Ann. Inst. Statist. Math.
36 (1984), Part A, 217-222

LARGE SAMPLE PROPERTIES OF THE MLE AND MCLE FOR THE NATURAL PARAMETER OF A TRUNCATED EXPONENTIAL FAMILY

SHAUL K. BAR-LEV

(Received Feb. 25, 1983; revised Sept. 27, 1983)

Summary

Consider a truncated exponential family of absolutely continuous distributions with natural parameter θ and truncation parameter γ. Strong consistency and asymptotic normality are shown to hold for the maximum likelihood and maximum conditional likelihood estimates of θ with γ unknown. Moreover, these two estimates are also shown to have the same limiting distribution, coinciding with that of the maximum likelihood estimate for θ when γ is assumed to be known.

1. Introduction

Let

$$f(x:\theta,\gamma)=a(x)\exp\{\theta u(x)\}/b(\theta,\gamma)\,,\qquad c<\gamma\leqq x<d \tag{1.1}$$

be a two-parameter density with respect to Lebesgue measure on the real line. Here, $-\infty\leqq c<d\leqq\infty$ are known, $a(x)$ is nonnegative and continuous a.s., and $u(x)$ is absolutely continuous with $du(x)/dx\not\equiv 0$ over (γ, d) for $\gamma\in\Gamma\equiv(c, d)$. For each fixed $\gamma\in\Gamma$, let $\Theta(\gamma)$ denote the set of all values of θ for which (1.1) is a density, i.e.

$$\Theta(\gamma)=\left\{\theta:\ 0<b(\theta,\gamma)=\int_\gamma^d a(x)\exp\{\theta u(x)\}dx<\infty\right\}\,,\qquad \gamma\in(c, d)\,.$$

Clearly, $\Theta(\gamma_1)\subset\Theta(\gamma_2)$ for $\gamma_1,\gamma_2\in\Gamma$ with $\gamma_1<\gamma_2$. Throughout this paper we shall assume that for any $\gamma\in\Gamma$, $\Theta(\gamma)\equiv\Theta$ is a nonempty open subset of the real line.

The family of distributions with densities of the form (1.1), where (θ,γ) ranges over $\Theta\times\Gamma$, is generally referred to as a truncated exponential family with natural parameter θ and truncation parameter γ. Such a family, with γ fixed, constitutes a regular exponential family of order 1. Thus, for any fixed $\gamma\in\Gamma$, $\log b(\theta,\gamma)$ is strictly convex and infinitely differentiable as a function of $\theta \in \Theta$ and $\lambda_k(\theta, \gamma) \equiv \partial^k \log b(\theta, \gamma)/\partial\theta^k$, $k=1, 2, \cdots$, is the kth cumulant corresponding to (1.1).

Examples of such a model are given by:

Example 1.1. (i) Truncated $N(\mu, \sigma^2=1)$ distribution; $a(x)=\exp(-x^2/2)$, $u(x)=x$, $\theta=\mu$, $\Gamma=R$, $\Theta(\gamma)=R$, $\forall \gamma \in R$. (ii) Truncated $N(\mu=0, \sigma^2)$ distribution; $a(x)=1/(2\pi)^{1/2}$, $u(x)=-x^2/2$, $\theta=1/\sigma^2$, $\Gamma=R$, $\Theta(\gamma)=R^+$, $\forall \gamma \in R$.

Example 1.2. $a(x)=jx^{j-1}$, $u(x)=-x^j$, $j=1, 2, \cdots$, $\Gamma=R$ if j is odd and R^+, j is even, $\Theta(\gamma)=R^+$, $\forall \gamma \in \Gamma$.

The truncated exponential family plays an important role in life testing. It was first introduced by Hogg and Craig [5]. Some later references are, Fraser [4], Lwin [9], Huzurbazar [7], and Barndorff-Nielsen [2]. Other references, which deal with special cases of this family, can be found in Johnson and Kotz ([8], Ch. 17, Section 7.1 and Ch. 13, Section 7). The latter are mainly concerned with providing iterative methods for solving the maximum likelihood equation for θ.

This paper first compares the asymptotic behaviour of two estimation procedures for the parameter θ with γ considered as a nuisance parameter. The two estimates considered are the ordinary maximum likelihood estimate (MLE) and the maximum conditional likelihood estimate (MCLE), which is proposed as a reasonable competitor of the MLE. The asymptotic behaviour of these two estimates is then compared with that of the MLE for θ when γ is considered to be known. Asymptotic behaviour of maximum conditional likelihood estimates for various families of distributions is discussed in a number of references: e.g. Andersen [1], Huque and Katti [6] and Barndorff-Nielsen and Cox [3].

Several remarks concerning the problem at hand should be made at this stage.

1) Whereas there is a substantial literature treating the asymptotic behaviour of the MLE for γ (in connection with the asymptotic theory of extreme values), there does not seem to be any literature concerning the asymptotic behaviour of the MLE for θ in the presence of unknown truncation parameter.

2) In the two-parameter regular case (such as a two-parameter exponential family) the asymptotic normality of the MLE for one of these parameters can be obtained from the asymptotic joint distribution (i.e., bivariate normal) of the MLE's of both of them. In contrast to the regular case, such a procedure cannot directly be used in a truncated exponential family since the limiting joint distribution of the MLE's for θ and γ is not, in general, bivariate normal. Thus, alternative

tools are needed for deriving the asymptotic behaviour of the MLE for θ.

The main results of the paper are presented in Theorem 2.1. It is shown there that for samples of size $n \geqq 2$, the MLE and MCLE for θ exist with probability 1 and are given as the unique roots of the appropriate maximum likelihood equations. These two estimates are then shown to be strongly consistent for θ with limiting distributions coinciding with that of the MLE for θ when γ is known. Most derivations relying on standard techniques will be omitted for the sake of brevity.

2. Main results

Consider a random sample $X_n=(X_1,\cdots,X_n)$ of i.i.d. r.v.'s with common density (1.1), and let $X_{(1)} \leqq \cdots \leqq X_{(n)}$ be the corresponding order statistics of the sample X_n. Let $L_n(\theta, \gamma)$ denote the likelihood function of θ and γ based on the sample X_n, and $\hat{\theta}_n$ and $\hat{\gamma}_n$ the MLE's for θ and γ, respectively. Clearly, $\hat{\gamma}_n = x_{(1)}$ and $L_n(\hat{\theta}_n, x_{(1)}) = \sup_{\theta \in \Theta} L_n(\theta, x_{(1)})$.

The conditional model for θ is obtained by elimination of γ via conditioning on $x_{(1)}$. The conditional likelihood function of θ, denoted by $L_n^c(\theta, x_{(1)})$, is then proportional to the density of $(X_{(2)},\cdots,X_{(n)})$ conditional on $x_{(1)}$, that is,

$$(2.1) \qquad L_n^c(\theta, x_{(1)}) \propto (n-1)! \prod_{i=2}^{n} \{a(x_{(i)}) \exp[\theta u(x_{(i)})]/b(\theta, x_{(1)})\}, \qquad x_{(n)} \geqq \cdots \geqq x_{(1)}.$$

The expression on the right-hand side of (2.1) can be interpreted as the joint density of the order statistics of a random sample of size $(n-1)$ from a parent density,

$$(2.2) \qquad g(y:\theta, x_{(1)}) = a(y) \exp\{\theta u(y)\}/b(\theta, x_{(1)}), \qquad y > x_{(1)}.$$

From this viewpoint it is clear that there exists a random permutation, say $Y_2,\cdots,Y_n$, of the $(n-1)!$ permutations of $(X_{(2)},\cdots,X_{(n)})$ such that *conditional* on $x_{(1)}$, the $Y_2,\cdots,Y_n$ are independently and identically distributed r.v.'s with common density of the form (2.2) (c.f. Quesenberry [10]).

Since the form of $g(y:\theta, x_{(1)})$ is as in (1.1) with $x_{(1)}$ playing the role of γ, it follows that for all $\theta \in \Theta$ and almost all $x_{(1)}$, the kth cumulant $(k=1, 2, \cdots)$ and the characteristic function of the conditional distribution of $u(Y_i)$ $(i=2,\cdots,n)$ given $x_{(1)}$, exist and are given by $\lambda_k(\theta, x_{(1)})$ and $b(\theta+is, x_{(1)})/b(\theta, x_{(1)})$, respectively.

The conditional likelihood function of θ can now be written, without loss of generality, as $L_n^c(\theta, x_{(1)})=\prod_{i=2}^{n} g(y_i:\theta, x_{(1)})$, and the MCLE $\hat{\theta}_n^c$ is that value of θ for which L_n^c attains its supremum. Three different alternative approaches can be suggested to prove the asymptotic properties of the MCLE and MLE for θ. The first utilizes the asymptotic behaviour of $X_{(1)}$. The other two make use of the representation of L_n^c based on the random permutation $(Y_2,\cdots,Y_n)$ of $(X_{(2)},\cdots,X_{(n)})$. We shall adopt one of the latter approaches, which seems to be intrinsically interesting and may also lend itself to other problems.

The main results of this paper are contained in the following theorem, in which $l_n^c(\theta, x_{(1)})$ and $l_n(\theta, x_{(1)})$ are used for $(1/(n-1))\partial \log L_n^c(\theta, x_{(1)})/\partial\theta$ and $(1/n)\partial \log L_n(\theta, x_{(1)})/\partial\theta$, respectively. Furthermore, $\tilde{\theta}_n$ will denote the MLE for θ when γ is known.

THEOREM 2.1. *Under the assumptions imposed in Section* 1,

(i) *For samples of size* $n\geqq 2$, $\hat{\theta}_n^c$ *and* $\hat{\theta}_n$ *exist with probability* 1 *and are given as the unique roots of the maximum likelihood equations,* $l_n^c(\hat{\theta}_n^c, x_{(1)})=0$ *and* $l_n(\hat{\theta}_n, x_{(1)})=0$, *respectively.*

(ii) *Let* $\theta_0\in\Theta$ *be the true parameter value, then as* $n\to\infty$, $\hat{\theta}_n^c$ *and* $\hat{\theta}_n$ *converge almost surely to* θ_0.

(iii) *Each of the following three random sequences has a limiting* $N(0, 1/\lambda_2(\theta_0,\gamma))$ *distribution:* $n^{1/2}(\tilde{\theta}_n-\theta_0)$, $n^{1/2}(\hat{\theta}_n^c-\theta_0)$, *and* $n^{1/2}(\hat{\theta}_n-\theta_0)$.

PROOF.

(i) First consider the problem of the existence and uniqueness of $\tilde{\theta}_n$. For given γ, the exponential model (1.1) is regular with the Lebesgue measure as the underlying dominating measure. Thus, by Corollaries 9.4 and 9.6 in Barndorff-Nielsen [2], $\tilde{\theta}_n$ exists with the probability 1 and is given as the unique solution of the maximum likelihood equation, $\sum_{i=1}^{n} u(x_i)/n-\lambda_1(\tilde{\theta}_n,\gamma)=0$. Now note that exactly the same argument works for $\hat{\theta}_n$ and $\hat{\theta}_n^c$ with, in the first case γ replaced by $x_{(1)}$, and in the second case with $x_{(1)}$ replacing γ and $\sum_{i=2}^{n} u(y_i)/(n-1)$ replacing $\sum_{i=1}^{n} u(x_i)/n$.

(ii) Henceforth we shall use $X_{(1)}^n$ and Y_i^n instead of $X_{(1)}$ and Y_i, respectively, to indicate the sample size according to which these r.v.'s are defined. We first show that $\hat{\theta}_n^c \xrightarrow{\text{a.s.}} \theta_0$. Let $\delta>0$ and define, $R_n^c(\pm\delta)\equiv\{\log L_n^c(\theta_0\pm\delta, X_{(1)}^n)-\log L_n^c(\theta_0, X_{(1)}^n)\}/(n-1)=\pm\delta\sum_{i=2}^{n} u(Y_i^n)/(n-1)-\log b(\theta_0\pm\delta, X_{(1)}^n)+\log b(\theta_0, X_{(1)}^n)$. Since, $X_{(1)}^n\xrightarrow{\text{a.s.}}\gamma$ and $\sum_{i=2}^{n} u(Y_i^n)/(n-1)=(n/(n-1))\left\{\sum_{i=1}^{n} u(X_i)/n-u(X_{(1)}^n)/(n-1)\right\}\xrightarrow{\text{a.s.}}\lambda_1(\theta_0,\gamma)$, it follows that

(2.3) $$R_n^c(\pm\delta)\xrightarrow{\text{a.s.}}\pm\delta\lambda_1(\theta_0,\gamma)-\log b(\theta_0\pm\delta,\gamma)+\log b(\theta_0,\gamma)<0\ ,$$

where the inequality in (2.3) is obtained by using the strict convexity of $\log b(\theta,\gamma)$ as a function of $\theta\in\Theta$. The proof can now be completed by using arguments similar to those in Rao ([11], Theorem (i), p. 364). The method of proof of the strong consistency of $\hat{\theta}_n$ is analogous.
(iii) The limiting distribution of $n^{1/2}(\tilde{\theta}_n-\theta_0)$ is, clearly, $N(0, 1/\lambda_2(\theta_0,\gamma))$. We show that this is also the limiting distribution of the other two random sequences considered. Taylor expansions of $l_n^c(\hat{\theta}_n^c, X_{(1)}^n)$ and $l_n(\hat{\theta}_n, X_{(1)}^n)$ at $\theta=\theta_0$ yield, respectively,

(2.4) $$\begin{aligned} n^{1/2}&\left\{\sum_{i=2}^n u(Y_i^n)/(n-1)-\lambda_1(\theta_0, X_{(1)}^n)\right\}\\ &=n^{1/2}(\hat{\theta}_n^c-\theta_0)\lambda_2(\theta_0, X_{(1)}^n)[1+(\hat{\theta}_n^c-\theta_0)\lambda_3(\xi_n^c, X_{(1)}^n)/2\lambda_2(\theta_0, X_{(1)}^n)]\ ,\\ &\qquad |\xi_n^c-\theta_0|\leqq|\hat{\theta}_n^c-\theta_0|\ \text{a.s.}\ ,\end{aligned}$$

and

(2.5) $$\begin{aligned} n^{1/2}&\left\{\sum_{i=1}^n u(X_i)/n-\lambda_1(\theta_0, X_{(1)}^n)\right\}\\ &=n^{1/2}(\hat{\theta}_n-\theta_0)\lambda_2(\theta_0, X_{(1)}^n)[1+(\hat{\theta}_n-\theta_0)\lambda_3(\xi_n, X_{(1)}^n)/2\lambda_2(\theta_0, X_{(1)}^n)]\ ,\\ &\qquad |\xi_n-\theta_0|\leqq|\hat{\theta}_n-\theta_0|\ \text{a.s.}\ .\end{aligned}$$

Clearly, the terms in the square brackets on the right-hand sides of equations (2.4) and (2.5) converge almost surely to 1. Let C_n and M_n denote the expressions on the left-hand sides of (2.4) and (2.5), respectively. Then, $C_n-M_n=(1/n^{1/2})\left\{\sum_{i=2}^n u(Y_i^n)/(n-1)-u(X_{(1)}^n)\right\}\xrightarrow{\text{a.s.}}0$. The proof can now be completed by proving that $C_n\xrightarrow{D}N(0, \lambda_2(\theta_0,\gamma))$. For this purpose let $\beta_n(s, x_{(1)}^n)$ denote the characteristic function of C_n conditional on $x_{(1)}^n$. An expression for $\beta_n(s, x_{(1)}^n)$ can be derived by exploiting the fact that the r.v.'s $u(Y_i^n)$, $i=2,\cdots,n$, are conditionally independent given $x_{(1)}^n$. Then, by a Taylor expansion of $\log b(\theta_0+isn^{1/2}/(n-1), x_{(1)}^n)$ at $\theta=\theta_0$, we obtain

$$\begin{aligned}\beta_n(s, x_{(1)}^n)&=\exp\{-(n/(n-1))(s^2/2)\lambda_2(\theta_n^0, x_{(1)}^n)\}\ ,\\ &\qquad |\theta_n^0-\theta_0|\leqq|s|n^{1/2}/(n-1)\ \forall s\ .\end{aligned}$$

Now, since for all s, $\beta_n(s, X_{(1)}^n)\xrightarrow{\text{a.s.}}\exp\{-(s^2/2)\lambda_2(\theta_0,\gamma)\}$, $|\beta_n(s, X_{(1)}^n)|\leqq 1$ a.s. and $\mathrm{E}\{\exp(iC_n)\}=\mathrm{E}\{\beta_n(s,X_{(1)}^n)\}$, the required result follows from the dominated convergence theorem.

Further generalizations of the results of this paper to higher dimensional θ and γ can be made by considering an absolutely continuous parametric family of distributions possessing densities of the form (see Hogg and Craig [5]),

$$f(x:\ \theta_1,\cdots,\theta_k,\gamma_1,\gamma_2)=a(x)\exp\left\{\sum_{i=1}^{k}\theta_i u_i(x)\right\}\Big/ b(\theta_1,\cdots,\theta_k,\gamma_1,\gamma_2)\ ,\qquad \gamma_1<x<\gamma_2$$

for $-\infty\leqq c<\gamma_1<\gamma_2<d\leqq\infty$, where $(\theta_1,\cdots,\theta_k,\gamma_1,\gamma_2)$ are assumed to be unknown. It would seem that the derivations of the asymptotic properties of the MLE and MCLE for $(\theta_1,\cdots,\theta_k)$ could be completed in a manner analogous to that of the case of a density of the form (1.1).

Acknowledgement

I would like to thank the referee for helpful comments.

UNIVERSITY OF HAIFA, ISRAEL

REFERENCES

[1] Andersen, E. B. (1973). *Conditional Inference and Methods for Measuring*, Mentalhygienisk Forlag, Copenhagen.

[2] Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*, John Wiley.

[3] Barndorff-Nielsen, O. and Cox, D. R. (1979). Edgeworth and saddle-point approximation with statistical applications, *J. R. Statist. Soc.*, B, 41, 279-312.

[4] Fraser, D. A. S. (1963). On sufficiency and the exponential family, *J. R. Statist. Soc.*, B, **25**, 115-123.

[5] Hogg, R. V. and Craig, A. T. (1956). Sufficient statistics in elementary distribution theory, *Sankhyā*, **17**, 209-216.

[6] Huque, F. and Katti, S. K. (1976). A note on maximum conditional likelihood estimate, *Sankhyā*, B, 28, 1-13.

[7] Huzurbazar, V. S. (1976). *Sufficient Statistics*, Marcel Dekker, Inc. N.Y. and Basel.

[8] Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions*—1, Houghton Mifflin Company, Boston.

[9] Lwin, T. (1975). Exponential family distribution with a truncation parameter, *Biometrika*, **62**, 218-220.

[10] Quesenberry, C. P. (1975). Transforming samples from truncation parameter distributions to uniformity, *Communications in Statist.*, **4**, 1149-1155.

[11] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Second Edition, John Wiley, London.