Statistica Sinica

A modified closed-form maximum likelihood estimator

Pedro Luiz Ramos[a], Eduardo Ramos[a],

Francisco A. Rodrigues[a], and Francisco Louzada[a]

[a]Institute of Mathematical Science and Computing

University of São Paulo, São Carlos, Brazil.

Abstract: The maximum likelihood estimator plays a fundamental role in statistics. However, for many models, the estimators do not have closed-form expressions. This limitation can be significant in situations where estimates and predictions need to be computed in real-time, such as in applications based on embedded technology, in which numerical methods can not be implemented. This paper provides a modification in the maximum likelihood estimator that allows us to obtain the estimators in closed-form expressions under some conditions. Under mild conditions, the estimator is invariant under one-to-one transformations, consistent, and has an asymptotic normal distribution. The proposed modified version of the maximum likelihood estimator is illustrated on the Gamma, Nakagami, and Beta distributions and compared with the standard maximum likelihood estimator.

Key words and phrases: Closed-form estimators; Maximum Likelihood estimators; Modified maximum likelihood estimator.

1 Introduction

Introduced by Ronald Fisher (Aldrich et al., 1997), the maximum likelihood method is one of the most well-known and used inferential procedures to estimate the unknown parameters of a given distribution. Alternative methods to the maximum likelihood estimator (MLE) have been proposed in the literature, such as those based on statistical moments (Hosking, 1990), percentile (Kao, 1958, 1959), product of spacings (Cheng and Amin, 1983), or goodness of fit measures, to list a few. Although alternative inferential methods are popular nowadays, the MLEs are the most widely used due to their flexibility in incorporating additional complexity (such as random effects, covariates, censoring, among others) and their properties: asymptotical efficiency, consistent and invariance under one-to-one transformation. These properties are achieved when the MLEs satisfy some regularity conditions (Bierens, 2004; Lehmann and Casella, 2006; Redner et al., 1981).

It is now well established from various studies that the MLEs do not always return closed-form expressions for most common distributions. In these cases, numerical methods, such as Newton-Rapshon or its variants, are usually considered to find the values that maximize the likelihood function. Important variants of the maximum likelihood estimator, such as profile (Murphy and Van der Vaart, 2000), pseudo (Gourieroux et al., 1984), conditional (Andersen, 1970), penalized (Anderson and Blair, 1982; Firth, 1993) and marginal likelihoods (Cox, 1975), have been presented to eliminate nuisance parameters and decrease the computational cost. Another important procedure to achieve the MLEs is the expectation-maximization (EM) algorithm (Dempster et al., 1977), which involves unobserved latent variables jointly with unknown parameters. The expectation and maximization steps also involve, in most cases, the use of numerical methods that may have a high computational cost. However, there is a need to use closed-form estimators to estimate the unknown parameters in many situations. For instance, in embed technology, small components need to compute the estimates without using maximization procedures, and in real-time applications, it is necessary to provide an immediate answer.

In the proposed study, we discuss a modification in the maximum likelihood method to obtain closed-form expressions for the distribution parameters' estimators. Moreover, we propose conditions under which the proposed estimator is strongly consistent and asymptotic normal. Additionally, we show such conditions are greatly simplified and easier to verify for distributions that belong to a specific general family of distributions. Further, the proposed approach is extended for a generic baseline distribution. Overall, to find the closed-form estimator, the method depends on a generalized version of the standard distribution and is based on its likelihood equations.

The proposed method is illustrated regarding the Gamma, Beta, and Nakagami distributions. In these cases, the standard ML does not have closed-form expressions, and numerical methods or approximations are necessary to find these distributions' solutions. Hence our approach does not require iterative numerical methods, and, computationally, the work required by using our estimators is less complicated than that required in ML estimators. The remainder of this paper is organized as follows. Section 2 presents the new modified maximum likelihood estimator and its properties. Section 3 considers the application of the Gamma, Nakagami, and Beta distributions. Finally, Section 4 summarizes the study.

2 Modified Maximum Likelihood Estimator

The method we propose here can be applied to obtain closed-form expressions for distributions with a given density $f(x\,;\boldsymbol{\theta})$. In order to formulate the method, let $\Theta\subset\mathbb{R}^s$ be an open set containing the vector parameter $\boldsymbol{\theta}$, which need to be estimated, and $\mathcal{A}\subset\mathbb{R}^r$, $0\leq r\leq s$, be an open set containing an additional parameter α, which will be used during the procedure to obtain the estimators. Moreover, Ω will represent the sample space, and $\mathcal{X}_\alpha\subset\mathbb{R}$ will be a measurable set which depends on the additional parameter α and represents the space of the data x, where we denote $\mathcal{X}=\mathcal{X}_{\boldsymbol{\alpha}_0}$

Now, given a parameter $\boldsymbol{\theta}_0\in\Theta$ to be estimated and a fixed $\boldsymbol{\alpha}_0\in\mathcal{A}$, suppose X_1, X_2, $\cdots$, X_n are independent and identically distributed (iid) random variables, which can be either discrete or continuous, with a strictly positive density function $f(x\,;\boldsymbol{\theta}_0)$ defined for all $x\in\mathcal{X}$, such that $f(x\,;\boldsymbol{\theta})$ can be viewed as a special case of a density probability function with additional parameters $g(x\,;\boldsymbol{\theta},\boldsymbol{\alpha})$ defined for all $x\in\mathcal{X}_\alpha$ in the sense that $f(x\,;\boldsymbol{\theta})=g(x\,;\boldsymbol{\theta},\boldsymbol{\alpha}_0)$ for all $x\in\mathcal{X}$ and $\boldsymbol{\theta}\in\Theta$. Then we define the *modified maximum likelihood equations for* g *over* $\boldsymbol{\alpha}=\boldsymbol{\alpha}_0$ to be the set of equations

$$\begin{aligned}\sum_{i=1}^{n}\frac{\partial}{\partial\theta_j}\log\, g(X_i\,;\boldsymbol{\theta},\boldsymbol{\alpha}_0)=0,\quad 1\leq j\leq s-r,\\ \sum_{i=1}^{n}\frac{\partial}{\partial\alpha_j}\log\, g(X_i\,;\boldsymbol{\theta},\boldsymbol{\alpha}_0)=0,\quad 1\leq j\leq r,\end{aligned}\tag{2.1}$$

as long as these partial derivatives are well defined and exist. In particular, for the partial derivative $\frac{\partial}{\partial\alpha_j}\log\, g(x\,;\boldsymbol{\theta},\boldsymbol{\alpha}_0)$ to be well defined, we assume henceforth that for each $x\in\mathcal{X}$ there exists an open set $\mathcal{A}_x\subset\mathcal{A}$ containing $\boldsymbol{\alpha}_0$ such that $x\in\mathcal{X}_\alpha$ for all $\boldsymbol{\alpha}\in\mathcal{A}_x$. Thus, this implies that for each fixed $x\in\mathcal{X}$ the function $g(x\,;\boldsymbol{\theta},\boldsymbol{\alpha})$ is well defined for all $\boldsymbol{\alpha}$

close enough to $\boldsymbol{\alpha}_0$.

From now on, our goal shall be that of giving conditions such that an obtained solution $\hat{\boldsymbol{\theta}}_n(X)$ of the modified maximum likelihood equations is a consistent estimator for the true parameter $\boldsymbol{\theta}_0$, and is asymptotically normal. In order to formulate the result, given a fixed $\boldsymbol{\alpha}_0 \in \Theta$, define $J(\boldsymbol{\theta}) = (J_{i,j}(\boldsymbol{\theta})) \in M_s(\mathbb{R})$ and $K(\boldsymbol{\theta}) = (K_{i,j}(\boldsymbol{\theta})) \in M_s(\mathbb{R})$, by

$$\begin{aligned} J_{i,j}(\boldsymbol{\theta}) &= \mathrm{E}_{\boldsymbol{\theta}}\left[-\frac{\partial^2 \log\, g(X_1\,;\,\boldsymbol{\theta},\boldsymbol{\alpha}_0)}{\partial\theta_i\beta_j}\right] \text{ and} \\ K_{i,j}(\boldsymbol{\theta}) &= \mathrm{cov}_{\boldsymbol{\theta}}\left[\frac{\partial \log\, g(X_1\,;\,\boldsymbol{\theta},\boldsymbol{\alpha}_0)}{\partial\beta_i}, \frac{\partial \log\, g(X_1\,;\,\boldsymbol{\theta},\boldsymbol{\alpha}_0)}{\partial\beta_j}\right], \end{aligned} \tag{2.2}$$

for all $1 \leq i \leq s$ and $1 \leq j \leq s$, where $(\beta_1, \cdots, \beta_s) = (\theta_1, \cdots, \theta_{s-r}, \alpha_1, \cdots, \alpha_r)$. These matrices shall play the role that the Fisher information matrix I plays in the classical maximum likelihood method. In the above formulas and the next result, some partial derivatives of g are necessary to formulate the equations, and thus we suppose such partial derivatives exist in the indicated points. Additionally, in the next result we suppose $g(x\,;\,\boldsymbol{\theta},\boldsymbol{\alpha}_0)$ is a strictly positive probability density function in $x \in \mathcal{X}$ for all $\boldsymbol{\theta} \in \Theta$.

Theorem 1. *Let $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$, where $X_1, \cdots, X_n$ are iid with density $g(x\,;\,\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0) : \mathcal{X} \to \mathbb{R}$ and suppose $\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_{1n}, \cdots, \hat{\theta}_{sn})$ is a measurable function in x satisfying the following properties for all $n \geq 2$:*

- *(A) With probability one in Ω, the modified maximum likelihood equations have $\hat{\boldsymbol{\theta}}_n(\boldsymbol{X})$ as its unique solution.*

- *(B) $J(\boldsymbol{\theta}_0)$ and $K(\boldsymbol{\theta}_0)$, as defined in (2.2), exist and $J(\boldsymbol{\theta}_0)$ is invertible.*

(C) The following hold

$$\mathrm{E}_{\boldsymbol{\theta}_0}\left[\frac{\partial \log g(X_1\,;\,\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)}{\partial \beta_i}\right]=0,$$

for all $1 \le i \le s$*, where* $(\beta_1,\cdots,,\beta_s)=(\theta_1,\cdots,\theta_{s-r},\alpha_1,\cdots,\alpha_r)$*.*

(D) There exist measurable functions $M_{ij}(x)$ *and an open set* Θ_0 *containing the true parameter* $\boldsymbol{\theta}_0$ *such that* $\overline{\Theta}_0 \subset \Theta$ *and for all* $\boldsymbol{\theta} \in \overline{\Theta}_0$ *we have*

$$\left|\frac{\partial^2 \log g(X_1\,;\,\boldsymbol{\theta},\boldsymbol{\alpha}_0)}{\partial\theta_i\partial\beta_j}\right| \le M_{ij}(X_1) \text{ and } E_{\boldsymbol{\theta}_0}\left[M_{ij}(X_1)\right]<\infty,$$

for all $1 \le i \le s$ *and* $1 \le j \le s$*, where* $(\beta_1,\cdots,,\beta_s)=(\theta_1,\cdots,\theta_{s-r},\alpha_1,\cdots,\alpha_r)$*.*

Then:

I) $\hat{\boldsymbol{\theta}}_n(\boldsymbol{X})$ *is a strongly consistent estimator for* $\boldsymbol{\theta}_0$*.*

II) $\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)^T \xrightarrow{D} N_s\left(0,(J(\boldsymbol{\theta}_0)^{-1})^T K(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1}\right)$*, where here* A^T *denotes the transpose of* A*.*

Usually in applications, when the random variables are continuous, the modified maximum likelihood equations (2.1) will have a unique solution for $n \ge 2$ unless $X_1 = X_2 = \cdots = X_n$. Since this equality has probability zero of occurring when case the random variables are continuous, it follows that condition (A) is satisfied in such setting. For discrete random variables, we assume $X_i \neq X_j$ that at least one $i \neq j$. Additionally, notice that if $r = 0$ then condition (B) corresponds to asking the Fischer information matrix $I(\boldsymbol{\theta}_0)$ to be invertible, since in such case $J(\boldsymbol{\theta}_0) = I(\boldsymbol{\theta}_0)$.

As a consequence of the above theorem we have the following result, which reflects the fact that conditions (C) and (D) of Theorem 1 are greatly simplified when $g(x : \boldsymbol{\theta}, \boldsymbol{\alpha}_0)$ is contained in a certain family of distributions.

Theorem 2. *Let $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ be iid with density $g(x\,;\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0) : \mathcal{X} \to \mathbb{R}$ where $g(x\,;\boldsymbol{\theta}, \boldsymbol{\alpha}_0)$ is a strictly positive probability density function for all $x \in \mathcal{X}$, $\boldsymbol{\theta} \in \Theta$, satisfying*

$$g(x\,;\boldsymbol{\theta}, \boldsymbol{\alpha}_0) = V(x) \exp\left(\sum_{i=1}^{s} \eta_i(\boldsymbol{\theta}) T_i(x, \boldsymbol{\alpha}_0) + L(\boldsymbol{\theta}, \boldsymbol{\alpha}_0)\right) \text{ for all } x \in \mathcal{X},\ \boldsymbol{\theta} \in \Theta,$$

where η_i and L are C^2 functions, $V(x)$ is measurable in x, $T_i(x, \boldsymbol{\alpha})$ is measurable on x for each $\boldsymbol{\alpha} \in \mathcal{A}$, the partial derivatives $\frac{\partial T_i(x, \boldsymbol{\alpha}_0)}{\partial \alpha_i}$ exist for each $x \in \mathcal{X}$, and suppose $\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_{1n}, \cdots, \hat{\theta}_{sn})$ is measurable in x satisfying the following properties for all $n \geq 2$:

(A) With probability one in Ω, the modified maximum-likelihood equations have $\hat{\boldsymbol{\theta}}_n(\boldsymbol{X})$ as its unique solution.

(B) $J(\boldsymbol{\theta}_0)$ and $K(\boldsymbol{\theta}_0)$, as defined in (2.2), exist and $J(\boldsymbol{\theta}_0)$ is invertible.

(C) For all $1 \leq i \leq r$ and $\boldsymbol{\theta} \in \Theta$ the function $\dfrac{\partial \log g(X_1\,;\boldsymbol{\theta}, \boldsymbol{\alpha}_0)}{\partial \alpha_i}$ has finite expectation and

$$\mathrm{E}_{\boldsymbol{\theta}_0}\left[\frac{\partial \log g(X_1\,;\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0)}{\partial \alpha_i}\right] = 0.$$

(D) The function $\pi : \Theta \to \mathbb{R}^s$ defined by $\pi(\boldsymbol{\theta}) = (\eta_1(\boldsymbol{\theta}), \cdots, \eta_s(\boldsymbol{\theta}))$ is a local diffeomorphism.

Then:

I) $\hat{\boldsymbol{\theta}}_n(\boldsymbol{X})$ *is a strongly consistent estimator for* $\boldsymbol{\theta}_0$*.*

II) $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^T \xrightarrow{D} N_s\big(0, (J(\boldsymbol{\theta}_0)^{-1})^T K(\boldsymbol{\theta}_0) J(\boldsymbol{\theta}_0)^{-1}\big)$*.*

Proof. The proof is available in the Appendix. □

Proposition 1 (One-to-one invariance)**.** *Let* $\Theta_1,\ \Lambda_1 \subset \mathbb{R}^{s-r}$ *and* $\Theta_2,\ \Lambda_2 \subset \mathbb{R}^r$ *be open sets, let* $\Theta = \Theta_1 \times \Theta_2$*,* $\Lambda = \Lambda_1 \times \Lambda_2$ *and let* $\pi : \Lambda \to \Theta$ *be of the form*

$$\pi(\boldsymbol{\lambda}) = (\pi_1(\boldsymbol{\lambda_1}), \pi_2(\boldsymbol{\lambda_2})),\ \textit{for all}\ \boldsymbol{\lambda} = (\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}) \in \Lambda_1 \times \Lambda_2,$$

where $\pi_1 : \Lambda_1 \to \Theta_1$ *and* $\pi_2 : \Lambda_2 \to \Theta_2$ *are diffeomorphism. Moreover, let* $f_1(x\,;\,\boldsymbol{\theta}, \boldsymbol{\alpha}) : \mathcal{X}_{\boldsymbol{\alpha}} \to \mathbb{R}$ *be a probability density function in* $x \in \mathcal{X}_{\boldsymbol{\alpha}}$ *for all* $(\boldsymbol{\theta}, \boldsymbol{\alpha}) \in \Theta \times \mathcal{A}$*, let*

$$f_2(x\,;\,\boldsymbol{\lambda}, \boldsymbol{\alpha}) = f_1(x\,;\,\pi(\boldsymbol{\lambda}), \boldsymbol{\alpha})\ \textit{for all}\ \boldsymbol{\lambda} \in \Lambda,\ \boldsymbol{\theta} \in \Theta,$$

and suppose that for some $n \in \mathbb{N}$*, with probability one on* Ω*,* $\hat{\boldsymbol{\theta}}_n(\boldsymbol{X})$ *and* $\hat{\boldsymbol{\lambda}}_n(\boldsymbol{X})$ *are the only solutions for the modified maximum likelihood equations (2.1) over, respectively, the functions* $f_1(x\,;\,\boldsymbol{\theta}, \boldsymbol{\alpha})$ *and* $f_2(x\,;\,\boldsymbol{\lambda}, \boldsymbol{\alpha})$ *over* $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$*. Then, with probability one in* Ω*, we have* $\hat{\boldsymbol{\theta}}_n(\boldsymbol{X}) = \pi(\hat{\boldsymbol{\lambda}}_n(\boldsymbol{X}))$*.*

Proof. By hypothesis, with probability one in Ω, $\hat{\boldsymbol{\theta}}_n(\boldsymbol{X})$ satisfies

$$\begin{aligned} \sum_{i=1}^n \frac{\partial}{\partial \theta_j} \log f_1(X_i\,;\,\hat{\boldsymbol{\theta}}_n(\boldsymbol{X}), \boldsymbol{\alpha}_0) = 0,\ 1 \le j \le s-r, \\ \sum_{i=1}^n \frac{\partial}{\partial \alpha_j} \log f_1(X_i\,;\,\hat{\boldsymbol{\theta}}_n(\boldsymbol{X}), \boldsymbol{\alpha}_0) = 0,\ 1 \le j \le r. \end{aligned} \tag{2.3}$$

Thus, denoting $l(\boldsymbol{\theta}, \boldsymbol{\alpha}\,;\,\boldsymbol{X}) = \sum_{i=1}^s \log f_1(X_i\,;\,\boldsymbol{\theta}, \boldsymbol{\alpha})$, by hypothesis $\sum_{i=1}^s \log f_2(X_i\,;\,\boldsymbol{\lambda}, \boldsymbol{\alpha}_0) = l(\pi(\boldsymbol{\lambda}), \boldsymbol{\alpha_0}\,;\,\boldsymbol{X})$, and thus, given $1 \le j \le s-r$ and letting $\pi_1(\boldsymbol{\lambda}_1) = (\pi_{11}(\boldsymbol{\lambda}_1), \cdots, \pi_{1(s-r)}(\boldsymbol{\lambda}_1))$, since by hypothesis π_2 does not depend on λ_j for $1 \le j \le s-r$, from the chain rule it follows that

$$\sum_{i=1}^{n} \frac{\partial \log f_2(X_i\,;\,\boldsymbol{\lambda}, \boldsymbol{\alpha}_0)}{\partial \lambda_j} = \sum_{k=1}^{s-r} \frac{\partial\, l(\pi(\boldsymbol{\lambda}), \boldsymbol{\alpha}_0\,;\, \boldsymbol{X})}{\partial \theta_k} \frac{\partial \pi_{1k}(\boldsymbol{\lambda}_1)}{\partial \lambda_j}, \tag{2.4}$$

and thus it follows combining (2.3) and (2.4) that

$$\sum_{i=1}^{n} \frac{\partial\, \log f_2(X_i\,;\, \pi^{-1}(\hat{\boldsymbol{\theta}}_n(\boldsymbol{X})), \boldsymbol{\alpha}_0)}{\partial \lambda_j} = 0 \text{ for } 1 \le j \le s-r.$$

Additionally, given $1 \le j \le r$, since by hypothesis π does not depend on the variable α_j it follows that

$$\sum_{i=1}^{n} \frac{\partial\, \log f_2(X_i\,;\, \pi^{-1}(\hat{\boldsymbol{\theta}}_n(\boldsymbol{X})), \boldsymbol{\alpha}_0)}{\partial \alpha_j} = \sum_{i=1}^{n} \frac{\partial\, \log f_1(X_i\,;\, \hat{\boldsymbol{\theta}}_n(\boldsymbol{X}), \boldsymbol{\alpha}_0)}{\partial \alpha_j} = 0,$$

for all $1 \le j \le r$. Thus, it follows from unicity that $\hat{\boldsymbol{\lambda}}_n(\boldsymbol{X}) = \pi^{-1}(\hat{\boldsymbol{\theta}}_n(\boldsymbol{X}))$ a.e. in Ω, which proves the proposition. □

In general, MML estimators will not necessarily be functions of sufficient statistics. Additionally, from an intuitive point of view, we will depend on generalized versions of the distribution to achieve the estimators, where some of the MLEs can be isolated in closed-form expressions. Although this idea seems not so familiar in practice, due to the high number of new distributions introduced in the past decades, it is not difficult to fulfill such conditions. In the next section, we present applications of the proposed method.

3 Examples

We illustrate the previous section's discussion by applying the Gamma, Nakagami-m, and Beta distribution approaches. The examples are for illustration, so we shall not present their backgrounds. The standard MLEs for the cited distributions are widely discussed in statistical books, which shows that no closed-form expression can be achieved using the MLE method.

The Gamma and the Nakagami distributions are particular cases of the generalized Gamma distribution. Therefore, we will consider this generalized distribution to obtain the modified maximum-likelihood equations used to obtain the closed-form estimators.

Example 1: Let us consider that X_1, X_2, $\cdots$ are iid random variables (RV) following a gamma distribution with probability density function (PDF) given by:

$$f(x\,;\,\lambda,\phi) = \frac{1}{\Gamma(\phi)}\left(\frac{\phi}{\lambda}\right)^{\phi} x^{\phi-1}\exp\left(-\frac{\phi}{\lambda}x\right) \text{ for all } x > 0, \tag{3.5}$$

where $\phi > 0$ is the shape parameter, $\lambda > 0$ is the scale parameter and $\Gamma(\alpha) = \int_0^\infty e^{-x}x^{\alpha-1}dx$ is the gamma function.

We can apply the modified maximum likelihood approach for this distribution by considering the density function $g(x\,;\,\lambda,\phi,\alpha)$ representing the generalized gamma distribution, where $\lambda > 0$, $\phi > 0$ and $\alpha > 0$, given by

$$g(x\,;\,\lambda,\phi,\alpha) = \frac{\alpha}{\Gamma(\phi)}\left(\frac{\phi}{\lambda}\right)^{\phi} x^{\alpha\phi-1}\exp\left(-\frac{\phi}{\lambda}x^{\alpha}\right) \text{ for all } x > 0. \tag{3.6}$$

Notice that $f(x\,;\,\lambda,\phi) = g(x\,;\,\lambda,\phi,1)$ for all $\lambda > 0$ and $\phi > 0$, and thus the modified maximum-likelihood method is applicable. Now, given $n \geq 2$, letting $s = 1$ and $r = 1$, the modified likelihood equations for g over $\alpha_0 = 1$ are given by

$$\sum_{i=1}^{n} \frac{\partial}{\partial \lambda} \log g(X_i\,;\,\lambda, \phi, 1) = \frac{n\phi}{\lambda} - \frac{\phi}{\lambda^2} \sum_{i=1}^{n} X_i \text{ and}$$

$$\sum_{i=1}^{n} \frac{\partial}{\partial \alpha} \log g(X_i\,;\,\lambda, \phi, 1) = n + \phi \left(\sum_{i=1}^{n} \log\left(X_i\right) - \frac{1}{\lambda} \sum_{i=1}^{n} X_i \log\left(X_i\right) \right) = 0.$$

Following Louzada et al. (2019), as long as the equality $X_1 = \cdots = X_n$ does not hold, we have $\sum_{i=1}^{n} X_i \log\left(X_i\right) - \frac{1}{n} \sum_{i=1}^{n} X_i \sum_{i=1}^{n} \log\left(X_i\right) \neq 0$, in which case the modified likelihood equations above has as only solution

$$\hat{\lambda}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \text{ and } \hat{\phi}_n = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i \log\left(X_i\right) - \frac{1}{n} \sum_{i=1}^{n} X_i \sum_{i=1}^{n} \log\left(X_i\right)}. \tag{3.7}$$

On the other hand, the MLE for ϕ and λ would be obtained by solving the non-linear system of equations

$$\lambda = \frac{1}{n} \sum_{i=1}^{n} X_i \text{ and } \log(\phi) - \psi(\phi) = \log(\lambda) - \frac{1}{n} \sum_{i=1}^{n} \log\left(X_i\right), \tag{3.8}$$

where $\psi(k) = \frac{\partial}{\partial k} \log \Gamma(k) = \frac{\Gamma'(k)}{\Gamma(k)}$ is the digamma function.

Proposition 2. *$\hat{\phi}_n$ and $\hat{\lambda}_n$ are strongly consistent estimators for the true parameters ϕ and λ, and asymptotically normal with $\sqrt{n}\left(\hat{\lambda}_n - \lambda\right) \xrightarrow{D} N\left(0, \lambda^2/\phi\right)$ and $\sqrt{n}\left(\hat{\phi}_n - \phi\right) \xrightarrow{D} N\left(0, \phi^3 \psi'(\phi+1) + \phi^2\right)$.*

Proof. In order to apply Theorem 2 we notice $g(x\,;\,\lambda, \phi, \alpha_0)$ can be written as

$$g(x\,;\,\lambda, \phi, \alpha_0) = V(x) \exp\left[\eta_1(\lambda, \phi) \left(\alpha_0 \log(x)\right) + \eta_2(\lambda, \phi) x^{\alpha_0} + L(\lambda, \phi, \alpha_0)\right]$$

for all $x > 0$, $\lambda > 0$, $\theta > 0$ and $\alpha_0 > 0$, where

$$V(x) = \frac{1}{x},\ \eta_1(\lambda, \phi) = \phi,\ \eta_2(\lambda, \phi) = -\frac{\phi}{\lambda} \text{ and}$$

$$L(\lambda, \phi, \alpha_0) = \log\left(\alpha_0\right) - \log\left(\Gamma(\phi)\right) + \phi \log\left(\frac{\phi}{\lambda}\right).$$

Now we notice that condition (A) of Theorem 2 is automatically satisfied, since we have already verified that the modified-likelihood has, with probability one, $(\hat{\lambda}_n, \hat{\phi}_n)$ as its only solution for all $n \geq 2$. To check condition (B) of Theorem 2 notice that, for $\alpha = \alpha_0$ and using reparametrization over the Fisher information matrix from the GG distribution available in Hager and Bain (1970) we have

$$I(\lambda,\phi,\alpha) = \begin{bmatrix} \dfrac{\phi}{\lambda^2} & 0 & \frac{\phi\log\left(\frac{\phi}{\lambda}\right)-\phi\psi(\phi)-1}{\lambda} \\ 0 & \frac{\phi\psi'(\phi)-1}{\phi} & \frac{1}{\phi} \\ \frac{\phi\log\left(\frac{\phi}{\lambda}\right)-\phi\psi(\phi)-1}{\lambda} & \frac{1}{\phi} & I_{\alpha,\alpha}(\boldsymbol{\theta}) \end{bmatrix}, \tag{3.9}$$

where

$$I_{\alpha,\alpha}(\boldsymbol{\theta}) = \log\left(\frac{\phi}{\lambda}\right)\left(\phi\log\left(\frac{\phi}{\lambda}\right) - 2\phi\psi(\phi) - 2\right) + \phi\psi'(\phi) + 2\psi(\phi) + \phi\psi(\phi)^2 + 1.$$

Therefore, we have

$$J(\lambda,\phi) = \begin{bmatrix} \dfrac{\phi}{\lambda^2} & \frac{\phi\log\left(\frac{\phi}{\lambda}\right)-\phi\psi(\phi)-1}{\lambda} \\ 0 & \frac{1}{\phi} \end{bmatrix} \text{ and } K(\lambda,\phi) = \begin{bmatrix} \dfrac{\phi}{\lambda^2} & \frac{\phi\log\left(\frac{\phi}{\lambda}\right)-\phi\psi(\phi)-1}{\lambda} \\ \frac{\phi\log\left(\frac{\phi}{\lambda}\right)-\phi\psi(\phi)-1}{\lambda} & I_{\alpha,\alpha}(\boldsymbol{\theta}) \end{bmatrix}, \tag{3.10}$$

and thus, since $\det(J(\lambda,\phi)) = \frac{1}{\lambda^2}$, it follows that $J(\lambda,\phi)$ is invertible for all $\phi > 0$ and $\lambda > 0$ with

$$J(\lambda,\phi)^{-1} = \begin{bmatrix} \frac{\lambda^2}{\phi} & -\lambda\left(\phi\log\left(\frac{\phi}{\lambda}\right) - \phi\psi(\phi) - 1\right) \\ 0 & \phi \end{bmatrix}.$$

In special it follows that condition (B) is verified and moreover, after some algebraic computations, one can verify that

$$(J(\lambda,\phi)^{-1})^T K(\lambda,\phi) J(\lambda,\phi)^{-1} = \begin{bmatrix} \frac{\lambda^2}{\phi} & 0 \\ 0 & \phi^3\psi'(\phi+1) + \phi^2 \end{bmatrix}. \tag{3.11}$$

To check condition (C) of of Theorem 2 notice that

$$\mathrm{E}_{\lambda,\phi}\left[\log(X_1)\right] = \psi(\phi) - \log\left(\frac{\phi}{\lambda}\right) \text{ and}$$
$$\mathrm{E}_{\lambda,\phi}\left[X_1\log(X_1)\right] = \lambda\left(\psi(\phi) + \frac{1}{\phi} - \log\left(\frac{\phi}{\lambda}\right)\right),$$

from which it follows that $\frac{\partial}{\partial\alpha}\log g(X_1\,;\,\lambda^*,\phi^*,1)$ has finite expectation for all $\lambda^* > 0$ and $\phi^* > 0$ and

$$\mathrm{E}_{\lambda,\phi}\left[\frac{\partial}{\partial\alpha}\log g(X_1\,;\,\lambda,\phi,1)\right] = \mathrm{E}_{\lambda,\phi}[1] + \phi\,\mathrm{E}_{\lambda,\phi}\left[\log\left(X_1\right)\right] - \frac{\phi}{\lambda}\,\mathrm{E}_{\lambda,\phi}\left[X_1\log\left(X_1\right)\right] = 0.$$

Finally, condition (D) of Theorem 2 follows by noticing that $\pi:(0,\infty)^2\to(0,\infty)^2$ given by $\pi(\lambda,\phi) = (\eta_1(\lambda,\phi),\eta_2(\lambda,\phi)) = (\phi,-\frac{\phi}{\lambda})$ is a local diffeomorphism, since it has non singular Jacobian for $\lambda > 0$ and $\phi > 0$. Thus, we verified conditions (A) to (D) of Theorem 2, and therefore the proposition follows from the conclusion of Theorem 2 combined with (3.11).

□

Note that the MLE of ϕ differs from the obtained using our approach, which leads to a closed-form expression. Figure 1 presents the Bias and root of the mean square error (RMSE) obtained from $100,000$ replications assuming $n = 10, 15, \ldots, 100$ and $\phi = 2$ and $\lambda = 1.5$. We presented only the results related to ϕ, since the estimator of λ is the same using both approaches. It can be seen from the obtained results that both estimators' results in similar (although not the same) results.

Example 2: Let X_1, X_2, $\cdots$ be iid random variables following a Nakagami-m distribution

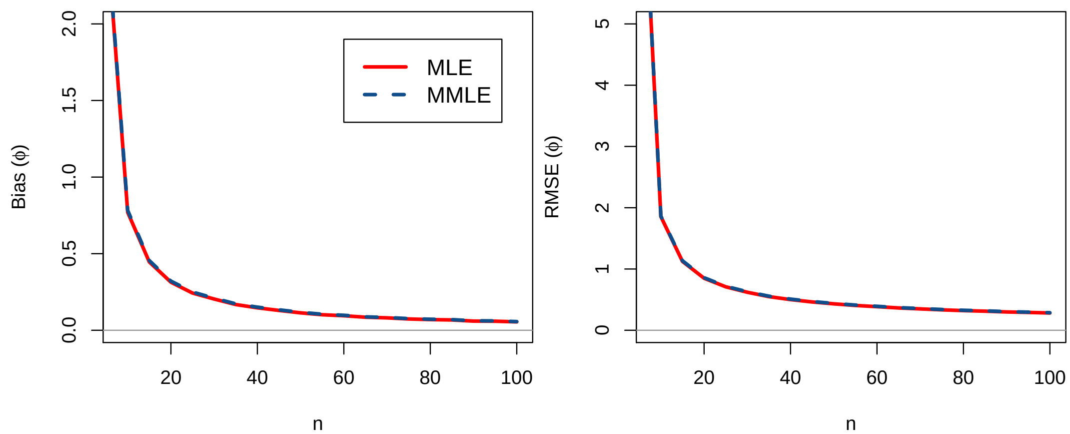

Figure 1: Bias and RMSE for ϕ for samples sizes of $10, 15, \ldots, 100$ elements considering $\phi = 2$ and $\lambda = 1.5$.

with PDF given by

$$f(x\,;\,\lambda,\phi) = \frac{2}{\Gamma(\phi)}\left(\frac{\phi}{\lambda}\right)^{\phi} t^{2\phi-1}\exp\left(-\frac{\phi}{\lambda}t^2\right),$$

for all $t > 0$, where $\phi > 0.5$ and $\lambda > 0$.

Once again letting g as in (3.6), just as in Example 1, following Ramos et al. (2020), as long as $X_1 = \cdots = X_n$ does not hold, it follows that $\sum_{i=1}^n X_i^2 \log\left(X_i^2\right) - \frac{1}{n}\sum_{i=1}^n X_i^2 \sum_{i=1}^n \log\left(X_i^2\right) \neq 0$, in which case the modified maximum likelihood equations for g over $\alpha = 2$ for $r = 1$ ans $s = 1$ has as only solution

$$\hat{\lambda}_n = \frac{1}{n}\sum_{i=1}^n X_i^2 \quad \text{and} \quad \hat{\phi}_n = \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i^2 \log\left(X_i^2\right) - \frac{1}{n}\sum_{i=1}^n X_i^2 \sum_{i=1}^n \log\left(X_i^2\right)}.$$

The estimator has a similar expression as that of the MMLEs of the Gamma distribution. Once again, we notice these estimators are strongly consistent and asymptotically normal:

Proposition 3. *$\hat{\phi}_n$ and $\hat{\lambda}_n$ are strongly consistent estimators for the true parameters ϕ and λ, and asymptotically normal with $\sqrt{n}\left(\hat{\lambda}_n - \lambda\right) \xrightarrow{D} N\left(0, \lambda^2/\phi\right)$ and $\sqrt{n}\left(\hat{\phi}_n - \phi\right) \xrightarrow{D} N\left(0, \phi^3\psi'(\phi+1) + \phi^2\right)$.*

Proof. The proof is analogous to that of Proposition 2. $\square$

Here, we also compare the proposed estimators with the standard MLE. In Figure 2 we present the Bias and RMSE obtained from $100,000$ replications assuming $n = 10, 15, \ldots, 100$ and $\phi = 4$ and $\lambda = 10$. We also presented only the results related to ϕ. It can be seen from the obtained results that both estimators returned very close estimates.

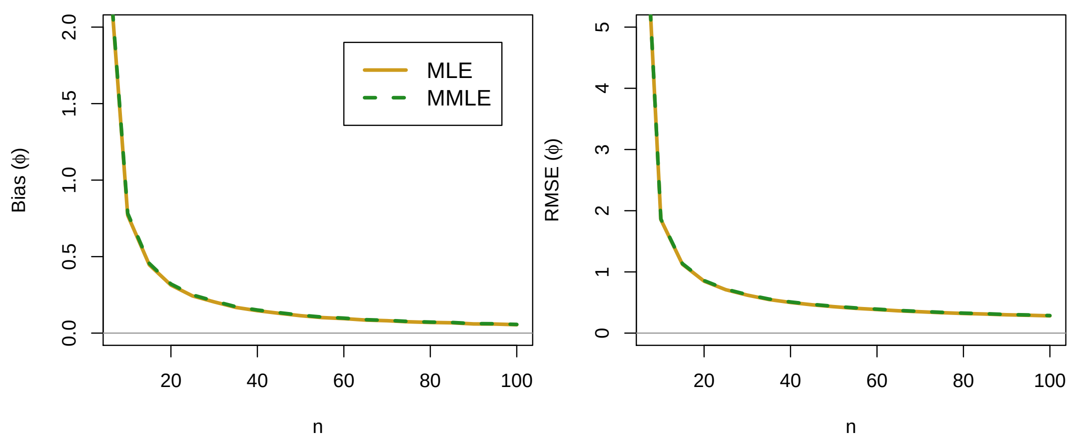

Figure 2: Bias and RMSE for ϕ for samples sizes of $10, 15, \ldots, 100$ elements considering $\phi = 4$ and $\lambda = 10$.

Note that the approach given above can be considered for other particular cases. For instance, the Wilson-Hilferty distributions are obtained when $\alpha = 3$. Hence, we can obtain closed-form estimators for cited distribution as well. It is essential to mention that, in the above examples, we do not claim that the GG distribution is the unique distribution to obtain the closed-form estimators for the Gamma and Nakagami. Different generalized distributions may lead to different closed-form estimators. The proposed approach is applied in a generalized version of the beta distribution that will return a closed-form estimator for both parameters.

Example 3: Let us assume that the chosen beta distribution has the PDF given by

$$f(x\,;\,\alpha,\beta)=\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}\quad 0<x<1, \tag{3.12}$$

where $\mathrm{B}(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the beta function, $\alpha>0$, $\beta>0$.

We can apply the modified maximum likelihood approach for this distribution by considering the function $g(x\,;\,\alpha,\beta,a,c)$ representing the generalized beta distribution, where $\alpha>2$, $\beta>2$ and $a<c$, given by:

$$g(x\,;\,\alpha,\beta,a,c)=\frac{(x-a)^{\alpha-1}\,(c-x)^{\beta-1}}{(c-a)^{\alpha+\beta-1}\,\mathrm{B}(\alpha,\beta)}\ \text{for all}\ x\in\mathcal{X}_{a,c}=\,]a,c[\,.$$

Notice $f(x\,;\,\alpha,\beta)=g(x\,;\,\alpha,\beta,0,1)$. Thus we can consider the modified likelihood equations for g over $(a,c)=(0,1)$ for $s=2$ and $r=2$, which are given by

$$\begin{aligned}
\sum_{i=1}^{n}\frac{\partial}{\partial a}\log g(X_i\,;\,\alpha,\beta,0,1)&=-(\alpha-1)\sum_{i=1}^{n}\frac{1}{X_i}\,+n(\alpha+\beta-1)=0,\ \text{and}\\
\sum_{i=1}^{n}\frac{\partial}{\partial c}\log g(X_i\,;\,\alpha,\beta,0,1)&=(\beta-1)\sum_{i=1}^{n}\frac{1}{1-X_i}\,-n(\alpha+\beta-1)=0.
\end{aligned}$$

Note that, from the harmonic-arithmetic inequality, as long as the equality $X_1=\cdots=X_n$ does not hold, we have $\sum_{i=1}^{n}\frac{1-X_i}{X_i}-\frac{n^2}{\sum_{i=1}^{n}\frac{X_i}{1-X_i}}>0$ and $\sum_{i=1}^{n}\frac{X_i}{1-X_i}-\frac{n^2}{\sum_{i=1}^{n}\frac{1-X_i}{X_i}}>0$, in which case, after some algebraic manipulations it is seem that the only solutions for the above system of linear equations are given by

$$\hat{\alpha}_n = \left(\sum_{i=1}^{n} \frac{1}{X_i}\right)\left(\sum_{i=1}^{n} \frac{1-X_i}{X_i} - \frac{n^2}{\sum_{i=1}^{n} \frac{X_i}{1-X_i}}\right)^{-1}. \tag{3.13}$$

$$\hat{\beta}_n = \left(\sum_{i=1}^{n} \frac{1}{1-X_i}\right)\left(\sum_{i=1}^{n} \frac{X_i}{1-X_i} - \frac{n^2}{\sum_{i=1}^{n} \frac{1-X_i}{X_i}}\right)^{-1}. \tag{3.14}$$

In the following, we prove that these estimators are consistent and asymptotically normal.

Proposition 4. *$\hat{\alpha}_n$ and $\hat{\beta}_n$ are strongly consistent estimators for the true parameters α and β, and asymptotically normal with $\sqrt{n}\left(\hat{\alpha}_n - \alpha\right) \xrightarrow{D} N\left(0, Q(\alpha,\beta)\right)$ and $\sqrt{n}\left(\hat{\beta}_n - \beta\right) \xrightarrow{D} N\left(0, Q(\beta,\alpha)\right)$, where*

$$Q(y,z) = \frac{y(y-1)^2(4yz^2 - 6z^2 - 10yz + 5y + 16z - 10)}{(y-2)(z-2)(y+z-1)} \text{ for all } y > 2 \text{ and } z > 2.$$

Proof. In order to apply Theorem 2 we notice $g(x\,;\,\alpha,\beta,a_0,c_0)$ can be written as

$$g(x\,;\,\alpha,\beta,a_0,c_0) = V(x)\exp\left[\eta_1(\alpha,\beta)\log(x-a_0) + \eta_2(\alpha,\beta)\log(c_0-x) + L(\alpha,\beta,a_0,c_0)\right]$$

for all $x \in \left]a_0, c_0\right[$, $\alpha > 2$, $\beta > 2$ and $a_0 < c_0$, where

$$V(x) = 1,\ \eta_1(\alpha,\beta) = \alpha - 1,\ \eta_2(\alpha,\beta) = \beta - 1 \text{ and}$$

$$L(\alpha,\beta,a_0,b_0) = -(\alpha+\beta-1)\log(c_0-a_0) - \log(\mathrm{B}(\alpha,\beta)).$$

In order to check condition (A) of Theorem 2 we notice that from the discussion above, the modified likelihood equations has $(\hat{\alpha}_n, \hat{\beta}_n)$ as its only solution unless $X_1 = X_2 = \cdots = X_n$, which has probability zero of occurring if $n \geq 2$. In order to check condition (B) of

Theorem 2 notice that for $a = a_0$ and $c = c_0$, following the computations of Aryal and Nadarajah (2004), we have

$$J(\alpha,\beta) = \begin{bmatrix} \frac{\beta}{(\alpha-1)} & -1 \\ 1 & -\frac{\alpha}{(\beta-1)} \end{bmatrix} \text{ and } K(\alpha,\beta) = \begin{bmatrix} \frac{\beta(\alpha+\beta-1)}{(\alpha-2)} & \alpha+\beta-1 \\ \alpha+\beta-1 & \frac{\alpha(\alpha+\beta-1)}{(\beta-2)} \end{bmatrix}. \tag{3.15}$$

Thus, since $\alpha + \beta - 1 > 0$, it is easy to see $J(\alpha,\beta)$ is invertible with

$$(J(\alpha,\beta))^{-1} = \begin{bmatrix} \frac{\alpha(\alpha-1)}{\alpha+\beta-1} & -\frac{(\alpha-1)(\beta-1)}{\alpha+\beta-1} \\ \frac{(\alpha-1)(\beta-1)}{\alpha+\beta-1} & -\frac{\beta(\alpha-1)}{\alpha+\beta-1} \end{bmatrix}.$$

Therefore we conclude condition (B) is satisfied, and after some algebraic computations one may find that

$$(J(\alpha,\beta)^{-1})^T K(\alpha,\beta) J(\alpha,\beta)^{-1} = \begin{bmatrix} Q(\alpha,\beta) & Q_1(\alpha,\beta) \\ Q_1(\alpha,\beta) & Q(\beta,\alpha) \end{bmatrix}. \tag{3.16}$$

where $Q(y,z)$ is as in the proposition and $Q_1(y,z)$ is a rational function on y and z. Condition (C) of Theorem 2 follows analogously as in the proof of Proposition 2, by noticing the relations:

$$\mathrm{E}_{\alpha,\beta}\left[\frac{1}{X_1}\right] = \frac{\alpha+\beta-1}{\alpha-1} \text{ and } \mathrm{E}_{\alpha,\beta}\left[\frac{1}{1-X_1}\right] = \frac{\alpha+\beta-1}{\beta-1}.$$

Finally, condition (D) of Theorem 2 follows by noticing that $\pi : (2,\infty)^2 \to (1,\infty)^2$ given by $\pi(\alpha,\beta) = (\eta_1(\alpha,\beta), \eta_2(\alpha,\beta)) = (\alpha-1, \beta-1)$ is a global diffeomorphism. Thus, we verified conditions (A) to (D) of Theorem 2, and therefore the proposition follows from the conclusion of Theorem 2 combined with (3.16). $\square$

Figure 3 provides the Bias and RMSE obtained from $100,000$ replications assuming $n = 10, 15, \ldots, 100$ and $\alpha = 3$ and $\beta = 2.5$. Here we considered the proposed estimator and compared with the standard MLE that does not have a closed-form expression.

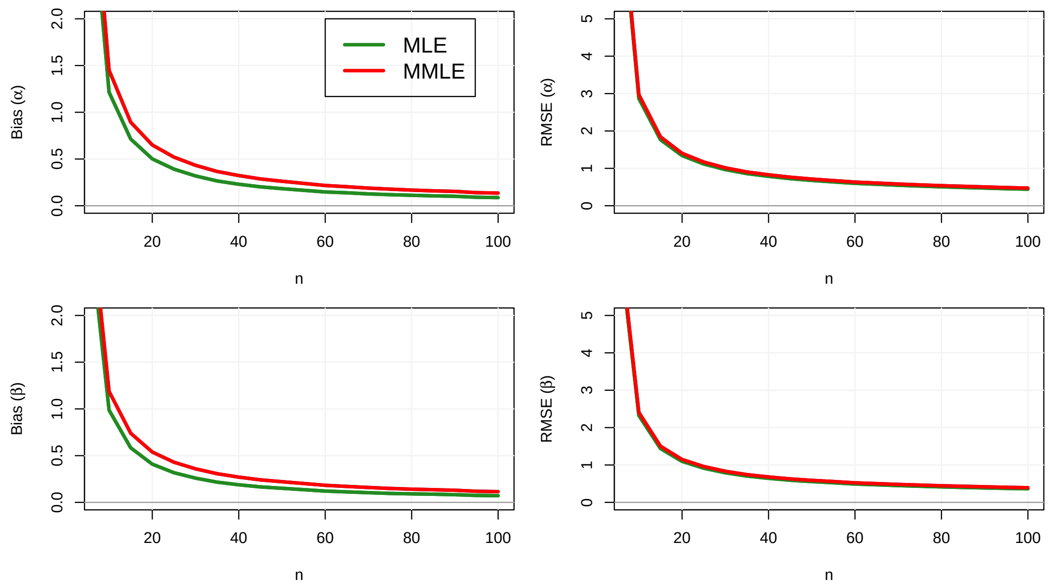

Figure 3: Bias and RMSE for α and β for samples sizes of $10, 15, \ldots, 100$ elements considering $\alpha = 3$ and $\beta = 2.5$.

Unlike the Gamma and Nakagami distributions, we observed that the closed-form estimator has an additional bias. Although they are obtained from different distributions for many parameter values, they returned similar results. A major drawback of the estimators (3.13) and (3.14) is that the properties that ensure the consistency and asymptotic normality do not hold when the values of α and β are smaller than 2.

4 Final Remarks

We have shown that the proposed modified version of the maximum likelihood estimators provides a vital alternative to achieve closed-form expressions when the standard MLE approach fails. The proposed approach can also be used with discrete distributions, the results remain valid, and the obtained estimators are still consistent, invariant, and is asymptotic normally distributed. Due to the likelihood function's flexibility, additional complexity can be included in the distribution and the inferential procedure such as censoring, long-term survival, covariates, and random effects.

The introduced method depends on the use of generalized versions of the baseline distribution. Hence, it provides another important motivation for applying many new distributions that have been introduced in the past few decades. Additionally, as the estimators are not unique and depend on the generalized distribution, comparisons between the different estimators are encouraged to find the one with better performance in terms of a specific metric.

As shown in Examples 1 and 2, the estimators' behaviors in terms of Bias and RMSE are similar to those obtained under the MLE for the Gamma and Nakagami distributions. Therefore, corrective bias approaches can also be used to remove the bias of the modified estimators. For the Beta distribution, the comparison showed different behavior for the proposed estimators. We observed that for specific small values of the parameters, the results might not be consistent. This example illustrates what happens in situations where, for some parameter values, the Fisher information of the generalized distribution has singularity problems. Although we have focus on three examples, our proposed approach can be extended beyond the examples and applied for a large number of distributions.

References

Aldrich, J. et al. (1997). Ra fisher and the making of maximum likelihood 1912-1922. *Statistical science 12*(3), 162–176.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society: Series B (Methodological) 32*(2), 283–301.

Anderson, J. and V. Blair (1982). Penalized maximum likelihood estimation in logistic regression and discrimination. *Biometrika 69*(1), 123–136.

Aryal, G. and S. Nadarajah (2004). Information matrix for beta distributions. *Serdica Mathematical Journal 30*(4), 513p–526p.

Bierens, H. J. (2004). *Introduction to the mathematical and statistical foundations of econometrics*. Cambridge University Press.

Cheng, R. and N. Amin (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society. Series B (Methodological)*, 394–403.

Cox, D. R. (1975). Partial likelihood. *Biometrika 62*(2), 269–276.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological) 39*(1), 1–22.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika 80*(1), 27–38.

Gourieroux, C., A. Monfort, and A. Trognon (1984). Pseudo maximum likelihood methods: Theory. *Econometrica: journal of the Econometric Society*, 681–700.

Hager, H. W. and L. J. Bain (1970). Inferential procedures for the generalized gamma distribution. *Journal of the American Statistical Association 65*(332), 1601–1609.

Hosking, J. R. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B (Methodological) 52*(1), 105–124.

Kao, J. H. (1958). Computer methods for estimating weibull parameters in reliability studies. *IRE Transactions on Reliability and Quality Control*, 15–22.

Kao, J. H. (1959). A graphical estimation of mixed weibull parameters in life-testing of electron tubes. *Technometrics 1*(4), 389–407.

Lehmann, E. L. and G. Casella (2006). *Theory of point estimation.* Springer Science & Business Media.

Lehmann, E. L. and J. P. Romano (2008). *Testing statistical hypotheses*. Springer Science & Business Media.

Louzada, F., P. L. Ramos, and E. Ramos (2019). A note on bias of closed-form estimators for the gamma distribution derived from likelihood equations. *The American Statistician 73*(2), 195–199.

Murphy, S. A. and A. W. Van der Vaart (2000). On profile likelihood. *Journal of the American Statistical Association 95*(450), 449–465.

Ramos, P. L., F. Louzada, and E. Ramos (2020). Bias reduction in the closed-form maximum likelihood estimator for the nakagami-m fading parameter. *IEEE Wireless Communications Letters 9*(10), 1692–1695.

Redner, R. et al. (1981). Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions. *Annals of Statistics 9*(1), 225–228.

Appendix

In order to prove Theorem 1 we shall need the lemma that follows, which answers us the question: If a sequence of functions F_n satisfy $\lim_{n\to\infty} F_n(\boldsymbol{\theta}_0) = 0$, under which conditions can we guarantee the existence of zeros $\boldsymbol{\theta}_n$ of F_n such that $\lim_{n\to\infty} \boldsymbol{\theta}_n = \boldsymbol{\theta}_0$? Notice this question is central when trying to prove the strong consistency of estimators.

In the following, given a differentiable function $F:\Theta \to \mathbb{R}^m$, for $\Theta \subset \mathbb{R}^m$ open, we denote by $F'(\boldsymbol{\theta}) \in M_m(\mathbb{R})$ the Jacobian of F at $\boldsymbol{\theta} \in \Theta$, that is, $F'(\boldsymbol{\theta}) = \left(\frac{\partial F_i}{\partial \theta_j}(\boldsymbol{\theta})\right) \in M_m(\mathbb{R})$ for all $\boldsymbol{\theta} \in \Theta$.

Lemma 1. *Let $\Theta \subset \mathbb{R}^m$ be open, let $F_i : \Theta \to \mathbb{R}^m$ be differentiable for all $i \in \mathbb{N}$, let $\boldsymbol{\theta}_0 \in \Theta$ and suppose that*

$$\lim_{n\to\infty} F_n(\boldsymbol{\theta}_0) = 0 \text{ and } \lim_{n\to\infty} F_n'(\boldsymbol{\theta}) = G(\boldsymbol{\theta}) \text{ uniformly on } \Theta,$$

where $G : \Theta \to M_m(\mathbb{R})$ is continuous at $\boldsymbol{\theta}_0$ and $G(\boldsymbol{\theta}_0)$ is invertible. Then there exists $N > 0$ and a sequence $\{\boldsymbol{\theta}_n\}_{n\geq N}$ in Θ such that

$$\lim_{n\to\infty} \boldsymbol{\theta}_n = \boldsymbol{\theta}_0 \text{ and } F_n(\boldsymbol{\theta}_n) = 0 \text{ for all } n \geq N, \tag{.17}$$

Proof. Denote $\lambda = \|G(\boldsymbol{\theta}_0)^{-1}\| > 0$. Since Θ is open and G is continuous at $\boldsymbol{\theta}_0$, there exists $\delta > 0$ such that

$$B(\boldsymbol{\theta}_0, \delta) \subset \Theta \text{ and } \|G(\boldsymbol{\theta}) - G(\boldsymbol{\theta}_0)\| < \frac{1}{4\lambda} \text{ for all } \boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \delta). \tag{.18}$$

Moreover, due to the hypothesis, it follows that there exists $N > 0$ such that $n \geq N$ implies in

$$\|F_n(\boldsymbol{\theta}_0)\| < \frac{\delta}{2\lambda} \text{ and } \|F_n'(\boldsymbol{\theta}) - G(\boldsymbol{\theta})\| < \frac{1}{4\lambda} \text{ for all } \boldsymbol{\theta} \in \Theta. \tag{.19}$$

Notice the inequality in (.18) combined with the second inequality in (.19) and the triangle inequality implies that

$$\|F_n'(\boldsymbol{\theta}) - G(\boldsymbol{\theta}_0)\| < \frac{1}{2\lambda} \text{ for all } \boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \delta) \text{ and } n \geq N. \tag{.20}$$

Moreover, if we denote $\epsilon_n = 2\lambda \|F_n(\boldsymbol{\theta}_0)\|$ for all $n \geq N$, from the first inequality in (.19) it follows that $\epsilon_n < \delta$ for all $n \geq N$, which combined with $B(\boldsymbol{\theta}_0, \delta) \subset \Theta$ implies in

$$\bar{B}(\boldsymbol{\theta}_0, \epsilon_n) \subset \Theta \text{ for all } n \geq N.$$

Now, given $n \geq N$ and letting $L_n : \Theta \to \mathbb{R}^m$ be defined by $L_n(\boldsymbol{\theta}) = \boldsymbol{\theta} - G(\boldsymbol{\theta}_0)^{-1}F_n(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$ we shall prove that $L_n(\bar{B}(\boldsymbol{\theta}_0, \epsilon_n)) \subset \bar{B}(\boldsymbol{\theta}_0, \epsilon_n)$. Indeed, notice that from the chain rule L_n is differentiable in Θ with

$$L_n'(\boldsymbol{\theta}) = I - G(\boldsymbol{\theta}_0)^{-1}F_n'(\boldsymbol{\theta}) = G(\boldsymbol{\theta}_0)^{-1}\left(G(\boldsymbol{\theta}_0) - F_n'(\boldsymbol{\theta})\right) \text{ for all } \boldsymbol{\theta} \in \Theta.$$

Thus for all $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \delta)$ we have from the inequality (.20) that

$$\|L_n'(\boldsymbol{\theta})\| = \left\|G(\boldsymbol{\theta}_0)^{-1}\left(G(\boldsymbol{\theta}_0) - F_n'(\boldsymbol{\theta})\right)\right\| \leq \left\|G(\boldsymbol{\theta}_0)^{-1}\right\| \|G(\boldsymbol{\theta}_0) - F_n'(\boldsymbol{\theta})\| < \frac{1}{2},$$

which, due to the mean value inequality, implies in

$$\|L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}_0)\| \leq \frac{1}{2}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \text{ for all } \boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \delta). \tag{.21}$$

Moreover, notice that

$$\|L_n(\boldsymbol{\theta}_0) - \boldsymbol{\theta}_0\| = \left\|G(\boldsymbol{\theta}_0)^{-1}F_n(\boldsymbol{\theta}_0)\right\| \leq \left\|G(\boldsymbol{\theta}_0)^{-1}\right\| \|F_n(\boldsymbol{\theta}_0)\| = \frac{\epsilon_n}{2}. \tag{.22}$$

Thus, given $\boldsymbol{\theta} \in \bar{B}(\boldsymbol{\theta}_0, \epsilon_n)$ from the triangle inequality and the inequalities (.21) and (.22) we have

$$\|L_n(\boldsymbol{\theta}) - \boldsymbol{\theta}_0\| \leq \|L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}_0)\| + \|L_n(\boldsymbol{\theta}_0) - \boldsymbol{\theta}_0\| \leq \frac{\epsilon_n}{2} + \frac{\epsilon_n}{2} = \epsilon_n,$$

that is, $L_n(\boldsymbol{\theta}) \in \bar{B}(\boldsymbol{\theta}_0, \epsilon_n)$, which proves that $L_n(\bar{B}(\boldsymbol{\theta}_0, \epsilon_n) \subset \bar{B}(\boldsymbol{\theta}_0, \epsilon_n)$.

Now, since L_n is continuous and since we proved that $L_n(\bar{B}(\boldsymbol{\theta}_0, \epsilon_n) \subset \bar{B}(\boldsymbol{\theta}_0, \epsilon_n)$, from the Browder fixed point theorem we conclude L_n has at least one fixed point $\boldsymbol{\theta}_n$ in $\bar{B}(\boldsymbol{\theta}_0, \epsilon_n)$ for all $n \geq N$, and thus

$$L_n(\boldsymbol{\theta}_n) = \boldsymbol{\theta}_n \Rightarrow G(\boldsymbol{\theta}_0)^{-1} F_n(\boldsymbol{\theta}_n) = 0 \Rightarrow F_n(\boldsymbol{\theta}_n) = 0 \text{ for all } n \geq N.$$

Moreover, notice that by hypothesis $\lim_{n\to\infty} \epsilon_n = \lim_{n\to\infty} 2\lambda \left\| F_n(\boldsymbol{\theta}_0) \right\| = 0$, and since $\boldsymbol{\theta}_n \in \bar{B}(\boldsymbol{\theta}_0, \epsilon_n)$ for $n \geq N$, it follows that $\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\| \leq \epsilon_n$ for all $n \geq N$ and thus

$$\lim_{n\to\infty} \|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\| = 0 \Rightarrow \lim_{n\to\infty} \boldsymbol{\theta}_n = \boldsymbol{\theta}_0,$$

which concludes the proof. □

Using the above lemma we are ready to prove Theorem 1.

Proof. **Strong Consistency:**

During this proof we shall always denote $(\beta_1, \cdots, \beta_s) = (\theta_1, \cdots, \theta_{s-r}, \alpha_1, \cdots, \alpha_r)$. By hypothesis (A) there exists a set Ω_0 of probability one in Ω such that $\hat{\boldsymbol{\theta}}(\boldsymbol{X}(w))$ is the unique solution of the modified maximum likelihood equations for all $w \in \Omega_0$. Now, letting $F_n : \Theta \times \Omega \to \mathbb{R}^s$ be defined by $F_n = (F_{n1}, \cdots, F_{ns})$ where

$$F_{nj}(\boldsymbol{\theta}, w) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \beta_j} \log\, g\left(X_i(w)\, ;\, \boldsymbol{\theta}, \boldsymbol{\alpha}_0\right),$$

for all $w \in \Omega$, $\boldsymbol{\theta} \in \Theta$ and $1 \leq j \leq s$, notice due to the strong law of the large numbers and hypothesis (C) that

$$\lim_{n\to\infty} F_{nj}(\boldsymbol{\theta}_0, w) = E_{\boldsymbol{\theta}_0}\left[-\frac{\partial}{\partial \beta_j} \log\, g\left(X_1\, ;\, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0\right)\right] = 0 \text{ a.e. in } \Omega,$$

or, in other words, there exists a set Ω_1 of probability one in Ω such that

$$\lim_{n\to\infty} F_n(\boldsymbol{\theta}_0, w) = 0 \text{ for all } w \in \Omega_1. \tag{.23}$$

Now, letting $G : \Theta \to M_s(\mathbb{R})$ be defined by $G(\boldsymbol{\theta}) = (G_{i,j}(\boldsymbol{\theta})) \in M_s(\mathbb{R})$ where

$$G_{i,j}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}_0}\left[-\frac{\partial^2 \log\, g(X_1\,;\,\boldsymbol{\theta},\boldsymbol{\alpha}_0)}{\partial\theta_i\beta_j}\right],$$

notice from condition (D) and the dominated convergence theorem that $G(\boldsymbol{\theta})$ is continuous at $\boldsymbol{\theta}_0$. Moreover, from condition (D) and the uniform strong law of the large numbers it follows that

$$\lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\bar{\Theta}_0} \left|\frac{\partial F_{nj}(\boldsymbol{\theta}, w)}{\partial\theta_i} - G_{i,j}(\boldsymbol{\theta})\right| = 0, \text{ a.e. in } \Omega,$$

for all i and j, or, in other words, there exists a set Ω_2 of probability one in Ω such that

$$\lim_{n\to\infty} \frac{\partial F_n(\boldsymbol{\theta}, w)}{\partial\theta_i} = G(\boldsymbol{\theta}) \text{ uniformly on } \Theta_0 \text{ for all } i \text{ and } w \in \Omega_2.$$

Combining the obtained results with the fact that $G(\boldsymbol{\theta}_0) = J(\boldsymbol{\theta}_0)$ is invertible by hypothesis (B) it follows that, for each fixed $w \in \Omega_1 \cap \Omega_2$, the sequence of functions $F_n(\boldsymbol{\theta}, w)$ satisfies Lemma 2.1 and thus we conclude that for each $w \in \Omega_1 \cap \Omega_2$, there exists $N(w) > 1$ and a sequence $(\boldsymbol{\theta}_n(w))_{n\geq N(w)}$ in Θ such that

$$F_n(\boldsymbol{\theta}_n(w), w) = 0 \text{ and } \lim_{n\to\infty} \boldsymbol{\theta}_n(w) = \boldsymbol{\theta}_0 \text{ for all } w \in \Omega_1 \cap \Omega_2 \text{ and } n \geq N(w). \tag{.24}$$

Hence, from (.24) we conclude that $\hat{\boldsymbol{\theta}}_n(\boldsymbol{X}(w)) = \boldsymbol{\theta}_n(w)$ for all $w \in \Omega_0 \cap \Omega_1 \cap \Omega_2$ and $n \geq N(w)$, which combined with (.24) implies in

$$\lim_{n\to\infty} \hat{\boldsymbol{\theta}}_n(\boldsymbol{X}(w)) = \lim_{n\to\infty} \boldsymbol{\theta}_n(w) = \boldsymbol{\theta}_0 \text{ for all } w \in \Omega_0 \cap \Omega_1 \cap \Omega_2.$$

Thus, since $\Omega_0 \cap \Omega_1 \cap \Omega_2$ is a set of probability one in Ω, it follows by definition that $\hat{\boldsymbol{\theta}}_n(\boldsymbol{X})$ is a strongly consistent estimator for $\boldsymbol{\theta}_0$.

Asymptotic normality:

First notice that, denoting $l(\boldsymbol{\theta}, \boldsymbol{\alpha}\,;\,\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{s} g(X_i\,;\,\boldsymbol{\theta}, \boldsymbol{\alpha})$, from the mean value theorem, for each fixed $1 \le i \le s$, there must exist $\boldsymbol{y}_i(\boldsymbol{X}) \in \mathbb{R}^s$ contained in the segment connecting $\boldsymbol{\theta}_0$ to $\hat{\boldsymbol{\theta}}_n(\boldsymbol{X})$ such that

$$\frac{\partial}{\partial \beta_i} l\left(\hat{\boldsymbol{\theta}}_n(\boldsymbol{X}), \boldsymbol{\alpha}_0\,;\,\boldsymbol{X}\right) = \frac{\partial}{\partial \beta_i} l\left(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0\,;\,\boldsymbol{X}\right) + \sum_{j=1}^{s} \frac{\partial l(\boldsymbol{y}_i(\boldsymbol{X}), \boldsymbol{\alpha}_0\,;\,\boldsymbol{X})}{\partial \theta_j \partial \beta_i}(\hat{\theta}_{nj}(\boldsymbol{X}) - \theta_{0j}),$$

which, combined with the fact that by hypothesis $\dfrac{\partial}{\partial \beta_i} l\left(\hat{\boldsymbol{\theta}}_n(\boldsymbol{X}), \boldsymbol{\alpha}_0\,;\,\boldsymbol{X}\right) = 0$ a.e. in Ω, implies in

$$-\sum_{j=1}^{s} \frac{\partial\, l(\boldsymbol{y}_i(\boldsymbol{X}), \boldsymbol{\alpha}_0\,;\,\boldsymbol{X})}{\partial \theta_j \partial \beta_i}(\hat{\theta}_{nj}(\boldsymbol{X}) - \theta_{0j}) = \frac{\partial}{\partial \beta_i} l\left(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0\,;\,\boldsymbol{X}\right) \text{ a.e. in } \Omega.$$

Thus, letting $A_n(\boldsymbol{X})$ be defined by $A_n(\boldsymbol{X}) = (a_{ij}(\boldsymbol{X})) \in M_s(\mathbb{R})$ where $a_{ij}(\boldsymbol{X}) = -\frac{\partial\, l(\boldsymbol{y}_i(\boldsymbol{X}), \boldsymbol{\alpha}_0\,;\,\boldsymbol{X})}{\partial \theta_i \partial \beta_j}$ for all $1 \le i \le s$ and $1 \le j \le s$, and letting $v_n(\boldsymbol{X}) = \left(\frac{\partial}{\partial \beta_1} l\left(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0\,;\,\boldsymbol{X}\right), \cdots, \frac{\partial}{\partial \beta_s} l\left(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0\,;\,\boldsymbol{X}\right)\right)^T$, the equation above can be rewritten as

$$A_n(\boldsymbol{X})^T(\hat{\boldsymbol{\theta}}_n(\boldsymbol{X}) - \boldsymbol{\theta}_0)^T = v_n(\boldsymbol{X}) \text{ a.e. in } \Omega.$$

Now, since we already proved that $\hat{\boldsymbol{\theta}}_n(\boldsymbol{X})$ converges strongly to $\boldsymbol{\theta}_0$, it follows that $\boldsymbol{y}_i(\boldsymbol{X})$ converges strongly to $\boldsymbol{\theta}_0$ as well, and thus it follows from the strong law of large numbers that $A_n(\boldsymbol{X})$ converges strongly to $J(\boldsymbol{\theta}_0)$. Additionally, from the central limit theorem we know that $v_n(\boldsymbol{X})$ converges in distribution to $N_s(0, K(\boldsymbol{\theta}_0))$. Thus, from Lemma 5.2

of Lehman (see Lehmann and Casella (2006)) it follows that $(\hat{\boldsymbol{\theta}}_n(\boldsymbol{X}) - \boldsymbol{\theta}_0)^T$ converges in distribution to

$$(J(\boldsymbol{\theta}_0)^T)^{-1} N_s(0, K(\boldsymbol{\theta}_0)) = N_s\left(0, (J(\boldsymbol{\theta}_0)^{-1})^T K(\boldsymbol{\theta}_0) J(\boldsymbol{\theta}_0)^{-1}\right),$$

which concludes the proof. □

We now proceed to prove Theorem 2. The proof shall be a consequence of the following lemma, which gives us some properties such as differentiation under the integral sign for exponential maps.

Lemma 2. *Let $h(x\,;\,\boldsymbol{\theta})$ be a stictly positive function satisfying*

$$h(x\,;\,\boldsymbol{\theta}) = V(x) \exp\left(\sum_{i=1}^{s} \eta_i(\boldsymbol{\theta}) T_i(X) - L(\boldsymbol{\theta})\right) \text{ for all } x \in \mathcal{X},\ \boldsymbol{\theta} \in \Theta,$$

where η_i and L are C^2, V and T_i are measurable on x, and $h(x\,;\,\boldsymbol{\theta})$ is integrable for all $\boldsymbol{\theta} \in \Theta$, and suppose that $\pi : \Theta \to \mathbb{R}^s$ given by $\pi(\boldsymbol{\theta}) = (\eta_1(\boldsymbol{\theta}), \cdots, \eta_s(\boldsymbol{\theta}))$ is a local diffeomorphism. Then there exists an open set Θ_0 containing $\boldsymbol{\theta}_0$ such that $\overline{\Theta}_0 \subset \Theta$, and measurable functions $M_{i,j}(x)$ such that

$$\left|\frac{\partial^2 \log h(x\,;\,\boldsymbol{\theta})}{\partial\theta_i\theta_j}\right| \leq M_{i,j}(x) \text{ for all } x \in \mathcal{X},\ \boldsymbol{\theta} \in \overline{\Theta}_0 \text{ and } \int_{\mathcal{X}} M_{i,j}(x) h(x\,;\,\boldsymbol{\theta_0})\, dx < \infty. \tag{.25}$$

Moreover, $\int_{\mathcal{X}} h(x\,;\,\boldsymbol{\theta})\, dx$ can be differentiated in relation to θ_i under the integral sign for all i and $\boldsymbol{\theta} \in \Theta_0$.

Proof. In order to prove the lemma we shall first prove that

$$\int_{\mathcal{X}} |T_i(x)|\, h(x\,;\,\boldsymbol{\theta}_0)\, dx < \infty \text{ for all } 1 \leq i \leq s. \tag{.26}$$

Indeed, from π being a local diffeomorphism it follows that there exists a neighborhood Θ_0 of $\boldsymbol{\theta}_0$ such that $\overline{\Theta}_0 \subset \Theta$ and such that $\pi^* : \Theta_0 \to V$ is a diffeomorphism for some open set $V \subset \mathbb{R}^s$, where π^* is the restriction of π to Θ_0. Thus, letting h^* be defined by

$$h^*(x\,;\,\boldsymbol{\lambda}) = V(x) \exp\left(\sum_{i=1}^{s} \lambda_i T_i(x)\,(\boldsymbol{\lambda}))\right) \text{ for all } x \in \mathcal{X} \text{ and } \boldsymbol{\lambda} \in V,$$

it follows that

$$h^*(x\,;\,\boldsymbol{\lambda}) = \exp\left[L((\pi^*)^{-1}(\boldsymbol{\lambda}))\right] h(x\,;\,(\pi^*)^{-1}(\boldsymbol{\lambda})) \text{ for all } x \in \mathcal{X} \text{ and } \boldsymbol{\lambda} \in V$$

and thus, from the integrability of $h(x\,;\,\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$ it follows that $h^*(x\,;\,\boldsymbol{\lambda})$ is integrable for all $\boldsymbol{\lambda} \in V$. Now, letting $\epsilon > 0$ be small enough such that $B(\pi(\boldsymbol{\theta}_0), 2\epsilon) \subset V$, since for all real $d \geq 0$ we have

$$|d| = \frac{1}{\epsilon}\,|\epsilon d| \leq \frac{1}{\epsilon} \exp(\epsilon\,|d|) \leq \frac{1}{\epsilon}\left(\exp(\epsilon d) + \exp(-\epsilon d)\right),$$

it follows that

$$|T_i(x)|\, h(x\,;\,\boldsymbol{\theta}_0) \leq \frac{1}{\epsilon}\left(\exp(\epsilon T_i(x)) + \exp(-\epsilon T_i(x))\right) h(x\,;\,\boldsymbol{\theta}_0), \tag{.27}$$

and since

$$\exp(\epsilon T_i(x))h(x\,;\,\boldsymbol{\theta}_0) = V(x) \exp\left(\epsilon \sum_{i=1}^{s} \lambda_i T_i(x) + L(\boldsymbol{\theta}_0)\right) = \exp(L(\boldsymbol{\theta}_0))h^*(x\,;\,\boldsymbol{\lambda}^*),$$

where $\boldsymbol{\lambda}^* \in B(\pi(\boldsymbol{\theta}_0), 2\epsilon)$ is defined by $\lambda_i^* = \eta_i(\boldsymbol{\theta}_0) + \epsilon$ and $\lambda_j^* = \eta_i(\boldsymbol{\theta}_0)$ for $j \neq i$, it follows that $\exp(\epsilon T_i(x))h(x\,;\,\boldsymbol{\theta}_0)$ is integrable and, analogously, one can prove that $\exp(-\epsilon T_i(x))h(x\,;\,\boldsymbol{\theta}_0)$ is integrable as well. Thus, from (.27) we conclude that $|T_i(x)h(x\,;\,\boldsymbol{\theta}_0)|$ is bounded from above by an integrable function, and thus is integrable as well, which proves (.26) is valid.

Now, letting $M_{i,j}(x) = \sum_{k=1}^{s} \max_{\boldsymbol{\theta}\in\overline{\Theta}_0} \left|\frac{\partial^2 \eta_i(\boldsymbol{\theta})}{\partial\theta_k \partial\theta_j}\right| |T_k(x)| + \max_{\boldsymbol{\theta}\in\overline{\Theta}_0} \left|\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right|$ for $1 \le i \le s$, $1 \le j \le s$ and $1 \le k \le s$, it follows that

$$\left|\frac{\partial^2 \log h(x\,;\,\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right| = \sum_{k=1}^{s} \left|\frac{\partial^2 \eta_k(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j} T_k(x) + \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right| \le M_j(x) \text{ for all } i,\ j,\ k \text{ and } x \in \mathcal{X},$$

and moreover from (.26) we have

$$\int_{\mathcal{X}} M_{i,j}(x) h(x\,;\,\boldsymbol{\theta}_0)\, dx = \sum_{k=1}^{s} \max_{\boldsymbol{\theta}\in\overline{\Theta}_0} \left|\frac{\partial^2 \eta_i(\boldsymbol{\theta})}{\partial\theta_k\partial\theta_j}\right| \int_{\mathcal{X}} |T_k(x)|\, h(x\,;\,\boldsymbol{\theta}_0)\, dx + \max_{\boldsymbol{\theta}\in\overline{\Theta}_0} \left|\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right| < \infty,$$

for all i and j, which proves (.25) is valid.

To prove $\int_{\mathcal{X}} h(x\,;\,\boldsymbol{\theta})\, dx$ can be differentiated under the integral sign at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, first notice that, from the classical theory of exponential maps (see Theorem 2.7.1 of Lehmann and Romano (2008)) we know the map $\rho : V \to \mathbb{R}$ defined by $\rho(\boldsymbol{\lambda}) = \int_{\mathcal{X}} h^*(x\,;\,\boldsymbol{\lambda})\, dx$ is differentiable in V with

$$\frac{\partial\rho(\boldsymbol{\lambda})}{\partial\lambda_i} = \int_{\mathcal{X}} \frac{\partial h^*(x\,;\,\boldsymbol{\lambda})}{\partial\lambda_i}\, dx = \int_{\mathcal{X}} T_i(x) h^*(x\,;\boldsymbol{\lambda})\, dx \text{ for all } i \text{ and } \boldsymbol{\lambda} \in V.$$

Thus, since $\int_{\mathcal{X}} h(x\,;\,\boldsymbol{\theta})\, dx = \rho(\pi^*(\boldsymbol{\theta})) \exp(L(\boldsymbol{\theta}))$ for all $\boldsymbol{\theta} \in \Theta_0$, we conclude from the chain rule that $\int_{\mathcal{X}} h(x\,;\,\boldsymbol{\theta})\, dx$ is differentiable in Θ_0 with

$$\begin{aligned}
\frac{\partial \int_{\mathcal{X}} h(x\,;\,\boldsymbol{\theta})\, dx}{\partial\theta_j} &= \left(\sum_{i=1}^{s} \frac{\partial\eta_i(\boldsymbol{\theta})}{\partial\theta_j}\frac{\partial\rho(\pi^*(\boldsymbol{\theta}))}{\partial\lambda_i} + \frac{\partial L(\boldsymbol{\theta})}{\partial\theta_j}\rho(\pi^*(\boldsymbol{\theta}))\right) \exp(L(\boldsymbol{\theta})) \\
&= \int_{\mathcal{X}} \left(\sum_{i=1}^{s} \frac{\partial\eta_i(\boldsymbol{\theta})}{\partial\theta_j} T_i(x) + \frac{\partial L(\boldsymbol{\theta})}{\partial\theta_j}\right) h(x\,;\boldsymbol{\theta})\, dx = \int_{\mathcal{X}} \frac{\partial h(x\,;\,\boldsymbol{\theta})}{\partial\theta_j}\, dx,
\end{aligned}$$

which concludes the proof. □

Using the above lemma, we can now prove Theorem 2.

Proof. We shall prove conditions (A) to (D) of Theorem 1 are satisfied. While it is clear from the hypothesis that conditions (A) and (B) of Theorem 1 are valid, to prove condition (C) of Theorem 1 hold we use Lemma 2 to differentiate the equation

$$\int_{\mathcal{X}} g(x\,;\,\boldsymbol{\theta},\boldsymbol{\alpha_0})\,dx = 1$$

in relation to θ_j under the integral sign at $\boldsymbol{\theta}=\boldsymbol{\theta}_0$ to find

$$E_{\boldsymbol{\theta}_0}\left[\frac{\partial\,\log g(X_1\,;\,\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)}{\partial\theta_j}\right] = \int_X \frac{\partial g(x\,;\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)}{\partial\theta_j}\,dx = 0,$$

which together with the hypothesis proves item (C) is valid. To prove item (D) of Theorem 1 is satisfied we first apply Lemma 2 to g to conclude there exist $M_{i,j}(x)$, $1\le i\le s$, $1\le j\le s$ and an open set Θ_0 containing $\boldsymbol{\theta}_0$ such that $\overline{\Theta}_0\subset\Theta$, and moreover

$$\left|\frac{\partial^2\,\log\,g(X_1\,;\,\boldsymbol{\theta},\boldsymbol{\alpha}_0)}{\partial\theta_i\partial\theta_j}\right| \le M_{ij}(X_1) \text{ and } E_{\boldsymbol{\theta}_0}\left[M_{ij}(X_1)\right]<\infty, \tag{.28}$$

for all i, j and $\boldsymbol{\theta}\in\Theta^*$. On the other hand, since π is by hypothesis a local diffeomorphism, given $1\le k\le s$ and denoting $\pi(\boldsymbol{\theta}_0)=\boldsymbol{\lambda}_0=(\lambda_{01},\cdots,\lambda_{0s})$, there must exist $\epsilon>0$ and $\boldsymbol{\theta}_{(k)}\in\Theta$ such that $\pi(\boldsymbol{\theta}_{(k)})=(\lambda_{01},\cdots,\lambda_{0k}+\epsilon,\cdots,\lambda_{0s})$. Therefore it follows that

$$\frac{1}{\epsilon}\left(\frac{\partial\,\log\,g(x\,;\,\boldsymbol{\theta}_{(k)},\boldsymbol{\alpha}_0)}{\partial\alpha_j} - \frac{\partial L(\boldsymbol{\theta}_{(k)},\boldsymbol{\alpha}_0)}{\partial\alpha_j} - \frac{\partial\,\log\,g(x\,;\,\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)}{\partial\alpha_j} + \frac{\partial L(\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)}{\partial\alpha_j}\right) = \frac{\partial\,T_k(x,\boldsymbol{\alpha}_0)}{\partial\alpha_j} \tag{.29}$$

and thus, since by the hypothesis in (C) the functions $\dfrac{\partial\,\log\,g(x\,;\,\boldsymbol{\theta}_{(k)},\boldsymbol{\alpha}_0)}{\partial\alpha_j}g(x\,;\,\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)$ and $\dfrac{\partial\,\log\,g(x\,;\,\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)}{\partial\alpha_j}g(x\,;\,\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)$ are integrable for all j and k, we conclude from (.29) that $\frac{\partial\,T_k(x,\boldsymbol{\alpha}_0)}{\partial\alpha_j}g(x\,;\,\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)$ is integrable as well for all j and k. Finally, letting

$$N_{i,j}(x) = \sum_{k=1}^{s}\sup_{\boldsymbol{\theta}\in\overline{\Theta}_0}\left|\frac{\partial\eta_k(\boldsymbol{\theta})}{\partial\theta_i}\right|\left|\frac{\partial T_k(x)}{\partial\alpha_j}\right| + \sup_{\boldsymbol{\theta}\in\overline{\Theta}_0}\left|\frac{\partial^2 L(\boldsymbol{\theta},\boldsymbol{\alpha}_0)}{\partial\theta_i\partial\alpha_j}\right|,$$

for $1 \leq i \leq s$ and $1 \leq j \leq r$ it follows that

$$\left|\frac{\partial^2 \log g(X_1\,;\,\boldsymbol{\theta},\boldsymbol{\alpha}_0)}{\partial\theta_i\partial\alpha_j}\right| = \left|\sum_{k=1}^{s}\frac{\partial\eta_k(\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial T_k(X_1)}{\partial\alpha_j} + \frac{\partial^2 L(\boldsymbol{\theta},\boldsymbol{\alpha_0})}{\partial\theta_i\partial\alpha_j}\right| \leq N_{i,j}(X_1),$$

and moreover from the integrability of $\frac{\partial T_k(x,\boldsymbol{\alpha}_0)}{\partial\alpha_j}g(x\,;\,\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)$ for all j and k we have

$$\mathrm{E}_{\boldsymbol{\theta}}\left[N_{i,j}(X_1)\right] = \int_{\mathcal{X}} N_{i,j}(x)g(x\,;\,\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0) < \infty,$$

for all i and j, which combined with (.28) proves that item (D) of Theorem 1 is satisfied.

Thus we verified items (A) to (D) of Theorem 1, and therefore the conclusions of Theorem 2 follow from the conclusions of Theorem 1. $\square$