ON MAXIMUM LIKELIHOOD AND SAMPLE MOMENT ESTIMATORS FOR THE MTH (CENTRAL) MOMENT IN A NORMAL AND GENERALIZED GAMMA POPULATION

by
HADI ABBASZADEHPEIVASTI

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfilment of
the requirements for the degree of Master of Science

Sabancı University
December 2020

ON MAXIMUM LIKELIHOOD AND SAMPLE MOMENT ESTIMATORS FOR THE MTH (CENTRAL) MOMENT IN A NORMAL AND GENERALIZED GAMMA POPULATION

Approved by:

Date of Approval: 11/12/2020.

ABSTRACT

ON MAXIMUM LIKELIHOOD AND SAMPLE MOMENT ESTIMATORS FOR THE MTH (CENTRAL) MOMENT IN A NORMAL AND GENERALIZED GAMMA POPULATION

HADI ABBASZADEHPEIVASTI

Industrial Engineering M.Sc. THESIS, December 2020

Thesis Supervisor: Prof. Dr. J.B.G Frenk

Keywords: maximum likelihood, moment, central moment, normal distribution, generalized gamma distribution, Pearson method

In this thesis, we consider the maximum likelihood and sample moment estimator of the mth (central) moment for a normal and generalized gamma population. We also propose using the method of moments approach new estimators for the parameters of a generalized gamma population.
To introduce these maximum likelihood estimators for the mth (central) moments in a generalized gamma population we first discuss the properties of the maximum likelihood optimization problem formulated for this class and propose an efficient algorithm to solve this optimization problem. As an application of this algorithm, we show how it can be used for a given sample to estimate the maximum likelihood estimator of the mth moment of a generalized gamma distributed random variable. By means of simulation experiments, we compare in the computational section its mean squared error with the mean squared error of the mth sample moment estimator.
An alternative estimator of these parameters is also proposed by applying the method of moment approach applied to the logarithmic transformation of a generalized gamma distributed random variable. Although the associated system of nonlinear equations is for small sample sizes inconsistent with a high probability, the system of nonlinear equations has a unique solution (if there is a solution) and this unique solution is easy to determine. For larger sample sizes this probability goes to zero and this is related to the accuracy of the sample moment estimator

of the skewness of a generalized gamma population. Hence it is easy to determine evaluating only the sample whether the system is inconsistent. These properties are not proved for other proposals of moment estimators of the parameters of this class which appeared in the literature.

Finally, we propose for any positive integer m a maximum likelihood-based estimator of the mth (central) moment in a normal population and compare the behavior of this estimator with the (classical) sample mth (central) moment estimator. In particular, we give for every computable expression for the mean and the variance of these different estimators for both the moment and the central moment estimation problem. For the mth central moment estimation problem it is shown that in a normal population one can compute a threshold value (independent of the unknown parameters) of the sample size such that beyond this sample size the mean squared error of the maximum likelihood-based estimator is smaller than the mean squared error of the sample mth central moment estimator. At the same time, this shows using the mean squared error objective that for sample sizes below a certain value the nonparametric sample moment estimator outperforms the parametric maximum likelihood-based estimator. Finally, in the computational section, we perform for these two estimation problems some simulation experiments and give some rule of thumbs for which sample sizes it is better to use the nonparametric moment estimator.

ÖZET

NORMAL VE GENELLEŞTIRILMIŞ BIR GAMMA POPÜLASYONUNDAKI M'INCI (MERKEZI) MOMENT IÇIN MAKSIMUM OLABILIRLIK VE ÖRNEK MOMENTI TAHMIN EDICISI ÜZERINE

HADI ABBASZADEHPEIVASTI

ENDÜSTRİ MÜHENDİSLİĞİ YÜKSEK LİSANS TEZİ, December 2020

Tez Danışmanı: Prof. Dr. J.B.G Frenk

Anahtar Kelimeler: maksimum olabilirlik, moment, merkezi moment, normal dağılım, genelleştirilmiş gamma dağılımı, Pearson yöntemi

Bu tezde, normal ve genelleştirilmiş bir gamma popülasyonu için m'inci (merkezi) momentin maksimum olasılığını ve örnek moment tahmin edicisini ele alıyoruz. Ayrıca, genelleştirilmiş bir gamma popülasyonunun parametreleri için yeni tahmin ediciler yaklaşımı yöntemini kullanmayı da öneriyoruz.
Genelleştirilmiş bir gamma popülasyonundaki mth (merkezi) anlar için bu maksimum olasılık tahmin edicilerini tanıtmak için, ilk olarak bu sınıf için formüle edilen maksimum olasılık optimizasyon probleminin özelliklerini tartışıyoruz ve bu optimizasyon problemini çözmek için verimli bir algoritma öneriyoruz. Bu algoritmanın bir uygulaması olarak, genelleştirilmiş bir gamma dağıtılmış rasgele değişkenin m'inci momentinin maksimum olasılık tahmin edicisini tahmin etmek için belirli bir örnek için nasıl kullanılabileceğini gösteriyoruz. Simülasyon deneyleri aracılığıyla, hesaplama bölümünde ortalama kare hatası ile m'inci örnek moment tahmin edicisinin ortalama kare hatası ile karşılaştırıyoruz.
Bu parametrelerin alternatif bir tahmincisi, genelleştirilmiş bir gamma dağıtılmış rasgele değişkenin logaritmik dönüşümüne uygulanan moment yaklaşımı yönteminin uygulanmasıyla da önerilmektedir. İlişkili doğrusal olmayan denklem sistemi, yüksek olasılıkla tutarsız küçük örnek boyutları için olmasına rağmen, doğrusal olmayan denklem sisteminin benzersiz bir çözümü vardır (bir çözüm varsa) ve bu benzersiz çözümün belirlenmesi kolaydır. Daha büyük örneklem boyutları için bu olasılık sıfıra gider ve bu, genelleştirilmiş bir gamma popülasyonunun çarpıklığının

örnek moment tahmin edicisinin doğruluğu ile ilgilidir. Bu nedenle, sistemin tutarsız olup olmadığını yalnızca numuneyi değerlendirerek belirlemek kolaydır. Bu özellikler, literatürde ortaya çıkan bu sınıfa ait parametrelerin moment tahmin edicilerinin diğer önerileri için ispatlanmamıştır.

Son olarak, herhangi bir pozitif tam sayı için, normal bir popülasyonda m'inci (merkezi) momentin maksimum olasılığa dayalı bir tahmin edicisini öneriyoruz ve bu tahmin edicinin davranışını (klasik) örnek mth (merkezi) moment tahmin edicisi ile karşılaştırıyoruz. Özellikle, hem moment hem de merkezi moment tahmin problemi için bu farklı tahmin edicilerin ortalaması ve varyansı için her hesaplanabilir ifade veriyoruz. Mnci merkezi moment tahmin problemi için, normal bir popülasyonda, örnek büyüklüğünün bir eşik değeri (bilinmeyen parametrelerden bağımsız olarak) hesaplanabileceği ve bu örnek büyüklüğünün ötesinde, maksimum olasılığa dayalı tahmin edicinin ortalama kare hatasının şu şekilde hesaplanabileceği gösterilmiştir. orta moment tahmin edicisinin ortalama kare hatasından daha küçük. Aynı zamanda, bu, belirli bir değerin altındaki örnek büyüklükleri için ortalama kare hata hedefini kullanarak, parametrik olmayan örnek moment tahmin edicisinin parametrik maksimum olasılığa dayalı tahmin ediciden daha iyi performans gösterdiğini gösterir. Son olarak, hesaplama bölümünde, bu iki tahmin problemi için bazı simülasyon deneyleri gerçekleştiriyoruz ve hangi örnek boyutları için parametrik olmayan moment tahmin ediciyi kullanmanın daha iyi olduğu bazı temel kurallar veriyoruz.

ACKNOWLEDGEMENTS

This thesis is dedicated to my beloved family who supported me during the process. Overcoming this academic ordeal was impossible if I did not have the privilege of their mental and physical support.

My gratitude towards my advisor, Dr. J.B.G Frenk, is impossible to be expressed in words. He not only taught me academic ethics but also the fundamentals of being a better scientist to make the world a better place to live. I wish him the best of luck in his career and hope our academic paths cross again someday in future.

This thesis is dedicated to my beloved family

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1. Introduction

In this thesis, we consider the properties of (central) moment estimators. We analyze two different estimators for the (central) moment. The first one is the well-known sample (central) moment estimator and the second one is the maximum likelihood estimator (MLE). In the sample (central) moment estimator, we have no prior information about parameters or the distribution of the sample and intend to calculate the mth (central) moment without any knowledge about the data. However, In the maximum likelihood estimator, we have some knowledge about the data including the distribution which the sample comes from. Then, we use this data to estimate the parameters of the distribution. Knowing the parameters of a distribution provides us with the opportunity to calculate the mth (central) moment analytically. Since in most cases the exact values of the parameters are not known in advance, we need to estimate them. One of the most important methods to estimate the parameters is using the maximum likelihood estimator. In this thesis, we consider this method to estimate the unknown parameters and substitute them in the analytical formula for mth (central) moment estimator, which we call this method as maximum likelihood estimator. We will analyse these methods for two widely used distributions including generalized gamma and normal distributions. Moreover, in the generalized gamma case we benefit the properties of methods of moment to estimate the parameters rather than using the maximum likelihood estimator.

In Chapter 2, we focus on the generalized gamma distribution. First, we provide a specific algorithm to calculate the unknown parameters of the gamma distribution. Next, we analyse the generalized gamma distribution when the parameter (α) is known. We carefully utilize the properties of generalized gamma distribution and develop an algorithm based on the MLE method to find an estimation for the unknown parameters. For the generalized gamma distribution, we also consider the case in which all of the parameters are unknown. Moreover, for this distribution, we propose an algorithm to find the unknown parameters. For all of these cases, we draw a comparison between the results of the sample (central) moment estimator and the MLE by means of simulations.

Since the maximum likelihood estimator does not always provide a decent estimation for the parameters in generalized gamma case, in chapter 3, we use the methods of moment and Pearson coefficient to calculate the parameters.

In Chapter 4, we consider normal distribution and develop an analytical formula to calculate the mean, variance, and mean squared error (MSE) for sample (central) moment estimator and MLE. Finally, we carry out a number of simulations and compute the analytical formula for specific examples with the aim of comparing the two previously mentioned estimators.

2. On the maximum likelihood approach applied to a (generalized) gamma population

2.1 Introduction.

The gamma function is introduced by the Swiss mathematician Leonhard Euler in 1738 in a paper on transcendental progression (cf.[1]). The related gamma distribution became much later a popular distribution in statistics, probability theory and stochastic processes. Next to the normal distribution it is probably the most used distribution within statistics. To start our discussion of the literature on the maximum likelihood approach applied to the (generalized) gamma distribution we observe that in probability theory most likely the first author discussing the gamma distribution (as a posterior distribution arising in a certain problem) is Laplace in 1812 (cf.[2]). In statistics Pearson (cf.[3]) was one of the first authors mentioning it as an continuous approximation of the chi-square statistic used for tests in contingency tables. For a more extensive survey on the gamma distribution in the history of statistics one should consult [4]. Due to its flexibility and mathematical simplicity it gained a lot of popularity over the years as a mathematical model to fit a sample of continuous non-negative data. In fitting data to a parametric class of distributions by means of the maximum likelihood principle one needs to solve for the parametric class of gamma distributions a two parameter (scale and shape parameter) continuous maximization problem over the positive two-dimensional orthant. In most textbooks on statistics (for example see [5]) one discusses the theoretical properties of the maximum likelihood approach. As an example one mostly considers the maximum likelihood estimation problem for a normal population by setting the gradient of the log likelihood function equal to zero. Due to its special form in the normal case this set of equations has a unique closed form solution. In more computational oriented books on statistics (see for example ([6] one refers to a numerical method like a Newton or gradient type method to find a solution of this set of equations. However, for a gamma population it is possible eliminate the scale parameter and write it as a function of the shape parameter. This shows that the two parameter system of two nonlinear equations can be written as a simple nonlinear function of the shape parameter. Using this approach one reduces the two parameter optimization problem to a one (shape) parameter optimization problem. Solving the associated first order conditions satisfied by any optimal solution one then applies a zero-root finding algorithm from numerical mathematics. Also tables are used to determine a solution of this one dimensional nonlinear system (cf.[4]). In this paper we analyze this optimization problem in more depth using global properties of the digamma and related gamma-type functions. Next to verifying the uniqueness of the optimal solution under certain general conditions we propose a very stable and fast special purpose algorithm to solve this problem. In a recent paper (cf.[7]) one also discusses the problem of finding the parameters of a gamma distribution by proposing a related set of nonlinear equations which have a closed form solution. These equations are given by the set of stationary equations of the three parameter generalized gamma distribution with the third parameter τ set equal to 1 (see Definition 2.2.1). By setting $\tau = 1$ a generalized gamma distribution reduces to a gamma distribution.

To extend the flexibility of the gamma distribution (motivated by applications in reliability theory) a more general three parameter class of distributions called the generalized gamma distribution was proposed by Stacy (cf.[8]). Again applying the principle of maximum likelihood to this class of distributions one need to solve a three parameter optimization problem over the non-negative orthant.

Stacy and Mihram (cf.[9]) were among the first to discuss parameter estimation by means of maximum likelihood of the parameters of a generalized gamma distribution. In this paper it is mentioned that a corresponding set of stationary equations need to be considered without supplying any detailed analysis of these equations. However, instead of focusing on the maximum likelihood approach the main focus of this paper was on applying the method of moments principle based on the expectation of the logarithm of a generalized gamma distributed random variable to obtain estimates of these parameters. Other examples of studies considering the maximum likelihood estimation problem for a generalized gamma distribution (among related topics) are [10] and [11]. In both papers the first order conditions of the associated loglikelihood function are worked out in more detail than done by Stacey and Mihran and a general iterative procedure is applied to solve these equations. A similar less detailed approach is followed by Harter (cf.[12]) covering both complete and sensored samples for a generalized gamma distribution with an unknown shift.

Considering the log-generalized gamma distribution and applying the log transformation to the given sample Prentice (cf.([13]) and Lawless (cf.[14]) analysed the maximum likelihood estimation problem for a generalised gamma distribution by means of the maximum likelihood estimation problem for a log generalized gamma distribution. For the last distribution the first order conditions are written down and in both papers different iterative procedures are used to solve the corresponding system of nonlinear equations. Since for some setting of the parameters the previous mentioned papers reported numerical difficulties Gomes et al [15] proposed in 2008 a heuristic approach to solve approximately the maximum likelihood estimation problem for a generalized gamma distribution. Recently Ling (cf.[16]) discussed in a specific application (one shot device testing under accelerated life test) the different methods (Maximum likelihood estimation, least squares method) to estimate the parameters of a generalized gamma distribution.

To conclude our introduction of the maximum likelihood approach for a generalized gamma population we mention that in most of the above papers also method of moments estimators are proposed for the generalized gamma distribution. An example of a paper only discussing these type of estimators is [17]. In general method of moments estimators are much easier to implement but lack certain desirable asymptotic statistical properties satisfied by estimators based on Maximum likelihood estimation (cf.[18]). These estimators mostly serve as an alternative if the Maximum likelihood estimation approach cannot be applied.

In this paper the main focus is on analyzing the properties of the maximum likelihood optimization problem for a (generalized gamma) population. The analysis of of this optimization problem and its subcases is much more detailed as seen in the literature and seems to be new making use of some properties of the so-called digamma function. A byproduct of this analysis are the special purpose algorithms to solve this problem for different subcases. Also we discuss an application in the computational section comparing the mean squared errors of the mth sample moment estimator with the maximum likelihood type moment estimator of the mth moment in such a population. An important reason to compare these mean squared errors of both estimators is the following. As observed in the computational section the optimal solution of the maximum likelihood optimization problem is not very close to the value of the real parameters in a generalized gamma population. Due to this relative big error the maximum likelihood estimator for the mth moment is much more inaccurate as the estimates given by the mth sample moment estimator. Due to the relative big error for small sample sizes of the maximum likelihood approach estimating the real parameters and the relative good performance of sample moment estimators this suggests that for small sample sizes it might be advisable to use the method of moment approach of Pearson to estimate the unknown parameters instead oıf the maximum likelihood approach. This is also observed in [17]. In Section 2 we will introduce the different estimators and analyse in detail the maximum likelihood optimization problem associated with a (generalized) gamma population starting with the gamma distribution and extending it to the generalized gamma distribution.We also show how to use these algorithms to derive alternative maximum likelihood estimators for the mth moment. Finally in Section 3 we will report some simulation experiments comparing the efficiency of both estimators.

2.2 On the maximum likelihood and sample moment estimator for the mth moment.

We first list the following well-known definition (cf.[18], [8]). Observe the notation $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$ means that the random variables $\mathbf{X}$ and $\mathbf{Y}$ have the same cumulative distribution function.

Definition 2.2.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space.*

1.1 The nonnegative random variable $\mathbf{X}$ defined on this probability space has a gamma distribution with scale parameter 1 and shape parameter $\alpha > 0$ if the density f of the random variable $\mathbf{X}$ is given by

$$f(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} 1_{(0,\infty)}(x)$$

with

$$\Gamma(\alpha) = \int_0^\infty x^\alpha e^{-x} dx$$

the so-called gamma function. This is denoted by $\mathbf{X} \sim G(\alpha, 1)$. The random variable $\mathbf{X}$ has a gamma distribution with scale parameter $\beta > 0$ and shape parameter $\alpha > 0$ if $\mathbf{X} \stackrel{d}{=} \beta\mathbf{Y}$ with $\mathbf{Y} \sim G(\alpha, 1)$.

1.2 A nonnegative random variable $\mathbf{X}$ has an inverse gamma distribution with parameters $\alpha, \beta > 0$ if $\mathbf{X} \stackrel{d}{=} \frac{1}{\mathbf{Y}}$ with $\mathbf{Y} \sim G(\alpha, \beta)$.

1.3 A nonnegative random variable $\mathbf{X}$ has a generalized gamma distribution with parameters $\beta > 0, \alpha > 0, \tau > 0$ if

$$\mathbf{X} \stackrel{d}{=} \beta\mathbf{Y}^{\tau^{-1}}$$

with $\mathbf{Y} \sim G(\alpha, 1)$. *This is denoted by* $\mathbf{X} \sim GG(\alpha, \beta, \tau)$

Clearly the family of generalized gamma distributions is closed under power transformations and scaling and applying standard calculus its density function is given by

$$f_{\mathbf{X}}(x) = \frac{\tau x^{\tau\alpha - 1} e^{-(\frac{x}{\beta})^{\tau}}}{\beta^{\tau\alpha}\Gamma(\alpha)} 1_{(0,\infty)}(x). \tag{2.1}$$

This class of density functions is studied in [8].The main reason to introduce this generalization of the gamma distribution is to have more flexibility in modelling life time data. Observe the Weibull distribution and lognormal distribution (being a limiting case) are subfamilies of this class. Using Definition 2.2.1 it follows for $\mathbf{X} \sim GG(\alpha, \beta, \tau)$ that its mth moment $\mu'_m(\mathbf{X}) := \mathbb{E}(\mathbf{X}^m), m \in \mathbb{N}$ is given by

$$\mu'_m(\mathbf{X}) = \beta^m \mu'_m(\mathbf{Y}^{\tau^{-1}}) = \frac{\beta^m \Gamma(m\tau^{-1} + \alpha)}{\Gamma(\alpha)}, m \in \mathbb{N}. \tag{2.2}$$

Applying the invariance principle of maximum likelihood estimators (cf [18], [5]) and relation (2.2) a maximum likelihood estimator $\widehat{\mu}'_{m,M}$ of the mth moment $\mu'_m(\mathbf{X})$ of a gamma distribution is given by

$$\widehat{\mu}'_{m,ML}(\mathbf{X}_1, ..., \mathbf{X}_n) = \frac{\widehat{\theta}_1^m \Gamma(m + \widehat{\theta}_2)}{\Gamma(\widehat{\theta}_2)} \tag{2.3}$$

with $\widehat{\theta}_1, \widehat{\theta}_2$ the maximum likelihood estimators of the scale parameter $\theta_1 = \beta > 0$ and shape parameter $\theta_2 = \alpha > 0$. For the generalized gamma distribution this maximum likelihood estimator $\widehat{\mu}'_{m,M}$ is given by

$$\widehat{\mu}'_{m,ML}(\mathbf{X}_1, ..., \mathbf{X}_n) = \frac{\widehat{\theta}_1^m \Gamma(m\widehat{\theta}_3^{-1} + \widehat{\theta}_2)}{\Gamma(\widehat{\theta}_2)}, \tag{2.4}$$

with $\widehat{\theta}_1, \widehat{\theta}_2$ and $\widehat{\theta}_3$ the MLE estimators of the parameters $\theta_1 = \beta, \theta_2 = \alpha$ and $\theta_3 = \tau$. To apply these estimators we need a fast algorithm to compute the maximum likelihood estimates of the parameters and so in the next subsection we will analyse in detail the maximum likelihood optimization problem for a (generalised) gamma population. Clearly an alternative estimator is to use the sample mth moment estimator $\widehat{\mu}'_m$ given by

$$\widehat{\mu}'_m(\mathbf{X}_1, ..., \mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i^m. \tag{2.5}$$

We will compare the efficiency of both estimators in the computational section.

2.2.1 On the maximum likelihood optimization problem for a (generalized) gamma population.

In this subsection we will study the properties of the maximum likelihood optimization problem for a (generalized) gamma population and start with the special case of a a gamma population. Introducing for any function $p:(0,\infty)\to\mathbb{R}$ the notation $\overline{p(\mathbf{x})}:=\frac{1}{n}\sum_{i=1}^{n}p(x_i)$ the loglikelihood function of a gamma population with unknown parameters $\alpha,\beta>0$ and sample $\mathbf{x}=(x_1,...,x_n)>\mathbf{0}$ is given by

$$(2.6)\qquad LL(\alpha,\beta;\mathbf{x})=n\left((\alpha-1)\overline{\ln(\mathbf{x})}-\beta^{-1}\overline{\mathbf{x}}-\ln(\Gamma(\alpha))-\alpha\ln(\beta)\right)$$

and applying the maximum likelihood principle we need to solve the maximum likelihood optimization problem

$$(P_1)\qquad v(P_1)=\sup_{\alpha,\beta>0}\{LL(\alpha,\beta;\mathbf{x})\}.$$

Before discussing this optimization problem we introduce for any random sample $\mathbf{x}>\mathbf{0}$ the so-called L_τ-norm, $0<\tau<\infty$ given by

$$(2.7)\qquad \parallel\mathbf{x}\parallel_\tau:=(\sum_{i=1}^{n}x_i^\tau)^{\frac{1}{\tau}}$$

and the maxnorm given by

$$\parallel\mathbf{x}\parallel_\infty=\max\{x_1,...,x_n\}.$$

Also we introduce the function $G:(0,\infty)\to\mathbb{R}$ given by

$$(2.8)\qquad G(\tau):=\ln(\overline{\mathbf{x}^\tau})-\overline{\ln(\mathbf{x}^\tau)}$$

and the function $K:(0,\infty)\to\mathbb{R}$ given by

$$(2.9)\qquad K(\tau):=\ln\left(\parallel\parallel\mathbf{x}\parallel_\infty^{-1}\mathbf{x}\parallel_\tau^\tau\right).$$

It is easy to check for every $\tau>0$

$$(2.10)\qquad G(\tau)=\tau(\ln(\parallel\mathbf{x}\parallel_\infty)-\overline{\ln(\mathbf{x})})+K(\tau)-\ln(n).$$

The function G plays an important role in the maximum likelihood optimization problems to be discussed and so we first derive its main properties.

Lemma 2.2.1. *If the sample $\mathbf{x}=(x_1,...,x_n)>\mathbf{0}$ contains at least two different* *realisations, then*

2.1 *The function* $G:(0,\infty)\to\mathbb{R}$ *given by relation (2.8) is strictly convex, positive and strictly increasing satisfying*

$$(2.11)\qquad \lim_{\tau\downarrow 0}\frac{G(\tau)}{\tau^2}=\frac{\overline{\ln^2(\mathbf{x})}-\overline{\ln(\mathbf{x})}^2}{2}>0$$

and

$$(2.12)\qquad \lim_{\tau\uparrow\infty} G(\tau)-\tau\left(\ln(\|\mathbf{x}\|_\infty)-\overline{\ln(\mathbf{x})}\right)=\ln\left(\frac{|S|}{n}\right)<0.$$

with $S=\{1\leq i\leq n: x_i=\|\mathbf{x}\|_\infty\}$ *and* $|S|$ *denoting the the cardinality of the set* S*. Its derivative* $G^{(1)}$ *satisfies*

$$(2.13)\qquad G^{(1)}(0^+)=0, G^{(1)}(\infty)=\ln(\|\mathbf{x}\|_\infty)-\overline{\ln(\mathbf{x})}>0.$$

2.2 *The function* $\tau\to G(\tau)-\tau\left(\ln(\|\mathbf{x}\|_\infty-\overline{\ln(\mathbf{x})}\right)$ *is negative, strictly decreasing and strictly convex.*

Proof. Since the random sample $\mathbf{x}=(x_1,...,x_n)>\mathbf{0}$ contains at least two different realisations it follows by the strict concavity of the logarithm that $G(\tau)>0$ for every $\tau>0$. It is easy to check that the function G is twice continuously differentiable on $(0,\infty)$ having first derivative

$$(2.14)\qquad G^{(1)}(\tau)=\sum_{i=1}^n\frac{x_i^\tau}{\|\mathbf{x}\|_\tau^\tau}\ln(x_i)-\overline{\ln(\mathbf{x})}$$

and second derivative

$$(2.15)\qquad G^{(2)}(\tau)=\sum_{i=1}^n\frac{x_i^\tau}{\|\mathbf{x}\|_\tau^\tau}(\ln(x_i))^2-\left(\sum_{i=1}^n\frac{x_i^\tau}{\|\mathbf{x}\|_\tau^\tau}\ln(x_i)\right)^2=Var(\mathbf{X}_\tau).$$

Observe the random variable $\mathbf{X}_\tau$ defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$ has the probability distribution function

$$(2.16)\qquad \mathbb{P}(\mathbf{X}_\tau=\ln(x_i))=\frac{x_i^\tau}{\|\mathbf{x}\|_\tau^\tau}, i=1,...,n.$$

Since the sample $\mathbf{x}>\mathbf{0}$ has at least two different realizations it follows using relation (2.14) and $\lim_{\tau\downarrow 0}x_i^\tau\|\mathbf{x}\|_\tau^{-\tau}=\frac{1}{n}$ that $G^{(1)}(0^+)=0$. Also this condition implies $\mathbb{P}(\mathbf{X}_\tau=\mathbb{E}(\mathbf{X}_\tau))<1$ and so by relation (2.15) $G^{(2)}(\tau)>0$ for every τ. Hence by Theorem A and Theorem C in Section 12 of [19] the function G is strictly convex and strictly increasing. It is easy to verify that $G(0+)=0$ and this implies G is also positive showing the first statement of this lemma. To verify relation (2.11) we conclude from relation (2.15) that

$$\lim_{\tau\downarrow 0} G^{(2)}(0^+) = \overline{\ln^2(\mathbf{x})} - \overline{\ln(\mathbf{x})}^2 \tag{2.17}$$

and relation (2.11) follows applying Taylors theorem and $G(0+) = G^{(1)}(0^+) = 0$. To show relation (2.12) we observe with S^c the complement of the set S that

$$\begin{aligned}
\lim_{\tau\uparrow\infty} \|\| \mathbf{x} \|_\infty^{-1} \mathbf{x} \|_\tau^\tau &= \lim_{\tau\uparrow\infty} \sum_{i=1}^n (\| \mathbf{x} \|_\infty^{-1} x_i)^\tau \\
&= \lim_{\tau\uparrow\infty} \sum_{i\in S^c} (\| \mathbf{x} \|_\infty^{-1} x_i)^\tau + \mid S \mid \\
&= \mid S \mid
\end{aligned}$$

and applying relations (2.9) and (2.10) we obtain relation (2.12). To check the second part of relation (2.13) we observe for $i \in S$ and using $\lim_{\tau\uparrow\infty} \| \mathbf{x} \|_\tau \downarrow \| \mathbf{x} \|_\infty$ (cf.[19]) that

$$0 \le \lim_{\tau\uparrow\infty} \| \mathbf{x} \|_\tau^{-\tau} x_i^\tau \le \lim_{\tau\uparrow\infty} (\| \mathbf{x} \|_\infty x_i)^\tau = 0. \tag{2.18}$$

Since $\sum_{i=1}^n \| \mathbf{x} \|_\tau^{-\tau} x_i^\tau = 1$ this implies by (2.18) that $\lim_{\tau\uparrow\infty} \sum_{i\in S} \| \mathbf{x} \|_\tau^{-\tau} x_i^\tau = 1$ and by relation (2.14)

$$G^{(1)}(\infty) = \lim_{\tau\uparrow\infty} G^{(1)}(\tau) = \ln(\| \mathbf{x} \|_\infty) - \overline{\ln(\mathbf{x})}. \tag{2.19}$$

The proof of part 2 is obvious applying part 1 and relation (2.10) and we have verified the lemma. □

Since with probability 1 any sample of a gamma population has a unique maximum we obtain from relation (2.12) with probability 1 over the space of all samples that

$$\lim_{\tau\uparrow\infty} G(\tau) - \tau G^{(1)}(\infty) = -\ln(n) < 0. \tag{2.20}$$

One can now show the following result for the loglikelihood optimization problem (P_1) of a gamma population.

Lemma 2.2.2. *If $\mathbf{x} = (x_1, ..., x_n) > \mathbf{0}$ is a random sample from a gamma population with unknown parameters α, β, then the maximum likelihood optimization problem (P_1) reduces to*

$$v(P_1) = -n\overline{\ln(\mathbf{x})} + n \sup_{\alpha>0}\{H_0(\alpha) - \alpha G(1)\} \tag{2.21}$$

with the function H_0 defined in relation (A.10) and the function G in relation (2.8)

.

Proof. By relation (2.6) we obtain for every $\alpha > 0$
(2.22)
$\sup_{\beta>0}\{LL(\alpha,\beta;\mathbf{x}\} = -n\overline{\ln(\mathbf{x})} + n\left(\alpha\overline{\ln(\mathbf{x})} - \ln(\Gamma(\alpha) + \sup_{\beta>0}\{-\beta^{-1}\overline{\mathbf{x}} - \alpha\ln(\beta)\}\right).$

Since for every $\alpha > 0$ and $\mathbf{x} > \mathbf{0}$

$$\lim_{\beta\downarrow 0} -\beta^{-1}\overline{\mathbf{x}} - \alpha\ln(\beta) = \lim_{\beta\uparrow\infty} -\beta^{-1}\overline{\mathbf{x}} - \alpha\ln(\beta) = -\infty$$

it follows by standard arguments taking the derivative that the optimal solution of the optimization problem $\sup_{\beta>0}\{-\beta^{-1}\overline{\mathbf{x}} - \alpha\ln(\beta)\}$ is given by $\beta_{opt}(\alpha) = \frac{\overline{\mathbf{x}}}{\alpha}$ and

$$\sup_{\beta>0}\{-\beta^{-1}\overline{\mathbf{x}} - \alpha\ln(\beta)\} = -\alpha\ln(\overline{\mathbf{x}}) + \alpha(\ln(\alpha) - 1).$$

This shows by relations (2.8) and (2.22) that

$$\sup_{\alpha,\beta>0}\{LL(\alpha,\beta;\mathbf{x})\} = -n\overline{\ln(\mathbf{x})} + n\sup_{\alpha>0}\{H_0(\alpha) - \alpha G(1)\}$$

and we have verified the result. □

To show that the optimization problem $\sup_{\alpha>0}\{H_0(\alpha) - \alpha G(1)\}$ has a unique optimal solution and to determine this optimal solution we need the following result.

Lemma 2.2.3. *If the sample* $\mathbf{x} > \mathbf{0}$ *from a gamma population contains at least two different realisations then* $\sup_{\alpha>0}\{H_0(\alpha) - \alpha G(1)\}$ *has an unique optimal solution* $\alpha_{opt}(1)$ *with* $\alpha_{opt}(\tau) = h_0^{\leftarrow}(G(\tau))$ *and* $h_0^{\leftarrow}$ *the inverse function of the function* h_0 *listed in relation (A.9).*

Proof. We observe by Lemma A.0.6 in the Appendix A that H_0 is a strictly decreasing, continuous strictly concave function satisfying

$$\lim_{\alpha\downarrow 0} H_0(\alpha) - \ln(\alpha) = 0 \tag{2.23}$$

and

$$\lim_{\alpha\uparrow\infty} H_0(\alpha) - \frac{1}{2}\ln(\alpha) + \frac{1}{2}\ln(2\pi) = 0 \tag{2.24}$$

Since the random sample $\mathbf{x} > \mathbf{0}$ contains at least two different realisations this implies by the first part of Lemma 2.2.1 that

$$\lim_{\alpha\uparrow\infty} H_0(\alpha) - \alpha G(1) = \lim_{\alpha\downarrow 0} H_0(\alpha) - \alpha G(1) = -\infty.$$

Hence the optimization problem $\sup_{\alpha>0}\{H_0(\alpha)-\alpha G(1)\}$ has a unique optimal solution $0<\alpha_{opt}(1)<\infty$ and taking the derivative of the objective function we obtain the desired result. $\square$

To solve the strictly concave maximization problem $\sup_{\alpha>0}\{H_0(\alpha)-\alpha G(1)\}$ in an efficient way we observe for $k\in\mathbb{N}$ using relation (A.10) and $\Gamma(k)=(k-1)!$ hat

$$H_0(k)=k(\ln(k)-1))-\ln((k-1)!).$$

Introducing the first order difference operator given by

$$\Delta p(k):=p(k+1)-p(k), k\in\mathbb{N}$$

we obtain for $p(\alpha):=H_0(\alpha)-\alpha G(1)$ that

$$\Delta p(k)=(k+1)\ln\left(1+\frac{1}{k}\right)-1+\overline{\ln(\mathbf{x})}-\ln(\overline{\mathbf{x}}). \tag{2.25}$$

Since the function p is strictly concave this shows by a similar proof as in Lemma 2.2.3 that an optimal solution of the related optimization problem $\sup_{k\in\mathbb{N},\beta>0}LL(k,\beta;\mathbf{x})$ with the objective function listed in relation (2.6) is given by $(k_{opt},\beta_{opt}(k_{opt})$ with

$$\beta_{opt}(k)=\frac{\overline{\mathbf{x}}}{k}$$

and using relation (2.25)

$$\begin{aligned} k_{opt} &= \min\{k\in\mathbb{N}:\Delta p(k)\leq 0\} \\ &= \min\left\{k\in\mathbb{N}:(1+\tfrac{1}{k})^{k+1}\leq e^{1+\ln(\overline{x})-\overline{\ln(x)}}\right\} \end{aligned}$$

One can now show the following result for the optimization problem $\sup_{\alpha>0}\{H_0(\alpha)-\alpha G(1)\}$.

Lemma 2.2.4. *The optimal solution of optimization problem* $\sup_{\alpha>0}\{H_0(\alpha)-\alpha G(1)\}$ *is located in the interval* $(k_{opt}-1,k_{opt}+1]$.

Proof. Since by Lemma A.0.6 the function $p(\alpha)=H_0(\alpha)-\alpha G(1)$ is differentiable and strictly concave with a strictly decreasing derivative it follows by the supergradient inequality for every $k\in\mathbb{N}$ that $p^{(1)}(k)\geq\Delta p(k)=p(k+1)-p(k)$. This shows by the definition of k_{opt} that

$$p^{(1)}(k_{opt}-1)\geq\Delta p(k_{opt}-1)>0$$

Also by the mean value theorem (cf.[20]) there exists some $\zeta \in (k_{opt}, k_{opt}+1)$ satisfying

$$0 \geq \Delta p(k_{opt}) = p(k_{opt}+1) - p(k_{opt}) = p'(\zeta)$$

and this shows using p is concave and hence $p^{(1)}$ decreasing that $p^{(1)}(k_{opt}+1) \leq p^{(1)}(\zeta) \leq 0$. Since $p^{(1)}$ is continuous and strictly decreasing and $p^{(1)}(\alpha_{opt}(1)) = 0$ the result follows. □

By the above result we start our search for the unique optimal solution $\alpha_{opt}(1)$ of optimization problem $\sup_{\alpha>0}\{H_0(\alpha) - \alpha G(1)\}$ generated by a a sample consisting of at least two different realisations (observe this event happens with probability 1) at k_{opt}. Recall from Lemma 2.2.3 we need to solve the nonlinear system

$$(S_0) \qquad h_0(\alpha) = G(1)$$

with h_0 defined in relation (A.9). Since by Lemma A.0.5 the function h_0 is strictly completely monotone and hence strictly decreasing and convex it follows that the unique optimal solution $h_0^{\leftarrow}(G(1))$ of the system in (S_0) can be found by the Newton Raphson method (cf.[21],[22]) starting at $k_{opt}(\mathbf{x})$. Observe we may also solve the nonlinear system

$$(S_1) \qquad h_1(\alpha) - \alpha G(1) = \alpha(h_0(\alpha) - G(1)) = 0$$

and by Lemma A.0.4 the function $\alpha \to h_1(\alpha) - \alpha G(1)$ is strictly convex. Since by Lemma A.0.4 we know that $\frac{1}{2} < h_1(\alpha) < 1$ and $h_0(0+) = \infty$ we apply the Newton Raphson method for numerical stability to system (S_1) in case $k_{opt} = 1$ and to system (S_0) in case $k_{opt} > 1$. Also it is possible to apply the bisection method restricted to the interval $(k_{opt} - 1, k_{opt} + 1]$ if we do not have a procedure to compute the derivative of the function h_0. In our computational experiments using the Newton Raphson method we encountered for α close to zero some numerical instabilities solving system (S_0) due to $h_0(\alpha)$ large. Since $\frac{1}{2} < \alpha h_0(\alpha) < 1$ we did not encounter these numerical instabilities solving the equivalent system (S_1). To identify beforehand a region containing the solution $\alpha_{opt}(1)$ we observe by Lemma A.0.4 that $\frac{1}{2} < h_1(\alpha) < 1$ for every $\alpha > 0$ and this shows

$$(2\alpha_{opt}(1))^{-1} < h_0(\alpha_{opt}(1)) < \alpha_{opt}(1)^{-1}.$$

Since $h_0(\alpha_{opt}(1)) = G(1)$ we conclude $(2\alpha_{opt}(1))^{-1} < G(1) < \alpha_{opt}(1)^{-1}$ and this implies

$$\frac{1}{2G(1)} < \alpha_{opt}(1) < \frac{1}{G(1)}. \tag{2.26}$$

By Lemmas 2.2.2, 2.2.3 and 2.2.4 the maximum likelihood optimization problem for a gamma population with unknown parameters α, β is given by the following algorithm.

Algorithm 2.2.1. *Algorithm maximum likelihood for gamma population with unknown α and β.*

3.1 Compute $k_{opt} = \min\left\{k \in \mathbb{N} : (1+\frac{1}{k})^{k+1} \leq e^{1+\ln(\overline{\mathbf{x}})-\overline{\ln(\mathbf{x})}}\right\}$

3.2 Start Newton Raphson method in k_{opt} and for $k_{opt} = 1$ find optimal solution $\alpha_{opt}(1) \in (0,2]$ of system

$$h_1(\alpha) - \alpha G(1) = 0$$

Otherwise find optimal solution $\alpha_{opt}(1) \in (k_{opt}-1, k_{opt}+1]$ of the system

$$h_0(\alpha) - G(1) = 0$$

3.3 Set $\widehat{\theta}_2 = \alpha_{opt}(1) > 0$, $\widehat{\theta}_1 = \overline{\mathbf{x}}\widehat{\theta}_2^{-1}$

To estimate the mth moment in a gamma population we apply the above algorithm and use relation (2.3). In case we need to estimate the parameters α, β of an inverse gamma distribution using maximum likelihood it is obvious from the definition of an inverse gamma distribution that we replace the sample $\mathbf{x}$ by the sample $(\frac{1}{x_1}, ..., \frac{1}{x_n})$ in the definition of the function G and apply the above procedure.

We will now discuss in detail the maximum likelihood estimation problem for a generalized gamma population. It follows by relation (2.1) that the loglikelihood function of a sample of size n from a generalized gamma population is given by

$$LL(\alpha,\beta,\tau;\mathbf{x}) = n\left(\ln(\tau) + (\tau\alpha - 1)\overline{\ln(x)} - \ln(\Gamma(\alpha)\right) - \beta^{-\tau} \parallel \mathbf{x} \parallel_\tau^\tau - n\tau\alpha\ln(\beta) \tag{2.27}$$

The first optimization problem we discuss is for a generalized gamma population with unknown $\beta > 0, \tau > 0$ but known $\alpha > 0$ and so we need to analyse the optimization problem

$$(P_2) \qquad \upsilon(P_2) = \sup_{\beta>0,\tau>0}\{LL(\alpha,\beta,\tau;\mathbf{x})\},$$

while the second is for a generalized gamma population with unknown $\beta > 0, \tau > 0$

and $\alpha > 0$ given by

$$(P_3) \qquad \upsilon(P_3) = \sup_{\alpha>0,\beta>0,\tau>0}\{LL(\alpha,\beta,\tau;\mathbf{x})\}.$$

Optimization problem (P_2) can be used if we estimate the two unknown parameters of a Weibull distribution. This corresponds to the case $\alpha = 1$. In a similar way as for optimization problem (P_1) for a gamma population one can show the following result for optimization problem (P_2).

Lemma 2.2.5. *If $\mathbf{x} = (x_1,...,x_n) > \mathbf{0}$ is a sample from a generalized gamma population with α known, then*

$$\upsilon(P_2) = -n\overline{\ln(\mathbf{x})} + nH_0(\alpha) + n\sup_{\tau>0}\{\ln(\tau) - \alpha G(\tau)\} \tag{2.28}$$

with the functions G and H_0 defined in relations (2.8), respectively (A.10).

Proof. It follows by relation (2.27) that

$$\begin{aligned} \upsilon(P_2) &= \sup_{\tau>0}\{\sup_{\beta>0} LL(\alpha,\beta,\tau;\mathbf{x})\} \\ &= -n\overline{\ln(\mathbf{x})} + \sup_{\tau>0}\{n\ln(\tau) + n\tau\alpha\overline{\ln(\mathbf{x})} - n\ln(\Gamma(\alpha)) + \vartheta(\alpha,\tau)\} \end{aligned} \tag{2.29}$$

with

$$(P(\alpha,\tau)) \qquad \vartheta(\alpha,\tau) := \sup_{\beta>0}\{-\beta^{-\tau} \parallel \mathbf{x} \parallel_\tau^\tau - n\tau\alpha\ln(\beta)\}.$$

By standard techniques observing for any $\alpha,\tau > 0$ that

$$\lim_{\beta\downarrow 0} -\beta^{-\tau} \parallel \mathbf{x} \parallel_\tau^\tau - n\tau\alpha\ln(\beta) = -\infty, \lim_{\beta\uparrow\infty} -\beta^{-\tau} \parallel \mathbf{x} \parallel_\tau^\tau - n\tau\alpha\ln(\beta) = -\infty$$

and taking the derivative of the function $\beta \to \beta^{-\tau} \parallel \mathbf{x} \parallel_\tau^\tau - n\tau\alpha\ln(\beta)$ it follows that the optimal solution $\beta_{opt}(\alpha,\tau)$ of optimization problem $(P(\alpha,\tau))$ is given by

$$\beta_{opt}(\alpha,\tau)^\tau = \frac{\parallel \mathbf{x} \parallel_\tau^\tau}{n\alpha}.$$

Substituting this optimal solution in optimization problem $(P(\alpha,\tau))$ we obtain

$$\vartheta(\alpha,\tau) = n\alpha(\ln(n\alpha) - 1) - n\tau\alpha\ln(\parallel \mathbf{x} \parallel_\tau).$$

This shows using relation (2.29) that

$$\upsilon(P_2) = -n\overline{\ln(\mathbf{x})} + n\sup_{\tau>0}\{H_0(\alpha) + \ln(\tau) - \alpha G(\tau)\} \tag{2.30}$$

Applying relation (2.10) yields the desired result. $\square$

By Lemma 2.2.5 we need to solve the optimization problem $\sup_{\tau>0}\{\ln(\tau)-\alpha G(\tau)\}$.

Lemma 2.2.6. *If the sample* $\mathbf{x}>\mathbf{0}$ *from a (generalized) gamma population contains at least two different realisations, then the optimization problem*

$$\sup_{\tau>0}\{\ln(\tau)-\alpha G(\tau)\}$$

has an unique optimal solution $\tau_{opt}(\alpha)$ *given by* $\tau_{opt}(\alpha)=g_1^{\leftarrow}(\frac{1}{\alpha})$ *with* $g_1^{\leftarrow}$ *the inverse function of the strictly increasing function* $g_1:(0,\infty)\to(0,\infty)$ *given by*

$$g_1(\tau)=\tau G^{(1)}(\tau). \tag{2.31}$$

Proof. By Lemma 2.2.1 we may conclude for every $\alpha>0$ that the function $\tau\to\ln(\tau)-\alpha G(\tau)$ is strictly concave and

$$\lim_{\tau\downarrow 0}\ln(\tau)-\alpha G(\tau)=\lim_{\tau\uparrow\infty}\ln(\tau)-\alpha G(\tau)=-\infty$$

for every $\alpha>0$. This shows the result using first order conditions and Lemma 2.2.1.

$\square$

To solve the optimization problem $\sup_{\tau>0}\{\ln(\tau)-\alpha G(\tau)\}$ in a computational fast and efficient way we apply the same approach as used in the maximum likelihood optimization problem for the gamma population. First we observe for $k\in\mathbb{N}$ that

$$G(k)=\ln\left(\overline{\mathbf{x}^k}\right)-\overline{\ln(\mathbf{x}^k)}=\ln\left(\frac{1}{n}\|\mathbf{x}\|_k^k\right)-k\overline{\ln(\mathbf{x})}$$

and so

$$\Delta G(k)=\ln\left(\frac{\|\mathbf{x}\|_{k+1}^{k+1}}{\|\mathbf{x}\|_k^k}\right)-\overline{\ln(\mathbf{x})}.$$

This shows introducing $p(\tau):=\ln(\tau)-\alpha G(\tau)$ that

$$\Delta p(k)=\ln\left(\frac{k+1}{k}\right)-\alpha\left(\ln\left(\frac{\|\mathbf{x}\|_{k+1}^{k+1}}{\|\mathbf{x}\|_k^k}\right)-\overline{\ln(\mathbf{x})}\right).$$

Since the function p is strictly concave an optimal solution k_{opt} of the related optimization problem $\sup_{k\in\mathbb{N}}\{\ln(k)-\alpha g(k)\}$ is given by

$$\begin{aligned} k_{opt} &= \min\{k\in\mathbb{N}:\Delta p(k)\leq 0\} \\ &= \min\left\{k\in\mathbb{N}:\ln\left(\tfrac{k+1}{k}\right)-\alpha\left(\ln\left(\frac{\|\mathbf{x}\|_{k+1}^{k+1}}{\|\mathbf{x}\|_k^k}\right)-\overline{\ln(\mathbf{x})}\right)\leq 0\right\} \end{aligned}$$

As in Lemma 2.2.4 one can show that the unique optimal solution of optimization problem

$$\sup_{\tau>0}\{\ln(\tau)-\alpha G(\tau)\}$$

is contained in the interval $(k_{opt}(\mathbf{x})-1,k_{opt}(\mathbf{x})+1]$ and by the previous results the algorithm to compute the MLE estimate for a generalized gamma population with unknown $\beta,\tau>0$ and α known is given by the following special procedure.

Algorithm 2.2.2. *Algorithm maximum likelihood for generalized gamma population with α known.*

4.1 Determine

$$k_{opt}(\mathbf{x})=\min\left\{k\in\mathbb{N}:\ln\left(\frac{k+1}{k}\right)-\alpha\left(\ln\left(\frac{\|\mathbf{x}\|_{k+1}^{k+1}}{\|\mathbf{x}\|_k^k}\right)-\overline{\ln(\mathbf{x})}\right)\leq 0\right\}$$

4.2 Start bisection method in k_{opt} restricted to the interval $(k_{opt}-1,k_{opt}+1]$ to find the unique solution $\tau_{opt}(\alpha)$ of the equation

$$g_1(\tau)=\tau G^{(1)}(\tau)=\frac{1}{\alpha}. \tag{2.32}$$

4.3 set $\widehat{\theta}_3=\tau_{opt}(\alpha)>0$, $\widehat{\theta}_1=\dfrac{\|\mathbf{x}\|_{\widehat{\theta}_3}}{(n\alpha)^{\widehat{\theta}_3^{-1}}}$

To estimate the mth moment in a generalised gamma population with α known we apply the above algorithm and use relation (2.4) with $\widehat{\theta}_2=\alpha$.

Before considering the maximum likelihood estimation problem for a generalised gamma distribution with unknown parameters α,β and τ we need to analyse the behaviour of the function $\alpha\to\tau_{opt}(\alpha)$ in the neighborhood of zero and infinity.

Lemma 2.2.7. *The function $\alpha\to\tau_{opt}(\alpha)$ is strictly decreasing and differentiable satisfying*

$$\lim_{\alpha\uparrow\infty 0}\alpha\tau_{opt}^2(\alpha)=\frac{1}{\overline{\ln^2(\mathbf{x})}-\overline{\ln(\mathbf{x})}^2} \tag{2.33}$$

and

$$(2.34) \qquad \lim_{\alpha\downarrow 0}\alpha\tau_{opt}(\alpha) = \frac{1}{\ln(\|\mathbf{x}\|_\infty) - \overline{\ln(\mathbf{x})}}.$$

Proof. By Lemma 2.2.1 it follows that the derivative $G^{(1)}$ of the function G is strictly increasing, nonnegative and differentiable and so the function $g_1(t) = tG^{(1)}(t)$ is strictly increasing, differentiable and nonnegative.This shows $g_1^{\leftarrow}$ is strictly increasing and differentiable and so $\tau_{opt}(\alpha) = g_1^{\leftarrow}(\alpha^{-1})$ is strictly decreasing and differentiable. It is easy to check that

$$\lim_{t\uparrow\infty}\frac{g_1(t)}{t} = G^{(1)}(\infty) = \ln(\|\mathbf{x}\|_\infty) - \overline{\ln(\mathbf{x})} > 0$$

and

$$\lim_{t\downarrow 0}\frac{g_1(t)}{t} = G^{(2)}(0^+) = \overline{\ln^2(\mathbf{x})} - \overline{\ln(\mathbf{x})}^2 > 0.$$

This implies

$$(2.35) \qquad \lim_{u\uparrow\infty}\frac{g_1^{\leftarrow}(u)}{u} = \frac{1}{\ln(\|\mathbf{x}\|_\infty) - \overline{\ln(\mathbf{x})}}, \lim_{u\downarrow 0}\frac{g_1^{\leftarrow}(u)}{\sqrt[2]{u}} = \frac{1}{\overline{\ln^2(\mathbf{x})} - \overline{\ln(\mathbf{x})}^2}.$$

and this shows the results in relations (2.33) and (2.34). □

We now analyse the maximum likelihood optimization problem for a generalized gamma population with unknown parameters α, β and τ. For an overview on the different sometimes heuristic procedures the reader is referred to [15]. We follow an exact approach analyzing in detail the properties of this optimization problem. By a similar proof as in Lemma 2.2.5 one can verify the following result.

Lemma 2.2.8. *If* $\mathbf{x} = (x_1, ..., x_n) > \mathbf{0}$ *is a random sample from a generalized gamma population with* α, β, τ *unknown then*

$$(2.36) \qquad \upsilon(P_3) = -n\overline{\ln(\mathbf{x})} + n\sup_{\alpha>0,\tau>0}\{H_0(\alpha) + \ln(\tau) - \alpha G(\tau)\}$$

with the function H_0 *listed in relation (A.10) and the function* G *and* K *in relation (2.8), respectively (2.9).*

By Lemma 2.2.8 we obtain that the maximum likelihood optimization problem for the generalized gamma distribution reduces to

$$(P_4) \qquad \upsilon(P_4) = \sup_{\alpha,\tau>0}\{H_0(\alpha) + \ln(\tau) - \alpha G(\tau)\}$$

and its optimal objective value satisfies

$$v(P_3) = -n\overline{\ln(\mathbf{x})} + nv(P_4).$$

Before analysing this optimization problem in detail we will indicate that this problem might lack desirable concavity properties. It ıs shown in the computational section by means of an example that this is indeed the case. Clearly by substituting $b = \frac{\tau}{\alpha}$ we obtain

$$\begin{aligned}
\sup_{\alpha>0,\tau>0}\{H_0(\alpha)+\ln(\tau)-\alpha G(\tau)\} &= \sup_{b>0,\alpha>0}\{H_0(\alpha)+\ln(\tfrac{b}{\alpha})-\alpha G(\tfrac{b}{\alpha})\} \\
&= \sup_{b>0,\alpha>0}\{H_0(\alpha)-\ln(\alpha)+\ln(b)-\alpha G(\tfrac{b}{\alpha})\}
\end{aligned}$$

By the perspective property of convex functions (cf.[23]) and G convex it follows that the function $(\alpha,b)\to \alpha G(\frac{b}{\alpha})$ is convex on $\mathbb{R}^2_+$. Also by relation (A.14) it follows that

$$H_0(\alpha)-\frac{1}{2}\ln(\alpha) = -\frac{1}{2}\ln(\sqrt[2]{2\pi})-\theta(\alpha)$$

and by Lemma A.0.2 the function $\alpha \to H_0(\alpha)-\frac{1}{2}\ln(\alpha)$ is concave. This shows that

$$\sup_{b>0,\alpha>0}\{H_0(\alpha)-\ln(\alpha)+\ln(b)-\alpha G(\frac{b}{\alpha})\} = \sup_{\alpha>0}\{H_0(\alpha)-\ln(\alpha)+\psi(\alpha)\} \tag{2.37}$$

with

$$\psi(\alpha) = \sup_{b>0}\{\ln(b)-\alpha G(\frac{b}{\alpha})\}. \tag{2.38}$$

Since $(\alpha,\beta)\to \ln(b)-\alpha G(\frac{b}{\alpha})$ is concave on $\mathbb{R}^2_+$ we obtain that ψ is concave and so the above optimization problem

$$\sup_{\alpha>0}\{H_0(\alpha)-\ln(\alpha)+\psi(\alpha)\}$$

consists of an objective function being the sum of the concave function $\alpha \to H_0(\alpha) - \frac{1}{2}\ln(\alpha)+\psi(\alpha)$ and the convex function $\alpha \to -\frac{1}{2}\ln(\alpha)$. Again by a bilevel approach first optimizing over τ for fixed α it follows by Lemma 2.2.6 that

$$\begin{aligned}
v(P_4) &= \sup_{\alpha>0}\{H_0(\alpha)+\sup_{\tau>0}\{\ln(\tau)-\alpha G(\tau)\}\} \\
&= \sup_{\alpha>0}\{H_0(\alpha)+\ln(\tau_{opt}(\alpha))-\alpha G(\tau_{opt}(\alpha))\} \qquad (2.39) \\
&= \sup_{\alpha>0} p(\alpha)
\end{aligned}$$

with the continuous objective function $p:(0,\infty)\to\mathbb{R}$ given by

$$p(\alpha) := H_0(\alpha)+\ln(\tau_{opt}(\alpha))-\alpha G(\tau_{opt}(\alpha)) \tag{2.40}$$

Observe it is easy to show that the function $\alpha\to\sup_{\tau>0}\{\ln(\tau)-\alpha G(\tau)\}$ is decreasing convex while H_0 is increasing concave. In the next lemma we compute the value of the objective function p at zero and infinity.

Lemma 2.2.9. *The objective function p in relation (2.40) satisfies*

$$p(0^+) = -1-\ln\left(\ln(\| \mathbf{x} \|_\infty)-\overline{\ln(\mathbf{x})}\right) \tag{2.41}$$

and

$$p(\infty) = -\frac{1}{2}-\frac{1}{2}\ln(2\pi)-\frac{1}{2}\ln\left(\overline{\ln^2(\mathbf{x})}-\overline{\ln(\mathbf{x})}^2\right). \tag{2.42}$$

Proof. By relation (A.11) and (A.12) we obtain

$$p(0^+) = \lim_{\alpha\downarrow 0} p(\alpha) = \lim_{\alpha\downarrow 0}\ln(\alpha\tau_{opt}(\alpha))-\alpha G(\tau_{opt}(\alpha)) \tag{2.43}$$

and

$$\begin{aligned} p(\infty) &= \lim_{\alpha\uparrow\infty} p(\alpha) \\ &= \lim_{\alpha\uparrow\infty}\tfrac{1}{2}\ln(\alpha)-\tfrac{1}{2}\ln(2\pi)+\ln(\tau_{opt}(\alpha))-\alpha G(\tau_{opt}(\alpha)) \\ &= -\tfrac{1}{2}\ln(2\pi)+\lim_{\alpha\uparrow\infty}\ln(\sqrt[2]{\alpha}\tau_{opt}(\alpha))-\alpha G(\tau_{opt}(\alpha)) \end{aligned} \tag{2.44}$$

Applying Lemma 10 we obtain

$$\lim_{\alpha\downarrow 0}\ln(\alpha\tau_{opt}(\alpha)) = -\ln\left(\ln(\| \mathbf{x} \|_\infty)-\overline{\ln(\mathbf{x})}\right)$$

and

$$\lim_{\alpha\downarrow 0}\alpha G(\tau_{opt}(\alpha) = \lim_{\alpha\downarrow 0}\alpha\tau_{opt}(\alpha)\frac{G(\tau_{opt}(\alpha))}{\tau_{opt}(\alpha)} = 1$$

This shows by relation (2.43) that

$$p(0^+) = -\ln\left(\ln(\| \mathbf{x} \|_\infty)-\overline{\ln(\mathbf{x})}\right)-1.$$

Again by Lemma 10 we know that

$$\lim_{\alpha\uparrow\infty}\ln(\sqrt[2]{\alpha}\tau_{opt}(\alpha)) = -\frac{1}{2}\ln\left(\overline{\ln^2(\mathbf{x})}-\overline{\ln(\mathbf{x})}^2\right)$$

and

$$\lim_{\alpha\uparrow\infty}\alpha G(\tau_{opt}(\alpha)) = \lim_{\alpha\uparrow\infty}\alpha\tau_{opt}^2(\alpha)\frac{G(\tau_{opt}(\alpha))}{\tau_{opt}^2(\alpha)} = \frac{1}{2}.$$

Hence by relation (2.43)

$$p(\infty) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2} - \frac{1}{2}\ln\left(\overline{\ln^2(\mathbf{x})} - \overline{\ln(\mathbf{x})}^2\right)$$

and we have verified the result. □

We may also use an alternative approach by first optimizing over α for τ fixed and then vary τ. This shows

$$\upsilon(P_4) = \sup_{\tau>0}\{\ln(\tau) + \sup_{\alpha>0}\{H_0(\alpha) - \alpha G(\tau)\}\}$$

It is easy to verify that for every $\tau > 0$ the unique optimal solution of $\sup_{\alpha>0}\{H_0(\alpha) - \alpha G(\tau)\}$ is given by

$$\alpha_{opt}(\tau) = h_0^{\leftarrow}(G(\tau)) \tag{2.45}$$

and this shows

$$\sup_{\alpha>0,\tau>0}\{H_0(\alpha) + \ln(\tau) - \alpha G(\tau))\} = \sup_{\tau>0}\{k(\tau)\}$$

with

$$k(\tau) = \ln(\tau) + H_0(h_0^{\leftarrow}(G(\tau))) - h_0^{\leftarrow}(G(\tau))G(\tau). \tag{2.46}$$

In the next lemma we will compute the value $k(0^+)$ and $k(\infty)$.

Lemma 2.2.10. *The objective function k in relation (2.46) satisfies*

$$k(0^+) = -\frac{1}{2} - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\overline{\ln^2(\mathbf{x})} - \overline{\ln(\mathbf{x})}^2\right) \tag{2.47}$$

and

$$k(\infty) = -1 - \ln\left(\ln(\| \mathbf{x} \|_\infty) - \overline{\ln(\mathbf{x})}\right). \tag{2.48}$$

Proof. We know by Lemma A.0.4 that $\lim_{\alpha\downarrow 0}\alpha h_0(\alpha) = 1$ and $\lim_{\alpha\uparrow\infty}\alpha h_0(\alpha) = \frac{1}{2}$. This shows substituting α by $h_0^{\leftarrow}(u)$ that

$$\lim_{u\uparrow\infty} h_0^{\leftarrow}(u)u = 1, \lim_{u\downarrow 0} h_0^{\leftarrow}(u)u = \frac{1}{2}. \tag{2.49}$$

By this observation we obtain using $G(\infty) = \infty$ and $G(0^+) = 0$ that

$$\lim_{\tau\uparrow\infty} h_0^{\leftarrow}(G(\tau))G(\tau) = 1, \lim_{\tau\downarrow 0} h_0^{\leftarrow}(G(\tau))G(\tau) = \frac{1}{2}. \tag{2.50}$$

By relation (2.46) to compute $k(0^+)$ and $k(\infty)$ we need to evaluate the behaviour of the function

$$\tau \to \ln(\tau) + H_0(h_0^{\leftarrow}(G(\tau)))$$

at zero and infinity. By Lemma A.0.6 it follows using $h_0^{\leftarrow}(G(\infty)) = 0$ that

$$\lim_{\tau\uparrow\infty} H_0(h_0^{\leftarrow}(G(\tau)) - \ln(h_0^{\leftarrow}(G(\tau))G(\tau)) + \ln(G(\tau)) = 0$$

and this implies by relation (2.50) that

$$\lim_{\tau\uparrow\infty} H_0(h_0^{\leftarrow}(G(\tau)) + \ln(G(\tau)) = 0$$

By this observation

$$\begin{aligned}
k(\infty) &= -1 + \lim_{\tau\uparrow\infty} \ln(\tau) + H_0(h_0^{\leftarrow}(G(\tau))) \\
&= -1 + \lim_{\tau\uparrow\infty} \ln(\tau) - \ln(G(\tau)) \\
&= -1 - \lim_{\tau\uparrow\infty} \ln\left(\tfrac{G(\tau)}{\tau}\right) \\
&= -1 - \ln\left(\ln(\| \mathbf{x} \|_\infty) - \overline{\ln(\mathbf{x})}\right)
\end{aligned}$$

and we have verified relation (2.48). Also by Lemma A.0.6

$$\lim_{\tau\downarrow 0} H_0(h_0^{\leftarrow}(G(\tau))) - \frac{1}{2}\ln(h_0^{\leftarrow}(G(\tau))G(\tau)) + \frac{1}{2}\ln(G(\tau)) = -\frac{1}{2}\ln(2\pi)$$

and this shows applying relation (2.50)

$$\lim_{\tau\downarrow 0} H_0(h_0^{\leftarrow}(G(\tau))) + \frac{1}{2}\ln(G(\tau)) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(2) = -\frac{1}{2}\ln(4\pi).$$

Hence we obtain using Lemma 2.2.1 and relation (2.46)

$$
\begin{aligned}
k(0^+) &= -\tfrac{1}{2} + \lim_{\tau\downarrow 0}\ln(\tau) + H_0(h_0^{\leftarrow}(G(\tau)) \\
&= -\tfrac{1}{2} - \tfrac{1}{2}\ln(4\pi) + \lim_{\tau\downarrow 0}\ln(\tau) - \tfrac{1}{2}\ln(G(\tau)) \\
&= -\tfrac{1}{2} - \tfrac{1}{2}\ln(4\pi) - \tfrac{1}{2}\ln\left(\tfrac{G(\tau)}{\tau^2}\right) \\
&= -\tfrac{1}{2} - \tfrac{1}{2}\ln(4\pi) - \tfrac{1}{2}\ln\left(\frac{\overline{\ln^2(\mathbf{x})} - \overline{\ln(\mathbf{x})}^2}{2}\right) \\
&= -\tfrac{1}{2} - \tfrac{1}{2}\ln(2\pi) - \tfrac{1}{2}\ln\left(\overline{\ln^2(\mathbf{x})} - \overline{\ln(\mathbf{x})}^2\right)
\end{aligned}
$$

This verifies relation (2.47) and concludes the proof. □

Observe by Lemma 2.2.9 and 2.2.10 that $k(0^+) = p(\infty)$. Also with probability one on the space of all samples $k(0^+) \neq k(\infty)$. Since both functions k and p do lack the concavity property as seen in Figure 2.3 (seems to be quasiconcave but this conjecture could not be verified) we compute first the function values of the selected function p or k on the finite lattice $\mathcal{D} = \{h, 2h, 3h, ..., Mh\}$ with mesh size h and $M < \infty$ to be determined by the algorithm. Thereafter we select that point on this lattice $\mathcal{D}$ having the largest objective value. Based on Lemma 2.2.8 up to 2.2.10 the following (approximation) algorithm can be executed. Due to the first step the generated outputs are approximations of the optimal solution. By selecting the mesh size smaller these approximations become more and more accurate. In the computational section h was given the value 0.05.

Algorithm 2.2.3. *Approximation algorithm maximum likelihood for generalized gamma population with all parameters unknown.*

5.1 If $p(0^+) > p(+\infty)$ go to step 2, otherwise go to step 3.

5.2 Select a step size h and compute on the grid mh, $m \in \mathbb{N}$

$$p(mh) = H_0(mh) + \sup_{\tau>0}\{\ln(\tau) - mhG(\tau)\}$$

for $m = 1, ..., M$ until $p(Mh)$ is close to the finite value $p(\infty)$. Choose the point $\widehat{m}h$ having the maximum value of the function p on mh, $m = 1, ..., M$ and select (see Lemma 2.2.6)

$$\tau_{opt}(\widehat{m}h) = \arg\max_{\tau>0}\{\ln(\tau) - \widehat{m}hG(\tau)\}$$

and compute

$$\widehat{\beta} = \frac{\| \mathbf{x} \|_{\tau_{opt}(\widehat{m}h)}}{(n\widehat{m}h)^{\frac{1}{\tau_{opt}(\widehat{m}h)}}}. \tag{2.51}$$

Go to step 4.

5.3 *Select a step size h and compute on the grid mh, $m \in \mathbb{N}$ the value*

$$k(mh) = \ln(mh) + \sup_{\alpha>0}\{H_0(\alpha) - \alpha G(mh)\}$$

for $m = 1,...,M$ until $k(Mh)$ is close to the finite value $k(\infty)$. Choose the point $\widehat{m}h$ having the maximum value of the function p on mh, $m = 1,...,M$ and select

$$\alpha_{opt}(\widehat{m}h) = \arg\max_{\alpha>0}\{H_0(\alpha) - \alpha G(\widehat{m}h)\}$$

and compute

$$\widehat{\beta} = \frac{\| \mathbf{x} \|_{\widehat{m}h}}{(n\alpha_{opt}(\widehat{m}h))^{\frac{1}{\widehat{m}h}}} \tag{2.52}$$

Go to step 5.

5.4 *Output $\widehat{\alpha} = \widehat{k}h, \widehat{\tau} = \tau_{opt}(\widehat{\alpha}), \widehat{\beta}$*

5.5 *Output $\widehat{\tau} = \widehat{k}h, \widehat{\alpha} = \alpha_{opt}(\widehat{\tau}), \widehat{\beta}$.*

In the next section we will perform some numerical experiments.

2.3 Comparison of the maximum likelihood and sample moment estimator for the mth moment of a (generalized) gamma distribution.

In this computational section we present the computational results of the maximum likelihood and sample moment estimator for the mth moment, $m = 2,4,6,8$ of a gamma or generalized gamma population with α known and unknown. Knowing the parameters for each of these distributions, we generate samples of size 10000, 1000, 100, and 10 and compute both the maximum likelihood estimator and sample moment estimator of these moments. To evaluate the maximum likelihood estimator of the mth moment of a gamma distribution we apply algorithm 2.2.1 for the computation of the maximum likelihood estimators of the unknown parameters α and β and use relation (2.3). A similar approach is used for the generalised gamma distribution applying Algorithm 2.2.2 or Algorithm 2.2.3 and relation (2.4) for a generalized gamma distribution with either α known or unknown. The true moments for the (generalized) gamma distribution are obtained from relation (2.2). To calculate the average estimated mean and the variance of these estimators 10000 simulation runs are executed and these values are shown in the tables under sample mean and sample variance. To compare their relative efficiency the estimated mean squared error (MSE) is also computed for both estimators and the ratio of these estimates is reported in the same table under estimated MSE ratio. Notice that the MSE-ratio is defined as the ratio of estimated MSE of the maximum likelihood and sample moment estimator. It is well known that the MSE of any estimator $\hat{\theta}$ of θ is given by (cf.[5]):

$$MSE = E_\theta(\hat{\theta} - \theta)^2 = Var_\theta(\hat{\theta}) + (Bias_\theta \hat{\theta})^2, \tag{2.53}$$

and the bias of an estimator equals

$$Bias_\theta \hat{\theta} = |E_\theta(\hat{\theta}) - \theta|. \tag{2.54}$$

Moreover, the code is written in Python 3 [24] and executed on a laptop having a 8.00 GB RAM, Intel(R) Core(TM) i5-8250U CPU and a 64-bit operating system.

2.3.1 Computational results for gamma distribution

In this subsection paper, we report the mth moment estimation with $m \in \{2,4,6,8\}$ for a gamma population having $\beta = 1$ and $\alpha \in \{0.02, 0.1, 1, 10\}$. Since β is a scale parameter we set β in all four different scenarios equal to 1. Due to limited space we only report the results in Table 2.1 for $\alpha = 0.1, \beta = 1$. For these parameters the true 8th, 6th, 4th, and 2nd moments are given by 648.50882, 14.97365, 0.71609, and 0.11000, respectively. As observed from Table 2.1 the MSE ratio of both estimators shows that the maximum likelihood estimator is more efficient than the sample moment estimator when the sample size exceeds 100. For small sample sizes, the sample estimator should be preferred over the maximum likelihood estimator due to its unbiasedness. The bias of the maximum likelihood estimator for a given sample size increases as m increases. However, for larger sample sizes, the bias of the maximum likelihood estimator decreases (see Figure 2.1) and this improves its performance compared to the sample moment estimator. This result is not surprising since the maximum likelihood estimator is asymptotically a consistent minimum variance estimator (cf.[18]). Although the mth sample moment estimator is always unbiased, we also observe from Figure 2.1 as the value of m increases that the histogram of this sample moment estimator becomes left skewed and its variance with increasing m increases.

		Maximum likelihood estimator		Sample moment estimator		MSE ratio	bias_{MLE}
	n	sample mean	sample var	sample mean	sample var		
m=8	10000	667.9724	31898.993	631.8879	8238460.7361	0.0039	19.46359
	1000	878.4645	682882.2006	867.3921	269831587.51	0.0027	229.9556
	100	7418.0093	3117399725	491.4093	129160373.8	24.486	6769.50
	10	32155360	8.226647e+17	427.4697	616300271.8	1336415772	32154712.39
m=6	10000	15.2084	9.1422	14.8258	441.2843	0.0208	0.23475
	1000	17.6441	137.1051	16.1922	10940.0076	0.0132	2.67044
	100	58.1131	39374.1718	13.1314	19424.8563	2.1224	43.13949
	10	12452.2309	49969710691	11.3953	141608.3952	353932.8	12437.25
m=4	10000	0.7204	0.0089	0.7143	0.0619	0.1447	0.00430
	1000	0.7655	0.1044	0.7217	0.8605	0.1242	0.04941
	100	1.2579	3.9395	0.6863	4.8963	0.8644	0.54178
	10	16.9307	18107.6158	0.6325	42.5427	431.7428	16.21456
m=2	10000	0.1101	0.0001	0.11	0.0001	0.7236	9.93212e-05
	1000	0.1114	0.0005	0.1103	0.0007	0.7251	0.00135
	100	0.1212	0.0063	0.11	0.0067	0.9549	0.01120
	10	0.2069	0.2149	0.1096	0.0621	3.6121	0.09688

Table 2.1 Estimated mth moment of the gamma distribution with parameters $\alpha = 0.1$ and $\beta = 1$

We also observe for $\alpha = 0.02$ in our more extensive experiments not reported in this paper that increasing the value of m will deteriorate the efficiency of the maximum likelihood estimator over the sample moment estimator if its efficiency is above 1 for $m = 2$. However, its efficiency will improve for m increasing if its efficiency is already below 1 for the same value of m. The first event almost always occurs for very small sample sizes ($n = 10$) and is less likely to happen for intermediate sample sizes between 10 and 100. Also this behaviour is not shown as prominent for α much larger. Increasing the sample size leads in general to a more accurate estimation by the maximum likelihood estimator than the sample moment estimator. This means for small sample sizes, our computational results suggest that the sample moment estimator is more accurate than the maximum likelihood estimator especially for α close to zero. The reason for this behavior is that the maximum likelihood estimates of the parameters α and β are not accurate for small sample sizes (although the algorithm gives the exact minimum of the log likelihood function). Due to these errors high errors occur in the used formula for the mth moment. For $\alpha > 1$ and small m, again this behaviour is less prominent. Sometimes even for small sample sizes the maximum likelihood estimator is at least as accurate as the sample moment estimator. This suggests that one should apply the following rule of thumb. The maximum likelihood estimator is always to be preferred above the sample moment estimator unless α is small and the sample size is not large enough ($n = 10$). If the sample size is much larger in all cases the maximum likelihood estimator due to a better estimate of the parameters should be used. This behaviour is not surprising due to the asymptotic theoretical properties of the MLE approach.

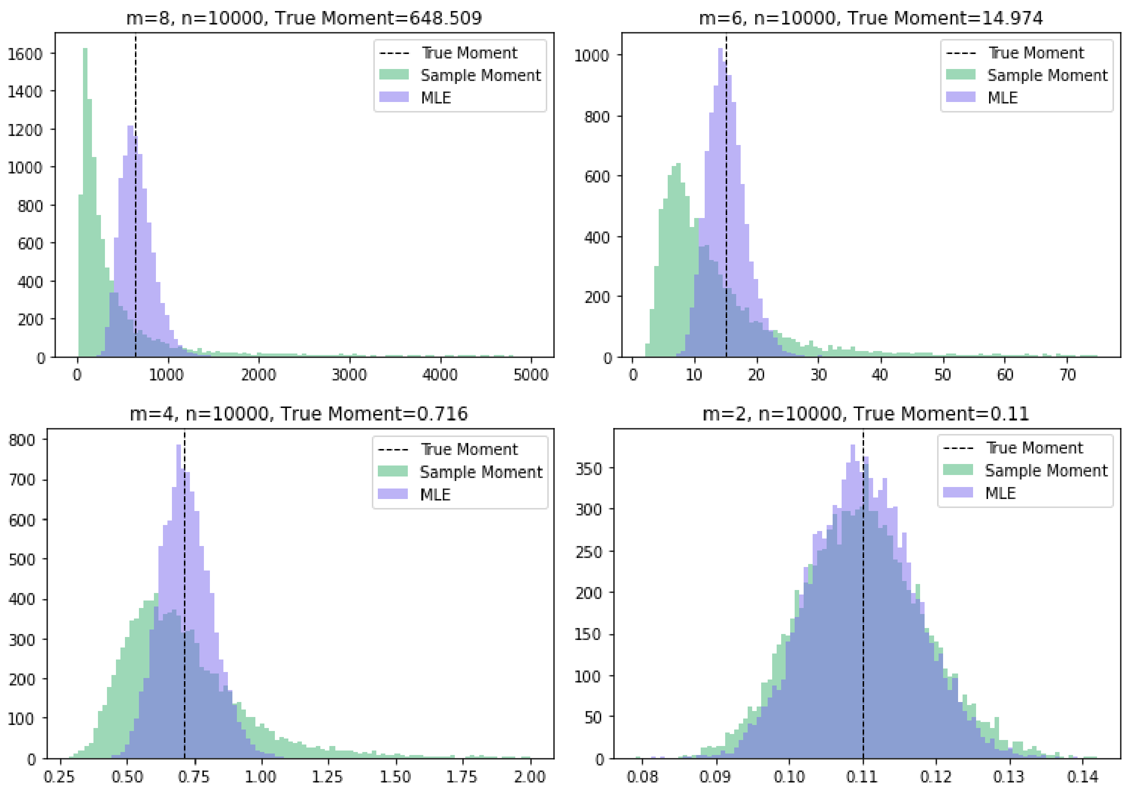

Figure 2.1 Histogram of the maximum likelihood and mth sample moment estimator of the gamma distribution with parameters $\alpha = 0.1$ and $\beta = 1$.

2.3.2 Computational results for generalized gamma distribution with known parameter α.

In this subsection we report the simulation results for a (generalized) gamma population with α known and β and τ to be estimated. We list in Table 2.2 the different executed scenarios. Again due to the limited space we only list extensively for $\alpha = 1$, $\beta = 1$, and $\tau = 0.5$ the results in Table 2.3 and Figure 2.2. For this scenario the true

8th, 6th, 4th and 2nd moment equal 20922789888000.0, 479001600.0, 40320.0, and 24.0, respectively. Also the average time for calculating the parameters is equal to 1.6685 seconds when the sample size is equal to 10000.

$(\alpha,\ \tau)$	(0.1, 10)	(1, 0.5)	(1, 1)	(1, 10)	(10, 0.5)	(10, 10)

Table 2.2 Different setting of the parameters α and τ for generalized gamma distribution with known parameter α used in experiments when β equals to 1

As for the gamma population discussed extensively in the previous subsection we observe for all scenarios a similar behavior.

		maximum likelihood estimator		Sample moment estimator		MSE ratio	bias_{MLE}
	n	sample mean	sample var	sample mean	sample var		
m=8	10000	22030491658671	6.0090e+25	20106226954895	2.5978e+29	0.0002	1107701770671
	1000	36061034635909	2.4887e+27	77517931006121	2.3795e+31	0.0001	15138244747909
	100	3024289070593550	3.0861e+33	438545230024213	1.8913e+33	1.6363	3003366280704969
	10	1.1362e+29	9.2181e+610	3247979131631	5.6891e+28	1.6116e+33	1.1362e+29
m=6	10000	490011352.0	1.3386e+16	490016661.4	1.6554e+19	0.0008	11009752.09
	1000	614072657	2.5940e+17	912514522	1.2579e+21	0.0002	135071057.34
	100	4980302454	1.2167e+21	3263866801	9.0706e+22	0.0136	4501300854.32
	10	1.2454e+18	9.6436e+39	237098447	1.2156e+20	7.9305e+19	1.2454e+18
m=4	10000	40589.0787	30224526.4241	40324.8978	2017630473.8430	0.015	269.0787
	1000	43602.2389	375064734.7530	43542.1869	77580539806.7348	0.005	3282.2389
	100	83618.1503	29949625135.4838	55619.1795	4386939099014.7607	0.0073	43298.15
	10	324454530	2.1430e+20	33742.8817	323869172827	661930276	324414210
m=2	10000	24.0124	1.5969	23.9761	4.0955	0.3899	0.0124
	1000	24.2101	16.4689	24.0103	43.4332	0.3802	0.2101
	100	25.9637	205.6182	23.6245	551.2352	0.3799	1.9637
	10	63.6485	81460.3836	22.9583	3278.3600	25.319	39.6485

Table 2.3 Estimated mth moment of a Weibull or generalized gamma distribution (with known $\alpha = 1$) and parameters $\beta = 1$, $\tau = 0.5$.

Due to the same theoretical considerations as for the gamma distribution we may conclude as a rule of thumb it is preferable to use the maximum likelihood estimator instead of the sample estimator in case the sample size is reasonable large and the known α is not very close to zero. In case α is close to zero then we should only use the maximum likelihood estimator for a reasonable large sample size.

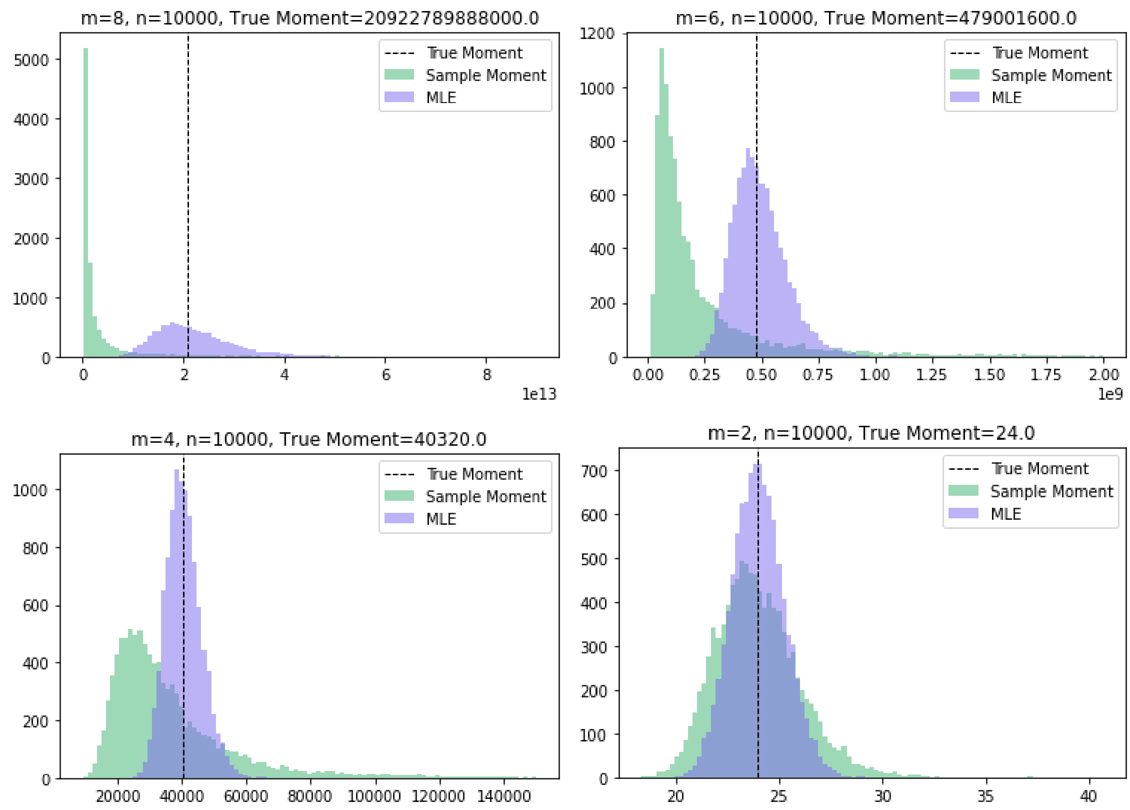

Figure 2.2 Histogram of the maximum likelihood and mth sample moment estimator of a Weibull or generalized gamma (with known $\alpha = 1$) distribution with parameters $\tau = 0.5$ and $\beta = 1$.

2.3.3 Computational resuls for generalized gamma distribution with all parameters unknown.

In this subsection we report the simulation results for a (generalized) gamma population with all parameters unknown. We list in Table 2.4 the different considered scenarios.

$(\alpha,\ \tau)$	(0.1, 5)	(0.5, 0.5)	(1, 1)	(1, 10)	(10, 1)	(10, 10)

Table 2.4 Different setting of the parameters, α and τ, for generalized gamma distribution with all parameters unknown used in experiments when β equals to 1

Again due to limited space we only discuss extensively for $\alpha = 1$, $\beta = 1$, and $\tau = 1$ the results in Table 2.5 and Figure 2.4. For this scenario the true 8th, 6th, 4th and 2nd moment equal 40320.0, 720.0, 24.0, and 2.0, respectively. Also the average time for calculating the parameters is equal to 2.3261 seconds for a sample size of 10000 and the mesh size of the lattice in Algorithm 2.2.3 is $h = 0.05$.

		MLE		sample mean estimator		MSE ratio	bias_{MLE}
	n	sample mean	sample var	sample mean	sample var		
m=8	10000	41253.4287	89647435.8654	40324.8978	2017630473.8430	0.0449	933.4287
	1000	51019.207	1743247911.32	43542.1869	77580539806	0.0239	10699.20
	100	253609419.5095	3.3624e+20	55653.9109	4390449749970	76596405	253569099
m=6	10000	725.58185	10048.1709	718.5998	49798.4325	0.2024	5.5818
	1000	783.4919	121615.6538	727.9244	917862.1643	0.1369	63.4919
	100	5800.6937	34811103063	733.1264	32757859	1063.46	5080.69
m=4	10000	24.0328	2.6556	23.9761	4.0955	0.6486	0.0328
	1000	24.4489	26.5143	24.0103	43.4332	0.6151	0.4489
	100	29.8940	1356.9375	23.6294	551.5836	2.5224	5.8940
m=2	10000	1.9999	0.0020	1.9997	0.002	1.006	4.6025e-05
	1000	2.0012	0.0199	1.9994	0.0200	0.9944	0.0012
	100	2.0140	0.2064	1.9916	0.1955	1.056	0.0140

Table 2.5 Estimated mth moment of a generalised gamma distribution with parameters $\alpha = 1$, $\beta = 1$, and $\tau = 1$.

Our algorithm 2.2.3 terminated in all instances and returned a solution of the two dimensional optimization problem (P_4). To compare the performance of our algorithm we also used a general optimization package. The often used Python package scipy.com implementing the Cobyla method (cf.[25]) in some scenario did not terminate and if this package returned a solution the solution proposed by Algorithm 2.2.3 was much more accurate. However, if the sample size equals 10, Algorithm 2.2.3 returned solutions far away from the real parameter values and so big differences occur between the true mth moment and the estimated mth moment. This is most likely due to the poor quality of the maximum likelihood estimator for small samples and so we decided to report only in Table 2.5 the results for sample sizes of at least 100. For the chosen m values even for sample of size 100 the sample moment estimator is to be preferred above the maximum likelihood estimator of the mth moment. As for the previous considered subcases we also observe that for larger sample sizes much more accurate estimates of the unknown parameters occur and hence we obtain much more accurate estimates of the mth moment. As can be seen from Figure 2.3 the functions p and k are not concave. However, since the asymptotic behaviour of the function p and k are known one can still find an optimal solution and not a local optimal solution.

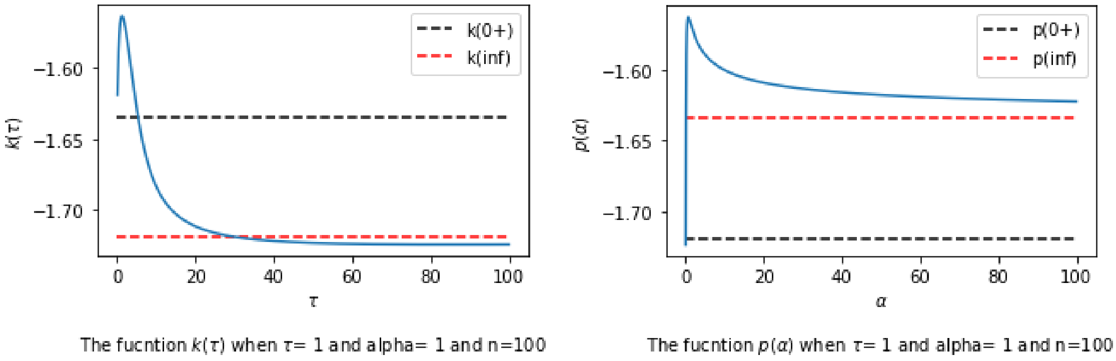

Figure 2.3 Behaviour of function p and k for $\tau = 1$, $\alpha = 1$, and $\beta = 1$ and sample size is 100.

Increasing the parameter m and the sample size we observe that the maximum likelihood estimator performs better than the sample moment estimator (see Table 2.5). Also, similar like in the previous discussed cases, the histogram of the sample moment estimator tends to be left skewed and this effect becomes bigger as m increases (see Figure 2.4). On the other hand, for a small sample size and larger values of m the sample moment estimator has a better performance rather than the maximum likelihood estimator (see Table 2.5).

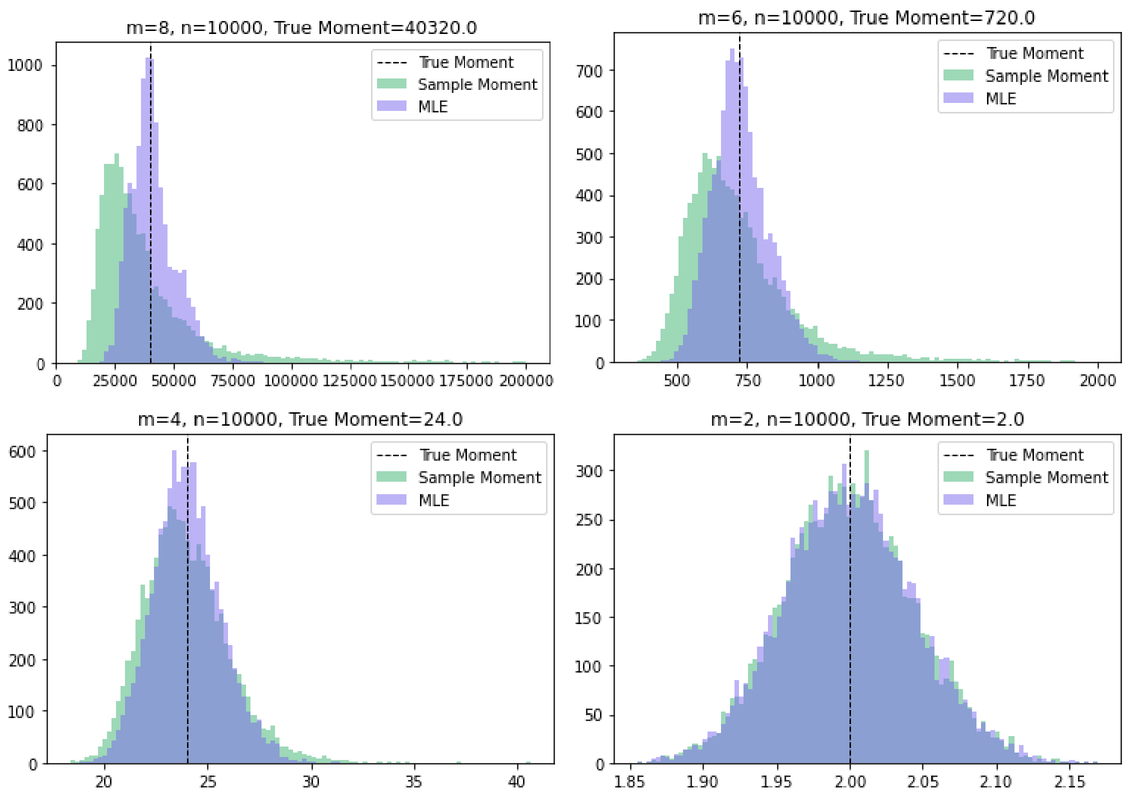

Figure 2.4 Histogram of the maximum likelihood and mth sample moment estimator of a generalised gamma distribution with parameters $\tau = 1$, $\alpha = 1$, and $\beta = 1$.

Due to similar theoretical considerations as for the previous subcases we may conclude as a rule of thumb it is preferable to use the maximum likelihood estimator instead of the sample moment estimator if the sample size is reasonable large and both α and τ are not very close to zero or τ is close to zero and α is not large at the same time.

2.4 Conclusion

In this paper we discuss in detail the properties of the maximum likelihood optimization problem for a (generalized) gamma population. We started analyzing this optimization problem for a gamma population. Based on some global properties of the digamma function we reduced it to a one dimensional strictly concave maximization problem and proposed a special purpose algorithm. After that we considered this problem for a generalized gamma population with parameter α known. Based again on global properties of the digamma function this optimization problem is reduced to a one dimensional strictly concave maximization problem and we proposed an easy special purpose algorithm. A special case of this problem is given by the maximum likelihood estimation problem for a Weibull population. Finally the maximum likelihood optimization problem for a generalized gamma population with all three parameters unknown is analyzed and we propose a special purpose algorithm to solve this case. Unfortunately this optimization problem is not a concave maximization problem. To test the performance of this algorithm we compared the behaviour of a standard maximum likelihood estimator of the mth moment in a (generalized) gamma population against the behaviour of the mth sample moment estimator. From our simulation experiments we identified under which conditions the maximum likelihood estimator is to be preferred above the mth sample moment estimator. As to be expected for a sample size relatively large the maximum likelihood estimator has a lower mean square error then the mth sample moment estimator. Only for relatively small sample sizes the mth sample moment estimator should be used. We also plotted the histogram of these different estimators for different scenarios. We may also conclude from our simulation results that for small sample sizes the maximum likelihood approach is not that accurate in estimating the unknown parameters of the distribution. Since the mth sample moment estimator is to be preferred for small sample sizes to estimate the mth moment this suggest for small sample sizes it might be a good strategy to use the method of moment approach of Pearson in stead of the maximum likelihood approach to estimate the unknown parameters of a (generalized) gamma population.

3. On the method of moments approach for a generalized gamma population

3.1 Introduction.

One of the oldest point estimating methods dating back to Pearson (cf.[26]) is to estimate unknown parameters of a cumulative distribution function by the so-called *method of moments* approach replacing the unknown moments by sample moments arising from a population sample. The original method of moments approach in probability theory used to determine the underlying cdf (in this case the normal distribution) is to identify all integer moments of a random variable and use this information to determine the cdf of this random variable. This approach was used in the proof of the central limit theorem by Pafnuty Chebyshev in 1887 (cf.[27]). Before discussing the related method of moments approach in parametric statistics we introduce the following notation. For any random variable $\mathbf{X}_1$ on some given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we denote by $\mu'_m(\mathbf{X}_1)$, $m \in \mathbb{N}$ the mth moment of the random variable $\mathbf{X}_1$ given by

$$\mu'_m(\mathbf{X}_1) = \mathbb{E}(\mathbf{X}_1^m), \tag{3.1}$$

and by $\mu_m(\mathbf{X}_1)$ the mth central moment given by

$$\mu_m(\mathbf{X}_1) := \mu'_m(\mathbf{X}_1 - \mathbb{E}(\mathbf{X}_1)) \tag{3.2}$$

If $m = 2$ the 2nd central moment (mostly denoted by $\sigma^2(\mathbf{X}_1)$) is also called the *variance* of the random variable $\mathbf{X}_1$. Its positive root $\sigma(\mathbf{X}_1) := \sqrt[2]{\sigma^2(\mathbf{X}_1)}$ is known as the *standard deviation* of $\mathbf{X}_1$. In parametric statistics the main problem is to identify using a so-called realisation of a random sample $\mathbf{X} = (\mathbf{X}_1, ..., \mathbf{X}_n)$, consisting of independent and identically distributed random variables, from which cumulative distribution function F_{θ_0} belonging to a set of cumulative distributions functions $(F_\theta)_{\theta\in\Theta}$ with $\theta = (\theta_1, ..., \theta_m) \in \Theta$, $m \in \mathbb{N}$ and Θ a given set (the so-called parameter space) this sample is generated. The kth moment $\mu'_k(\mathbf{X}_1), k \in \mathbb{N}, k \leq m$ is clearly a function of this unknown cdf $F_{\theta_0}, \theta_0 = (\theta_{01}, ..., \theta_{0m}) \in \Theta$ and so we know

$$\mu'_k(\mathbf{X}_1) = f_k(\theta_{01}, ..., \theta_{0m}), k = 1, ..., m. \tag{3.3}$$

Since in statistics the value $\mu'_k(\mathbf{X}_1), k = 1, ..., m$ is not known (the problem is to select a candidate cdf from this parametric family) the information contained in the sample is used to identify the unknown cdf F_{θ_0}. As proposed by Pearson we replace the true unknown kth moments by their sample kth moment estimators. Remember the sample kth moment estimator $\widehat{\mu}'_k(\mathbf{X}), k \in \mathbb{N}$ is given by

$$\widehat{\mu}'_k(\mathbf{X}) := \frac{1}{n}\sum\nolimits_{i=1}^{n} \mathbf{X}_i^k$$

Related is the sample kth central moment estimator $\widehat{\mu}_k(\mathbf{X}), k \in \mathbb{N}$ given by

$$\widehat{\mu}_k(\mathbf{X}) = \frac{1}{n}\sum\nolimits_{i=1}^{n} (\mathbf{X}_i - \widehat{\mu}'_1(\mathbf{X}))^k$$

A realisation of the sample moment estimator is denoted by $\widehat{\mu}'_k(X)$, respectively $\widehat{\mu}_k(X)$ with $X = (x_1, ..., x_n)$ a realized sample. Justified by the asymptotic consistency of these sample moment estimators one should (if possible) identify now a hopefully unique solution of the system of nonlinear equations

$$\widehat{\mu}'_k(X) = f_k(\theta_{01}, ..., \theta_{0m}), k = 1, ..., m, \theta_0 = (\theta_{01}, ..., \theta_{0m}) \in \Theta \tag{3.4}$$

and use this (unique) solution as an estimate of the unknown vector θ_0. In general it can happen that this system is inconsistent or has multiple solutions. In the next section we will apply this procedure to the parametric family of generalized gamma distributions and give a detailed analysis of how to solve the above system. Contrary to other proposals of method of moment estimators ([17] for a generalized gamma population we will propose a method of moment estimator of the generalized gamma distribution for which we can easily identify for which values of the estimator the system in relation (3.4) has a solution and in this holds the solution is unique. This estimator is not as accurate as the maximum likelihood estimator and might serve as a quick approximation of the unknown parameters. Analysing the corresponding system of equations we also propose for this estimator an approximate analytical solution of the system.

3.2 The method of moments approach for a (generalized) gamma population.

In this section we start with the following general observation about the method of moments approach. The next result shows that one can replace the original set of nonlinear equations by a sometimes simpler set of of nonlinear equations. To this simpler set of equations one can then apply an easier algorithm to find its solution (if it exists).

Lemma 3.2.1. *For any sample $X=(x_1,...,x_n)$ the set of solutions of the system*

$$\widehat{\mu}'_k(X)=\mu'_k(\mathbf{X}_1), k=1,..,m$$

equals the set of solutions of the system

$$\widehat{\mu}'_1(X)=\mu'_1(\mathbf{X}_1), \widehat{\mu}_k(X)=\mu_k(\mathbf{X}_1), k=2,...,m$$

Proof. In the method of moments approach one need to find for a given sample $X=(x_1,...,x_n)$ a solution of the following system of equations (nonlinear in its parameters)

$$\widehat{\mu}'_k(X)=\mu'_k(\mathbf{X}_1)=f_k(\theta_{01},...,\theta_{0m}), k=1,...,m \tag{3.5}$$

Clearly for for any solution of this system it follows for any $j\leq m$ that

$$\begin{aligned}
\widehat{\mu}_j(X) &= \tfrac{1}{n}\textstyle\sum_{i=1}^n (x_i-\widehat{\mu}'_1(X))^j \\
&= \tfrac{1}{n}\textstyle\sum_{i=1}^n\sum_{k=0}^j \binom{j}{k}(-1)^{j-k}x_i^j\widehat{\mu}'_1(X)^{j-k} \\
&= \textstyle\sum_{k=0}^j \binom{j}{k}(-1)^{j-k}\widehat{\mu}'_j(X)\widehat{\mu}'_1(X)^{j-k} \\
&= \textstyle\sum_{k=0}^j \binom{j}{k}(-1)^{j-k}\mu'_j(\mathbf{X}_1)\mu'_1(\mathbf{X}_1)^{j-k} \qquad (3.6)\\
&= \mathbb{E}\left(\textstyle\sum_{k=0}^j \binom{j}{k}(-1)^{j-k}\mathbf{X}_1^j\mu'_1(\mathbf{X})^{j-k}\right) \\
&= \mathbb{E}((\mathbf{X}_1-\mu'_1(\mathbf{X}))^j) \\
&= \mu_j(\mathbf{X}_1)
\end{aligned}$$

Since central moments determine uniquely moments and vice versa this shows for $j=2$ that the set of solutions of the system $\widehat{\mu}'_k(X)=\mu'_k(\mathbf{X}_1), k=1,...,m$ equals the set of solutions of the system

$$\widehat{\mu}_2(X) = \mu_2(\mathbf{X}_1), \widehat{\mu}'_k(X) = \mu'_k(\mathbf{X}_1), k = 1, ..., m, k \neq 2$$

By iterating over the index set j we obtain the desired result. □

If a sample is generated from a population having a cdf $F_\theta, \theta = (a, b) \in \mathbb{R} \times \mathbb{R}_+$ given by

$$F_\theta(x) = F\left(\frac{x-a}{b}\right)$$

and the cumulative distribution F is known it is easy to see that this is the same as

$$\mathbf{X}_1 \stackrel{d}{=} a + b\mathbf{Z}_1 \tag{3.7}$$

with $a \in \mathbb{R}$ and $b > 0$ and the cdf of the random variable $\mathbf{Z}_1$ is known. By the definition of the mth (central) moment it is obvious that

$$\mu'_m(b\mathbf{Z}_1) = b^m \mu'_m(\mathbf{Z}_1) \tag{3.8}$$

for any random variable $\mathbf{Z}_1$ and $b > 0$ and

$$\mu_m(a + b\mathbf{Z}_1) = b^m \mu_m(\mathbf{Z}_1) \tag{3.9}$$

for any $a \in \mathbb{R}$ and $b > 0$. Applying now the method of moments approach to such a class of parametric cdfs it is easy to verify the following result.

Lemma 3.2.2. *If $\mathbf{X}_1 \stackrel{d}{=} a + b\mathbf{Z}_1$ with unknown location parameter $a \in \mathbb{R}$ and unknown scale parameter $b > 0$ and $\sigma^2(\mathbf{Z}_1) > 0$ then the methods of moment estimators $\widehat{b}_{MM}(\mathbf{X}), \widehat{a}_{MM}(\mathbf{X})$ of b,respectively a are given by*

$$\widehat{b}_{MM}(\mathbf{X}) = \sqrt[2]{\widehat{\mu}_2(\mathbf{X})}\sigma(\mathbf{Z}_1)^{-1}, \widehat{a}_{MM}(\mathbf{X}) = \widehat{\mu}'_1(\mathbf{X}) - \sqrt[2]{\widehat{\mu}_2(\mathbf{X})}\frac{\mathbb{E}(\mathbf{Z}_1)}{\sigma(\mathbf{Z}_1)} \tag{3.10}$$

Proof. By Lemma 3.2.1 and relations (3.8) and (3.9) it is equivalent to solve the system

$$\widehat{\mu}'_1(\mathbf{X}) = a + b\mathbb{E}(\mathbf{Z}_1), \widehat{\mu}_2(\mathbf{X}) = b^2\sigma^2(\mathbf{Z}_1)$$

and this shows the desired result. □

It follows immediately from Lemma 3.2.2 that $\widehat{b}^2_{MM}(\mathbf{X})$ is an unbiased estimator of b^2 and in case $\mathbb{E}(\mathbf{Z}_1) = 0$ that $\widehat{a}_{MM}(\mathbf{X})$ is an unbiased estimator of a. Another consequence of Lemma 3.2.2 is given by the following result. Observe we denote by $\ln(\mathbf{X})$ the random vector $(\ln(\mathbf{X}_1),...,\ln(\mathbf{X}_n))$.

Lemma 3.2.3. *If* $\mathbf{X}_1 \stackrel{d}{=} a\mathbf{Z}_1^b$ *with* $\mathbf{Z}_1$ *a nonnegative random variable having a known cdf satisfying* $\sigma^2(\ln(\mathbf{Z}_1)) > 0$ *and* $a, b > 0$ *unknown parameters then the method of moments estimators* $\widehat{b}_{MM}(\mathbf{X})$,$\widehat{a}_{MM}(\mathbf{X})$ *of* b, *respectively* a *are given by*

$$(3.11)\quad \widehat{b}_{MM}(\mathbf{X}) = \sqrt[2]{\widehat{\mu}_2(\ln(\mathbf{X}))}\sigma(\ln(\mathbf{Z}_1))^{-1}, \widehat{a}_{MM}(\mathbf{X}) = e^{\widehat{\mu}_1(\ln(\mathbf{X})) - \sqrt[2]{\widehat{\mu}_2(\ln(\mathbf{X}))}\frac{\mathbb{E}(\ln(\mathbf{Z}_1))}{\sigma(\ln(\mathbf{Z}_1))}}$$

Proof. Clearly $\ln(\mathbf{X}_1) \stackrel{d}{=} \ln(a) + b\ln(\mathbf{Z}_1)$ and this shows applying Lemma 3.2.2 the desired result. □

The methods of moment approach can result in different estimators. For $\mathbf{X}_1 \stackrel{d}{=} a\mathbf{Z}_1^b, a, b > 0$ one could also try to solve the set of nonlinear equations

$$(3.12)\qquad \widehat{\mu}_1'(X) = a\mathbb{E}(\mathbf{Z}_1^b), \widehat{\mu}_2(X) = a^2\sigma^2(\mathbf{Z}_1^b)$$

for $X = (x_1, ..., x_n)$ a realized sample of the random vector $\mathbf{X} = (\mathbf{X}_1, ..., \mathbf{X}_n)$. This leads to a different set of estimators as the one proposed in Lemma 3.2.3. One can use the method of moments estimators listed in relation (3.11) if it is easy to compute either by simulation or analysis the expectations $\mathbb{E}(\ln(\mathbf{Z}_1))$ and $\mathbb{E}(\ln(\mathbf{Z}_1)^2)$. To choose the moment estimators in relation (3.12) one needs to be able to calculate $\mathbb{E}(\mathbf{Z}_1^\alpha)$ as a function of α.In general it might also be more difficult algorithmically to find a solution (if it exists) of the system in relation (3.12).

We will now focus on the method of moments approach applied to a (generalized) gamma population. The next definition is well known (cf.[8]). Observe the class of gamma distributions was extended to the class of generalized gamma distributions to gain more flexibility in fitting distributions occurring in reliability theory.

Definition 3.2.1. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a given probability space.*

6.1 *The nonnegative random variable* $\mathbf{Y}_1$ *has a gamma distribution with parameter* $\alpha > 0$ *and scale parameter* $\beta = 1$ *if its density is given by*

$$f(y) = \frac{e^{-y}y^{\alpha-1}}{\Gamma(\alpha)}1_{(0,\infty)}(y)$$

with $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-xdx}$ *the well known gamma function. This is denoted by* $\mathbf{Y} \sim G(\alpha, 1)$.

6.2 *The random variable* $\mathbf{X}_1$ *has a gamma distribution with parameters* $\alpha > 0$ *and scale parameter* $\beta > 0$ *if* $\mathbf{X}_1 \stackrel{d}{=} \beta\mathbf{Y}_1$ *with* $\mathbf{Y}_1 \sim G(\alpha, 1)$.

6.3 The random variable $\mathbf{X}_1$ *has a generalised gamma distribution with scale parameter* $\beta > 0$ *and positive parameters* α *and* τ *if*

$$\mathbf{X}_1 \stackrel{d}{=} \beta \mathbf{Y}_1^{\tau^{-1}}$$

with $\mathbf{Y}_1 \sim G(\alpha, 1)$. *This is denoted by* $\mathbf{X}_1 \sim G(\alpha, \beta, \tau)$.

6.4 The random variable $\mathbf{X}_1$ *has a Weibull distribution with scale parameters* β *and positive* $\tau > 0$ *if* $\mathbf{X}_1 \sim G(1, \beta, \tau)$.

If $\mathbf{Y}_1$ has a gamma distribution with parameters $\alpha > 0$ and $\beta = 1$ it is easy to check that for every $m \in \mathbb{N}$

$$\mu'_m(\mathbf{Y}_1) = \frac{\Gamma(\alpha + m)}{\Gamma(\alpha)} = \Pi_{j=1}^m (\alpha + m - j) \tag{3.13}$$

To give an interpretation of the unknown parameters $\alpha, \tau > 0$ occurring within the class of (generalized) gamma distributions we list the following well known definition (cf. [4]).

Definition 3.2.2. *Let* $\mathbf{X}_1$ *be a nonnegative random variable on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$.

7.1 If $\mathbb{E}(\mathbf{X}_1) > 0$ *the squared coefficient* $c^2(\mathbf{X}_1)$ *of variation is defined by*

$$c^2(\mathbf{X}_1) = \mathbb{E}\left(\left(\frac{\mathbf{X}_1 - \mathbb{E}(\mathbf{X}_1)}{\mathbb{E}(\mathbf{X}_1)}\right)^2\right) = \frac{\sigma^2(\mathbf{X}_1)}{(\mu'_1(\mathbf{X}_1))^2} > 0.$$

7.2 If $\sigma^2(\mathbf{X}_1) > 0$ *the Pearson coefficient skewness*$(\mathbf{X}_1)$ *of skewness is defined by*

$$\mathit{skewness}(\mathbf{X}_1) = \mathbb{E}\left(\left(\frac{\mathbf{X}_1 - \mathbb{E}(\mathbf{X}_1)}{\sigma(\mathbf{X}_1)}\right)^3\right) = \frac{\mu_3(\mathbf{X}_1)}{\sigma^3(\mathbf{X}_1)}.$$

Clearly by the definition of squared coefficient of variation it is easy to check for any $b > 0$ that

$$c^2(b\mathbf{Z}_1) = c^2(\mathbf{Z}_1) \tag{3.14}$$

for any nonnegative random variable $\mathbf{Z}_1$ satisfying $\mathbb{E}(\mathbf{Z}_1) > 0$. This means that the squared coefficient of variation is invariant under a linear mapping. Also it easily follows for any random variable $\mathbf{Z}_1$ having a positive variance that for any $a \in \mathbb{R}$

and $b > 0$

$$\text{Skewness}(a + b\mathbf{Z}_1) = \text{Skewness}(\mathbf{Z}_1) \tag{3.15}$$

and so the skewness measure is invariant under affine mappings. Also it is well known (see page 112 of [28]) that for any random variable $\mathbf{Z}_1$ it follows that

$$\mu_1'(\mathbf{Z}_1) = \kappa_1(\mathbf{Z}_1), \mu_m(\mathbf{Z}_1) = \kappa_m(\mathbf{Z}_1) \tag{3.16}$$

for $m = 2,3$ with $\kappa_m(\mathbf{Z}_1)$ the mth cumulant of the random variable $\mathbf{Z}_1$ given by $\kappa_m(\mathbf{Z}_1) = K^{(m)}(0)$ with $K(s) := \ln(\mathbb{E}(e^{s\mathbf{Z}_1}))$ the so-called *cumulant generating function* of the random variable $\mathbf{Z}_1$. This shows that an alternative representation of skewness and squared coefficient of variation is given by

$$c^2(\mathbf{Z}_1) = \frac{\kappa_2(\mathbf{Z}_1)}{\kappa_1(\mathbf{Z}_1)^2}, \text{skewness}(\mathbf{Z}_1) = \frac{\kappa_3(\mathbf{Z}_1)}{\kappa_2(\mathbf{Z}_1)^{\frac{3}{2}}} \tag{3.17}$$

It is easy to verify using $\kappa_1(\mathbf{Z}_1) = \mathbb{E}(\mathbf{Z}_1)$ that for $\mathbf{Y}_1 \sim G(\alpha, 1)$ and hence

$$K(s) = \mathbb{E}(e^{s\ln(\mathbf{Y}_1)}) = \mathbb{E}(\mathbf{Y}_1^s) = \frac{\Gamma(\alpha + s)}{\Gamma(\alpha)} \tag{3.18}$$

that

$$\mathbb{E}(\ln(\mathbf{Y}_1)) = K^{(1)}(0) = \psi_0(\alpha) \tag{3.19}$$

with ψ_0 the so-called polygamma function of order zero or bigamma function introduced in the Appendix A. The next result gives an interpretation of the parameters $\alpha, \tau > 0$ for $\mathbf{X}_1$ having a (generalized) gamma distribution.

Lemma 3.2.4. *Let $\mathbf{X}_1$ be a nonnegative random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$*

8.1 If $\mathbf{X}_1 \sim G(\alpha, \beta)$ then $c^2(\mathbf{X}_1) = \alpha^{-1}$

8.2 If $\mathbf{X}_1 \sim G(\alpha, \beta, \tau)$ then skewness$(\ln(\mathbf{X}_1)) = \psi_2(\alpha)\psi_1(\alpha)^{-\frac{3}{2}}$.

Proof. Applying relation (3.14) and the definition of a gamma distributed random variable we obtain

$$c^2(\mathbf{X}_1) = c^2(\beta\mathbf{Y}_1) = c^2(\mathbf{Y}_1)$$

with $\mathbf{Y}_1 \sim G(\alpha, 1)$. Applying relation (3.13) shows the desired result. To verify part

2 we observe for $\mathbf{X}_1 \sim G(\alpha, \beta, \tau)$ that

$$\ln(\mathbf{X}_1) \stackrel{d}{=} \ln(\beta) + \tau^{-1} \ln(\mathbf{Y}_1)$$

with $\mathbf{Y} \sim G(\alpha, 1)$.Applying relation (3.15) yields

$$\text{skewness}(\ln(\mathbf{X}_1)) = \text{skewness}(\ln(\mathbf{Y}_1)).$$

To compute skewness($\ln(\mathbf{Y}_1)$) we observe by relation 3.17 that

$$\text{skewness}(\ln(\mathbf{Y}_1)) = \frac{\kappa_3(\ln(\mathbf{Y}_1))}{\kappa_2(\ln(\mathbf{Y}_1))^{\frac{3}{2}}}$$

Clearly the cumulant generating function of the random variable $\ln(\mathbf{Y}_1))$ with $\mathbf{Y}_1 \sim G(\alpha, 1)$ is given by

$$K(s) = \mathbb{E}(e^{s \ln(\mathbf{Y}_1)}) = \mathbb{E}(\mathbf{Y}_1^s) = \frac{\Gamma(\alpha + s))}{\Gamma(\alpha)}$$

and so

$$\kappa_2(\ln(\mathbf{Y}_1)) = K^{(2)}(0) = \psi_1(\alpha), \kappa_3(\ln(\mathbf{Y}_1)) = \psi_2(\alpha) \tag{3.20}$$

with ψ_m the polygamma function of order m defined in the Appendix A. This shows the desired result. $\square$

By Lemma A.0.8 it follows that skewness($\mathbf{X}_1$) for any $\mathbf{X}_1 \sim G(\alpha, \beta, \tau)$ is a strictly increasing function of α with limit at $\alpha = 0$ given by -2 and at $\alpha = \infty$ equal to zero. We will now apply the method of moments approach to a (generalized) gamma population and start with the well known case of a gamma population.

Lemma 3.2.5. *If $\mathbf{X}_1 \sim G(\alpha, \beta)$ then $\widehat{\mu}_2(\mathbf{X}) > 0$ and $\widehat{\mu}'_1(\mathbf{X}) > 0$ with probability 1 and the methods of moment estimator $\widehat{\beta}_{MM}(\mathbf{X})$ of β, respectively $\widehat{\alpha}_{MM}(\mathbf{X})$ of α are given by*

$$\widehat{\beta}_{MM}(\mathbf{X}) \stackrel{a.s}{=} \frac{\widehat{\mu}_2(\mathbf{X})}{\widehat{\mu}'_1(\mathbf{X})}, \widehat{\alpha}_{MM}(\mathbf{X}) \stackrel{a.s}{=} \frac{\widehat{\mu}'_1(\mathbf{X})^2}{\widehat{\mu}_2(\mathbf{X})}.$$

Proof. Since the gamma distribution is continuous the first part of the lemma is obvious. By relation (3.13) it follows that $\mu'_1(\mathbf{X}_1) = \alpha\beta$ and $\mu'_2(\mathbf{X}_1) = \beta^2(\alpha+1)\alpha$ and so $\mu_2(\mathbf{X}_1) = \beta^2\alpha$. By the method of moments approach and Lemma 3.2.1 we need to solve the system

$$\widehat{\mu}'_1(X) = \alpha\beta, \widehat{\mu}_2(X) = \beta^2\alpha$$

Since $0 < \widehat{\mu}_2(\mathbf{X})$ with probability 1 we obtain

$$\frac{\widehat{\mu}_1'(X)^2}{\widehat{\mu}_2(X)} = \alpha, \widehat{\mu}_1'(X) = \alpha\beta$$

and this shows the desired result. □

Clearly the estimator $\widehat{\mu}_1'(\mathbf{X})^2\widehat{\mu}_2(\mathbf{X})^{-1}$ is the sample estimator of the inverse squared coefficient of variation of a gamma distribution and this result is clear due to Lemma 3.2.4. In the remainder of this section we will derive method of moment estimators for a generalized gamma population by applying the method of moments approach to the random variable $\ln(\mathbf{X}_1)$. In [17] another set of method of moments estimators are proposed by applying the methods of moment approach to the random variable $\mathbf{X}_1$. However, the corresponding set of equations derived in [17] are much more complicated and it is unclear, contrary to our proposal, under which conditions this system has solution and if so, how one can easily find such a solution. Also, if the system in [17] has a solution, it is not obvious whether this solution is unique. We first list the following auxiliary result.

Lemma 3.2.6. *If $\mathbf{X}_1 \sim G(\alpha, \beta, \tau)$ then applying the method of moments approach to the random variable $\ln(\mathbf{X}_1)$ one needs to find for every realized sample $X = (X_1, ..., X_n)$ a solution of the system*

$$\frac{\widehat{\mu}_3(\ln(X))}{\widehat{\mu}_2(\ln(X))^{\frac{3}{2}}} = \psi_2(\alpha)\psi_1(\alpha)^{-\frac{3}{2}}. \tag{3.21}$$

Proof. Clearly we obtain by Definition 3.2.1 that

$$\ln(\mathbf{X}_1) \stackrel{d}{=} \ln(\beta) + \tau^{-1}\ln(\mathbf{Y}_1)$$

with $\mathbf{Y}_1 \sim G(\alpha, 1)$. Also by relations (3.9),(3.19), (3.20) and (3.16) we obtain

$$\mathbb{E}(\ln(\mathbf{X}_1)) = \ln(\beta) + \tau^{-1}\psi_0(\alpha), \mu_2(\ln(\mathbf{X}_1)) = \tau^{-2}\psi_1(\alpha), \mu_3(\ln(\mathbf{X}_1)) = \tau^{-3}\psi_2(\alpha). \tag{3.22}$$

To determine for each realized sample $X = (x_1, ..., x_n)$ an estimate of the unknown parameters using the method of moment approach applied to the random variable $\ln(\mathbf{X}_1)$ it follows by Lemma 3.2.1 and relation (3.22) that one therefore needs to find a solution of the nonlinear system of equations

$$\begin{aligned}\widehat{\mu}_1'(\ln(X)) &= \mathbb{E}(\ln(\mathbf{X}_1)) &= \ln(\beta)+\tau^{-1}\psi_0(\alpha)\\ \widehat{\mu}_2(\ln(X)) &= \mu_2(\ln(\mathbf{X}_1)) &= \tau^{-2}\psi_1(\alpha)\\ \widehat{\mu}_3(\ln(X)) &= \mu_3(\ln(\mathbf{X}_1)) &= \tau^{-3}\psi_2(\alpha)\end{aligned} \tag{3.23}$$

This shows by Lemma 3.2.4 that

$$\frac{\widehat{\mu}_3(\ln(X))}{\widehat{\mu}_2(\ln(X))^{\frac{3}{2}}} = \frac{\mu_3(\ln(\mathbf{X}_1))}{\mu_2(\ln(\mathbf{X}_1))^{\frac{3}{2}}} = \text{skewness}(\mathbf{X}_1) = \psi_2(\alpha)\psi_1(\alpha)^{-\frac{3}{2}}$$

and we have shown the desired result. □

Using Lemma 3.2.6 it is easy to show the following result.

Lemma 3.2.7. *The method of moments system for the estimation of α,β,τ in a generalized gamma population is consistent if and only if the sample $X=(x_1,...,x_n)$ satisfies*

$$-2 < \frac{\widehat{\mu}_3(\ln(X))}{\widehat{\mu}_2(\ln(X))^{\frac{3}{2}}} < 0.$$

If the system is consistent then $\widehat{\alpha}_{MM}(X)$ is the unique solution of the system

$$\frac{\widehat{\mu}_3(\ln(X))}{\widehat{\mu}_2(\ln(X))^{\frac{3}{2}}} = \psi_2(\alpha)\psi_1(\alpha)^{-\frac{3}{2}} \tag{3.24}$$

and the method of moments estimator $\widehat{\beta}_{MM}(\mathbf{X}),\widehat{\tau}_{MM}(\mathbf{X})$ of β, respectively τ are given by

$$\widehat{\beta}_{MM}(\mathbf{X}) = \exp\left(\widehat{\mu}_1'(\ln(\mathbf{X})) - \sqrt[2]{\frac{\widehat{\mu}_2(\ln(\mathbf{X}))}{\psi_1(\widehat{\alpha}_{MM}(\mathbf{X}))}}\psi_0(\widehat{\alpha}_{MM}(\mathbf{X}))\right) \tag{3.25}$$

and

$$\widehat{\tau}_{MM}(\mathbf{X}) = \sqrt[2]{\frac{\psi_1(\widehat{\alpha}_{MM}(\mathbf{X}))}{\widehat{\mu}_2(\ln(X))}} \tag{3.26}$$

Proof. By Lemma A.0.8 and relation (A.0.8) the first part immediately follows. To prove the second part we need to solve by relation 3.23 the system of equations

$$\begin{aligned}\widehat{\mu}_1'(\ln(X)) &= \ln(\beta)+\tau^{-1}\psi_0(\widehat{\alpha}_{MM}(\ln(X))\\ \widehat{\mu}_2(\ln(X)) &= \tau^{-2}\psi_1(\widehat{\alpha}_{MM}(\ln(X))\end{aligned}$$

This shows the result. $\square$

In the next algorithm we list the procedure to calculate the proposed method on moments estimates in a a generalized gamma population.

Algorithm 3.2.1. *Algorithm to compute method of moment estimates of unknown parameters β, α, τ*

9.1 *Check for the given sample $X = (x_1, ..., x_n)$ whether*

$$-2 < \frac{\widehat{\mu}_3(\ln(X))}{\widehat{\mu}_2(\ln(X))^{\frac{3}{2}}} < 0.$$

If not stop, otherwise continue

9.2 *Determine by bisection the unique solution $\widehat{\alpha}_{MM}(X)$ of the system*

$$\frac{\widehat{\mu}_3(\ln(X))}{\widehat{\mu}_2(\ln(X))^{\frac{3}{2}}} = \psi_2(\alpha)\psi_1(\alpha)^{-\frac{3}{2}}.$$

9.3 *Evaluate*

$$\widehat{\beta}_{MM}(X) = \exp\left(\widehat{\mu}_1'(\ln(X)) - \sqrt[2]{\frac{\widehat{\mu}_2(\ln(X))}{\psi_1(\widehat{\alpha}_{MM}(X))}}\psi_0(\widehat{\alpha}_{MM}(X))\right)$$

and

$$\widehat{\tau}_{MM}(X) = \sqrt[2]{\frac{\psi_1(\widehat{\alpha}_{MM}(X))}{\widehat{\mu}_2(\ln(X))}}.$$

To start the bisection one could use the following approach. By relation (A.17) it follows that

$$\lim_{\alpha \uparrow \infty} p(\alpha)\alpha^{\frac{1}{2}} = -1 \tag{3.27}$$

Using this asymptotic result start the bisection in the initial point $\alpha_0 = \widehat{\mu}_2(\ln(X))^6 \widehat{\mu}_3(\ln(X))^{-2}$. For α is large this value approximates clearly the solution of the system in relation (3.24).

To compute the method of moment estimators of a Weibull distribution we can apply the same procedure as above replacing $\widehat{\alpha}_{MM}(X)$ by 1. This shows that the method of moments estimators of the parameters β and τ in a Weibull population are given by

$$\widehat{\beta}_{MM}(\mathbf{X}) = \exp\left(\widehat{\mu}_1'(\ln(\mathbf{X})) - \sqrt[2]{\frac{\widehat{\mu}_2(\ln(\mathbf{X}))}{\psi_1(1)}}\psi_0(1)\right), \widehat{\tau}_{MM}(\mathbf{X}) = \sqrt[2]{\frac{\psi_1(1)}{\widehat{\mu}_2(\ln(\mathbf{X}))}} \tag{3.28}$$

Introducing the event

$$C = \left\{-2 < \frac{\widehat{\mu}_3(\ln(\mathbf{X}))}{\widehat{\mu}_2(\ln(\mathbf{X}))^{\frac{3}{2}}} < 0\right\}$$

conditionally on this event the proposed method of moments approach gives a unique solution. In the next section we will by means of simulation give an estimate of the probability of occurrence of this event. Clearly the probability of this event might depend on all the unknown parameters and so we estimate in the next computational section this probability for different values of the unknown parameters.

3.3 Computational results

In this section we will perform some simulation experiments comparing the behaviour of the classical maximum likelihood estimators and the method of moments estimator. In the first subsection we report the results for a gamma population and a Weibull population using the method of moments estimators derived in Lemma 3.2.5 for the gamma distribution and Lemma 3.2.3 for the Weibull distribution. In this particular case the random variable $\mathbf{Z}_1$ has a exponential distribution with parameter 1. In the last subsection we report the performance of the method of moment estimator proposed in this paper for a generalized gamma population and compare this performance with the classical maximum likelihood estimators. We also give an estimation of the probability for different sizes of the sample that the set of equations related to the method of moments estimators has a unique solution. To calculate the maximum likelihood estimators for both distributions, we use the efficient special purpose algorithms proposed in Chapter 2.

3.3.1 Computational results for a gamma and Weibull population.

In this section, we report the results for a gamma and Weibull population. In general, comparing the absolute estimated bias and mean square error the maximum likelihood estimators yields a better performance than the method of moment estimators for both the gamma and Weibull distributions. For both estimators the estimated parameters are not that accurate for small sample sizes. This can be seen from Tables 3.1 for the gamma distribution and from Tables 3.2 for the Weibull distribution. Moreover, our experiments show that changing the value of the scale parameter β does not affect the relative efficiency of the estimators in both cases. From our set of experiments we may conclude that both estimators have a good performance unless α is large and the sample size is small for the gamma distribution and τ is small and the sample size is small for the Weibull distribution.

Sample Size (n)		α	$\lvert\hat{\alpha}_{ML}-\alpha\rvert$	$\lvert\hat{\alpha}_{MM}-\alpha\rvert$	$\lvert\hat{\beta}_{ML}-\beta\rvert$	$\lvert\hat{\beta}_{MM}-\beta\rvert$
10	Bias	0.1	0.0216	0.1765	0.0650	0.5490
		1	0.3554	0.5887	0.0980	0.1906
		10	4.1581	4.3913	0.1031	0.1122
	MSE	0.1	0.0028	0.0524	1.0641	0.5748
		1	0.7217	1.1780	0.2454	0.3003
		10	90.4770	94.5732	0.1893	0.2018
100	Bias	0.1	0.0016	0.0241	0.0026	0.1024
		1	0.0259	0.0613	0.0106	0.0253
		10	0.3003	0.3301	0.0097	0.0104
	MSE	0.1	0.0001	0.0020	0.1122	0.2057
		1	0.0179	0.0424	0.0253	0.0455
		10	2.2881	2.5469	0.0207	0.0230

Table 3.1 Mean absolute bias and mean square error for the estimated parameter α and β of the gamma distribution using 10000 simulation with different samples sizes for $\beta = 1$ and different values of α

.

Sample Size (n)		τ	$\lvert\hat{\tau}_{ML}-\tau\rvert$	$\lvert\hat{\tau}_{MM}-\tau\rvert$	$\lvert\hat{\beta}_{ML}-\beta\rvert$	$\lvert\hat{\beta}_{MM}-\beta\rvert$
10	Bias	0.1	0.0171	0.0223	50.6055	46.2939
		1	0.1710	0.2234	0.0107	0.0041
		10	1.7108	2.2341	0.0038	0.0054
	MSE	0.1	0.0015	0.0021	400854	343294
		1	0.1538	0.2184	0.1113	0.1109
		10	15.3849	21.8437	0.0011	0.0011
100	Bias	0.1	0.0014	0.0020	0.6429	0.6507
		1	0.0147	0.0209	0.0012	0.0005
		10	0.1475	0.2098	0.0003	0.0005
	MSE	0.1	6.6697e-5	0.0001	5.0306	5.4106
		1	0.0066	0.0115	0.0110	0.0115
		10	0.6669	1.1535	0.0001	0.0001

Table 3.2 Mean absolute bias and mean square error for the estimated parameters τ and β for the Weibull distribution using 10000 simulation with different samples sizes for $\beta = 1$ and different values of τ.

3.3.2 Computational results for a generalized gamma population.

To calculate for the proposed method of moments estimator a unique solution of the corresponding system of nonlinear equations it must hold by Lemma 3.2.7 that the sample estimator of skewness($\mathbf{X}_1$) has a realisation strictly between 0 and -2 . In Figure 3.1 we draw the graph of skewness ($\mathbf{X}_1$) as a function of α. In the same

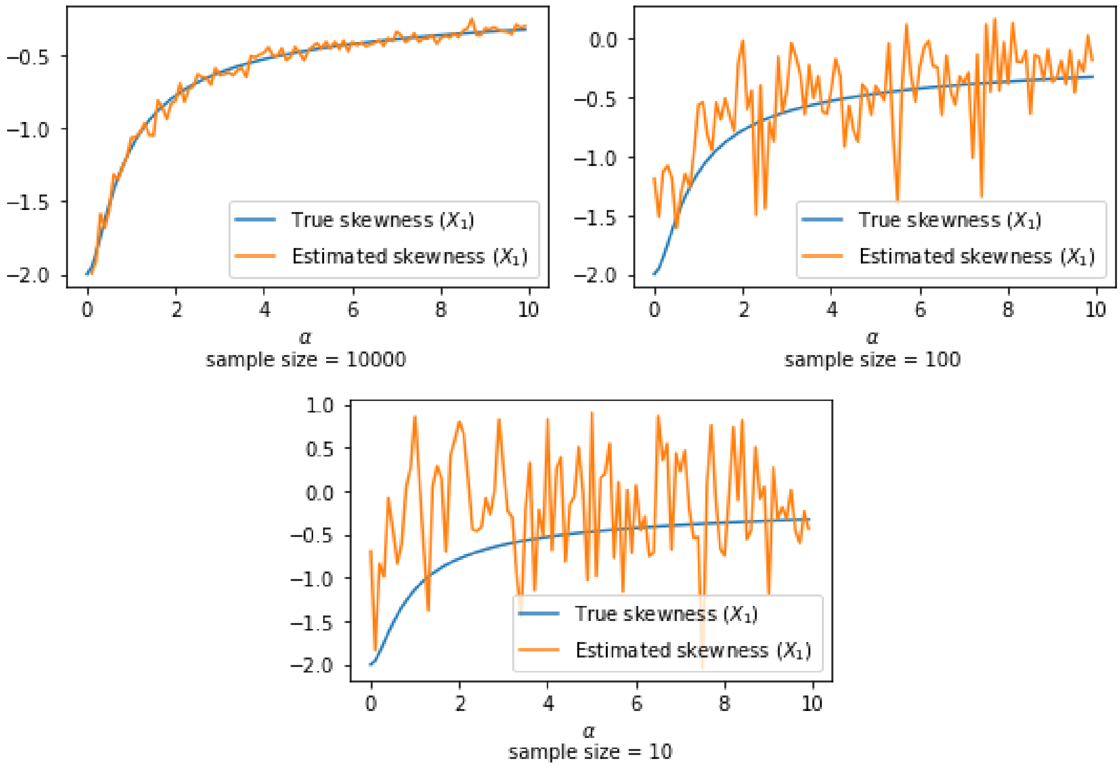

Figure 3.1 the graph of the function p and the estimation of this function for given values of α using sample moments with different sample sizes for parameters $\beta = 1$ and $\tau = 1$

figure we show for different sample sizes for the scenarios $\beta = 1$, $\tau = 1$ and different values of α realisations of the sample moment estimator of skewness $(\mathbf{X}_1)$. For large sample sizes this behaviour is robust and with probability close to 1 one can find the method of moments estimators for α and the other parameters using Algorithm 3.2.1. If the sample size is small the probability that the sample moment estimator of the skewness$(\mathbf{X}_1)$ is between 0 and -2 and hence the proposed method of moment estimator cannot be evaluated is smaller. Observe in Figure (Figure 3.3) we have drawn the graph of the estimated probability that the sample moment estimator of the skewness is between the values 0 and -2 for different values of α.

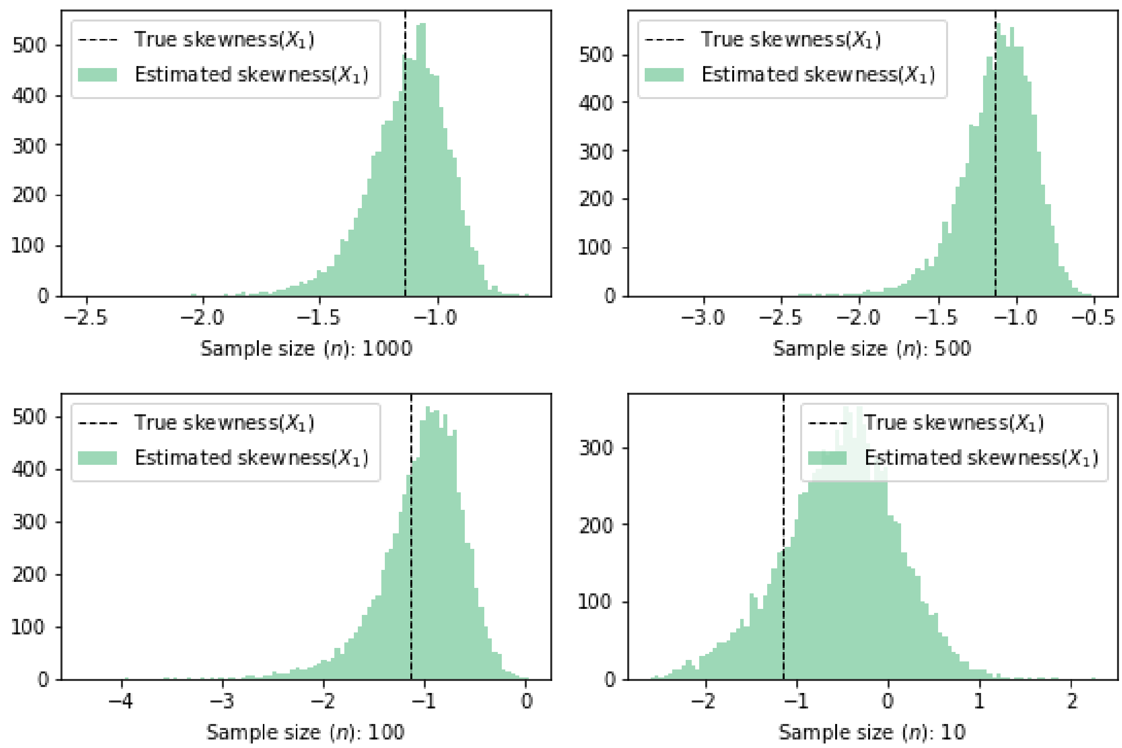

Figure 3.2 the histogram of 10000 simulation of estimated function p with different sample sizes for parameters $\alpha = 1$, $\beta = 1$, and $\tau = 1$ (the real $p(\alpha) = -1.1395$)

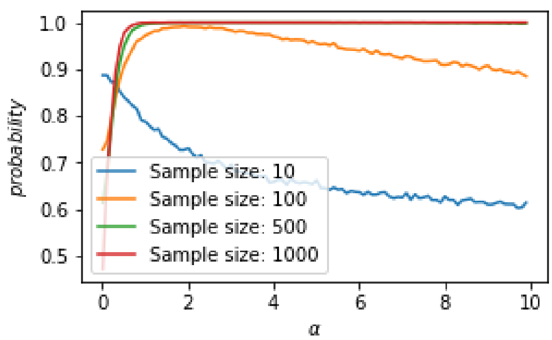

Figure 3.3 the graph of the estimated probability that the sample moment estimator of the skewness is between -2 and 0 using 10000 simulation for each values of α for the scenario $\beta = 1$ and $\tau = 1$

3.4 Conclusion.

In general, for a gamma or Weibull population one should apply, if possible, the maximum likelihood estimators for estimating the unknown parameters However, if one cannot do this due to the nonavailability of an efficient algorithm, the proposed analytical solution of the method of moment estimator gives a reasonable estimation especially for larger sample values. Within a generalized gamma population, especially for larger sample sizes. the newly proposed easy to compute method of moment estimator also yields reasonable estimates (again not as accurate as the maximum likelihood estimators but these estimates can only be found by applying a much more complicated algorithm). In general one could also use the easy to obtain estimates of the proposed method of moment estimators as an initial guess in the maximum likelihood algorithm to speed up the computation.

4. On the sample mth moment and the maximum likelihood estimator of the mth (central) moment in a normal population.

4.1 Introduction.

In this paper we will compare for a sample arising from a normal population the sample mth (central) moment estimator for the mth (central) moment with a maximum likelihood based estimator. The sample mth moment estimator does not use information about the underlying parametric class of distributions and their application is justified by the strong law of large numbers. However, if we estimate a (central) moment we might also use (if available) the information from which parametric class of distributions the sample arises. We could apply that information in our estimation problem and use more complicated estimators based on for example the maximum likelihood principle. The easiest example to which one can apply this idea is to assume that the sample arises from a normal population.

In statistics courses one introduces the sample variance and the sample mean as estimators for the (unknown) variance and first moment. In only a few books (cf.[29], [30]) related to large sample theory the problem of estimating higher moments and the use of higher sample moment estimators are discussed. The authors did not encounter a recent paper in the literature showing more properties of these higher sample moment estimators. This in particular applies to estimating higher (central) moments in a normal population and the alternative use of different estimators based on the maximum likelihood principle. Observe these higher (central) moment estimators can be used in estimating the skewness or kurtosis of a certain distribution. Within a normal population these two different type of estimators are the same in estimating the mean and the variance. Since they differ for higher (central) moments the question arises which estimator (maximum likelihood based or sample moment) is more efficient in estimating these higher (central) moments. Since sample higher moments averages are easier to compute than the proposed maximum likelihood estimators one would select due to computational efficiency the first class of estimators. However the question remains whether this leads to a loss of statistical efficiency by ignoring from which parametric class of distributions the sample did arise. To answer this question for the normal family is the research question in this paper. In particular we derive a closed form expression for the mean square error of both estimators of the mth (central) moment for arbitrary integer m. An outcome of this analysis is that for the estimation of the mth central moment in any normal distribution it is possible to determine the threshold value of the sample for which the mean square error of the maximum likelihood estimator is below the mean square error of the sample mth central moment estimator.

The outline of this paper is as follows. In the next section we will discuss some well known results for the normal distribution and introduce the proposed maximum likelihood based estimators and sample moment estimators for the higher (central) moments. At the same time we derive for all considered estimators an exact computable expression for the variance and the mean of these estimators thereby enabling us to compute the mean square error of these estimators. In the last section some computational experiments are performed and the outcomes of both estimators are compared. Also by means of repeated simulations we draw the histogram of both estimators and visually compare the statistical efficiency of these estimators.

4.2 On the mth sample moment and the maximum likelihood estimators for the mth (central) moment

Let $(\Omega, \mathbb{P}, \mathcal{F})$ be a given probability space. It is assumed in this section that all random variables are defined on this probability space. Before starting our analysis we introduce some notation. If the random variables $\mathbf{Z}_1$ and $\mathbf{Z}_2$ have the same cumulative distribution function this is denoted by $\mathbf{Z}_1 \stackrel{d}{=} \mathbf{Z}_2$. Also for any random variable $\mathbf{Z}$ its mth moment is given by

$$\mu'_m(\mathbf{Z}) := \mathbb{E}(\mathbf{Z}^m) \tag{4.1}$$

and its mth central moment by

$$\mu_m(\mathbf{Z}) := \mu'_m(\mathbf{Z} - \mathbb{E}(\mathbf{Z})) = \mathbb{E}((\mathbf{Z} - \mathbb{E}(\mathbf{Z}))^m), m \in \mathbb{N}.$$

Clearly for $m = 2$ the 2nd central moment is also known as the variance $Var(\mathbf{Z})$.

Definition 4.2.1. *Let $(\Omega, \mathbb{P}, \mathcal{F})$ be a probability space. A random variable $\mathbf{Z}$ has a normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ denoted by $\mathbf{Z} \sim N(\mu, \sigma)$ if its density f is given by*

$$f(z) = \frac{1}{\sigma \sqrt[2]{2\pi}} e^{-\frac{1}{2}(\frac{z-\mu}{\sigma})^2}, z \in \mathbb{R}$$

In case $\mu = 0$ and $\sigma = 1$ the normal distribution is called a standard normal distribution.

In case we observe a sample $\mathbf{X} = (\mathbf{X}_1, ..., \mathbf{X}_n)$ of size n from a normal population and we need to estimate the mth moment for $m \in \mathbb{N}$ or the mth central moment for m even (for m odd we know that the mth central moment is zero and we do not need to estimate it!) we mostly use the sample mth moment estimator

$$\widehat{\mu}'_{m,n}(\mathbf{X}) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i^m \tag{4.2}$$

for estimating the mth moment or the sample mth central moment estimator

$$\widehat{\mu}_{m,n}(\mathbf{X}) := \frac{1}{n} \sum_{i=1}^{n} (\mathbf{X}_i - \overline{\mathbf{X}}_n)^m, \overline{\mathbf{X}}_n := \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \tag{4.3}$$

for estimating the mth central moment. An important measure to estimate the efficiency of a estimator $T(\mathbf{X})$ with $\mathbf{X}$ some sample of size n estimating some unknown parameter θ is the mean squared deviation $\mathrm{MSE}(T(\mathbf{X}))$ given by (cf.[5])

$$\mathrm{MSE}(T(\mathbf{X})) = \mathbb{E}((T(\mathbf{X}) - \theta)^2) = \mathrm{Var}(T(\mathbf{X})) + \mathrm{bias}(T(\mathbf{X}))^2. \tag{4.4}$$

with

$$\mathrm{bias}(T(\mathbf{X}) := \mathbb{E}(T(\mathbf{X})) - \theta \tag{4.5}$$

A useful property of the sample mth central moment estimator is given by the following. Since the random variables $\mathbf{X}_i, i = 1, ..., n$ are independent and normally distributed with parameter $\mu \in \mathbb{R}$ and $\sigma > 0$ it follows that (cf.[5])

$$\mathbf{X}_i \stackrel{d}{=} \mu + \sigma \mathbf{Y}_i$$

with $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n)$ a random vector consisting of independent standard normal distributed random variables. This implies

$$\widehat{\mu}_{m,n}(\mathbf{X}) \stackrel{d}{=} \frac{\sigma^m}{n} \sum_{i=1}^{n} (\mathbf{Y}_i - \overline{\mathbf{Y}}_n)^m = \sigma^m \widehat{\mu}_{m,n}(\mathbf{Y}) \tag{4.6}$$

and so

$$\mathbb{E}(\widehat{\mu}_{m,n}(\mathbf{X})) = \sigma^m \mathbb{E}(\widehat{\mu}_{m,n}(\mathbf{Y})), \mathrm{Var}(\widehat{\mu}_{m,n}(\mathbf{X})) = \sigma^{2m}\mathrm{Var}(\widehat{\mu}_{m,n}(\mathbf{Y})) \tag{4.7}$$

and

$$\mathrm{MSE}(\widehat{\mu}_{m,n}(\mathbf{X})) = \sigma^{2m}\mathrm{MSE}(\widehat{\mu}_{m,n}(\mathbf{Y})). \tag{4.8}$$

In most textbooks on statistics (see for example [5]) only estimating the mean and or the variance of a normal population is considered. In this case we mostly use for the variance the unbiased sample variance $S_n^2(\mathbf{X})$ given by

$$S_n(\mathbf{X}) = \frac{n}{n-1}\widehat{\mu}_{2,n}(\mathbf{X})$$

and for the mean the sample mean $\overline{\mathbf{X}}_n$. Before introducing an alternative maximum likelihood estimator for the mth central moment we discuss some properties of the sample mth moment estimator. The next result is listed in [30] and its proof is standard.

Lemma 4.2.1. *It follows for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$ that*

$$\mathbb{E}(\widehat{\mu}'_{m,n}(\mathbf{X})) = \mu'_m(\mathbf{X}_1), MSE(\widehat{\mu}'_{m,n}(\mathbf{X})) = \frac{1}{n}\,Var(\mathbf{X}_1^m) = \frac{\mu'_{2m}(\mathbf{X}_1) - \mu'_m(\mathbf{X}_1)^2}{n}. \tag{4.9}$$

with $\mu'_{2m}(\mathbf{X}_1)$ and $\mu'_m(\mathbf{X}_1)$ given in relation (A.18).

We will now discuss some properties of the sample mth central moment estimator. In the next result we show for a normal population that the estimator $(\frac{n}{n-1})^{\frac{m}{2}}\widehat{\mu}_m(\mathbf{X})$ is an unbiased estimator of the mth central moment.

Lemma 4.2.2. *It follows for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$ that*

$$\mathbb{E}(\widehat{\mu}_{m,n}(\mathbf{X})) = (1-n^{-1})^{\frac{m}{2}}\sigma^m \mu_m(\mathbf{Z}) = (1-n^{-1})^{\frac{m}{2}}\mu_m(\mathbf{X}_1). \tag{4.10}$$

with $\mathbf{Z}$ a standard normal distributed random variable.

Proof. By relation (4.7) we only need to compute $\mathbb{E}(\widehat{\mu}_{m,n}(\mathbf{Y}))$. By the definition of the mth sample central moment estimator and the random variables $\mathbf{Y}_i, i = 1, ..., n$ are independent and identically distributed we obtain

$$\mathbb{E}(\widehat{\mu}_{m,n}(\mathbf{Y})) = \mathbb{E}((\mathbf{Y}_1 - \overline{\mathbf{Y}}_n)^m) = n^{-m}\mathbb{E}(((n-1)\mathbf{Y}_1 - \textstyle\sum_{i=2}^n \mathbf{Y}_i)^m)$$

Since the random variables $\mathbf{Y}_i, i = 1, ..., n$ are independent and standard normal distributed it follows by the addition property of the sum of independent normally distributed random variables that

$$(n-1)\mathbf{Y}_1 - \sum_{i=2}^{n} \mathbf{Y}_i \overset{d}{=} \sqrt[2]{(n-1)^2+n-1}\mathbf{Z} = \sqrt[2]{(n^2-n)}\mathbf{Z}$$

with $\mathbf{Z}$ having a standard normal distribution. This shows

$$\mathbb{E}(\widehat{\mu}_{m,n}(\mathbf{Y})) = n^{-m}(n^2-n)^{\frac{m}{2}}\mu'_m(\mathbf{Z})$$

and using $\sigma^m\mu'_m(\mathbf{Z}) = \sigma^m\mu_m(\mathbf{Z}) = \mu_m(\mathbf{X}_1)$ we obtain relation (4.10). □

In the next result we list a formula for the variance of the sample mth central moment estimator. To show this result we need a result about the cdf of a quadratic form proved in the Appendix A.

Lemma 4.2.3. *It follows that*

$$Var(\widehat{\mu}_{m,n}(\mathbf{X})) = \sigma^{2m}(1-n^{-1})\left[\mathbb{E}(\mathbf{V}_n^m) - \left(1-n^{-1}\right)^{m-1}\left(\mu_m(\mathbf{Z}_1)^2 - \frac{1}{n}\mu_{2m}(\mathbf{Z}_1)\right)\right]$$

with $\mathbf{V}_n \overset{d}{=} (2^{-1}-n^{-1})\mathbf{Z}_1 - 2^{-1}\mathbf{Z}_2$ *and* $\mathbf{Z}_1, \mathbf{Z}_2$ *independent standard normal distributed random variables.*

Proof. It is obvious by relation (4.7) that

$$\text{Var}(\widehat{\mu}_{m,n}(\mathbf{X})) = \frac{\sigma^{2m}}{n^2}\text{Var}\left(\sum_{i=1}^{n}(\mathbf{Y}_i - \overline{\mathbf{Y}}_n)^m\right) \tag{4.11}$$

and

$$\begin{aligned} &\text{Var}\left(\sum_{i=1}^{n}(\mathbf{Y}_i-\overline{\mathbf{Y}}_n)^m\right) \\ = \ &\sum_{i=1}^{n}\text{Var}\left((\mathbf{Y}_i-\overline{\mathbf{Y}}_n)^m\right) + \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\text{Cov}((\mathbf{Y}_i-\overline{\mathbf{Y}}_n)^m, (\mathbf{Y}_j-\overline{\mathbf{Y}}_n)^m). \end{aligned} \tag{4.12}$$

Since the random variables $\mathbf{Y}_i, i=1,..,n$ are independent and identically distributed we obtain for every $1 \le i \le n$

$$\text{Var}\left((\mathbf{Y}_i-\overline{\mathbf{Y}}_n)^m\right) = \text{Var}\left((\mathbf{Y}_1-\overline{\mathbf{Y}}_n)^m\right) = \mathbb{E}((\mathbf{a}_n^\top\mathbf{Y})^{2m}) - \mathbb{E}((\mathbf{a}_n^\top\mathbf{Y})^m)^2 \tag{4.13}$$

and $\mathbf{a}_n := \mathbf{e}_{1,n} - \frac{1}{n}\mathbf{i}_n, \mathbf{b}_n := \mathbf{e}_{2,n} - \frac{1}{n}\mathbf{i}_n$ with $\mathbf{i}_n$ denoting the nth dimensional vector consisting only of ones and $\mathbf{e}_{i,n}, i=1,2$ the ith unit vector in $\mathbb{R}^n$. By the same argument it follows for every $1 \le i, j \le n$

(4.14)

$$\begin{aligned} \operatorname{Cov}((\mathbf{Y}_i - \overline{\mathbf{Y}}_n)^m, (\mathbf{Y}_j - \overline{\mathbf{Y}}_n)^m) &= \operatorname{Cov}((\mathbf{Y}_1 - \overline{\mathbf{Y}}_n)^m, (\mathbf{Y}_2 - \overline{\mathbf{Y}}_n)^m) \\ &= \mathbb{E}(((\mathbf{Y}_1 - \overline{\mathbf{Y}}_n)(\mathbf{Y}_2 - \overline{\mathbf{Y}}_n)^m)) - \mathbb{E}((\mathbf{Y}_1 - \overline{\mathbf{Y}}_n)^m \mathbb{E}((\mathbf{Y}_2 - \overline{\mathbf{Y}}_n)^m) \\ &= \mathbb{E}(((\mathbf{Y}_1 - \overline{\mathbf{Y}}_n)(\mathbf{Y}_2 - \overline{\mathbf{Y}}_n)^m)) - \mathbb{E}((\mathbf{Y}_1 - \overline{\mathbf{Y}}_n)^m)^2 \\ &= \mathbb{E}(\left(\mathbf{Y}^\top \mathbf{a}_n \mathbf{b}_n^\top \mathbf{Y}\right)^m) - \mathbb{E}((\mathbf{a}_n^\top \mathbf{Y})^m)^2. \end{aligned}$$

This shows substituting relation (4.13) and 4.14 into relation (4.12) that

$$\operatorname{Var}\left(\sum_{i=1}^n (\mathbf{Y}_i - \overline{\mathbf{Y}}_n)^m\right) = n(n-1)\mathbb{E}\left(\left(\mathbf{Y}^\top \mathbf{a}_n \mathbf{b}_n^\top \mathbf{Y}\right)^m\right) + n\mathbb{E}((\mathbf{a}_n^\top \mathbf{Y})^{2m}) - n^2 \mathbb{E}((\mathbf{a}_n^\top \mathbf{Y})^m)^2$$

and we obtain by relation (4.11)

$$\operatorname{Var}(\widehat{\mu}_{m,n}(\mathbf{X})) = \sigma^{2m}\left[(1-n^{-1})\alpha_m + n^{-1}\mathbb{E}((\mathbf{a}_n^\top \mathbf{Y})^{2m}) - \mathbb{E}((\mathbf{a}_n^\top \mathbf{Y})^m)^2\right] \tag{4.15}$$

with $\alpha_m := \mathbb{E}\left(\left(\mathbf{Y}^\top \mathbf{a}_n \mathbf{b}_n^\top \mathbf{Y}\right)^m\right)$. Since $\| \mathbf{a}_n \|_2^2 = \| \mathbf{b}_n \|_2^2 = 1 - n^{-1}$ it follows by the addition property of independent standard normal distributed random variables that

$$\mathbf{a}_n^\top \mathbf{Y} \stackrel{d}{=} \sqrt[2]{1-n^{-1}}\mathbf{Z}_1, \mathbf{b}_n^\top \mathbf{Y} \stackrel{d}{=} \sqrt[2]{1-n^{-1}}\mathbf{Z}_2 \tag{4.16}$$

with $\mathbf{Z}_i, i = 1,2$ having a standard normal distribution. This shows

$$\mathbb{E}((\mathbf{a}_n^\top \mathbf{Y})^m) = \mathbb{E}((\mathbf{b}_n^\top \mathbf{Y})^m) = (1-n^{-1})^{\frac{m}{2}} \mu_m(\mathbf{Z}_1) \tag{4.17}$$

and so by relation (4.15)

$$\operatorname{Var}(\widehat{\mu}_{m,n}(\mathbf{X})) = \sigma^{2m}\left[(1-n^{-1})\alpha_m + (1-n^{-1})^m (n^{-1}\mu_{2m}(\mathbf{Z}) - \mu_m(\mathbf{Z})^2)\right] \tag{4.18}$$

Since in Lemma A.0.11 we have shown

$$\mathbf{Y}^\top \mathbf{a}_n \mathbf{b}_n^\top \mathbf{Y} \stackrel{d}{=} \left(\frac{1}{2} - \frac{1}{n}\right)\mathbf{Z}_1^2 - \frac{1}{2}\mathbf{Z}_2^2$$

with $\mathbf{Z}_1, \mathbf{Z}_2$ independent standard normal distributed random variables the desired result follows using relation (4.15). □

To compute $\operatorname{Var}(\widehat{\mu}_{m,n}(\mathbf{X}))$ it follows by Lemma 4.2.3 that we need to compute the mth moment of the random variable $\mathbf{V}_n$. Clearly by Newtons binomial formula and the random variables $\mathbf{Z}_i, i = 1,2$ are independent we obtain

$$\begin{aligned}\mathbb{E}(\mathbf{V}_n^m) &= \mathbb{E}(((2^{-1}-n^{-1})\mathbf{Z}_1^2 - 2^{-1}\mathbf{Z}_2^2)^m) \\ &= \sum_{k=0}^m \binom{m}{k}(-1)^{m-k}(2^{-1}-n^{-1})^k 2^{-(m-k)}\mu_{2k}(\mathbf{Z}_1)\mu_{2(k-m)}(\mathbf{Z}_2).\end{aligned}$$

By relation A.20 we know that

$$\mu_{2k}(\mathbf{Z}_1) = \mu_{2k}(\mathbf{Z}_2) = \frac{(2k)!}{2^m k!}$$

To investigate the asymptotic behaviour of the mean squared error of the maximum likelihood estimator of the mth central moment as the sample size goes to infinity we observe using a different technique that for all populations having a finite $2m$th central moment (cf.[30],[29])

$$\lim_{n\uparrow\infty} n\mathrm{Var}(\widehat{\mu}_{m,n}) = \alpha_m \tag{4.19}$$

with

(4.20)

$$\alpha_m := \mu_{2m}(\mathbf{X}_1) - 2m\mu_{m-1}(\mathbf{X}_1)\mu_{m+1}(\mathbf{X}_1) - \mu_m(\mathbf{X}_1)^2 + m^2\mu_2(\mathbf{X}_1)\mu_{m-1}(\mathbf{X}_1)^2.$$

Using the asymptotic result in relation (4.19) and Lemma 4.2.2 the next asymptotic result for the mean squared error of the sample mth central moment estimator can now be easily verified.

Lemma 4.2.4. *It follows for any normal population that*

$$\lim_{n\uparrow\infty} n\mathit{MSE}(\widehat{\mu}_{m,n}(\mathbf{X})) = \alpha_m. \tag{4.21}$$

with α_m defined in relation (4.20).

Proof. We know

$$\mathrm{MSE}(\widehat{\mu}_{m,n}(\mathbf{X})) = Var(\widehat{\mu}_{m,n}(\mathbf{X})) + (\mathbb{E}(\widehat{\mu}_{m,n}(\mathbf{X})) - \mu_m(\mathbf{X}_1))^2. \tag{4.22}$$

By Lemma 4.2.2 we obtain

$$n(\mathbb{E}(\widehat{\mu}_{m,n}(\mathbf{X})) - \mu_m(\mathbf{X}_1)) = n\left[(1-n^{-1})^{\frac{m}{2}} - 1\right]\mu_m(\mathbf{X}_1)$$

and this shows

$$\lim_{n\uparrow\infty} n(\mathbb{E}(\widehat{\mu}_{m,n}(\mathbf{X})) - \mu_m(\mathbf{X}_1)) = -\frac{m}{2}.$$

Applying now relation (4.22) and (4.19) yields the desired result. □

An alternative maximum likelihood based estimator for the mth moment is given by the following. Clearly we assume for this estimator that $\mu \neq 0$. For $\mu = 0$ the mth moment estimation problem reduces to the mth central moment estimation problem to be discussed in the last part of this section. It is well known (cf.[5]) for a normal population that the maximum likelihood estimator of μ is given by the sample mean $\overline{\mathbf{X}}_n$ and the maximum likelihood estimator of σ^2 is given by $\widehat{\mu}_{2,n}(\mathbf{X})$. This implies by the invariance property of maximum likelihood estimators (cf.[31]) and Lemma A.0.9 that we may also use the (maximum likelihood) estimator

$$\widehat{\mu}'_{m,ML,n}(\mathbf{X}) = m! \sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \frac{\widehat{\mu}^k_{2,n}(\mathbf{X}) \overline{\mathbf{X}}_n{}^{m-2k}}{(m-2k)! 2^k k!} \tag{4.23}$$

for estimating the mth moment of the normal distribution. For the maximum likelihood estimator of the mth moment the following result holds.

Lemma 4.2.5. *It follows for every $m \in \mathbb{N}$ and $n \in \mathbb{N}, n \geq 2$ that*

$$\mathbb{E}(\widehat{\mu}'_{m,ML,n}(\mathbf{X})) = m! \sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \frac{C_{k,n} \sigma^{2k} \mathbb{E}(\overline{\mathbf{X}}_n^{m-2k})}{(m-2k)! 2^k k!} \tag{4.24}$$

with

$$C_{k,n} := \begin{cases} 1 & \text{if } k = 0 \\ \Pi_{j=0}^{k-1} (1 + \frac{2j-1}{n}) & \text{if } k \in \mathbb{N} \end{cases} \tag{4.25}$$

and $\mathbb{E}(\overline{\mathbf{X}}_n^{m-2k})$ given in relation (A.21).

Proof. It is shown in [32] that a sample arises from a normal population if and only if $\widehat{\mu}_{2,n}(\mathbf{X})$ and the sample mean $\overline{\mathbf{X}}_n$ are independent. This implies by relation (4.23) and the well known properties of expectations for products of independent random variables that

$$\mathbb{E}(\widehat{\mu}'_{m,ML,n}(\mathbf{X})) = m! \sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \frac{\mathbb{E}(\widehat{\mu}^k_{2,n}(\mathbf{X})) \mathbb{E}(\overline{\mathbf{X}}_n{}^{m-2k})}{(m-2k)! 2^k k!}. \tag{4.26}$$

To evaluate $\mathbb{E}(\widehat{\mu}^k_{2,n}(\mathbf{X}))$ in relation (4.26) we observe (cf.[5]) that

$$n \widehat{\mu}_{2,n}(\mathbf{X}) \stackrel{d}{=} \sigma^2 \mathbf{W}_n$$

with $\mathbf{W}_n$ having a Gamma distribution with parameter $\alpha = \frac{n-1}{2}$ and $\beta = 2$. This shows

$$\mathbb{E}(\widehat{\mu}_{2,n}^{k}(\mathbf{X})) = \frac{\sigma^{2k}}{n^k}\mathbb{E}(\mathbf{W}_n^k). \tag{4.27}$$

Since for any $k \in \mathbb{N}$ it follows for $\mathbf{W}_n$ having a Gamma distribution with parameter $\alpha = \frac{n-1}{2}$ and $\beta = 2$ that

$$\mathbb{E}(\mathbf{W}_n^k) = 2^k \frac{\Gamma(k+\frac{n-1}{2})}{\Gamma(\frac{n-1}{2})} = \Pi_{j=0}^{k-1}(2j-1+n)$$

we obtain by relation (4.27) that

$$\mathbb{E}(\widehat{\mu}_{2,n}^{k}(\mathbf{X})) = C_{k,n}\sigma^{2k}. \tag{4.28}$$

Hence we have verified relation (4.24). □

In the next result we discuss the asymptotic behavior of the expectation of the maximum likelihood estimator of the mth moment as the sample size goes to infinity.

Lemma 4.2.6. *It follows for any normal population satisfying $\mu \neq 0$ that for every $m \in \mathbb{N}, m \geq 2$*

$$\lim_{n\uparrow\infty} n(\mathbb{E}(\widehat{\mu}'_{m,ML,n}(\mathbf{X})) - \mu'_m(\mathbf{X}_1)) = m!\beta_m \tag{4.29}$$

with

$$\beta_m := \sum_{k=1}^{\lfloor \frac{m}{2} \rfloor} \frac{\sigma^{2k}\mu^{m-2k}}{2^k(k-1)!}\frac{k}{(m-2k)!} \tag{4.30}$$

Proof. Applying Lemma 4.2.5 and A.0.10 we obtain

$$n(\mathbb{E}(\widehat{\mu}'_{m,ML,n}(\mathbf{X})) - \mu'_m(\mathbf{X}_1)) = I_1(m,n) + I_2(m,n) \tag{4.31}$$

with

$$I_1(m,n) := m!\sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \frac{n(C_{k,n}-1)\sigma^{2k}\mathbb{E}(\overline{\mathbf{X}}_n^{m-2k})}{(m-2k)!2^k k!} \tag{4.32}$$

and

$$I_2(m,n) := m!\sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \sigma^{2k}\frac{n(\mathbb{E}(\overline{\mathbf{X}}_n^{m-2k}) - \mu^{m-2k})}{(m-2k)!2^k k!} \tag{4.33}$$

To analyse the asymptotic behavior of $I_1(m,n)$ we introduce for every $k \in \mathbb{N}$ the function $f_k : [0,1] \to \mathbb{R}$ given by

$$f_k(x) = \Pi_{j=0}^{k-1}(1 + x(2j-1)) \tag{4.34}$$

Introducing the related function $g_k(x) = \ln f_k(x)$ it follows that

$$g_k'(x) = \sum_{j=0}^{k-1} \frac{2j-1}{1 + x(2j-1)}$$

and this shows for every $0 < x < 1$

$$f_k'(x) = g_k'(x) f_k(x). \tag{4.35}$$

Clearly by relation (4.25) and (4.35) we now obtain

$$\lim_{n\uparrow\infty} n(C_{k,n} - 1) = f_k'(0) = \sum_{j=0}^{k-1}(2j-1) = k^2 - 2k \tag{4.36}$$

and this implies using relation (4.32) and $\lim_{n\uparrow\infty} \mathbb{E}(\overline{\mathbf{X}}_n^{m-2k}) = \mu^{m-2k}$ that

$$\lim_{n\uparrow\infty} I_1(m,n) = m! \sum_{k=1}^{\lfloor \frac{m}{2} \rfloor} \frac{(k-1)\sigma^{2k}\mu^{m-2k}}{(k-1)!2^k(m-2k)!}$$

(Observe empty summation is by definition 0). To verify the asymptotic behaviour of $I_2(m,n)$ it follows by relation (A.23) that

$$\begin{aligned}
\lim_{n\uparrow\infty} I_2(m,n) &= m! \sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \sigma^{2k} \frac{(m-2k)(m-2k-1)\mu^{m-2k-2}\sigma^2}{(m-2k)!2^{k+1}k!} \\
&= m! \sum_{k=0}^{\lfloor \frac{m}{2} \rfloor - 1} \frac{\mu^{m-2k-2}\sigma^{2k+2}}{(m-2k-2)!2^{k+1}k!} \\
&= m! \sum_{k=1}^{\lfloor \frac{m}{2} \rfloor} \frac{\mu^{m-2k}\sigma^{2k}}{(m-2k)!2^k(k-1)!}
\end{aligned}$$

and this shows by relation (4.31) the desired result. □

In the next result we will compute the variance of the maximum likelihood estimator for the $2m$th moment. We will only consider this case. A similar formula can be shown for the $(2m+1)$th moment. Before mentioning this result we introduce the matrices

$$A_n = (a_{ij}^{(n)}), B_n = (b_{ij}^{(n)})$$

given by

$$a_{ij}^{(n)} := \sigma^{2(i+j)} C_{i+j,n} \mathbb{E}\left(\overline{\mathbf{X}}_n^{4m-2(i+j)}\right), i,j = 0,...,m \tag{4.37}$$

and

$$(4.38) \qquad b_{ij}^{(n)} := \sigma^{2(i+j)} C_{i,n} C_{j,n} \mathbb{E}\left(\overline{\mathbf{X}}_n^{2m-2j}\right) \mathbb{E}\left(\overline{\mathbf{X}}_n^{2m-2i}\right), i,j = 0,...,m$$

By a standard application of the variance formula for a sum of random variables one can show the following result.

Lemma 4.2.7. *If the vector* $\mathbf{x}$ *is given by* $\mathbf{x}^\top = (x_0,...,x_m), x_k = \frac{(2m)!}{(2m-2k)!k!2^k}, k = 0,...,m$, *then*

$$(4.39) \qquad Var(\widehat{\mu}'_{2m,ML,n}(\mathbf{X})) = \mathbf{x}^\top A_n \mathbf{x} - \mathbf{x}^\top B_n \mathbf{x}$$

Proof. By relation (4.23) we conclude

$$\mathrm{Var}(\widehat{\mu}'_{2m,ML,n}(\mathbf{X})) = \mathrm{Var}\left(\sum_{k=0}^m x_k \widehat{\mu}_{2,n}^k(\mathbf{X}) \overline{\mathbf{X}}_n^{2(m-k)}\right) = \mathbf{x}^\intercal \Lambda_n \mathbf{x}$$

with $\Lambda_n = (\lambda_{ij}^{(n)})_{i,j=0}^m$ and

$$\lambda_{ij}^{(n)} := \mathrm{Cov}\left(\widehat{\mu}_{2,n}^i(\mathbf{X}) \overline{\mathbf{X}}_n^{2(m-i)}, \widehat{\mu}_{2,n}^j(\mathbf{X}) \overline{\mathbf{X}}_n^{2(m-j)}\right)$$

To evaluate these covariances we observe by the independence of the random variables $\widehat{\mu}_{2,n}(\mathbf{X})$ and $\overline{\mathbf{X}}_n$ (cf.[32]) that

$$\begin{aligned} \sigma_{ij}^{(n)} &= \mathbb{E}\left(\widehat{\mu}_{2,n}^{i+j}(\mathbf{X}) \overline{\mathbf{X}}_n^{4m-2(i+j)}\right) - \mathbb{E}\left(\widehat{\mu}_{2,n}^i(\mathbf{X}) \overline{\mathbf{X}}_n^{2(m-i)}\right) \mathbb{E}\left(\widehat{\mu}_{2,n}^j(\mathbf{X}) \overline{\mathbf{X}}_n^{2(m-j)}\right) \\ &= \mathbb{E}\left(\widehat{\mu}_{2,n}^{i+j}(\mathbf{X})\right) \mathbb{E}\left(\overline{\mathbf{X}}_n^{4m-2(i+j)}\right) - \Pi_{k\in\{i,j\}} \mathbb{E}\left(\widehat{\mu}_{2,n}^k(\mathbf{X})\right) \Pi_{k\in\{i,j\}} \mathbb{E}\left(\overline{\mathbf{X}}_n^{2(m-k)}\right). \end{aligned}$$

This shows by relation (4.28) the desired result. □

Since by Lemma 4.2.5 and lemma 4.2.7 one can evaluate the mean and the variance of the maximum likelihood estimator $\widehat{\mu}'_{2m,ML,n}(\mathbf{X})$ of the mth moment one can also compute the mean squared error of the estimator $\widehat{\mu}'_{m,ML,n}(\mathbf{X})$ given by

$$(4.40) \qquad \mathrm{MSE}(\widehat{\mu}'_{m,ML,n}(\mathbf{X})) = \mathrm{Var}(\widehat{\mu}'_{m,ML,n}(\mathbf{X})) + \left(\mathbb{E}(\widehat{\mu}'_{m,ML,n}(\mathbf{X})) - \mu'_m(\mathbf{X})\right)^2.$$

In the next result we derive an asymptotic formula for the mean squared error of the maximum likelihood estimator of the mth moment. Before proving this result we need the following auxiliary limit result.

Lemma 4.2.8. *It follows for $\mu \neq 0$ and $k_1, k_2 \in \mathbb{Z}_+$ and $p_1, p_2 \in \mathbb{Z}_+$ that*

$$\lim_{n\uparrow\infty} n\left(\Pi_{i=1}^2 C_{k_i,n}\mathbb{E}(\overline{\mathbf{X}}_n^{p_i}) - \mu^{p_1+p_2}\right) = \gamma \tag{4.41}$$

with

$$\gamma := \mu^{p_1+p_2}\sum_{i=1}^2 (k_i^2 - 2k_i) + \mu^{p_1+p_2-2}\sigma^2 \sum_{i=1}^2 \frac{\left(p_i^2 - 2p_i\right)}{2} \tag{4.42}$$

Proof. We observe

$$\begin{aligned} & n(\Pi_{i=1}^2 C_{k_i,n}\mathbb{E}(\overline{\mathbf{X}}_n^{p_i}) - \mu^{p_1+p_2}) \\ = \quad & n(\Pi_{i=1}^2 C_{k_i,n} - 1)\Pi_{i=1}^2\mathbb{E}(\overline{\mathbf{X}}_n^{p_i}) + n(\Pi_{i=1}^2\mathbb{E}(\overline{\mathbf{X}}_n^{p_i}) - \mu^{p_1+p_2}). \end{aligned} \tag{4.43}$$

Since $\lim_{n\uparrow\infty} n(\Pi_{i=1}^2 C_{k_i,n} - 1) = (f_{k_1} f_{k_2})'(0)$ with f_k defined in relation (4.34) we obtain

$$\lim_{n\uparrow\infty} n(\Pi_{i=1}^2 C_{k_i,n} - 1)\Pi_{i=1}^2\mathbb{E}(\overline{\mathbf{X}}_n^{p_i}) = \mu^{p_1+p_2}\sum_{i=1}^2 (k_i^2 - 2k_i) \tag{4.44}$$

Also it follows using

$$n(\Pi_{i=1}^2\mathbb{E}(\overline{\mathbf{X}}_n^{p_i}) - \mu^{p_1+p_2}) = \mathbb{E}(\overline{\mathbf{X}}_n^{p_1})n(\mathbb{E}(\overline{\mathbf{X}}_n^{p_2}) - \mu^{p_2}) + n\mu^{p_2}(\mathbb{E}(\overline{\mathbf{X}}_n^{p_1}) - \mu^{p_1})$$

and applying relation (A.23) that

$$\lim_{n\uparrow\infty} n(\Pi_{i=1}^2\mathbb{E}(\overline{\mathbf{X}}_n^{p_i}) - \mu^{p_1+p_2}) = \mu^{p_1+p_2-2}\sigma^2\sum_{i=1}^2 \frac{\left(p_i^2 - 2p_i\right)}{2}$$

This shows using relation (A.13) the limit result in relation (4.41). □

Since $C_{0,n} = 1$ and $\mathbb{E}(\overline{\mathbf{X}}_n^0) = 1$ we obtain by Lemma 4.2.8 that

$$\lim_{n\uparrow\infty} n(C_{k,n}\mathbb{E}(\overline{\mathbf{X}}_n^p) - \mu^p) = (k^2 - 2k)\mu^p + \frac{(p^2 - p)}{2}\mu^{p-2}\sigma^2. \tag{4.45}$$

Also by Lemma 4.2.8 we obtain for every $k_1, k_2 \in \mathbb{N}$ and $p_1, p_2 \in \mathbb{N}$ and $\mu \neq 0$ that

$$\lim_{n\uparrow\infty}(C_{k_1+k_2,n}\mathbb{E}(\overline{\mathbf{X}}_n^{p_1+p_2}) - \Pi_{i=1}^2 C_{k_i}\mathbb{E}(\overline{\mathbf{X}}_n^{p_i})) = 2\mu^{p_1+p_2}k_1k_2 + \mu^{p_1+p_2-2}\sigma^2 p_1p_2 \tag{4.46}$$

An implication of relation 4.46 and Lemma 4.2.7 is given by the following asymptotic result for the mean squared error of the maximum likelihood estimator of the mth moment for $\mu \neq 0$.

Lemma 4.2.9. *It follows for any normal population satisfying $\mu \neq 0$ and $m \in \mathbb{N}, m \geq$*

2 *that*

$$\lim_{n\uparrow\infty} nMSE(\widehat{\mu}'_{2m,ML,n}(\mathbf{X})) = \mathbf{x}^\top \Delta_\infty \mathbf{x}$$

with $\Delta_\infty = (\delta_{ij}^{(\infty)})_{i,j=0}^m$ *given by*

$$\delta_{ij}^{(\infty)} := 2\sigma^{2(i+j)}\mu^{4m-2(i+j)}ij + 4\sigma^{2(i+j)+2}\mu^{4m-2(i+j)-2}(m-i)(m-j)$$

and the vector $\mathbf{x}$ *defined in Lemma 4.2.7.*

Proof. Easy application of Lemma 4.2.8 and 4.2.7 replacing k_i by i and p_i by $2(m-i)$ in relation (4.46). □

In the computational section we show for the scenario $\mu = 3, \sigma = 4$ in Figure 4.2 in a graph the functions $f : \mathbb{N} \to \mathbb{R}$ and $g : \mathbb{N} \to \mathbb{R}$ given by

$$f(n) = nMSE\left(\widehat{\mu}'_{m,ML,n}(\mathbf{X})\right), g(n) = nMSE\left(\widehat{\mu}'_{m,n}(\mathbf{X})\right)$$

Clearly this graph shows for which values of n the maximum likelihood estimator of the mth moment is more efficient than the sample mth moment estimator.

By a similar argument using relation (A.20) we may use the maximum likelihood estimator

$$\widehat{\mu}_{2m,ML,n}(\mathbf{X}) = \frac{(2m)!\widehat{\mu}_{2,n}^m(\mathbf{X})}{m!2^m} \tag{4.47}$$

for estimating the $(2m)$th central moment of the normal distribution. Observe the odd central moments are all zero and so we do not need to estimate them. Also for this estimator we obtain applying relation (4.6)

$$\widehat{\mu}_{m,ML,n}(\mathbf{X}) \stackrel{d}{=} \sigma^m \widehat{\mu}_{m,ML,n}(\mathbf{Y})$$

with $\mathbf{Y} = (\mathbf{Y}_1, ...\mathbf{Y}_n)$ and $\mathbf{Y}_i, i = 1, ..., n$ independent and standard normal distributed. This shows

$$\mathbb{E}(\widehat{\mu}_{m,ML,n}(\mathbf{X})) = \sigma^m \mathbb{E}(\widehat{\mu}_{m,ML,n}(\mathbf{Y})), \mathrm{Var}(\widehat{\mu}_{m,ML,n}(\mathbf{X})) = \sigma^{2m}\mathrm{Var}(\widehat{\mu}_{m,ML,n}(\mathbf{Y})) \tag{4.48}$$

and

$$\mathrm{MSE}(\widehat{\mu}_{m,ML,n}(\mathbf{X})) = \sigma^{2m} MSE(\widehat{\mu}_{m,ML,n}(\mathbf{Y})) \tag{4.49}$$

We now show the following result for the maximum likelihood estimator $\widehat{\mu}_{2m,ML}(\mathbf{X})$ for m any positive integer. Next to the value of the variance of this estimator it shows that the estimator $C_{m,n}^{-1}\widehat{\mu}_{2m,ML,n}(\mathbf{X})$ with $C_{m,n}$ given in relation (4.25) is an unbiased estimator of the $(2m)$th central moment of the normal distribution.

Lemma 4.2.10. *It follows for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$ that*

(4.50)
$$\mathbb{E}(\widehat{\mu}_{2m,ML,n}(\mathbf{X})) = \mu_{2m}(\mathbf{X}_1)C_{m,n},\ Var(\widehat{\mu}_{2m,ML,n}(\mathbf{X})) = \mu_{2m}(\mathbf{X}_1)^2(C_{2m,n} - C_{m,n}^2)$$

and

$$\begin{aligned} MSE(\widehat{\mu}_{2m,ML,n}(\mathbf{X})) &= \mu_{2m}(\mathbf{X}_1)^2(1-2C_{m,n}+C_{2m,n}) \\ &= \sigma^{4m}\mu_{2m}(\mathbf{Z})^2(1-2C_{m,n}+C_{2m,n}) \end{aligned} \tag{4.51}$$

with $\mathbf{Z}$ having a standard normal distribution and $C_{k,n}$ given in relation (4.25).

Proof. Clearly by relation (4.47) we obtain

$$\mathbb{E}(\widehat{\mu}_{2m,ML,n}(\mathbf{X})) = \frac{(2m)!}{2^m m!}\mathbb{E}(\widehat{\mu}_{2,n}^m(\mathbf{X})), \mathrm{Var}(\widehat{\mu}_{2m,ML,n}) = \left(\frac{2m!}{2^m m!}\right)^2 \mathrm{Var}\left(\widehat{\mu}_{2,n}^m\right).$$

This shows by Lemma A.0.9 and relation (4.28) the first formula. Since $\mathrm{Var}\left(\widehat{\mu}_2^m\right) = \mathbb{E}(\widehat{\mu}_2^{2m}) - \mathbb{E}(\widehat{\mu}_2^m)^2$ the second part can be verified in a similar way. The relation for the mean squared error follows easily from relation (4.50). □

An immediate consequence of the above lemma and relation (4.36) is given by the following asymptotic result for the mean squared error of the maximum likelihood estimator of the mth central moment.

Lemma 4.2.11. *It follows for any normal population*

$$\lim_{n\uparrow\infty} n MSE(\widehat{\mu}_{2m,ML,n}(\mathbf{X})) = 2m^2\mu_{2m}(\mathbf{X}_1)^2. \tag{4.52}$$

Proof. Easy application of relation (4.36) and (4.51). □

By relation (4.49) and (4.8) it follows that the ratio of the mean squared error of the maximum likelihood estimator and the mean squared error of the sample moment estimator is independent of σ. Hence for all normal populations with unknown μ and σ it is possible to determine analytically a threshold value $n_*(m)$ for the sample size. If the sample size is above this threshold value the maximum likelihood estimator of the mth central moment has a lower mean squared error than the mean squared error of the sample mth central moment estimator and if it is below the inequality is reversed. Clearly this threshold value of the sample size depends on m. In Figure 4.1 we have plotted the graph of the function $n \mapsto n\mathrm{MSE}\left(\widehat{\mu}_{m,ML,n}(\mathbf{X})\right)$ and

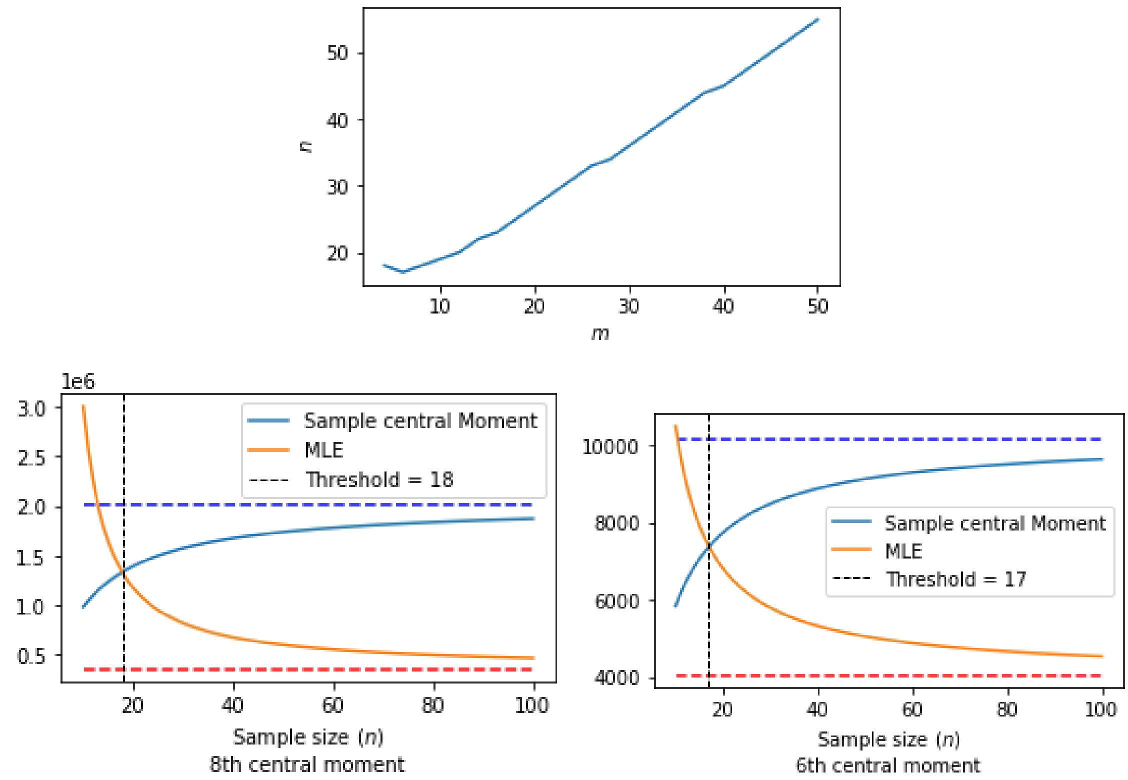

Figure 4.1 graph of the function $n \mapsto n_*(m)$ and $n \mapsto n\text{MSE}\left(\widehat{\mu}_{m,ML,n}(\mathbf{X})\right)$ and $n \mapsto n\text{MSE}\left(\widehat{\mu}_{m,n}(\mathbf{X})\right)$ for $m = 6, 8$ and $\sigma = 1$.

$n \mapsto n\text{MSE}\left(\widehat{\mu}_{m,n}(\mathbf{X})\right)$ for the base scenario $\sigma = 1$. Clearly the dotted line represent this threshold value for $m = 6$ and $m = 8$. At the same time we have plotted in the same figure the function $m \mapsto n_*(m)$. As observed from this figure the threshold value seems to be increasing in m. This means that for higher values of m and a given sample size the use of the sample mth central moment estimator becomes more attractive. Observe for any normal population it is shown in this figure for $m = 6$ and $m = 8$ that the threshold value equals 18 while for $m = 50$ it is given by 54. Also the intersection point of the blue line with the Y-axis equals the asymptotic value α_m given in Lemma 4.2.4, while the intersection point of the red line with the Y-axis represent the asymptotic first order term of the mean squared error of the maximum likelihood estimator computed in Lemma 4.2.11. .

Since computing the maximum likelihood or the sample mth (central) moment estimator for a given sample is in general very fast it is good to know applying the MSE objective for which sample sizes one estimator is preferred above the other. Clearly for $m = 2$ both estimators coincide. As we have shown the sample mth moment estimator is always unbiased, while the maximum likelihood estimator $\widehat{\mu}'_{m,ML}(\mathbf{X})$ is biased. For the mth central moment both estimators are biased but asymptotically unbiased. On the other hand, the celebrated maximum likelihood principle yields in theory (cf.[31]) for large sample sizes high quality estimations with the smallest asymptotic variance. As already observed for the mth central moment ($m \geq 3$) there is a threshold value sample size separating both estimators if one uses the mean squared error objective. In the next section we will by means of simulation experiments compare the efficiency of both estimators with respect to this objective. Since for estimating moments a possible threshold value also depends on the unknown σ (and so this threshold problem for estimating moments is less clear than for estimating central moments) we will focus more attention in the next section on the estimation of moments. However, we still report our simulation results for central moments. At the same time we will show in the next section the histograms of the maximum likelihood and sample moment estimator for both the moment and central moment estimation problem.

4.3 Computational results.

In this section we compare the behaviour of the sample moment and the maximum likelihood estimator of the mth moment and the mth central moment, $m = 4,6,8$, using simulation. Since for $m = 2$ both estimators are the same we do not report the results for this case. For each n we performed 10.000 simulation runs. We only show in detail the results for a normal distribution having parameters $\mu = 3$ and $\sigma = 4$. Similar experiments for the 9 different scenarios $(\mu,\sigma), \mu \in \{0,10,100\}$ and $\sigma \in \{1,10,100\}$ were also conducted, but since the conclusions about the performance of both estimators are similar we do not report these results. In the first subsection we report the simulation results for the moment estimation problem, while in the second subsection we report the results for the central moment estimation problem.

4.3.1 Computational results for moments

To evaluate for a given sample the realisation of the sample moment estimator of the mth moment we apply relation (4.2) and relation (4.23) for the maximum likelihood estimator. For a normally distributed random variable $\mathbf{X}_1$ with $\mu = 3$ and $\sigma = 4$ it follows by Lemma A.0.9 that

$$\mu_8'(\mathbf{X}_1) = 27051873, \mu_6'(\mathbf{X}_1) = 185289, \mu_4'(\mathbf{X}_1) = 1713 \tag{4.53}$$

The average over 10.000 runs of the realisations of the sample moment and maximum likelihood estimator of the mth moment (rounded down to integers) in a normal population with $\mu = 3$ and $\sigma = 4$ are provided in the second and fifth row of Table 4.1. In the first row the size of the sample is reported. The mean squared error ratio shown in the seventh row of Table 4.1 denotes the ratio of the mean squared error of the maximum likelihood estimator over the mean squared error of the sample moment estimator. Both mean squared errors were calculated using the formulas in relation (4.9) and (4.40). If in this row the ratio is smaller than one the maximum likelihood estimator has a lower mean squared error then the sample moment estimator for the given sample size n. In the eight row we list for the biased maximum likelihood estimator the bias divided by the mth moment again for the given sample size n. These are also calculated using relations (4.9) and (4.40). Finally in the fourth and sixth row of Table 4.1 we calculate (rounded down to integers) the sample standard deviations of both estimators.

As expected the variance of both estimators become smaller as the sample size increases. However, the variance of the maximum likelihood estimator is highly nonlinear in the reciprocal of the sample size and this shows for small sample sizes ($n = 10$) that its variance is higher or comparable with the variance of the sample moment estimator. Adding the unbiasedness of the maximum likelihood estimator for small sample sizes listed in the eighth row of Table 4.1 this implies that for $n = 10$ the maximum likelihood estimator has a higher or comparable mean squared error than the mean squared error of the unbiased sample moment estimator. The ratio increases in favour of the sample moment estimator as m increases and so for $m = 6, 8$ and $n = 10$ one should use the sample moment estimator. Due to the same nonlinear behaviour in the reciprocal of the sample size the variance of the maximum likelihood estimator yields for larger sample sizes a much lower variance than the variance of the sample moment estimator (linear in the reciprocal of the sample size). Combining this with the improving behaviour of the bias of the maximum likelihood estimator as the sample size increases (see eighth row of Table 4.1) the mean squared error of the maximum likelihood estimator becomes much smaller than the mean squared error of the sample moment estimator. Since the density of the maximum likelihood estimator has a smaller kurtosis than the density of the sample moment estimator for $n = 10.000$ this improvement of the maximum likelihood estimator can also be observed from the histograms in Figure 4.3. At the same time the almost identical mean square error of both the maximum likelihood

		MLE		sample moment estimator		MSE ratio	$\frac{bias_{MLE}}{\mu'_m(\mathbf{X}_1)}$
	n	sample mean	sample std. dev.	sample mean	sample std. dev.		
m=8	10000	27074669	1341236	27066457	2846382	0.2313	0.0006
	1000	27182461	4348205	26944263	8855927	0.2360	0.0068
	100	29031519	15308445	27281404	28474737	0.2865	0.0689
	10	48059570	102244804	27074969	89459298	1.3065	0.7629
m=6	10000	185368	6919	185325	10188	0.4649	0.0003
	1000	185565	22281	184791	32173	0.4690	0.0031
	100	191428	74401	186200	103919	0.5113	0.0310
	10	241452	333596	184009	331398	1.0412	0.3036
m=4	10000	1713	43	1713	48	0.7882	0
	1000	1712	138	1710	154	0.7900	0.0008
	100	1729	447	1715	499	0.8075	0.0088
	10	1843	1552	1700	1579	0.9771	0.0807

Table 4.1 Estimation of mth moment in a normal population with parameters $\mu = 3$ and $\sigma = 4$ for different values of the sample size and m.

estimator and the sample moment estimator for $n = 10$ and $m = 4$ can be observed from Figure 4.4.

If the value of m increases we observe from Table 4.1 that the estimates of both estimators become less accurate. This can also be observed from Figure 4.3 considering the histograms of the different estimators for sample size $n = 10000$ and different values of m. From Table 4.1 we notice that for larger values of m the mean squared error ratio decreases. This means for larger values of m that the maximum likelihood estimator is more efficient than the sample moment estimator for a given sample size. All these observations imply that one should use for all considered values of m the maximum likelihood estimator if the sample size exceeds a certain threshold value. In our particular chosen scenario of $\mu = 3$ and $\sigma = 4$ we show these threshold values in Figure 4.2 for $m = 4, 6, 8$.

However, contrary to estimating the mth central moment (the threshold value only depends on m!), the threshold value of the sample size for estimating moments also depends on the unknown σ and μ. In this case a rule of thumb to choose between the two different estimators could be the following: Estimate beforehand for a given sample the unknown μ and σ and for these estimated parameters compute the mean squared error of both estimators. If the estimated mean squared error of the maximum likelihood estimator is smaller than the mean squared error of the sample moment estimator use the maximum likelihood estimator and otherwise the sample moment estimator.

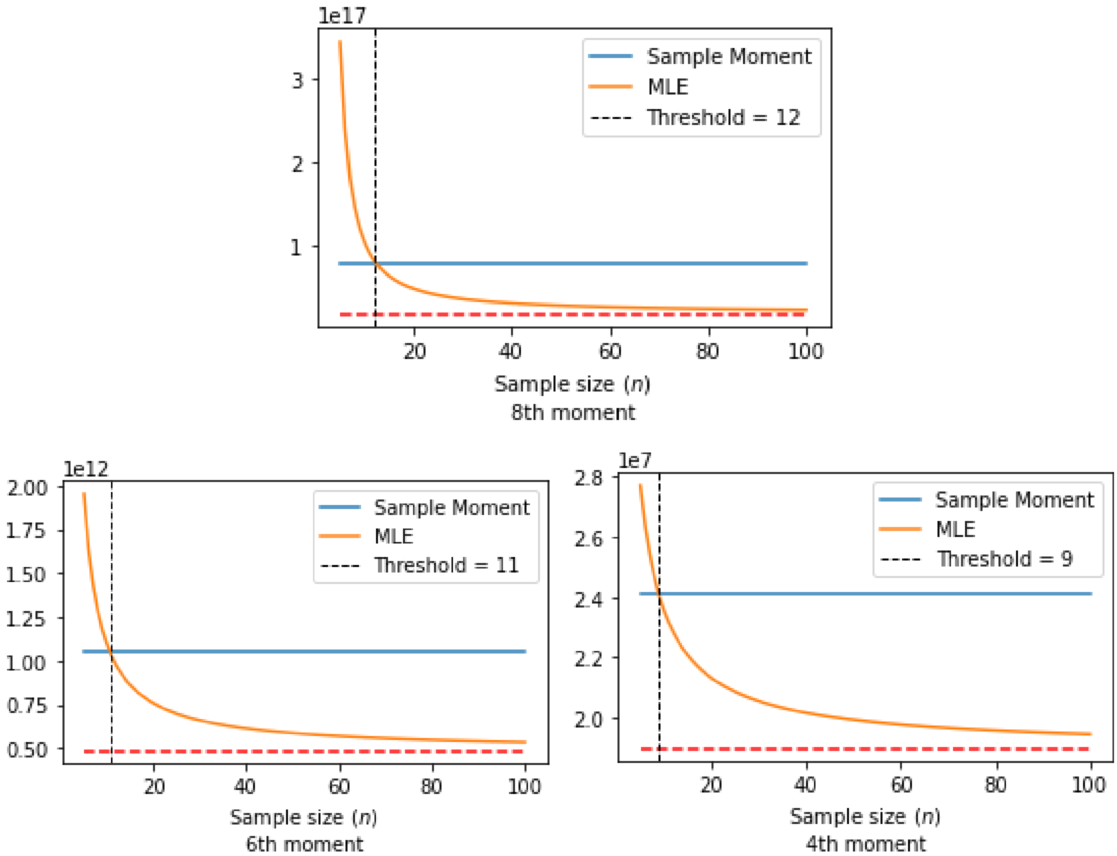

Figure 4.2 graph of the function $n \mapsto n\mathrm{MSE}\left(\widehat{\mu}'_{m,ML,n}(\mathbf{X})\right)$ and $n \mapsto n\mathrm{MSE}(\widehat{\mu}'_{m,n}(\mathbf{X}))$ in a normal population having parameters $\mu = 3$ and $\sigma = 4$.

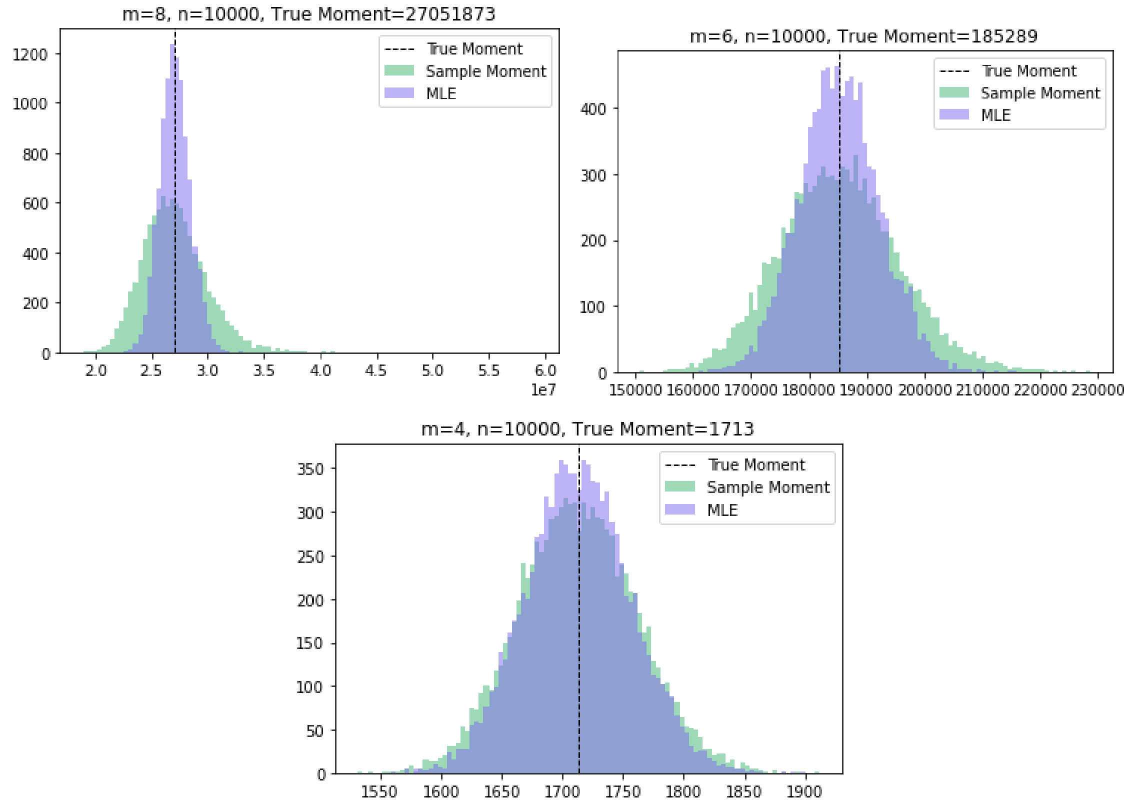

Figure 4.3 Histogram of the density of the maximum likelihood estimator and the sample mth moment estimator, $m = 2, 4, 6$, for a normal population with parameters $\mu = 3$ and $\sigma = 4$ and sample size $n = 10.000$.

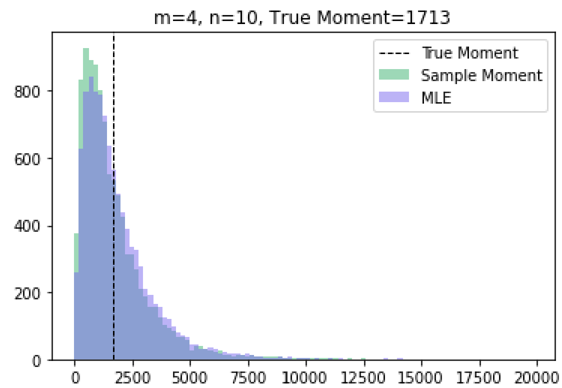

Figure 4.4 Histogram of the density of the sample moment estimator and the maximum likelihood estimator of the 4th moment for a normal population with parameters $\mu = 3$ and $\sigma = 4$ and sample size $n = 10$

		MLE		sample central moment estimator		MSE ratio	$\frac{bias_{MLE}}{\mu_m(\mathbf{X}_1)}$
	n	sample mean	sample std. dev.	sample mean	sample std. dev.		
m=8	10000	6888682	385577	6882680	942069	0.1756	0.0008
	1000	6929961	1258024	6835907	3093529	0.1813	0.0080
	100	7489963	4511146	6657733	8713352	0.2470	0.0813
	10	13601159	42395693	4514047	19616118	3.0639	0.9305
m=6	10000	61471	2580	61443	4104	0.3989	0.0002
	1000	61578	8347	61118	13163	0.4050	0.0029
	100	63533	28036	59920	40303	0.4704	0.0298
	10	79598	141189	44419	99696	1.7980	0.2870
m=4	10000	768	21	768	24	0.7503	0
	1000	767	69	765	79	0.7539	0
	100	769	223	754	250	0.7903	-0.0001
	10	760	766	618	675	1.2531	-0.0099

Table 4.2 Estimation of mth central moment in a normal population with parameters $\mu = 3$ and $\sigma = 4$ for different values of the sample size and m.

4.3.2 Computational results for central moments

To calculate the mth central moment we use relation (4.3) for the sample central moment estimator and relation (4.47) for the maximum likelihood estimator. For a normally distributed random variable $\mathbf{X}_1$ with $\mu = 3$ and $\sigma = 4$ it follows by relation (A.20) that

$$\mu_8(\mathbf{X}_1) = 6881280, \mu_6(\mathbf{X}_1) = 61440, \mu_4(\mathbf{X}_1) = 768. \tag{4.54}$$

In Table 4.2 it is shown that the bias of the maximum likelihood estimator of the central moment decreases as the sample size increases. Looking in detail at this table one can draw similar conclusions as done for the case of estimating moments. In the two extreme cases $n = 10$ and $n = 10.000$ this is also shown in Figure 4.5 and Figure 4.6. In this particular case we have also shown at the end of the previous section that one can compute easily a threshold value $n_*(m)$ of the sample size satisfying $n \geq n_*(m)$ if and only if the maximum likelihood estimator should be applied. A graph of these values up to $m = 50$ is given in Figure 4.1.

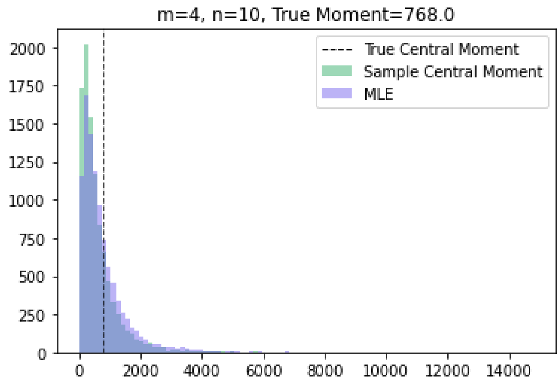

Figure 4.5 Histogram of the density of the maximum likelihood and sample central moment estimator of the 4th central moment in a normal population with parameters $\mu = 3$ and $\sigma = 4$ and sample size $n = 10$

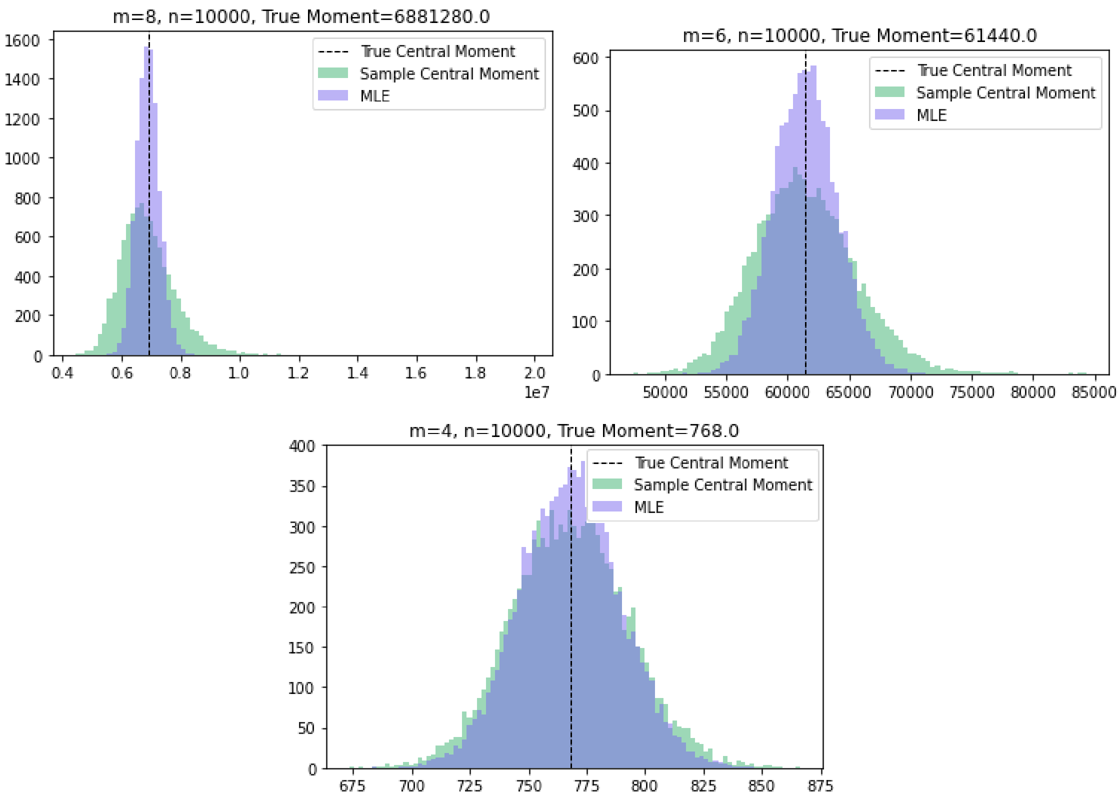

Figure 4.6 Histogram of the density of the the maximum likelihood and sample mth moment estimator of the mth central moment in a normal population with parameters $\mu = 3$ and $\sigma = 4$ and sample size 10.000.

4.4 Conclusion

In this paper we compared maximum likelihood and sample moment estimators for the mth (central) moments in a normal population. Our analytical results as well as simulation results show that the maximum likelihood estimator performs better with respect to the mean square error objective than the sample estimator beyond a certain sample size. In particular this threshold value can be easily computed for estimating central moments. In estimating moments we propose a heuristic approach to compute this threshold value.

BIBLIOGRAPHY

[1] Leonhard Euler. De progressionibus transcendentibus seu quarum termini generales algebraice dari nequeunt. *Commentarii Academiae Scientiarum Petropolitanae*, 5:36–57, 1738.

[2] Pierre Simon Laplace. *Théorie Analytique des Probabilités.* Courcier, Paris, 1812.

[3] Karl Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900.

[4] N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions Vol 1 (second edition).* Wiley, New York, 1994.

[5] George Casella and Roger L. Berger. *Statistical inference (second edition).* Duxbury, Pacific Grove, 2002.

[6] Kenneth Lange. *Optimization.* Springer Verlag, New York, 2013.

[7] Zhi-Shang Ye and Nan Cheng. Closed Form Estimators for the Gamma distibution derived from Likelihood Equations. *The American Statistician*, 71(2):177–181, 2017.

[8] E.W. Stacy. A generalization of the Gamma distribution. *Annals of Mathematical Statistics*, 33(3):1187–1192, 1962.

[9] E.W.Stacy and G.A. Mihram. Parameter Estimation for a Generalzed Gamma distribution. *Technometrics*, 7(3):349–358, 1965.

[10] Van. B. Parr and J.T. Webster. A method for discriminating between failure density functions used in reliablity predictions. *Technometrics*, 7(1):1–10, 1965.

[11] Harold W. Hager and Lee. J. Bain. Inferential procedures for the generalized Gamma distribution. *Journal of the American Statistical Association*, 65(332):1601–1609, 1970.

[12] H. Leon Harter. Maximum likelihood estimation of the paramaters of a four-parameter generalized Gamma population from complete and censored samples. *Technometrics*, 9(1):159–165, 1967.

[13] JR.L. Prentice. A Log-Gamma model and its maximum likelihood estimation. *Biometrika*, 61(3):539–544, 1974.

[14] J.F. Lawless. Inference in the Generalized Gamma and Log-Gamma distributions. *Technometrics.*, 22(3):409–419, 1980.

[15] O.Gomes, C. Combes, and A.Dussauchy. Parameter estimation of the genralized Gamma distribution. *Mathematics and Computers in Simulation*, 79:955–963, 2008.

[16] Man Ho Ling. A comparison of estimation methods for generalized gamma disrtibution with one-shot device testing data. *International Journal of Applied and Experimental Mathematics*, 3:124–131, 2018.

[17] Ping-Huang Huang and Tea-Yuan Hwang. On new moment estimation of parameters of the generalized Gamma distribution using its characterization. *Taiwanse Journal of Mathematics*, 10(4):1083–1093, 2006.

[18] Steven F. Arnold. *Mathematical Statistics.* Prentice Hall, Englewood Cliffs, 1990.

[19] A.Wayne Roberts and Dale.E.Varberg. *Convex Functions.* Academic Press, New York, 1973.

[20] W.Rudin. *Principles of Mathematical Analysis.* McGraw-Hill, New York, 1976.

[21] Ward Cheney and David Kincaid. *Numerical Mathematics and Computing (fourth edition).* Brooks/Cole, Pacific Grove, 1999.

[22] F.Stummel and K.Hainer. *Introduction to Numerical Analysis.* Scottish Academic Press, 1980.

[23] Stephen Boyd and Lieven Vanderberghe. *Convex Optimization.* Cambridge University Press, Cambridge, 2004.

[24] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual.* CreateSpace, Scotts Valley, CA, 2009.

[25] Michael JD Powell. A view of algorithms for optimization without derivatives. *Mathematics Today-Bulletin of the Institute of Mathematics and its Applications*, 43(5):170–174, 2007.

[26] Karl Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London A*, 185:71–110, 1894.

[27] P. Chebycheff. Sur deux théorèmes relatifs aux probabilités. *Acta Mathemetica*, 14:305–315, 1890.

[28] Thomas A Severini. *Elements of distribution theory.* Cambridge University Press, Cambridge, 2005.

[29] E.L. Lehmann. *Elements of Large-Sample Theory.* Springer, Berlin, 2010.

[30] R.J.Serfling. *Approximation Theorems in Statistics.* Wiley, New York, 1980.

[31] S.F.Arnold. *Mathematical Statistics.* Prentice Hall, Englewood Cliffs, 1990.

[32] Eugene Lukacs. A characterization of the normal distribution. *Annals of Mathematical Statistics*, 13(1):91–93, 1942.

[33] E.T.Whittaker and G.N.Watson. *A course in Modern Analysis (Fourth edition)*. Cambridge University Press, Cambridge, 1927.

[34] Zoltan Sasvari. An elementary proof of Binets formula for the Gamma function. *The American Mathematical Monthly*, 106(2):156–158, 1999.

[35] Fred W. Steutel and Klaas van Harn. *Infinite divisibility of Probability distributions on the real line*. Marcel Dekker, New York, 2004.

[36] Horst Alzer. On some inequalities for the gamma and psi functions. *Mathematics of Computation*, 66:337–389, 1997.

[37] Emil Artin. *The Gamma function*. Holt, Rinehard and Winston, New York, 1964.

[38] Necdet Batir. On some properties of digamma and polygamma functions. *Journal of Mathematical Analysis and Applications*, 328:452–465, 2007.

[39] Horst Alzer and Jim Wells. Inequaliteis for the polygamma functions. *SIAM J. Math. Anal.*, 29(6):1459–1466, 1998.

[40] N.L Johnson and S.Kotz. *Continuous Univariate Distributions-1*. Wiley, New York, 1970.

[41] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1990.

APPENDIX

Appendix of on the maximum likelihood approach applied to a (generalized)gamma population

In this appendix we provide for completeness a property of the digamma function useful in the analysis of the maximum likelihood optimization problem for a generalized gamma population. We first mention its definition (cf.[33]).

Definition A.0.1. *The function* $\psi : (0,\infty) \to \mathbb{R}$ *given by*

$$\psi(\alpha) := \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \tag{A.1}$$

with $\Gamma(\alpha) := \int_0^\infty t^{\alpha-1} e^{-\alpha t} dt, \alpha > 0$ *the well known gamma function is called the digamma function.*

To start our analysis we introduce Binets formula of the gamma function (for an elementary proof see [34]) given by

$$\Gamma(\alpha+1) = (\frac{\alpha}{e})^{\alpha} \sqrt[2]{2\pi\alpha} e^{\theta(\alpha)}, \alpha > 0 \tag{A.2}$$

with

$$\theta(\alpha) := \int_0^\infty \left(\frac{1}{e^t - 1} - \frac{1}{t} + \frac{1}{2}\right) \frac{e^{-\alpha t}}{t} dt. \tag{A.3}$$

In the next lemma we analyse in more detail the function θ listed in relation (A.3). Before discussing this result we introduce the following well known class of functions (cf.[35]).

Definition A.0.2. *A function* $f : I \to \mathbb{R}$ *with* I *some open interval is said to be completely monotone if all derivatives of the function* f *exists in* I *and satisfy* $(-1)^m f^{(m)}(\alpha) \geq 0$ *for every* $\alpha \in I$ *and* $m \in \mathbb{Z}_+$. *The function* $f : I \to \mathbb{R}$ *is called strictly completely monotone if* $(-1)^m f^{(m)}(\alpha) > 0$ *for every* $\alpha \in I$ *and* $m \in \mathbb{Z}_+$

Lemma A.0.1. *The function* $F : [0,\infty) \to \mathbb{R}$ *given by*

$$F(t) := \frac{2}{1-e^{-t}} - \frac{2}{t} - 1 \tag{A.4}$$

is a cumulative distribution function satisfying $F(\infty) = 1$ *and* $F(0^+) := \lim_{t \downarrow 0} F(t) =$

0.

Proof. Clearly $F(\infty)=1$. Introducing now the function $\omega:[0,\infty)\to\mathbb{R}$ given by

$$\omega(t)=\frac{1}{1-e^{-t}}-\frac{1}{t} \tag{A.5}$$

we obtain $F(t)=2\omega(t)-1$ and the result is proved showing ω is strictly increasing and $\omega(0^+)=\frac{1}{2}$. To verify this we observe computing the derivative that the function $x\to e^x+e^{-x}$ is a strictly increasing function on $(0,\infty)$. This shows $2<e^x+e^{-x}$ for every $x>0$ and hence for every $t>0$

$$t^2=\int_0^t 2vdv=\int_0^t\int_0^v 2dxdv<\int_0^t e^v-e^{-v}dv=e^t+e^{-t}-2. \tag{A.6}$$

It is easy to check for every $t>0$ that the derivative of the function ω is given by

$$\omega^{(1)}(t)=\frac{1}{t^2}-\frac{e^{-t}}{(1-e^{-t})^2}=\frac{1}{t^2}-\frac{1}{e^t+e^{-t}-2}.$$

and so by relation (A.6) the function ω is strictly increasing. To show $\omega(0^+)=\frac{1}{2}$ we observe

$$\omega(t)=\frac{t-(1-e^{-t})}{t(1-e^{-t})}=\frac{t^{-1}\int_0^t 1-e^{-v}dv}{1-e^{-t}}=\frac{\int_0^1 1-e^{-ut}du}{1-e^{-t}}.$$

This implies

$$\lim_{t\downarrow 0}\omega(t)=\lim_{t\downarrow 0}\frac{\int_0^1 1-e^{-ut}du}{1-e^{-t}}=\lim_{t\downarrow 0}\frac{\int_0^1\frac{1-e^{-ut}}{tu}udu}{\frac{1-e^{-t}}{t}}=\int_0^1 udu=\frac{1}{2}$$

and we have verified the desired properties for ω. □

Applying Lemma A.0.1 the next result follows immediately.

Lemma A.0.2. *The function θ listed in relation (A.3) is strictly completely monotone and has the alternative representation*

$$\theta(\alpha)=\frac{1}{2}\int_0^\infty F(t)\frac{e^{-\alpha t}}{t}dt,\alpha>0$$

with F the cdf listed in relation (A.4).

Proof. Since $(1-e^{-t})^{-1}=1+(e^t-1)^{-1}$ we obtain by relation (A.3) and the definition of the cdf F in relation (A.4) implying $\omega(t)=\frac{1}{2}F(t)+\frac{1}{2}$ that

$$\theta(\alpha)=\int_0^\infty\left(\omega(t)-\frac{1}{2}\right)\frac{e^{-\alpha t}}{t}dt=\frac{1}{2}\int_0^\infty F(t)\frac{e^{-\alpha t}}{t}dt.$$

Applying the monotone convergence theorem we obtain by induction for every $m \in \mathbb{Z}_+$ that

$$(-1)^m \theta^{(m)}(\alpha) = \frac{(-1)^{2m}}{2} \int_0^\infty F(t) t^{m-1} e^{-\alpha t} dt$$

and this shows the result. □

Following the same approach as in [36] it is easy to derive the following representation for the digamma function ψ. This representation is useful in determining the asymptotic behaviour of the function ψ for $\alpha \uparrow \infty$ and for $\alpha \downarrow 0$.

Lemma A.0.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space and $\mathbf{X}$ a random variable having cdf F listed in relation (A.4). Then it follows for every $\alpha > 0$ that*

$$2\alpha(\ln(\alpha) - \psi(\alpha)) - 1 = \mathbb{E}(e^{-\alpha \mathbf{X}})$$

Proof. Since $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ for every $\alpha > 0$ we obtain by relation (A.2) that

$$\ln(\Gamma(\alpha)) = \ln(\Gamma(\alpha+1)) - \ln(\alpha) = (\alpha - \frac{1}{2}) \ln(\alpha) - \alpha + \frac{1}{2} \ln(2\pi) + \theta(\alpha) \tag{A.7}$$

Taking the derivative in (A.7) and applying the representation of θ in Lemma A.0.2 shows

$$2\alpha(\psi(\alpha) - \ln(\alpha)) - 1 = \alpha \int_0^\infty F(t) e^{-\alpha t} dt = \mathbb{E}(e^{-\alpha \mathbf{X}})$$

and we have verifed the result. □

Using Lemma A.0.3 the following result follows immediately. This result is also shown in [36] by means of a more lengthy but related proof.

Lemma A.0.4. *It follows that the function $h_1 : (0, \infty) \to \mathbb{R}$ given by*

$$h_1(\alpha) = \alpha(\ln(\alpha) - \psi(\alpha)) \tag{A.8}$$

is strictly completely monotone satisfying $h_1(0+) = 1$ and $h_1(\infty) = \frac{1}{2}$.

Proof. Apply Lemma A.0.3. □

The next result is an easy consequence of Lemma A.0.4. Actually in [36] it is verified that h_υ is strictly completely monotone if and only if $\upsilon \leq 1$.

Lemma A.0.5. *The function $h_\upsilon : (0, \infty) \to \mathbb{R}$ given by*

$$h_\upsilon(\alpha) = \alpha^\upsilon (\ln(\alpha) - \psi(\alpha)) \tag{A.9}$$

is strictly completely monotone for every $\upsilon \leq 1$.

Proof. Clearly by Lemma A.0.4 the result holds for $\upsilon = 1$. Since it is easy to check by Leibniz formula for the differentiation of the product of two differentiable functions that the product of strictly completely monotonic functions is strictly completely monotone (cf.[35]) we obtain using $\alpha \to \alpha^{\upsilon-1}$ is strictly completely monotone for every $\upsilon < 1$ that by Lemma A.0.4 the result follows for $\upsilon < 1$. □

Using the above lemmas it is easy to show the following result needed for our analysis of the maximum likelihood problem for a generalized gamma distributed random variable.

Lemma A.0.6. *If the function* $H_0 : (0,\infty) \to \mathbb{R}$ *is given by*

$$H_0(\alpha) = \alpha(\ln(\alpha) - 1) - \ln(\Gamma(\alpha)) \tag{A.10}$$

then $(-1)^n H_0^{(n+1)}(\alpha) > 0$ *for every* $n \in \mathbb{Z}_+$ *and*

$$\lim_{\alpha\downarrow 0} H_0(\alpha) - \ln(\alpha) = 0 \tag{A.11}$$

and

$$\lim_{\alpha\uparrow\infty} H_0(\alpha) - \frac{1}{2}\ln(\alpha) + \frac{1}{2}\ln(2\pi) = 0. \tag{A.12}$$

Proof. Since it is easy to check that $H_0^{(1)}(\alpha) = h_0(\alpha)$ for every $\alpha > 0$ the first part follows by Lemma A.0.5. To verify relation (A.11) observe using $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ that

$$H_0(\alpha) = \alpha(\ln(\alpha) - 1) - \ln(\Gamma(\alpha+1) + \ln(\alpha) \tag{A.13}$$

and using $\Gamma(1) = 1$ this shows relation (A.11). To show relation (A.12) we observe by relation (A.2) and (A.13) that

$$H_0(\alpha) = (\alpha+1)\ln(\alpha) - \alpha - \ln\Gamma(\alpha+1) = \frac{1}{2}\ln(\alpha) - \frac{1}{2}\ln(2\pi) - \theta(\alpha) \tag{A.14}$$

Applying now the monotone convergence theorem and Lemma A.0.2 we obtain $\lim_{\alpha\uparrow\infty}\theta(\alpha) = 0$ and this shows the result. □

Appendix of on the method of moments approach for a generalized gamma population

In this Appendix we list some results for the polygamma functions. The next definition is well known (cf.[37],[38]).

Definition A.0.3. *For any* $m \in \mathbb{Z}_+$ *the function*

$$\psi_m(\alpha) = \psi^{(m)}(\alpha) = \frac{d^{m+1} \ln \Gamma}{d\alpha^{m+1}}(\alpha)$$

is called the polygamma function of order m

It is well know (cf.[37]) that for every $\alpha > 0$ and $n \in \mathbb{N}$

$$(-1)^{n-1}\psi_n(x) = \int_0^\infty \frac{t^n}{1-e^{-t}} e^{-\alpha t} dt = n! \sum_{k=0}^\infty \frac{1}{(\alpha+k)^{n+1}} \tag{A.15}$$

The next result is shown in [39]

Lemma A.0.7. *It follows for every* $\alpha > 0$ *and* $n \in \mathbb{N}, n \geq 2$ *that*

$$\frac{n-1}{n} < \frac{\psi_n(\alpha)^2}{\psi_{n-1}(\alpha)\psi_{n+1}(\alpha)} < \frac{n}{n+1}$$

Introducing the function $p : (0,\infty) \to \mathbb{R}$ given by

$$p(\alpha) = \frac{\psi_2(\alpha)}{\psi_1(\alpha)^{\frac{3}{2}}} \tag{A.16}$$

we will shown the following result.

Lemma A.0.8. *The function* p *defined in relation (A.16) is a negative strictly increasing function on* $(0,\infty)$ *satisfying* $p(\infty) = 0$ *and* $p(0+) = \lim_{\alpha \downarrow 0} p(\alpha) = -2$.

Proof. By relation (A.15) it follows that the function p is negative on $(0,\infty)$. Clearly its derivative is given by

$$p^{(1)}(\alpha) = \frac{-\frac{3}{2}\psi_2(\alpha)^2}{\psi_1(\alpha)^{\frac{5}{2}}} + \frac{\psi_3(\alpha)}{\psi_1(\alpha)^{\frac{3}{2}}} = \frac{\psi_3(\alpha)}{\psi_1(\alpha)^{\frac{3}{2}}}\left(1 - \frac{\frac{3}{2}\psi_2(\alpha)^2}{\psi_3(\alpha)\psi_1(\alpha)}\right)$$

Since by relation A.15 it follows that $\frac{\psi_3(\alpha)\psi_1(\alpha)}{\psi_1(\alpha)^{\frac{5}{2}}} > 0$ we obtain by Lemma A.0.7 that

$$p^{(1)}(\alpha) > \frac{\psi_3(\alpha)\psi_1(\alpha)}{\psi_1(\alpha)^{\frac{5}{2}}}(1-1) = 0$$

and we have shown the result. To show the asymptotic limits we observe by relation (A.15) that

$$\psi_1(\alpha) = \sum_{k=0}^{\infty}(\alpha+k)^{-2}$$

and

$$\psi_2(\alpha) = -2\sum_{k=0}^{\infty}(\alpha+k)^{-3}$$

for every $\alpha > 0$. This shows that the function p is negative on $(0,\infty)$. Since it is easy to check for every $\alpha > 1$ that

$$\alpha^{-1} \leq \sum_{k=0}^{\infty}(\alpha+k)^{-2} \leq (\alpha-1)^{-1}$$

and

$$\frac{1}{2}\alpha^{-2} \leq \sum_{k=0}^{\infty}(\alpha+k)^{-3} \leq \frac{1}{2}(\alpha-1)^{-2}$$

we obtain

$$-(\alpha-1)^{-2}\alpha^{\frac{3}{2}} \leq p(\alpha) \leq -(\alpha-1)^{\frac{3}{2}}\alpha^{-2}. \tag{A.17}$$

This shows $p(\infty) = 0$. To verify that $p(0+) = -2$ we observe for every $\alpha > 0$ that $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$. This implies by differentiation that $\psi_1(\alpha+1) = -\alpha^{-2} + \psi_1(\alpha)$ and $\psi_2(\alpha+1) = 2\alpha^{-3} + \psi_2(\alpha)$ for every $\alpha > 0$. Hence it follows that

$$\begin{aligned} p(0^+) &= \lim_{\alpha\downarrow 0}\frac{\psi_2(\alpha+1)-2\alpha^{-3}}{(\psi_1(\alpha+1)+\alpha^{-2})^{\frac{3}{2}}} &&= \lim_{\alpha\downarrow 0}\frac{\psi_2(1)-2\alpha^{-3}}{(\psi_1(1)+\alpha^{-2})^{\frac{3}{2}}} \\ &= -2\lim_{\alpha\downarrow 0}\frac{1}{\alpha^3(\psi_1(1)+\alpha^{-2})^{\frac{3}{2}}} &&= -2\lim_{\alpha\downarrow 0}\frac{1}{(\alpha^2\psi_1(1)+1)^{\frac{3}{2}}} \\ &= -2 \end{aligned}$$

and this shows the desired result. □

Appendix of on the sample mth moment and the maximum likelihood estimator of the mth (central) moment in a normal population.

In this appendix we list in the first two lemmas some known results (cf.[40] for the normal distribution. For completeness a short proof of these results is given. Observe the lower entire function $\lfloor . \rfloor : \mathbb{R} \to \mathbb{N}$ is defined by

$$\lfloor y \rfloor = \max\{n \in \mathbb{Z}_+ : n \leq y\}.$$

Lemma A.0.9. *If* $\mathbf{X}$ *has a normal distribution with parameter* $\mu \in \mathbb{R}$ *and* $\sigma > 0$ *then its moments are given*

$$\mu'_m(\mathbf{X}) = m! \sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \frac{\sigma^{2k}\mu^{m-2k}}{(m-2k)!2^k k!}. \tag{A.18}$$

Proof. It is well known that $\mathbf{X} \stackrel{d}{=} \mu + \sigma \mathbf{Y}$ with $\mathbf{Y}$ having a standard normal distribution. Applying Newton's binomial formula this shows for every $m \in \mathbb{Z}_+$ that

$$\mu'_m(\mathbf{X}) = \mathbb{E}((\mu + \sigma \mathbf{Y})^m) = \sum_{k=0}^{m} \binom{m}{k} \sigma^k \mu'_k(\mathbf{Y}) \mu^{m-k}. \tag{A.19}$$

Also it is known (cf.[5]) for every $s \in \mathbb{R}$ that the moment generating function $\varphi(s) = \mathbb{E}(e^{s\mathbf{Y}}) = e^{\frac{s^2}{2}}$. Applying now Taylor's formula for the exponential function we obtain

$$\sum_{k=0}^{\infty} \frac{\mu'_k(\mathbf{Y})}{k!} s^k = \mathbb{E}(e^{s\mathbf{Y}}) = e^{\frac{s^2}{2}} = \sum_{k=0}^{\infty} \frac{s^{2k}}{2^k k!}.$$

and this shows $\mu'_{2k}(\mathbf{Y}) = \frac{(2k)!}{2^k k!}$ and $\mu'_{2k+1}(\mathbf{Y}) = 0$ for every $k \in \mathbb{Z}_+$. Hence by relation (A.19) replacing m by $2m$, respectively $2m+1$ it follows that

$$\mu'_{2m}(\mathbf{X}) = \sum_{k=0}^{m} \binom{2m}{2k} \frac{(2k)!\sigma^{2k}\mu^{2m-2k}}{2^k k!}$$

and

$$\mu'_{2m+1}(\mathbf{X}) = \sum_{k=0}^{m} \binom{2m+1}{2k} \frac{(2k)!\sigma^{2k}\mu^{2m+1-2k}}{2^k k!}$$

showing the desired result. □

From Lemma A.0.9 it is easy to derive the centralized moments of a normal distribution observing $\mathbf{X} - \mathbb{E}(\mathbf{X})$ has a normal distribution with parameter $\mu = 0$ and $\sigma > 0$ if $\mathbf{X}$ has a normal distribution with parameter $\mu \in \mathbb{R}$ and $\sigma > 0$. Since

$\mu_m(\mathbf{X}) = \mu'_m(\mathbf{X} - \mathbb{E}\mathbf{X})$ for every $m \in \mathbb{Z}_+$ we obtain by Lemma A.0.9 that

$$\mu_{2m}(\mathbf{X}) = \frac{(2m)!\sigma^{2m}}{2^m m!}, \mu_{2m+1}(\mathbf{X}) = 0 \tag{A.20}$$

Another result needed in our analysis is the following.

Lemma A.0.10. *If the random variables $\mathbf{X}_i, i = 1,..,n$ are independent and normal distributed with parameter $\mu \in \mathbb{R}$ and $\sigma > 0$ then for every $m \in \mathbb{Z}_+$*

$$\mathbb{E}(\overline{\mathbf{X}}_n^m) = m! \sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \frac{1}{2^k (m-2k)! k!} \mu^{m-2k} \frac{\sigma^{2k}}{n^k} \tag{A.21}$$

with $\overline{\mathbf{X}}_n := \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$ denoting the sample mean estimator of the random variables $\mathbf{X}_1, ..., \mathbf{X}_n$.

Proof. Since $\overline{\mathbf{X}}_n \stackrel{d}{=} \mu + \sigma n^{-\frac{1}{2}} \mathbf{Z}$ with $\mathbf{Z}$ having a standard normal distribution we obtain

$$\mathbb{E}(\overline{\mathbf{X}}_n^{2m}) = \mathbb{E}((\mu + \sigma n^{-\frac{1}{2}} \mathbf{Z})^{2m}) = \sum_{k=0}^{2m} \binom{2m}{k} \sigma^k n^{-\frac{k}{2}} \mathbb{E}(\mathbf{Z}^k) \mu^{2m-k} \tag{A.22}$$

This shows by relation (A.22) and Lemma A.0.9 the desired result. A similar proof applies to $\mathbb{E}(\overline{\mathbf{X}}_n^{2m+1})$ and so we omit it. □

It is easy to see applying relation (A.21) that for $\mu \neq 0$

$$\lim_{n \uparrow \infty} n(\mathbb{E}(\overline{\mathbf{X}}_n^m) - \mu^m) = \frac{m(m-1)}{2} \mu^{m-2} \sigma^2. \tag{A.23}$$

Also for $\mu = 0$ we obtain

$$n\mathbb{E}(\overline{\mathbf{X}}_n^m) = \sigma^m n^{1-\frac{m}{2}} \mu_m(\mathbf{Z}) \tag{A.24}$$

with $\mathbf{Z}$ a standard normal distributed random variable.

In the next result we evaluate the cdf of the quadratic form

$$\mathbf{W}_n = \mathbf{Y}^\top \mathbf{a}_n \mathbf{b}_n^\top \mathbf{Y} \tag{A.25}$$

with $\mathbf{Y}^\top = (\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n)$ a random vector consisting of independent and standard normal distributed random variables and the vectors $\mathbf{a}_n$ and $\mathbf{b}_n$ given by

$$\mathbf{a}_n = \mathbf{e}_{1,n} - \frac{\mathbf{i}_n}{n}, \mathbf{b}_n = \mathbf{e}_{2,n} - \frac{\mathbf{i}_n}{n} \tag{A.26}$$

for every $n \in \mathbb{N}, n \geq 3$. (for the definition see after formula (4.13)).

Lemma A.0.11. *If* $\mathbf{W}_n = \mathbf{Y}^\top \mathbf{a}_n \mathbf{b}_n^\top \mathbf{Y}$ *with* $\mathbf{Y}^\top = (\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n)$ *a random vector consisting of independent and standard normal distributed random variables then for every* $n \in \mathbb{N}$, $n \geq 3$

$$\mathbf{W}_n \stackrel{d}{=} \left(\frac{1}{2} - \frac{1}{n}\right) \mathbf{Z}_1^2 - \frac{1}{2} \mathbf{Z}_2^2$$

with $\mathbf{Z}_1, \mathbf{Z}_2$ *independent standard normal distributed random variables.*

Proof. Since by relation (A.26) it is obvious that $a_{1n} b_{2n} = (1 - n^{-1})^2 \neq n^{-2} = b_{1n} a_{2n}$ the $n \times n$ matrix $A_n = \mathbf{a}_n \mathbf{b}_n^\top$ is not symmetric. Introducing the symmetric matrix $\overline{A}_n$ given

$$\overline{A}_n := \frac{1}{2}(A_n + A_n^\top)$$

we obtain $\mathbf{W}_n = \mathbf{Y}^\top \overline{A}_n \mathbf{Y}$. Since the $n \times n$ matrix $\overline{A}_n$ is symmetric it is well known (cf.[41]) that $\overline{A}_n$ has n real eigenvalues (counting multiciplicity) $\lambda_{1n}, ..., \lambda_{nn}$ and there exists a orthogonal $n \times n$ matrix P_n ($P_n^\top = P_n^{-1}$) satisfying

$$P_n^\top \overline{A}_n P_n = \Lambda_n \tag{A.27}$$

with Λ_n the diagonal matrix consisting of the real eigenvalues $\lambda_{1n}, ...\lambda_{nn}$ of the matrix $\overline{A}_n$. By the independence of the vectors $\mathbf{a}_n$ and $\mathbf{b}_n$ it also follows that the rank $r(\overline{A}_n)$ of the matrix $\overline{A}_n$ equals 2. Applying now the rule $r(\overline{A}_n B) = r(B)$ for any $n \times n$ invertible matrix B (cf.[41]) we obtain by relation (A.27) that $r(\Lambda_n) = 2$. This shows that only two eigenvalues say $\lambda_{1n}, \lambda_{2n}$ are nonzero. By relation (A.27) we obtain $\overline{A}_n = P_n \Lambda_n P_n^\top$ and this yields

$$\mathbf{W}_n = \mathbf{Z}_n^\top \Lambda_n \mathbf{Z}_n, \mathbf{Z}_n = P_n^\top \mathbf{Y} \tag{A.28}$$

Since the random vector $P_n^\top \mathbf{Y}$ has a multivariate normal distribution with mean $\mu = \mathbf{0}$ and

$$\text{Var}(P_n^\top \mathbf{Y}) = P_n^\top \text{Cov}(\mathbf{Y}, \mathbf{Y}) P_n = \sigma^2 P_n^\top I_n P_n = \sigma^2 I_n$$

it follows that $P_n^\top \mathbf{Y} \stackrel{d}{=} (\mathbf{Z}_1, ..., \mathbf{Z}_n)$ with $\mathbf{Z}_i, i = 1, ..., n$ independent standard normal distributed random variables. Substituting this into relation (A.28) we obtain by the above observations that

$$\mathbf{W}_n \stackrel{d}{=} \sum_{i=1}^n \lambda_i \mathbf{Z}_i^2 = \lambda_{1n} \mathbf{Z}_1^2 + \lambda_{2n} \mathbf{Z}_2^2$$

In the remainder of this proof we determine the nonzero eigenvalues of the matrix

$\overline{A}_n$. We first observe with $\mathbf{e}_i$ denoting the ith unit vector of $\mathbb{R}^n$ that

$$\mathbf{e}_n^\top \overline{A}_n \mathbf{e}_n = a_{nn} b_{nn} = n^{-2} > 0, \mathbf{e}_1^\top \overline{A}_n \mathbf{e}_1 = a_{1n} b_{1n} = -\frac{1}{n}(1-\frac{1}{n}) < 0$$

and so one of nonzero eigenvalues $\lambda_{1n}, \lambda_{2n}$ should be positive and one negative. Select λ_{2n}, as the negative one and λ_{1n} as the positive one. We will now determine a set of nonlinear equations satisfied by the eigenvalues $\lambda_{1n}, \lambda_{2n}$. Introducing the trace $\mathrm{tr}(C)$ of a $n \times n$ matrix C given by

$$\mathrm{tr}(C) = \sum\nolimits_{i=1}^{n} c_{ii}$$

it is well known (cf.[41]) for a symmetric matrix C that

$$\mathrm{tr}(C) = \sum\nolimits_{i=1}^{n} \mu_i$$

with μ_i the real eigenvalues of the matrix C. By this observation we obtain for the symmetric matrix $\overline{A}_n$ that

$$\text{(A.29)} \qquad \lambda_{1n} + \lambda_{2n} = \mathrm{tr}\left(\frac{\mathbf{a}_n \mathbf{b}_n^\top + \mathbf{b}_n \mathbf{a}_n^\top}{2}\right) = \mathbf{a}_n^\top \mathbf{b}_n = \frac{-2(1-n^{-1})}{n} + \frac{n-2}{n^2} = \frac{-1}{n}$$

Also it follows that the matrix $\overline{A}_n^2$ satisfies

$$\begin{aligned}
A_n^2 &= \left(\tfrac{\mathbf{a}_n \mathbf{b}_n^\top + \mathbf{b}_n \mathbf{a}_n^\top}{2}\right)\left(\tfrac{\mathbf{a}_n \mathbf{b}_n^\top + \mathbf{b}_n \mathbf{a}_n^\top}{2}\right) \\
&= \tfrac{1}{4}[\mathbf{a}_n \mathbf{b}_n^\top \mathbf{a}_n \mathbf{b}_n^\top + \mathbf{a}_n \mathbf{b}_n^\top \mathbf{b}_n \mathbf{a}_n^\top + \mathbf{b}_n \mathbf{a}_n^\top \mathbf{a}_n \mathbf{b}_n^\top + \mathbf{b}_n \mathbf{a}_n^\top \mathbf{b}_n \mathbf{a}_n^\top] \\
&= \tfrac{1}{4}\left[\mathbf{b}_n^\top \mathbf{a}_n \mathbf{a}_n \mathbf{b}_n^\top + \|\mathbf{b}_n\|_2^2 \, \mathbf{a}_n \mathbf{a}_n^\top + \|\mathbf{a}_n\|_2^2 \, \mathbf{b}_n \mathbf{b}_n^\top + \mathbf{a}_n^\top \mathbf{b}_n \mathbf{b}_n \mathbf{a}_n^\top\right] \\
&= \tfrac{1}{2} \mathbf{b}_n^\top \mathbf{a}_n A_n + \tfrac{1}{4} \|\mathbf{b}_n\|_2^2 \, \mathbf{a}_n \mathbf{a}_n^\top + \|\mathbf{a}_n\|_2^2 \, \mathbf{b}_n \mathbf{b}_n^\top
\end{aligned}$$

Hence its trace $\mathrm{tr}(\overline{A}_n^2)$ is given by

$$\text{(A.30)} \qquad \begin{aligned}
\mathrm{tr}(\overline{A}_n^2) &= -\tfrac{1}{2n}(\mathbf{b}_n^\top \mathbf{a}_n)^2 + \tfrac{1}{2}\|\mathbf{b}_n\|_2^2 \|\mathbf{a}_n\|_2^2 \\
&= \tfrac{1}{2n^2} + \tfrac{1}{2}(1-\tfrac{1}{n})^2 \\
&= \tfrac{1}{2} - n^{-1} + n^{-2}
\end{aligned}$$

Since A_n^2 is again symmetric and $\lambda_{i,n}^2, i = 1, ..., n$ are its eigenvalues we obtain by relation (A.30) that

$$\lambda_{1,n}^2 + \lambda_{2n}^2 = n^{-2} - n^{-1} + \frac{1}{2}$$

Hence the two nonzero eigenvalues $\lambda_{1n}, \lambda_{2n}$ must satisfy the system

$$\lambda_{1n} + \lambda_{2n} = \frac{-1}{n}, \lambda_{1,n}^2 + \lambda_{2n}^2 = n^{-2} - n^{-1} + \frac{1}{2}, \lambda_{1,n} > 0, \lambda_{2,n} < 0$$

This is equivalent to the system

$$\lambda_{1n} = -\frac{1}{n} - \lambda_{2n}, (-\frac{1}{n} - \lambda_{2n})^2 + \lambda_{2n}^2 = n^{-2} - n^{-1} + \frac{1}{2}, \lambda_{2,n} < 0$$

and this implies

$$2\lambda_{2n}^2 + \frac{2}{n}\lambda_{2n} + \frac{1}{n^2} = n^{-2} - n^{-1} + \frac{1}{2}, \lambda_{2,n} < 0$$

or equivalently

$$\left(\lambda_{2n} + \frac{1}{2n}\right)^2 = \frac{1}{4} - \frac{1}{2n} + \frac{1}{4n^2} = \frac{1}{4}(1 - \frac{2}{n} + \frac{1}{n^2}) = \frac{1}{4}(1 - n^{-1})^2$$

Since $\frac{1}{4} - \frac{1}{2n} + \frac{1}{4n^2} \geq \frac{1}{4n^2}$ for $n \geq 2$ it must follow that $\lambda_{2n} = -\frac{1}{2}(1 - n^{-1}) - \frac{1}{2n} = -\frac{1}{2}$ and $\lambda_{1:n} = \frac{1}{2} - \frac{1}{n}$ showing the desired result. □