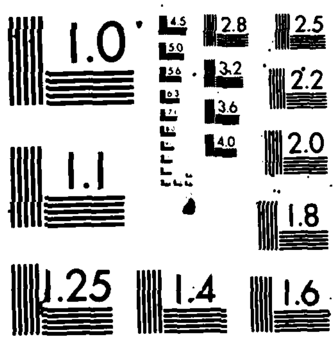
AD-A171 860 LIKELIHOOD PRINCIPLE AND MAXIMUM LIKELIHOOD ESTIMATOR OF LOCATION PARAMET (U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS Z D BAI ET AL MAY 86 TR-86-13 1/1

UNCLASSIFIED AFOSR-TR-86-0602 F49620-85-C-0008 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 86-0602
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): Likelihood principle and maximum likelihood estimator of location parameter for Cauchy distribution
5. TYPE OF REPORT & PERIOD COVERED: Technical - May 1986
6. PERFORMING ORG. REPORT NUMBER: 86-13
7. AUTHOR(s): Z.D. Bai and J.C. Fu
8. CONTRACT OR GRANT NUMBER(s): F49620-85-C-0008
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Air Force of Scientific Research, Department of the Air Force, Bolling Air Force Base, DC 20332
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 6.1102F-2304-A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Same as 9
12. REPORT DATE: May 1986
13. NUMBER OF PAGES: 17
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

DTIC ELECTE SEP 15 1986

18. SUPPLEMENTARY NOTES:
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): Likelihood function, maximum likelihood estimator, likelihood equation, Cauchy distribution, consistent estimator, first-order efficient, second-order efficient.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In the literature of point estimation, Cauchy distribution with location parameters was often cited as an example for the failure of maximum likelihood method and hence the failure of likelihood principle in general. Contrary to the above notion, we proved, even in this case that the likelihood equation has multiple roots, that the maximum likelihood estimator (the global maximum) remains as an asymptotically optimal estimator in the Bahadur sense.

DD FORM 1 JAN 73 1473

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AFOSR-TR. 86-0602

Approved for public release,
distribution unlimited

Center for Multivariate Analysis

University of Pittsburgh

86 9 15 15

LIKELIHOOD PRINCIPLE AND MAXIMUM LIKELIHOOD ESTIMATOR OF LOCATION PARAMETER FOR CAUCHY DISTRIBUTION

by

Z.D. Bai[1]
Center for Multivariate Analysis
University of Pittsburgh

J.C. Fu[2]
University of Manitoba
Canada

Technical Report No. 86-13

May 1986

Center for Multivariate Analysis
515 Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

[1]The work of this author is supported by Contract F49620-85-C-0008 of the Air Force of Scientific Research. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copy notation hereon.

[2]The work of this author was supported by the Natural Science and Engineering Research Council of Canada under grant NSERC A-9216.

ABSTRACT

In the literature of point estimation, Cauchy distribution with location parameters was often cited as an example for the failure of maximum likelihood method and hence the failure of likelihood principle in general. Contrary to the above notion, we proved, even in this case that the likelihood equation has multiple roots, that the maximum likelihood estimator (the global maximum) remains as an asymptotically optimal estimator in the Bahadur sense.

AMS 1970 Subject Classification: Primary 62F20 Secondary 62E10.

Key Words and Phrases: Likelihood function, maximum likelihood estimator, likelihood equation, Cauchy distribution, consistent estimator, first-order efficient, second-order efficient.

1. INTRODUCTION

For point estimation the parameter θ, the likelihood principle (see Fisher (1922, 1925)) yields the maximum likelihood estimator (m.l.e.) $\hat{\theta}$. There are considerable literatures studying various properties of m.l.e. The use of m.l.e. has a long history and may go back to Gauss and Edgeworth. For general review, it can be found from recent articles by Edwards (1972) and Norton (1972). The maximum likelihood method is a very controversial and emotional issue throughout the history of statistical point estimation. Recently there have been many articles still interested in this issue, to name a few, Berkson (1953, 1980), Efron (1975, 1982), Kraft and LeCam (1956), Rao (1980), Ferguson (1982), Fu (1982), and Reeds (1985).

Let $X_1,\ldots,X_n$ be n independent identically distributed (i.i.d.) observations having density function $f(x|\theta)$, where θ is a fixed value in the parameter space Θ. Given the data $s = (x_1,\ldots,x_n)$, the likelihood function of θ is defined by

$$(1.1) \qquad L_n(\theta|s) = \prod_{i=1}^{n} f(x_i|\theta).$$

For given s, the maximum likelihood estimator $\hat{\theta}_n(s)$ for θ is a value in the parameter space Θ which maximizes the likelihood function (1.1): i.e.

$$(1.2) \qquad L_n(\hat{\theta}_n(s)|s) = \max_{\theta\in\Theta} \prod_{i=1}^{n} f(x_i|\theta).$$

The standard method to obtain the maximum likelihood estimator $\hat{\theta}_n(s)$ is to find the root (or roots) of the following equation

$$(1.3) \qquad \ell_n^{(1)}(\theta|s) = \frac{d}{d\theta} \log L_n(\theta|s) = 0.$$

The equation (1.3) will be referred to as the likelihood equation.

Cramér (1946, p.500) proved that, under certain regularity conditions, there exists a sequence of roots $\hat{\theta}_n(s)$ of likelihood equation (1.3) which converges in probability to θ as n tends to infinite. Since then the consistency (or strong consistency) of maximum likelihood estimator has been studied by many researchers, for example, Wald (1949), Wolfowitz (1953, 1965) and LeCam (1953, 1970). Under certain regularity conditions, the maximum likelihood estimator $\hat{\theta}_n(s)$ is also asymptotically normally distributed with asymptotic variance $v(\theta)$ achieving the Cramer-Rao lower bound; i.e., for any estimator $T_n(s)$

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

then

(1.5) $$v(\theta) \geq 1/I(\theta)$$

where $I(\theta)$ is Fisher information and the equality holds when $T_n(s)$ is maximum likelihood estimator. Hence the m.l.e. is an asymptotically efficient estimator.

An example which was mostly cited in the literature for the failure of likelihood principle is when observations were sampled from a Cauchy distribution with a location parameter θ. For given $s = (x_1,\ldots,x_n)$ it has likelihood function

(1.6) $$L_n(\theta|s) = \prod_{i=1}^{n} \frac{1}{\pi(1+(x_i-\theta)^2)}$$

and likelihood equation

(1.7) $$\ell_n^{(1)}(\theta|s) = -\sum_{i=1}^{n} \frac{2(\theta-x_i)}{1+(x_i-\theta)^2} = 0.$$

The major reasons that the Cauchy distribution is often cited as an example for the failure of maximum likelihood method of estimation, hence the failure of likelihood principle in general (for example, Berkson (1980), Ferguson (1978) and Reeds (1985)), are as follows:

(a) The likelihood equation (1.7) associates with a polynominal with (2n-1) degrees. Hence it has (2n-1) roots (real and complex). The number of roots increases as the sample size increases.

(b) Neither analytical nor numerical solutions of the likelihood equation (1.7) can be obtained easily when sample size is moderately large.

(c) All the real roots, but one (the global maximum: m.l.e.), of the likelihood equation tend to $+\infty$ or $-\infty$ in probability as $n\to\infty$ (see Reeds (1985)).

(d) The asymptotic efficiency of maximum likelihood estimator (the global maximum) still remains unknown.

The main purpose of this paper is to show that the maximum likelihood estimator $\hat{\theta}_n(s)$ (the global maximum) converges to θ exponentially and is an asymptotically efficient estimator in the Bahadur sense.

2. Main Results

Fu (1971, 1973) proved that for any consistent estimator $T_n(s)$ and $\varepsilon > 0$

$$(2.1) \qquad \underline{\lim}_{n\to\infty} \frac{1}{n} \log P(|T_n(s)-\theta|\geq\varepsilon) \geq -B(\theta,\varepsilon)$$

and

(2.2) $$\underline{\lim}_{\varepsilon\to 0}\ \underline{\lim}_{n\to\infty}\ \frac{1}{n\varepsilon^2} \log P(|T_n(s)-\theta|\geq\varepsilon) \geq -I(\theta)/2$$

where $I(\theta)$ is Fisher information and $B(\theta,\varepsilon)$ is Bahadur bound defined by

(2.3) $$B(\theta,\varepsilon) = \inf_{\theta'}\{K(\theta',\theta): |\theta'-\theta|>\varepsilon\}$$

and the Kullback-Liebler information $K(\theta',\theta)$ is given by

(2.4) $$K(\theta',\theta) = \int_{-\infty}^{\infty} \{\log \frac{f(x|\theta')}{f(x|\theta)}\} f(x|\theta')dx.$$

The inequalities (2.1) and (2.2) provide an important conclusion in large sample theory of estimation that for any consistent estimator $T_n(s)$ the ε-tail probability

(2.5) $$\alpha_n(T_n,\theta,\varepsilon) = P(|T_n(s)-\theta|\geq\varepsilon)$$

cannot tend to zero faster than the rate $\exp\{-n[B(\theta,\varepsilon)+o(1)]\}$ (for fixed ε) or the rate $\exp\{-nI(\theta)\varepsilon^2/2\}$ (for ε near the zero) tends to zero. The estimator achieving the bound (2.2) is referred to as first-order efficient estimator in the Bahadur sense.

For given s, let $\bar{\theta}_n(s)$ and $\underline{\theta}_n(s)$ be the largest and the smallest real roots of the likelihood equation (1.3) respectively and the maximum likelihood estimator $\hat{\theta}_n(s)$ be a root of (1.3) which maximizes the likelihood function (1.7) (i.e. $\hat{\theta}_n(s)$ is a global maximum). It follows

(2.6) $$\underline{\theta}_n(s) \leq \hat{\theta}_n(s) \leq \bar{\theta}_n(s)$$

and for any $\varepsilon > 0$, the following inequalities hold

(2.7) $$P(\underline{\theta}_n(s)\geq\theta+\varepsilon) \leq P(\ell_n^{(1)}(s|\theta+\varepsilon)\geq 0) \leq P(\bar{\theta}_n(s)\geq\theta+\varepsilon),$$

and

$$(2.8)\qquad P(\bar{\theta}_n(s)\le\theta-\varepsilon)\ \le\ P(\ell_n^{(1)}(s|\theta-\varepsilon)\le 0)\ \le\ P(\underline{\theta}_n(s)\le\theta-\varepsilon).$$

If the likelihood equation (1.1) has a unique root for every s then $\bar{\theta}_n(s) = \hat{\theta}_n(s) = \underline{\theta}_n(s)$. Hence inequalities (2.7) and (2.8) yield the following inequalities

$$(2.9)\qquad P(\hat{\theta}_n\le\theta+\varepsilon) = P(\ell_n^{(1)}(s|\theta+\varepsilon)\ge 0),$$

and

$$(2.10)\qquad P(\hat{\theta}_n(s)\le\theta-\varepsilon) = P(\ell_n^{(1)}(s|\theta-\varepsilon)\le 0).$$

Under this assumption of unique root of likelihood equation (hence unique m.l.e. $\hat{\theta}_n(s)$), Fu (1973) proved

$$(2.11)\qquad \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n\varepsilon^2}\log P(|\hat{\theta}_n(s)-\theta|\ge\varepsilon) = -I(\theta)/2.$$

Hence the m.l.e. $\hat{\theta}_n(s)$ is an asymptotically efficient estimator in the Bahadur sense.

For the Cauchy distribution, it is clear that the condition of unique root of likelihood equation is violated. Whether, in this case, the m.l.e. $\hat{\theta}_n(s)$ remains as an asymptotically efficient estimator was listed as a conjecture in the papers of Fu (1982) and Rubin and Rukin (1983). The following theorem gave a positive answer to the conjecture.

<u>Theorem</u>: <u>For</u> $\varepsilon > 0$ <u>sufficiently small, we have</u>

$$(2.12)\qquad \lim_{n\to\infty}\frac{1}{n}\log P(|\hat{\theta}_n(s)-\theta|\ge\varepsilon) = -\beta(\hat{\theta},\theta,\varepsilon)$$

<u>where</u>

$$(2.13)\qquad \beta(\hat{\theta},\theta,\varepsilon) = \varepsilon^2(1+0(\sqrt{\varepsilon}))/4.$$

<u>and</u> $0(\sqrt{\varepsilon})$ <u>stands for</u> $0(\sqrt{\varepsilon})/\sqrt{\varepsilon}\to$ <u>constant as</u> $\varepsilon\to 0$.

The mathematical proof of this theorem is extremely hard and tedious. We leave the proof in the next section.

One may note that the above exponential rate $\beta(\hat{\theta},\theta,\varepsilon)$ is independent of θ. This is due to the fact that θ is a location parameter. Similarly the Fisher information $I(\theta)$ of the Cauchy distribution is

$$(2.14) \qquad I(\theta) = E\left(\frac{2(\theta-X)}{1+(X-\theta)^2}\right)^2 = \frac{1}{2}$$

which is also independent of θ. Results (2.13) and (2.14) together give

$$(2.15) \qquad \lim_{\varepsilon\to 0} \beta(\hat{\theta},\theta,\varepsilon)/\varepsilon^2 = I(\theta)/2 = \frac{1}{4}.$$

Hence the maximum likelihood estimator $\hat{\theta}_n$ for location parameter of Cauchy distributions is an asymptotically efficient estimator in the Bahadur sense.

The properties (a) and (b) of likelihood equation (1.7) of Cauchy distribution couldn't be altered. In view of Reeds' and our results, the m.l.e. $\hat{\theta}_n$ converges to θ exponentially with the optimal rate. One could not consider the Cauchy distribution as a major example for the failure of maximum likelihood method of estimation.

3. Proof of Main Theorem

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. bounded random variables with mean zero, $\mathrm{Var}(X_1) = \sigma^2 > 0$, and $|X_1|<M$. Write $S_n = \sum_{i=1}^{n} X_i$. To prove our theorem we need the following lemmas.

<u>Lemma 1</u>. <u>For any</u> $\varepsilon>0$, <u>when</u> n <u>is sufficiently large then</u>

$$(3.1) \qquad P(|S_n|\geq\varepsilon n) \leq 2\exp\{-\frac{\varepsilon^2 n}{2\sigma^2}(1-\frac{\varepsilon M}{2\sigma^2})\}.$$

<u>Lemma 2</u>. <u>For any</u> $0<\varepsilon\leq\frac{\sigma^2}{M}$, <u>when</u> n <u>is sufficiently large then</u>

$$(3.2) \qquad P(|S_n|\geq\varepsilon n) \geq 2\exp\{-\frac{\varepsilon^2 n}{2\sigma^2}(1+5\sqrt{\frac{2\varepsilon M}{\sigma^2}})\}.$$

The above lemmas can be proved by the same methods used for Lemma 1 and Lemma 2 in Chapter X, Petrov (1975). We omit the proofs.

<u>Proof of Main Theorem</u>:

Without loss of generality, we can assume that the true value of θ equals zero and that $\{X_n\}$, $n = 1,2,\ldots$ will represent a sequence of i.i.d. standard Cauchy random variables with common density function

$$(3.3) \qquad f(x) = \frac{1}{\pi(1+x^2)}, \qquad \text{for all } x\in(-\infty,\infty).$$

The E will stand for expectation with respect to the standard Cauchy distribution. Let θ be the variable of following log-likelihood function

$$(3.4) \qquad S_n(\theta) = \sum_{i=1}^{n} \log\,(1+(X_i-\theta)^2).$$

It follows

$$(3.5) \qquad \frac{1}{n} E\, S_n(\theta) = \log\,(\theta^2+4).$$

Take

$$(3.6) \qquad \delta = \frac{1}{4}(\log(4+\frac{1}{4}) - \log 4) = \frac{1}{4}\log\,(1+\frac{1}{16}) \cong 0.0152,$$

and define

$$(3.7) \qquad A_n(\theta) = \{S_n(\theta) \leq n(\log 4 + 2\delta)\},$$

then we have, for any $t \epsilon (0,1/4)$,

$$(3.8) \qquad P(A_n(\theta) = P\{\sum_{i=1}^{n} \log \frac{\theta^2+4}{1+(X_i-\theta)^2} \quad n(\log \frac{4+\theta^2}{4} - 2\delta)\}$$

$$\leq [\exp\{-nt(\log \frac{4+\theta^2}{4} - 2\delta)\}]E\{\exp[t \log \frac{\theta^2+4}{1+(X_1- \ )^2}]\}^n$$

$$= [\exp\{-nt(\log \frac{4+\theta^2}{4} - 2\delta)\}]\{1+ \sum_{k=2}^{\infty} \frac{t^k}{k!} E(\log \frac{\theta^2+4}{1+(X_1-\theta)^2})^k\}^n.$$

If $\theta > 0$ and $X_1 < \frac{\theta}{2}$ then

$$(3.9) \qquad \log \frac{\theta^2+4}{1+(X_1-\theta)^2} \leq \log \frac{\theta^2+4}{1+\theta^2/4} = \log 4.$$

If $\theta > 0$ and $X_1 > \frac{\theta}{2}$ then

$$(3.10) \qquad P(X_1 > \frac{\theta}{2}) \leq \frac{2}{\pi\theta} < \frac{1}{\theta} \text{ and}$$

$$(3.11) \qquad \log \frac{\theta^2+4}{1+(X_1-\theta)^2} \leq \log (\theta^2+4).$$

Thus (3.9), (3.10) and (3.11) yields

$$(3.12) \qquad E(\log \frac{\theta^2+4}{1+(X_1-\theta)^2)})^k \leq (\log 4)^k + \frac{1}{\theta} (\log(\theta^2+4))^k.$$

Inequalities (3.8) and (3.12) imply that for $\theta > \frac{1}{2}$,

$$(3.13) \qquad P(A_n(\theta)) \leq \exp\{-n[t(\log \frac{\theta^2+4}{4} - 2\delta) - At^2]\},$$

where

$$(3.14) \qquad A = [\log^2 4 + \frac{1}{2} \sup_{\theta>1/2} (\frac{1}{\theta}(\theta^2+4)^{1/4} \log^2(\theta^2+4))].$$

Taking $t = \delta/A$ and inserting it into the inequality (3.13), we get, for $\theta > \frac{1}{2}$,

$$(3.15) \qquad P(A_n(\theta)) \le \exp\{-n\delta A^{-1}(\log \frac{\theta^2+4}{4} - 3\delta)\}.$$

Similarly, the inequality (3.15) also holds for $\theta < -\frac{1}{2}$.

Let $\theta_k = \frac{1}{2} + k\delta$, $k = 0,1,2,\ldots.$ It follows from (3.15) that we have for n sufficiently large,

$$(3.16) \qquad \sum_{k=0}^{\infty} P(A_n(\theta_k)) \le \sum_{k=0}^{\infty} \exp\{-n\delta A^{-1}(\log \frac{4+\theta_k^2}{4+1/4} + \delta)\}$$

$$\le 2 \exp\{-n\delta^2 A^{-1}\}.$$

Note that

$$(3.17) \qquad \left|\frac{1}{n} S_n^{(1)}(\theta)\right| = \left|\frac{1}{n} \sum_{i=1}^{n} \frac{2(\theta-X_i)}{1+(\theta-X_i)^2}\right| \le 1.$$

If $\theta \in (\theta_k, \theta_{k+1})$ and $\frac{1}{n} S_n(\theta) \le \log 4 + \delta$ then

$$(3.18) \qquad \frac{1}{n} S_n(\theta_k) = \frac{1}{n} S_n(\theta) + (\frac{1}{n} S_n(\theta_k) - \frac{1}{n} S_n(\theta))$$

$$= \frac{1}{n} S_n(\theta) - (\theta-\theta_k)(\frac{1}{n} S_n^{(1)}(\xi_k))$$

$$\le \log 4 + 2\delta.$$

Hence

$$(3.19) \qquad P(\sup_{\theta \ge \frac{1}{2}} \frac{1}{n} S_n(\theta) \le \log 4 + \delta)$$

$$\le P(\bigcup_{k=0}^{\infty} A(\frac{1}{n} S_n(\theta_k) \le \log 4 + 2\delta))$$

$$\le \sum_{k=0}^{\infty} P(\ _n(\theta_k)) \le 2 \exp\{-n\delta^2/A\}.$$

By the same token we have

(3.20) $$P(\sup_{\theta<-\frac{1}{2}} \frac{1}{n} S_n(\theta) \leq \log 4 + \delta) \leq 2 \exp\{-n\delta^2/A\}.$$

Inequalities (3.19) and (3.20) imply that

(3.21) $$P(\sup_{|\theta|>\frac{1}{2}} \frac{1}{n} S_n(\theta) \leq \log 4 + \delta) \leq 4 \exp\{-n\delta^2/A\}.$$

On the other hand, we have for $t\epsilon(0,1/4)$

(3.22) $$P(\frac{1}{n} S_n(0) \geq \log 4 + \delta) = P(\frac{1}{n} \sum_{i=1}^{n} \log \frac{1+X_1{}^2}{4} \geq \delta)$$

$$[\exp\{-n\delta t\}](E \exp\{t \log \frac{1+X_1{}^2}{4})^n.$$

Since

$$E \exp\{t \log \frac{1+X_1{}^2}{4}\} < \infty, \qquad \text{for all } t < \frac{1}{2}$$

and

$$E \log \frac{1+X_1{}^2}{4} = 0$$

thus there exists a constant $\Delta > 0$ such that

(3.23) $$E \exp\{t \log \frac{1+X_1{}^2}{4}\} \leq 1 + \frac{t^2}{2} \Delta \leq \exp\{\frac{t^2}{2}\Delta\}$$

for all sufficiently small positive t. Therefore

(3.24) $$P(\frac{1}{n} S_n(0) > \log 4 + \delta) \leq \exp\{-n\delta t +n\frac{t^2}{2}\Delta\}.$$

Taking $t = \delta/\Delta$, we have

(3.25) $$P(\frac{1}{n} S_n(0) \geq \log 4 + \delta) \leq \exp\{-n\delta^2/2\Delta\}.$$

Now we consider the case when $\theta\epsilon(- \frac{1}{2},\frac{1}{2})$. Note that

(3.26) $$S_n^{(2)}(\theta) = 2\sum_{i=1}^{n} \frac{1-(\theta-X_i^2)}{(1+(\theta-X_i)^2)^2}$$

and

(3.27) $$E\,\frac{1}{n}\,S_n^{(2)}(\theta) = \frac{8-2\theta^2}{(\theta^2+4)^2}\,.$$

Hoeffding (1965) proved that if $X_1,..,X_n$ are independent and $a_i \leq X_i \leq b_i$ then, for $t > 0$ the following inequality holds

$$P\Big(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} EX_i \geq nt\Big) \leq \exp\{-2n^2t^2/\sum_{i=1}^{n}(b_i-a_i)\}.$$

Since

(3.28) $$-\frac{1}{4} \leq \frac{2(1-(X_i-\theta)^2)}{(1+(X_i-\theta)^2)^2} \leq 2$$

it follows from Hoeffding's inequality that we have

(3.29)
$$P\Big(\frac{1}{n}\,S_n^{(2)}(\theta) \leq \frac{1}{5}\Big)$$

$$= P\Big(\frac{1}{n}\sum_{i=1}^{n}\Big[\frac{8-2\theta^2}{(\theta^2+4)^2} - \frac{2(1-(X_i-\theta)^2)}{(1+(X_i-\theta)^2)^2}\Big] \geq \frac{8-2\theta^2}{(\theta^2+4)^2} - \frac{1}{5}\Big)$$

$$\leq P\Big(\frac{1}{n}\sum_{i=1}^{n}\Big[\frac{8-2\theta^2}{(\theta^2+4)^2} - \frac{2(1-(X_i-\theta)^2)}{(1+(X_i-\theta)^2)^2}\Big] \geq \frac{1}{25}\Big)$$

$$\leq \exp\{-2n(\tfrac{1}{75})^2], \qquad \text{for all } |\theta|<1.$$

Again, since

(3.30) $$\Big|\frac{1}{n}\,S_n^{(3)}(\theta)\Big| = \Big|\frac{4}{n}\sum_{i=1}^{n}\frac{(\theta-X_i)(3-(\theta-X_i)^2}{(1+(X_i-\theta)^2)3}\Big| \leq 6,$$

if $\theta\epsilon(\frac{2k}{75},\frac{2(k+1)}{75})$ and $\frac{1}{n}\, S_n^{(2)}(\theta) \le \frac{1}{25}$, then

$$\frac{1}{n}\, S_n^{(2)}(\tfrac{2k}{75}) = \frac{1}{n}\, S_n^{(2)}(\theta) + (\frac{1}{n}\, S_n^{(2)}(\tfrac{2k}{75}) - \frac{1}{n}\, S_n^{(2)}(\theta))$$

$$= \frac{1}{n}\, S_n^{(2)}(\theta) + (\frac{2k}{75} - \theta)\, \frac{1}{n}\, S_n^{(3)}(\xi_k) \tag{3.31}$$

$$< \frac{1}{25} + \frac{2}{75}\cdot 6 = \frac{1}{5}$$

for k = -36, -35,...,35,36. Hence

$$P(\inf_{|\theta|\le \frac{24}{25}} \frac{1}{n}\, S_n^{(2)}(\theta) \le \frac{1}{25}) \le \sum_{k=-36}^{36} P(\frac{1}{n}\, S_n^{(2)}(\tfrac{2k}{75}) \le \frac{1}{5}) \tag{3.32}$$

$$\le 75 \exp\{-2n(\tfrac{1}{75})^2\}.$$

Now we define the following events

$$A_{n1} = \{ \sup_{|\theta|\ge\frac{1}{2}} \frac{1}{n}\, S_n(\theta) \le \log 4 + \delta\},$$

$$A_{n2} = \{\frac{1}{n}\, S_n(0) \ge \log 4 + \delta\}, \text{ and}$$

$$A_{n3} = \{ \inf_{|\theta|\le\frac{24}{25}} \frac{1}{n}\, S_n^{(2)}(\theta) \le \frac{1}{25}\}.$$

Note that the m.l.e. $\hat{\theta}_n(s)$ minimizes $S_n(\theta)$ and satisfies the likelihood equation $S_n^{(1)}(\theta) = 0$. If A_{n1}^c and A_{n2}^c occur simultaneously, then $\hat{\theta}_n\epsilon(-\frac{1}{2},\frac{1}{2})$. If A_{n3}^c occurs then $S_n^{(1)}(\theta)$ is strictly increasing on the interval $(-\frac{24}{25}\ \frac{24}{25})$. If A_{n1}^c, A_{n2}^c, and A_{n3}^c simultaneously occur then, for $\varepsilon\epsilon(0,\frac{1}{2})$, $\hat{\theta}n > \varepsilon$ if and only if $S_n^{(1)}(\varepsilon) < 0$. Hence

(3.34) $$P(\hat{\theta}_n(s) \geq \varepsilon,\ A^c_{n1},\ A^c_{n2},\ \text{and } A^c_{n3})$$

$$= P(S_n^{(1)}(\varepsilon) < 0,\ A^c_{n1},\ A^c_{n2},\ \text{and } A^c_{n3}).$$

Note that

(3.35) $$E\frac{2(\varepsilon-X_1)}{1+(\varepsilon-X_1)^2} = \frac{2\varepsilon}{\varepsilon^2+4},$$

by Lemmas 1 and 2, we have for $\varepsilon > 0$ sufficiently small

(3.36) $$P(S_n^{(1)}(\varepsilon)<0) = P(\sum_{i=1}^{n}(\frac{2\varepsilon}{\varepsilon^2+4} - \frac{2(\varepsilon-X_i)}{1+(\varepsilon-X_i)^2}) > n\frac{2\varepsilon}{\varepsilon^2+4})$$

$$= \exp\{-\frac{n}{2}(\frac{2\varepsilon}{\varepsilon^2+4})^2(\frac{(\varepsilon^2+4)^2}{8-2\varepsilon} + 0(\sqrt{\varepsilon}))\}$$

$$= \exp\{-\frac{n}{4}\varepsilon^2(1 + 0(\sqrt{\varepsilon}))\}.$$

From the above results (3.20), (3.25), (3.32), (3.34) and (3.36), we have

(3.37) $$P(\hat{\theta}_n(s)>\varepsilon) = \exp\{-\frac{n}{4}\varepsilon^2(1 + 0(\sqrt{\theta}))\},$$

for $\varepsilon > 0$ sufficiently small.

Similarly we have

(3.38) $$P(\hat{\theta}_n(s)<-\varepsilon) = P(S_n^{(1)}(-\varepsilon)>0)$$

$$= \exp\{-\frac{n}{4}\varepsilon^2(1 + 0(\sqrt{\varepsilon}))\},$$

for $\varepsilon > 0$ sufficiently small.

Since Cauchy distribution is continuous and satisfies Bahadur's condition (Bahadur 1971 p.9) hence equations (3.37) and (3.38) yield

$$P(|\hat{\theta}_n(s)| \geq \varepsilon) = \exp\{-\frac{n}{4}\varepsilon^2(1 + 0(\sqrt{\varepsilon}))\}$$

for $\varepsilon > 0$ sufficiently small. This completes the proof.

REFERENCES

[1] Bahadur, R.R. (1971). Some limit theorems in statistics. Regional Conference Series in Applied Mathematics, 4, Philadelphia: SIAM.

[2] Berkson, J. (1955). Maximum likelihood and minimum χ^2 estimates of logistic function. J. Amer. Statist. Assoc., Vol. 50, 130-162.

[3] Berkson, J. (1980). Minimum chi-square, not maximum likelihood: (with discussion). Ann. Statist., Vol. 8, 457-487.

[4] Cramer, H. (1946). Mathematical methods of statistics. Princeton University Press.

[5] Edwards, A.W.F. (1972). Likelihood. Cambridge University Press.

[6] Efron, B. (1975). Defining the curvature of a statistical problem: (with discussion). Ann. Statist., Vol. 3, 1189-1242.

[7] Efron, B. (1982). Maximum likelihod and decision theory. Ann. Statist. Vol. 10, 340-356.

[8] Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. Phil. Trans. Royal Soc. of London A. Vol. 222, 309-360.

[9] Fisher, R.A. (1925). Theory of statistical estimations. Proc. Cambridge Phil. Soc., Vol. 22, 700-725.

[10] Ferguson, T.S. (1978). Maximum likelihood estimates of the parameters of the Cauchy distribution for samples of size 3 and 4. J. Amer. Statist. Assoc., Vol. 73, 211-213.

[11] Furguson, T.S. (1982). An inconsistent maximum likelihood estimate. J. Amer. Statist. Assoc., Vol. 77, 831-834.

[12] Fu, J.C. (1971). Some large sample methods in statistical point estimation. Technical Report, The Johns Hopkins University.

[13] Fu, J.C. (1973). On a theorem of Bahadur on the rate of convergence of point estimators. Ann. Statist., Vol. 1, 741-749.

[14] Fu, J.C. (1982). Large sample point estimation: A large deviation theory approach. Ann. Statist., Vol. 10, 762-771.

[15] Hoeffding, W. (1965). On probability of large deviations. Proc. Fifth Berkeley Symp. Math. Statist. Prob., Vol. 1, 203-219.

[16] Kraft, C.H. and LeCam, L.M. (1956). A remark on the roots of the maximum likelihood equation. Ann. Math. Statist., Vol. 27, 1174-1177.

[17] LeCam, L.M. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. University of California Publications in Statistics. Vol. 1, 277-328.

[18] LeCam, L.M. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. Ann. Math. Statist., Vol. 41, 802-828.

[19] Norton, R.H. (1972). A survey of maximum likelihood estimations. Review of the International Statistical Institute. Vol., 329-354, and part II (1973) Vol. 41, 39-58.

[20] Peterov, V.V. (1975), Sum of independent random variables. Springer-Verlay.

[21] Rao, C.R. (1980). Discussion of Berkson's (1980) paper.

[22] Reeds, J.A. (1985). Asymptotic number of roots of Cauchy location likelihood equations. Ann. Statist. Vol. 13, No. 2, 775-784.

[23] Rubin, H. and Rukin, A.L. (1983). Convergence rates of large deviation probabilities for point estimators. Statistics and Probability Letters Vol. 1, No. 3, 197-202.

[24] Wald, A. (1949). Note on the consistency of maximum likelihood estimates. Ann. Math. Statist., Vol. 20, 595-601.

[25] Wolfowitz, J. (1953). The method of maximum likelihood and Wald theory of decision functions. Indag. Math., Vol. 15, 114-119.

[26] Wolfowitz, J. (1965). Asymptotic efficiency of maximum likelihood estimator. Theory Prob. Appl. Vol. 10, 247-260.

END
10-86
DTIC