The Annals of Statistics
1982, Vol. 10, No. 3, 762–771

LARGE SAMPLE POINT ESTIMATION: A LARGE DEVIATION THEORY APPROACH[1]

BY JAMES C. FU

University of Manitoba

In this paper the exponential rates of decrease and bounds on tail probabilities for consistent estimators are studied using large deviation methods. The asymptotic expansions of Bahadur bounds and exponential rates in the case of the maximum likelihood estimator are obtained. Based on these results we have obtained a result parallel to the Fisher-Rao-Efron result concerning second-order efficiency (see Efron, 1975). Our results also substantiate the geometric observation given by Efron (1975) that if the statistical curvature of the underlying distribution is small, then the maximum likelihood estimator is nearly optimal.

1. Introduction. In large sample point estimation one needs to be concerned only with consistent estimators. For any consistent sequence of estimators T_n, and any given $\varepsilon > 0$, the tail probability

$$\alpha(T_n, \theta, \varepsilon) = P(|T_n - \theta| \geq \varepsilon) \tag{1.1}$$

tends to zero as $n \to \infty$. It has been suggested (Basu, 1956; Bahadur, 1971; Fu, 1973) that the rate of convergence to zero of $\alpha(T_n, \theta, \varepsilon)$ be used as a criterion for evaluating the asymptotic performance of consistent estimators. Typically, for consistent estimators, this rate of convergence is exponential and has an asymptotic expansion given by

$$\alpha(T_n, \theta, \varepsilon) = e^{-n\beta(T,\theta,\varepsilon)}(c_{0,n} + n^{-1}c_{1,n} + \cdots). \tag{1.2}$$

Here $\beta(T, \theta, \varepsilon)$ and $c_{i,n}$ are constants that may depend on ε, the sequence of estimators T_n, and the underlying distribution. The positive constant $\beta(T, \theta, \varepsilon)$ is called an exponential rate.

Recently, there have been several papers in the literature which examined the rates of convergence of consistent estimators with regard to the probabilities of large deviations, for example, Bahadur (1971), Basu (1956), Chernoff (1952), Fu (1973, 1975), Kester (1981), and Sievers (1978). This paper is a continuation of the author's previous work (1973, 1975) studying the exponential rates $\beta(T, \theta, \varepsilon)$ and their upper bounds (Bahadur bounds) for constant estimators.

In Section 2, we study the Bahadur bound and the exponential rate for the maximum likelihood estimator (m.l.e.) $\hat{\theta}_n$. We begin by deriving the first four terms of the respective Taylor expansions. The results show that the exponential rate of an m.l.e. has a third-order contact with respect to the Bahadur bound at $\varepsilon = 0$. This yields a Fisher-Rao-Efron type second-order efficiency for $\hat{\theta}_n$ in the Bahadur sense. The second-order Bahadur efficiency depends on the statistical curvature. In Section 3 we derive these results. Section 4 studies the connection between the rate of convergence and the Fisher-Rao-Efron results in second-order efficiency. The final section contains examples for illustrating the results in Sections 2 and 3.

Received January 1981; revised February 1982.

[1] This work was supported in part by the National Sciences and Engineering Research Council of Canada under grant NSERC A-9216.

AMS 1980 *subject classifications.* Primary 62E20; secondary 62E10.

Key words and phrases. Consistent estimator, probability of large deviation, exponential rate of convergence, statistical curvature, second-order efficiency, maximum likelihood estimator, Bahadur bound.

2. Main results. Let $S = (X_1, \cdots, X_n)$ be a sequence of i.i.d. random variables with common density function $f(x|\theta)$, where $\theta \in \Theta$ and the parameter space Θ is an open interval of real line. A sequence of estimators $T_n(X_1, \cdots, X_n)$ will be referred to here as an estimator T_n. It has been proved by Bahadur (1971), Fu (1971), and Sievers (1978) that for any consistent estimator T_n and $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha(T_n, \theta, \varepsilon) \geq -B(\theta, \varepsilon), \tag{2.1}$$

where $B(\theta, \varepsilon)$, the Bahadur bound, is given by

$$B(\theta, \varepsilon) = \inf_{\theta'}\{K(\theta', \theta) : |\theta' - \theta| > \varepsilon\}, \tag{2.2}$$

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

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

The inequality (2.1) suggests that there exists no consistent estimator T_n whose exponential rate $\beta(T, \theta, \varepsilon)$ is greater than the Bahadur bound $B(\theta, \varepsilon)$. Bahadur (1971) indirectly and Fu (1971) directly proved that for any consistent estimator T_n,

$$\lim_{\varepsilon\to 0} \varepsilon^{-2}\beta(T, \theta, \varepsilon) \leq I(\theta)/2, \tag{2.3}$$

where $I(\theta)$ is the Fisher information. Furthermore, the m.l.e. $\hat{\theta}_n$ was shown to be asymptotically efficient in the sense that

$$\lim_{\varepsilon\to 0} \varepsilon^{-2}\beta(\hat{\theta}, \theta, \varepsilon) = I(\theta)/2. \tag{2.4}$$

Loosely speaking, (2.3) and (2.4) imply that for any consistent estimator T_n, the tail probability $\alpha(T_n, \theta, \varepsilon)$ cannot tend to zero faster than the rate given by $\exp\{-\frac{1}{2}n\varepsilon^2 I(\theta)\}$, but that the tail probability $\alpha(\hat{\theta}_n, \theta, \varepsilon)$ does tend to zero nearly at this optimal rate. We shall refer to estimators satisfying equation (2.4) as being first-order efficient in the Bahadur sense. It is known (Fu, 1973, 1975; Sievers, 1978) that there are many consistent estimators satisfying the above optimal asymptotic criteria. This leads to further study of the asymptotic behaviour of the Bahadur bound $B(\theta, \varepsilon)$ and the exponential rate $\beta(T, \theta, \varepsilon)$ as ε nears zero.

In the following, we state the main results of this paper and leave the conditions and proofs in the next section.

From a direct expansion of the density function $f(x|\theta')$ for θ' near θ, we can derive a four-term Taylor expansion for the Bahadur bound $B(\theta, \varepsilon)$ as follows:

$$\begin{aligned} B(\theta, \varepsilon) = \min \Big\{ &\frac{\varepsilon^2}{2!} I(\theta) \pm \frac{\varepsilon^3}{3!} (3\mu_{110} - \mu_{300}) \\ &+ \frac{\varepsilon^4}{4!} (2\mu_{400} - 6\mu_{210} + 4\mu_{101} + 3\mu_{020}) + o(\varepsilon^4) \Big\}, \end{aligned} \tag{2.5}$$

where μ_{ijk} is given by, in Fisher's notation,

$$\mu_{ijk} = E\left(\frac{f^{(1)}}{f}\right)^i \left(\frac{f^{(2)}}{f}\right)^j \left(\frac{f^{(3)}}{f}\right)^k. \tag{2.6}$$

Note that the Fisher information, $I(\theta) = \mu_{200}$, is the leading term of the expansion. Thus $I(\theta)$ provides the most important contribution to the exponential bound. The second most important contribution is provided by $(3\mu_{110} - \mu_{300})$. When θ is a location parameter, this second term is zero if the underlying distribution is symmetric, and non-zero if the underlying distribution is asymmetric. In the location parameter case, it seems that $(3\mu_{110} - \mu_{300})$ has a connection with the skewness of the underlying distribution. When θ is more general the connection is not so clear. The third most important term is $(2\mu_{400} - 6\mu_{210} + 4\mu_{101} + 3\mu_{020})$. At present we do not fully understand what characteristic of the underlying distribution is represented by this term.

By extending the technique used in Fu (1973) which is a combination of results and ideas due to Daniels (1961), Chernoff (1952), Bahadur (1971) and Hoeffding (1965), we obtain the following four-term Taylor expansion of the exponential rate $\beta(\hat{\theta}, \theta, \varepsilon)$ for the m.l.e. $\hat{\theta}_n$:

$$\beta(\hat{\theta}, \theta, \varepsilon) = \min \left\{ \frac{\varepsilon^2}{2!} I(\theta) \pm \frac{\varepsilon^3}{3!} (3\mu_{110} - \mu_{300}) + \frac{\varepsilon^4}{4!} [(2\mu_{400} - 6\mu_{210} + 4\mu_{101} + 3\mu_{020}) - 3I^2(\theta)\gamma^2(\theta)] + o(\varepsilon^4) \right\}, \tag{2.7}$$

where $\gamma^2(\theta)$ is the statistical curvature given by Efron (1975), namely

$$\gamma^2(\theta) = \frac{1}{I^2(\theta)} (\mu_{020} - 2\mu_{210} + \mu_{400}) - 1 - \frac{1}{I^3(\theta)} (\mu_{110} - \mu_{300})^2. \tag{2.8}$$

It follows from (2.5) and (2.7) that

$$\lim_{\varepsilon\to 0} \varepsilon^{-i}[B(\theta, \varepsilon) - \beta(\hat{\theta}, \theta, \varepsilon)] = \begin{cases} 0, & i = 1, 2, 3 \\ \tfrac{1}{8} I^2(\theta)\gamma^2(\theta), & i = 4. \end{cases} \tag{2.9}$$

Furthermore, if θ is a location parameter for a symmetric distribution, then for any translation invariant consistent estimator T_n,

$$\lim_{\varepsilon\to 0} \varepsilon^{-4}[B(\theta, \varepsilon) - \beta(T, \theta, \varepsilon)] \geq \tfrac{1}{8} I^2(\theta)\gamma^2(\theta), \tag{2.10}$$

with equality when $T_n = \hat{\theta}_n$.

The results (2.1), (2.5), (2.7), and (2.10) are parallel to the results given by Fisher (1925), Rao (1963) and Efron (1975) in the theory of second-order efficiency. Efron (1975) gave an elegant and penetrating geometric interpretation for the statistical curvature $\gamma^2(\theta)$ and its roles in large sample inference. The connection between the exponential rate of convergence and the Fisher-Rao-Efron results will be studied in Section 4.

The equations (2.5) and (2.7) imply that the exponential rate $\beta(\hat{\theta}, \theta, \varepsilon)$ for the m.l.e. $\hat{\theta}_n$ has a third order contact with the Bahadur bound $B(\theta, \varepsilon)$ at $\varepsilon = 0$. The statistical curvature $\gamma^2(\theta)$ is the first term which separates the exponential rate of $\hat{\theta}_n$ from the Bahadur bound. It has a fourth order contact only when the statistical curvature $\gamma^2(\theta)$ of underlying distribution is zero. Since the statistical curvature $\gamma^2(\theta)$ is the first term in the exponential rate expansion for $\hat{\theta}_n$ that differs from the Bahadur bound expansion, it follows from inequality (2.1) that for a distribution with small statistical curvature, $\hat{\theta}_n$ tends to θ at a fast rate (nearly fourth order contact at $\varepsilon = 0$). A large statistical curvature indicates that the corresponding m.l.e. $\hat{\theta}_n$ tends to θ at a slow rate. For natural exponential families, the statistical curvatures are zero. Hence, our results substantiate Efron's (1975, page 1189) basic assertions that families with small curvature enjoy the good properties of exponential families whereas large curvature indicates a breakdown of these properties.

When θ is a location parameter, the inequality (2.10) indicates a preference for the m.l.e. $\hat{\theta}_n$ among all first-order efficient translation invariant estimators. We refer to the estimators which satisfy (2.10) as second-order efficient estimators in the Bahadur sense. Loosely speaking, if the underlying distribution has a small statistical curvature, then the m.l.e. $\hat{\theta}_n$ has a rate of convergence close to the optimal rate (i.e., close to fourth-order contact with the Bahadur bound).

3. Proofs of main results. Let $\ell(x \mid \theta) = \log f(x \mid \theta)$ and $\ell^{(i)}(x \mid \theta) = (\partial/\partial\theta)^i \ell(x \mid \theta)$ and $f^{(i)}(x \mid \theta) = (\partial/\partial\theta)^i f(x \mid \theta)$ for all $\theta \in \Theta$. For each n and $s = (x_1, \cdots, x_n)$, let $\ell_n(s \mid \theta) = \log \prod_{i=1}^n f(x_i \mid \theta)$, $\ell_n^{(1)}(s \mid \theta)$ be a continuous function of θ, $\hat{\theta}_n(s)$ be the m.l.e. which satisfies

$\ell_n^{(1)}(s \mid \theta) = 0$, $\underline{\theta}_n(s) = \inf\{\theta : \ell_n^{(1)}(s \mid \theta) = 0 \text{ and } \theta \in \Theta\}$ and $\bar{\theta}(s) = \sup\{\theta : \ell_n^{(1)}(s \mid \theta) = 0 \text{ and } \theta \in \Theta\}$. The quantities $\underline{\theta}_n(s)$ and $\bar{\theta}_n(s)$ are the smallest and the largest roots of $\ell_n^{(1)}(s \mid \theta) = 0$ respectively. It follows that $\underline{\theta}_n(s) \le \hat{\theta}_n(\mathrm{s}) \le \bar{\theta}_n(s)$, and the following inequalities hold:

$$P\{\underline{\theta}_n(s) \ge \theta + \varepsilon\} \le P\{\ell_n^{(1)}(s \mid \theta + \varepsilon) \ge 0\} \le P\{\bar{\theta}_n(s) \ge \theta + \varepsilon\}, \tag{3.1}$$

$$P\{\bar{\theta}_n(s) \le \theta - \varepsilon\} \le P\{\ell_n^{(1)}(s \mid \theta - \varepsilon) \le 0\} \le P\{\underline{\theta}_n(s) \le \theta - \varepsilon\}. \tag{3.2}$$

To complete the proofs of results in Section 2 we require the following conditions and lemmas.

CONDITION 1. For every $\theta \in \Theta$, $K(\theta', \theta)$ is a locally convex function for θ' in a δ-neighbourhood $N(\theta, \delta)$.

CONDITION 2. For every $\theta \in \Theta$, there exists a neighbourhood $N(\theta, \delta)$ and measurable functions $A_i(x, \theta)$ such that $E_\theta A_i(x, \theta) < \infty$ and the Lipschitz conditions $| \ell^{(i)}(x \mid \theta'') - \ell^{(i)}(x \mid \theta') | \le A_i(x, \theta) \mid \theta'' - \theta' \mid$ hold for all $i = 1, \cdots, 4$ and all $\theta', \theta'' \in N(\theta, \delta)$.

CONDITION 3. For each θ, there exist two constants $u = u(\theta) > 0$ and $v = v(\theta) > 0$ such that $P\{\ell^{(1)}(x \mid \theta + \varepsilon) > 0\} > 0$ and $P\{\ell^{(1)}(x \mid \theta - \varepsilon) < 0\} > 0$ for all ε, $0 < \varepsilon < u$, and the moment generating function $\phi(t, \theta, \varepsilon) = E_\theta[\exp\{t\ell^{(1)}(x \mid \theta + \varepsilon)\}]$ is finite for all $(t, \varepsilon) \in [-v, v] \times [-u, u]$.

CONDITION 4. The partial derivatives $(\partial / \partial t)^i(\partial / \partial\varepsilon)^j\phi(t, \theta, \varepsilon)$, $i, j = 1, \cdots, 4$ exist and are jointly continuous in t and ε for $(t, \varepsilon) \in [-v, v] \times [-u, u]$.

CONDITION 5. For each n and s, the m.l.e. $\hat{\theta}_n(s)$ is the unique solution of $\ell_n^{(1)}(s \mid \theta) = 0$.

To simplify the notation, we write $\ell^{(i)}(x \mid \theta)$, $\ell^{(i)}(x \mid \theta + \varepsilon)$, $I(\theta)$ and $E_\theta[\cdot]$ as $\ell^{(i)}$, $\ell^{(i)}(\varepsilon)$, I, and $E[\cdot]$ respectively.

LEMMA 3.1. *Under Condition 2, we have*

(i) $E\ell^{(1)}(\varepsilon) = \varepsilon E\ell^{(2)} + \frac{\varepsilon^2}{2!} E\ell^{(3)} + \frac{\varepsilon^3}{3!} E\ell^{(4)} + o(\varepsilon^3)$,

(ii) $E[\ell^{(1)}(\varepsilon)]^2 = E(\ell^{(1)})^2 + 2\varepsilon E\ell^{(1)}\ell^{(2)} + \varepsilon^2[E(\ell^{(2)})^2 + E\ell^{(1)}\ell^{(3)}] + o(\varepsilon^2)$,

(iii) $E[\ell^{(1)}(\varepsilon)]^3 = E(\ell^{(1)})^3 + 3\varepsilon E(\ell^{(1)})^2\ell^{(2)} + 3\varepsilon^2[E\ell^{(1)}(\ell^{(2)})^2 + \frac{1}{2} E(\ell^{(1)})^2\ell^{(3)}] + o(\varepsilon^2)$,

(iv) $E[\ell^{(1)}(\varepsilon)]^4 = E(\ell^{(1)})^4 + o(1)$.

PROOF. The results follow immediately from the Taylor expansion, Lebesgue's dominated integration theorem, and integration term by term. □

LEMMA 3.2. *Under Conditions* 3 *and* 4 *there exists a unique single-valued function* $\tau_\theta(\varepsilon)$ *defined on* $0 < \varepsilon < u$ *such that*

$$\phi^{(1)}(\tau_\theta(\varepsilon), \theta, \varepsilon) = 0, \quad \rho_1 = \phi(\tau_\theta(\varepsilon), \theta, \varepsilon), \quad and \quad \tau_\theta(\varepsilon) = \varepsilon + A\varepsilon^2 + B\varepsilon^3 + o(\varepsilon),$$

where $A = -E\ell^{(1)}\ell^{(2)}/2I$, *and*

$$B = -\frac{1}{I}\left[\frac{1}{3!} E\ell^{(4)} - \frac{1}{I}(E\ell^{(1)}\ell^{(2)})^2 + E(\ell^{(2)})^2 + E\ell^{(1)}\ell^{(3)} + \frac{3}{2} E(\ell^{(1)})^2\ell^{(2)} + \frac{1}{2I} E\ell^{(1)}\ell^{(2)}E(\ell^{(1)})^3 + \frac{1}{3!} E(\ell^{(1)})^4\right].$$

PROOF. Hoeffding (1965) proved that there exists a unique single-valued function $\tau_\theta(\varepsilon)$ which is a solution of equation $\phi^{(1)}(t, \theta, \varepsilon) = 0$, and

$$\rho_1 = \inf_{t \ge 0} \phi(t, \theta, \varepsilon) = \phi(\tau_\theta(\varepsilon), \theta, \varepsilon).$$

The first order approximation for $\tau_\theta(\varepsilon)$ when ε is near zero was shown by Fu (1973) to be given by $\tau_\theta(\varepsilon) = \varepsilon + o(\varepsilon)$. Further, let $\tau_\theta(\varepsilon) = \varepsilon + A\varepsilon^2 + B\varepsilon^3 + o(\varepsilon^3)$. Since $\tau_\theta(\varepsilon)$ satisfies the equation $\ell^{(1)}(t, \theta, \varepsilon) = 0$, it follows that

$$\text{(3.3)} \qquad \int_{-\infty}^{\infty} \ell^{(1)}(x \mid \theta + \varepsilon) \exp\{[\varepsilon + A\varepsilon^2 + B\varepsilon^3 + o(\varepsilon^3)]\ell^{(1)}(x \mid \theta + \varepsilon)\} f(x \mid \theta)\, dx \equiv 0.$$

We expand the exponential part of integrand into a series (four terms are sufficient) and integrate term by term. Applying Lemma 3.1 to the above series gives a power series in ε. The four-term expansion of $\tau_\theta(\varepsilon)$ follows directly from the fact that all coefficients of expansion (3.3) have to be zero. □

Let $k_i(\theta, \varepsilon)$, $i = 1, 2, 3$ and 4, be the first four cumulants of the random variable $\ell^{(1)}(x \mid \theta + \varepsilon)$ under P_θ.

LEMMA 3.3. *Under Condition* 2 *we have*

(i) $k_1(\theta, \varepsilon) = \varepsilon E\ell^{(2)} + \frac{\varepsilon^2}{2!} E\ell^{(3)} + \frac{\varepsilon^3}{3!} E\ell^{(4)} + o(\varepsilon^3)$,

(ii) $k_2(\theta, \varepsilon) = E(\ell^{(1)})^2 + 2\varepsilon E\ell^{(1)}\ell^{(2)} + \varepsilon^2[E(\ell^{(2)})^2 - (E\ell^{(2)})^2 + E\ell^{(1)}\ell^{(3)}] + o(\varepsilon^2)$,

(iii) $k_3(\theta, \varepsilon) = E(\ell^{(1)})^3 + 3\varepsilon[E(\ell^{(1)})^2\ell^{(2)} - E(\ell^{(1)})^2 E\ell^{(2)}] + o(\varepsilon)$,

(iv) $k_4(\theta, \varepsilon) = E(\ell^{(1)})^4 - 3(E(\ell^{(1)})^2)^2 + o(1)$.

The proof of this lemma follows from the definition of cumulants and from Lemma 3.1.

THEOREM 3.1. *For any consistent estimator T_n the inequality* (2.1) *holds. If Conditions* 1 *and* 2 *are satisfied, the Bahadur bound $B(\theta, \varepsilon)$ has an expansion at $\varepsilon = 0$ given by* (2.5).

PROOF. The proof of (2.1) was given by Fu (1971). It follows from Conditions 1 and 2 that, for each θ and ε, Kullback-Leibler information $K(\theta', \theta)$ is a continuous and locally convex function for $\theta' \in N(\theta, \delta)$, and

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

Hence, the local expansion (2.5) is a direct result of the Taylor expansion of $K(\theta \pm \varepsilon, \theta)$. □

THEOREM 3.2. (i) *Under Condition* 3 *the following inequalities hold*:

$$\text{(3.4)} \qquad \limsup_{n\to\infty} \frac{1}{n} \log P\{\underline{\theta}_n(s) \ge \theta + \varepsilon\} \le \log \rho_1 \le \liminf_{n\to\infty} \frac{1}{n} \log P\{\bar{\theta}_n(s) \ge \theta + \varepsilon\},$$

$$\text{(3.5)} \qquad \limsup_{n\to\infty} \frac{1}{n} \log P\{\bar{\theta}_n(s) \le \theta - \varepsilon\} \le \log \rho_2 \le \liminf_{n\to\infty} \frac{1}{n} \log P\{\underline{\theta}_n(s) \le \theta - \varepsilon\},$$

where

$$\rho_1 = \inf_{t \ge 0} \phi(t, \theta, \varepsilon) \quad \textit{and} \quad \rho_2 = \inf_{t \le 0} \phi(t, \theta, -\varepsilon).$$

(ii) *Under Conditions* 3 *and* 5 *the* m.l.e. $\hat{\theta}_n(s)$ *has an exponential rate given by*

$$\lim_{n\to\infty} \frac{1}{n} \log P(|\hat{\theta}_n(s) - \theta| \geq \varepsilon) = -\beta(\hat{\theta}, \theta, \varepsilon), \tag{3.6}$$

where

$$\beta(\hat{\theta}, \theta, \varepsilon) = -\log \rho \quad \textit{and} \quad \rho = \max(\rho_1, \rho_2).$$

(iii) *Under Conditions* 2 *to* 5 *the exponential rate* $\beta(\hat{\theta}, \theta, \varepsilon)$ *for the* m.l.e. $\hat{\theta}_n(s)$ *has a local expansion for* ε *near zero given by* (2.7).

PROOF. Result (i) is a direct consequence of inequalities (3.1) and (3.2), and the Bernstein-Chernoff-Bahadur Theorem (see Chernoff, 1952, or Bahadur, 1971). It follows from Condition 5 that the inequalities (3.1) and (3.2) reduce to the equalities

$$P\{\hat{\theta}_n(s) \geq \theta + \varepsilon\} = P\{\ell_n^{(1)}(s \mid \theta + \varepsilon) \geq 0\}, \quad \text{and} \tag{3.7}$$

$$P\{\hat{\theta}_n(s) \leq \theta - \varepsilon\} = P\{\ell_n^{(1)}(s \mid \theta - \varepsilon) \leq 0\}. \tag{3.8}$$

Result (ii) is derived from equalities (3.7) and (3.8), and the Bernstein-Chernoff-Bahadur Theorem. Using Lemma 3.2 it follows that

$$\log \rho_1(\varepsilon) = \log \phi(\tau_\theta(\varepsilon), \theta, \varepsilon) = \sum_{i=1}^{4} \frac{1}{i!} k_i(\theta, \varepsilon)\tau_\theta^i(\varepsilon) + o(\varepsilon^4). \tag{3.9}$$

Using Lemmas 3.2 and 3.3, we have a four-term Taylor expansion (3.9) given by

$$\begin{aligned} \log \rho_1 = -\Bigg\{ & \frac{\varepsilon^2}{2!} I + \frac{\varepsilon^3}{3!} (3\mu_{110} - \mu_{300}) \\ & + \frac{\varepsilon^4}{4!} [(2\mu_{400} - 6\mu_{210} + 4\mu_{101} + 3\mu_{020}) - 3I^2\gamma^2(\theta)] + o(\varepsilon^4) \Bigg\}, \end{aligned} \tag{3.10}$$

where $\gamma^2(\theta)$ is the statistical curvature (2.8). The quantity $\log \rho_2$ has a four-term Taylor expansion similar to (3.10) with ε replaced by $-\varepsilon$. Result (iii) follows directly from $\beta(\hat{\theta}, \theta, \varepsilon) = -\log \rho$, $\rho = \max(\rho_1, \rho_2)$, and from (3.10). □

THEOREM 3.3. *Under Conditions* 1 *through* 5 *we have that* (i) *the result* (2.9) *holds, and* (ii) *if* θ *is location parameter with the underlying density being log-concave and symmetric, then the inequality* (2.10) *holds for any translation invariant consistent estimator, and equality holds for the* m.l.e. $\hat{\theta}_n(s)$.

PROOF. The density function being log-concave implies Condition 5. Result (i) follows directly from Theorems 3.1 and 3.2. Result (ii) is an immediate consequence of result (i) and a result of Sievers (1978, page 612). □

REMARK 1. Conditions 1 through 5 can be easily verified. To verify Condition 1, one needs to check that for every θ and ε there exists a δ-neighbourhood $N(\theta, \delta)$ such that the Kullback-Leibler information $K(\cdot, \theta)$ has a continuous second derivative and $K^{(2)}(\theta, \theta) > 0$. Condition 5, that the m.l.e. $\hat{\theta}_n$ be a unique solution of the equation $\ell_n^{(1)}(s \mid \theta) = 0$, is essential in our proofs. It is usually satisfied by many distributions, particularly by those whose density functions are log-concave. For example: normal distribution with known variance, logistic distribution, Γ-distribution, and other members of the Koopmans-Darmois class of probability distributions. If the equation $\ell_n^{(1)}(s \mid \theta) = 0$ has multiple roots, or the number of roots depends on the sample size, then our proofs break down. In the case of the Cauchy distributions, for example, the problem remains open.

REMARK 2. An estimator may have an optimal exponential rate but may be inefficient in the exact rate. This substantiates a remark given by LeCam (see Efron, 1975, page 1224) that even for certain exponential families, the m.l.e. is difficult to compute and is inferior (in the sense of expected square deviation) to some alternatives. The phenomenon applies not only to the expected square deviation, but also to any loss function which is a monotonic increasing function of $|T_n - \theta|$. The existence of the phenomenon is due to the fact that the m.l.e. $\hat{\theta}_n$ does not always minimize the tail probability $\alpha(T_n, \theta, \varepsilon)$. The consistent estimator $\tilde{\theta}_n(s)$ which minimizes $\alpha(T_n, \theta, \varepsilon)$, or maximizes $1 - \alpha(T_n, \theta, \varepsilon)$ will be called maximum probability estimator (m.p.e.). Since the m.p.e. $\tilde{\theta}_n(s)$ has by definition the fastest rate of convergence, it is asymptotically efficient. For a given estimation problem, the m.p.e. may depend on ε, and be difficult to compute. In some cases, it may not even exist. With a suitable additional constraint, however, it usually does exist. For example, a m.p.e. may exist in the class of translation invariant estimators. Specific examples are given in Section 5.

REMARK 3. It is reasonable to expect results proved for the location parameter to hold for sufficiently smooth non-location parameter families, though no mathematical justification is available. It is of great interest from both a practical and a theoretical point of view that a more general theorem in this direction be extended to a wider class of distributions.

4. Exponential rate and Fisher-Rao-Efron second-order efficiency. For any consistent estimator T_n, Fisher in his fundamental 1925 paper on estimation theory stated that

$$\lim_{n\to\infty}[E_\theta\{\ell_n^{(1)}(s \mid \theta)\}^2 - E_\theta\{\ell^{(1)}(T_n(s) \mid \theta)\}^2] \geq I(\theta)\gamma^2(\theta), \tag{4.1}$$

where $\ell^{(1)}(T_n(s) \mid \theta)$ is the first derivative of the log-likelihood of estimator T_n, and the equality holds for the m.l.e. $\hat{\theta}_n$. Fisher believed the Fisher information, $I(\theta)$, to be a perfect measure of the amount of information available to a statistician, and also that asymptotically the m.l.e. $\hat{\theta}_n$ extracts all but $I(\theta)\gamma^2(\theta)$ of the information in the sample. This optimal property of the m.l.e. is said to be "second-order efficiency" among the class of first order efficient estimators T_n, those which satisfy the weaker condition

$$\lim_{n\to\infty} E_\theta\{\ell_n^{(1)}(s \mid \theta)\}^2 / E_\theta\{\ell^{(1)}(T_n(s) \mid \theta)\}^2 = 1.$$

Rao (1963) gave a similar result for square error loss which can be stated as:

$$\operatorname{Var}(T_n) \geq \frac{1}{nI(\theta)} + \frac{1}{n^2 I(\theta)}\left\{\gamma^2(\theta) + \frac{\Gamma^2}{I(\theta)}\right\} + O(n^{-3}), \tag{4.2}$$

for any consistent estimator T_n, where Γ^2 is the ordinary curvature and equality holds for the m.l.e. $\hat{\theta}_n$. Fisher and Rao tried to prove the result (4.1) for multinomial families. Efron (1975) gave a counterexample, showing that the result (4.1) is not true for multinomial families. In that paper Efron introduces the statistical curvature and also gives a proof of results (4.1) and (4.2) for curved exponential families. His proof relies on an ingenious geometrical interpretation of statistical curvature and a powerful large deviation result similar to the Bernstein-Chernoff-Bahadur theorem used in Section 3.

Our results (2.1), (2.9) and (2.10), obtained from the exponential rate of convergence approach using probability of large deviation theory, are parallel to their results. Before commenting further on this relationship we will show a mathematical connection between the rate of convergence and the Fisher-Rao-Efron result (4.2). Since a detailed proof of this is extremely tedious we only outline the important steps of the proof for the situation where θ is a location parameter with a symmetric density function $f(x \mid \theta)$ which satisfies the Conditions 1 to 5 stated in Section 3 and $\Gamma = 0$: Let $|\hat{\theta}_n(s) - \theta| = U_n(s)$, and $\psi(t) = \log$

$\phi(t, \theta, \varepsilon)$. It follows from Theorems 3.2 and 3.3, and the generalized version of Laplace's method, that the distribution of the random variable $U_n(s)$ has an asymptotic expansion (also see Petrov, 1975, page 248) given by

$$P\{U_n(s) \geq u\} = 1 - F(u) = \{\tau(u)\}^{-1}\{2\pi n\psi^{(2)}(\tau)\}^{-1/2}$$
$$\cdot\left\{\exp\left(-n\left[\frac{u^2}{2!}I(\theta) + \frac{u^4}{4!}\{2\mu_{400} - 6\mu_{210} + 4\mu_{101} + 3\mu_{020} - 3I^2(\theta)\gamma^2(\theta)\} + o(u^4)\right]\right)\right\} \tag{4.3}$$
$$\cdot\{1 + o_n(1)\},$$

where $\tau(u)$ is as defined in Lemma 3.2. If the m.l.e. $\hat{\theta}_n(s)$ is an unbiased estimator for θ, then $\mathrm{Var}(\hat{\theta}_n) = E_\theta U_n^2(s)$. Integrating by parts, it follows that with $1 - F(u)$ given by (4.3),

$$\mathrm{Var}(\hat{\theta}_n) = 2\int_0^\infty u\{1 - F(u)\}\,du. \tag{4.4}$$

We first expand the integrand, not including the term $\exp\{-nu^2I(\theta)/2\}$, as a series in u. We then use the transformation $u^2 = v$ and integrate term by term. Applying the generalized Laplace method to evaluate each term, the first two terms of the expansion of $\mathrm{Var}(\hat{\theta}_n)$ come out as (4.2).

Fisher's result (4.1) can perhaps be obtained by this same method. However, we have been unable to show mathematically that this is the case. It is noteworthy that the Kullback-Leibler information $K(\theta', \theta)$ is directly associated with the predominant (exponential) term of the asymptotic distribution of the consistent estimator, and that the Fisher information $I(\theta)$ is associated only with its second order derivative. Contrary to Fisher's claim, we believe that for higher order efficiency in large sample estimation the Kullback-Leibler information is more directly related to the performance of consistent estimators than the Fisher information $I(\theta)$ is. It is well-known (Rao, 1963, and Efron, 1975) that the central limit theorem in general is not capable of supporting such fine results (4.1), (4.2), and (2.9). From our results, it seems that the probability of large deviation theory does offer more in this direction.

5. Examples. The results in Section 3 are true for many sufficiently smooth families of distributions. To illustrate these results, we consider the examples of normal, logistic, exponential, double exponential and uniform distributions, which are of general interest in both statistical theory and practice. Among them, the normal, logistic, and exponential distributions satisfy the regularity conditions stated in Section 3. The double exponential and uniform distributions do not satisfy those conditions.

EXAMPLE 1. Let $\{X_i\}_{i=1}^n$ be a sequence of i.i.d. normal random variables with unknown mean θ and variance σ^2. It has Fisher information $I(\theta) = 1/\sigma^2$ and Kullback-Liebler information $K(\theta', \theta) = (\theta' - \theta)^2/2\sigma^2$. The sample mean $\hat{\theta}_n = \bar{X}_n$ is a minimum sufficient statistic for θ and is also m.l.e. One can show that $\tilde{\theta}_n = \bar{X}_n$ is the m.p.e. among all translation invariant estimators and is also the m.p.e. among all scale invariant estimators. In this case the m.l.e. coincides with the m.p.e. and therefore it enjoys the optimal properties of the m.p.e. The exponential rate of the m.l.e., $\beta(\hat{\theta}, \theta, \varepsilon)$, achieves the Bahadur bound, i.e.,

$$\beta(\hat{\theta}, \theta, \varepsilon) = B(\theta, \varepsilon) = \varepsilon^2/2\sigma^2 \quad \text{for all} \quad \varepsilon > 0 \quad \text{and} \quad \theta \in (-\infty, \infty). \tag{5.1}$$

EXAMPLE 2. Let $\{X_i\}_{i=1}^n$ be a sequence of i.i.d. logistic random variables having density functions

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

The density function is log-concave and satisfies all the regularity conditions. It has Fisher information $I(\theta) = 1/3$, Kullback-Leibler information

$$K(\theta + \varepsilon, \theta) = \varepsilon + 2 \int_0^\infty \{\log(1 + e^{-\varepsilon}u) - \log(1 + u)\} \frac{du}{(1 + u)^2} \tag{5.2}$$

and a small statistical curvature $\gamma^2(\theta) = 0.2$. The m.l.e. $\hat{\theta}_n$ is the solution of the equation $\ell_n^{(1)}(s) = n - 2 \sum_{i=1}^{n} \{1 + \exp(X_i - \theta)\}^{-1} = 0$. By the result (3.6) it has an exponential rate given by

$$\beta(\hat{\theta}, \theta, \varepsilon) = -\log\left\{\inf_{t\geq 0} \exp(-t/2) \int_{-\infty}^{\infty} (\exp[t/\{1 + \exp(x - \theta - \varepsilon)\}])f(x \mid \theta)\, dx\right\}. \tag{5.3}$$

Numerical computations of (5.2) and (5.3) show that

$$\lim_{\varepsilon\to 0}\varepsilon^{-4}[B(\theta, \varepsilon) - \beta(\theta, \theta, \varepsilon)] = 0.0028 = \tfrac{1}{8} I^2(\theta)\gamma^2(\theta).$$

The m.l.e. is second-order efficient estimator in the Bahadur sense among all translation invariant estimators. The smallness of the difference between $B(\hat{\theta}, \varepsilon)$ and $\beta(\hat{\theta}, \theta, \varepsilon)$ when θ is near zero indicates that the exponential rate of $\hat{\theta}_n$ is near the optimal rate (Bahadur bound).

EXAMPLE 3. Let $\{X_i\}_{i=1}^n$ be a sequence of i.i.d. exponential random variables with scale parameter θ and density function $f(x|\theta) = \theta e^{-\theta x}$ for $x \in (0, \infty)$ and $\theta > 0$. The Kullback-Leibler information is $K(\theta', \theta) = \log \theta'/\theta - (\theta' - \theta)/\theta'$. The m.l.e. $\hat{\theta}_n = \bar{X}_n$ has an exponential rate given by

$$\beta(\hat{\theta}, \theta, \varepsilon) = -\log \inf_{t\geq 0}\phi(t, \theta, \varepsilon) = \log(\theta + \varepsilon)/\theta - \varepsilon/(\theta + \varepsilon), \tag{5.4}$$

which achieves the Bahadur bound. Again the m.l.e. coincides with the scale invariant m.p.e. $\tilde{\theta}_n$.

EXAMPLE 4. Let $\{X_i\}_{i=1}^n$ be a sequence of i.i.d. double exponential random variables with location parameter θ and density function

$$f(x \mid \theta) = \tfrac{1}{2} \exp(-| x - \theta |), \qquad x \in (-\infty, \infty), \qquad \theta \in (-\infty, \infty).$$

The Kullback-Leibler information $K(\theta', \theta)$ is a locally convex function in θ'. The Bahadur bound has a Taylor expansion given by

$$B(\theta, \varepsilon) = \tfrac{1}{2} \varepsilon^2 - \tfrac{1}{6} \varepsilon^3 + \tfrac{1}{24} \varepsilon^4 + o(\varepsilon^4). \tag{5.5}$$

The density function $f(x|\theta)$ does not satisfy our regularity conditions. However, its m.l.e. $\hat{\theta}_n$ = median $(X_1, \cdots, X_n) = X_{(1/2)}$ does have an exponential rate given by

$$\beta(\hat{\theta}, \theta, \varepsilon) = \tfrac{1}{2} \varepsilon^2 - \tfrac{1}{2} \varepsilon^3 + \tfrac{13}{24} \varepsilon^4 + o(\varepsilon^4). \tag{5.6}$$

It follows from (5.5) and (5.6) that the m.l.e. $\hat{\theta}_n = X_{(1/2)}$ is a first-order efficient estimator in the Bahadur sense (2.3). It is neither a third order contact estimator nor a second-order efficient estimator in Bahadur sense (2.10). Furthermore, the parameter θ is a location parameter and $f(x \mid \theta)$ is symmetric and $3\mu_{110} - \mu_{300} = 0$. The irregular behaviour of $B(\theta, \varepsilon)$ and $\beta(\hat{\theta}, \theta, \varepsilon)$ is due to the fact that the density function is non-differentiable at $\theta = x$.

EXAMPLE 5. Let $\{X_i\}_{i=1}^n$ be a sequence of i.i.d. uniform random variables with density function $f(x \mid \theta) = 1/\theta$ for all $x \in [0, \theta]$ where θ is a scale parameter. The m.l.e. $\hat{\theta}_n = \max(X_1, \cdots, X_n) = X_{[n]}$ is scale invariant but does not coincide with the scale invariant m.p.e. $\hat{\theta}_n = [(n + 1)/n]X_{[n]}$. It follows that both the m.l.e. and m.p.e. have the same exponential rate given by

$$\beta(\hat{\theta}, \theta, \varepsilon) = \beta(\tilde{\theta}, \theta, \varepsilon) = \log(\theta - \varepsilon)/\theta, \tag{5.7}$$

for all $\theta > \varepsilon > 0$. Hence, the exponential rate would not distinguish between these two estimators. However, by comparing the exact rates, we have

$$\lim_{n\to\infty}\alpha(\hat{\theta}_n, \theta, \varepsilon)/\alpha(\tilde{\theta}_n, \theta, \varepsilon) = e^{-1} \tag{5.8}$$

for all $\theta > \varepsilon > 0$, which is less than one and independent of ε. Clearly the scale invariant m.p.e. $\tilde{\theta}_n$ is superior to the m.l.e. $\hat{\theta}_n$. The scale invariant m.p.e. is asymptotically efficient among all scale invariant estimators with respect to any loss function L which is a non-decreasing function of $|T_n - \theta|$. For example, if $L(|T_n - \theta|) = (T_n - \theta)^2$, then

$$E(X_{[n]} - \theta)^2 - E\left(\frac{n+1}{n}X_{[n]} - \theta\right)^2 = \frac{\theta^2}{(n+1)^2}\left\{1 + O\left(\frac{1}{n}\right)\right\} > 0. \tag{5.9}$$

6. Acknowledgment. A part of this research was carried out during the author's sabbatical leave at the Department of Statistics, University of Wisconsin. The author is grateful to Professor R. R. Bahadur, Professor R. A. Johnson, and Professor J. Klotz for their careful reading of the manuscript and useful comments, resulting in considerable improvement of the text. The author also wishes to express his gratitude to the editor and referees for their suggestions.

REFERENCES

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

BASU, D. (1956). On the concept of asymptotic efficiency. *Sankhyá* **17** 193–196.

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.

DANIELS, H. E. (1961). The asymptotic efficiency of a maximum likelihood estimator. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 151–164.

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

FISHER, R. A. (1925). Theory of statistical estimation. *Proc. Cambridge Philos. Soc.* **122** 700–725.

FU, J. C. (1971). Some large sample methods in statistical point estimation. Ph.D. Thesis. The Johns Hopkins University.

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.

HOEFFDING, W. (1965). On probability of large deviations. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 203–219.

KESTER, A. (1981). Large deviation optimality of MLE's in exponential families. Technical Report 160, Free Univ. of Amsterdam. (Unpublished).

PETROV, V. V. (1975). *Sums of Independent Random Variables.* Page 248. Springer, Berlin.

RAO, C. R. (1963). Criteria of estimation in large samples. *Sankhyā* **25** 189–206.

SIEVERS, G. (1978). Estimates of Location: A large deviations comparison. *Ann. Statist.* **6** 610–618.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MANITOBA
WINNIPEG, MANITOBA
CANADA, R3T 2N2