The Annals of Statistics
1978, Vol. 6, No. 5, 1048–1056

RATE OF CONVERGENCE OF ESTIMATORS BASED ON SAMPLE MEAN

BY S.-S. PERNG

Howard University

The rate of convergence of the point estimators for the parameter based on the sample mean of i.i.d. random vectors is obtained. The result is used to prove the Bahadur's efficiency of the regular best asymptotically normal estimator when the underlying distribution is in the exponential family. An example is given to show that if the distribution is not in the exponential family, then the regular best asymptotically normal estimator is not necessarily efficient in Bahadur's sense.

1. Introduction and summary. Let $X, X_1, X_2, \cdots,$ be i.i.d. random vectors with values in R^k and common distribution P_θ, $\theta \in \Theta \subset R^r$, where k and r are positive integers. Assume that Θ is open and that $E_\theta X = \mu(\theta) = \mu$ and $\operatorname{Cov}_\theta(X, X') = \Sigma(\theta) = \Sigma$. Write $\bar{X}_n = (1/n)(X_1 + \cdots + X_n)$. The estimator $\hat{\theta}_n$ of θ to be considered will depend on the X_i's only through $\bar{X}_n$ and possibly a consistent estimator S_n of Σ. It is well known (see, e.g., [4]) that if $h(\bar{X}_n, S_n)$ is regular (i.e., h has continuous partial derivatives) and consistent for θ, then it is asymptotically normal estimator for θ and furthermore it is regular best asymptotically normal (RBAN) estimator when certain conditions are satisfied.

The objective of this research is to establish the rate of convergence to θ of such estimator and consider its efficiency in Bahadur's sense. It is shown, for instance, under some conditions, that

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{\|h(\bar{X}_n) - h(\mu)\| \geqq \varepsilon\} = -\tfrac{1}{2}/\|A\Sigma A'\| ,$$

where A is the matrix of partial derivatives of h evaluated at μ. The norm of a matrix is defined in Section 2. In fact $\|A\Sigma A'\|$ equals the maximum eigenvalue of $A\Sigma A'$. A similar result for the case where h also depends on the sample covariance matrix is also established. The Bahadur efficiency of RBAN estimators for θ is proved for the case where P_θ is in the exponential family. An example is given to show that if P_θ is not in the exponential family, then the RBAN estimator is not necessarily efficient in Bahadur's sense.

2. Basic lemmas. In the k-dimensional Euclidean space R^k, let $\|\cdot\|$ be the Euclidean norm. It is noted that for $\varepsilon > 0$,

$$\{x : \|x\| \geqq \varepsilon\} = \{x : \max_{\|l\|=1} |l'x| \geqq \varepsilon\} . \tag{2.1}$$

LEMMA 2.1. *Let* $x \in R^k$. *If* $\varepsilon > 0$, *then for each* $0 < \lambda < 1$, *there are finite*

Received September 1976; revised October 1977.
AMS 1970 *subject classifications*. Primary 62E20; Secondary 62F20, 62H10.
Key words and phrases. Rate of convergence, asymptotically efficient, regular best asymptotically normal, sample mean, Bahadur's efficiency.

number of $l_1, l_2, \cdots, l_M \in R^k$, $\|l_i\| = 1$, *such that*

$$\{x : \|x\| \geqq \varepsilon\} \subset \bigcup_1^M \{x : |l_i x| \geqq \lambda\varepsilon\},$$

where the l_i's *and* M *depend only on* λ *and not on* ε.

PROOF. Let $S = \{x : \|x\| = \varepsilon\}$. For each $l \in R^k$ with $\|l\| = 1$, define $D_l = \{x : |l'x| > \lambda\varepsilon$ and $\|x\| = \varepsilon\}$. Then D_l is open relative to the usual topology in S. It suffices to show that there are finite number of D_l's covering S. Since S is compact the result follows by using Heine–Borel argument. Note that the D_l's depend on λ and ε but the l_i's and M depend only on λ and not on ε, i.e., if a set of the l_i's and M works for one $\varepsilon > 0$, then the same set works for all $\varepsilon > 0$.

DEFINITION. For any $p \times k$ matrix P, define the norm of P to be

$$\|P\| = \sup_{x \neq 0} \|Px\|/\|x\| = \max_{\|x\|=1} \|Px\|.$$

Note that by definition of $\|P\|$, $\|Px\| \leqq \|P\|\|x\|$. Note also that if $P = [I_p, 0]$ with I_p being the p^2 identity matrix, then $\|P\| = 1$. Furthermore the norm $\|P\|$ is a continuous function of its entries. The following lemma is readily proved.

LEMMA 2.2. *If* A *is a positive semidefinite matrix, then*

$$\|A\| = \max_{\|l\| \leqq 1} \|Al\| = \max_{\|l\|=1} l'Al = \max_{\|l\| \leqq 1} l'Al$$

which equals the maximal eigenvalue of A.

LEMMA 2.3. *If* Σ *is a* k^2 *positive definite matrix and* P *a* $r \times k$ *matrix with rank* r, *then there is a* $(k - r) \times k$ *matrix* P_1 *such that* $[P', P_1']$ *is nonsingular and* $P_1\Sigma P' = 0$. *Furthermore if* P_1 *satisfies the conclusion, so does* αP_1 *for* $\alpha \neq 0$.

PROOF. Let $Q = P\Sigma^{\frac{1}{2}}$. Then Q is of rank r. Hence there is a $(k - r) \times k$ matrix Q_1 such that $[Q', Q_1']$ is nonsingular and $QQ_1' = 0$. Let $P_1 = Q_1\Sigma^{-\frac{1}{2}}$. Then P_1 is the desired matrix. The second statement is obvious by examining the linear independence of the columns of $[P', \alpha P_1']$.

3. Rate of convergence. In this section the rate of convergence for estimators based on sample mean is obtained. The first theorem is Lemma 2.4 of Bahadur [1], page 235.

THEOREM 3.1. *Let* $Y, Y_1, \cdots,$ *be* i.i.d. *random variables with finite* mgf *and* $E(Y) = 0$ *and* $E(Y^2) = \sigma^2$. *Then for* $\varepsilon > 0$

$$\frac{1}{n} \log P\{|\bar{Y}_n| \geqq \varepsilon\} = -\varepsilon^2/(2\sigma^2)(1 + \delta(n, \varepsilon)),$$

where $\delta(n, \varepsilon) \to 0$ *as* $n \to \infty$ *and* $\varepsilon \to 0$.

For the remainder of the paper, let $X, X_1, X_2, \cdots,$ be i.i.d. random k-vectors with finite mgf and $E_\theta(X) = \mu(\theta) = \mu$ and $\mathrm{Cov}_\theta(X) = \Sigma(\theta) = \Sigma$. Assume that $\theta \in \Theta \subset R^r$, $r \leqq k$ and that Θ is open.

THEOREM 3.2. *If* $\mu(\theta) = \theta$, *then for each* $\theta \in \Theta$,

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{\|\bar{X}_n - \theta\| \geqq \varepsilon\} = -\tfrac{1}{2}/\|\Sigma\|. \tag{3.1}$$

PROOF. By Lemma 2.1, for each $0 < \lambda < 1$, there are M unit vectors $l_1, \cdots, l_M$ depending only on λ such that $\{x \in R^k : \|x - \theta\| \geqq \varepsilon\} \subset \bigcup_1^M \{x \in R^k : |l_i'(x - \theta)| \geqq \lambda\varepsilon\}$. For each n, let $l(n)$ be the l_i's such that $P_\theta\{|l_i'(\bar{X}_n - \theta)| \geqq \lambda\varepsilon\}$ is maximized. Then

$$P_\theta\{\|\bar{X}_n - \theta\| \geqq \varepsilon\} \leqq MP_\theta\{|l'(n)(\bar{X}_n - \theta)| \geqq \lambda\varepsilon\}$$

and hence

$$\begin{aligned}(3.2)\qquad &\limsup_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{\|\bar{X}_n - \theta\| \geqq \varepsilon\} \\ &\leqq \limsup_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{|l'(n)(\bar{X}_n - \theta)| \geqq \lambda\varepsilon\} .\end{aligned}$$

Note that $l(n)$ takes only finite number of values, namely $l_1, \cdots, l_M$, which do not depend on ε. Let l_0 be the l_i's which maximizes $l_i'\Sigma l_i$, $i = 1, 2, \cdots, M$. Then, by Theorem 3.1, the right-hand side of (3.2) is not greater than $-\frac{1}{2}\lambda^2/l_0'\Sigma l_0$. Letting $\lambda \to 1$, we have

$$(3.3)\qquad \limsup_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{\|\bar{X}_n - \theta\| \geqq \varepsilon\} \leqq -\tfrac{1}{2}/l_0'\Sigma l_0 \leqq -\tfrac{1}{2}/\|\Sigma\| .$$

On the other hand, by Lemma 2.1,

$$\{\|X_n - \theta\| \geqq \varepsilon\} = \{\max_{\|l\|=1} |l'(\bar{X}_n - \theta)| \geqq \varepsilon\} \supset \{|l'(\bar{X}_n - \theta)| \geqq \varepsilon\} ,$$

for all unit vector l. By Theorem 3.1 and Lemma 2.2

$$\begin{aligned}(3.4)\qquad \liminf_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{\|\bar{X}_n - \theta\| \geqq \varepsilon\} &\geqq -\tfrac{1}{2}/\max_{\|l\|=1} l'\Sigma l \\ &= \tfrac{1}{2}/\|\Sigma\| .\end{aligned}$$

The result follows from (3.3) and (3.4)

To apply the results in the following theorems to RBAN estimators, $h(\mu)$ (or $h(\mu, \Sigma)$) equals θ. But in the discussion here no such restriction is necessary.

THEOREM 3.3. *If $h: R^k \to R^r$ $(r \leqq k)$ has continuous partial derivatives, $A(x) = (\partial/\partial x)h(x)$, where $A(\cdot)$ is of rank r, then for each $\theta \in \Theta$*

$$(3.5)\qquad \liminf_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{\|h(\bar{X}_n) - h(\mu)\| \geqq \varepsilon\} = -\tfrac{1}{2}/\|A\Sigma A'\| ,$$

where $A = A(\mu)$.

PROOF. By Lemma 2.3, there is a $(k - r) \times k$ matrix A_1 such that $\left[\begin{smallmatrix} A \\ A_1 \end{smallmatrix}\right]$ is non-singular and $A_1\Sigma A' = 0$. Write $A_2^\alpha = A_2 = \left[\begin{smallmatrix} A \\ \alpha A_1 \end{smallmatrix}\right]$, $\alpha > 0$. Then

$$h(\bar{X}_n) - h(\mu) = A(\eta_n)(\bar{X}_n - \mu) = A(\eta_n)A_2^{-1}A_2(\bar{X}_n - \mu) ,$$

where η_n is in between μ and $\bar{X}_n$ and is a continuous function of $\bar{X}_n$. Thus for each $\varepsilon > 0$,

$$\begin{aligned}(3.6)\qquad \{\|h(\bar{X}_n) - h(\mu)\| \geqq \varepsilon\} &\subset \{\|A(\eta_n)A_2^{-1}\|\|A_2(\bar{X}_n - \mu)\| \geqq \varepsilon\} \\ &\subset \{\|A(\eta_n)A_2^{-1}\| > \lambda_1\} \cup \{\|A_2(\bar{X}_n - \mu)\| \geqq \varepsilon/\lambda_1\} ,\end{aligned}$$

where $\lambda_1 > 1$ is arbitrary but fixed. Let $B(x) = A(x)A_2^{-1}$. By the continuity of $A(\cdot)$, as $x \to \mu$, $B(x) \to AA_2^{-1} = [I_r, 0]$, by noting $[{}_{\alpha A_1}^{A}]A_2^{-1} = I_k$. Hence by the continuity of $B(\cdot)$ there is a neighborhood N (depending on λ_1) of μ such that $x \in N$ implies $|\|B(x)\| - 1| < \lambda_1 - 1$, which in turn implies $\|B(x)\| < \lambda_1$. Hence $\{\|B(x)\| \geqq \lambda_1\} \subset \{x \notin N\}$. Combining this with (3.6), we have

$$\{\|h(\bar{X}_n) - h(\mu)\| \geqq \varepsilon\} \subset \{\bar{X}_n \notin N\} \cup \{\|A_2(\bar{X}_n - \mu)\| \geqq \varepsilon/\lambda_1\}\,. \tag{3.7}$$

Since $\{x: \|A_2(x - \mu)\| < \varepsilon/\lambda_1\}$ is an ellipsoid centered at μ, there is an $\varepsilon_0 > 0$ depending on λ_1 such that $0 < \varepsilon \leqq \varepsilon_0$ implies

$$\{\|A_2(\bar{X}_n - \mu)\| < \varepsilon/\lambda_1\} \subset \{\bar{X}_n \in N\}\,.$$

This along with (3.7) implies that for $0 < \varepsilon \leqq \varepsilon_0$

$$\{\|h(\bar{X}_n) - h(\mu)\| \geqq \varepsilon\} \subset \{\|A_2(\bar{X}_n - \mu)\| \geqq \varepsilon/\lambda_1\}\,. \tag{3.8}$$

But by Theorem 3.2 we have

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{\|A_2(\bar{X}_n - \mu)\| \geqq \varepsilon/\lambda_1\} = -\frac{1}{2\lambda_1^2}\Big/\|A_2\Sigma A_2'\|\,. \tag{3.9}$$

By (3.8) and (3.9), after letting $\alpha \to 0$ and $\lambda_1 \to 1$, and using Lemma 2.2, we get

$$\limsup_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{\|h(\bar{X}_n) - h(\mu)\| \geqq \varepsilon\} \leqq -\tfrac{1}{2}/\|A\Sigma A'\|\,. \tag{3.10}$$

Next, for $\beta > 0$, let $A_2^\beta(\eta_n) = \binom{A(\eta_n)}{\beta A_1(\eta_n)}$ be a nonsingular augmented matrix of $A(\eta_n)$. Note that

$$\begin{aligned} \{\|h(\bar{X}_n) - h(\mu)\| \geqq \varepsilon\} &= \lim_{\beta\to 0} \{\|A_2^\beta(\eta_n)A_2^{-1}A_2(\bar{X}_n - \mu)\| \geqq \varepsilon\} \\ &= \lim_{\beta\to 0} \{\max_{\|l\|=1} |l'A_2^\beta(\eta_n)A_2^{-1}A_2(\bar{X}_n - \mu)| \geqq \varepsilon\} \\ &= \lim_{\beta\to 0} \{\max_{l\in E_n(\beta)} |l'A_2(\bar{X}_n - \mu)| \geqq \varepsilon\} \\ &\supset \lim_{\beta\to 0} \{\max_{l\in F_k(\beta)} |l'A_2(\bar{X}_n - \mu)| \geqq \varepsilon\} \end{aligned} \tag{3.11}$$

for $k \leqq n$, where for $\beta \geqq 0$ $E_n(\beta)$ is defined to be the union of the spaces of $l'A_2^\beta(\eta_n)A_2^{-1}$ over all l such that $\|l\| = 1$ and $F_k(\beta) = \bigcap_{i=k}^\infty E_i(\beta)$. Note that with probability one $\lim_{n\to\infty} \eta_n = \mu$ and $\lim_{n\to\infty} E_n(\beta) = E(\beta) = \lim_{k\to\infty} F_k(\beta)$ for $\beta \geqq 0$, where $E(\beta) = \{u: u' = l'A_2^\beta A_2^{-1}, \|l\| = 1\}$. Since $F_k(\beta)$ is nondecreasing in k, $F_k(\beta) \subset E(\beta)$ for $\beta \geqq 0$. Note that the elements in $E(\beta)$ do not depend on the random variable $\bar{X}_n$, as do the elements in $F_k(\beta)$. Note also that $\lim_{\beta\to 0} E_n(\beta) = E_n(0)$ and hence that $\lim_{\beta\to 0} F_k(\beta) = F_k(0)$. (In fact the $E_n(\beta)$'s and the $F_k(\beta)$'s are "continuous" for β in $[0, \frac{1}{2}]$ in the sense that $\lim_{\beta\to\beta_0} E_n(\beta) = E_n(\beta_0)$ and $\lim_{\beta\to\beta_0} F_k(\beta) = F_k(\beta_0)$.) By (3.11)

$$\begin{aligned} \{\|h(\bar{X}_n) - h(\mu)\| \geqq \varepsilon\} &\supset \{\max_{l\in F_k(0)} |l'A_2(\bar{X}_n - \mu)| \geqq \varepsilon\} \\ &\supset \{|l'A_2(\bar{X}_n - \mu)| \geqq \varepsilon\} \end{aligned} \tag{3.11'}$$

for every $l \in F_k(0)$. By Theorem 3.1 for every $l \in F_k(0)$,

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{|l'A_2(\bar{X}_n - \mu)| \geqq \varepsilon\} = -\tfrac{1}{2}/l'A_2\Sigma A_2'l\,. \tag{3.12}$$

Since $AA_2^{-1} = [I_r, 0]$, as $\beta = 0$ $A_2^{\beta}A_2^{-1} = [\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}]$ and $E(\beta) = E(0) = \{u : \|u\| \leqq 1$ and the last $(k - r)$ components of u are $0\}$. Recall that $\lim_{k\to\infty} F_k(0) = E(0)$. By (3.11′) and (3.12), after letting $k \to \infty$ and $\alpha \to 0$, we have

$$(3.13) \qquad \liminf_{\varepsilon\to0} \liminf_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{\|h(\bar{X}_n) - h(\mu)\| \geqq \varepsilon\} \geqq -\tfrac{1}{2}/l_1' A\Sigma A' l_1$$

for all $l_1 \in E$, where E is a subset of R^r obtained from $E(0)$ by dropping the last $(k - r)$ components of the elements in $E(0)$. The desired result follows from (3.10) and (3.13) with the help of Lemma 2.2.

Let $Y_i = X_i X_i'$. Let the column vector formed by the columns of the lower triangular matrix of $X_i X_i'$ also be denoted by Y_i. The exact meaning of Y_i will be clear from the context. Assume that Y_i has finite mgf. Let Σ_Y be the covariance matrix of Y_i and $\Sigma_{(X,Y)}$ the covariance matrix of $(X_i', Y_i')'$. Recall that $\Sigma = \Sigma_X$.

THEOREM 3.4. *If $h: R^{k+k(k+1)/2} \to R^r$ has continuous partial derivatives, $A(x, s) = (\partial/\partial x)h(x, s)$ and $B(x, s) = (\partial/\partial s)h(x, s)$, where the matrix $[A(\cdot, \cdot), B(\cdot, \cdot)]$ is of rank r and $B = B(\mu, \Sigma) = 0$, then for each $\theta \in \Theta$,*

$$\lim_{\varepsilon\to0} \lim_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{\|h(\bar{X}_n, S_n) - h(\mu, \Sigma)\| \geqq \varepsilon\} = -\tfrac{1}{2}/\|A\Sigma A'\|\,,$$

where $S_n = \bar{Y}_n - \bar{X}_n \bar{X}_n'$ (= sample covariance matrix) and $A = A(\mu, \Sigma)$.

PROOF. Let $h_1(x, y) = h(x, s)$, where $y = s + xx'$. Then at $(x, y) = (\mu, \Sigma + \mu\mu')$, $(\partial/\partial y)h_1 = (\partial/\partial s)h = B = 0$ and $(\partial/\partial x)h_1 = (\partial/\partial x)h = A$. The result follows from Theorem 3.3 by noting $(A, B)\Sigma_{(X,Y)}(A, B)' = A\Sigma A'$.

THEOREM 3.5. *If $h: R^{k+k(k+1)/2} \to R^r$ has continuous partial derivatives, $A(x, s) = (\partial/\partial x)h(x, s)$ and $B(x, s) = (\partial/\partial s)h(x, s)$, where $A(\mu, \Sigma)$ is of rank r and $B(\mu, \Sigma) = 0$ and if S_n is a consistent estimator of Σ such that for small $\varepsilon > 0$,*

$$(3.14) \qquad \limsup_{n\to\infty} [P_\theta\{\|S_n - \Sigma\| \geqq \varepsilon\}/P_\theta\{\|A(\bar{X}_n - \mu)\| \geqq \varepsilon\}] < \infty$$

then for each $\theta \in \Theta$,

$$\limsup_{\varepsilon\to0} \limsup_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{\|h(\bar{X}_n, S_n) - h(\mu, \Sigma)\| \geqq \varepsilon\} \leqq -\tfrac{1}{2}/\|A\Sigma A'\|\,,$$

where $A = A(\mu, \Sigma)$.

PROOF. Write

$$(3.15) \qquad h(\bar{X}_n, S_n) - h(\mu, \Sigma) = A_n(\bar{X}_n - \mu) + B_n(S_n - \Sigma)\,.$$

For $\varepsilon > 0$ and $0 < \lambda < 1$, we have

$$\{\|h(\bar{X}_n, S_n) - h(\mu, \Sigma)\| \geqq \varepsilon\} \subset \{\|A_n(\bar{X}_n - \mu)\| \geqq \lambda\varepsilon\} \cup \{\|B_n(S_n - \Sigma)\| \geqq (1 - \lambda)\varepsilon\}\,.$$

But

$$\{\|B_n(S_n - \Sigma)\| \geqq (1 - \lambda)\varepsilon\} \subset \{\|B_n\| > (1 - \lambda)\} \cup \{\|S_n - \Sigma\| \geqq \varepsilon\}$$

and

$$\{\|A_n(\bar{X}_n - \mu)\| \geqq \lambda\varepsilon\} \subset \{\|A_n\binom{A}{\alpha A_1}^{-1}\| > \lambda_1\} \cup \{\|\binom{A}{\alpha A_1}(\bar{X}_n - \mu)\| \geqq \lambda\varepsilon/\lambda_1\}\,,$$

where A_1 is defined as the P_1 in Lemma 2.3 and $\lambda_1 > 1$. Since $\|B_n\| \to 0$ and $\|A_n\binom{A}{\alpha A_1}^{-1}\| \to 1$ as $\bar{X}_n \to \mu$ and $S_n \to \Sigma$, there is a neighborhood $N = N_1 \times N_2$ of (μ, Σ) such that $(\bar{X}_n, S_n) \in N$ implies $\|B_n\| < 1 - \lambda$ and $\|A_n\binom{A}{\alpha A_1}^{-1}\| < \lambda_1$. Thus $\|B_n\| \geqq 1 - \lambda$ or $\|A_n\binom{A}{\alpha A_1}^{-1}\| \geqq \lambda_1$ implies $S_n \notin N_2$ or $X_n \notin N_1$. Choose $\varepsilon > 0$ so small that $S_n \notin N_2$ or $\bar{X}_n \notin N_1$ implies $\{\|\binom{A}{\alpha A_1}(\bar{X}_n - \mu)\| \geqq \lambda\varepsilon/\lambda_1\}$ or $\{\|S_n - \Sigma\| \geqq \varepsilon\}$. Then

$$\begin{aligned} P_\theta\{\|h(\bar{X}_n, S_n) - h(\mu, \Sigma)\| \geqq \varepsilon\} & \\ \leqq P_\theta\{\|\binom{A}{\alpha A_1}(\bar{X}_n - \mu)\| \geqq \lambda\varepsilon/\lambda_1\} &+ P_\theta\{\|S_n - \Sigma\| > \varepsilon\} \\ \leqq P_\theta\{\|\binom{A}{\alpha A_1}(\bar{X}_n - \mu)\| > \lambda\varepsilon/\lambda_1\} &+ LP_\theta\{\|A(\bar{X}_n - \mu)\| > \varepsilon\}\,, \end{aligned}$$

as n sufficiently large, by (3.14) where $L>0$. Taking $\limsup_{\varepsilon\to 0}\limsup_{n\to\infty} 1/\varepsilon^2 n \log$ on both sides, noting $\log(a + b) \leqq \max(\log 2a, \log 2b)$ and letting $\lambda_1 \to 1$, $\lambda \to 1$ and $\alpha \to 0$, the result follows from a similar argument as in the first part of the proof of Theorem 3.2.

REMARK 1. The assumption of $B = 0$ in Theorems 3.4 and 3.5 is inherited from a similar situation in the discussions of RBAN estimators. This assumption is discussed in the remark on page 343 in [4] and is explicitly assumed in [7]. It says essentially that the dependence of $h(\bar{X}_n, S_n) - h(\mu, \Sigma)$ on $(S_n - \Sigma)$ is negligible as $n \to \infty$ (see (3.15)). It is readily shown (and demonstrated in [7]) that such assumption holds for a wide class of estimators generated by minimizing quadratic equations of $(\bar{X}_n - \mu)$ (e.g., $(\bar{X}_n - \mu)'S_n^{-1}(\bar{X}_n - \mu)$).

REMARK 2. The assumption of (3.14) in Theorem 3.5 is satisfied if S_n is the sample covariance matrix given in Theorem 3.4. This is an immediate consequence of Theorem 3.3 by noting that the conclusion of the latter can be written in the form of that of Theorem 3.1.

REMARK 3. By the definition of the matrix norm, $\|A\Sigma A'\|$ is the maximum eigenvalue of $A\Sigma A'$. In other words, the rate of convergence of $h(\bar{X}_n)$ (or $h(\bar{X}_n, S_n)$) to $h(\mu)$ depends only on the variance of a normalized linear combination of the components of $A\bar{X}_n$ which has maximal variance, or roughly, on the variance of a normalized linear combination of the components of $h(\bar{X}_n)$ (or $h(\bar{X}_n, S_n)$) which has maximal variance. This is an interesting observation.

4. Bahadur's theorem. To prove the efficiency of consistent estimators in Bahadur's sense, the following theorems are useful. The first theorem can be found in [1], page 252 and the second is an extension of the first to multidimensional case. Let $I(\theta)$ be the information matrix of X. Assume that $I(\theta)$ is nonsingular. Assume that $g: \Theta \to R^s$ has continuous first partial derivatives. Let $U = (\partial/\partial\theta)g(\theta)$.

THEOREM 4.1. *If* $s = 1$ (i.e., $g = g(\theta)$ *is real*) *and* T_n *is a consistent estimator* *of g, then for each $\theta \in \Theta$*

$$\liminf_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{|T_n - g| \geqq \varepsilon\} \geqq -\tfrac{1}{2}/U' I^{-1}(\theta) U .$$

THEOREM 4.2. *For positive integers, if T_n is a consistent estimator of g, then for each $\theta \in \Theta$,*

$$\liminf_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{\|T_n - g\| \geqq \varepsilon\} \geqq -\tfrac{1}{2}/\|U' I^{-1}(\theta) U\| .$$

PROOF. By (2.1) for all unit vector l,

$$\{\|T_n - g\| \geqq \varepsilon\} \supset \{|l'(T_n - g)| \geqq \varepsilon\} .$$

Combining this with Theorem 4.1, we have

$$\liminf_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{\varepsilon^2 n} \log P_\theta\{\|T_n - g\| \geqq \varepsilon\} \geqq -\tfrac{1}{2}/l' U' I^{-1}(\theta) U l ,$$

for all unit vector l. The result follows by taking maximum on the right-hand side over all unit vector l.

Note that if the equality in the inequality of Theorems 4.1 or 4.2 holds, then T_n, as an estimator of $g(\theta)$, is efficient in Bahadur's sense.

5. Efficiency of RBAN estimators. Assume that $\mu : \Theta\ (\subset R^r) \to \mu(\Theta) \subset R^k$ is one-to-one and bicontinuous and has continuous first partial derivatives. Assume that $\Sigma = \Sigma(\theta)$ is positive definite (this condition may be weakened, see, e.g., [5]). Let $V = V(\theta) = (\partial/\partial\theta)\mu(\theta)$. Assume that V is of rank r. Consider the function h defined in Theorems 3.3, 3.4 or 3.5, if $\hat\theta_n = h(\bar X_n)$ (or $h(\bar X_n, S_n)$) is a (regular) consistent estimator of θ, then (see, e.g., [4] or [7]) $\hat\theta_n$ is asymptotically normal, i.e., $n^{\frac{1}{2}}(\hat\theta_n - \theta) \to N(0, A\Sigma A')$ in distribution and $h(\mu)$ (or $h(\mu, \Sigma)) = \theta$, where A is defined in the aforementioned theorems. Furthermore

$$\hat\theta_n \quad \text{is RBAN iff} \quad A = (V'\Sigma^{-1}V)^{-1}V'\Sigma^{-1} . \tag{5.1}$$

See, e.g., [4], [5] or [7]. RBAN is defined in the sense that if $d_n = d(\bar X_n)$ or $d(\bar X_n, S_n)$ satisfying the same conditions as h is any other consistent estimator of θ, then $\Sigma_d - \Sigma_h$ is positive semi-definite, where Σ_d and Σ_h are respectively the asymptotical covariance matrix of $n^{\frac{1}{2}}d_n$ and $n^{\frac{1}{2}}\hat\theta_n$. Note that if $\hat\theta_n$ is RBAN, then $\Sigma_h = (V'\Sigma^{-1}V)^{-1}$. Methods to generate RBAN estimators can be found in [4] and [5], among others.

EXAMPLE 1. *Exponential family.* Let $X, X_1, X_2, \cdots,$ be i.i.d. random k-vectors with pdf with respect to some σ-finite measure,

$$p_\theta(x) = \exp(q'(\theta)x - b(q(\theta))) , \tag{5.2}$$

where $\theta \in \Theta \subset R^r$ with $r \leqq k$. Assume that Θ is convex and open and that both $b(\cdot)$ and $q(\cdot)$ have continuous second partial derivatives. It is readily shown that $\mu = \mu(\theta) = (\partial/\partial q)b(q)$ and $\Sigma = \Sigma(\theta) = (\partial/\partial q)((\partial/\partial q)b(q))'$. Let $\hat\theta_n = h(\bar X_n)$ be a solution to

$$W'(\bar X_n - \mu) = 0 , \tag{5.3}$$

where $W = (\partial/\partial\theta)q(\theta)$. Note that the left-hand side of (5.3) is the partial derivatives with respect to θ of the exponent in (5.2). From (5.3), using implicit function theorem, it is found $A = (\partial/\partial x)h(x)|_{x=\mu} = (V'\Sigma^{-1}V)^{-1}V'\Sigma$, where $V = (\partial/\partial\theta)\mu = V(\theta)$. By (5.1), $\hat{\theta}_n = h(\bar{X}_n)$ is RBAN. (A more detailed discussion of this can be found in [7].) Further, by Theorem 3.3,

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{\varepsilon^2 n}\log P_\theta\{\|\hat{\theta}_n - \theta\| \geqq \varepsilon\} = -\tfrac{1}{2}/\|(V'\Sigma^{-1}V)^{-1}\| . \tag{5.4}$$

Next, by (5.3), the information matrix of X is given by

$$I(\theta) = W'\Sigma W = V'\Sigma^{-1}V , \tag{5.5}$$

because $V = \Sigma W$. By Theorem 4.2 with g being the identity function and (5.4) and (5.5), $\hat{\theta}_n$ is efficient in Bahadur's sense.

Note that $\hat{\theta}_n$ is also a maximum likelihood estimator (if one exists). Hence the result of this example confirms the conjecture mentioned on page 252 of [1] for distribution in exponential family.

Note also that Theorem 3.5 is useful in proving Bahadur's efficiency if the minimum modified χ^2 method is used to generate the RBAN estimator, e.g., in the multinomial case.

The following example shows that an RBAN estimator is not necessarily efficient in Bahadur's sense if the distribution is not in the exponential family.

EXAMPLE 2. *Double exponential distribution.* Consider the i.i.d. random variables $X, X_1, X_2, \cdots$ with pdf

$$f(x;\theta) = \tfrac{1}{2}e^{-|x-\theta|} , \qquad -\infty < x < \infty ,$$

where $\theta \in (-\infty, \infty)$. Consider $h(X_n) = \bar{X}_n$ as an estimator of $\mu(\theta) = \theta$. Then $A = V = 1$. By (5.1), $\bar{X}_n$ is RBAN. But by Theorem 4.2 of [6], $\bar{X}_n$ is not efficient in Bahadur's sense since the sample median is strictly more efficient than $\bar{X}_n$. See also Example 6.1 in [3].

Acknowledgment. I wish to thank Professor Robert A. Wijsman for suggesting the problem. I have been greatly benefited by his lecture note in the RBAN estimators [7]. I would also like to thank a referee for pointing out errors in the original manuscript and for comments which lead to improvement in the presentation of the paper.

REFERENCES

[1] BAHADUR, R. R. (1960). Asymptotic efficiency of tests and estimates. *Sankhyā* **22** 229-252.

[2] BAHADUR, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303-324.

[3] BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics*. SIAM, Philadelphia.

[4] CHIANG, C. L. (1956). On regular best asymptotically normal estimates. *Ann. Math. Statist.* **27** 336-351.

[5] FERGUSON, T. S. (1958). A method of generating best asymptotically normal estimates with applications to the estimation of bacterial densities. *Ann. Math. Statist.* **29** 1046-1062.

[6] Fu, James C. (1975). The rate of convergence of consistent point estimators. *Ann. Statist.* **3** 234-239.

[7] Wijsman, R. A. (1971). Regular best asymptotically normal estimators. Lecture notes, Fall, 1971. Univ. Illinois, Urbana, Illinois.

Department of Mathematics
Howard University
Washington, D.C. 20059