International Journal of Probability and Statistics 2015, 4(2): 37-41
DOI: 10.5923/j.ijps.20150402.01

Estimation of the Mean and Variance of a Univariate Normal Distribution Using Least-Squares via the Differential and Integral Techniques

C. R. Kikawa*, M. Y. Shatalov, P. H. Kloppers

Department of Mathematics and Statistics, Tshwane University of Technology, Pretoria, South Africa

Abstract Two new approaches (method I and II) for estimating parameters of a univariate normal probability density function are proposed. We evaluate their performance using two simulated normally distributed univariate datasets and their results compared with those obtained from the maximum likelihood (ML) and the method of moments (MM) approaches on the same samples, small n = 24 and large n = 1200 datasets. The proposed methods, I and II have shown to give significantly good results that are comparable to those from the standard methods in a real practical setting. The proposed methods have performed equally well as the ML method on large samples. The major advantage of the proposed methods over the ML method is that they do not require initial approximations for the unknown parameters. We therefore propose that in the practical setting, the proposed methods be used symbiotically with the standard methods to estimate initial approximations at the appropriate step of their algorithms.

Keywords Maximum likelihood, Method of moments, Normal distribution, Bootstrap samples

1. Introduction

Statistical inference is largely concerned with making logical conclusions about a population using an observed section or part of the entire population referred to as the sample [1]. The reference population can always be represented using an appropriate probability framework which is usually written in terms of unknown parameters. For instance the crop yield obtained when a certain fertilizer is applied can be assumed to follow a normal distribution with mean μ, and standard deviation, σ; it is thereafter required to make inferences about the parameters, μ and σ using the statistics $\bar{x}$ and s that are estimated based on the sample of crop yield and then inferences made on the total crop yield. Note that in this work we only deal with one aspect of statistical inference that is estimation and two novel approaches are discussed in this case.

Let x be a single realisation from a univariate normal density function with mean μ, and standard deviation σ, which implies that $x \sim N(\mu, \sigma)$ with $-\infty < \mu < \infty$, $\sigma > 0$. In this paper, simple and computationally attractive methods for estimating both μ and σ of a univariate normal distributionfunction are proposed. However, methods for estimating the sufficient parameters of a univariate normal density function are well known such as the method of moments and the maximum likelihood method [2, 3], but all these are computationally intensive. Again, much as the maximum likelihood estimators have higher probability of being in the neighbourhood of the parameters to be computed, in some instances the likelihood equations are intractable in the absence of high computing gadgets like computers. Though the method of moments could quickly be computed manually by hand, its estimators are usually far from the required quantities and for small samples the estimates are often times outside the parameter space [4, 5]. In all it is not worthwhile to rely on the estimates from the method of moments.

1.1. Generalized Probability Density Function

When a dataset is presented and critically observed for any characteristics that it may exhibit; statistically called exploratory data analysis, we usually want to study its pattern that can vaguely lead us to a possible probability density function (pdf) that can be taken as its probability frame-work for those data. However, if it requires one to build a whole new frame-work or model, then a lot of work has to be done which is quite demanding. In this section we present a frame-work that nearly suits all the pdfs of continuous random variables

$$\frac{1}{\wedge_r(s)} \emptyset_s \left(\frac{x-l}{ב}\right), \; L_b \le x \le L_c \qquad (1.1)$$

where L_b and L_c indicate the domain of applicability: often times from $-\infty$to ∞ or from 0 to ∞ depending on the framework under consideration. $\emptyset_s$, is the actual shape

* Corresponding author:
richard.kikawa@gmail.com (C. R. Kikawa)
Published online at http://journal.sapub.org/ijps

function of the pdf; $\wedge_r$ (the area under the function) is necessary to normalise the integral of the respective function, over the stated to one, s is the shape parameter, l and ב are the location and scaleparameters respectively.

On examining Eq. (1.1), we present the normal distribution function as

$$f_{X_i}(x_i;\boldsymbol{\theta}) = \frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_i-\theta_1}{\theta_2}\right)^2\right) \quad (1.2)$$

Eq. (1.2) represents a normal probability density function with mean θ_1 andstandard deviation θ_2, where X_i is a univariate random variable and θ is avector of length, say M comprising the unknown sufficient parameters [6]. The study presents descriptions and numerical evaluations of two proposed methods intended to estimate the unknown parameters of univariate normal density functions. The proposed methods are compared with two methods in current use, that is the maximum likelihood and method of moments.

These are preferred due to their robustness and it is also known that they are symbiotic in that; estimates by the method of moments may be used as the initial approximations to the solutions of the formulated likelihood equations, and successive improved approximations are found using the well-known numerical methods like the Newton-Raphson, Levenberg-Marquardt etc. [7].

2. Method Formulations

It is well known that parameter estimation is an integral part of statistical modelling [8]. In this section we describe two formulations of estimating the univariate normal distribution based on the least-squares method. Linearization of the transcendental model is performed via differentiation and integration methods.

2.1. Theoretical Approach

The main idea is to transform the original problem into a new problem which is linear with respect to a part of the original unknown parameters or their combinations. For instance in the case of the Gaussian density or commonly known as the normal pdf, the transformation is done as in the corresponding system in the following section 2.2. On formulation of a linear system through differential and integral techniques, we finally identify the formulated linear system using ordinary least squares method [9].

2.2. Method I

Considering Eq. (1.2) and taking its first derivative

$$\boldsymbol{f}'(\boldsymbol{x},\boldsymbol{\theta}) = -\frac{1}{2}\frac{\sqrt{2}}{\theta_2^3\sqrt{\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)(\boldsymbol{x}-\boldsymbol{\theta_1}) \quad (2.1)$$

$$= -\frac{1}{\theta_2^2}\frac{\sqrt{2}}{\theta_2\sqrt{\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)(\boldsymbol{x}-\boldsymbol{\theta_1}) \quad (2.2)$$

From Eq. (2.3), let $\emptyset = \left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)$ then,

$$\boldsymbol{f}'(\boldsymbol{x},\boldsymbol{\theta}) = -\frac{x}{\theta_2^2}\frac{1}{\theta_2\sqrt{2\pi}}\exp(\emptyset) + \frac{\theta_1}{\theta_2^2}\frac{1}{\theta_2\sqrt{2\pi}}\exp(\emptyset) \quad (2.3)$$

$$\boldsymbol{f}'(\boldsymbol{x},\boldsymbol{\theta}) = \frac{\theta_1}{\theta_2^2}\frac{1}{\theta_2\sqrt{2\pi}}\exp(\emptyset) - \frac{x}{\theta_2^2}\frac{1}{\theta_2\sqrt{2\pi}}\exp(\emptyset) \quad (2.4)$$

Hence Eq (2.4) now becomes,

$$\boldsymbol{f}'(\boldsymbol{x},\boldsymbol{\theta}) = \frac{\theta_1}{\theta_2^2}\boldsymbol{f}(\boldsymbol{x}) - \frac{x}{\theta_2^2}\boldsymbol{f}(\boldsymbol{x}) \quad (2.5)$$

$$\boldsymbol{f}'(\boldsymbol{x},\boldsymbol{\theta}) = \boldsymbol{K_1 f}(\boldsymbol{x}) - \boldsymbol{K_2 x f}(\boldsymbol{x}) \quad (2.6)$$

where,

$\boldsymbol{K_1} = \boldsymbol{\theta_1}(\boldsymbol{\theta_2^2})^{-1}$, $\boldsymbol{K_2} = (\boldsymbol{\theta_2^2})^{-1}$ and $\boldsymbol{f}(\boldsymbol{x}) = \frac{1}{\theta_2\sqrt{2\pi}}\exp(\emptyset)$

Numerical integral methods [10] are now used to integrate Eq. (2.6) over an interval $[\boldsymbol{x_0}, \boldsymbol{x}]$

$$\int_{x_0}^{x}\boldsymbol{f}'(\boldsymbol{x})\boldsymbol{dx} = \boldsymbol{K_1}\int_{x_0}^{x}\boldsymbol{f}(\boldsymbol{x})\boldsymbol{dx} - \boldsymbol{K_2}\int_{x_0}^{x}\boldsymbol{xf}(\boldsymbol{x})\boldsymbol{dx} \quad (2.7)$$

$$= \boldsymbol{K_1 I_1}(\boldsymbol{x}) - \boldsymbol{K_2 I_2}(\boldsymbol{x}) \quad (2.8)$$

where,

$$\boldsymbol{I_1}(\boldsymbol{x}) = \int_{x_0}^{x}\boldsymbol{f}(\boldsymbol{x})\boldsymbol{dx} \text{ and } \boldsymbol{I_2}(\boldsymbol{x}) = \int_{x_0}^{x}\boldsymbol{xf}(\boldsymbol{x})\boldsymbol{dx}\,.$$

Care should be taken on the numerical procedure for Eq. (2.7) as it is somewhat involved and requires a step by step approach before implementation in any suitable programing language.

We now write a complete linear regression function as

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{K_1 I_1}(\boldsymbol{x}) + \boldsymbol{K_2 I_2}(\boldsymbol{x}) + \boldsymbol{f}(\boldsymbol{x_0}) \quad (2.9)$$

Where, $\boldsymbol{f}(\boldsymbol{x_0})$ is a regression constant while $\boldsymbol{K_1}$ and $\boldsymbol{K_2}$, are the regression coefficients. It has been assume that $\boldsymbol{I_1}(\boldsymbol{x_0}) = \boldsymbol{0}$ and $\boldsymbol{I_2}(\boldsymbol{x_0}) = \boldsymbol{0}$ for the analytical illustration otherwise these could be included in Eq. (2.9). However, in real practise this assumption could be violated without compromising the accuracy of the method. There are a variety of methods for solving linear regression models of the form presented in Eq. (2.9), such as Gauss-elimination, QR-decomposition, least-squares and total least-squares [11]. In this work, the ordinary least-squares (OLS) method is preferred for its simplicity and it is applied to estimate the parameter coefficients $\boldsymbol{K_1}$ and $\boldsymbol{K_2}$. The OLS method is well known and available in a number of statistical literature, for the estimation procedure using OLS the reader is referred to [9]. The parameters of the univariate normal density function can then be computed by applying straight forward algebra, hence

$$\hat{\mu} = \frac{K_1}{K_2}$$

and

$$\hat{\sigma} = \frac{1}{\sqrt{-K_2}}\,.$$

Using an appropriate univariate dataset, the computed estimates can now be compared with those from the method of moments and maximum likelihood methods, and standard statistical measures applied to ascertain the accuracy of the proposed method.

2.3. Method II

The main difference between this method and that formulated in the preceding section 2.3, is that numerical integration techniques were applied on $I_2(x) = \int_{x_0}^{x} xf(x)dx$ Eq.(2.8). However, for the current method II, the well-known method of integration by parts [12] is applied at the same stage, to obtain,

$$\int_{x_0}^{x} xf(x)dx = xI(x) - x_0I(x_0) + I_*(x) + I(x_0) \quad (2.10)$$

where,

$$I(x) = \int_{x_0}^{x} f(x)dx \text{ and } I_*(x) = \int_{x_0}^{x}[\int f(x)\,dx]dx$$

From method I, Eq. (2.7) can now be written as

$$f(x) - f(x_0) = K_1I_x - K_2[xI(x) - I_*(x) - x_0I(x_0) + I_*(\boldsymbol{x}_0)] \quad (2.11)$$

$$f(x) = K_{1*}I_x - K_{2*}[xI(x) - I_*(x)] + \varepsilon \quad (2.12)$$

Where,

$\varepsilon = K_{2*}x_0I(x_0) - K_{2*}(x_0) + f(x_0)$. Note, that numerical integration is required at appropriate steps for this second method as well. Eq. (2.13) is then estimated using standard OLS method [9] and the required parameters estimated as

$$\hat{\mu}_* = \frac{K_{1*}}{K_{2*}}$$

and

$$\hat{\sigma}_* = \frac{1}{\sqrt{-K_{2*}}}$$

The main difference between these two approaches is that, in method I, we considered use of numerical integration at an earlier step, Eq. (2.7), but in method II, the conventional method of integration by parts is considered and numerical integration on Eq. (2.12). It is noticed that the application of the different approaches of integration at the relevant stages causes a significant difference in the accuracy of estimates from the two estimation methods. Two Monte Carlo numerical simulations are performed using Mathematica software. Mathematica provides an environment in which programming of the proposed approaches is performed and application of the maximum likelihood and the method of moments on the simulated datasets.

3. Simulations

To evaluate empirically the performance of the proposed methods, two normally distributed datasets were simulated with known μ and σ. These datasetsare considered to be random samples of some infinite hypothetical population of possible values. It was necessary to consider both the large $(n > 30)$ and small $(n < 30)$ samples as this could probably give a clue on the performance of the proposed methods when applied to samples of varying sizes.

It is known that the principal qualifications of acceptable statistics may most readily be seen by their behaviour when derived from large samples [13]. The aim was to ascertain how these different methods reproduced the known parameter estimates (i.e. $\hat{\mu}$ and $\hat{\sigma}$) and also provide base of supportto the proposed methods especially when used on large samples. It is stated that "a statistic is said to be a consistent estimate of any parameter, if when calculated from an indefinitely larges ample it tends to be accurately equal to that parameter" [13]. For our work the results from the large sample undoubtedly give a hint on the consistence of the estimates computed from the proposed methods see also Table 6.

We have considered the performance of the proposed methods I and II, see sections 2.2 and 2.3, by applying the maximum likelihood (ML) and the method of moments (MM) using two simulated datasets. The small sample $(n = 24)$ and the large sample $(n = 1200)$ methodological evaluations are as presented in the subsequent tables.

4. Results

Table 1. Performance of proposed methods I, II, the maximum likelihood and method of moments, n=24

Parameter	Exact values	Method I	Method II	MM	ML
$\hat{\mu}$	40.0000	39.4575	39.8927	39.4544	39.4544
$\hat{\sigma}$	1.20000	1.05777	1.19526	1.04627	1.04627

Table 2. Performance of proposed methods I, II, the maximum likelihood and method of moments, n=1200

Parameter	Exact values	Method I	Method II	MM	ML
$\hat{\mu}$	40.0000	40.0456	40.0013	40.0456	40.0456
$\hat{\sigma}$	1.20000	1.20688	1.20001	1.20633	1.20633

From Tables 1 and 2, we observe that the evaluated methods produce good and acceptable results when compared with the actual or "true" parameters i.e. μ = 40 and $\sigma = 1.2$ in both large and small samples. However, we haveto understand that these proposed methods cannot be very useful without quantitative statements about their accuracy; in this way it is imperative to evaluate their success. The simplest method of accuracy assessment is based on the confidence intervals of the parameters in question. Confidence intervals can be computed for the accuracy of the point estimate in this case for the mean values presented in Tables 1 and 2. When we require to measure accuracy based on the 95% confidence level, then the interval will be computed as

$$\hat{\mu} \pm 1.96\hat{\sigma},$$

where $\hat{\mu}$ *is* the estimated mean and $\hat{\sigma}$ is the estimated standard deviation. Tables 3 and 4 show the confidence intervals of the estimated mean from both the small and large samples of the estimates computed from methods I and II and the maximum likelihood. It should be noted that

we have only considered the estimates from ML method and not MM, since from the latest version of Mathematica (i.e. Mathematica 9), the results from ML and MM are virtually the same in all our computations. In Tables 5 and 6, we presented several results of $\hat{\mu}$ and $\hat{\sigma}$ respectively, computed bybootstrapping from the parent sample or re-sampling. This was aimed at giving a visual analysis on which methods could be better preferred and the sample size on which we could consider using the proposed method for acceptable results.

Table 3. The 95% confidence interval of the estimates from the evaluated methods, n=24

Method	Estimates	Confidence intervals
I	39.4575	37.3843;41.5307
II	39.8927	37.5500;42.2354
Maximum likelihood	39.4544	37.4037;41.5051

Table 4. The 95% confidence interval of the estimates from the evaluated methods, n=1200

Method	Estimates	Confidence intervals
I	40.0456	37.6801;42.5307
II	40.0013	37.6493;42.3533
Maximum likelihood	40.0456	37.6812;42.4100

Table 5. Performance of ML, I and II methods on various test bootstrap samples, n=24

Maximum likelihood $(\hat{\mu},\hat{\sigma})$	Method I $(\hat{\mu},\hat{\sigma})$	Method II $(\hat{\mu},\hat{\sigma})$
40.2391 1.18817	40.2313;1.19199	40.2135;1.19824
40.1509 0.87617	40.1495;0.88872	40.1383;0.89825
40.1939 1.15810	40.1801;1.15708	40.1800;1.16231
39.6678 1.21087	39.6688;1.21866	39.6789;1.29725
39.6678 1.16358	39.8248;1.15310	39.8824;1.67242
40.0976 1.20060	40.0961;1.22095	40.0722;1.21002
40.3297 1.05846	40.3332;1.07261	40.3112;1.04562
40.2005 1.29045	40.1948;1.31052	40.1892;1.24561
40.2820 0.98208	40.2800;0.99535	40.27901;0.9872
39.9402 1.40495	39.9463;1.41680	39.99231;1.4544
39.6930 1.18872	39.6947;1.20114	39.69501;1.2015
39.6399 1.27985	39.6377;1.29467	39.64772;1.2952
40.0148 1.45176	39.9892;1.45013	39.99821;1.4541
40.1256 1.19554	40.1315;1.20011	40.11245;1.2001
40.1411 1.28769	40.1398;1.29999	40.12971;1.2778

5. Discussion

From the current work, conventional methods, i.e. maximum likelihood and the method of moments are compared with the proposed methods I and II. The comparison was aimed at visualising how best each of them reproduced the known parameters μ and σ. In this case confidenceintervals were computed see Tables 3 and 4 for each of the point estimates (mean value) computed from either methods.

The confidence intervals were interpreted to mean that if we had repeated the same sampling scheme a large number of times, we would have expected that in 95% of these experiments the observed accuracy see Tables 1 and 2 for the point estimates; would be somewhere between the respective confidence limits as presented in Tables 3 and 4 in either methods. It should be noted here that we took a risk of 5% that the true means are either less than or greater than the lower and upper boundaries respectively. Therefore, we can narrow the confidence intervals at the expense of committing a greater risk of a Type I error [14]. From Tables 5 and 6, the results of 15 samples of sizes n = 24 and n = 1200 are presented respectively. This was intended to show the performance of either method on both the large and small samples.

We cannot over state that, visual inspection is not the best method to ascertain whether or not a given method produces acceptable results and we reserve as future work for an analytical proof that the proposed methods are consistent.

Table 6. Performance of ML, I and II methods on various test bootstrap samples, n=1200

Maximum likelihood $(\hat{\mu},\hat{\sigma})$	Method I $(\hat{\mu},\hat{\sigma})$	Method II $(\hat{\mu},\hat{\sigma})$
40.0357,1.18570	40.0357,1.18620	40.0135;1.19924
40.0273,1.17764	40.0273,1.17813	40.0121;1.99252
40.0160, 1.20916	40.0160,1.20865	40.0140;1.20414
39.9988, 1.23107	39.9988,1.23159	39.9998;1.22725
39.9997, 1.18629	39.9997,1.18681	39.9999;1.19724
40.0084, 1.20564	40.0084,1.20614	40.0062;1.20002
40.0336, 1.19816	40.0336,1.19869	40.0311;1.98845
39.9651, 1.17556	39.9651,1.17600	39.9792;1.18756
39.9619, 1.18619	39.9614,1.18670	39.9752;1.19877
39.9670, 1.17614	39.9670,1.17664	39.9723;1.17954
40.0251, 1.18668	40.0251,1.18722	40.0151;1.18845
39.9515, 1.17922	39.9515,1.17972	39.9575;1.17982
40.0551, 1.22794	40.0551,1.22848	40.02521;1.2211
39.9891, 1.19764	39.9891,1.19817	39.9981;1.19837
39.9540, 1.23894	39.9540,1.23947	39.9574;1.25787

6. Conclusions

Considering estimates obtained from the proposed methods and the maximum likelihood, on both the small and large samples, it can be observed that the proposed methods produced relatively acceptable estimates. For the large samples the mean values are the same for all the methods which shows that the proposed methods have the same accuracy as the more trusted and frequently applied maximum likelihood method also regarded as indispensable tool for many statistical modelling techniques [15]. However, the standard deviation estimates differ slightly in

each method. The strength of the proposed methods over the maximum likelihood is that the proposed methods do not require starting approximations for the unknown parameters while for the maximum likelihood, it is a requirement for the practitioner to provide starting approximations for the unknown parameters. These starting approximations may not guarantee convergence and may also result in longer computation time if they are far from the required minimum.

ACKNOWLEDGEMENTS

The authors would like to extend their sincere appreciation to the Directorate of Research and Innovation of Tshwane University of Technology for funding this research project under the Postdoctoral Scholarship fund 2014/2015. We also thank the editorial board and the anonymous reviewer who provided insightful comments that led to the improvement of the article.

REFERENCES

[1] Cox, D.R. and D.V. Hinkley, *Theoretical Statistics.* 1974: Chapman Hall.

[2] Prokhorov, Y.V. and Y.A. Rozanov, *Probability Theory, basic concepts. Limit theorems, random processes*. 1969: Springer.

[3] Dorfman, D.D. and A.J. Edward, *Maximum-likelihood estimation of Parameters of signal-detection theory and determination of confidence intervals rating-method data*. Journal of Mathematical Psychology, 1960. 6(3): p. 487-496.

[4] Hall, A.R., *Generalized Method of Moments (Advanced Texts in Economics)*. 2005: Oxford University Press.

[5] Hansen, L.P., *Method of Moments in International Encyclopedia of the Social and Behavior Science.* 2002, Pergamon: Oxford.

[6] Huber, P.J., *Robust Statistics*. 1981: Wiley.

[7] NIST, *e-Hand book of Statistical Methods*. 2003.

[8] Kloppers, P.H., C.R. Kikawa, and M.Y. Shatalov, *A new method for least squares identification of parameters of the transcendental equations*. International Journal of the Physical Sciences, 2012. 7: p. 5218-5223.

[9] Searle, S.R., *Linear Models*. 1971: John Wiley and Sons.

[10] Gerald, C.F. and P.O. Wheatley, *Applied Numerical Analysis*. Third ed. 1983: Addison-Wesley Publishing Company

[11] Elad, M., P. Milanfar, and G.H. Golub. *Shap from moments-an estimation theory perspective. in IEE Transactions on Signal Processing*. 2004.

[12] Fang, S., *Integration by parts for heat measures over loop groups*. Journal de Mathematique Pures et Appliques, 1999. 78(9): p. 877-894.

[13] Fisher, R.A. *Theory of Statistical estimation. in Mathematical Proceedings of the Cambridge Philosophical Society*. 1925.

[14] Rossiter, D.G., *Statistical methods for accuracy assesment of classified thematic maps*. 2004, International Institute for Geo-information Science and Earth Observation.

[15] Myung, I.J., *Tutorial on maximum likelihood estimation*. Journal of Mathematical Psychology, 2003. 47(1): p. 90-100.