function as n increases. The fact that the MSE converges to zero implies that these Gaussians tend to delta distributions for large n. The fact that the estimators are asymptotically unbiased implies that this delta distributions converge to the real value of the parameter for large n.

References

[1] K.O Bowman and L.R Shenton, "Estimator: Method of Moments", p. 2092-2098, *Encyclopedia of statistical sciences*, Wiley (1998).

[2] Rossi, Richard J., p. 227 *Mathematical Statistics: An Introduction to Likelihood Based Inference*, Wiley (2018)

[3] J. del Castillo, *Evolució dels estimadors segons la mida mostral*, Campus Virtual de la Universitat Autònoma de Barcelona (2020)

[4] J. del Castillo, *Estadística. Diapositives de la primera part del curs*, Campus Virtual de la Universitat Autònoma de Barcelona (2020)

[5] L. Orellana,*Regresión Lineal Simple*, Departamento de Matemática. Facultad de ciencias exactas y naturales. Universidad de Buenos Aires (2011)