Maximum integrated likelihood estimator of the interest parameter when the nuisance parameter is location or scale

A. Zaigraev*, A. Podraza-Karakulska
Nicolaus Copernicus University of Toruń, Poland

A R T I C L E   I N F O

Article history:
Received 2 September 2013
Received in revised form 13 January 2014
Accepted 23 January 2014
Available online 10 February 2014

MSC:
62F10

Keywords:
Maximum integrated likelihood estimator
Maximum likelihood estimator
Bias
Mean squared error

A B S T R A C T

The problem of estimation of an interest parameter in the presence of a nuisance parameter, which is either location or scale, is studied. Two estimators are considered: the usual maximum likelihood estimator and the estimator based on maximization of the integrated likelihood function. The estimators are compared, asymptotically, with respect to the bias and with respect to the mean squared error. The examples are given.

1. Introduction

Let a sample \( x = (x_1, x_2, \ldots, x_n) \) be drawn from an absolutely continuous distribution with the probability density function \( p(\cdot; \theta, \lambda) \). The parameters \( \theta \) and \( \lambda \) are assumed to be unknown: \( \theta \) is the interest parameter and \( \lambda \) is the nuisance parameter. We consider the case when \( \lambda \) is either location or scale.

To estimate the parameters one can adopt the well-known maximum likelihood (ML) method. It consists in taking the likelihood function

\[
L(\theta, \lambda; x) := \prod_{j=1}^{n} p(x_j; \theta, \lambda)
\]

and searching for the maximum likelihood estimators (MLE) \( \hat{\theta}(x), \hat{\lambda}(x) \) defined as \( (\hat{\theta}(x), \hat{\lambda}(x)) \in \text{Arg sup}_{(\theta, \lambda)} L(\theta, \lambda; x) \).

Recently different modifications of the ML method have been forced in the literature: see e.g. Berger et al. (1999) or Severini (2000, 2007, 2010) and references therein. Among those modifications there is that based on integration of the likelihood function, with a weight function, over the nuisance parameter. In general, the integrated likelihood function is written as

\[
\hat{L}(\theta; x) := \int L(\theta, \lambda; x) w(\lambda) d\lambda,
\]

where \( w(\cdot) \) is the weight function. Using Bayesian language, the weight function can be interpreted as a conditional a priori density corresponding to \( \lambda \) given \( \theta \) (i.e. \( w(\lambda) \propto \pi(\lambda | \theta) \)).

* Correspondence to: Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Chopin str. 12/18, 87-100 Toruń, Poland.
E-mail address: alzaig@mat.umk.pl (A. Zaigraev).

http://dx.doi.org/10.1016/j.spl.2014.01.024
0167-7152/© 2014 Elsevier B.V. All rights reserved.
So, the integrated likelihood automatically incorporates a nuisance parameter uncertainty by averaging over all \( \lambda \) (instead of calculation \( \hat{\lambda}(x; \theta) \in \text{Arg sup}_\lambda L(\theta; \lambda; x) \), the conditional MLE of \( \lambda \) given \( \theta \)), and then maximization of the profile likelihood function \( \tilde{L}(\theta; x) := L(\theta; \hat{\lambda}(x; \theta) | x) \) to obtain \( \theta^*(x) \).

We call \( \theta^*(x) \in \text{Arg sup}_\theta L(\theta; x) \) the maximum integrated likelihood estimator (MILE) of \( \theta \). What is the motivation of using this estimator, apart from the facts that sometimes the MLE may give the misleading results (see e.g. Berger et al., 1999, Example 3) and that, as we shall see, the MILE may have better properties than the MLE?

Typically, the MLE is invariant w.r.t. the nuisance parameter. Indeed, let \( \lambda \) be the location parameter and assume that

\[
p(u + c; \theta, \lambda) = p(u; \theta, \lambda - c) \quad \forall c.
\]

Then it is easy to see that \( \hat{\lambda}(x + c; \theta) = \hat{\lambda}(x; \theta) + c \quad \forall c \) and \( \theta^*(x + c) = \theta^*(x) \quad \forall c \).

A similar property holds for the scale parameter as nuisance, since if

\[
p(cu; \theta, \lambda) = \frac{1}{c}p(u; \theta, \lambda / c) \quad \forall c > 0,
\]

then \( \hat{\lambda}(cx; \theta) = c\hat{\lambda}(x; \theta) \quad \forall c > 0 \) and \( \theta^*(cx) = \theta^*(x) \quad \forall c > 0 \).

The invariance property of the MLE can be used to propose another method of estimation. Namely, let us apply the maximum integrated likelihood principle to the measure, defined on the \( \sigma \)-algebra of invariant sets (w.r.t. the nuisance parameter), induced by the distribution considered.

Let \( \lambda \) be the scale parameter. A set \( A \subset \mathbb{R}^n \) is called scale invariant, if \( cA = A \forall c > 0 \) (e.g. all the rays of the form \( \{ x : x = ce, c > 0, e \in \mathbb{S}^{n-1} \} \) are scale invariant). Let \( \mathcal{A} \) be the \( \sigma \)-algebra of such sets. Note that the Lebesgue measure of any \( A \in \mathcal{A} \) is either \( 0 \) or \( \infty \).

That is why, let us consider the density corresponding to the measure, induced by the distribution considered, not w.r.t. the Lebesgue measure, but w.r.t. the measure \( \Phi \) induced by e.g. the \( n \)-dimensional normal distribution with the density

\[
\phi(u) := (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{j=1}^{n} u_j^2 \right\}, \quad u = (u_1, \ldots, u_n) \in \mathbb{R}^n.
\]

The calculation of the above-mentioned density is given in Hájek et al. (1999, Chapter 3.2.2) or in Nagaev (1996, Chapter 8). For \( A \in \mathcal{A} \), starting with the equalities

\[
Q(A) := \frac{P_{\theta, \lambda}(A)}{\Phi(A)} = \frac{\lambda^n \int_A \prod_{j=1}^{n} p(\lambda^{-1} u_j; \theta, 1) du}{\int_A \phi(u) du} = \frac{\int_A \prod_{j=1}^{n} p(u_j; \theta, 1) du}{\int_A \phi(u) du},
\]

after routine calculations we obtain this density as

\[
q(u; \theta) = \frac{\lambda^n \int_0^{\infty} \prod_{j=1}^{n} p(t u_j; \theta, 1) dt}{\int_0^{\infty} \lambda^n \phi(t u_1, \ldots, u_n) dt}.
\]

Now let us take \( \theta^*(x) \in \text{Arg sup}_\theta q(x; \theta) \) as an estimator of \( \theta \).

It is evident that \( \text{Arg sup}_\theta q(x; \theta) = \text{Arg sup}_\theta \tilde{L}(\theta; x) \), where

\[
\tilde{L}(\theta; x) = \int_0^{\infty} t^{n-1} \prod_{j=1}^{n} p(t x_j; \theta, 1) dt = \int_0^{\infty} \frac{1}{\lambda^{n+1}} \prod_{j=1}^{n} p \left( \frac{x_j}{\lambda}; \theta, 1 \right) d\lambda = \int_0^{\infty} \prod_{j=1}^{n} p(x_j; \theta, \lambda) d\lambda = \int_0^{\infty} L(\theta, \lambda; x) \frac{1}{\lambda} d\lambda.
\]

Therefore, we come to the integrated likelihood function with the weight function \( w(\lambda) = 1/\lambda, \quad \lambda > 0 \).

Similarly, if \( \lambda \) is the location parameter, we obtain

\[
\tilde{L}(\theta; x) = \int_{-\infty}^{\infty} L(\theta, \lambda; x) d\lambda
\]

and, therefore, we come to the integrated likelihood function with the weight function \( w(\lambda) \equiv 1 \).

It should be noted that these weight functions are not obtained by chance. Indeed, if \( \lambda \) is the location parameter and (1) holds then, due to the invariance requirement, \( \pi(\lambda | \theta) = \pi(\lambda - c | \theta) \forall c \equiv \pi(\lambda | \theta) \propto 1 \) and \( \tilde{L}(\theta; x + c) = \tilde{L}(\theta; x) \forall c \).

Similarly, if \( \lambda \) is the scale parameter and (2) holds then, due to the invariance requirement, \( \pi(\lambda | \theta) = \frac{1}{\lambda} \pi(\lambda / \theta) \forall \lambda > 0 \equiv \pi(\lambda | \theta) \propto \frac{1}{\lambda} \) and \( \tilde{L}(\theta; cx) = \frac{1}{c} \tilde{L}(\theta; x) \forall c > 0 \).

There are other arguments to choose the weight function, e.g. via Jeffreys’s prior density or via the induced right invariant Haar density for \( \lambda \) (see Berger et al., 1999). But if the nuisance parameter is either location or scale, then all they lead to the same weight functions as above. From now on, we consider only such a choice.
We would like to underline that we do not perform the exhaustive study of the maximum integrated likelihood estimators and their properties. Our aim is only to consider two special cases of nuisance parameters (location or scale) and to compare the MILE and the MLE.

One of the most investigated in the literature cases is that of estimation of the shape parameter $\alpha > 0$ in the gamma distribution when the scale parameter $\sigma > 0$ is nuisance. Here the maximum integrated likelihood estimator, considered in Zaigraev and Podraza-Karakulska (2008a,b), coincides with the conditional MLE studied e.g. in Yanagimoto (1988). As it was proved in Zaigraev and Podraza-Karakulska (2008a), the MILE and the MLE.

$\theta$ is the MLE that, as $n \to \infty$, the MILE and the MLE.

$\lambda$ in Zaigraev and Podraza-Karakulska (2008b), the asymptotic comparison of both estimators was performed. It was shown that, as $n \to \infty$,

$$n(\alpha^{**} - \alpha) \to \frac{1}{2\alpha g'/(\alpha)} + O_p(n^{-1/2}),$$

$$n^2 \left[ E(\alpha^{**} - \alpha)^2 - E(\alpha^* - \alpha)^2 \right] = \frac{3}{4\alpha^2 g'/(\alpha)^3} + O(n^{-1/2}).$$

where $g(\alpha) := \ln \alpha - \psi(\alpha)$ and $\psi(\alpha) := \frac{1}{\alpha} \ln \Gamma(\alpha)$ is the so-called Euler psi (digamma) function.

In what follows, we use biases and mean squared errors of the estimators as the criteria for their comparison. Since, in general, the exact distributions of the estimators are not available due to non-explicit forms of these estimators, we are focused on the asymptotic comparison.

The structure of the paper is the following: in Section 2 the main results are given, while Section 3 contains a couple of examples. Schemes of proving the results are given in the Appendix.

2. Main results

We start with notations and assumptions. Let $l(\theta, \lambda; x) := \ln l(\theta, \lambda; x)$ be the log-likelihood function, $l_0(\theta, \lambda; x) := \frac{\partial}{\partial \theta} l(\theta, \lambda; x)$, $l_1(\theta, \lambda; x) := \frac{\partial}{\partial \lambda} l(\theta, \lambda; x)$ be the log-likelihood derivatives of the first order, $l_{00}(\theta, \lambda; x) := \frac{\partial^2}{\partial \theta^2} l(\theta, \lambda; x)$, $l_{01}(\theta, \lambda; x) := \frac{\partial^2}{\partial \theta \partial \lambda} l(\lambda, \lambda; x)$ be the log-likelihood derivatives of the second order, and so on.

Further on, we assume that (1) or (2) holds (depending on the case considered) and consider only regular models (cf. Severini, 2000, Chapter 3.4). Saying “regular models” we mean that

- the MLE $(\theta^{**}(x), \lambda^{**}(x))$ is unique;
- there exist all log-likelihood derivatives up to the fourth order; the normed vector of the log-likelihood derivatives $(Y_1, Y_2) := n^{-1/2} \{l_0(\theta, \lambda; x), l_1(\theta, \lambda; x)\}$ converges in distribution to a two-dimensional non-degenerate normal random vector, as $n \to \infty$, and, as a function of parameters, it admits the Taylor expansion in a neighborhood of the true parameters, as $n \to \infty$; the second moments of all log-likelihood derivatives of the third order, as well as the first moments of all log-likelihood derivatives of the fourth order, are finite for any fixed $\theta$ and $\lambda$;
- integration and differentiation in calculations can be interchanged whenever it is needed.

For regular models, the MLE $\theta^{**}(x)$ is consistent and asymptotically normal, i.e. $n^{-1/2} (\theta^{**}(x) - \theta) \to N(0, h^{**})$ in distribution, as $n \to \infty$, where $h^{**}$ denotes the $i,j$-element of the inverse matrix w.r.t. the Fisher information matrix $H = (h_{11}, h_{12}, h_{22}, b_{12})$ corresponding to a single observation, $i, j = 1, 2$.

The asymptotic properties of the MILE is quite close to that of the MLE.

**Proposition.** For regular models, the MILE $\theta^{**}(x)$ is consistent and asymptotically normal with the same asymptotic distribution as the MLE $\theta^{**}(x)$.

For a single observation $x_1$ let us denote:

$$G_{11,1} := \mathbb{E} \left[ l_{00}(\theta, \lambda; x_1) l_0(\theta, \lambda; x_1) \right], \quad G_{11,2} := \mathbb{E} \left[ l_{00}(\theta, \lambda; x_1) l_1(\theta, \lambda; x_1) \right], \ldots,$$

$$G_{11} := \mathbb{E} \left[ l_{00}(\theta, \lambda; x_1) \right], \quad G_{112} := \mathbb{E} \left[ l_{00}(\theta, \lambda; x_1) l_1(\theta, \lambda; x_1) \right], \ldots,$$

$$G_{111,1} := \mathbb{E} \left[ l_{000}(\theta, \lambda; x_1) l_0(\theta, \lambda; x_1) \right], \quad G_{111,2} := \mathbb{E} \left[ l_{000}(\theta, \lambda; x_1) l_1(\theta, \lambda; x_1) \right], \ldots,$$

$$G_{111} := \mathbb{E} \left[ l_{000}(\theta, \lambda; x_1) \right], \quad G_{1112} := \mathbb{E} \left[ l_{000}(\theta, \lambda; x_1) l_1(\theta, \lambda; x_1) \right], \ldots.$$

Asymptotic comparisons of both estimators (w.r.t. the bias and w.r.t. the MSE) are given in the next two theorems.
Theorem 1. For regular models, at least for all sufficiently large values of $n$:

- $n(\theta^{**}(x) - \theta^{*}(x)) = A + O_p(n^{-1/2}),$
  
  where $A = A(\theta) := \frac{1}{2h^2} \sum_{k=1}^{n} h^{i1} G_{22k} - \frac{h^{12}}{\lambda} \cdot I$ and $I = \begin{cases} 1, & \text{if } \lambda \text{ is scale; } \\ 0, & \text{if } \lambda \text{ is location; } \end{cases}$
- the sign of the difference $|E\theta^{**}(x) - \theta| - |E\theta^{*}(x) - \theta|$ coincides with that of $A(A + B)$, where

\[
B = B(\theta) := \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} h^{i1} h^{j1} (2G_{ij,k} + G_{jk}).
\]

Corollary. Under the conditions of Theorem 1,

(i) if $0 < A < -\frac{1}{2} B \forall \theta$ then $E\theta^{*}(x) < E\theta^{**}(x) < \theta$;
(ii) if $-\frac{1}{2} B < A < 0 \forall \theta$ then $\theta < E\theta^{**}(x) < E\theta^{*}(x)$.

Theorem 2. For regular models, at least for all sufficiently large values of $n$,

\[
n^2 \left( E(\theta^{**}(x) - \theta)^2 - E(\theta^{*}(x) - \theta)^2 \right) = A^2 + AB + 2A \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} h^{i1} h^{j1} (G_{jk,i} + G_{ji}) + \frac{1}{h^{22}} \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} h^{i1} h^{j1} (G_{22j,k} + G_{22k}) + G_{22j} \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} h^{i1} h^{j1} (G_{2j,i} + G_{2i}) + \left[ \frac{2(h^{12})^2}{\lambda^2} - \frac{2h^{11}}{\lambda h^{22}} \sum_{i=1}^{2} h^{i1} (G_{21,i} + G_{2i}) \right] \cdot I + O(n^{-1/2}).
\]

Now suppose that the parameter of interest is $\zeta = f(\theta)$, where $f : \mathbb{R} \mapsto \mathbb{R}$ is a monotone and sufficiently smooth given function (i.e. there exist the derivatives up to the third order). Then $\zeta^{*}(x) = f(\theta^{*}(x))$, $\zeta^{**}(x) = f(\theta^{**}(x))$, and the following statement holds.

Theorem 3. For regular models, at least for all sufficiently large values of $n$:

\[
n(\zeta^{**}(x) - \zeta^{*}(x)) = f'(\theta) n(\theta^{**}(x) - \theta^{*}(x)) + O_p(n^{-1/2});
\]
the sign of the difference $|E\zeta^{**}(x) - \zeta| - |E\zeta^{*}(x) - \zeta|$ coincides with that of $A(A + B + h^{11} f''(\theta)/f'(\theta))$:

\[
n^2 \left( E(\zeta^{**}(x) - \zeta)^2 - E(\zeta^{*}(x) - \zeta)^2 \right)
= (f'(\theta))^2 n^2 \left( E(\theta^{**}(x) - \theta)^2 - E(\theta^{*}(x) - \theta)^2 \right) + 3f'(\theta) f''(\theta) h^{11} A + O(n^{-1/2}).
\]

3. Examples

Example 1. Consider the Weibull $\mathcal{W}(\alpha, \sigma)$ distribution, where the shape parameter $\alpha > 0$ is of interest, while the scale parameter $\sigma > 0$ is nuisance, $n \geq 2$. The likelihood function looks as follows:

\[
L(\alpha, \sigma; x) = \frac{\alpha^n \left( \prod_{j=1}^{n} x_j \right)^{\alpha - 1}}{\sigma^{n\alpha}} \exp \left( -\frac{\sum_{j=1}^{n} x_j^{\alpha}}{\sigma^{\alpha}} \right), \text{ where } \min\{x_1, \ldots, x_n\} := x_{(1)} > 0.
\]

Here $\tilde{x}_i(x, \alpha) = \left( \frac{1}{n} \sum_{j=1}^{n} x_j^{\alpha} \right)^{1/\alpha}$, while the MLE $\alpha^{*}(x)$ is the unique root of the equation

\[
\frac{1}{\alpha} = \frac{\sum_{j=1}^{n} \frac{\ln x_j}{x_j^{\alpha}}}{\sum_{j=1}^{n} x_j^{\alpha / \alpha}} - \frac{1}{n} \sum_{j=1}^{n} \ln x_j.
\]
The integrated likelihood function has the form
\[ \widehat{L}(\alpha; x) = \int_0^\infty \frac{1}{\sigma} L(\alpha, \sigma; x) d\sigma = \frac{\Gamma(n) \alpha^{n-1} \left( \prod_{j=1}^n x_j \right)^{\alpha - 1}}{\left( \sum_{j=1}^n x_j^2 \right)^n} = e^n \Gamma(n) \widehat{L}(\alpha; x) = \frac{e^n}{n^n \alpha}, \]
and the MLE \( \alpha^{**}(x) \) is the unique root of the equation
\[ \frac{n}{n \alpha} = \frac{\sum_{j=1}^n x_j^2 \ln x_j}{\sum_{i=1}^n x_i} - \frac{1}{n} \sum_{i=1}^n \ln x_i. \]
In this case,
\[ h_{11} = \frac{1 + \Gamma''(2)}{\alpha^2}, \quad h_{12} = -\frac{\Gamma''(2)}{\sigma}, \quad h_{22} = \frac{\alpha^2}{\sigma^2}, \quad h^{11} = \frac{\alpha^2}{\sigma^2}, \quad h^{12} = \frac{\sigma \Gamma'(2)}{\Psi'(1)}, \]
\[ G_{111} = \frac{2 - \Gamma'''(2)}{\alpha^3}, \quad G_{112} = \frac{\Gamma'''(2) + 2 \Gamma''(2)}{\sigma \alpha}, \quad G_{122} = -\frac{2\alpha + (1 + \alpha) \Gamma'(2)}{\sigma^2}, \quad G_{222} = \frac{\alpha^2}{\Psi'(1)}, \]
\[ G_{111, 1} = \frac{\Gamma'''(2) + 2 \Gamma''(2)}{\sigma^2 \alpha}, \quad G_{112, 1} = G_{12, 1} = -G_{112}, \quad G_{122, 1} = \frac{2\alpha + \alpha \Gamma'(2)}{\sigma^2}, \quad G_{222, 1} = \frac{(1 + \alpha) \Gamma'(2)}{\sigma^2}, \]
\[ G_{122, 2} = -\frac{\alpha^2 (1 + \alpha)}{\sigma^3}, \quad G_{222, 2} = \frac{-\alpha^2 (2 \alpha + 6 \alpha + 11)}{\sigma^4} \]
Applying Theorems 1 and 2, we obtain, as \( n \to \infty \),
\[ n(\alpha^{**}(x) - \alpha^{*}(x)) = -\frac{\alpha}{\Psi'(1)} + O_p(n^{-1/2}), \quad \alpha < E\alpha^{**}(x) < E\alpha^{*}(x) \]
(case (ii) of Corollary),
\[ n^2 \left( E \left( \alpha^{**}(x) - \alpha \right)^2 - E \left( \alpha^{*}(x) - \alpha \right)^2 \right) = D(\alpha) + O(n^{-1/2}), \]
where
\[ D(\alpha) := -\frac{\alpha^2}{(\Psi'(1))^2} \left( 7 \Psi'(1) + \Psi''(1) \right) < 0. \]

**Example 2.** Consider the Gumbel \( g(\alpha, \sigma) \) distribution, where the scale parameter \( \sigma > 0 \) is of interest, while the location parameter \( \alpha \in \mathbb{R} \) is nuisance, \( n \geq 2 \). The likelihood function looks as follows:
\[ L(\sigma, \alpha; x) = \frac{1}{\sigma^n} \exp \left( -\frac{\sum_{j=1}^n x_j - \alpha}{\sigma} - \frac{n}{\sigma} \right). \]
Here \( \hat{\alpha}(x; \sigma) = \sigma (\ln n - \ln \sum_{j=1}^n e^{-x_j/\sigma}) \), while the MLE \( \sigma^{*}(x) \) is the unique root of the equation
\[ \sigma = \bar{x} - \frac{\sum_{j=1}^n x_j e^{-x_j/\sigma}}{\sum_{i=1}^n e^{-x_i/\sigma}}, \]
where \( \bar{x} \) is the sample mean.
The integrated likelihood function has the form
\[
\tilde{L}(\sigma; x) = \int_{-\infty}^{\infty} L(\sigma, a; x) da = \frac{\Gamma(n) e^{-n\overline{X}/\sigma}}{\sigma^{n-1} \left( \sum_{i=1}^{n} e^{-x_i/\sigma} \right)^n} = \frac{e^n \Gamma(n)\sigma \tilde{L}(\sigma; x)}{n^n},
\]
and the MILE \(\sigma^{**}(x)\) is the unique root of the equation
\[
\frac{n-1}{n} \sigma = \bar{x} - \frac{\sum_{i=1}^{n} x_i e^{-x_i/\sigma}}{\sum_{i=1}^{n} e^{-x_i/\sigma}}.
\]
In this case,
\[
\begin{align*}
h_{11} &= \frac{1 + \Gamma''(2)}{\sigma^2}, & h_{12} &= -\frac{\Gamma'(2)}{\sigma^2}, & h_{22} &= \frac{1}{\sigma^2}, & h^{11} &= \frac{\sigma^2}{\Psi'(1)}, \\
h_{12} &= \frac{\sigma^2 \Gamma'(2)}{\Psi'(1)}, & h_{22} &= \frac{\sigma^2 (1 + \Gamma''(2))}{\Psi'(1)}, \\
h_{11} &= \frac{4 + 6 \Gamma''(2) + \Gamma'''(2)}{\sigma^3}, & h_{12} &= -\frac{4 \Gamma'(2) + \Gamma''(2)}{\sigma^3}, & h_{22} &= \frac{2 + \Gamma'(2)}{\sigma^3}, & h_{222} &= \frac{-1}{\sigma^3}, \\
h_{122} &= -\frac{6 + 6 \Gamma''(2) + \Gamma'''(2)}{\sigma^4}, & h_{122} &= \frac{3 + \Gamma'(2)}{\sigma^4}, & h_{222} &= -\frac{1}{\sigma^4}, & h_{122} &= -G_{122}, \\
h_{111} &= -\frac{2 + 4 \Gamma''(2) + \Gamma'''(2)}{\sigma^3}, & h_{112} &= -G_{112}, & h_{121} &= \frac{2 \Gamma'(2) + \Gamma''(2)}{\sigma^3}, & h_{221} &= -\frac{\Gamma'(2)}{\sigma^3}, \\
h_{222} &= -G_{222}, & h_{222} &= \frac{3 \Gamma'(2) + \Gamma''(2)}{\sigma^4}, & h_{222} &= \frac{-3 + \Gamma'(2)}{\sigma^4}. \\
h_{222,1} &= -\frac{\Gamma'(2)}{\sigma^4}, & G_{222} &= \frac{1}{\sigma^4}.
\end{align*}
\]
Applying Theorems 1 and 2, we obtain, as \(n \to \infty\),
\[
n(\sigma^{**}(x) - \sigma^*(x)) = \frac{\sigma}{\Psi'(1)} + O_p(n^{-1/2}), \quad \text{E} \sigma^*(x) < \text{E} \sigma^{**}(x) < \sigma
\]
(case i) of Corollary),
\[
n^2 \left( E \left( \sigma^{**}(x) - \sigma \right)^2 - E \left( \sigma^*(x) - \sigma \right)^2 \right) = D(\sigma) + O(n^{-1/2}), \quad n \to \infty,
\]
where
\[
D(\sigma) := -\frac{\sigma^2}{(\Psi'(1))^3} \left( \Psi'(1) + \Psi''(1) \right) > 0.
\]

Note that the same results can be obtained via Theorem 3, applying it to the Weibull distribution and taking \(z = 1/\alpha\) (with substituting afterwards \(1/\sigma\) instead of \(\omega\)), since if \((x_1, x_2, \ldots, x_n)\) is a sample drawn from \(g(a, \sigma)\) distribution, then \((e^{-x_1}, e^{-x_2}, \ldots, e^{-x_n})\) is a sample drawn from \(W(1/\sigma, e^{-\sigma})\) distribution.

**Appendix. On proving Theorems 1–3**

From the consistency of \(\hat{\theta}^*\) and the regularity conditions on the distribution, it follows that \(\theta^{**}\) is also consistent. We shall calculate \(n^{1/2}(\theta^* - \hat{\theta})\), \(n^{1/2}(\lambda^* - \hat{\lambda})\) up to the order \(O_p(n^{-3/2})\) (Takeuchi and Akahira, 1979, made such a calculation up to the order \(O_p(n^{-3})\)).

Taking the Taylor expansion of the functions \(n^{-1/2}l_0(\theta^*; x)\) and \(n^{-1/2}l_2(\theta^*, \lambda^*; x)\) in a neighborhood of the true parameters \(\theta, \lambda\), and substituting the representations \(n^{1/2}(\theta^* - \theta) = U_1 + n^{-1/2}V_1 + n^{-1}W_1 + O_p(n^{-3/2})\) and \(n^{1/2}(\lambda^* - \lambda) = U_2 + n^{-1/2}V_2 + n^{-1}W_2 + O_p(n^{-3/2})\), after equating to zero the coefficients under the corresponding powers of \(n\), we obtain:
\[
\begin{align*}
U := \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} &= H^{-1} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}, & V := \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} &= H^{-1} \begin{pmatrix} YV + T/2 \end{pmatrix}, \\
\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} &= H^{-1} \begin{pmatrix} YV + P/2 + R + Q/6 \end{pmatrix},
\end{align*}
\]
(3)
where \( Y_1 := n^{-1/2} l_0(\theta, \lambda; x), \ Y_2 := n^{-1/2} l_0(\theta, \lambda; x), \ Y = (Y_0)_{\thetaj=1}^2 \).

\[
Y_{11} := n^{-1/2} \left( l_{00}(\theta, \lambda; x) + h h_{11} \right), \quad Y_{21} = Y_{12} := n^{-1/2} \left( l_{0\lambda}(\theta, \lambda; x) + nh_{12} \right), \quad Y_{22} := n^{-1/2} \left( l_{\lambda\lambda}(\theta, \lambda; x) + nh_{22} \right),
\]

\[ T := (T_1, T_2)^T, \quad T_i := \sum_{j=1}^2 \sum_{k=1}^2 G_{ijk} U_j U_k, \quad i = 1, 2, \quad P := (P_1, P_2)^T, \]

\[
P_i := \sum_{j=1}^2 \sum_{k=1}^2 Y_{ijk} U_j U_k, \quad i = 1, 2,
\]

\[
R := (R_1, R_2)^T, \quad R_i := \sum_{j=1}^2 \sum_{k=1}^2 G_{ijk} U_j V_k, \quad i = 1, 2,
\]

\[
Q := (Q_1, Q_2)^T, \quad Q_i := \sum_{j=1}^2 \sum_{k=1}^2 \sum_{m=1}^2 G_{ikm} U_j U_k U_m, \quad i = 1, 2,
\]

\[
Y_{111} := n^{-1/2} \left( l_{000}(\theta, \lambda; x) - nG_{111} \right), \quad Y_{211} = Y_{121} := n^{-1/2} \left( l_{0\lambda\lambda}(\theta, \lambda; x) - nG_{112} \right), \quad Y_{221} = Y_{222} := n^{-1/2} \left( l_{\lambda\lambda\lambda}(\theta, \lambda; x) - nG_{222} \right).
\]

Since \( EU = 0 \), after routine calculations we obtain the formula for the bias: 
\[ E\theta^* - \theta = EV_1/n + O(n^{-3/2}) = B/(2n) + O(n^{-3/2}). \]

As to the MILE, first we prove that under the regularity conditions
\[
\int L(\theta, s; x) w(s) ds = \int_{|x|<|\delta|} L(\theta, s; x) w(s) ds(1 + o_p(1))
\]

for any fixed small \( \delta > 0 \). Substituting the Taylor expansions of the functions \( n^{-1/2}(\partial/\partial \theta) L(\theta^{**}, s; x) \) in a neighborhood of the true parameters \( (\theta, \lambda) \) and \( w(s) \) in a neighborhood of \( \lambda \) into the equation \( n^{-1/2}(\partial/\partial \theta) \int_{|x|<|\delta|} L(\theta, s; x) w(s) ds = 0 \), and calculating all the integrals, we obtain the representation
\[
n^{1/2}(\theta^{**} - \theta) = U_1 + n^{-1/2}(V_1 + A) + n^{-1}(W_1 + Z) + O_p(n^{-3/2}). \quad (4)
\]

where
\[
Z = A \sum_{j=1}^2 \sum_{k=1}^2 \left[ Y_{ij} + \sum_{k=1}^2 G_{ijk} U_k \right] + \frac{1}{2h_{12}} \sum_{j=1}^2 \sum_{k=1}^2 h_{ij} \left[ Y_{22j} + \sum_{k=1}^2 G_{22k} U_k \right] + \left[ \sum_{j=1}^2 \sum_{k=1}^2 G_{2jk} U_k \right] \frac{h_{11}}{\lambda h_{22}} \left[ Y_{21j} + \sum_{k=1}^2 G_{21k} U_k + \frac{h_{12}}{\lambda} U_k \right] .
\]

Thus, \( \theta^* \) and \( \theta^{**} \) have the same asymptotic distribution, and the first statement of Theorem 1 follows. We have \( E\theta^{**} - \theta = n^{-1}(A + B/2) + O(n^{-3/2}) \), therefore for sufficiently large values of \( n \) the sign of the difference \( n^{1/2}|E\theta^{**}(x) - \theta| - n^{1/2}|E\theta^*(x) - \theta| \) coincides with that of \( nA(A + B) \).

Note that in view of (3) and (4)
\[
n(\theta^{**} - \theta)^2 - n(\theta^* - \theta)^2 = \frac{2AU_1}{\sqrt{n}} + \frac{2U_1Z + 2AV_1 + A^2}{n} + O_p(n^{-3/2}).
\]

Taking expectations, we obtain Theorem 2.

To prove Theorem 3 we only note that
\[
\zeta^* = f(\theta) + \frac{f(\theta)}{\sqrt{n}} \sqrt{n}(\theta^* - \theta) + \frac{f''(\theta)}{2n} \left( \sqrt{n}(\theta^* - \theta) \right)^2 + \frac{f''(\theta)}{6n\sqrt{n}} \left( \sqrt{n}(\theta^* - \theta) \right)^3 + O_p(n^{-2})
\]

\[
\implies \sqrt{n}(\zeta^* - \zeta) = U'_1 + \frac{V_1'}{\sqrt{n}} + \frac{W_1'}{n} + O_p(n^{-3/2}),
\]

where
\[
U'_1 = f'(\theta)U_1, \quad V'_1 = f'(\theta)V_1 + \frac{f''(\theta)}{2} U_1^2, \quad W'_1 = f'(\theta)W_1 + f''(\theta)U_1 V_1 + \frac{f'''(\theta)}{6} U_1^3.
\]
and similarly
\[ \sqrt{n(\hat{\xi}^{**} - \xi)} = U''_1 + \frac{V''}{\sqrt{n}} + \frac{W''}{n} + O_p(n^{-3/2}), \]

where
\[ U''_1 = U'_1, \quad V'' = V'_1 + f'(\theta)A, \quad W'' = W'_1 + f'(\theta)Z + f''(\theta)AU_1. \]

It remains to repeat the reasonings led to the statements of Theorems 1 and 2.

References