Ann. Inst. Statist. Math.
Vol. 45, No. 3, 477–498 (1993)

ON LARGE DEVIATION EXPANSION OF DISTRIBUTION OF MAXIMUM LIKELIHOOD ESTIMATOR AND ITS APPLICATION IN LARGE SAMPLE ESTIMATION

J. C. FU[1]*, GANG LI[2] AND L. C. ZHAO[3]**

[1] *Department of Statistics, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2*
[2] *Department of Math. Sciences, SUNY at Binghamton, Binghamton, NY 13902, U.S.A.*
[3] *Department of Mathematics, Univ. of Science and Technology of China, Hefie, Anhui, China*

(Received April 10, 1991; revised September 29, 1992)

Abstract. For estimating an unknown parameter θ, the likelihood principle yields the maximum likelihood estimator. It is often favoured especially by the applied statistician, for its good properties in the large sample case. In this paper, a large deviation expansion for the distribution of the maximum likelihood estimator is obtained. The asymptotic expansion provides a useful tool to approximate the tail probability of the maximum likelihood estimator and to make statistical inference. Theoretical and numerical examples are given. Numerical results show that the large deviation approximation performs much better than the classical normal approximation.

Key words and phrases: Large deviation expansion, maximum likelihood estimator, exponential rate.

1. Introduction

Let $s = (x_1, \dots, x_n)$ be a sample of independently identically distributed (i.i.d.) observations with a common density function $f(x \mid \theta)$, where the parameter space Θ is an open interval of the real line. Let $l(x \mid \theta) = \log f(x \mid \theta)$, $l_n(s \mid \theta) = \sum_{i=1}^{n} l(x_i \mid \theta)$ be the log-likelihood function of the sample s, and $l_n^{(i)}(s \mid \theta) = (d^i/d\theta^i)l_n(s \mid \theta)$ for $i = 1, 2, 3, \dots$ be the corresponding derivatives.

The likelihood principle yields the maximum likelihood estimator (mle) $\hat{\theta}_n$. It is usually favoured by many statisticians for its large sample optimal properties. Under certain regularity conditions, the mle $\hat{\theta}_n$ is a consistent and asymptotically normally distributed estimator whose asymptotic variance achieves the Cramér-Rao lower bound: i.e., for any estimator $T_n(s)$, if

$$\sqrt{n}(T_n - \theta) \to N(0, v(\theta)) \quad \text{as} \quad n \to \infty, \tag{1.1}$$

* This work is supported in part by the Natural Science and Engineering Research Council of Canada under grant NSERC A-9216.

** This author is also partially supported by the National Science Foundation of China.

then

$$v(\theta) \geq 1/I(\theta), \tag{1.2}$$

where $I(\theta)$ is the Fisher information, and the equality holds when $T_n(s)$ is mle. Consequently, the mle is an asymptotically efficient estimator in the sense of (1.2). Its tail probability is traditionally approximated by its limiting distribution; i.e.,

$$\alpha_n(\hat{\theta}, \theta, \epsilon) = P_\theta(\hat{\theta}_n(s) - \theta \geq \epsilon) \sim 1 - \Phi(\sqrt{nI(\theta)}\epsilon), \quad \text{as} \quad n \to \infty, \tag{1.3}$$

where Φ stands for the distribution function (df) of a standard normal. The above normal approximation often performs poorly, especially for estimating the tail probability. Further, the normal approximation is not found to be satisfactory for statistical inference by many statisticians for various reasons, both in practice and in theory (Cramér (1937), Weiss and Wolfowitz (1966, 1967, 1970)). Recently, several different approximations have been developed by, for instance, Cramér (1937), Daniel (1954), Barndorff-Neilsen and Cox (1979), Field (1982), and Hougaard (1985).

Typically, for a good consistent estimator T_n, the rate of convergence is usually exponential and its tail probability has the following large deviation expansion: for $\epsilon > 0$,

$$\alpha_n(T, \theta, \epsilon) = e^{-n\beta(T,\theta,\epsilon)} \frac{1}{\sqrt{n}} \{a_0 + a_1 n^{-1} + \cdots + a_m n^{-m} + O(n^{-m-1})\}, \tag{1.4}$$

as $n \to \infty$. The positive constant $\beta(T, \theta, \epsilon)$ is called the exponential rate for the consistent estimator T_n. The β is an indicator of the performance of the estimator, the larger the exponential rate, the better the estimator. It has been studied by many researchers, for example, Bahadur (1971), Chernoff (1952), Fu (1973, 1975), Kester and Kallenberg (1986), Rubin and Rukhin (1983), and Sievers (1978). For any consistent estimator T_n and $\epsilon > 0$, the exponential rate obeys the following inequality

$$\beta(T, \theta, \epsilon) \leq B(\theta, \epsilon), \tag{1.5}$$

where

$$B(\theta, \epsilon) = \inf_{\theta'} \{K(\theta', \theta) : |\theta' - \theta| > \epsilon\} \tag{1.6}$$

and the Kullback-Leibler information of $f(x \mid \theta')$ with respect to $f(x \mid \theta)$ is defined as

$$K(\theta', \theta) = \int_{-\infty}^{\infty} \left(\log \frac{f(x \mid \theta')}{f(x \mid \theta)}\right) f(x \mid \theta') dx. \tag{1.7}$$

The exponential rate $\beta(\hat{\theta}, \theta, \epsilon)$ of the mle $\hat{\theta}$ achieves the lower bound (1.5) if and only if the underlying distribution belongs to the exponential family (Kester and Kallenberg (1986), Cheng and Fu (1986)). Kester and Kallenberg (1986) also showed that if the parameter space is not convex, which occurs, for instance, in curved exponential families, then in general, no estimator achieves the lower bound (1.5).

The objectives of this manuscript are two-folds, (i) to obtain a large derivation expansion similar to (1.4) for the mle, which provides a good approximation of its tail probability, and (ii) to show its applications in large sample statistical inference. In order to illustrate our results, several examples and numerical comparisons are given in Section 3. The numerical results show that the large deviation approximation performs much better than the usual normal approximation. The detailed mathematical proof for the large deviation expansion of the mle is given in Section 4.

2. Large deviation expansion of mle

In order to get the large deviation expansion (1.4) for the mle, we first obtain a large deviation expansion for the sum of i.i.d. random variables. Let $\{X_i\}$ be a sequence of i.i.d. random variables having distribution function $F(x)$ defined on the real line and characteristic function $\psi(t)$. We assume that the following conditions hold:

Condition A. F satisfies Cramér's Condition (C) (Cramér (1937), p. 81), i.e.

$$\limsup_{|t|\to\infty} |\psi(t)| < 1. \tag{2.1}$$

Condition B. Let $\phi(t) = Ee^{tX_1}$ be the moment generating function (m.g.f.) of X_1. Assume that

$$\sup\{0 \le t < \infty : \phi(t) \le \infty\} = t_0 > 0. \tag{2.2}$$

Note that t_0 can be finite or infinite.

Assume that $EX_1 < \epsilon < \lim_{t\to t_0^-} \phi'(t)/\phi(t)$. Let $h(z) = \epsilon z - \log\phi(z)$. From the Cramér-Hoeffding theorem (Petrov (1975), p. 234) there exists τ, $0 < \tau < t_0$, such that (i) $h'(\tau) = 0$, and (ii) $h(\tau) > 0$ and $h''(\tau) < 0$. The point τ is called saddle point. Denote $\sigma = \sqrt{-h''(\tau)}$.

Theorem 2.1. *Under Conditions* A *and* B *we have*

$$P\left(\sum_{i=1}^{n} X_i \ge n\epsilon\right) = e^{-nh(\tau)}\frac{1}{\sqrt{n}}(a_0 + a_1 n^{-1} + \cdots + a_m n^{-m} + O(n^{-m-1})), \tag{2.3}$$

where the coefficients a_i, $i = 0, 1, \ldots, m$ *are independent of* n; *in particular*

$$\begin{aligned} a_0 &= \frac{1}{\sqrt{2\pi}\sigma\tau} \quad \textit{and} \\ a_1 &= \frac{1}{\sqrt{2\pi}\sigma\tau}\left[-\frac{1}{\sigma^2\tau^2} + \frac{h^{(3)}(\tau)}{2\tau\sigma^4} - \frac{h^{(4)}(\tau)}{8\sigma^4} - \frac{5(h^{(3)}(\tau))^2}{24\sigma^6}\right]. \end{aligned} \tag{2.4}$$

Let $\{X_n\}$ be a sequence of i.i.d. nondegenerate random variables defined on $\{ra : r = 0, \pm 1, \ldots\}$ which satisfies the following conditions:

CONDITION A$'$. Let $\{k_n\}$ be a sequence of integers satisfying

$$\frac{k_n}{n}a \to \epsilon, \qquad \text{as } n \to \infty \text{ for some } \epsilon \text{ such that } EX_1 < \epsilon < \lim_{t \to t_0^-} \phi'(t)/\phi(t). \tag{2.5}$$

Let $h(\epsilon, z) = \epsilon z - \log \phi(z)$, τ_0 be a solution of $h'(\tau_0) = 0$, and $\{\tau_n\}$ be a sequence of positive numbers such that

$$\frac{\partial}{\partial z} h\left(\frac{k_n}{n}a, z\right) |_{z=\tau_n} = 0, \tag{2.6}$$

$$h\left(\frac{k_n}{n}a, \tau_n\right) > 0 \quad \text{and} \quad \frac{\partial^2}{\partial z^2} h\left(\frac{k_n}{n}a, z\right) |_{z=\tau_n} < 0. \tag{2.7}$$

It follows from (2.6) and (2.7) that $\tau_0 < t_0$ and

$$\tau_n \to \tau_0 > 0, \qquad \text{as} \quad n \to \infty. \tag{2.8}$$

Denote

$$\sigma_n = \sqrt{-\frac{\partial^2}{\partial z^2} h\left(\frac{k_n}{n}a, z\right) |_{z=\tau_n}} \, . \tag{2.9}$$

THEOREM 2.2. *Under Conditions* A$'$ *and* B, *we have*

$$P\left(\sum_{i=1}^n X_i \geq k_n a\right) = e^{-nh((k_n/n)a, \tau_n)} \frac{1}{\sqrt{n}} [a_0 + a_1 n^{-1} + \cdots + a_m n^{-m} + O(n^{-m-1})], \tag{2.10}$$

where a_j, $j = 0, 1, \ldots, m$ *are real numbers; in particular*

$$a_0 = \frac{a}{\sqrt{2\pi}\sigma_n (1 - e^{-\tau_n a})}. \tag{2.11}$$

If $k_n/n = k$, *then* a_j, $j = 0, \ldots, m$ *are independent of* n.

The above two theorems are modified versions of results due to Cramér (1937) and Bahadur and Rao (1960). Our proofs are based on a modification of Laplace's method (Oliver (1974)) applied directly to the inverse formulas, which are *new and quite different* from the original proofs given by Cramér (1937) and Bahadur and Rao (1960). Theoretically speaking, these results are basically equivalent. The major advantage of our expansion is that the first two coefficients in our expansion are given *explicitly* for easier use in statistical applications. On the other hand, the coefficients of third order and above in the Bahadur and Rao method are given by recurrence formulae but not in our method. The details of our proofs are given in Section 4.

Let $\hat{\theta}_n(s)$ be the mle, and $\underline{\theta}_n(s) = \inf\{\theta : l_n^{(1)}(s \mid \theta) = 0\}$ and $\bar{\theta}_n(s) = \sup\{\theta : l_n^{(1)}(s \mid \theta) = 0\}$ be the smallest and the largest roots of the likelihood equation $l_n^{(1)}(s \mid \theta) = 0$, respectively. Since

$$\underline{\theta}_n(s) \leq \hat{\theta}_n(s) \leq \bar{\theta}_n(s), \qquad \text{for every } s \text{ and } n, \tag{2.12}$$

the following two inequalities hold:

$$P_\theta(\underline{\theta}_n(s) \geq \theta + \epsilon) \leq P_\theta(l_n^{(1)}(s \mid \theta + \epsilon) \geq 0) \leq P_\theta(\bar{\theta}_n(s) \geq \theta + \epsilon) \tag{2.13}$$

and

$$P_\theta(\bar{\theta}_n(s) \leq \theta - \epsilon) \leq P_\theta(l_n^{(1)}(s \mid \theta - \epsilon) \leq 0) \leq P_\theta(\underline{\theta}_n(s) \leq \theta - \epsilon). \tag{2.14}$$

To obtain our main result (1.4), we need the following additional conditions:

CONDITION C. For given n and every s, the maximum likelihood estimator $\hat{\theta}_n(s)$ is the unique solution of the likelihood equation $l_n^{(1)}(s \mid \theta) = 0$.

CONDITION D. For each $\theta \in \Theta$ and $\epsilon > 0$, the moment generating function

$$\phi(t, \theta, \epsilon) = E_\theta[\exp\{tl^{(1)}(x \mid \theta + \epsilon)\}] < \infty \tag{2.15}$$

exists for $0 < |t| < t_0$, and the characteristic function $\psi(t, \theta, \epsilon)$ of the random variable $l^{(1)}(x \mid \theta + \epsilon)$ satisfies Condition A.

Let, for $\epsilon > 0$, $h_\epsilon(z) = -\log \phi(z, \theta, \epsilon)$. By Cramér-Hoeffding theorem, there exists a saddle point τ_ϵ, $0 < \tau_\epsilon < \tau_0$, such that

$$h'_\epsilon(\tau_\epsilon) = 0 \tag{2.16}$$

and

$$h_\epsilon(\tau_\epsilon) > 0 \qquad \text{and} \qquad h''_\epsilon(\tau_\epsilon) < 0. \tag{2.17}$$

In the following, we assume that the τ_ϵ exists, and denote

$$\sigma_\epsilon = \sqrt{-h''_\epsilon(\tau_\epsilon)}. \tag{2.18}$$

THEOREM 2.3. *For any $\epsilon > 0$, under Conditions* C *and* D, *the ϵ-tail probability of the maximum likelihood estimator $\hat{\theta}_n$ has an asymptotic expansion given by*

$$\alpha_n(\hat{\theta}, \theta, \epsilon) = e^{-n\beta(\hat{\theta},\theta,\epsilon)} \frac{1}{\sqrt{n}}\{b_0 + b_1 n^{-1} + \cdots + b_m n^{-m} + O(n^{-m-1})\}, \tag{2.19}$$

where

$$\beta(\hat{\theta}, \theta, \epsilon) = -\log \phi(\tau_\epsilon, \theta, \epsilon) \tag{2.20}$$

is the exponential rate of the maximum likelihood estimator $\hat{\theta}_n$ and the coefficients b_i, $i = 0, 1, \ldots, m$ are independent of n; in particular

$$b_0 = \frac{1}{\sqrt{2\pi}\sigma_\epsilon \tau_\epsilon} \tag{2.21}$$

and

$$b_1 = \frac{1}{\sqrt{2\pi}\sigma_\epsilon \tau_\epsilon}\left[-\frac{1}{\sigma_\epsilon^2\tau_\epsilon^2} + \frac{h_\epsilon^{(3)}(\tau_\epsilon)}{2\tau_\epsilon\sigma_\epsilon^4} - \frac{h_\epsilon^{(4)}(\tau_\epsilon)}{8\sigma_\epsilon^4} - \frac{5(h^{(3)}(\tau_\epsilon))^2}{24\sigma_\epsilon^6}\right]. \tag{2.22}$$

PROOF. It follows from Condition C and the inequality (2.13) and (2.14) that we have $\underline{\theta}_n(s) = \hat{\theta}_n(s) = \bar{\theta}_n(s)$ and $P_\theta(\hat{\theta}_n \geq \theta + \epsilon) = P_\theta(l_n^{(1)}(s \mid \theta + \epsilon) \geq 0)$. For given θ and $\epsilon > 0$, $l^{(1)}(X_i \mid \theta + \epsilon)$, $i = 1, 2, \ldots, n$ are i.i.d. random variables. Hence, the result follows immediately from Theorem 2.1. □

Similarly, the large deviation expansion for the left hand ϵ-tail probability can be obtained by replacing all the ϵ in the Theorem 2.3 with the value $-\epsilon$. For small ϵ, the saddle point τ_ϵ defined by equation (2.16) has a Taylor expansion (Fu (1982)) given by

$$\tau_\epsilon = \epsilon + A\epsilon^2 + B\epsilon^3 + o(\epsilon^3), \tag{2.23}$$

where

$$\begin{aligned} A &= -El^{(1)}l^{(2)}/2I, \\ B &= -\frac{1}{I}\left[\frac{1}{3!}El^{(4)} - \frac{1}{I}(El^{(1)}l^{(2)})^2 + El^{(1)}l^{(3)} + \frac{3}{2}E(l^{(1)})^2 l^{(2)}\right. \\ &\qquad\left. + \frac{1}{2I}El^{(1)}l^{(2)}E(l^{(1)})^3 + \frac{1}{3!}E(l^{(1)})^4\right], \end{aligned}$$

and E and $l^{(i)}$ stand for E_θ and $l^{(i)}(X \mid \theta)$ respectively.

Usually, the saddle point τ_ϵ cannot be obtained explicitly from the equation $h_\epsilon'(t) = 0$. We suggest to replace τ_ϵ in the Theorem 2.3 with $\epsilon + A\epsilon^2 + B\epsilon^3$. For n large and ϵ small, (2.19) yields the well-known result

$$\alpha_n(\hat{\theta}, \theta, \epsilon) \sim \frac{1}{\sqrt{2\pi n I}\epsilon} e^{-(nI/2)\epsilon^2}. \tag{2.24}$$

3. Large sample point estimation

The large deviation expansion (2.19) of the mle $\hat{\theta}$ contains two parts: (a) the exponential rate $\beta(\hat{\theta}, \theta, \epsilon)$ and (b) the non-exponential term $(1/\sqrt{n})\{b_0 + b_1 n^{-1} + \cdots + b_m n^{-m} + O(n^{-m-1})\}$. The predominate exponential term $\exp\{-n\beta(\hat{\theta}, \theta, \epsilon)\}$ is directly associated with the asymptotic exact distribution of the mle which plays an important role in large sample point estimation. For approximating the exact tail probability of the mle, both exponential and non-exponential terms are vital. Four examples, involving the normal, Poisson, exponential and binomial distributions which satisfy the regularity conditions, are given to illustrate our results. They are of interest in both statistical theory and practice.

Traditionally, the distribution of the mle is approximated by a normal distribution (central limiting theorem). To our knowledge, at present, there has been no direct theoretical comparison between the large deviation approximation and the normal approximation. Numerical comparisons are very much needed; therefore, we provided numerical comparisons between the normal approximation and the large deviation approximation in two of our examples (exponential and binomial distributions). Since most statistical inferences and decisions such as 95% (or 99%) confidence intervals and 5% (or 1%) critical regions are often involved in computing the tail probabilities, our numerical comparisons will concentrate around tail probabilities of .025 or less.

Example 1. Let $\{X_i\}$ be a sequence of i.i.d. normal random variables with unknown mean θ and known variance σ^2. The mle for θ is the sample mean $\bar{X}_n$. Taking $m = 1$, it follows from Theorem 2.3 that the ϵ-tail probability of the mle $\hat{\theta}_n$ has an asymptotic expansion given by

$$\alpha_n(\hat{\theta}, \theta, \epsilon) = \frac{\sigma}{\sqrt{2\pi n}\epsilon} e^{-n\epsilon^2/2\sigma^2} \left[1 - \frac{\sigma^2}{n\epsilon^2} + O\left(\frac{1}{n^2}\right)\right]. \tag{3.1}$$

Example 2. Let $\{X_i\}$ be a sequence of i.i.d. Poisson random variables with mean λ. The maximum likelihood estimator for λ is $\hat{\lambda}_n = \bar{X}_n$ the sample mean. By Theorem 2.2, for an integer sequence $\{k_n\}$ satisfying $k_n/n \to \epsilon > \lambda$, we have

$$P_\lambda\left(\bar{X}_n \geq \frac{k_n}{n\lambda}\right) = e^{-(k_n \log(k_n/n\lambda) - k_n + n\lambda)} \frac{\sqrt{k_n}}{\sqrt{2\pi}(k_n - n\lambda)} \left(1 + O\left(\frac{1}{n}\right)\right).$$

Example 3. Let $\{X_i\}$ be a sequence of i.i.d. random variables with common density function

$$f(x \mid \lambda) = \lambda e^{-\lambda x}, \qquad \lambda, x > 0. \tag{3.2}$$

The sample mean $\bar{X}_n$ is the mle for $1/\lambda$. Since the sum S_n has a Gamma distribution with parameters n and λ, it follows that the exact ϵ-tail probability of $\bar{X}_n$ is

$$P\left(\bar{X}_n - \frac{1}{\lambda} > \epsilon\right) = \sum_{j=0}^{n} (n\epsilon\lambda + n)^j e^{-(n\epsilon\lambda + n)}/j!. \tag{3.3}$$

The normal approximation for the ϵ-tail probability of $\bar{X}_n$ is

$$\begin{aligned} P\left(\bar{X}_n - \frac{1}{\lambda} > \epsilon\right) &= P\left\{(\sqrt{n})^{-1}\lambda\left(\sum_{i=1}^{n} X_i - n/\lambda\right) \geq \sqrt{n}\lambda\epsilon\right\} \\ &\sim 1 - \Phi(\sqrt{n}\lambda\epsilon). \end{aligned} \tag{3.4}$$

Applying Theorem 2.3, we obtain a two-term large deviation approximation for the ϵ-tail probability of $\bar{X}_n$

$$P\left(\bar{X}_n - \frac{1}{\lambda} > \epsilon\right) \sim e^{-nh(\tau)}(\sqrt{n})^{-1}[a_0 + n^{-1}a_1], \tag{3.5}$$

where

$$h(\tau) = \lambda\epsilon - \log(1+\lambda\epsilon), \qquad a_0 = (\sqrt{2\pi}\lambda\epsilon)^{-1},$$
$$a_1 = -(\sqrt{2\pi}\lambda\epsilon)^{-1}[(\lambda\epsilon)^{-2} + (\lambda\epsilon)^{-1} + (12)^{-1}].$$

The numerical comparisons among the exact ϵ-tail probability P_E given by (3.3), the normal approximation P_N given by (3.4), and the large deviation approximation P_L given by (3.5) are provided by the following table for various values of λ and ϵ.

The numerical results in Table 1 show that the large deviation approximation performs well against the normal approximation in almost all cases, especially for the extreme tail probabilities.

Table 1. The exact and approximate ϵ-tail probabilities of mle $\bar{X}_n$ with $n = 11$ and 51.

n	λ	ϵ	P_E	P_N	P_L
11	0.5	1.320	.0267	.0143	.0218
	0.5	2.100	.0026	.0002	.0024
	1.0	0.660	.0267	.0143	.0218
	1.0	0.800	.0120	.0040	.0107
	2.0	0.325	.0282	.0155	.0228
	2.0	0.400	.0120	.0040	.0107
	3.0	0.230	.0226	.0111	.0190
	3.0	0.320	.0046	.0007	.0043
	4.0	0.170	.0239	.0121	.0199
	4.0	0.230	.0058	.0011	.0054
51	0.5	0.575	.0268	.0200	.0228
	0.5	0.650	.0156	.0101	.0140
	1.0	0.300	.0225	.0161	.0195
	1.0	0.420	.0034	.0014	.0033
	2.0	0.150	.0225	.0161	.0195
	2.0	0.180	.0091	.0051	.0084
	3.0	0.096	.0266	.0199	.0227
	3.0	0.150	.0020	.0007	.0020
	4.0	0.072	.0266	.0199	.0227
	4.0	0.084	.0132	.0082	.0120

Example 4. Let $\{X_i\}$ be a sequence of i.i.d. Bernoulli random variables with

$$P(X = x) = p^x(1-p)^{1-x}, \qquad x = 0, 1.$$

The maximum likelihood estimator for the parameter p is $\hat{p} = \bar{X}_n$. In this example we give only the numerical comparison.

The numerical results in Table 2 also show that the large deviation approximation is far superior to the classical normal approximation. Even at the sample size $n = 11$, the large deviation approximation performs reasonably well while the normal approximation performs poorly. Furthermore, an important phenomenon (which can be seen especially from Table 2) is that the normal approximation becomes worse as the underlying distribution becomes more skewed. The accuracy of the large deviation approximation is much less sensitive to the skewness of underlying distribution. This is because that the normal approximation involves only the variance of the underlying distribution where as the large deviation approximation involves not only the variance but also the skewness of the underlying distribution. For instance, the dominant term of the expansion, the exponential rate, has a strong connection with the skewness (Fu (1982), p. 764). Based on our limited numerical experience, a two-terms large deviation expansion of (1.4) is sufficiently accurate for statistical applications.

Table 2. The exact and approximate ϵ-tail probabilities for binomial random variables.

n	p	ϵ	P_E	P_N	P_L
10	0.5	0.300	0.0107	0.0057	0.0125
10	0.7	0.250	0.0282	0.0192	0.0308
25	0.6	0.175	0.0294	0.0207	0.0269
25	0.8	0.150	0.0274	0.0228	0.0270
50	0.7	0.120	0.0183	0.0154	0.0178
50	0.9	0.080	0.0053	0.0091	0.0057
100	0.6	0.100	0.0148	0.0123	0.0138
100	0.8	0.100	0.0023	0.0030	0.0023
200	0.6	0.070	0.0173	0.0152	0.0160
200	0.8	0.050	0.0283	0.0262	0.0267

All the conditions, except Condition C, are very mild and easy to verify. Condition C, which states that the mle $\hat{\theta}_n(s)$ is an unique root of the likelihood equation $l_n^{(1)}(s \mid \theta) = 0$, is somewhat strong, but is vital to the proof of our results. This condition is required by almost all other types of expansions (see, for instance, Field ((1982), p. 673) and Hougaard ((1985), p. 162)). The following two remarks pertain to those cases where the condition of uniqueness fails.

Remark 1. For the Cauchy distribution with location parameter, the log-likelihood function is very smooth and all derivatives exist. The likelihood equation

$$l_n^{(1)}(s \mid \theta) = -\sum_{i=1}^{n} \frac{2(\theta - x_i)}{1 + (x_i - \theta)^2} = 0 \tag{3.6}$$

is a polynomial of degree $2n-1$. It has $(2n-1)$ roots (real and complex), and the number of roots increases with the sample size. Hence it does *not satisfy* our conditions. Reeds (1985) shows that all the real roots but one (the global maximum: mle) of the likelihood equations $l_n^{(1)}(s \mid \theta) = 0$ tend to $+\infty$ or $-\infty$ almost surely as $n \to \infty$. Bai and Fu (1986) have proved that the mle $\hat{\theta}_n(s)$ still converges to θ exponentially; i.e., for ϵ small,

$$P_\theta(\hat{\theta}_n \geq \theta + \epsilon) = e^{-n\{\epsilon^2[1+O(\sqrt{\epsilon})]/4\}} \tag{3.7}$$

as $n \to \infty$. The same results also have been stated by Kester and Kallenberg ((1986), Remark 3.2). Although the likelihood equation has multiple roots we do believe that the asymptotic expansion exists in this case. Mathematically, we have not been able to obtain its asymptotic expansion.

Remark 2. If there is more than one mle, they all usually converge to the unknown parameter θ (i.e., they are all consistent estimators). It is traditionally believed that the statistician can use any one of them. Contrary to the above notion, they very often have different exponential rates, and hence, different asymptotic expansions. For example, consider the uniform distribution $U[\theta - 1/2, \theta + 1/2]$ with location parameter θ. Any point between the points $X_{(n)} - 1/2$ and $X_{(1)} + 1/2$ is a maximum likelihood estimator. They are all consistent, have different rates of convergence to θ. For example, for $p \in [0,1]$, the mle $\hat{\theta}_p = p(X_{(1)} + 1/2) + (1-p)(X_{(n)} - 1/2)$ has an exponential rate given by

$$\beta(\hat{\theta}_p, \theta, \epsilon) = -\lim_{n\to\infty} \frac{1}{n} \log P(|\hat{\theta}_p - \theta| \geq \epsilon) = -\log(1 - a_p\epsilon),$$

where $a_p = 1/\max(p, 1-p)$. Consequently, the mle's $X_{(1)} + 1/2$ and $X_{(n)} - 1/2$ have same exponential rate $-\log(1-\epsilon)$, the mle $(X_{(n)}/8 + 7X_{(1)}/8 - 3/8)$ has an exponential rate $-\log(1 - 8\epsilon/7)$, and the mle $(X_{(n)} + X_{(1)})/2$ has an exponential rate, $-\log(1-2\epsilon)$, *which is the fastest among all the maximum likelihood estimators*. Clearly, one should only use the optimal mle $(X_{(1)} + X_{(n)})/2$. Note that neither the mle $X_{(n)} - 1/2$ nor the mle $X_{(1)} - 1/2$ is asymptotically normally distributed. The asymptotic normality criterion of selecting an optimal consistent estimator collapses completely, at least in this simple example. On the other hand, the large deviation approach is applicable and provides a good solution to this problem.

Remark 3. Even in the case where there is a unique mle, Kraft and LeCam (1956) gave a very disturbing example that the mle satisfies the likelihood equation, but it is not consistent. Note that the large deviation approach remains applicable and has an exponential rate of zero.

Under smooth conditions, it is well known that the mle has all the classical optimal properties. For example, it is a first and second-order efficient estimator (see Efron (1975)). However, in a large deviation context, the picture is rather different. If the underlying distribution belongs to an exponential family of distributions, then its exponential rate $\beta(\hat{\theta}, \theta, \epsilon)$ achieves the Bahadur bound (1.5)

(Kester and Kallenberg (1986), Cheng and Fu (1986)). Outside the exponential family of distributions, the exponential rate $\beta(\hat{\theta}, \theta, \epsilon)$ given by mle is not always optimal. Let us give the following example.

Example 5. Let $\{X_i\}_{i=1}^n$ be a sequence of i.i.d. logistic random variables having common density function

$$f(x \mid \theta) = e^{-x+\theta}/(1+e^{-x+\theta})^2, \qquad x \in (-\infty, \infty).$$

For given ϵ, let $\tilde{\theta}_n$ be the likelihood ratio estimator for θ (which is the best translation invariant estimator), the unique solution of the following equation:

$$\lambda_n(s, \theta, \epsilon) = 2\epsilon + \frac{2}{n} \sum_{i=1}^{n} \log[(1+e^{-x_i+\theta-\epsilon})/(1+e^{x_i+\theta+\epsilon})] = 0.$$

It has an exponential rate given by

$$\beta(\tilde{\theta}, \theta, \epsilon) = \epsilon + \{\log[1 - \exp(-2\epsilon)]/2\epsilon\}$$

which reaches the Chernoff bound (see Fu (1985) and Kester and Kallenberg (1986)).

The mle $\hat{\theta}_n$ is the solution of the following equation:

$$\sum_{i=1}^{n} [1 + \exp(x_i - \theta)]^{-1} - \frac{n}{2} = 0.$$

It has an exponential rate given by

$$\beta(\hat{\theta}, \theta, \epsilon) = -\log \left\{ \inf_{t>0} e^{-t/2} \int_{-\infty}^{\infty} [\exp(t/(1+\exp(x-\epsilon)))] f(x) dx \right\}.$$

As ϵ tending to zero, we have

$$\lim_{\epsilon \to 0} \epsilon^{-i} \{\beta(\tilde{\theta}, \theta, \epsilon) - \beta(\hat{\theta}, \theta, \epsilon)\} = \begin{cases} 0, & i = 1, 2, 3, 4, 5 \\ 0.000079, & i = 6. \end{cases}$$

The exponential rates differ only at the coefficient of ϵ^6. This shows that for ϵ small, the likelihood ratio estimator $\tilde{\theta}_n$ is superior to mle $\hat{\theta}_n$. Our results substantiates that, in general the mle $\hat{\theta}$ is second order efficient estimator. Both Efron and Rao (Efron (1975)) seemed to suggest that if a theory of third-order efficiency were to be developed, mle would still emerge as the "optimal" estimator. The above counterexample shows this claim to be false.

Remark 4. Note that the above mentioned likelihood ratio estimator $\tilde{\theta}_n$ is an M-estimator. If the underlying density and the ϕ function are sufficiently smooth, then the large deviation expansion of M-estimator derived from the ϕ function can perhaps be obtained by a simple modification of the method developed in Section 2.

Remark 5. We would like to point out an important fact that the large deviation approximation (2.20) for the tail probability of maximum likelihood estimator has a pole at $\tau_\epsilon = 0$ ($\epsilon = 0$). Hence, if the large deviation approximation (2.20) is misused, for instance fixed n and as $\epsilon \to 0$, then the approximation could be inaccurate. There are several approaches, for example, Lugannani and Rice (1980) and Daniels (1987) to overcome this problem. The conditions in their paper (see Lugannani and Rice (1980) Condition (ii) on p. 481) are much stronger than that used in our paper. Their approximation performs very well numerically in the region as ϵ nears the zero and uniformly up to a factor $O(1/\sqrt{n})$. The approximation becomes less accurate in the region when the deviation of S_n is of order $O(n)$. This region is one of the most important region especially for the statistical inferences. For example, the statistical inferences such as confidence interval and critical region of testing hypothesis, based on mle's are always involved in computing the tail probabilities of S_n at the order $O(n)$.

In view of all the above examples, numerical results, and remarks it suggests that the large deviation theoretical approach is more applicable than the central limit theory as a method of selecting the optimal estimator. It is also superior for approximating the tail probabilities than the normal approximation.

4. Proofs

To prove Theorem 2.1 and Theorem 2.2, we need the following lemmas.

LEMMA 4.1. *If F_1 satisfies Condition* A, *and if F_1 is absolutely continuous with respect to F_2, then F_2 also satisfies Condition* A.

PROOF. See Lemma 4 of Bahadur and Rao (1960). □

LEMMA 4.2. *Suppose that G is a distribution function and its m.g.f. exists*

$$\phi(t) = \int_{-\infty}^{\infty} e^{tu} dG(u) < \infty, \qquad \text{for} \qquad 0 < |t| < t_0.$$

If $\epsilon > 0$ is a continuity point of G, then for any $0 < a < t_0$

$$P(Y \geq \epsilon) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} \frac{1}{z} e^{-\epsilon z} \phi(z) dz. \tag{4.1}$$

PROOF. Denote the right hand side of (4.1) by $\beta(\epsilon)$. We have

$$\begin{aligned} \beta(\epsilon) &= \lim_{N\to\infty} \frac{1}{2\pi i} \int_{a-iN}^{a+iN} \frac{1}{z} e^{-\epsilon z} \int_{-\infty}^{\infty} e^{zu} dG(u) dz \\ &= \lim_{N\to\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi i} \int_{a-iN}^{a+iN} \frac{1}{z} e^{(u-\epsilon)z} dz dG(u). \end{aligned} \tag{4.2}$$

For $u > \epsilon$, let

$$J_N(u) = 1 - \frac{1}{2\pi i}\int_{\pi/2}^{3\pi/2} \frac{1}{a + Ne^{i\xi}} e^{(u-\epsilon)a} e^{(u-\epsilon)Ne^{i\xi}} Nie^{i\xi} d\xi.$$

For $u < \epsilon$, let

$$J_N(u) = \frac{1}{2\pi i}\int_{-\pi/2}^{\pi/2} \frac{1}{a + Ne^{i\xi}} e^{(u-\epsilon)a} e^{-(\epsilon-u)Ne^{i\xi}} Nie^{i\xi} d\xi.$$

From the residue theorem, we have

$$J_N(u) = \frac{1}{2\pi i}\int_{a-iN}^{a+iN} \frac{1}{z} e^{(u-\epsilon)z} dz, \qquad u \neq \epsilon. \tag{4.3}$$

Now, when $u < \epsilon$, note that $|a + Ne^{i\xi}| > |Ne^{i\xi}|$ for $\xi \in [-\pi/2, \pi/2]$, we have

$$|J_N(u)| \le e^{(u-\epsilon)a} \frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} e^{-(\epsilon-u)N\cos\xi} d\xi = e^{(u-\epsilon)a}\frac{1}{\pi}\int_0^{\pi/2} e^{-(\epsilon-u)N\sin\xi} d\xi.$$

Further, since $\sin\xi \ge (2/\pi)\xi$ ($\xi \in [0, \pi/2]$), it follows

$$\begin{aligned} |J_N(u)| &\le \frac{1}{\pi} e^{(u-\epsilon)a}\int_0^{\pi/2} e^{-N(\epsilon-u)2\xi/\pi} d\xi \\ &= \frac{1}{2N(\epsilon-u)} e^{(u-\epsilon)a}(1 - e^{-N(\epsilon-u)}) \to 0, \qquad \text{as} \quad N \to \infty. \end{aligned}$$

Similarly, for $u > \epsilon$, we have

$$|J_N(u) - 1| \le \frac{1}{\pi} e^{(u-\epsilon)a}\int_0^{\pi/2} e^{-N(\epsilon-u)2\xi/\pi} d\xi \qquad \text{and} \qquad \lim_{N\to\infty} J_N(u) = 1.$$

Then we obtain that

$$\lim_{N\to\infty} J_N(u) = \begin{cases} 1, & \text{if } u > \epsilon \\ 0, & \text{if } u < \epsilon \end{cases} \tag{4.4}$$

and

$$|J_N(u)| < 1 + \frac{1}{2}e^{(u-\epsilon)a} \qquad \text{with} \qquad \int_{-\infty}^{\infty}\left(1 + \frac{1}{2}e^{(u-\epsilon)a}\right) dG(u) < \infty. \tag{4.5}$$

It follows from (4.2), (4.3), (4.4) and (4.5), and the dominated convergence theorem that

$$\beta(\epsilon) = \lim_{N\to\infty}\int_{-\infty}^{\infty} J_N(u)dG(u) = \int_{-\infty}^{\infty} \lim_{N\to\infty} J_N(u)dG(u) = P(Y \ge \epsilon). \qquad \square$$

LEMMA 4.3. *For $m = 1, 2, \dots$ and any complex number z*

$$\left| e^z - 1 - \frac{z}{1!} - \cdots - \frac{z^{m-1}}{(m-1)!} \right| \le \frac{|z|^m}{m!} e^{|z|}.$$

PROOF. Let $\xi_m(z) = e^z - 1 - \cdots - z^{m-1}/(m-1)!$. Then

$$\xi_1(z) = z \int_0^1 e^{zt} dt \quad \text{and} \quad \xi_m(z) = z \int_0^1 \xi_{m-1}(zt) dt \quad \text{for} \quad m > 1.$$

The desired result follows from induction. □

PROOF OF THEOREM 2.1. Let $\{Y_n\}$ be independent of $\{X_n\}$ where Y_n has the density function

$$\xi_n(u) = \begin{cases} n^{m+1}/\epsilon, & u \in [-\epsilon/(2n^{m+1}), \epsilon/(2n^{m+1})], \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$\begin{aligned} \beta_n(\epsilon) &= P(X_1 + X_2 + \cdots + X_n + Y_n \ge n\epsilon) \quad \text{and} \\ \alpha_n(\epsilon) &= P(X_1 + \cdots + X_n \ge n\epsilon). \end{aligned} \tag{4.6}$$

Then

$$\beta_n(\epsilon + \epsilon/(2n^{m+2})) \le \alpha_n(\epsilon) \le \beta_n(\epsilon - \epsilon/(2n^{m+2})). \tag{4.7}$$

It is easily seen that

$$Ee^{z(X_1 + \cdots + X_n + Y_n)} = \phi^n(z) \frac{n^{m+1}}{\epsilon z} (e^{z\epsilon/(2n^{m+1})} - e^{-z\epsilon/(2n^{m+1})}), \tag{4.8}$$

exists for $|z| < t_0$. Since $X_1 + \cdots + X_n + Y_n$ has an absolutely continuous d.f., by Lemma 4.2 we have

$$\begin{aligned} \beta_n(\epsilon + \epsilon/(2n^{m+2})) &= \frac{1}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} \frac{1}{z} e^{-(n\epsilon + \epsilon/(2n^{m+1}))z} \phi^n(z) Ee^{zY_n} dz \\ &= \frac{e^{-nh(\tau)}}{2\pi} \int_{-\infty}^{\infty} \frac{1}{(\tau + iy)^2} e^{n\epsilon iy} \cdot e^{-\epsilon(\tau + iy)/(2n^{m+1})} \\ &\quad \cdot \left(\frac{\phi(\tau + iy)}{\phi(\tau)} \right)^n \frac{n^{m+1}}{\epsilon} \\ &\quad \cdot [e^{\epsilon(\tau + iy)/2n^{m+1}} - e^{-\epsilon(\tau + iy)/2n^{m+1}}] dy. \end{aligned} \tag{4.9}$$

Define $\tilde{F}(u) = \int_{-\infty}^u (1/\phi(\tau)) e^{\tau v} dF(v)$. Then F is absolutely continuous with respect to $\tilde{F}$, and the c.f. of $\tilde{F}$ is $\phi(\tau + iy)/\phi(\tau)$. By Lemma 4.1, $\phi(\tau + iy)/\phi(\tau)$

satisfies Condition A. Thus, for any $\delta > 0$, there exists a constant $c' > 0$ which depends on δ such that

$$|\phi(\tau + iy)/\phi(\tau)| \leq e^{-c'} \quad \text{for} \quad |y| > \delta. \tag{4.10}$$

Denote the integrand of (4.9) by $I_n(y)$. It follows from (4.9) and (4.10), we have

$$\begin{aligned} \left|\frac{1}{2\pi} e^{-nh(\tau)} \int_{|y|\geq\delta} I_n(y)dy\right| &\leq \frac{n^{m+1}}{2\pi\epsilon} e^{-nh(\tau)-nc'} 2e^{\tau\epsilon} \int \frac{dy}{\tau^2 + y^2} \\ &= e^{-nh(\tau)} O(e^{-nc}) \end{aligned} \tag{4.11}$$

for some constant c such that $0 < c < c'$.

Let $\delta > 0$ be sufficiently small such that $\delta < \min(t_0 - \tau, \tau)$. It is easily seen that

$$\begin{aligned} &\frac{(e^{(\tau+iy)\epsilon/2n^{m+1}} - e^{-(\tau+iy)\epsilon/2n^{m+1}})}{[(\tau + iy)\epsilon/n^{m+1}]} = 1 + O(n^{-m-1}) \quad \text{and} \\ &e^{-\epsilon(\tau+iy)/2n^{m+1}} = 1 + O(n^{-m-1}) \end{aligned} \tag{4.12}$$

uniformly for $|y| \leq \delta$. It follows from (4.9), (4.11) and (4.12) that

$$\begin{aligned} &\beta_n(\epsilon + \epsilon/2n^{m+2}) \\ &\quad = \frac{1}{2\pi} e^{-nh(\tau)} \int_{-\delta}^{\delta} \frac{1}{\tau + iy} e^{-n[h(\tau+iy)-h(\tau)]} dy(1 + O(n^{-m-1})) \\ &\quad + e^{-nh(\tau)} O(e^{-nc}). \end{aligned} \tag{4.13}$$

Similarly, we have

$$\begin{aligned} &\beta_n(\epsilon - \epsilon/2n^{m+2}) \\ &\quad = \frac{1}{2\pi} e^{-nh(\tau)} \int_{-\delta}^{\delta} \frac{1}{\tau + iy} e^{-n[h(\tau+iy)-h(\tau)]} dy(1 + O(n^{-m-1})) \\ &\quad + e^{-nh(\tau)} O(e^{-nc}). \end{aligned} \tag{4.14}$$

By (4.7), (4.13) and (4.14),

$$\begin{aligned} \alpha_n(\epsilon) = e^{-nh(\tau)} \frac{1}{2\pi} \int_{-\delta}^{\delta} \frac{1}{\tau + iy} e^{-n[h(\tau+iy)-h(\tau)]} dy(1 + O(n^{-m-1})) \\ + e^{-nh(\tau)} O(e^{nc}). \end{aligned} \tag{4.15}$$

For $|y| \leq \delta$, we have the Taylor expansion

$$\begin{aligned} &-\left[h(\tau + iy) - h(\tau) - \frac{1}{2}\sigma^2 y^2\right] \\ &\quad = C_3(iy)^3 + C_4(iy)^4 + \cdots + C_{2m+3}(iy)^{2m+3} + \theta_1 |y|^{2m+4}, \end{aligned} \tag{4.16}$$

where $\sigma^2 = -h''(\tau)$, $|\theta_1| \le M$, $C_j = -h^{(j)}(\tau)/j!$, $|C_j| \le M$ for $j = 3, 4, \ldots, 2m+3$, and M represents a positive constant. Let

$$\Psi_m(iy) = C_3(iy)^3 + \cdots + C_{2m+3}(iy)^{2m+3} \tag{4.17}$$

and

$$\eta_m(iy) = 1 + n\Psi_m(iy) + \frac{1}{2!}n^2\Psi_m^2(iy) + \cdots + \frac{1}{(2m+1)!}n^{2m+1}\Psi_m^{2m+1}(iy). \tag{4.18}$$

Inserting equations (4.16) and (4.17) into (4.15), it yields

$$\begin{aligned}\alpha_n(\epsilon) = e^{-nh(\tau)}\frac{1}{2\pi}\int_{-\delta}^{\delta}\frac{1}{\tau+iy} &\cdot \exp(-n\sigma^2y^2/2 + n\Psi_m(iy) + n\theta_1 y^{2m+4})dy \\ &\cdot (1 + O(n^{-m-1})) + e^{-nh(\tau)}O(e^{-nc}).\end{aligned} \tag{4.19}$$

Let $\delta < \sigma^2/(4M + \sigma^2)$. For $|y| < \delta$, we have

$$|\Psi_m(iy)| + M|y|^{2m+4} \le M|y|^3/(1-\delta) \le M\delta y^2/(1-\delta) \le \sigma^2 y^2/4. \tag{4.20}$$

By Lemma 4.3, $|e^{\theta_1 n|y|^{2m+4}} - 1| \le Mn|y|^{2m+4}e^{Mn|y|^{2m+4}}$. Hence,

$$\begin{aligned}&\left|\frac{1}{2\pi}\int_{-\delta}^{\delta}\frac{1}{\tau+iy}\exp(-n\sigma^2y^2/2 + n\Psi_m(iy))(e^{\theta_1 n|y|^{2m+4}} - 1)dy\right| \\ &\le \frac{M}{\pi\tau}\int_0^{\delta} n|y|^{2m+4}\exp\{-n\sigma^2y^2/2 + nM\delta y^2/(1-\delta)\}dy \\ &\le \frac{M}{\pi\tau}\int_0^{\delta} n|y|^{2m+4}e^{-n\sigma^2y^2/4}dy \\ &\le \frac{M}{\pi\tau}\int_0^{\infty}\frac{nu^{2m+4}}{n^{m+2}}e^{-\sigma^2u^2/4}\cdot\frac{1}{\sqrt{n}}du = O(n^{-(m+3/2)}).\end{aligned} \tag{4.21}$$

Similarly, applying Lemma 4.3 and using (4.18) and (4.20), it yields

$$\left|\frac{1}{2\pi}\int_{-\delta}^{\delta}\frac{1}{\tau+iy}e^{-n\sigma^2y^2/2}(e^{n\Psi_m(iy)} - \eta_m(iy))dy\right| = O(n^{-(m+3/2)}). \tag{4.22}$$

Let

$$\xi_m(iy) = \frac{1}{\tau}\left[1 - \frac{iy}{\tau} + \left(\frac{-iy}{\tau}\right)^2 + \cdots + \left(\frac{-iy}{\tau}\right)^{2m+1}\right].$$

Then, for $|y| \le \delta$,

$$\frac{1}{\tau+iy} = \xi_m(iy) + \theta_2|y|^{2m+2},$$

where $|\theta_2| \le M$. Hence

$$\left|\frac{1}{2\pi}\int_{-\delta}^{\delta}\left(\frac{1}{\tau+iy}-\xi_m(iy)\right)\exp(-n\sigma^2y^2/2+\eta_m(iy))dy\right| = O(n^{-(m+3/2)}). \tag{4.23}$$

From equations (4.19), (4.21), (4.22) and (4.23), and the fact that

$$\begin{aligned}\left|\int_{|y|>\delta}\xi_m(iy)\eta_m(iy)e^{-n\sigma^2y^2/2}dy\right| &< e^{-(n-1)\sigma^2\delta^2/2}\int_{-\infty}^{\infty}|\xi_m(iy)\eta_m(iy)|e^{-\sigma^2y^2/2}dy \\ &= O(n^{-(m+3/2)}),\end{aligned}$$

we obtain

$$\begin{aligned}\alpha_n(\epsilon) = e^{-nh(\tau)}\frac{1}{2\pi}\int_{-\infty}^{\infty}\xi_m(iy)\eta_m(iy)e^{-n\sigma^2y^2/2}dy(1+O(n^{-m-1})) \\ + e^{-nh(\tau)}O(n^{-(m+3/2)}).\end{aligned} \tag{4.24}$$

Integrating term by term, the integrals of the odd power of (iy) in the product $\xi_m(iy)\eta_m(iy)$ are zero. Since $\mathrm{Re}[\xi_m(iy)\eta_m(iy)]$ is a polynomial of power y^2 it follows

$$\alpha_n(\epsilon) = e^{-nh(\tau)}\frac{1}{\sqrt{n}}\left\{a_0+\frac{a_1}{n}+\cdots+\frac{a_m}{n^m}+O(n^{-m-1})\right\},$$

where $a_0 > 0$ and a_j, $j = 0, 1, \dots$, are real numbers independent of n.

Taking $m = 1$ and after integrating term by term, the equation (4.24) becomes

$$\begin{aligned}\alpha_n(\epsilon) = e^{-nh(\tau)}\frac{1}{\sqrt{2\pi n}\sigma\tau}\left\{1+\frac{1}{n}\left(-\frac{1}{\sigma^2\tau^2}+\frac{h^{(3)}(\tau)}{2\tau\sigma^4}\right.\right. \\ \left.\left.-\frac{h^{(4)}(\tau)}{8\sigma^4}-\frac{5(h^{(3)}(\tau))^2}{24\sigma^6}\right)+O(n^{-2})\right\}.\end{aligned}$$

This completes our proof. □

LEMMA 4.4. *Suppose* $t_0 > 0$, G *is a d.f. with*

$$\phi(t) = \int_{-\infty}^{\infty} e^{tu}dG(u) < \infty \quad \textit{for} \quad 0 \le t < t_0,$$

and $\{\tau_j, j \ge 0\} \subset (0, t_0)$ *satisfies* $\tau_j \to \tau_0$. *Then*

$$\frac{\phi(\tau_j+iy)}{\phi(\tau_j)} \to \frac{\phi(\tau_0+iy)}{\phi(\tau_0)}$$

uniformly for $y \in (-\infty, +\infty)$ *as* $j \to \infty$.

PROOF. Since $\tau_0 \in (0, t_0)$ and $\tau_j \to \tau_0$, there exists b_1 and b_2 such that $0 < b_1 < \tau_0 < b_2 < t_0$ and $|e^{\tau_j u} - e^{\tau_0 u}| \leq e^{b_1 u} + e^{b_2 u}$ for j large. Note $\int (e^{b_1 u} + e^{b_2 u}) dG(u) \leq \infty$, by the dominated convergence theorem. Then

$$|\phi(\tau_j + iy) - \phi(\tau_0 + iy)| \leq \int |e^{\tau_j u} - e^{\tau_0 u}| dG(u) \to 0 \qquad \text{as} \quad j \to 0.$$

That is, $\phi(\tau_j + iy) \to \phi(\tau_0 + iy)$ uniformly for $y \in (-\infty, \infty)$. Therefore, Lemma 4.4 holds. □

PROOF OF THEOREM 2.2. Let Y be independent of $\{X_j\}$, with density

$$f_y(u) = \begin{cases} 1/a, & \text{if } u \in (0, a) \\ 0, & \text{otherwise.} \end{cases}$$

We have

$$Ee^{zY} = \frac{e^{za} - 1}{za}.$$

Let $\alpha_n = P\{X_1 + \cdots + X_n \geq k_n a\}$. Then $\alpha_n = P\{X_1 + \cdots + X_n + Y \geq k_n a\}$ and the d.f. of $X_1 + \cdots + X_n + Y$ is continuous. By Lemma 4.2, we have

$$\begin{aligned}
\alpha_n &= \frac{1}{2\pi i} \int_{\tau_n - i\infty}^{\tau_n + i\infty} \frac{1}{z} e^{-k_n a z} \phi^n(z) \frac{e^{za} - 1}{za} dz \\
&= \frac{1}{2\pi a} e^{-nh((k_n/n)a, \tau_n)} \int_{-\infty}^{\infty} \frac{e^{(\tau_n + iy)a} - 1}{(\tau_n + iy)^2} e^{-k_n a i y} \left[\frac{\phi(\tau_n + iy)}{\phi(\tau_n)} \right]^n dy \\
&= \frac{1}{2\pi a} e^{-nh((k_n/n)a, \tau_n)} \\
&\quad \cdot \int_{-\pi/a}^{\pi/a} \sum_{l=-\infty}^{\infty} \left\{ \frac{e^{(\tau_n + iy + 2\pi l i/a)a} - 1}{(\tau_n + 2\pi l i/a + iy)^2} e^{-k_n a i(2\pi l/a + y)} \right. \\
&\qquad\qquad \left. \cdot \left[\frac{\phi(\tau_n + iy + 2\pi l i/a)}{\phi(\tau_n)} \right]^n \right\} dy.
\end{aligned}$$

Note that the d.f. of X_1 is of the lattice type, so

$$\phi(\tau_n + iy + 2\pi l i/a) = \phi(\tau_n + iy), \qquad l = \pm 1, \pm 2, \ldots$$

Thus,

$$\begin{aligned}
\alpha_n &= \frac{1}{2\pi a} e^{-nh((k_n/n)a, \tau_n)} \\
&\quad \cdot \int_{-\pi/a}^{\pi/a} \sum_{l=-\infty}^{\infty} \frac{e^{(\tau_n + iy)a} - 1}{(\tau_n + 2\pi l i/a + y)^2} e^{-k_n a i y} \left[\frac{\phi(\tau_n + iy)}{\phi(\tau_n)} \right]^n dy \\
&= \frac{1}{2\pi a} e^{-nh((k_n/n)a, \tau_n)} \int_{-\pi/a}^{\pi/a} \sum_{l=-\infty}^{\infty} \frac{e^{(\tau_n + iy)a} - 1}{(\tau_n + 2\pi l i/a + y)^2} \\
&\quad \cdot e^{-n(h((k_n/n)a, \tau_n + iy) - h((k_n/n)a, \tau_n))} dy.
\end{aligned} \tag{4.25}$$

For each n, $\phi(\tau_n+iy)/\phi(\tau_n)$ is a characteristic function of a lattice distribution. Then by Theorem 6.4.7 of Chung (1974) and its Corollary, we have

$$\left|\frac{\phi(\tau_n+iy)}{\phi(\tau_n)}\right| < 1 \qquad \text{for all} \quad 0<|y|\le \pi/a. \tag{4.26}$$

Let $\delta>0$ such that $\delta<\min(\pi/a,a)$. Thus, it follows from (2.8) and Lemma 4.4 that there exists c' which depends on δ, such that

$$\left|\frac{\phi(\tau_n+iy)}{\phi(\tau_n)}\right| \le e^{-c'}, \qquad \delta\le|y|\le\pi/a, \qquad \text{for all } n. \tag{4.27}$$

Thus, we have

$$\begin{aligned}(4.28)\quad &\left|\frac{1}{2\pi a}e^{-nh((k_n/n)a,\tau_n)}\right.\\ &\quad\left.\cdot\int_{\delta\le|y|\le\pi/a}\sum_{l=-\infty}^{\infty}\frac{e^{(\tau_n+iy)a}-1}{(\tau_n+2\pi li/a+iy)^2}e^{-k_n aiy}\left[\frac{\phi(\tau_n+iy)}{\phi(\tau_n)}\right]^n dy\right|\\ &\quad\le \frac{1}{2\pi a}e^{-nh((k_n/n)a,\tau_n)}\left[\int_{|y|\le\pi/a}\sum_{l=-\infty}^{\infty}\frac{e^{\tau_n a}+1}{\tau_n^2+(2\pi l/a+y)^2}dy\right]e^{-nc'}\\ &\quad= e^{-nh((k_n/n)a,\tau_n)}O(e^{-nc'}).\end{aligned}$$

It is easily seen that there is an $M>0$, such that

$$\left|\frac{1}{(1+z)^2}-(1+2(-z)+3(-z)^2+\cdots+(2m+2)(-z)^{2m-1})\right|\le M|z|^{2m+2}$$

for all $|z|\le c<1$. Thus, for $|y|\le\delta$,

$$\begin{aligned}(4.29)\quad &\sum_{l=-\infty}^{\infty}\frac{1}{(\tau_n+2\pi li/a+iy)^2}\\ &\quad=\sum_{l=-\infty}^{\infty}\frac{1}{(\tau_n+2\pi li/a)^2}\cdot\frac{1}{\left(1+\dfrac{iy}{\tau_n+2\pi li/a}\right)^2}\\ &\quad=\sum_{l=-\infty}^{\infty}\frac{1}{(\tau_n+2\pi li/a)^2}\left(1+2\frac{(-iy)}{(\tau_n+2\pi li/a)}+\cdots\right.\\ &\qquad\qquad\left.+(2m+2)\left(\frac{-iy}{\tau_n+2\pi li/a}\right)^{2m+1}\right)\\ &\qquad+y^{2m+2}\theta\sum_{l=-\infty}^{\infty}\frac{1}{(\tau_n+2\pi li/a)^{2m+4}}\end{aligned}$$

where $|\theta| < M$. Let

$$(4.30)\quad \xi_m^*(iy) = \sum_{l=-\infty}^{\infty} \frac{1}{(\tau_n + 2\pi li/a)^2} + 2\sum_{l=-\infty}^{\infty} \frac{1}{(\tau_n + 2\pi li/a)^3}(-iy) + \cdots + (2m+2)\sum_{l=-\infty}^{\infty} \frac{1}{(\tau_n + 2\pi li/a)^{2m+3}}(-iy)^{2m+1}.$$

For $|y| \le \delta$ and $\delta < \min((t_0 - \tau_0)/2, \tau_0/2, \pi/a)$, we have

$$(4.31)\quad -\left[h\left(\frac{k_n}{h}a, \tau_n + iy\right) - h\left(\frac{k_n}{n}a, \tau_n\right) - \frac{1}{2}\sigma_n^2 y^2\right] = C_3^*(iy)^3 + C_4^*(iy)^4 + \cdots + C_{2m+3}^*(iy)^{2m+3} + \theta_1 |y|^{2m+4},$$

where

$$C_j^* = -\frac{1}{j!}\frac{\partial^j h(\epsilon, z)}{\partial z^j}\,|_{(k_n/n, \tau_n)},$$

and σ_n^2 is defined as (2.16). Let

$$\Psi_m^*(iy) = C_3^*(iy)^3 + C_4^*(iy)^4 + \cdots + C_{2m+3}^*(iy)^{2m+3}$$

and

$$\eta_m^*(iy) = 1 + n\Psi_m^*(iy) + \frac{1}{2!}n^2\Psi_m^{*2}(iy) + \cdots + \frac{1}{(2m+1)!}n^{2m+1}\Psi_m^{*2m+1}(iy).$$

With the same argument which we used in the proof of Theorem 2.1, we obtain

$$\begin{aligned}\alpha_n &= e^{-nh((k_n/n)a, \tau_n)}\frac{1}{2\pi a}\int_{-\infty}^{\infty}\left(e^{\tau_n a}(1 + iya + \frac{(iya)^2}{2!} + \cdots + \frac{(iya)^m}{m!} - 1\right) \\ &\quad \cdot \xi_m^*(iy)\eta_m^*(iy)e^{-n\sigma_n^2 y^2/2}dy(1 + O(n^{-m-1})) + (e^{-ny(\tau)}n^{-(m+3/2)}) \\ &= e^{-nh((k_n/n)a, \tau_n)}\frac{1}{\sqrt{n}}\left\{a_0 + \frac{a_1}{n} + \cdots + \frac{a_m}{n^m} + O(n^{-m-1})\right\},\end{aligned}$$

where $a_0 > 0$, and a_j, $j = 0, 1, \ldots$, are real numbers depending on τ_n and σ_n which are determined by k_n. If $k_n/n = k$, then $\tau_n = \tau_0$ and $\sigma_n = \sigma_0$, hence a_j, $j = 0, 1, \ldots$, are independent of n. Applying Formula 1.422.4 of Gradshteyn and Ryzhik (1965), gives

$$a_0 = \sum_{-\infty}^{\infty} \frac{1}{(\tau_n + 2\pi li/a)^2}(e^{a\tau_n} - 1)\frac{1}{\sqrt{2\pi}\sigma_n a} = \frac{a}{\sqrt{2\pi}\sigma_n(1 - e^{-\tau_n a})}. \qquad \square$$

Remark 6. Our Theorem 2.2 is clearly the case $\theta_n = 0$ of Bahadur and Rao (1960). It also covers case $0 < \theta_n < 1$ of Bahadur and Rao. This is immediately consequence of that, let $k_n = [n\epsilon]$ ($k_n/n \to \epsilon$, as $n \to \epsilon$),

$$P\left(\sum_{i=1}^{n} X_i \ge n\epsilon\right) = P\left(\sum_{i=1}^{n} X_i \ge [n\epsilon]\right)$$

for all n and $\epsilon > 0$.

Acknowledgements

The authors are grateful to the editor and referee for their detailed comments and suggestions resulting in considerable improvement of the text.

References

Bahadur, R. R. (1971). *Some limit theorems in statistics*, Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia.

Bahadur, R. R. and Rao, R. R. (1960). On deviations of the sample means, *Ann. Math. Statist.*, **31**, 1015–1027.

Bai, Z. C. and Fu, J. C. (1986). Likelihood principle and maximum likelihood estimator of location parameter for Cauchy distribution, *Canad. J. Statist.*, **15**, 137–146.

Barndorff-Nielsen, O. and Cox, D. R. (1979). Edgeworth and saddlepoint approximations with statistical applications, *J. Roy. Statist. Soc. Ser. B*, **41**, 279–312.

Cheng, P. and Fu, J. C. (1986). On a fundamental optimality of the maximum likelihood estimator, *Statist. Probab. Lett.*, **4**, 173–178.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of an hypothesis based on the sum of observations, *Ann. Math. Statist.*, **23**, 493–502.

Chung, K. L. (1974). *A Course of Probability Theory*, Academic Press, New York.

Cramer, H. (1937). Random variables and probability distributions, *Cambridge Tracts in Math.*, **36**.

Daniels, H. E. (1954). Saddlepoint approximations in statistics, *Math. Math. Statist.*, **25**, 631–650.

Daniels, H. E. (1987). Tail probability approximations in statistics, *Internat. Statist. Rev.*, **55**, 37–48.

Efron, B. (1975). Defining the curvature of a statistical problem, *Ann. Statist.*, **3**, 1189–1242.

Field, C. A. (1982). Small sample asymptotic expansions for multivariate M-estimators, *Ann. Statist.*, **10**, 672–689.

Fu, J. C. (1973). On a theorem of Bahadur on the rate of convergence of point estimators, *Ann. Statist.*, **1**, 741–749.

Fu, J. C. (1975). The rate of convergence of consistent point estimators, *Ann. Statist.*, **3**, 234–240.

Fu, J. C. (1982). Large sample point estimator: a large deviation theory approach, *Ann. Statist.*, **10**, 762–771.

Fu, J. C. (1985). On exponential rates of likelihood ratio estimators for location parameters, *Statist. Probab. Lett.*, **3**, 101–106.

Gradshteyn, I. S. and Ryzhik, I. M. (1965). *Tables of Integrals, Series and Products*, Academic Press, New York.

Hougaard, P. (1985). Saddlepoint approximation for curved exponential families, *Statist. Probab. Lett.*, **3**, 161–166.

Kester, A. and Kallenberg, W. C. M. (1986). Large deviations of estimators, *Ann. Statist.*, **14**, 648–664.

Kraft, C. H. and LeCam, L. M. (1956). A remark on the roots of maximum likelihood equation, *Ann. Math. Statist.*, **27**, 1174–1177.

Lugannani, R. and Rice, S. O. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables, *Adv. in Appl. Probab.*, **12**, 475–490.

Olver, F. W. J. (1974). *Asymptotic and Special Functions*, Academic Press, New York.

Petrov, V. V. (1975). *Sums of Independent Random Variables*, Springer, Berlin.

Reeds, J. A. (1985). Asymptotic number of roots of Cauchy location likelihood equations, *Ann. Statist.*, **13**, 775–784.

Rubin, H. and Rukhin, A. L. (1983). Convergence rates of large deviation probabilities for point estimators, *Statist. Probab. Lett.*, **1**, 197–202.

Sievers, G. (1978). Estimates of location: a large deviation comparison, *Ann. Statist.*, **6**, 610–618.

Weiss, L. and Wolfowitz, J. (1966). Generalized maximum likelihood estimators, *Teoriya Vyerogatnostey*, **11**, 68–93.

Weiss, L. and Wolfowitz, J. (1967). Maximum probability estimators, *Ann. Inst. Statist. Math.*, **19**, 193–200.

Weiss, L. and Wolfowitz, J. (1970). Maximum probability estimators and asymptotic sufficiency, *Ann. Inst. Statist. Math.*, **22**, 225–244.