The Annals of Statistics
1986, Vol. 14, No. 2, 648–664

LARGE DEVIATIONS OF ESTIMATORS[1]

By A. D. M. Kester and W. C. M. Kallenberg

University of Limburg, Maastricht and Twente University of Technology

The performance of a sequence of estimators $\{T_n\}$ of $g(\theta)$ can be measured by its inaccuracy rate $-\lim\inf_{n\to\infty} n^{-1}\log \mathbb{P}_\theta(\|T_n - g(\theta)\| > \varepsilon)$. For *fixed* $\varepsilon > 0$ optimality of consistent estimators wrt the inaccuracy rate is investigated. It is shown that for exponential families in standard representation with a convex parameter space the maximum likelihood estimator is optimal. If the parameter space is not convex, which occurs for instance in curved exponential families, in general no optimal estimator exists.

For the location problem the inaccuracy rate of M-estimators is established. If the underlying density is sufficiently smooth an optimal M-estimator is obtained within the class of translation equivariant estimators.

Tail-behaviour of location estimators is studied. A connection is made between gross error and inaccuracy rate optimality.

1. Introduction. Let $\mathscr{X}$ be a set of points x and $\mathscr{B}$ a σ-field of subsets of $\mathscr{X}$. The parameter space Θ is an index set of points θ and for each $\theta \in \Theta$, P_θ is a probability measure on $\mathscr{B}$. Let $X_1, X_2, \dots$ be a sequence of i.i.d. random variables, each defined on $\mathscr{X}$. The distribution of $S = (X_1, X_2, \dots)$ is denoted by $\mathbb{P}_\theta$, $\theta \in \Theta$. Let g be a mapping of the abstract space Θ into $\mathbb{R}^d$ and let $\{T_n\}$ denote a sequence of estimators of $g(\theta)$, where T_n is based on n observations. Note that T_n takes values in $g(\Theta)$ only. The performance of $\{T_n\}$ is measured by its *inaccuracy rate*

$$e(\varepsilon, \theta, \{T_n\}) = -\liminf_{n\to\infty} n^{-1}\log \mathbb{P}_\theta(\|T_n - g(\theta)\| > \varepsilon); \tag{1.1}$$

the larger the inaccuracy rate the better the estimator. Due to the fact that the large deviation probabilities involved are hard to handle inaccuracy rates have been discussed mainly for $\varepsilon \to 0$. See for example Bahadur, Gupta and Zabell (1980) and Fu (1982 and references therein). In this paper however, the inaccuracy rate is investigated for *fixed* $\varepsilon > 0$.

Two main themes are considered:

(i) optimality of consistent estimators;
(ii) the inaccuracy rate of M-estimators for the location problem.

To investigate optimality of a sequence of consistent estimators the inaccuracy rate of the sequence is compared with an upper bound of (1.1). As usual in large

Received December 1984; revised June 1985.

[1] This research was done while the authors were affiliated with the Free University at Amsterdam.

AMS 1980 *subject classifications.* Primary 62F10; secondary 60F10.

Key words and phrases. Large deviations, inaccuracy rate, exponential convexity, maximum likelihood estimator, M-estimator, translation equivariance, tail-behaviour.

deviation theory the upper bound is obtained essentially by application of the Neyman–Pearson lemma; cf. Bahadur et al. (1980).

PROPOSITION 1.1 (Bahadur). *If $\{T_n\}$ is a consistent estimator of $g(\theta)$ for each $\theta \in \Theta$, then*

$$e(\varepsilon, \theta, \{T_n\}) \le b(\varepsilon, \theta), \tag{1.2}$$

where $b(\varepsilon, \theta) = \inf\{K(\eta, \theta)\colon \eta \in \Theta, \|g(\eta) - g(\theta)\| > \varepsilon\}$ and K is the Kullback–Leibler information

$$K(\eta, \theta) = \begin{cases} E_\eta \log(dP_\eta/dP_\theta) & \text{when } P_\eta \ll P_\theta, \\ \infty & \text{otherwise}. \end{cases} \tag{1.3}$$

In view of Proposition 1.1 a sequence of estimators $\{T_n\}$ is called *optimal* wrt the inaccuracy rate (IR-optimal) at θ_0 for $\varepsilon > 0$ if $\{T_n\}$ is a consistent estimator of $g(\theta)$ for each $\theta \in \Theta$ and

$$-\lim_{n\to\infty} n^{-1} \log \mathbb{P}_{\theta_0}(\|T_n - g(\theta_0)\| > \varepsilon) = b(\varepsilon, \theta_0). \tag{1.4}$$

Note that IR-optimality at θ_0 depends on Θ in two ways: Bahadur's bound $b(\varepsilon, \theta_0)$ depends on Θ and $\{T_n\}$ has to be a consistent estimator of $g(\theta)$ *for each* $\theta \in \Theta$; cf. Proposition 1.1. The important role of Kullback–Leibler information in large deviation theory is apparent from the following simple but useful proposition of Bahadur (1980, 1983 Section 2), which states that IR-optimality of $\{K(T_n, \theta_0)\}$ as an estimator of $K(\theta, \theta_0)$ yields IR-optimality of $\{g(T_n)\}$ as an estimator of $g(\theta)$.

PROPOSITION 1.2 (Bahadur). *If g is continuous and $\{T_n\}$ is a consistent estimator of θ for each $\theta \in \Theta$ such that for each $b < b_0 = b_0(\theta_0)$*

$$-\limsup_{n\to\infty} n^{-1} \log \mathbb{P}_{\theta_0}(K(T_n, \theta_0) \ge b) \ge b, \tag{1.5}$$

then $\{g(T_n)\}$ is an inaccuracy rate optimal estimator of $g(\theta)$ at θ_0 for each $\varepsilon > 0$ with $b(\varepsilon, \theta_0) < b_0$.

It is well-known that the likelihood ratio test is often an optimal test in a large deviation context. One might guess that the maximum likelihood estimator (MLE) plays a similar role in large deviation estimation theory. It turns out that for exponential families in standard representation with a convex parameter space MLE's are indeed IR-optimal. *Exponential convexity* (cf. Section 2) is here the key point. As soon as exponential convexity fails, (1.5) cannot hold true for all θ_0 and b which are of interest; cf. Lemma 2.4. This occurs for instance in curved exponential families.

In Section 3 shift families $\{P_\theta\} = \{p(x - \theta)\colon \theta \in \mathbb{R}\}$ are considered. Here p is a Lebesgue density. Only in some exceptional cases shift families are exponentially convex. We may therefore expect that IR-optimal estimators usually do not exist. Sievers (1978) came to the same conclusion, be it apparently on a more empirical basis.

In shift families it is natural to restrict attention to (translation) equivariant estimators. When $p(x-\varepsilon)/p(x+\varepsilon)$ is nondecreasing in x, Sievers (1978) obtained an upper bound for the inaccuracy rate of equivariant estimators by application of the Neyman–Pearson lemma. Remarkably, Sievers' bound can be *higher* than Bahadur's bound. The reason is that Sievers' bound concerns equivariant estimators, which are not necessarily consistent; cf. Example 3.1. However, when p is symmetric and sufficiently regular, Sievers' bound is not larger than Bahadur's; cf. Kester (1985), page 71, and Fu (1985).

For a wider class of shift families we derive in Section 3 an upper bound, which coincides with Sievers' bound when p satisfies his condition. When p is sufficiently smooth and ε is small enough (but fixed), an M-estimator is constructed which attains this bound. This result (Theorem 3.4) provides optimal equivariant estimators even for such densities as the Cauchy density; cf. Example 3.2. In contrast to Sievers, who applies a finite sample result, we employ a typical large deviation approach. It is also shown that in the double exponential family a trimmed mean attains Sievers' bound.

Apart from investigating optimality of equivariant estimators the inaccuracy rate of quite general M-estimators is obtained.

In the last section tail-behaviour of location estimators is discussed, mainly for a fixed sample size. Jurečková (1979) has shown that the sample mean has a certain gross error optimality property when the distribution of the observations has exponentially decreasing tails. Here it is shown that the inaccuracy rate optimal estimator for a fixed error, say ε, converges to the sample mean when $\varepsilon \to \infty$, i.e. when gross errors come in.

2. Exponential families, exponential convexity. Let $\{P_\theta\colon \theta \in \Theta\}$ be a k-parameter exponential family in standard representation given by its densities wrt a σ-finite measure μ on $\mathbb{R}^k$

$$(2.1)\quad dP_\theta(x) = \exp\{\theta' x - \psi(\theta)\}\, d\mu(x), \qquad x \in \mathbb{R}^k,\ \theta \in \Theta \subset \Theta^* \subset \mathbb{R}^k,$$

where Θ is a *subset* of $\Theta^* = \{\theta \in \mathbb{R}^k\colon \int \exp(\theta' x)\, d\mu(x) < \infty\}$ and $\psi(\theta) = \log \int \exp(\theta' x)\, d\mu(x)$, $\theta \in \Theta^*$. Here $\theta' x$ denotes the inner product of θ and x. Assume without loss of generality that the covariance matrix of P_θ is nonsingular for $\theta \in \operatorname{int} \Theta^*$. Define $\Theta_1 = \{\theta \in \Theta^*\colon E_\theta \|X_1\| < \infty\}$, then $\operatorname{int}\Theta^* \subset \Theta_1 \subset \Theta^*$ and the mapping $\lambda\colon \theta \to E_\theta X_1$ is 1-1 on Θ_1; cf. Berk (1972). The likelihood of a sample $X_1, \dots, X_n$ is maximized over Θ^* at the point

$$(2.2)\qquad \hat\theta_n^* = \lambda^{-1}(\overline{X}_n)$$

when $\overline{X}_n \in \Lambda = \{\lambda(\theta)\colon \theta \in \Theta_1\}$. Noting that the Kullback–Leibler information $K(\eta,\theta) = (\eta - \theta)'\lambda(\eta) - \psi(\eta) + \psi(\theta)$ when $\eta \in \Theta_1$ and $\theta \in \Theta^*$, it is seen that maximizing the likelihood over a subset Θ of Θ^* is equivalent to minimizing $K(\lambda^{-1}(\overline{X}_n), \theta) = K(\hat\theta_n^*, \theta)$ over $\theta \in \Theta$ when $\overline{X}_n \in \Lambda$; cf. Efron (1978). If it exists the unique point $\hat\theta(\eta)$ minimizing $K(\eta,\theta)$ over $\theta \in \Theta$ is called the *Kullback–Leibler projection* of η on Θ. Thus when $\overline{X}_n \in \Lambda$ and $\hat\theta$ exists at $\lambda^{-1}(\overline{X}_n)$,

$$(2.3)\qquad \hat\theta_n = \hat\theta(\hat\theta_n^*)$$

is the MLE of θ on Θ. Before stating a lemma which establishes existence of the MLE we introduce the "Kullback–Leibler distance" $K(\theta)$ of the boundary of Θ^* to θ by

(2.4) $K(\theta) = \sup\{a: \{\eta: K(\eta,\theta) \le a\} \subset C_a \subset \operatorname{int}\Theta^*, \text{ where } C_a \text{ is compact}\}.$

LEMMA 2.1. *Let Θ be a relatively closed convex subset of Θ^* and let $\eta \in \operatorname{int}\Theta^*$. If $K(\eta,\theta) < K(\theta)$ for some $\theta \in \Theta$, then the Kullback–Leibler projection $\hat{\theta}(\eta)$ exists; thus $\hat{\theta}_n$ exists when*

$$\overline{X}_n \in \lambda\Big(\bigcup_{\theta\in\Theta}\{\eta: K(\eta,\theta) < K(\theta)\}\Big). \tag{2.5}$$

Moreover, if $K(\eta,\theta) < K(\theta)$ for some $\theta \in \Theta$, then $K(\hat{\theta}(\eta),\theta) \le K(\eta,\theta)$ and hence $\hat{\theta}(\eta) \in \operatorname{int}\Theta^*$.

PROOF. Let $\eta \in \operatorname{int}\Theta^*$ and $\theta \in \Theta$ satisfy $K(\eta,\theta) < K(\theta)$. On the compact set $\Theta \wedge \{\zeta \in \Theta^*: K(\zeta,\theta) \le K(\eta,\theta)\}$ the infimum of $K(\eta,\cdot\,)$ is attained at π, say. Consider $\xi \in \Theta$ satisfying $K(\xi,\theta) > K(\eta,\theta)$ and let $\delta > 0$. Define $\xi_\alpha = \alpha\xi + (1-\alpha)\theta$ and let $0 < \tilde{\alpha} < 1$ be such that $K(\eta,\xi_{\tilde{\alpha}}) < K(\eta,\xi) + \delta$. Note that $\{\xi_\alpha: 0 \le \alpha \le \tilde{\alpha}\} \subset \operatorname{int}\Theta^*$. Now $K(\eta,\cdot\,)$ attains its infimum on the compact set $\{\xi_\alpha: 0 \le \alpha \le \tilde{\alpha}\}$ at ξ_{α^*}, say. Since $K(\eta, t\xi_{\alpha^*} + (1-t)\theta)$, $0 \le t \le 1$, is minimal for $t = 1$, its derivative is nonpositive at $t = 1$: $(\xi_{\alpha^*} - \theta)'(\lambda(\xi_{\alpha^*}) - \lambda(\eta)) \le 0$. In combination with $K(\xi_{\alpha^*},\theta) - K(\eta,\theta) = (\xi_{\alpha^*} - \theta)'(\lambda(\xi_{\alpha^*}) - \lambda(\eta)) - K(\eta,\xi_{\alpha^*})$ it follows that $\xi_{\alpha^*} \in \Theta \wedge \{\zeta \in \Theta^*: K(\zeta,\theta) \le K(\eta,\theta)\}$ and hence $K(\eta,\pi) \le K((\eta,\xi_{\alpha^*}) \le K(\eta,\xi_{\tilde{\alpha}}) < K(\eta,\xi) + \delta$. Since $\delta > 0$ was arbitrarily chosen we have $K(\eta,\pi) \le K(\eta,\xi)$, implying that the infimum of $K(\eta,\cdot)$ on Θ is attained at π. Unicity of $\hat{\theta}(\eta) = \pi$ follows from the convexity of Θ and the strict convexity of $K(\eta,\cdot\,)$. □

One of the main results of the paper, optimality of MLE's in convex exponential families, is presented in the following theorem.

THEOREM 2.2. *Suppose Θ is a convex relatively closed subset of* $\operatorname{int}\Theta^*$, *where Θ^* is the full parameter space of an exponential family in standard representation. Let $\hat{g}_n = g(T_n)$ where $T_n = \hat{\theta}_n$ whenever the* MLE *$\hat{\theta}_n$ of θ exists. If g is continuous then $\{\hat{g}_n\}$ is inaccuracy rate optimal at θ for each $\varepsilon > 0$ satisfying*

$$b(\varepsilon,\theta) < K(\theta). \tag{2.6}$$

Before proving Theorem 2.2 we mention the following special case.

COROLLARY 2.3. *Let $\Theta = \Theta^*$ be open. Suppose that*

$$\Big\{x \in \mathbb{R}^k: \sup_{\theta\in\Theta^*}\{\theta'x - \psi(\theta)\} < \infty\Big\} \quad \textit{is open}, \tag{2.7}$$

then $\{\hat{\theta}_n^*\}$ is inaccuracy rate optimal for all $\varepsilon > 0$ and $\theta \in \Theta^*$.

PROOF. Apply Theorem 2.2 with $\Theta = \Theta^*$, $g(\theta) = \theta$ and note that (2.7) implies $K(\theta) = \infty$ for each $\theta \in \Theta^*$; cf. Kourouklis (1984). □

PROOF OF THEOREM 2.2. Since Θ is locally compact and $\Theta \subset \text{int}\,\Theta^*$, it follows by Theorem 3.1 and Corollary 3.3 of Berk (1972) that $\{\hat{\theta}_n\}$ and hence $\{T_n\}$ is a consistent estimator of θ for each $\theta \in \Theta$. Let $\theta \in \Theta$ and $\varepsilon > 0$ satisfy (2.6). For each $b < K(\theta)$ we have

$$\mathbb{P}_\theta(K(T_n, \theta) \geq b) \leq \mathbb{P}_\theta(\lambda^{-1}(\overline{X}_n) \in A, K(T_n, \theta) \geq b) + \mathbb{P}_\theta(\overline{X}_n \notin \lambda(A)) \tag{2.8}$$

with $A = \{\eta \in \Theta^*\colon\ K(\eta, \theta) < b\}$. By Lemma 2.1 $\lambda^{-1}(\overline{X}_n) \in A$ implies that $\hat{\theta}_n$ exists and $K(T_n, \theta) = K(\hat{\theta}_n, \theta) \leq K(\hat{\theta}_n^*, \theta) = K(\lambda^{-1}(\overline{X}_n), \theta) < b$. It follows that the first term in the right-hand side of (2.8) equals zero. The second term equals $\exp\{-nb + o(n)\}$ as $n \to \infty$ by Theorem 6 of Efron and Truax (1968). Application of Proposition 1.2 completes the proof. □

REMARK 2.1. At first sight one might guess that IR-optimality of $\{\hat{\theta}_n\}$ can be proved as follows: Define $g(\theta) = \hat{\theta}(\theta)$, show that g is continuous, prove directly IR-optimality of $\{\hat{\theta}_n^*\}$ in the full parameter space Θ^* and apply Proposition 1.2. As a result one obtains

$$-\lim_{n\to\infty} n^{-1} \log \mathbb{P}_\theta(\|\hat{\theta}_n - \theta\| > \varepsilon) = \inf\{K(\eta, \theta)\colon\ \eta \in \Theta^*, \|\hat{\theta}(\eta) - \theta\| > \varepsilon\},$$

while one has to show

$$-\lim_{n\to\infty} n^{-1} \log \mathbb{P}_\theta(\|\hat{\theta}_n - \theta\| > \varepsilon) = \inf\{K(\eta, \theta)\colon\ \eta \in \Theta, \|\eta - \theta\| > \varepsilon\}.$$

(Bahadur's bound depends on the parameter space; cf. Section 1.)

In general the second infimum is larger than the first one. Under the condition that Θ is a relatively cosed convex subset of Θ^* it is seen by Lemma 2.1 that both infima are equal. This argument is also used in the preceding proof applying the crucial inequality $K(\hat{\theta}_n, \theta) \leq K(\hat{\theta}_n^*, \theta)$. Convexity is the key point as is further elaborated in a wider context in the rest of this section.

In general one cannot expect IR-optimality of the MLE when Θ is not convex. To explain this we need the concept of exponential convexity.

We return to the general framework of Section 1.

Let $\{P_\theta\colon\ \theta \in \overline{\Theta}\}$ be the class of all probability measures on $(\mathscr{X}, \mathscr{B})$. For $\eta, \theta \in \overline{\Theta}$ denote by dP_η and dP_θ the densities of P_η and P_θ wrt a dominating measure μ. The family $\{P_{\gamma(\alpha)}\colon\ \gamma(\alpha) \in \overline{\Theta}, \alpha \in [0,1]\}$ *between* P_η and P_θ is defined by its μ-densities

$$dP_{\gamma(\alpha)}(x) = \exp\left\{\alpha \log \frac{dP_\theta}{dP_\eta}(x) - \psi^{\eta,\theta}(\alpha)\right\} dP_\eta(x) 1_{\{dP_\theta > 0\}}(x), \tag{2.9}$$

where $\psi^{\eta,\theta}(\alpha)$ is a normalizing constant. Further for $\eta, \theta \in \overline{\Theta}$ we denote by $\Gamma(\eta, \theta)$ the set $\{\gamma(\alpha)\colon\ \alpha \in [0,1]\}$, where $\gamma(\alpha)$ is determined by (2.9). A family $\{P_\theta\colon \theta \in \Theta\}$ is called *exponentially convex* when $\eta, \theta \in \Theta$ implies $\Gamma(\eta, \theta) \subset \Theta$.

When P_η, P_θ are members of an exponential family $\{P_\theta:\theta \in \Theta\}$, the family between P_η and P_θ is a *linear* subfamily of $\{P_\theta:\ \theta \in \Theta^*\}$: $\Gamma(\eta,\theta) = \{\alpha\theta + (1-\alpha)\eta:\ \alpha \in [0,1]\}$. Note that here $\Gamma(\eta,\theta) \subset \Theta$ for all $\eta,\theta \in \Theta$ iff Θ is convex. So if the parameter space of an exponential family is convex, the family is exponentially convex.

Furthermore we define for $\eta,\theta \in \Theta$ the number $C(\eta,\theta)$ by

$$C(\eta,\theta) = \inf\{\max[K(\zeta,\eta),K(\zeta,\theta)] : \zeta \in \overline{\Theta}\}. \tag{2.10}$$

This number is strongly related to the Chernoff index; cf. (2.15) and Chernoff (1952). When restricted to a subfamily Θ of $\overline{\Theta}$ we define

$$C_\Theta(\eta,\theta) = \inf\{\max[K(\zeta,\eta),K(\zeta,\theta)] : \zeta \in \Theta\}. \tag{2.11}$$

The following lemma indicates that (1.5) cannot hold for each $\theta_0 \in \Theta$ when $\{P_\theta:\ \theta \in \Theta\}$ is not exponentially convex.

LEMMA 2.4. (i) *If* $\eta,\theta \in \Theta$ *satisfy*

$$C(\eta,\theta) < C_\Theta(\eta,\theta) \tag{2.12}$$

then for each estimator $\{T_n\}$ *of* θ *and each* b *with* $C(\eta,\theta) < b < C_\Theta(\eta,\theta)$ *condition* (1.5) *fails at least at one of the points* η *and* θ.

(ii) *If* $\{P_\theta:\ \theta \in \Theta\}$ *is exponentially convex then* $C(\eta,\theta) = C_\Theta(\eta,\theta)$ *for all* $\eta,\theta \in \Theta$.

(iii) *If* $\{P_\theta:\ \theta \in \Theta\}$ *is closed in total variation and* $\{P_\theta:\ \theta \in \Theta\}$ *is not exponentially convex, then there are* $\eta,\theta \in \Theta$ *such that* (2.12) *holds*.

PROOF. (i) Let $C(\eta,\theta) < b < C_\Theta(\eta,\theta)$ and choose $\zeta \in \overline{\Theta}$ such that $\max\{K(\zeta,\eta),K(\zeta,\theta)\} < b$. Since T_n takes values in Θ only we have $\mathbb{P}_\zeta(\max\{K(T_n,\eta),K(T_n,\theta)\} > b) = 1$ for each n, implying

$$\limsup_{n\to\infty} \mathbb{P}_\zeta(K(T_n,\eta) > b) > 0 \quad\text{or}\quad \limsup_{n\to\infty} \mathbb{P}_\zeta(K(T_n,\theta) > b) > 0.$$

A slight modification of Theorem 2.1 in Bahadur et al. (1980) yields the result.

(ii) Without loss of generality assume $C(\eta,\theta) < \infty$ or, equivalently, $\mu(dP_\eta\, dP_\theta > 0) > 0$. Noting that $\alpha \to K(\gamma(\alpha),\eta)$ and $\alpha \to K(\gamma(\alpha),\theta)$ are continuous and monotone, an $\tilde{\alpha}$ exists which minimizes $\max[K(\gamma(\alpha),\eta),K(\gamma(\alpha),\theta)]$ over $\alpha \in [0,1]$. The probability measure $P_{\gamma(\tilde{\alpha})}$ is unique. We shall prove that for each $\zeta \in \overline{\Theta}$ with $K(\zeta,\eta)$ and $K(\zeta,\theta)$ finite

$$\max[K(\zeta,\eta),K(\zeta,\theta)] \ge K(\zeta,\gamma(\tilde{\alpha})) + \max[K(\gamma(\tilde{\alpha}),\eta),K(\gamma(\tilde{\alpha}),\theta)]. \tag{2.13}$$

Let $\zeta \in \overline{\Theta}$ with $K(\zeta,\eta)$ and $K(\zeta,\theta)$ finite, then $P_\zeta \ll P_{\gamma(\alpha)}$. Without loss of generality assume $K(\zeta,\eta) \ge K(\zeta,\theta)$. In view of (2.9) we have

$$\begin{aligned} &K(\zeta,\eta) - K(\zeta,\gamma(\tilde{\alpha})) - K(\gamma(\tilde{\alpha}),\eta) \\ &\quad = \tilde{\alpha}[K(\zeta,\eta) - K(\zeta,\theta) - \{K(\gamma(\tilde{\alpha}),\eta) - K(\gamma(\tilde{\alpha}),\theta)\}] \\ &\quad \ge \tilde{\alpha}\{K(\gamma(\tilde{\alpha}),\theta) - K(\gamma(\tilde{\alpha}),\eta)\}. \end{aligned} \tag{2.14}$$

First assume $\tilde{\alpha} = 0$, then $K(\gamma(0),\eta) \ge K(\gamma(0),\theta)$ and (2.14) implies (2.13). In case

$0 < \tilde{\alpha} < 1$ we have $K(\gamma(\tilde{\alpha}), \eta) = K(\gamma(\tilde{\alpha}), \theta)$ and (2.14) implies (2.13). If $\tilde{\alpha} = 1$, then $K(\gamma(1), \theta) \geq K(\gamma(1), \eta)$ and again (2.14) implies (2.13). Having established (2.13) it follows that *the infimum in* (2.10) *is attained at the unique probability measure* $P_{\gamma(\tilde{\alpha})}$.

Since exponential convexity implies $\gamma(\tilde{\alpha}) \in \Theta$, the proof of (ii) is complete.

(iii) Let $\{P_\theta: \theta \in \Theta\}$ be not exponentially convex. Then there exist $\eta^*, \theta^* \in \Theta$ such that $\Gamma(\eta^*, \theta^*) - \Theta$ is nonempty. The set $\{\alpha: \gamma^*(\alpha) \in \Theta\}$, where γ^* is associated with $\Gamma(\eta^*, \theta^*)$, is closed because $K(\cdot, \cdot)$ is continuous on $\Gamma(\eta^*, \theta^*)$ and $K(\zeta_n, \gamma^*(\alpha)) \to 0$ implies convergence in total variation of P_{ζ_n} to $P_{\gamma^*(\alpha)}$; cf. Pinsker (1964) page 20. Now let α_1, α_2 be the endpoints of the (a) largest open (relative to $[0,1]$) interval U with $\{\gamma^*(\alpha): \alpha \in U\} \wedge \Theta = \varnothing$. Define

$$P_\eta = \begin{cases} P_{\gamma^*(\alpha_1)} & \text{when } \alpha_1 > 0, \\ P_{\gamma^*(0)} & \text{when } \alpha_1 = 0 \text{ and } P_{\gamma^*(0)} \in \Theta, \\ P_{\eta^*} & \text{when } \alpha_1 = 0 \text{ and } P_{\gamma^*(0)} \notin \Theta, \end{cases}$$

and P_θ similarly for α_2, $P_{\gamma^*(1)}$, and P_{θ^*}. Note that $\gamma(\tilde{\alpha}) \notin \Theta$, where γ is associated with $\Gamma(\eta, \theta)$, and that $C(\eta, \theta) < \infty$ since $\Gamma(\eta^*, \theta^*) \neq \varnothing$. Noting that the infimum in (2.10) is attained at the unique probability measure $P_{\gamma(\tilde{\alpha})}$, combination of (2.11) and (2.13) yields

$$C_\Theta(\eta, \theta) \geq \inf\{K(\zeta, \gamma(\tilde{\alpha})): \zeta \in \Theta\} + C(\eta, \theta) > C(\eta, \theta),$$

because $K(\zeta_n, \gamma(\tilde{\alpha})) \to 0$ with $\zeta_n \in \Theta$ implies that P_{ζ_n} converges in total variation to $P_{\gamma(\tilde{\alpha})}$ and hence that $\gamma(\tilde{\alpha}) \in \Theta$, which is a contradiction. This completes the proof of the lemma. □

REMARK 2.2. When $\{P_\theta: \theta \in \Theta\}$ is not exponentially convex, closed in total variation, and connected, usually there exist $\eta, \theta \in \Theta$ with $C(\eta, \theta)$ arbitrarily small, thereby refuting (1.5) for arbitrarily small values of b.

REMARK 2.3. Consider a *curved exponential family* with statistical curvature unequal to zero. Although it may be possible to obtain an IR-optimal estimator for each $\varepsilon > 0$ at a fixed θ_0 or for a fixed $\varepsilon_0 > 0$ at each $\theta \in \Theta$ (cf. Examples 3.6 and 3.7 in Kester, 1985, pages 41, 42), IR-optimal estimators for a *class* of ε's and θ's usually do not exist, due to the fact that a curved exponential family is not exponentially convex unless the curve is a straight line in the natural parameter space.

REMARK 2.4. Let $\eta, \theta \in \overline{\Theta}$ be such that $C(\eta, \theta) < \infty$, i.e., $\mu(dP_\eta\, dP_\theta > 0) > 0$. The function $\psi^{\eta, \theta}$ as defined in (2.9) is convex and continuous on [0,1]. Moreover we have for all $0 < \alpha < 1$

$$\frac{d}{d\alpha}\psi^{\eta, \theta}(\alpha) = E_{\gamma(\alpha)}\log\left(\frac{dP_\theta}{dP_\eta}\right) = K(\gamma(\alpha), \eta) - K(\gamma(\alpha), \theta).$$

Inspection of the proof of Lemma 2.4 (ii) now yields

$$(2.15)\quad C(\eta,\theta) = -\psi^{\eta,\theta}(\tilde{\alpha}) = -\inf_{0<\alpha<1}\psi^{\eta,\theta}(\alpha) = -\log\inf_{0<\alpha<1}\int dP_\theta^\alpha\, dP_\eta^{1-\alpha}\, d\mu.$$

EXAMPLE 2.1. (i) Let $\{P_\theta\colon \theta\in\Theta\}$ be the class of all probability measures having a positive and continuous Lebesgue density on $\mathbb{R}$ and let g map θ onto the median of P_θ. Note that $\{P_\theta\colon \theta\in\Theta\}$ is exponentially convex. Bahadur et al. (1980) proved that the sample median is IR-optimal for all $\theta\in\Theta$ and $\varepsilon>0$.

(ii) Let $\{P_\theta\colon \theta\in\Theta\}$ and g as in (i), but with the restriction that P_θ is *symmetric* about $g(\theta)$. This class is *not* exponentially convex. Indeed, Example 2.2 in Kester (1985, pages 27–30) shows that IR-optimal estimators do not exist in this example as already presumed in Bahadur et al. (1980); this parallels the known fact (cf. Pfanzagl (1976)) that the sample median is not optimal wrt the asymptotic variance in this class.

3. Shift families. In this section let $\{P_\theta\colon \theta\in\mathbb{R}\}$ be a shift family of probability measures on $\mathbb{R}$ with Lebesgue densities

$$(3.1)\qquad p_\theta(x) = p(x-\theta),\qquad x,\theta\in\mathbb{R},$$

and let $g(\theta)=\theta$.

Only in some exceptional cases shift families are exponentially convex. We may therefore expect that usually Bahadur's bound is not attained. On the other hand translation equivariance is a natural restriction for location estimators in shift families. In the following lemma an upper bound for the inaccuracy rate of equivariant estimators is derived. This result generalizes previous work of Sievers (1978). Note that for equivariant estimators the inaccuracy rate is independent of θ; it is denoted by $e(\varepsilon,\{T_n\})$.

LEMMA 3.1. *If T_n is equivariant then $e(\varepsilon,\{T_n\})\le C(-\varepsilon,\varepsilon)$; cf.* (2.10).

REMARK 3.1. By (2.15) we have

$$(3.2)\qquad C(-\varepsilon,\varepsilon) = -\log\inf_{0<\alpha<1}\int p^\alpha(x-\varepsilon)p^{1-\alpha}(x+\varepsilon)\,dx,$$

which is the expression for the bound in Sievers (1978). The bound $C(-\varepsilon,\varepsilon)$ will be called Sievers' bound.

PROOF OF LEMMA 3.1. Let $\zeta\in\overline{\Theta}$ satisfy $K(\zeta,-\varepsilon)<\infty$ and $K(\zeta,\varepsilon)<\infty$; then $P_\zeta \ll P_\varepsilon$; hence P_ζ has a Lebesgue density. The equivariance of $\{T_n\}$ now implies $\mathbb{P}_\zeta(T_n=0)=0$ since the Lebesgue measure of the same event is zero.

Let $\{n_i\}$ be a subsequence such that $\lim_{i\to\infty} n_i^{-1}\log\mathbb{P}_0(|T_{n_i}|>\varepsilon) = -e(\varepsilon,\{T_n\})$. If $\limsup_i \mathbb{P}_\zeta(T_{n_i}>0)>0$, there exists a subsequence $\{m_i\}$ of $\{n_i\}$ such that $\lim_i \mathbb{P}_\zeta(T_{m_i}>0)>0$ and hence by equivariance and Theorem 2.1 in Bahadur et al. (1980)

$$\begin{aligned} -e(\varepsilon, \{T_n\}) &= \lim_i m_i^{-1} \log \mathbb{P}_0(|T_{m_i}| > \varepsilon) \geq \liminf_i m_i^{-1} \log \mathbb{P}_0(T_{m_i} > \varepsilon) \\ &= \liminf_i m_i^{-1} \log \mathbb{P}_{-\varepsilon}(T_{m_i} > 0) \geq -K(\zeta, -\varepsilon) \\ &\geq -\max\{K(\zeta, -\varepsilon), K(\zeta, \varepsilon)\}. \end{aligned} \tag{3.3}$$

If $\limsup_i \mathbb{P}_\zeta(T_{m_i} > 0) = 0$, then $\lim_i \mathbb{P}_\zeta(T_{m_i} < 0) = 1 > 0$ and hence

$$\begin{aligned} -e(\varepsilon, \{T_n\}) &= \lim_i m_i^{-1} \log \mathbb{P}_0(|T_{m_i}| > \varepsilon) \geq \liminf_i m_i^{-1} \log \mathbb{P}_0(T_{m_i} < -\varepsilon) \\ &= \liminf_i m_i^{-1} \log \mathbb{P}_\varepsilon(T_{m_i} < 0) \geq -K(\zeta, \varepsilon) \\ &\geq -\max\{K(\zeta, -\varepsilon), K(\zeta, \varepsilon)\}. \end{aligned} \tag{3.4}$$

Since $\zeta \in \overline{\Theta}$ is arbitrarily chosen, combination of (3.3) and (3.4) yields the result. □

In contrast to Sievers' claim (1978, page 611) Bahadur's bound can be less than Sievers' bound as is shown by the following example.

EXAMPLE 3.1. Let $p(x) = e^{-x} 1_{[0,\infty)}(x)$, then $\{P_\theta: \theta \in \mathbb{R}\}$ defined by (3.1) is the exponential shift family. Bahadur's bound $b(\varepsilon) = \varepsilon$ and this bound is attained by $\min\{X_i: 1 \leq i \leq n\}$; Sievers' bound $C(-\varepsilon, \varepsilon) = 2\varepsilon$ and this bound is attained by $\min\{X_i: 1 \leq i \leq n\} - \varepsilon$, which estimator is not consistent. Examples to the same effect have been given by Kester (1985, page 62) and Fu (1985).

Our next aim is to derive the inaccuracy rate for a wide class of M-estimators and to investigate at which of these estimators Sievers' bound is attained. An M-estimator is defined here as a suitable zero or change of sign of

$$\lambda_n(t) = \sum_{i=1}^{n} \psi(X_i - t),$$

where ψ is a function into the extended real line which attains positive as well as negative values, but not both $-\infty$ and $+\infty$. We consider two classes of functions ψ requiring either

$$\psi \text{ is nondecreasing} \tag{3.5}$$

or

(3.6) ψ is bounded, continuous, and such that λ_n has at least one zero for each $n[P_0]$.

The condition on λ_n holds when $x\psi(x)$ is nonnegative for $|x|$ large enough. When ψ satisfies (3.5) the M-estimator $\{T_n\} = \{T_n^{(\psi)}\}$ is defined by

$$T_n = \sup\{t: \lambda_n(t) \geq 0\}. \tag{3.7}$$

When ψ satisfies (3.6), $\{T_n\}$ is defined by

$$T_n = \begin{cases} t^+ & \text{when } t^+ - M_n \le M_n - t^-, \\ t^- & \text{when } t^+ - M_n > M_n - t^-, \end{cases} \tag{3.8}$$

where

$$t^+ = \inf\{t\colon t \ge M_n, \lambda_n(t) = 0\},$$
$$t^- = \sup\{t\colon t \le M_n, \lambda_n(t) = 0\},$$

and where $M_n = X_{[n/2]:n}$ is the sample median. Note that definitions (3.7) and (3.8) render $\{T_n\}$ translation equivariant.

The inaccuracy rates of these estimators involve the log-moment generating functions of $\psi(X)$ under P_ε and $P_{-\varepsilon}$; we define

$$\rho_\theta(\tau) = \log \int e^{\tau\psi(x)}\, dP_\theta(x)$$

and the quantity $e_\psi(\varepsilon)$ by

$$e_\psi(\varepsilon) = \min\Big\{-\inf_{\tau \ge 0} \rho_{-\varepsilon}(\tau), -\inf_{\tau \le 0} \rho_\varepsilon(\tau)\Big\}.$$

In the following two theorems the inaccuracy rate of M-estimators is determined.

THEOREM 3.2. *Let ψ satisfy* (3.5) *and let $\{T_n\}$ be defined by* (3.7). *If $P_\varepsilon(\psi(X_1) < 0) > 0$ or $P_\varepsilon(\psi(X_1) = 0) = 0$ then*

$$e(\varepsilon, \{T_n\}) = e_\psi(\varepsilon). \tag{3.9}$$

Since this result is very similar to Theorem 2 in Rubin and Rukhin (1983), the proof of Theorem 3.2, which is essentially an application of Chernoff's theorem, is omitted.

THEOREM 3.3. *Assume that p is positive in a neighbourhood of* 0 *and that $P_0((-\infty, 0)) = \frac{1}{2}$. If ψ satisfies* (3.6) *and is moreover continuously differentiable with bounded derivative such that $|\psi'(x) - \psi'(y)| \le c|x - y|$ for some $c < \infty$ and all $x, y \in \mathbb{R}$, and such that $\int \psi'(x)p(x)\,dx > 0$, then for each $0 < \varepsilon < \varepsilon_0$*

$$e(\varepsilon, \{T_n\}) = e_\psi(\varepsilon),$$

where $\{T_n\}$ is defined by (3.8).

REMARK 3.2. Rubin and Rukhin (1983) remark that for the MLE of the Cauchy shift family (3.9) does not hold. Noting that this MLE is obtained as the M-estimator with $\psi(x) = 2x(1 + x^2)^{-1}$, Theorem 3.3 implies however that (3.9) does hold in this case when ε is sufficiently small. Together with Theorem 3.4 (cf. Example 3.2) this also provides an answer to the open problem mentioned in Fu (1985, Remark 2).

PROOF OF THEOREM 3.3. Let $c_1 = \int \psi' p\, dx > 0$. Since the Lipschitz condition on ψ' is "inherited" by $n^{-1}\lambda'_n$ we have writing $\delta = \frac{1}{2}c^{-1}c_1$

$$n^{-1}\lambda'_n(0) < -\tfrac{1}{2}c_1 \Rightarrow n^{-1}\lambda'_n(t) < 0 \quad \text{on } (-\delta, \delta). \tag{3.10}$$

Let $0 < \varepsilon < \frac{1}{2}\delta$. If $|T_n| > \varepsilon$, $\lambda_n(-\varepsilon) > 0$, $\lambda_n(\varepsilon) < 0$, and $\lambda'_n(t) < 0$ on $(-\delta, \delta)$ then $|M_n| \geq \frac{1}{4}\delta$; hence we obtain

$$\begin{aligned}\mathbb{P}_0(|T_n| > \varepsilon) \leq \mathbb{P}_0(\lambda_n(-\varepsilon) \leq 0 \text{ or } \lambda_n(\varepsilon) \geq 0) + \mathbb{P}_0(|M_n| \geq \tfrac{1}{4}\delta) \\ + \mathbb{P}_0(n^{-1}\lambda'_n(0) \geq -\tfrac{1}{2}c_1).\end{aligned} \tag{3.11}$$

By Chernoff's theorem and translation equivariance we have

$$-\lim_{n\to\infty} n^{-1}\log \mathbb{P}_0(\lambda_n(-\varepsilon) \leq 0 \text{ or } \lambda_n(\varepsilon) \geq 0) = e_\psi(\varepsilon). \tag{3.12}$$

Writing $p_+ = P_0((\frac{1}{4}\delta, \infty))$, $p_- = P_0((-\infty, -\frac{1}{4}\delta))$ it is readily seen (cf. Example 6.1 in Bahadur (1971)) that

$$\begin{aligned}-\lim_{n\to\infty} n^{-1}\log \mathbb{P}_0(|M_n| \geq \tfrac{1}{4}\delta) = \min\{-\tfrac{1}{2}\log[4p_+(1-p_+)], \\ -\tfrac{1}{2}\log[4p_-(1-p_-)]\} > 0.\end{aligned} \tag{3.13}$$

The derivative of

$$\tau \to \log \int e^{\tau(\psi' - c_1/2)} p\, dx$$

at $\tau = 0$ equals $\frac{1}{2}c_1 > 0$, and hence

$$-\inf_{\tau \leq 0} \log \int e^{\tau(\psi' - c_1/2)} p\, dx = c_2 > 0.$$

By Chernoff's theorem it follows that

$$-\lim_{n\to\infty} n^{-1}\log \mathbb{P}_0(n^{-1}\lambda'_n(0) \geq -\tfrac{1}{2}c_1) = c_2 > 0. \tag{3.14}$$

Since ψ is bounded and continuous, $\rho_\theta(\tau)$ is continuous in θ and τ by dominated convergence. Moreover, by strict convexity of ρ_0 we have $\rho_0(\tau) > 0$ for each $\tau > 0$ or $\rho_0(\tau) > 0$ for each $\tau < 0$. Without loss of generality assume the latter; by pointwise convergence of ρ_ε to ρ_0 and convexity of ρ_ε it follows that $\inf_{\tau \leq 0}\rho_\varepsilon(\tau) \to 0$ as $\varepsilon \to 0$, implying

$$e_\psi(\varepsilon) \to 0 \quad \text{as } \varepsilon \to 0. \tag{3.15}$$

Combination of (3.11), (3.12), (3.13), (3.14), and (3.15) yields that there exists $\varepsilon_1 > 0$ such that for each $0 < \varepsilon < \varepsilon_1$

$$-\limsup_{n\to\infty} n^{-1}\log \mathbb{P}_0(|T_n| > \varepsilon) \geq e_\psi(\varepsilon).$$

On the other hand

$$\begin{aligned}\mathbb{P}_0(|T_n| > \varepsilon) = \mathbb{P}_0(|T_n| \geq \varepsilon) \geq \mathbb{P}_0(\lambda_n(-\varepsilon) \leq 0 \text{ or } \lambda_n(\varepsilon) \geq 0) \\ -\mathbb{P}_0(n^{-1}\lambda'_n(0) \geq -\tfrac{1}{2}c_1)\end{aligned}$$

and hence there exists $\varepsilon_0 < \varepsilon_1$ such that for each $0 < \varepsilon < \varepsilon_0$

$$-\liminf_{n\to\infty} n^{-1}\log \mathbb{P}_0(|T_n| > \varepsilon) \leq e_\psi(\varepsilon). \square$$

REMARK 3.3. The only property of the sample median M_n we need in the above proof is (3.13). Therefore if we define $\{\tilde{T}_n\}$ by (3.8) using another preliminary estimator $\{\tilde{M}_n\}$ which satisfies

$$-\lim_{n\to\infty} n^{-1}\log \mathbb{P}_0(|\tilde{M}_n| > \delta) > 0$$

for each $\delta > 0$, Theorem 3.3 holds for $\{\tilde{T}_n\}$.

Having established an expression for the inaccuracy rate of M-estimators (Theorems 3.2 and 3.3) and an upper bound for the inaccuracy rate of equivariant estimators (Lemma 3.1) the question arises naturally which, if any, of the M-estimators attains Sievers' bound. In vew of (3.2) Sievers' bound $C(-\varepsilon, \varepsilon)$ can be written as

$$-\log \inf_{0<\alpha<1} \int \exp\left(\alpha \log \frac{p(x-\varepsilon)}{p(x+\varepsilon)}\right) p(x+\varepsilon)\, dx.$$

Define

$$\psi_\varepsilon(x) = \log \frac{p(x-\varepsilon)}{p(x+\varepsilon)}. \tag{3.16}$$

We assume that either $\psi_\varepsilon < \infty$ a.e. or $\psi_\varepsilon > -\infty$ a.e. If $\psi_\varepsilon > -\infty$ a.e. then $\inf_{\tau\le 0}\rho_\varepsilon(\tau) = \inf_{0<\tau<1}\rho_{-\varepsilon}(\tau)$ and hence $e_{\psi_\varepsilon}(\varepsilon) = C(-\varepsilon, \varepsilon)$. If $\psi_\varepsilon < \infty$ a.e. then $\inf_{\tau\ge 0}\rho_{-\varepsilon}(\tau) = \inf_{0<\tau<1}\rho_\varepsilon(\tau)$ and hence $e_{\psi_\varepsilon}(\varepsilon) = C(-\varepsilon, \varepsilon)$. Therefore as a rule the M-estimator based on ψ_ε given by (3.16) is inaccuracy rate optimal within the class of equivariant estimators. For instance, when ψ_ε is nondecreasing and either $> -\infty$ a.e. or $< \infty$ a.e., indeed $\{T_n^{(\psi_\varepsilon)}\}$ attains Sievers' bound; cf. Sievers (1978), Theorem 2.1 and Fu (1985). Note that in general the optimal M-estimator will depend on ε.

For a nonmonotone ψ_ε we cannot apply Theorem 3.3 directly, since in general ε_0 will depend on $\psi = \psi_\varepsilon$; so for a fixed ε^*, say, $\varepsilon_0 = \varepsilon_0(\psi_{\varepsilon^*})$ may be smaller than ε^*. Nevertheless the next theorem states that if p is sufficiently smooth (3.9) holds for $\psi = \psi_\varepsilon$ when ε is small enough, with the important implication that Sievers' bound is attainable in these situations. Even for the Cauchy shift family an inaccuracy rate optimal M-estimator is obtained by this result; cf. Example 3.2.

THEOREM 3.4. *Let $p > 0$ on $\mathbb{R}$ with $\int_{-\infty}^{0} p\,dx = \frac{1}{2}$. If p is three times differentiable such that $p'(x) > 0$ (< 0) for each small (large) enough x, such that the first three derivatives of* $\log p$ *are bounded and such that*

$$\int (\log p)'' p\, dx < 0,$$

then

$$e\left(\varepsilon, \left\{T_n^{(\psi_\varepsilon)}\right\}\right) = e_{\psi_\varepsilon}(\varepsilon) = C(-\varepsilon, \varepsilon) \tag{3.17}$$

when ε is small enough.

PROOF. For $\varepsilon \to 0$ the function ψ_ε strongly depends on ε. Therefore we define the "standardized" $\tilde{\psi}_\varepsilon = (2\varepsilon)^{-1}\psi_\varepsilon$; then obviously $T_n^{(\psi_\varepsilon)} = T_n^{(\tilde{\psi}_\varepsilon)}$. It is easily checked that $\tilde{\psi}_\varepsilon$ satisfies (3.6). Further we have

$$e_{\tilde{\psi}_\varepsilon}(\varepsilon) = e_{\psi_\varepsilon}(\varepsilon) = C(-\varepsilon, \varepsilon) \to 0 \quad \text{as } \varepsilon \to 0.$$

Noting that

$$\lim_{\varepsilon \to 0} \int \tilde{\psi}_\varepsilon' p\, dx = -\int (\log p)'' p\, dx = c_1 > 0$$

and

$$\begin{aligned}\liminf_{\varepsilon \to 0} \; - \inf_{\tau \le 0} \log \int e^{\tau(\tilde{\psi}_\varepsilon' - c_1/2)} p\, dx &\ge - \inf_{\tau \le 0} \lim_{\varepsilon \to 0} \log \int e^{\tau(\tilde{\psi}_\varepsilon' - c_1/2)} p\, dx \\ &= - \inf_{\tau \le 0} \log \int e^{\tau(-(\log p)'' - c_1/2)} p\, dx > 0,\end{aligned}$$

we can follow for $\tilde{\psi}_\varepsilon$ the same line of argument as in the proof of Theorem 3.3, using the mean value theorem to bound $\tilde{\psi}_\varepsilon$, $\tilde{\psi}_\varepsilon'$, and $\tilde{\psi}_\varepsilon''$. We omit further details. □

EXAMPLE 3.2. Consider the Cauchy density $p(x) = \pi^{-1}(1 + x^2)^{-1}$. We have

$$\psi_\varepsilon(x) = \log \frac{1 + (x + \varepsilon)^2}{1 + (x - \varepsilon)^2}.$$

The conditions of Theorem 3.4 hold, hence $\{T_n^{(\psi_\varepsilon)}\}$ attains Sievers' bound when ε is sufficiently small.

It is seen in the proof of Lemma 2.4(ii) that $C(-\varepsilon, \varepsilon) = \max\{K(\gamma(\tilde{\alpha}), -\varepsilon), K(\gamma(\tilde{\alpha}), \varepsilon)\}$. It can be shown that in regular cases $e(\varepsilon\{T_n\}) = C(-\varepsilon, \varepsilon)$ implies that the influence curve at $P_{\gamma(\tilde{\alpha})}$ of the estimator $\{T_n\}$ is a.e. proportional to ψ_ε given by (3.16), and hence $\{T_n^{(\psi_\varepsilon)}\}$ is the essentially unique M-estimator which attains Sievers' bound; cf. Theorem 4.7 in Kester (1985), page 76. However, possibly also an L-estimator attains Sievers' bound. The next example shows that this occurs in the double exponential family, where a trimmed mean attains Sievers' bound.

EXAMPLE 3.3. Let $p(x) = \frac{1}{2}e^{-|x|}$. Sievers' bound $C(-\varepsilon, \varepsilon) = \varepsilon - \log(1 + \varepsilon)$ is attained by the M-estimator $\{T_n^{(\psi_\varepsilon)}\}$ with ψ_ε given by

$$\psi_\varepsilon(x) = \begin{cases} -2\varepsilon & \text{when } x < -\varepsilon, \\ 2x & \text{when } |x| \le \varepsilon, \\ 2\varepsilon & \text{when } x > \varepsilon; \end{cases}$$

cf. Sievers (1978). The probability measure $P_{\gamma(\tilde{\alpha})}$ is given by

$$\frac{dP_{\gamma(\tilde{\alpha})}(x)}{dx} = \begin{cases} \frac{1}{2}(1 + \varepsilon)^{-1}\exp\{-|x| + \varepsilon\} & \text{when } |x| > \varepsilon, \\ \frac{1}{2}(1 + \varepsilon)^{-1} & \text{when } |x| \le \varepsilon. \end{cases}$$

The influence curve at $P_{\gamma(\tilde{\alpha})}$ of the $\frac{1}{2}(1+\varepsilon)^{-1}$-trimmed mean L_n, say, is a.e. proportional to the influence curve of $\{T_n^{(\psi_\varepsilon)}\}$. So this L-estimator is a candidate for being optimal. By symmetry the inaccuracy rate of $\{L_n\}$ equals

$$-\lim_{n\to\infty} n^{-1}\log \mathbb{P}_0(L_n > \varepsilon).$$

In view of Theorem 6.3 and formula (6.13) of Groeneboom et al. (1979) this can be expressed as (writing $\alpha = \frac{1}{2}(1+\varepsilon)^{-1}$)

$$\begin{aligned} &2\alpha\log\alpha + (1-2\alpha)\log(1-2\alpha) \\ &\quad + \inf\Big\{\sup_{t\geq 0} f(a,b,t)\colon -\infty < a < b < \infty,\, b > \varepsilon\Big\} \end{aligned} \tag{3.18}$$

with

$$\begin{aligned} f(a,b,t) &= (1-2\alpha)\left[t\varepsilon - \log\int_a^b e^{tx}p(x)\,dx\right] \\ &\quad -\alpha[\log F(a) + \log(1-F(b))], \end{aligned} \tag{3.19}$$

where F denotes the distribution function of P_0. We shall show that $f(0,2\varepsilon,1)$ attains the sup and inf in (3.18). Consider

$$\begin{aligned} f(a,b,1) - f(0,2\varepsilon,1) &= (1-2\alpha)\left[-\log\int_a^b \tfrac{1}{2}e^{x-|x|}\,dx + \log\int_0^{2\varepsilon}\tfrac{1}{2}\,dx\right] \\ &\quad -\alpha\left[\log F(a) - \log\tfrac{1}{2} + \log(1-F(b)) - \log\tfrac{1}{2}e^{-2\varepsilon}\right]. \end{aligned}$$

Multiplying by $(1+\varepsilon)/\varepsilon = (1-2\alpha)^{-1}$ we obtain, when $a < 0$ and $b > \varepsilon$,

$$\begin{aligned} &\log\varepsilon - \log\frac{1}{2} - \log\left[\frac{1-e^{2a}}{2} + b\right] - \frac{1}{2\varepsilon}[a - b + 2\varepsilon] \\ &\quad = -\log\left[\frac{1}{2\varepsilon}\left(\frac{1-e^{2a}}{2} + b\right)\right] - \frac{1}{2\varepsilon}[a-b] - 1 \\ &\quad \geq \frac{1}{4\varepsilon}[e^{2a} - 2a - 1] \geq 0, \end{aligned}$$

where the inequality $\log x \leq x - 1$ was used. When $a \geq 0$ we find in a similar way that

$$f(a,b,1) \geq f(0,2\varepsilon,1).$$

It follows that

$$\inf_{a,b}\sup_{t\geq 0} f(a,b,t) \geq f(0,2\varepsilon,1)$$

and it suffices to remark that $f(0,2\varepsilon,1) \geq f(0,2\varepsilon,t)$ by symmetry and convexity of

$$t \to \int_0^{2\varepsilon} e^{(t-1)(x-\varepsilon)}\,dx.$$

It remains to evaluate $f(0,2\varepsilon,1)$. Together with the other part of (3.18) this indeed equals $\varepsilon - \log(1+\varepsilon)$. So the $\frac{1}{2}(1+\varepsilon)^{-1}$-trimmed mean attains Sievers' bound.

4. Tail-behaviour of location estimators. The asymptotic behaviour as $\varepsilon \to \infty$ of

$$B(\varepsilon, T_n) = \frac{-\log \mathbb{P}_0(|T_n| > \varepsilon)}{-\log \mathbb{P}_0(|X_1| > \varepsilon)} \tag{4.1}$$

is proposed by Jurečková (1979, 1981) as a basis for comparison of translation equivariant estimators T_n in location families $\{p_\theta(x) = p(x - \theta)\colon x, \theta \in \mathbb{R}\}$. For symmetric p and translation equivariant estimators T_n satisfying

$$\min\{X_i\colon 1 \le i \le n\} > 0 \Rightarrow T_n(X_1, \dots, X_n) > 0,$$

$$\max\{X_i\colon 1 \le i \le n\} < 0 \Rightarrow T_n(X_1, \dots, X_n) < 0,$$

it holds that

$$1 \le \liminf_{\varepsilon \to \infty} B(\varepsilon, T_n) \le \limsup_{\varepsilon \to \infty} B(\varepsilon, T_n) \le n; \tag{4.2}$$

cf. Jurečková (1979, 1981). Moreover, if

$$\lim_{\varepsilon \to \infty} \frac{-\log P_0((\varepsilon, \infty))}{b\varepsilon^r} = 1 \tag{4.3}$$

for some $b > 0$, $r \ge 1$, then the sample mean attains the upper bound in (4.2), i.e.,

$$\lim_{\varepsilon \to \infty} B(\varepsilon, \overline{X}_n) = n; \tag{4.4}$$

cf. Jurečková (1979). In this section we connect the results of Jurečková and the inaccuracy rates as treated in the previous section.

First consider just as above a *fixed sample size n*. Suppose that $\psi_\varepsilon(x) = \log\{p(x - \varepsilon)/p(x + \varepsilon)\}$ is nondecreasing; then $T_n^{(\psi_\varepsilon)}$ minimizes $\mathbb{P}_0(|T_n| > \varepsilon)$ over the class of translation equivariant estimators; cf. Huber (1968). We prove that under a similar condition as (4.3) the optimal estimator $T_n^{(\psi_\varepsilon)}$ converges pointwise to $\bar{x}_n$ as $\varepsilon \to \infty$.

THEOREM 4.1. *If p is* log *concave and satisfies*

$$-\log p(x) = b|x|^r(1 + f(x)) \tag{4.5}$$

with $b > 0$, $r \ge 1$, $f(x) \to 0$ as $|x| \to \infty$ and for each x

$$f(x - \varepsilon) = f(x + \varepsilon) + o(\varepsilon^{-1}) \quad \text{as } \varepsilon \to \infty, \tag{4.6}$$

then

$$\lim_{\varepsilon \to \infty} T_n^{(\psi_\varepsilon)}(x_1, \dots, x_n) = n^{-1} \sum_{i=1}^n x_i \quad \text{for each } (x_1, \dots, x_n) \in \mathbb{R}^n.$$

The proof hinges on the following lemma.

LEMMA 4.2. *Under the conditions* (4.5) *and* (4.6)

$$\psi_\varepsilon(x) = 2br\varepsilon^{r-1}x(1 + o(1)) \quad \text{as } \varepsilon \to \infty.$$

PROOF. Fix $x \in \mathbb{R}$. As $\varepsilon \to \infty$ we obtain

$$\begin{aligned}\psi_\varepsilon(x) &= -b(\varepsilon - x)^r(1 + f(x+\varepsilon) + o(\varepsilon^{-1})) + b(\varepsilon + x)^r(1 + f(x+\varepsilon)) \\ &= b\{(\varepsilon + x)^r - (\varepsilon - x)^r\}(1 + o(1)) + o(\varepsilon^{r-1}) = 2brx\varepsilon^{r-1}(1 + o(1)).\end{aligned}$$

□

PROOF OF THEOREM 4.1. Fix $\delta > 0$ and $(x_1, \dots, x_n) \in \mathbb{R}^n$. By Lemma 4.2 we have, writing $\bar{x} = n^{-1}\sum_{j=1}^n x_j$,

$$\begin{aligned}\sum_{i=1}^n \psi_\varepsilon(x_i - \bar{x} - \delta) &= 2br\varepsilon^{r-1}\sum_{i=1}^n (x_i - \bar{x} - \delta)(1 + o(1)) \\ &= -2br\varepsilon^{r-1}n\delta(1 + o(1)) < 0\end{aligned}$$

and similarly $\sum_{i=1}^n \psi_\varepsilon(x_i - \bar{x} + \delta) > 0$ when ε is sufficiently large, implying

$$|T_n^{(\psi_\varepsilon)}(x_1, \dots, x_n) - \bar{x}| < \delta.$$

□

Next consider the double exponential distribution. For each $0 < \varepsilon < \infty$ the $\frac{1}{2}(1+\varepsilon)^{-1}$-trimmed mean minimizes among translation equivariant estimators the inaccuracy rate $-\liminf_{n\to\infty} n^{-1}\log \mathbb{P}_0(|T_n| > \varepsilon)$;cf. Example 3.3. This estimator converges to the sample mean if $\varepsilon \to \infty$ and to the sample median if $\varepsilon \to 0$, thus providing a bridge between gross error optimality and strictly local optimality.

Acknowledgment. The authors are much indebted to J. Oosterhoff for his stimulating advice and his continuous interest during the preparation of this paper.

REFERENCES

BAHADUR, R. R. (1980). On large deviations of maximum likelihood and related estimates. Technical Report No. 121, Univ. of Chicago.

BAHADUR, R. R. (1983). Large deviations of the maximum likelihood estimate in the Markov chain case. In *Recent Advances in Statistics* (M. H. Rizvi, J. S. Rustagi and D. Siegmund, eds.) 273–286. Academic, New York.

BAHADUR, R. R., GUPTA, J. C. and ZABELL, S. L. (1980). Large deviations, tests and estimates. In *Asymptotic Theory of Statistical Tests and Estimation* (I.M. Chakravarti, ed.) 33–64. Academic, New York.

BERK, R. H. (1972). Consistency and asymptotic normality of MLE's for exponential models. *Ann. Math. Statist.* **43** 193–204.

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on sums of observations. *Ann. Math. Statist.* **23** 493–507.

EFRON, B. (1978). The geometry of exponential families. *Ann. Statist.* **6** 362–376.

EFRON, B. and TRUAX, D. (1968). Large deviations theory in exponential families. *Ann. Math. Statist.* **39** 1402–1424.

FU, J. C. (1982). Large sample point estimation: 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** 101–105.

GROENEBOOM, P., OOSTERHOFF, J. and RUYMGAART, F. H. (1979). Large deviation theorems for empirical probability measures. *Ann. Probab.* **7** 553–586.

HUBER, P. J. (1968). Robust confidence limits. *Z. Wahrsch. verw. Gebiete* **10** 269–278.

JUREČKOVÁ, J. (1979). Finite-sample comparison of L-estimators of location. *Comment. Math. Univ. Carolin.* **20** 509–518.

JUREČKOVÁ, J. (1981). Tail-behavior of location estimators. *Ann. Statist.* **9** 578–585.

KESTER, A. D. M. (1985). Some large deviation results in statistics. CWI Tract 18. Mathematisch Centrum, Amsterdam.

KOUROUKLIS, S. (1984). A large deviation result for the likelihood ratio statistic in exponential families. *Ann. Statist.* **12** 1510–1521.

PFANZAGL, J. (1976). Investigating the quantile of an unknown distribution. In *Contributions to Statistics* (W. J. Ziegler, ed.) 111–126. Birkhäuser, Basel.

PINSKER, M. S. (1964). *Information and Information Stability of Random Variables and Processes.* Holden-Day, San Francisco.

RUBIN, H. AND RUKHIN, A. L. (1983). Convergence rates of large deviations probabilities for point estimators. *Statist. Probab. Lett.* **1** 197–202.

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

DEPARTMENT OF MEDICAL INFORMATICS
AND STATISTICS
UNIVERSITY OF LIMBURG
P.O. BOX 616
6200 MD MAASTRICHT
THE NETHERLANDS

DEPARTMENT OF APPLIED MATHEMATICS
TWENTE UNIVERSITY OF TECHNOLOGY
P.O. BOX 217
7500 AE ENSCHEDE
THE NETHERLANDS