Continuous Statistical Models: With or Without Truncation Parameters?

V. Vancak[1*], Y. Goldberg[1**], S. K. Bar-Lev[1***], and B. Boukai[2****]

[1]*Dept. Statist., Univ. Haifa, Israel*
[2]*Dept. Math. Sci., IUPUI, US*
Received June 29, 2014; in final form, February 10, 2015

Abstract—Lifetime data are usually assumed to stem from a continuous distribution supported on $[0, b)$ for some $b \leq \infty$. The continuity assumption implies that the support of the distribution does not have atom points, particularly not at 0. Accordingly, it seems reasonable that with an accurate measurement tool all data observations will be positive. This suggests that the true support may be truncated from the left. In this work we investigate the effects of adding a left truncation parameter to a continuous lifetime data statistical model. We consider two main settings: right truncation parametric models with possible left truncation, and exponential family models with possible left truncation. We analyze the performance of some optimal estimators constructed under the assumption of no left truncation when left truncation is present, and vice versa. We investigate both asymptotic and finite-sample behavior of the estimators. We show that when left truncation is not assumed but is, in fact present, the estimators have a constant bias term, and therefore will result in inaccurate and inefficient estimation. We also show that assuming left truncation where actually there is none, typically does not result in substantial inefficiency, and some estimators in this case are asymptotically unbiased and efficient.

Keywords: truncation parameter, estimation, order statistics, efficiency, model selection.

2000 Mathematics Subject Classification: 62F35

DOI: 10.3103/S1066530715010044

1. INTRODUCTION

When modeling lifetime data it is usually assumed that the distribution is continuous and supported on $\Omega_1 = [0, b)$, $b \leq \infty$. Here b is either a known constant or a parameter (Elandt-Johnson, 1999; Lawless, 2003). However, is this the right support? The continuity assumption implies that the support Ω_1 of the distribution does not have atom points, particularly not at 0. Thus one can expect that all observations will be positive. Indeed, when the lifetime data measure time to events such as death or remission, it seems reasonable to assume that one observes only positive time to event values. Even when the observed data is on events of small time scale (such as time of detection of motion sensors), as measured with an accurate measuring tool, it is expected that all observations would be positive. This suggests that the true support may be truncated from the left and is, in fact, of the form $\Omega_2 = [\gamma, b)$, $\gamma > 0$, with γ unknown.

In this work we consider models that allow the support of the distribution to be chosen adaptively from the data by using truncation parameters. Each of these proposed models can be considered as generalization of a model in which the support includes $\{0\}$. A natural question that arises is which of these two models should we use? With this, we need to consider two possible errors. To describe these errors, let Model I denote the statistical model of which the true support is $\Omega_1 = [0, b)$, and let Model II denote the truncated statistical model of which the true support is $\Omega_2 = [\gamma, b)$, $\gamma > 0$. We say that a False

[*]E-mail: `valentin.vancak@gmail.com`
[**]E-mail: `ygoldberg@stat.haifa.ac.il`
[***]E-mail: `barlev@stat.haifa.ac.il`
[****]E-mail: `bboukai@iupui.edu`

Model I error has occurred if Model I has been incorrectly used for inference while the correct model is Model II. A False Model II error is defined similarly. Then the question arises: Which of the two types of errors is more severe? The answer to this question is useful when the model's underlying true support is unknown.

It seems reasonable that even if $\Omega_1 = [0, b)$ is the correct support, using Model II will not result in substantial loss of information. Conversely, if $\Omega_2 = [\gamma, b)$ for some $\gamma > 0$ is the correct support, there will be substantial loss of information when using Model I. This claim can be justified in terms of sufficiency. Assume that Model I depends on some unknown parameter θ (possibly a vector), and is associated with a minimal sufficient statistic $t_n = t(\mathbf{X}_n)$. Model II, which is obtained by a left-truncation of Model I, is therefore parameterized by $\eta = (\gamma, \theta)$ and is associated with the minimal sufficient statistic $(X_{(1)}, t_n)$. Note that $(X_{(1)}, t_n)$, while being a minimal and sufficient statistic for Model II, is still sufficient for Model I; whereas t_n alone, while being minimal and sufficient for Model I, is not sufficient for Model II, hence we expect the False Model I error to be more critical. As we later show, this understanding, while essentially correct, requires some further clarifications.

Two main settings are investigated in this work. In the first one, we assume that the density function is known up to a right truncation parameter. In this setting, under Model I, there is only a right truncation parameter θ. In other words, $\Omega_1 = [0, \theta)$. Under Model II, we assume also left truncation and hence the support is $\Omega_2 = [\gamma, \theta)$. For this setting, two candidate estimators (Tate, 1959; Bar-Lev and Boukai 1985) will be compared for their cross-model bias and MSE, as well as for their asymptotic efficiency. More specifically, for the right truncation setting with a possible left truncation parameter, we are interested in the behavior of the Bar-Lev and Boukai (1985) (hereafter abbreviated BB) estimator when there is no left truncation, and the behavior of the Tate's estimator (Tate, 1959) when left truncation is indeed present.

The second setting deals with distributions having a 'regular' parameter with a possible left truncation. We begin with the Erlang distribution as a special case of the natural exponential family (NEF), and illustrate the effect of the possible truncation on the estimator of the 'regular' parameter. We proceed to discuss this problem in the general case of the NEF distribution, however in the asymptotic sense only.

The question discussed in this paper can be considered as a model selection problem. One can think about Model I as a narrow model and on Model II as a wider model since Model II includes an additional parameter. Selecting the right model was addressed considerably in the literature for maximum likelihood estimation, and in particular for linear regression, using tools such as AIC and BIC (see, for example, Burnham and Anderson, 2002). The consequences of choosing a misspecified model when using maximum likelihood estimation were discussed by White (1982), among others. Bickel (1984) considered the effect of misspecification for linear regression model. Claeskens and Hjort (2008) suggested criteria, such as tolerance radius, for choosing between a narrow model and a 'wider' one. We note that since the two possible models we consider have different supports, many of the results mentioned above do not hold for this setting (see, for example, White (1982), Assumption A7). Moreover, the approach we consider here, at least for the first setting, does not fall under the maximum likelihood estimation. Finally, most of our analysis is exact and not asymptotic. Hence, this paper offers a new approach for an interesting novel problem of model selection.

The paper is organized as follows. The analysis of continuous statistical right-truncated models with possible left truncation is presented in Section 2. In Section 3 we discuss in detail the Erlang distribution case, and conclude with some asymptotic aspects of possible truncation in the NEF case. Concluding remarks appear in Section 4. The proofs are presented in the Appendix.

2. RIGHT TRUNCATED MODELS WITH A POSSIBLE LEFT TRUNCATION

In Section 2.1 we present the model. We then discuss estimation in Section 2.2, and cross-model analysis in Section 2.3. Finally, in Section 2.4, we illustrate the results by examples.

2.1. General Setup

Let $h(\cdot)$ be a positive integrable function over $[0, \infty)$. For any $0 \leq \gamma < \theta$ we define

$$g_k(\gamma, \theta) = \int_{\gamma}^{\theta} x^k h(x)\, dx, \quad k = 0, 1, 2, \ldots . \tag{1}$$

Using (1), we construct the probability density function (p.d.f.) of a continuous type random variable X as

$$f(x; \eta) = \frac{h(x)}{g_0(\gamma, \theta)} I[\gamma < x < \theta]. \tag{2}$$

Here, $I[A]$ is the indicator function of the set A and γ and θ are the two possibly unknown parameters of $f(x; \eta)$. Accordingly, we consider two possible models for $\eta \equiv (\gamma, \theta)$:

- Model I: $\gamma \equiv \gamma_0 = 0$ is known, while $\theta > \gamma_0$ is an unknown parameter, so that $\eta_0 \equiv (\gamma_0, \theta)$ designates the model's only unknown parameter θ.

- Model II: Both γ and θ are unknown parameters, $0 < \gamma < \theta$, so that $\eta \equiv (\gamma, \theta)$ designates the model's two unknown parameters.

Note that with the notation in (1), the moments of X under η are easily defined by

$$E_\eta(X^k) = \frac{g_k(\gamma, \theta)}{g_0(\gamma, \theta)}, \quad k = 0, 1, 2, \ldots . \tag{3}$$

In particular, the expected value of X is $E_\eta(X) = g_1(\gamma, \theta)/g_0(\gamma, \theta)$. Similarly, the cumulative distribution function (c.d.f.) of X is given, for any $\tau \in \mathbb{R}$, by

$$F_\eta(\tau) \equiv P_\eta(X \leq \tau) = \frac{g_0(\gamma, \tau)}{g_0(\gamma, \theta)} I[\gamma < \tau < \theta] + I[\theta \leq \tau], \tag{4}$$

with the corresponding tail probability

$$1 - F_\eta(\tau) \equiv P_\eta(X > \tau) = I[\tau \leq \gamma] + \frac{g_0(\tau, \theta)}{g_0(\gamma, \theta)} I[\gamma < \tau < \theta]. \tag{5}$$

Following (4), the π-th quantile is given by τ_π that solves the equation $F_\eta(\tau_\pi) \equiv P(X \leq \tau_\pi) = \pi$ for $\gamma < \tau_\pi < \theta$, and can be expressed using the inverse of the cumulative distribution function

$$\tau_\pi = F_\eta^{-1}(\pi), \tag{6}$$

such that $\pi \equiv F_\eta(\tau_\pi) = g_0(\gamma, \tau_\pi)/g_0(\gamma, \theta)$.

As before, let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ be a sample of n i.i.d. observations from $f(x; \eta)$ in (2), and let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ denote the corresponding order statistics. It is a standard exercise to show that under Model I, the minimal sufficient statistic (MSS) for $\eta_0 \equiv (0, \theta)$ is $S_I = X_{(n)}$, while under Model II, with $\eta \equiv (\gamma, \theta)$, the MSS is $S_{II} = (X_{(1)}, X_{(n)})$. Under Model I, the p.d.f. of the MSS statistic $S_I = X_{(n)}$ is

$$f_{S_I}(t; \eta_0) = \frac{n h(t) \big(g_0(0, t)\big)^{n-1}}{\big(g_0(0, \theta)\big)^n} I[0 < t < \theta], \tag{7}$$

whereas, under Model II, the p.d.f. of $S_{II} = (X_{(1)}, X_{(n)})$ can be shown to be

$$f_{S_{II}}(y, t; \eta) = \frac{n(n-1) h(y) h(t) \big(g_0(y, t)\big)^{n-2}}{\big(g_0(\gamma, \theta)\big)^n} I[\gamma < y < t < \theta]. \tag{8}$$

We finally note that under Model I, when $\eta \equiv \eta_0 = (0, \theta)$, the statistic $S_{II} = (X_{(1)}, X_{(n)})$, while sufficient for θ, is not minimal, and its p.d.f. is given by

$$f_{S_{II}}(y, t; \eta_0) = \frac{n(n-1) h(y) h(t) \big(g_0(y, t)\big)^{n-2}}{\big(g_0(0, \theta)\big)^n} I[0 \leq y < t < \theta]. \tag{9}$$

However, under Model II, $\eta \equiv (\gamma, \theta)$, and the statistic $S_I = X_{(n)}$ is not sufficient for the unknown η, and its p.d.f. is given by

$$f_{S_I}(t; \eta) = \frac{n\, h(t) \left(g_0(\gamma,\, t)\right)^{n-1}}{\left(g_0(\gamma,\, \theta)\right)^n}\, I[\gamma \le t < \theta].$$

Regardless of the assumed model (Model I or II), it is easy to verify that the conditional p.d.f. of $X_{(n)}$ given $X_{(1)} = y$ (with $y > \gamma$) is given by

$$f_{X_{(n)}|X_{(1)}}(t;\, y, \theta) = \frac{(n-1)\, h(t) \left(g_0(y,\, t)\right)^{n-2}}{\left(g_0(y,\, \theta)\right)^{n-1}}\, I[y < t < \theta], \tag{10}$$

whereas the marginal p.d.f. of $X_{(1)}$ is

$$f_{X_{(1)}}(y;\, \eta) = \frac{n\, h(y) \left(g_0(y,\, \theta)\right)^{n-1}}{\left(g_0(\gamma,\, \theta)\right)^n}\, I[\gamma < y < \theta]. \tag{11}$$

2.2. UMVU Estimation

Let $\xi(\eta)$ be any estimable function of the model's unknown parameter η. For instance, $\xi(\eta) = E_\eta(X)$, or $\xi(\eta) = F_\eta(\tau)$, for some fixed $\tau \in \mathbb{R}$. Based on the sample data $\mathbf{x} = (x_1, x_2, \dots, x_n)$, we are interested in constructing a UMVUE $\hat{\xi}_n \equiv \hat{\xi}(S(\mathbf{x}))$ for $\xi(\eta)$. Clearly, for any η, this estimator should satisfy

$$E_\eta(\hat{\xi}_n) = \xi(\eta). \tag{12}$$

By repeatedly differentiating both sides of (12) with respect to the components of η, along with application of Leibniz's integral rule, one can obtain (in the case of distributions of the form in (2)), explicit expressions for such UMVU estimators.

Tate (1959) considered this problem under Model I (i.e., $\eta \equiv \eta_0 = (0, \theta)$ and $S_I(\mathbf{x}) = x_{(n)}$) and obtained that the general form of the UMVUE for $\xi(\theta)$ is

$$\hat{\xi}_n^I = \xi(x_{(n)}) + \frac{\xi'(x_{(n)})g_0(0,\, x_{(n)})}{nh(x_{(n)})}, \tag{13}$$

whenever the derivative $\xi'(\theta) = \partial\xi(\theta)/\partial\theta$ exists and is continuous almost everywhere on the support $\Omega_1 = \{(0, \theta)\colon a < \theta < b\}$.

Similarly, Bar-Lev and Boukai (1985) considered the same estimation problem under Model II (i.e., $\eta \equiv (\gamma, \theta)$ and $S_{II}(\mathbf{x}) = (x_{(1)},\, x_{(n)})$). They showed that the general form of the UMVUE for any estimable function $\xi(\gamma, \theta)$ is

$$\hat{\xi}_n^{II} = \xi(x_{(1)}, x_{(n)}) - \frac{g_0(x_{(1)}, x_{(n)})\xi_1(x_{(1)}, x_{(n)})}{(n-1)h(x_{(1)})}$$
$$+ \frac{g_0(x_{(1)}, x_{(n)})\xi_2(x_{(1)}, x_{(n)})}{(n-1)h(x_{(n)})} - \frac{g_0^2(x_{(1)}, x_{(n)})\xi_{12}(x_{(1)}, x_{(n)})}{n(n-1)h(x_{(1)})h(x_{(n)})}, \tag{14}$$

whenever the partial derivatives $\xi_1 = \partial\xi/\partial\gamma$, $\xi_2 = \partial\xi/\partial\theta$, and $\xi_{12} = \partial^2\xi/\partial\gamma\partial\theta$ exist and are continuous almost everywhere on $\Omega_2 = \{(\gamma, \theta)\colon a < \gamma < \theta < b\}$.

Remark 1. Assume that Model II holds, but $\xi(\gamma, \theta) \equiv \xi(\theta)$ for some estimable function of θ alone. Then, similarly to (13), the general form of the BB's UMVUE for $\xi(\theta)$ is reduced to

$$\hat{\xi}_n^{II} = \xi(x_{(n)}) + \frac{\xi'(x_{(n)})\, g_0(x_{(1)}, x_{(n)})}{(n-1)h(x_{(n)})}, \tag{15}$$

where $\xi' = \partial\xi/\partial\theta$. A comparison of $\hat{\xi}_n^I$ in (13) to $\hat{\xi}_n^{II}$ in (15) reveals the extent of the bias upon erroneously using Tate's estimator $\hat{\xi}_n^I$ instead of the UVMUE $\hat{\xi}_n^{II}$. In fact, it can be easily seen that

$$\hat{\xi}_n^I = \hat{\xi}_n^{II} + \frac{\xi'(x_{(n)})}{h(x_{(n)})}\left(\frac{g_0(0,\,x_{(n)})}{n} - \frac{g_0(x_{(1)},x_{(n)})}{n-1}\right) \equiv \hat{\xi}_n^{II} + b_n\,.$$

Hence, since $E_\eta(\hat{\xi}_n^{II}) = \xi(\theta)$, it immediately follows that $E_\eta(\hat{\xi}_n^I) = \xi(\theta) + B_n$, where $B_n \equiv E_\eta(b_n)$ represents the bias.

Remark 2. It can be shown (see (10) and (11)) that under Model II, the conditional expectation of $\hat{\xi}_n^{II}$, given $X_{(1)} = y$, is

$$E_\eta(\hat{\xi}_n^{II}|X_{(1)} = y) = \xi(y,\theta) - \frac{\xi_1(y,\,\theta)g_0(y,\theta)}{nh(y)} \equiv r(y,\,\theta)\,. \tag{16}$$

Hence, by Remark 1 and (15), $\hat{\xi}_n^{II}$ must be of the form

$$\hat{\xi}_n^{II}(y,\,t) = r(y,\,t) + \frac{r_2(y,\,t)g_0(y,t)}{(n-1)h(t)}, \tag{17}$$

where $r_2 = \partial r/\partial\theta$.

In the next section we provide a more general assessment of the bias term in (15). Examples for Tate's and for BB's UMVU estimators are provided below; we omit the derivations.

Example 1. Under Model I:

(a) For $\xi(\theta) = E_{\eta_0}(X^k) = g_k(\gamma_0,\theta)/g_0(\gamma_0,\theta)$, with known $\gamma_0 = 0$, it can be shown that Tate's UMVU estimator is

$$\hat{\xi}_n^I = \frac{g_k(0,x_{(n)})}{g_0(0,x_{(n)})}\left(1 - \frac{1}{n}\right) + \frac{x_{(n)}^k}{n}\,.$$

(b) In particular, for $k = 1$, we have that $\xi(\theta) = E_{\eta_0}(X) = g_1(\gamma_0,\theta)/g_0(\gamma_0,\theta)$, with known $\gamma_0 = 0$. It can be shown that Tate's UMVU estimator is

$$\hat{\xi}_n^I = \frac{g_1(0,x_{(n)})}{g_0(0,x_{(n)})}\left(1 - \frac{1}{n}\right) + \frac{x_{(n)}}{n}\,.$$

(c) For $\xi(\theta) = 1 - F_{\eta_0}(\tau) = g_0(\tau,\theta)/g_0(\gamma_0,\theta)$, with known $\gamma_0 = 0$, and $\tau \geq \gamma_0$, it can be shown that Tate's UMVU estimator is

$$\hat{\xi}_n^I = 1 - \left(1 - \frac{1}{n}\right)\frac{g_0(0,\tau)}{g_0(0,x_{(n)})}\,.$$

(d) For $\xi(\theta) \equiv \tau = F_{\eta_0}^{-1}(g_0(\gamma_0,\tau)/g_0(\gamma_0,\theta))$, with known $\gamma_0 = 0$, and $\tau \geq \gamma_0$, it can be shown that Tate's UMVU estimator for τ is

$$\hat{\xi}_n^I = F_{\eta_0}^{-1}\left(\frac{g_0(0,\tau)}{g_0(0x_{(n)})}\right) + \frac{g_0^2\left(0, F_{\eta_0}^{-1}(x_{(n)})\right)g_0(0,x_{(n)})}{nh(x_{(n)})g_0(0,\tau)h(F_{\eta_0}^{-1}(x_{(n)}))}\,. \tag{18}$$

Example 2. Under Model II:

(a) For $\xi(\theta) = E_\eta(X^k) = g_k(\gamma,\theta)/g_0(\gamma,\theta)$, with known $\gamma > 0$, it can be shown that BB's UMVU estimator is

$$\hat{\xi}_n^{II} = \frac{g_k(x_{(1)},x_{(n)})}{g_0(x_{(1)},x_{(n)})}\left(1 - \frac{2}{n}\right) + \frac{x_{(1)}^k + x_{(n)}^k}{n}\,.$$

(b) In particular, for $k = 1$, if $\xi(\gamma, \theta) = E_\eta(X) = g_1(\gamma, \theta)/g_0(\gamma, \theta)$, the general form of BB's UMVU estimator is

$$\hat{\xi}_n^{II} = \frac{g_k(x_{(1)}, x_{(n)})}{g_0(x_{(1)}, x_{(n)})}\left(1 - \frac{2}{n}\right) + \frac{x_{(1)} + x_{(n)}}{n}.$$

(c) For $\xi(\gamma, \theta) = 1 - F_\eta(\tau) = g_0(\tau, \theta)/g_0(\gamma, \theta)$, $\gamma \leq \tau \leq \theta$, one obtains that BB's UMVU estimator is

$$\hat{\xi}_n^{II} = \left(1 - \frac{1}{n}\right) - \left(1 - \frac{2}{n}\right)\frac{g_0(x_{(1)}, \tau)}{g_0(x_{(1)}, x_{(n)})}.$$

(d) It follows from (6) that for $\xi(\theta) = F_\eta^{-1}(g_0(\gamma, \tau)/g_0(\gamma, \theta))$, with $\gamma > 0$ and $\tau \geq \gamma$, BB's UMVU is

$$\hat{\xi}_n^{II} = F_\eta^{-1}\left(\frac{g_0(x_{(1)}, \tau)}{g_0(x_{(1)}, x_{(n)})}\right) - \frac{g_0(x_{(1)}, x_{(n)})g_0^2(F_\eta^{-1}(x_{(1)}), x_{(n)})}{(n-1)h(x_{(1)})h(F_\eta^{-1}(x_{(1)}))g_0(F_\eta^{-1}(x_{(1)}), \tau)}$$
$$- \frac{g_0(x_{(1)}, x_{(n)})g_0^2(x_{(1)}, F_\eta^{-1}(x_{(n)}))}{(n-1)h(x_{(n)})h(F_\eta^{-1}(x_{(n)}))g_0(x_{(1)}, \tau)} - \frac{g_0^2(x_{(1)}, x_{(n)})\xi_{12}(x_{(1)}, x_{(2)})}{n(n-1)h(x_{(1)})h(x_{(n)})},$$

where

$$\xi_{12}(x_{(1)}, x_{(n)}) = -\frac{F_{\eta 12}(F_\eta^{-1}(x_{(1)}), x_{(n)})}{(F_{\eta 1}(F_\eta^{-1}(x_{(1)}), x_{(n)}))^2 F_{\eta 2}(F_\eta^{-1}(x_{(1)}), F_\eta^{-1}(x_{(n)}))},$$

assuming that the partial derivatives $F_{\eta 1} = \partial F_\eta/\partial\gamma$, $F_{\eta 2} = \partial F_\eta/\partial\theta$, and $F_{\eta 12} = \partial^2 F_\eta/\partial\gamma\partial\theta$ exist and are continuous almost everywhere on $\Omega_2 = \{(\gamma, \theta)\colon a < \gamma < \theta < b\}$.

2.3. Cross-Model Analysis

The analysis in this section focuses on model misspecification, where the quantities of interest are (i) the estimators' expectations and (ii) the estimators' MSE w.r.t. the incorrect support. In other words, what is the reduction in efficiency (if any) when we derive the estimators w.r.t. Model I support while actually Model II support holds, and vice versa. More specifically, we are interested in the evaluation of $E(\hat{\xi}_n^I|\text{Model II}) = E_\eta(\hat{\xi}_n^I)$ when $\eta = (\gamma, \theta)$ is the unknown parameter, and of $E(\hat{\xi}_n^{II}|\text{Model I}) = E_{\eta_0}(\hat{\xi}_n^{II})$ when $\eta_0 = (0, \theta)$ and θ is the only unknown parameter. Similar cross-evaluations will be considered for the $MSE(\hat{\xi}_n^I|\text{Model II}) = E_\eta((\hat{\xi}_n^I - \xi)^2)$ and the $MSE(\hat{\xi}_n^{II}|\text{Model I}) = E_{\eta_0}((\hat{\xi}_n^{II} - \xi)^2)$. In the following theorem we evaluate the extent of the cross-model bias by straightforward calculations.

Theorem 1. *Let $\xi(\eta)$ be any estimable function, i.e., any function of the unknown parameters which possesses an unbiased estimator under both Model I and Model II (e.g., Examples 1, 2), and let $\hat{\xi}_n^I$ and $\hat{\xi}_n^{II}$ be its respective estimators.*

(i) *Under False Model I misspecification we have*

$$E(\hat{\xi}_n^I|\text{Model II}) = E_\eta(\hat{\xi}_n^I) = \xi(\theta)(1 + a) + b_n, \tag{19}$$

where, with $\eta = (\gamma, \theta)$, $a = g_0(0, \gamma)/g_0(\gamma, \theta)$ and

$$b_n = -a\int_\gamma^\theta \xi(t)(n-1)h(t)\frac{(g_0(\gamma, t))^{n-2}}{(g_0(\gamma, \theta))^{n-1}}\,dt.$$

(ii) *Under False Model* II *misspecification we have*

$$E(\hat{\xi}_n^{II}|\text{Model I}) = \xi(\eta) + d_{1n} + d_{2n},$$

where

$$d_{1n} \equiv E_{\eta_0}(\hat{\xi}_n^{II} I[0 < y < t < \gamma]) = \int_0^\gamma \int_0^t \hat{\xi}_n^{II}(y, t) f_{S_{II}}(y, t; \eta_0) dy\, dt$$

and

$$d_{2n} \equiv E_{\eta_0}(\hat{\xi}_n^{II} I[0 < y < \gamma < t < \theta]) = \int_\gamma^\theta \int_0^\gamma \hat{\xi}_n^{II}(y, t) f_{S_{II}}(y, t; \eta_0) dy\, dt.$$

The proof appears in Appendix A1.

Note that since $\hat{\xi}_n^I$ and $\hat{\xi}_n^{II}$ are UMVUE for $\xi(\theta)$ and $\xi(\eta)$, respectively, under the correct models, it holds that $E_\eta(\hat{\xi}_n^{II}) = \xi(\eta)$, $MSE(\hat{\xi}_n^{II}|\text{Model II}) = \text{Var}_\eta(\hat{\xi}_n^{II})$, and $E_{\eta_0}(\hat{\xi}_n^I) = \xi(\theta)$, with $MSE(\hat{\xi}_n^I|\text{Model I}) = \text{Var}_{\eta_0}(\hat{\xi}_n^I)$. Further, since by part (i) of Theorem 1, $\hat{\xi}_n^I$ is a biased estimator of $\xi(\eta) \equiv \xi(\theta)$ under Model II, it follows immediately that it is an inconsistent estimator and that

$$MSE(\hat{\xi}_n^I|\text{Model II}) = \text{Var}_\eta(\hat{\xi}_n^I) + (\xi(\theta)a + b_n)^2 > MSE(\hat{\xi}_n^{II}|\text{Model II}).$$

Proposition 1. *Let $\hat{\xi}_n^I$ be the Model* I *UMVUE of $\xi(\theta)$ as is given in* (13). *Then*

$$\text{Var}_{\eta_0}(\hat{\xi}_n^I) = \int_0^\theta \left[\frac{\xi'(t)g_0(0, t)}{nh(t)}\right]^2 f_{S_I}(t; \eta_0)\, dt. \tag{20}$$

See the proof in Appendix A2.

While the expression (20) for the $\text{Var}_{\eta_0}(\hat{\xi}_n^I)$ is exact, its explicit form depends much on the form of $\xi(\theta)$. Examples are provided below.

Example 3. Let $\xi(\theta) = E_{\eta_0}(X) = g_1(0, \theta)/g_0(0, \theta)$ as in Example 1. Then we have $\xi'(\theta) = (\theta - \xi(\theta))h(\theta)/g_0(0, \theta)$. Hence using (20), we obtain

$$\text{Var}_{\eta_0}(\hat{\xi}_n^I) = \frac{1}{n^2} \int_0^\theta (t - \xi(t))^2 f_{S_I}(t; \eta_0)\, dt.$$

Example 4. Let $\xi(\theta) = 1 - F_{\eta_0}(\tau) = g_0(\tau, \theta)/g_0(0, \theta)$ for some fixed τ, $0 < \tau < \theta$, as in Example 2. We have $\xi'(\theta) = h(\theta)g_0(0, \tau)/g_0^2(0, \theta)$. Using (7) and (13) in (20), we obtain that

$$\text{Var}_{\eta_0}(\hat{\xi}_n^I) = \frac{(1 - \xi(\theta))^2}{n(n-2)} \equiv \frac{(F_{\eta_0}(\tau))^2}{n(n-2)}.$$

Proposition 2. *Let $\hat{\xi}_n^{II}$ be the Model* II *UMVUE of $\xi(\gamma, \theta)$ as is given in* (15). *Then*

$$\text{Var}_\eta(\hat{\xi}_n^{II}) = \int_\gamma^\theta \left[\int_y^\theta \left[\frac{r_2(y, t)g_0(y,t)}{(n-1)h(t)}\right]^2 f_{X_{(n)}|X_{(1)}}(t; y, \theta)\, dt\right] f_{X_{(1)}}(y; \eta)\, dy$$

$$+ \int_\gamma^\theta \left[\frac{\xi_1(y, \theta)g_0(y, \theta)}{nh(y)}\right]^2 f_{X_{(1)}}(y; \eta)\, dy,$$

where by (17),

$$r_2(y, t) = \xi_2(y, t) - \frac{\xi_{12}(y, t)g_0(y, t)}{nh(y)} - \frac{\xi_1(y, t)h(t)}{nh(y)}.$$

See the proof in Appendix A3.

Example 5. Let $\xi(\gamma, \theta) = 1 - F_\eta(\tau) = g_0(\tau, \theta)/g_0(\gamma, \theta)$, $\gamma \leq \tau \leq \theta$ as in Example 2. Then

$$\hat{\xi}_n^{II}(y, t) = \left(1 - \frac{1}{n}\right) - \left(1 - \frac{2}{n}\right)\frac{g_0(y, \tau)}{g_0(y, t)}.$$

A direct application of Propositions 1 and 2, together with Remark 2 yields that

$$\text{Var}_\eta(\hat{\xi}_n^{II}) = \frac{1}{(n-1)(n-3)}\left(1 - 2\frac{n}{n-1}\xi + \frac{n}{n-2}\xi^2\right) + \frac{1}{n(n-2)}\xi^2$$

$$\approx \frac{(1-\xi)^2}{(n-1)(n-3)} + \frac{\xi^2}{n(n-2)},$$

where $\xi \equiv \xi(\eta) = P_\eta(X \geq \tau) = g_0(\tau, \theta)/g_0(\gamma, \theta)$.

2.4. Examples

We illustrate the details of this analysis in the case of the truncated $\text{Beta}(\alpha, 1)$ (with $\alpha \geq 1$ known) distribution. The p.d.f. of $\text{Beta}(\alpha, 1)$ is of the form given in (2) with $h(x) = \alpha x^{\alpha-1}$ and $g_0(\gamma, \theta) = \theta^\alpha - \gamma^\alpha$ for $0 < \gamma < \theta$, namely,

$$f(x; \eta) = \frac{\alpha x^{\alpha-1}}{\theta^\alpha - \gamma^\alpha}I[\gamma < x < \theta].$$

When Model I is assumed, $\gamma = \gamma_0 \equiv 0$ and hence

$$f(x; \eta_0) = \frac{\alpha x^{\alpha-1}}{\theta^\alpha}I[0 < x < \theta].$$

We are interested in estimating $\xi(\eta) = P_\eta(X > \tau)$ for some $\gamma < \tau < \theta$. Note that $\xi(\eta)$ is estimable under both Model I and Model II (simply take $I[X_1 > \tau]$ as the estimator). Clearly, in this case,

$$\xi(\eta) = \frac{\theta^\alpha - \tau^\alpha}{\theta^\alpha - \gamma^\alpha}I[\gamma < \tau < \theta], \tag{21}$$

which under Model I (with $\gamma = 0$) is expressed as

$$\xi(\eta_0) \equiv \xi(\theta) = 1 - \left(\frac{\tau}{\theta}\right)^\alpha I[0 < \tau < \theta]. \tag{22}$$

Following Examples 1 and 2, it is straightforward to verify that the UMVU estimators of this tail probability under Model I and Model II, respectively, are

$$\hat{\xi}_n^I = 1 - \left(1 - \frac{1}{n}\right)\left(\frac{\tau}{x_{(n)}}\right)^\alpha \qquad \text{for} \quad 0 < \tau < x_{(n)},$$

and

$$\hat{\xi}_n^{II} = \left(1 - \frac{1}{n}\right) - \left(1 - \frac{2}{n}\right)\frac{\tau^\alpha - x_{(1)}^\alpha}{x_{(n)}^\alpha - x_{(1)}^\alpha} \qquad \text{for} \quad x_{(1)} < \tau < x_{(n)}. \tag{23}$$

By construction, both estimators are unbiased for $\xi(\eta) = P_\eta(X > \tau)$ under their correct models. Furthermore, by Example 4,

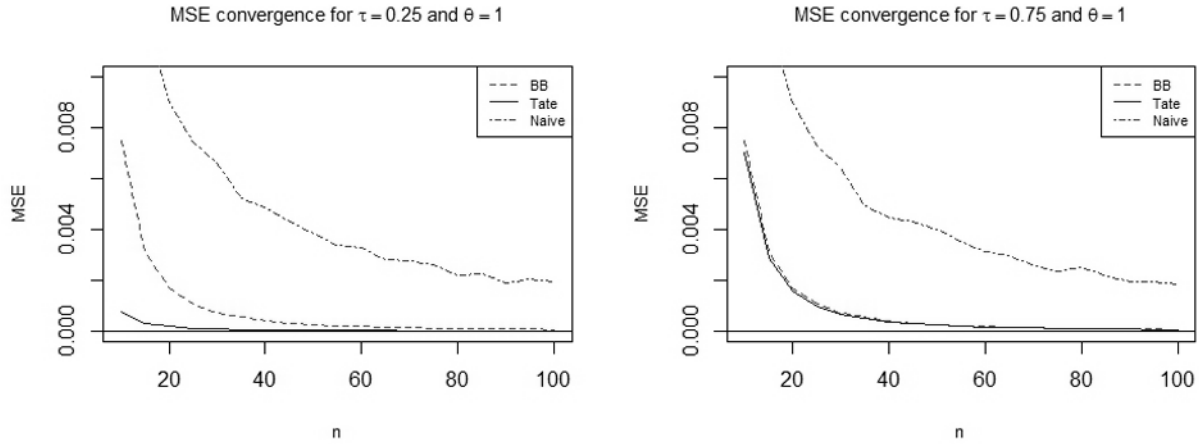$$MSE(\hat{\xi}_n^I|\text{Model I}) = \left(\frac{\tau^\alpha}{\theta^\alpha}\right)^2\frac{1}{n(n-2)},$$

and by Theorem 1, under Model II,

$$MSE(\hat{\xi}_n^I|\text{Model II}) = \text{Var}_\eta(\hat{\xi}_n^I) + (\xi(\theta)a + b_n)^2,$$

where b_n is as given in (19).

Using (9) one can show that $\hat{\xi}_n^{II}$ is a biased estimator w.r.t. Model I support for any finite sample size, where the bias term is of order $1/n$. Computing its MSE is complicated and an explicit expression was found only for the case that $\alpha = 1$, i.e., truncated uniform random variable.

Fig. 1. The MSE convergence of both estimators and the empirical quantile estimator (based on 1,000 iterations) for the tail probability w.r.t. Model I support, for $\tau \in \{0.25, 0.75\}$, $\theta = 1$ and $\alpha = 1$.
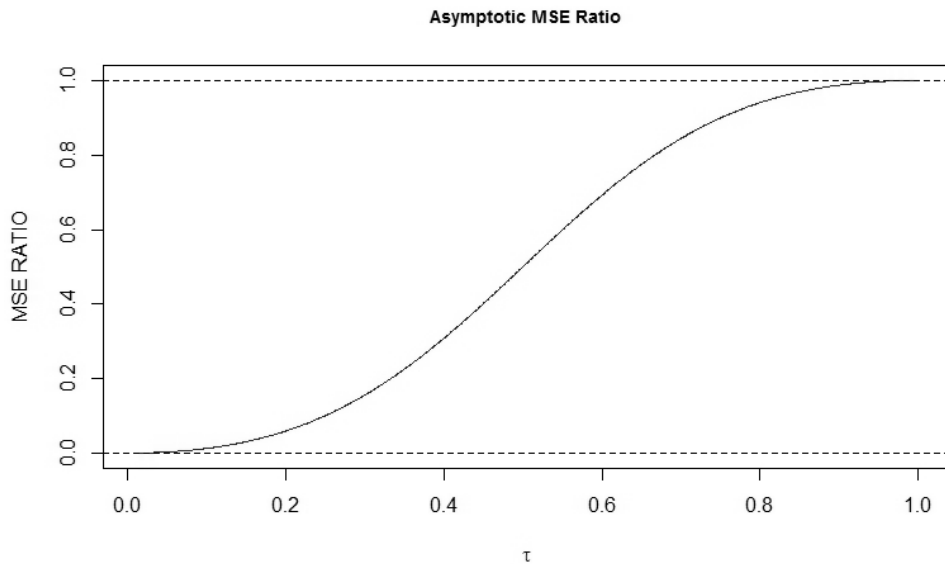
Fig. 2. Asymptotic MSE ratio (Tate/BB) of the estimators of the tail probability w.r.t. Model I support as a function of τ for $\theta = 1$ and $\alpha = 1$.

Lemma 1. *Let $\hat{\xi}_n^{II}$ be as in* (23) *with $\alpha = 1$. Then the MSE of $\hat{\xi}_n^{II}$ w.r.t. Model* I *support is given by*

$$\frac{2\,\tau^2\,n + \theta^2(n-1) - 2\tau\theta n}{\theta^2 n^2(n-3)}\,. \tag{24}$$

The proof appears in Appendix A4.

Finally, using Proposition 2, it follows that under Model II

$$MSE(\hat{\xi}_n^{II}\,|\,\text{Model II}) \approx \frac{1}{n-1}\left[\frac{(1-\xi)^2}{(n-3)} + \frac{\xi^2}{(n-2)}\right],$$

where $\xi \equiv \xi(\eta)$ is as given in (21). Figures 1 and 2 illustrate the behavior of the MSE of both estimators in Example 5 for finite and asymptotic sample sizes.

3. 'REGULAR' MODELS WITH A POSSIBLE LEFT TRUNCATION PARAMETER

In Section 3.1 we first present the Erlang distribution as a special case of the natural exponential family (NEF) and illustrate the effects of the possible truncation on the estimator of the 'regular' parameter. We then proceed, in Section 3.2 to discuss asymptotic aspects of this problem for general NEF. Finally, in Section 3.3 we illustrate the results with examples.

3.1. Case Study: the Erlang Distribution

Let $\gamma \in [0, \infty)$ be fixed, $\lambda > 0$, and let $k = 1, 2, \ldots$. Define

$$Q_k(\gamma, \lambda) = \int_\gamma^\infty \lambda^k\, x^{k-1}\, e^{-\lambda x}\, dx. \tag{25}$$

Note that Q_k is the k-fold incomplete gamma function, and, in particular, $Q_k(0, \lambda) = (k-1)!$ for any $\lambda > 0$. As such, we have

$$Q_k(\gamma, \lambda) = (k-1)! \sum_{i=0}^{k-1} \frac{(\lambda\gamma)^i}{i!} e^{-\lambda\gamma}. \tag{26}$$

We define the truncated version of the k-stage Erlang distribution, denoted here as Erlang(k, λ, γ), by the p.d.f.

$$f(x; \gamma, \lambda) = \frac{\lambda^k\, x^{k-1}}{Q_k(\gamma, \lambda)} e^{-\lambda x}\, I[\gamma < x < \infty]. \tag{27}$$

Note the similarity to the definition in Section 2 that appears in (2). By (27), the jth moment of the Erlang(k, λ, γ) distribution is

$$E(X^j | k, \gamma, \lambda) = \frac{1}{\lambda^j} \frac{Q_{k+j}(\gamma, \lambda)}{Q_k(\gamma, \lambda)}, \quad j = 0, 1, 2, \ldots. \tag{28}$$

We now consider the problem of estimating λ under Model I (i.e., $\gamma = 0$) versus Model II (i.e., $\gamma > 0$ unknown). To that end, we use (26) and (28) to calculate

$$\mu_0(\lambda) \equiv E(X | k, 0, \lambda) \qquad \text{and} \qquad \mu_\gamma(\lambda) \equiv E(X | k, \gamma, \lambda)$$

and to obtain that

$$\mu_0(\lambda) = \frac{k}{\lambda} \qquad \text{and} \qquad \mu_\gamma(\lambda) = \frac{k}{\lambda} \left[1 + \frac{(\lambda\gamma)^k e^{-\lambda\gamma}}{k Q_k(\gamma, \lambda)} \right]. \tag{29}$$

It can be easily seen from (29) (see also Section 3.2) that the maximum likelihood equation based on a sample of n observations from Erlang(k, λ, γ) should satisfy

$$\mu_0(\hat{\lambda}) = \bar{X}_n \qquad \text{and} \qquad \mu_\gamma(\hat{\lambda}) = \bar{X}_n.$$

That is, the MLE $\hat{\lambda}_n^I$ of λ under Model I (i.e., $\gamma = 0$) and the MLE $\hat{\lambda}_n^{II}$ and $\hat{\gamma}_n = X_{(1)}$ of λ and γ under Model II should satisfy, respectively,

$$\mu_0(\hat{\lambda}_n^I) = \bar{X}_n \qquad \text{and} \qquad \mu_{\hat{\gamma}_n}(\hat{\lambda}_n^{II}) = \bar{X}_n. \tag{30}$$

Hence, under Model II,

$$\hat{\lambda}_n^I = \frac{k}{\bar{X}_n} = \frac{k}{\mu_{\hat{\gamma}_n}(\hat{\lambda}_n^{II})} \xrightarrow{P_{II}} \frac{\lambda}{1 + \frac{(\lambda\gamma)^k e^{-\lambda\gamma}}{k\, Q_k(\gamma, \lambda)}} \neq \lambda,$$

and thus $\hat{\lambda}_n^I$ is inconsistent.

To assess the cross-model behavior of $\hat{\lambda}_n^{II}$ under Model I, we first define the bias when estimating the mean function. Define $B(\lambda, \gamma)$ by

$$B(\lambda, \gamma) \equiv \mu_\gamma(\lambda) - \mu_0(\lambda) = \frac{(\lambda\gamma)^k e^{-\lambda\gamma}}{k Q_k(\gamma, \lambda)}, \tag{31}$$

where the equality follows from (29). By Theorem 2 below, $\hat{\gamma}_n \equiv X_{(1)} \xrightarrow{P_{II}} \gamma$ under Model II and $\hat{\gamma}_n \xrightarrow{P_I} 0$ under Model I. Hence for every $\lambda > 0$ we have $B(\lambda, \hat{\gamma}_n) \xrightarrow{P_I} B(\lambda, 0) = 0$ under Model I. Since by (30) and (31)

$$\mu_{\hat{\gamma}_n}(\hat{\lambda}_n^{II}) = \mu_0(\hat{\lambda}_n^{II}) + B(\hat{\lambda}_n^{II}, \hat{\gamma}_n) = \bar{X}_n,$$

it follows that under Model I,

$$\mu_0(\hat{\lambda}_n^{II}) = \bar{X}_n - B(\hat{\lambda}_n^{II}, \hat{\gamma}_n) \xrightarrow{P_I} \mu_0(\lambda) - B(\lambda, 0) = \mu_0(\lambda). \qquad (32)$$

Since $\mu_0(\cdot)$ has a continuous inverse function we also obtain, by the continuous mapping theorem (van der Vaart, 2000) that $\hat{\lambda}_n^{II} \xrightarrow{P_I} \lambda$. In other words, we showed that $\hat{\lambda}_n^{II}$ is consistent and asymptotically unbiased estimator. To conclude this case study, based on the above results, it is clearly preferable, at least from asymptotic point of view, to use the MLE of λ w.r.t. Model II, i.e., $\hat{\lambda}_n^{II}$, to mitigate the possible existence of a left-truncation parameter. The following lemma, which is a special case of Theorem 3, summarizes the cross-model analysis of the case study presented above.

Lemma 2. *Let $X \sim \mathrm{Erlang}(k, \lambda, \gamma)$, where $\gamma \in [0, \infty)$ is a fixed parameter which designates the left truncation, $\lambda > 0$ is the parameter of interest and $k \in \mathbb{N}$. Let $\hat{\lambda}_n^I$ and $\hat{\lambda}_n^{II}$ be the MLE estimators of λ w.r.t. Model I and Model II support respectively. Then, under Model I support*

$$\hat{\lambda}_n^{II} \xrightarrow{P_I} \lambda,$$

while under Model II support

$$\hat{\lambda}_n^I \xrightarrow{P_{II}} \frac{\lambda}{1 + \frac{(\lambda\gamma)^k e^{-\lambda\gamma}}{k\,Q_k(\gamma,\lambda)}},$$

where Q_k is the k-fold incomplete gamma function (see (26)).

In the next subsection we discuss this observation for the general case of NEF distributions, for which the Erlang distribution is a special case.

3.2. Natural Exponential Families with Possible Left Truncation

Let $\gamma \in [0, \infty)$ be fixed, $h \colon [0, \infty) \to (0, \infty)$ be an absolutely continuous mapping with respect to the Lebesgue measure on the real line, and denote by $\mathcal{L}(\theta, \gamma)$ the Laplace transform of $h(x)\,dx$, i.e.,

$$\mathcal{L}(\theta, \gamma) = \int_\gamma^\infty e^{\theta x} h(x)\,dx.$$

Assume that $\Theta_\gamma \equiv \mathrm{int}\{\theta \in \mathcal{R}; \mathcal{L}(\theta, \gamma) < \infty\} \neq \phi$. Then the NEF generated by $h(x)\,dx$ is given by probability densities of the form

$$f(x; \theta, \gamma) = h(x) e^{\theta x - k(\theta, \gamma)} dx, \quad \theta \in \Theta_\gamma, \qquad (33)$$

where $k(\theta, \gamma) = \log \mathcal{L}(\theta, \gamma)$.

For any fixed $\gamma \in [0, \infty)$, it is well known that $k(\theta, \gamma)$ is a strictly convex real analytic function on Θ_γ and $k_j(\theta, \gamma) \equiv \partial^j k_j(\theta, \gamma)/\partial\theta^j$, $j = 1, 2, \ldots$, is the jth cumulant corresponding to $f(x; \theta, \gamma)$. Specifically, $\mu_\gamma(\theta) = k_1(\theta, \gamma)$ and $\sigma_\gamma^2(\theta) = k_2(\theta, \gamma)$ are the corresponding mean and variance. Note that if $0 \leq \gamma_1 < \gamma_2$ then $\Theta_{\gamma_1} \subseteq \Theta_{\gamma_2}$, and in particular, $\Theta_0 \subseteq \Theta_\gamma$ for all $\gamma > 0$. However, without loss of generality, we assume that $\Theta_0 = \Theta_\gamma$ for all $\gamma > 0$. Note that in the NEF terminology, $\mathcal{M}_\gamma \equiv k_1(\Theta_\gamma, \gamma)$ is the mean parameter space associated with the corresponding NEF, whereas $\mathcal{M}_0 \equiv k_1(\Theta_0, \gamma)$ satisfies $\mathcal{M}_0 \subset \mathcal{M}_\gamma$.

Unlike the Erlang distribution case, we are interested here only in the asymptotic behavior of the MLE of the natural parameter θ under Model I (with $\gamma = 0$) and under Model II (with $\gamma > 0$), as no explicit

relationship as in (29) generally exists. As in the previous section, the maximum likelihood equation based on a sample of n observations from (33) should satisfy, for Model I and Model II, respectively,

$$\mu_0(\hat{\theta}) = \bar{X}_n \qquad \text{and} \qquad \mu_\gamma(\hat{\theta}) = \bar{X}_n\,,$$

provided that $\bar{X}_n \in \mathcal{M}_\gamma$ (a.s.). That is, the MLE $\hat{\theta}_n^I$ of θ under Model I and the MLE $\hat{\theta}_n^{II}$ and $\hat{\gamma}_n = X_{(1)}$ of θ and γ under Model II should satisfy, respectively,

$$\mu_0(\hat{\theta}_n^I) = \bar{X}_n \qquad \text{and} \qquad \mu_{\hat{\gamma}_n}(\hat{\theta}_n^{II}) = \bar{X}_n\,.$$

More specifically, under Model II, the MLE $\hat{\theta}_n^{II}$ of θ satisfies $k_1(\hat{\theta}_n^{II}, \hat{\gamma}_n) = \bar{X}_n$, with $\hat{\gamma}_n \equiv X_{(1)}$, whereas under Model I, the MLE $\hat{\theta}_n^I$ for θ satisfies $k_1(\hat{\theta}_n^I, 0) = \bar{X}_n$.

In the following theorem we restate the results of Proposition 2 of Dubinin and Vardeman (2003).

Theorem 2. *If Model* I *holds (i.e.* $\gamma = 0$) *then, as* $n \to \infty$,

$$\sqrt{n}(\hat{\theta}_n^I - \theta) \xrightarrow{D} \mathcal{N}\left(0, \frac{1}{k_2(\theta, 0)}\right).$$

Furthermore, if Model II *holds (i.e.,* $\gamma > 0$) *and* $h(x)$ *is right-continuous at* 0*, then* $(\sqrt{n}(\hat{\theta}_n^{II} - \theta)', n(\hat{\gamma}_n - \gamma))'$ *are asymptotically independent with marginal distributions* $\mathcal{N}(0, k_2(\theta, \gamma)^{-1})$ *and* $\mathrm{Exp}(\alpha)$ *with* $\alpha = f(\gamma; \theta, \gamma) = h(\gamma)e^{\theta\gamma - k(\theta, \gamma)}$, *respectively.*

Example 6. Let $X_1, \ldots, X_n \sim \mathrm{Exp}(\lambda, \gamma)$, so that $\lambda = -\theta$, where θ is the natural parameter of the distribution, and where γ is the truncation parameter. It can be shown that

$$k_1(\theta, \gamma) = \gamma - \frac{1}{\theta}\,.$$

It can also be shown that $k_2(\theta, \gamma) = k_2(\theta, 0) = \frac{1}{\theta^2}$. This observation means that in the exponential case, truncation is equivalent to shifting by a factor of γ. Bar-Lev and Boukai (2009) showed that this is the only case for which this unusual property holds.

Consequently, the maximum likelihood estimating equation for Model I is $-\frac{1}{\theta} = \bar{X}_n$, and for Model II is $-\frac{1}{\theta} + X_{(1)} = \bar{X}_n$, since $X_{(1)}$ is the MLE of γ. Hence the MLE for the natural parameter is, under Model I and Model II, respectively,

$$\hat{\theta}_n^I = -\frac{1}{\bar{X}_n} \qquad \text{and} \qquad \hat{\theta}_n^{II} = \frac{1}{X_{(1)} - \bar{X}_n}\,.$$

Example 7. Let $X_1, \ldots, X_n \sim \mathrm{Erlang}(2, \lambda, \gamma)$, where $\lambda = -\theta$, θ is the natural parameter, and γ is the truncation parameter. Similarly to the exponential distribution example, $k_1(\theta, 0) = -\frac{2}{\theta}$, hence $\hat{\theta}_n^I = -\frac{2}{\bar{X}_n}$. For $\gamma > 0$, we obtain that $k_1(\theta, \gamma) = -\frac{2}{\theta} + \gamma - \frac{\gamma}{1 - \theta\gamma}$. Some algebra then yields

$$\hat{\theta}_n^{II} = -4\left(\bar{X}_n - 2X_{(1)} + \sqrt{4X_{(1)}(\bar{X}_n - X_{(1)}) + \bar{X}_n^2}\right)^{-1}.$$

Before we discuss cross-model results, we need the following notation. For every $\gamma > 0$, define the inverse function $\kappa_\gamma^{-1} \colon \mathcal{M}_\gamma \mapsto \Theta_\gamma$ by $\kappa_\gamma^{-1}(\mu) = \theta$ for the unique θ that satisfies $k_1(\theta, \gamma) = \mu_\gamma$. Note that this θ is indeed unique since $k_1(\cdot, \gamma)$ is strictly monotonically increasing. It follows from Proposition 2 of Dubinin and Vardeman (2003) that

$$\kappa_{\gamma_1}^{-1}(\mu) < \kappa_{\gamma_2}^{-1}(\mu) \tag{34}$$

for all $\gamma_1 < \gamma_2$, since $k_1(\theta, \gamma)$ is strictly monotonically increasing in γ. Finally, we note that using this notation we have

$$\hat{\theta}_n^I = \kappa_0^{-1}(\bar{X}_n) \qquad \text{and} \qquad \hat{\theta}_n^{II} = \kappa_{X_{(1)}}^{-1}(\bar{X}_n)\,.$$

We first show that $\hat{\theta}_n^I$ is a biased estimator of θ under Model II, with bias that does not vanish asymptotically. We then show that $\hat{\theta}_n^{II}$ is an asymptotically efficient estimator of θ under Model I.

Theorem 3. *Assume that Model* II *holds. Then*

$$\hat{\theta}_n^I \xrightarrow{a.s.} \kappa_0^{-1}(\mu_\gamma) < \kappa_\gamma^{-1}(\mu_\gamma) = \theta. \tag{35}$$

If Model I *holds, and* $h(x)$ *is right-continuous at* 0, *then*

$$\sqrt{n}(\hat{\theta}_n^{II} - \theta) \xrightarrow{D} \mathcal{N}\left(0, \frac{1}{k_2(\theta, 0)}\right).$$

See the proof in Appendix A5.

3.3. Examples

Consider first the scenario of the negative exponential distribution (i.e., as a special case of the k-stage Erlang distribution discussed above, but with $k = 1$). Recall that when the MLE for θ was derived under Model I, while Model II actually holds, we obtained that $\mu_\gamma = -\frac{1}{\theta} + \gamma = \mu_0 + \gamma$. Therefore,

$$-\frac{1}{\bar{X}_n} - \theta \xrightarrow{a.s} \frac{\theta^2 \gamma}{1 - \theta\gamma}, \tag{36}$$

which means that the sequence $\sqrt{n}(-\frac{1}{\bar{X}_n} - \theta)$ goes to infinity as $n \to \infty$.

MSE Ratio For Finite Sample Sizes
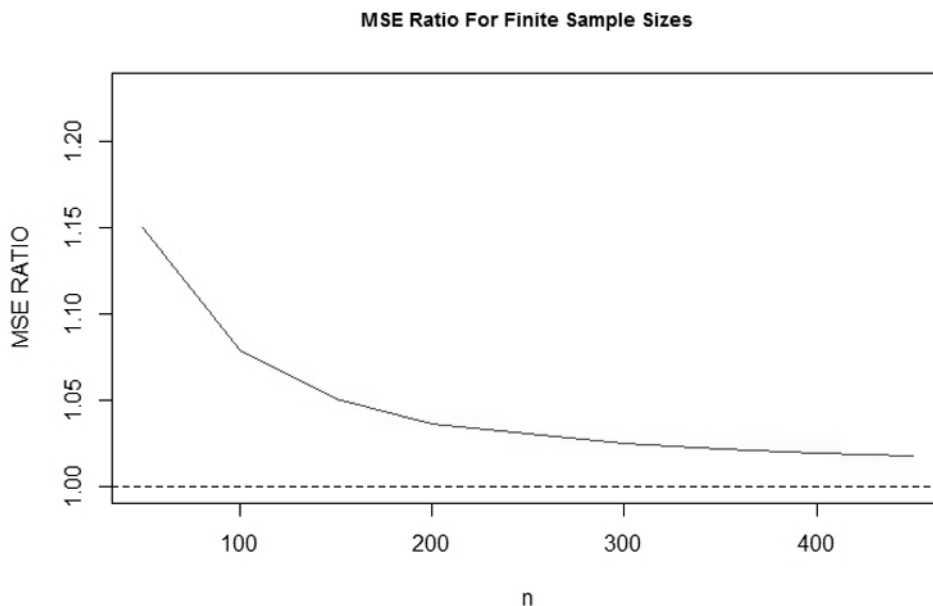
Fig. 3. The MSE ratio (Model II/Model I) for various sample sizes where the true model is Model I. For each sample size n, 100,000 times n random values from $\text{Erlang}(2, 1, 0)$ distribution were drawn and both MLEs were calculated. The MSEs were calculated by averaging the MLEs of these 100,000 runs, for each sample size.

Consider now the case that the MLE for θ was derived under Model II, while Model I actually holds. We start with the finite-sample behavior of the estimator $\hat{\theta}_n^{II}$ under Model I. It can be shown (see (32)) that its expected value under Model I is $\frac{n\theta}{n-2}$, and therefore its bias is given by $\frac{2}{n-2}$. Taking the limit as $n \to \infty$ shows that the estimator $\hat{\theta}_n^{II}$ is asymptotically unbiased under Model I. Furthermore, direct calculations show that

$$MSE(\hat{\theta}_n^{II}|\text{Model I}) = \frac{(n(n+4) - 12)\theta^2}{(n-2)^2(n-3)}.$$

The asymptotic behavior of the MLE for the natural parameter θ follows from Theorem 3. Specifically, we have

$$\sqrt{n}\left(\frac{1}{X_{(1)} - \bar{X}_n} - \theta\right) \xrightarrow{D} \mathcal{N}(0, \theta^2). \tag{37}$$

We now discuss the Erlang-2 distribution discussed in Example 7. Consider, for instance, the scenario in which the MLE was derived erroneously under Model I support, while Model II holds. It can be shown that in such a case,

$$-\frac{2}{\bar{X}_n} - \theta \xrightarrow{a.s} \frac{-\theta^3\gamma^2}{2 + \theta\gamma(\theta\gamma - 2)}. \tag{38}$$

Hence, the sequence $\sqrt{n}(-\frac{2}{\bar{X}_n} - \theta)$ goes to infinity as $n \to \infty$. We now consider the scenario in which the MLE was derived under Model II support, but Model I holds. The asymptotic behavior of the MLE for the natural parameter can be derived from Theorem 3 and is similar to (37). Note that the finite-sample behavior is analytically complicated and is demonstrated using simulations, see Figure 3.

4. CONCLUSIONS

In this work we analyzed the effect of addition of a left-truncation parameter on estimation in continuous distribution functions. We discussed two main settings: general continuous right-truncated models with possible left truncation, and exponential families with possible left truncation. We investigated the effects of model misspecification on UMVU estimators for the first setting, and on maximum likelihood estimators for the second setting. For both settings, we discussed both finite-sample properties and asymptotic behavior of the estimators.

In both settings we showed that mistakenly assuming Model I, when the true model is Model II, leads to a biased estimation with bias that does not vanish asymptotically. On the other hand, assuming Model II, when the true model is Model I, leads to an asymptotically unbiased estimation, which is, at least for the exponential family setting, also asymptotically efficient. Nonetheless, it is important to note that estimators constructed under Model II can be more complicated and that there can be a significant efficiency price for estimating under Model II when Model I is correct. In conclusion, based on the results described above, we recommend using Model II when there is a good reason to suspect that the model involves left truncation. When there is no reason to assume left truncation, we recommend to use Model I.

APPENDIX: PROOFS

A1. Proof of Theorem 1

Part (i): The expectation of Tate's UMVU estimator w.r.t. False Mode II can be written as follows

$$E(\hat{\xi}_n^I | \text{Model II}) = \int_\gamma^\theta \left(\xi(t) + \frac{\xi'(t)g_0(0,t)}{nh(t)}\right) f_{S_I}(t;\eta)\, dt = \int_\gamma^\theta \xi(t) f_{S_I}(t;\eta)\, dt$$
$$+ \int_\gamma^\theta \xi'(t)\left(\frac{g_0(0,t)}{nh(t)} - \frac{g_0(\gamma,t)}{nh(t)}\right) f_{S_I}(t;\eta)\, dt + \int_\gamma^\theta \xi'(t)\frac{g_0(\gamma,t)}{nh(t)} f_{S_I}(t;\eta)\, dt.$$

Note that $F_{S_I}(t;\eta) = \left(\frac{g_0(\gamma,t)}{g_0(\gamma,\theta)}\right)^n \equiv (F_\eta(t))^n$, thus $f_{S_I}(t;\eta) = \frac{nh(t)}{g_0(\gamma,\theta)}(F_\eta(t))^{n-1}$. Therefore, one can rewrite the equation above in the following way

$$E_\eta(\hat{\xi}_n^I) = \int_\gamma^\theta \xi(t) f_{S_I}(t;\eta)\, dt + \frac{g_0(0,\gamma)}{g_0(\gamma,\theta)}\int_\gamma^\theta \xi'(t)\frac{g_0(\gamma,\theta)}{nh(t)} f_{S_I}(t;\eta)\, dt + \int_\gamma^\theta \xi'(t)(F_\eta(t))^{n-1}\, dt.$$

By rearranging the equation, using the identities stated before and integration by parts for the last term, one can show that since $F_{S_I}(\gamma) = 0$, $F_{S_I}(\theta) = 1$,

$$E_\eta(\hat{\xi}_n^I) = \xi(t)F_{S_I}(t)\Big|_\gamma^\theta + \frac{g_0(0,\gamma)}{g_0(\gamma,\theta)}\int_\gamma^\theta \xi'(t)(F_\eta(t))^{n-1}\,dt = \xi(\theta) + b_n,$$

with $a = \frac{g_0(0,\gamma)}{g_0(\gamma,\theta)}$ and $b_n = a\int_\gamma^\theta \xi'(t)(F_\eta(t))^{n-1}\,dt$. Again, by integration by parts we obtain that

$$b_n = a\left(\xi(\theta) - \int_\gamma^\theta \xi(t)(n-1)h(t)\frac{(g_0(\gamma,t))^{n-2}}{(g_0(\gamma,\theta))^{n-1}}\,dt\right),$$

which completes the proof of the first part.

Part (ii): The proof is straightforward, and is given by

$$E(\hat{\xi}_n^{II}|\text{Model I}) = \int_0^\theta \int_0^t \hat{\xi}_n^{II}(y,t)f_{S_{II}}(y,t;\eta_0)\,dy\,dt = \int_\gamma^\theta \int_\gamma^t \hat{\xi}_n^{II}(y,t)f_{S_{II}}(y,t;\eta_0)\,dy\,dt$$

$$+ \int_0^\gamma \int_0^t \hat{\xi}_n^{II}(y,t)f_{S_{II}}(y,t;\eta_0)\,dy\,dt + \int_\gamma^\theta \int_0^\gamma \hat{\xi}_n^{II}(y,t)f_{S_{II}}(y,t;\eta_0)\,dy\,dt$$

$$= \xi(\eta) + d_{1n} + d_{2n}.$$

$\square$

A2. Proof of Proposition 1

The expectation of the squared Tate's UMVU estimator w.r.t. Model I support can be explicitly written as

$$E((\hat{\xi}_n^I)^2|\text{Model I}) = E_{\eta_0}\left(\xi^2(t) + \frac{2\xi(t)\xi'(t)g_0(0,t)}{nh(t)} + \left(\frac{\xi'(t)g_0(0,t)}{nh(t)}\right)^2\right).$$

Note that since $\frac{g(0,t)}{nh(t)}f_{S_I}(t;\eta_0) = F_{S_I}(t;\eta_0)$, one can show that

$$E_{\eta_0}\left((\hat{\xi}_n^I)^2\right) = \int_0^\theta \xi^2(t)f_{S_I}(t;\eta_0)\,dt + \int_0^\theta 2\xi(t)\xi'(t)F_{S_I}(t;\eta_0)\,dt + \int_0^\theta \left(\frac{\xi'(t)g_0(0,t)}{nh(t)}\right)^2 f_{S_I}(t;\eta_0)\,dt.$$

By using integration by parts and the facts that $F_{S_I}(0) = 0$ and $F_{S_I}(\theta) = 1$, one can show that

$$E_{\eta_0}\left((\hat{\xi}_n^I)^2\right) = \xi^2(t)F_{S_I}(t)\Big|_0^\theta + \int_0^\theta \left(\frac{\xi'(t)g_0(0,t)}{nh(t)}\right)^2 f_{S_I}(t;\eta_0)\,dt$$

$$= \xi^2(\theta) + \int_0^\theta \left(\frac{\xi'(t)g_0(0,t)}{nh(t)}\right)^2 f_{S_I}(t;\eta_0)\,dt.$$

$\square$

A3. Proof of Proposition 2

Let $\hat{\xi}_n^{II}$ be the Model II UMVUE of $\xi(\gamma,\theta)$ as is given in (15). Then using the law of total variance we can express the variance of $\hat{\xi}_n^{II}$ in the following way

$$\text{Var}_\eta(\hat{\xi}_n^{II}) = E_\eta\big(\text{Var}_\eta(\hat{\xi}_n^{II}|X_{(1)})\big) + \text{Var}_\eta\big(E_\eta(\hat{\xi}_n^{II}|X_{(1)})\big).$$

Starting with the first term, by using (17), we know that

$$\hat{\xi}_n^{II}(y,t) = r(y,t) + \frac{r_2(y,t)g_0(y,t)}{(n-1)h(t)},$$

where $r_2 = \partial r / \partial \theta$. Therefore, by using Proposition 1 we can immediately calculate $\mathrm{Var}_\eta(\hat{\xi}_n^{II}|X_{(1)} = y)$, which is simply an integration of the squared second term in the expression above w.r.t. the density function of $X_{(n)}$ over Model II support, i.e.,

$$\mathrm{Var}_\eta(\hat{\xi}_n^{II}|X_{(1)} = y) = \int_y^\theta \left(\frac{r_2(y,t)g_0(y,t)}{(n-1)h(t)} \right)^2 f_{X_{(n)}}(t;\eta)\, dt \,.$$

In order to compute $E_\eta\big(\mathrm{Var}_\eta(\hat{\xi}_n^{II}|X_{(1)})\big)$ we have to integrate $\mathrm{Var}_\eta(\hat{\xi}_n^{II}|X_{(1)} = y)$ over all possible values of $X_{(1)}$ w.r.t. Model II support. More specifically,

$$\int_\gamma^\theta \left[\int_y^\theta \left(\frac{r_2(y,\,t)g_0(y,t)}{(n-1)h(t)} \right)^2 f_{X_{(n)}}(t;\eta)\, dt \right] f_{X_{(1)}}(y;\eta)\, dy,$$

which completes the first term of $\mathrm{Var}_\eta(\hat{\xi}_n^{II})$.

Proceeding to the second term, using (16) we know that

$$E_\eta(\hat{\xi}_n^{II}|X_{(1)} = y) = \xi(y,\theta) - \frac{\xi_1(y,\theta)g_0(y,\theta)}{nh(y)} \,.$$

Utilizing Proposition 1 to calculate the variance of the expression presented above, one should integrate the squared second term of this expression w.r.t. the density function $X_{(1)}$ over Model II support, i.e.,

$$\mathrm{Var}_\eta\big(E_\eta(\hat{\xi}_n^{II}|X_{(1)})\big) = \int_\gamma^\theta \left(\frac{\xi_1(y,\theta)g_0(y,\theta)}{nh(y)} \right)^2 f_{X_{(1)}}(y;\eta)\, dy,$$

which completes the computation of the second term of $\mathrm{Var}_\eta(\hat{\xi}_n^{II})$ and concludes the proof. $\square$

A4. Proof of Lemma 1

Let $\xi(\eta)$ be as in (22) but with $\alpha = 1$, i.e., $\xi(\eta) = \frac{\theta - \tau}{\theta}$. We would like to compute

$$MSE(\hat{\xi}_n^{II}|\text{Model I}) = E_{\eta_0}\left(\hat{\xi}_n^{II} - \frac{\theta - \tau}{\theta} \right)^2 = E_{\eta_0}\big(\hat{\xi}_n^{II}\big)^2 - \frac{2(\theta - \tau)}{\theta} E_{\eta_0}\hat{\xi}_n^{II} + \left(\frac{\theta - \tau}{\theta} \right)^2. \quad (39)$$

Since $\alpha = 1$, we have $h(x) = 1$ and $g_0(0,\theta) = \theta$. Hence, from (9), we have

$$f_{S_{II}}(y,t;\eta_0) = \frac{n(n-1)(t-y)^{n-2}}{\theta^n}, \qquad \text{for} \quad 0 \le y < t \le \theta \,.$$

We also have

$$\hat{\xi}_n^{II}(y,t) = \left(1 - \frac{1}{n}\right) - \left(1 - \frac{2}{n}\right)\frac{\tau - y}{t - y} \qquad \text{for} \quad y < \tau < t \,.$$

Hence

$$\begin{aligned}
E_{\eta_0}\hat{\xi}_n^{II} &= \int_0^\theta \int_0^t \hat{\xi}_n^{II}(y,t) f_{S_{II}}(y,t;\eta_0)\, dy\, dt \\
&= \int_0^\theta \int_0^t \left(\left(1 - \frac{1}{n}\right) - \left(1 - \frac{2}{n}\right)\frac{\tau - y}{t - y} \right) \frac{n(n-1)(t-y)^{n-2}}{\theta^n}\, dy\, dt \\
&= \int_0^\theta \int_0^t \left(C_1(t-y)^{n-2} + C_2(t-y)^{n-3} + C_3 y(t-y)^{n-3} \right) dy\, dt \\
&= \int_0^\theta \left(C_1 D(n-2,0) + C_2 D(n-3,0) + C_3 D(n-3,1)|_{y=0}^t \right) dt \,,
\end{aligned}$$

where

$$C_1 \equiv \left(1 - \frac{1}{n}\right)\frac{n(n-1)}{\theta^n} = \frac{(n-1)^2}{\theta^n},$$

$$C_2 \equiv -\left(1 - \frac{2}{n}\right)\frac{\tau n(n-1)}{\theta^n} = -\tau\frac{(n-1)(n-2)}{\theta^n},$$

$$C_3 \equiv \left(1 - \frac{2}{n}\right)\frac{n(n-1)}{\theta^n} = -\frac{C_2}{\tau},$$

and where

$$D(n,0) \equiv \int (t-y)^n\, dy = -\frac{1}{(n+1)}(y-t)^{n+1},$$

$$D(n,1) \equiv \int (t-y)^n y\, dy = -\frac{(t-y)^{n+1}\,(t+(n+1)y)}{(n+1)(n+2)},$$

$$D(n,2) \equiv \int (t-y)^n y^2\, dy = -\frac{(t-y)^{n+1}\left(8t^2 + 8t(n+1)y + 4(n+1)(n+2)y^2\right)}{4(n+1)(n+2)(n+3)}.$$

Hence

$$E_{\eta_0}\hat{\xi}_n^{II}(y,t) = \frac{C_1}{n-1}\int_0^\theta t^{n-1}\,dt + \frac{C_2}{n-2}\int_0^\theta t^{n-2}\,dt + \frac{C_3}{(n-1)(n-2)}\int_0^\theta t^{n-1}\,dt$$

$$= \left(\frac{C_1}{n-1} + \frac{C_3}{(n-1)(n-2)}\right)\frac{\theta^n}{n} + \frac{C_2}{n-2}\frac{\theta^{n-1}}{n-1} = 1 - \frac{\tau}{\theta}.$$

We now compute $E_{\eta_0}\left((\hat{\xi}_n^{II})^2\right)$:

$$E_{\eta_0}\left((\hat{\xi}_n^{II})^2\right) = \int_0^\theta \int_0^t \left(\hat{\xi}_n^{II}(y,t)\right)^2 f_{S_{II}}(y,t;\eta_0)\,dy\,dt$$

$$= \int_0^\theta \int_0^t \left(\left(1-\frac{1}{n}\right) - \left(1-\frac{2}{n}\right)\frac{\tau-y}{t-y}\right)^2 \frac{n(n-1)(t-y)^{n-2}}{\theta^n}\,dy\,dt$$

$$= \int_0^\theta \left(E_1 D(n-2,0) + E_2 D(n-4,0) + E_3 D(n-4,2)\right)\big|_{y=0}^t \, dt$$

$$+ \int_0^\theta \left(E_{1,2} D(n-3,0) + E_{1,3} D(n-3,1) + E_{1,2} D(n-4,1)\right)\big|_{y=0}^t \, dt,$$

where

$$E_1 \equiv \left(1-\frac{1}{n}\right)^2\frac{n(n-1)}{\theta^n} = \frac{(n-1)^3}{\theta^n\, n},$$

$$E_2 \equiv \left(1-\frac{2}{n}\right)^2\frac{\tau^2 n(n-1)}{\theta^n} = -\frac{\tau^2\,(n-1)\,(n-2)^2}{\theta^n\, n},$$

$$E_3 \equiv \left(1-\frac{2}{n}\right)^2\frac{n(n-1)}{\theta^n} = \frac{(n-1)\,(n-2)^2}{\theta^n\, n},$$

$$E_{1,2} \equiv -2\tau\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\frac{n(n-1)}{\theta^n} = -\frac{2\,\tau(n-1)^2\,(n-2)}{\theta^n\, n},$$

$$E_{1,3} \equiv 2\tau\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\frac{n(n-1)}{\theta^n} = \frac{2\,(n-1)^2\,(n-2)}{\theta^n\, n},$$

$$E_{2,3} \equiv -2\tau\left(1-\frac{2}{n}\right)^2\frac{n(n-1)}{\theta^n} = -\frac{2\,\tau(n-2)^2(n-1)}{\theta^n\, n}.$$

Hence

$$E_{\eta_0}\left((\hat{\xi}_n^{II})^2\right) = -\int_0^\theta \left(E_1\frac{(t-y)^{n-1}}{n-1} + E_2\frac{(t-y)^{n-3}}{n-3}\right)\bigg|_{y=0}^t \, dt$$

$$- \int_0^\theta \left(E_3\frac{(t-y)^{n-3}(8t^2 + 8(n-3)ty + 4(n-3)(n-2)y^2)}{4(n-3)(n-2)(n-1)} + E_{1,2}\frac{(t-y)^{n-2}}{n-2}\right)\bigg|_{y=0}^t \, dt$$

$$- \int_0^\theta \left(E_{1,3} \frac{(t-y)^{n-2}(t+(n-2)y)}{(n-2)(n-1)} + E_{2,3} \frac{(t-y)^{n-3}(t+(n-3)y)}{(n-3)(n-2)} \right)\Bigg|_{y=0}^t dt$$

$$= \int_0^\theta \left(\frac{E_1 t^{n-1}}{n-1} + \frac{E_2 t^{n-3}}{n-3} + \frac{8 E_3 t^{n-1}}{4(n-3)(n-2)(n-1)} \right) dt$$

$$+ \int_0^\theta \left(\frac{E_{1,2} t^{n-2}}{n-2} + \frac{E_{1,3} t^{n-1}}{(n-2)(n-1)} + \frac{E_{2,3} t^{n-2}}{(n-3)(n-2)} \right) dt$$

$$= \left(\frac{E_1}{n-1} + \frac{2 E_3}{(n-3)(n-2)(n-1)} + \frac{E_{1,3}}{(n-2)(n-1)} \right) \frac{\theta^n}{n}$$

$$+ \left(\frac{E_{1,2}}{n-2} + \frac{E_{2,3}}{(n-3)(n-2)} \right) \frac{\theta^{n-1}}{n-1} + \frac{E_2 \theta^{n-2}}{(n-3)(n-2)}$$

$$= \frac{n^3 - 3 n^2 + n - 1}{n^2 (n-3)} - \frac{2 \tau (n^2 - 3 n + 1)}{\theta n (n-3)} + \frac{\tau^2 (n-1)(n-2)}{\theta^2 n (n-3)} .$$

Substituting the expressions we obtained for $E_{\eta_0} \hat{\xi}_n^{II}$ and $E_{\eta_0}\big((\hat{\xi}_n^{II})^2\big)$ in (39), and some simplifying we obtain the result given in (24). $\qquad\square$

A5. Proof of Theorem 3

The first assertion follows from the continuous mapping theorem, see van der Vaart (2000). Indeed,

$$\hat{\theta}_n^I = \kappa_0^{-1}(\bar{X}_n) \xrightarrow{a.s.} \kappa_0^{-1}(\mu_\gamma).$$

The inequality in (35) follows from (34).

We now move to the second assertion. Note that when Model I holds

$$\sqrt{n}(\hat{\theta}_n^{II} - \theta) = \sqrt{n}(\hat{\theta}_n^{II} - \hat{\theta}_n^I) + \sqrt{n}(\hat{\theta}_n^I - \theta).$$

Since by Theorem 2, $\sqrt{n}(\hat{\theta}_n^I - \theta) \xrightarrow{D} \mathcal{N}\big(0, \frac{1}{k_2(\theta,0)}\big)$, it is enough to show that $\sqrt{n}(\hat{\theta}_n^{II} - \hat{\theta}_n^I) \xrightarrow{D} 0$.

We define the function $g \colon \mathcal{M}_0 \times \mathbb{R} \mapsto \Theta_0$ by $g(\mu, \gamma) = \theta$ for the unique θ that solves $k_1(\theta, \gamma) = \mu$. It follows from Proposition 2 of Dubinin and Vardeman (2003) that this θ is unique and that $g(\mu, \gamma)$ is continuously differentiable. Note that $\hat{\theta}_n^I = g(\bar{X}_n, 0)$ and $\hat{\theta}_n^{II} = g(\bar{X}_n, X_{(1)})$. Write

$$\sqrt{n}(\hat{\theta}_n^{II} - \hat{\theta}_n^I) = \sqrt{n}\big(g(\bar{X}_n, X_{(1)}) - g(\bar{X}_n, 0)\big) = \sqrt{n} \frac{\partial}{\partial \gamma} g(\bar{X}_n, r) X_{(1)}$$

for some $0 < r < X_{(1)}$. Note that $X_{(1)} = o_p(n^{-1/2})$ and that $\frac{\partial}{\partial \gamma} g(\bar{X}_n, r) = O_p(1)$ because the function is bounded in the vicinity of 0. Hence we obtain that

$$\sqrt{n}(\hat{\theta}_n^{II} - \hat{\theta}_n^I) = n^{1/2} O_p(1) o_p(n^{-1/2}) \xrightarrow{P_I} 0,$$

which concludes the proof. $\qquad\square$

ACKNOWLEDGMENTS

REFERENCES

1. S. K. Bar-Lev and B. Boukai, "Minimum Variance Unbiased Estimation for Families of Distributions Involving Two Truncation Parameters", J. Statist. Plann. Inf. **12**, 379−384 (1985).
2. S. K. Bar-Lev and B. Boukai, "A Characterization of the Exponential Distribution by Means of Coincidence of Location and Truncated Densities", Statist. Papers **50** (2), 403−405 (2009).
3. P. J. Bickel, "Parametric Robustness: Small Biases can be Worthwhile", Ann. Statist. **12** (3), 864−879 (1984).
4. K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference : A Practical Information-Theoretic Approach* (Springer, 2002).
5. G. Claeskens and N. L. Hjort, *Model Selection and Model Averaging* (Cambridge Univ. Press, 2008).
6. T. M. Dubinin and S. B. Vardeman, "Likelihood-Based Inference in Some Continuous Exponential Families with Unknown Threshold Parameters", J. Amer. Statist. Assoc. **98** (463), (2003).
7. R. C. Elandt-Johnson, *Survival Models and Data Analysis* (Wiley, 1999).
8. J. F. Lawless, *Statistical Models and Methods for Lifetime Data* (Wiley, 2003).
9. R. F. Tate, "Unbiased Estimation: Functions of Location and Scale Parameters", Ann. Math. Statist. **30** (2), 341−366 (1959).
10. A. W. van der Vaart, *Asymptotic Statistics* (Cambridge Univ. Press, 2000).
11. H. White, "Maximum Likelihood Estimation of Misspecified Models", Econometrica **50**, 1−25 (1982).