CONVERGENCE RATES OF ESTIMATORS OF A FINITE PARAMETER: HOW SMALL CAN ERROR PROBABILITIES BE?*

by

Andrew L. Rukhin
Purdue University, W. Lafayette, IN

Mimeograph Series #81-39

Department of Statistics
Division of Mathematical Sciences
August 1981

*This work was supported by National Science Foundation grant MCS-8101670.

ABSTRACT

In this paper we investigate the rates of convergence to zero of error probabilities in the estimation problem of a finite-valued parameter. It is shown that if a consistent estimator attains Bahadur's bound for the probability of incorrect decision at some parametric point then the error probability does not tend to zero exponentially fast for some other value of the parameter. We evaluate the minimal possible value of the rate of this probability at a fixed parametric point for all asymptotically minimax procedures, and establish a necessary and sufficient condition for any of these procedures to have a constant risk. A simple example demonstrating the asymptotical inadmissibility of the usual maximum likelihood estimator is constructed.

AMS 1980 subject classifications. Primary 62F12, Secondary 62F10, 62C20 62B10

Key words and phrases: Finite-valued parameter, error probabilities, weighted maximum likelihood estimator, maximin procedures.

SUMMARY

CONVERGENCE RATES OF ESTIMATORS
OF A FINITE PARAMETER: HOW SMALL CAN
ERROR PROBABILITIES BE?

In this paper we investigate the rates of convergence to zero of error probabilities in the estimation problem of a finite-valued parameter. It is shown that if a consistent estimator attains Bahadur's bound for the probability of incorrect decision at some parametric point then the error probability does not tend to zero exponentially fast for some other value of the parameter. We evaluate the minimal possible value of the rate of this probability at a fixed parametric point for all asymptotically minimax procedures, and establish a necessary and sufficient condition for any of these procedures to have a constant risk. A simple example demonstrating the asymptotical inadmissibility of the usual maximum likelihood estimator is constructed.

AMS 1980 subject classifications. Primary 62F12, Secondary 62F10, 62C20 62B10

Key words and phrases: Finite-valued parameter, error probabilitites, weighted maximum likelihood estimator, maximin procedures.

1. Introducution

Let P_θ, $\theta \in \Theta = \{1,\ldots,m\}$ be a finite collection of (different) probability distributions with positive densities f_θ, and let $\underline{x} = (x_1,\ldots,x_n)$ be a random sample from a member of this family. If $\delta = \{\delta^n = \delta^n(\underline{x}),\ n=1,2,\ldots\}$ is a sequence of estimators of the finite-valued parameter θ then the error probability, $P_\theta(\delta^n \neq \theta)$, is an important characteristic of δ. The rate at which this probability tends to zero for a consistent estimator δ is of interest. Typically, this rate is exponential, so that the following quantities (asymptotic efficiencies), $G(\theta,\delta) = -\liminf n^{-1} \log P_\theta(\delta^n \neq \theta)$, are positive and finite.

In this paper we are concerned with how small the error probabilities $P_\theta(\delta^n \neq \theta)$ can be, i.e., how large the quantities $G(\theta,\delta)$ can be, for consistent estimator δ.

It follows from results of Bahadur (1960), (1967), (1971) and Bahadur, Gupta and Zabell (1980) that for any consistent δ

$$G(\theta,\delta) \leq \min_{\eta:\eta\neq\theta} K(\eta,\theta), \tag{1.1}$$

where $K(\eta,\theta) = E_\eta[f_\eta(x)/f_\theta(x)]$ is the Kullback-Leibler information number. However equality in (1.1) cannot be attained for all θ, since if δ_m is an asymptotically maximin, then

$$\min_\theta G(\theta,\delta_m) < \min_{\eta:\eta\neq\theta} K(\eta,\theta).$$

Indeed it is known that if for all δ

$$\min_\theta G(\theta,\delta_m) \geq \min_\theta G(\theta,\delta),$$

then

$$\min_\theta G(\theta,\delta_m) = -\max_{\eta\neq\theta} \inf_{0<s<1} \log E_\theta[f_\eta(x)/f_\theta(x)]^s \tag{1.2}$$

(see Krafft and Puri (1974) and Renyi (1969)).

By convexity

$$\inf_{0<s<1} \log E_\theta[f_\eta(x)/f_\theta(x)]^s > - E_\eta \log[f_\eta(x)/f_\theta(x)] = -K(\eta,\theta).$$

This inequality also follows from the comparison of most powerful test and minimax test of hypothesis f_θ against alternative f_η (see Chernoff 1972) p.49).

In this paper it is shown that if (1.1) reduces to an equality for some θ, then

$$\min_{\eta:\eta\neq\theta} G(\eta,\delta) = 0 .$$

This result is proved in Section 2 by means of a useful lemma generalizing (1.2).

For all n there exists unique maximin estimator δ^n, and it has a constant risk $P_\theta(\delta^n \neq \theta)$ (see Wald (1950), Theorems 5.3 and 5.4). However there is a wealth of asymptotically maximin estimators, and it is of interest to find among them a procedure, which minimizes the risk at some parametric point, i.e., which maximizes $G(\theta,\delta)$ at some θ.

In Section 3 we study efficiencies $G(\theta,\delta)$ for an asymptotically maximin estimator δ. A sharp upper bound for these quantities is established, and a method of finding an estimator, which attains this bound, is given. The latter procedure belongs to the class of weighted maximum likelihood estimators with prior probabilities depending on the sample size. We also give a necessary and sufficient condition for the constancy of risk of all maximin rules. A simple example, which demonstrates the asymptotical inadmissibility of the usual maximum likelihood estimator for the gain $G(\theta,\delta)$, is constructed.

2. Lower Bounds for Error Probabilities

We start with the following lemma.

Lemma. *Let* $c_1,\ldots,c_m$ *be real constants. Then for any estimator* δ

$$\min_{\theta} [G(\theta,\delta) - c_\theta] \le - \max_{\eta\neq\theta} \inf_{s>0}[c_\theta + s(c_\eta - c_\theta) + \log E_\theta[f_\eta(x)/f_\theta(x)]^s]. \quad (2.1)$$

For any subset N of Θ

$$\min_{\theta\in N} [G(\theta,\delta) - c_\theta] \le - \max_{\substack{\eta\neq\theta \\ \eta,\theta\in N}} \inf_{s>0}[c_\theta + s(c_\eta - c_\theta) + \log E_\theta[f_\eta(x)/f_\theta(x)]^s]. \quad (2.2)$$

The first part of Lemma follows from (1.2) by replacing probability distributions P_θ in Krafft and Puri (1974) by positive measures $e^{c_\theta}P_\theta$. By assigning prior probabilities which give zero mass to elements of Θ outside a certain subset N we obtain (2.2).

There exist procedures for which (2.1) and (2.2) hold with equality. These asypmtotically maximin estimators (for the shifted efficiencies $G(\theta,\delta) - c_\theta$) can be taken to be weighted maximum likelihood estimators $\hat{\delta}_c$, which, ignoring ties, maximize $\exp(nc_\theta) \prod_1^n f_\theta(x_j)$.

It is easy to see that

$$\left[P_\theta \{\hat{\delta}_c \neq \theta\}\right]^{1/n} \sim \max_{\eta:\eta\neq\theta} \left[P_\theta \{e^{nc_\eta} \prod_1^n f_\eta(x_j) > e^{nc_\theta} \prod_1^n f_\theta(x_j)\}\right]^{1/n}$$

$$= \max_{\eta:\eta\neq\theta} \inf_{s>0} e^{s(c_\eta - c_\theta)} E_\theta[f_\eta(x)/f_\theta(x)]^s,$$

because of Chernoff's theorem. Therefore

$$G(\theta,\hat{\delta}_c) = - \max_{\eta:\eta\neq\theta} \inf_{s>0} [s(c_\eta - c_\theta) + \log E_\theta[f_\eta(x)/f_\theta(x)]^s] \quad (2.3)$$

and equalities hold in (2.1) or (2.2).

Also, if

$$c_\eta - c_\theta + E_\eta \log[f_\eta(x)/f_\theta(x)] \ge 0,$$

which means that the derivative of the convex function

$s(c_\eta - c_\theta) + \log E_\theta[f_\eta(x)/f_\theta(x)]^s$ is nonnegative at $s = 1$, then

$$\inf_{s>0}[c_\theta + s(c_\eta - c_\theta) + \log E_\theta[f_\eta(x)/f_\theta(x)]^s]$$

$$= \inf_{0<s<1} [c_\eta + s(c_\theta - c_\eta) + \log E_\eta[f_\theta(x)/f_\eta(x)]^s].$$

It is easy to see that

$$\inf_{s>0}[c_\theta + s(c_\eta - c_\theta) + \log E_\theta[f_\eta(x)/f_\theta(x)]^s] \leq \min\{c_\theta, c_\eta\}, \tag{2.4}$$

and that the estimator $\hat{\delta}_c$ is consistent if and only if for all θ, η, $\theta \neq \eta$

$$c_\eta - c_\theta + K(\eta,\theta) > 0 .$$

<u>Theorem 1</u>. <u>Let</u> δ <u>be an estimator such that</u>

$$G(\theta,\delta) = \min_{\eta:\eta\neq\theta} K(\eta,\theta)$$

<u>for some</u> θ. <u>Then</u>

$$\min_{\eta:\eta\neq\theta} G(\eta,\delta) = 0 .$$

<u>Proof</u>. Let

$$G(\theta,\delta) = \min_{\eta:\eta\neq\theta} K(\eta,\theta) = K(\xi,\theta),$$

and in (2.1) put $c_\theta = K(\xi,\delta)$, $c_\eta = 0$ for $\eta \neq 0$. Then the left side of (2.1) is equal to

$$\min[0, \min_{\eta:\eta\neq\theta} G(\eta,\delta)] = \min_{\eta:\eta\neq\theta} G(\eta,\delta).$$

But because of (2.4)

$$\max_{\rho\neq\eta} \inf_{s>0}[c_\rho + s(c_\eta - c_\rho) + \log E_\rho[f_\eta(x)/f_\rho(x)]^s] \leq \max_{\rho\neq\eta} \min\{c_\rho, c_\eta\} = 0$$

and

$$\inf_{s>0}[c_\theta + s(c_\xi - c_\theta) + \log E_\theta[f_\xi(x)/f_\theta(x)]^s]$$

$$= K(\xi,\theta) + \inf_{0<s<1} [-sK(\xi,\theta) + \log E_\theta[f_\xi(x)/f_\theta(x)]^s]$$

$$= K(\xi,\theta) - K(\xi,\theta) = 0.$$

Thus

$$\max_{\rho\neq\eta} \inf_{s>0}[c_\rho + s(c_\eta - c_\rho) + \log E_\rho[f_\eta(x)/f_\rho(x)]^s] = 0$$

and because of (2.1)

$$\min_{\eta:\eta\neq\theta} G(\eta,\delta) = 0.$$

Remark. A slight modification of the proof shows that if for some positive ε

$$G(\theta,\delta) = \min_{\eta:\eta\neq\theta} K(\eta,\theta) + \varepsilon = K(\xi,\theta) + \varepsilon,$$

then

$$\min_{\eta:\eta\neq\theta} G(\eta,\delta) \leq \varepsilon^2 / \mathrm{Var}_\xi[\log(f_\xi(x)/f_\theta(x))],$$

where the variance of this inequality is assumed to be finite. Thus ε-closeness to the bound (1.1) implies that the function $G(\eta,\delta)$ is of order ε^2 for some η.

3. Maximin Procedures and Weighted Maximum Likelihood Estimators

In this section we are interested in the behavior of a maximin procedure δ_m, i.e. a procedure such that

$$\min_\theta G(\theta,\delta_m) = -\max_{\theta\neq\eta} \inf_{s>0} \log E_\theta[f_\eta(x)/f_\theta(x)]^s = g. \quad (3.1)$$

More exactly how large can be the asymptotic efficiency $G(\theta,\delta_m)$ under condition (3.1), and how does one evaluate $\sup_{\delta_m} G(\theta,\delta_m)$ for a particular value of θ and find δ_m for which this supremum is attained?

To answer these questions we define real $c(\eta,\theta)$ for $\theta \neq \eta$ as the real solution of the following equation

$$\inf_{s>0}[c(\eta,\theta) + \log E_\eta[f_\theta(x)/f_\eta(x)]^s] = -g.$$

We also denote

$$\bar{c}_\theta = \min_{\eta:\eta\neq\theta} c(\eta,\theta) .$$

<u>Theorem 2</u>. <u>For any</u> θ

$$\sup_{\delta_m} G(\theta, \delta_m) = g + \bar{c}_\theta. \tag{3.2}$$

<u>Proof</u>. First we show that for any maximin procedure δ_m

$$G(\theta,\delta_m) \leq g + \bar{c}_\theta .$$

For any η, $\eta \neq \theta$

$$\inf_{0<s<1}[- s\bar{c}_\theta + \log E_\theta[f_\eta(x)/f_\theta(x)]^s]$$

$$\geq \inf_{0<s<1} [- c(\eta,\theta) + \log E_\theta[f_\eta(x)/f_\theta(x)]^s] = -c_{\eta\theta} - g$$

and

$$\inf_{0<s<1}[- s\bar{c}_\theta + \log E_\theta[f_\eta(x)/f_\theta(x)]^s]$$

$$\leq - \bar{c}_\theta + \inf_{0<s<1} [s\bar{c}_\theta + \log E_\eta[f_\theta(x)/f_\eta(x)]^s]$$

$$\leq - \bar{c}_\theta + \inf_{0<s<1} [c(\eta,\theta)+ \log E_\eta[f_\theta(x)/f_\eta(x)]^s] = - \bar{c}_\theta - g.$$

Combining these inequalities we see that

$$-c(\eta,\theta) - g \leq \inf_{0<s<1} [-s\bar{c}_\theta + \log E_\theta[f_\eta(x)/f_\theta(x)]^s] \leq -\bar{c}_\theta - g.$$

Therefore if ξ is defined by the formula $c(\xi,\theta) = \bar{c}_\theta$, then

$$\inf_{0<s<1} [-s\bar{c} + \log E_\theta[f_\xi(x)/f_\theta(x)]^s]$$

$$= \max_{\eta:\eta\neq\theta} \inf_{0<s<1} [-s\bar{c}_\theta + \log E_\theta[f_\eta(x)/f_\theta(x)]^s] = -\bar{c}_\theta - g. \tag{3.3}$$

Because of Lemma 2 and (3.3)

$$\min[G(\theta,\delta_m) - \bar{c}_\theta, G(\xi,\delta_m)]$$

$$\leq -\inf_{s\geq 0}[\bar{c}_\theta - s\bar{c}_\theta + \log E_\theta[f_\xi(x)/f_\theta(x)]^s]$$

$$= -\bar{c}_\theta + \bar{c}_\theta + g = g.$$

But because of (3.1), $G(\xi, \delta_m) \geq g$, so that

$$G(\theta,\delta_m) - \bar{c}_\theta \leq g.$$

Now we exhibit a maximin procedure δ such that for a particular value of θ

$$G(\theta,\delta) = g + \bar{c}_\theta .$$

Namely, define $\delta = \hat{\delta}_c$ to be the weighted maximum likelihood estimator with $c_\theta = \bar{c}_\theta$ and $c_\eta = 0$ for $\eta\neq\theta$.

Because of (2.3) and (3.3)

$$G(\theta,\delta) = -\max_{\eta:\eta\neq\theta} \inf_{s>0}[-s\bar{c}_\theta + \log E_\theta[f_\eta(x)/f_\theta(x)]^s] = g + \bar{c}_\theta$$

and for $\xi\neq\theta$

$$G(\xi,\delta) = -\max_{\eta:\eta\neq\xi} \inf_{s>0}[sc_\eta + \log E_\xi[f_\eta(x)/f_\xi(x)]^s]$$

$$\geq -\max_{\eta:\eta\neq\xi} \inf_{s>0}[c(\xi,\eta) + \log E_\xi[f_\eta(x)/f_\xi(x)]^s] = g .$$

Thus δ is a maximin procedure, for which the supremum in (3.2) is attained. Theorem 2 is proven.

Corollary. Let

$$\Theta_0 = \{\theta: \inf_{s>0} \log E_\theta[f_\eta(x)/f_\theta(x)]^s = -g \text{ for some } \eta\}.$$

Every maximin procedure δ_m has a constant asymptotic efficiency $G(\theta,\delta_m) \equiv g$, if and only if $\Theta_0 = \Theta$.

This result is clear since $\bar{c}_\theta = 0$ if and only if $\theta \in \Theta_0$. Thus if $\Theta_0 \neq \Theta$ any maximin procedure with constant risk is inadmissible.

Theorem 2 also provides a method which in some situations (namely when $\Theta_0 \neq \Theta$) allows one to improve upon traditional maximum likelihood estimator. To illustrate this method let us consider an example in which f_θ is a normal density with mean a_θ and variance 1.

Easy calculation shows that

$$-\inf_{s>0}[sc + \log E_\theta[f_\eta(x)/f_\theta(x)]^s] = [(a_\eta - a_\theta)^2 - 2c]^2(a_\eta - a_\theta)^{-2}/8$$

for $c < (a_\eta - a_\theta)^2/2$. Therefore

$$g = \min_{\eta\neq\theta} (a_\eta - a_\theta)^2/8, \quad c_{\theta\eta} = [(a_\eta - a_\theta)^2 - 2|a_\eta - a_\theta|\,(2g)^{1/2}]/2.$$

Using Theorem 2, for any maximin procedure δ

$$G(\theta,\delta) \leq g + \min_{\eta:\eta\neq\theta} (a_\eta - a_\theta)^2/2 - (2g)^{1/2} \min_{\eta:\eta\neq\theta} |a_\eta - a_\theta|, \tag{3.4}$$

and for the maximum likelihood estimator $\hat{\delta}$

$$G(\theta,\hat{\delta}) = \min_{\eta:\eta\neq\theta} (a_\eta - a_\theta)^2/8,$$

which of course does not exceed (3.4). Moreover $G(\theta,\hat{\delta})$ is equal to the right side of (3.4) if and only if

$$\min_{\eta:\eta\neq\theta} (a_\eta - a_\theta)^2/8 = g,$$

i.e. when $\Theta_0 = \Theta$. If $\Theta_0 \neq \Theta$, then for any $\theta \notin \Theta_0$ there exists a maximin estimator δ such that

$$G(\theta,\delta) > G(\theta,\hat{\delta}).$$

In fact $\hat{\delta}$ can be inadmissible with respect to the gain $G(\theta,\delta)$. Let $m = 3$, $a_1 < a_2 < a_3$, and, say, $a_1 + a_3 < 2a_2$. Then $\hat{\delta} = 1$, $\bar{x} < (a_1 + a_2)/2$; $= 3$, $\bar{x} > (a_2 + a_3)/2$; $= 2$, otherwise, and the weighted maximum likelihood estimator δ has the form $\delta = 1$, $\bar{x} < (3a_2 - a_3)/2$; $= 3$, $\bar{x} > (a_2 + a_3)/2$; $= 2$, otherwise. Here $\bar{x} = \sum x_j/n$. Simple calculation shows that

$$G(1,\delta) = (3a_2 - 2a_1 - a_3)^2/8 > (a_2 - a_1)^2/8 = G(1,\hat{\delta})$$

and for $\theta \neq 1$

$$G(\theta,\delta) = G(\theta,\hat{\delta}) = (a_3 - a_2)^2/8,$$

which implies inadmissibility of the maximum likelihood estimator.

Since the maximum likelihood estimator is the Bayes estimator with respect to the uniform prior distribution over Θ, it is admissible for any finite sample size. Indeed

$$P_1\{\hat{\delta}^n \neq 1\} = \phi(-bn^{1/2}/2) > P_1\{\delta^n \neq 1\} = \phi((a/2 - b)n^{1/2}),$$
$$P_2\{\hat{\delta}^n \neq 2\} = \phi(-bn^{1/2}/2) + \phi(-an^{1/2}) < P_2\{\delta^n \neq 2\} = 2\phi(-an^{1/2}/2),$$
$$P_3\{\hat{\delta}^n \neq 3\} = \phi(-an^{1/2}/2) = P_3\{\delta^n \neq 3\},$$

where $a = a_3 - a_2$, $b = a_2 - a_1$ and ϕ is the standard normal distribution function. Thus δ does not improve upon $\hat{\delta}$ for any n. However

$$\lim \log P_2(\hat{\delta}^n \neq 2)/\log P_2(\delta^n \neq 2) = 1,$$

while

$$\lim \log P_1(\hat{\delta}^n \neq 1)/\log P_1(\delta^n \neq 1) = (2 - a/b)^2 > 1.$$

Therefore for large values of n, δ is preferable to $\hat{\delta}$.

In Table 1 the efficiencies $e_n = -n^{-1} \log P_\theta(\delta^n \neq \theta)$ and $d_n = n^{-1} \log P_\theta(\hat{\delta}^n \neq \theta)$ of procedures δ and $\hat{\delta}$ are evaluated for different sample sizes when a = b/2 =1. In this case the estimator δ is considerably better than $\hat{\delta}$ at $\theta = 1$ with their efficiencies almost equal at $\theta = 2$ even for relatively small sample sizes.

Table 1

Efficiencies e_n and d_n of Estimators δ and $\hat{\delta}$ for Different Sample Sizes

	$\theta = 1$		$\theta = 2$		$\theta = 3$
n	e_n	d_n	e_n	d_n	$e_n = d_n$
2	2.039	1.271	0.368	0.572	0.714
4	1.652	0.946	0.287	0.427	0.460
6	1.506	0.823	0.252	0.357	0.367
10	1.377	0.715	0.217	0.285	0.287
50	1.178	0.558	0.156	0.170	0.170
100	1.155	0.532	0.144	0.150	0.150
1000	1.129	0.504	0.128	0.128	0.128
∞	1.125	0.500	0.125	0.125	0.125

REFERENCES

Bahadur, R.R. (1960). Asymptotic efficiency of tests and estimates. Sankhyā, 22 229-252.

Bahadur, R.R. (1967). Rates of convergence of estimates and test statistics. Ann. Math. Statist. 38 304-324.

Bahadur, R.R. (1971). Some Limit Theorems in Statistics. Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia.

Bahadur, R.R., Gupta, J.C. and Zabell, S.L. (1980). Large Deviations, Tests, and Estimates, in "Asymptotic Theory of Statistical Tests and Estimation", edited by I.M. Chakravarti, Academic Press, 33-64.

Chernoff, H. (1972). Sequential Analysis and Optimal Design. Regional Conference Series in Appied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia.

Krafft, O. and Puri, M.L. (1974). The asymptotical behavior of the minimax risk for multiple decision problems. Sankhyā, 36, 1-12.

Renyi, A. (1969). On some problems of statistics from the point of view of information theory, Proceedings of the Colloquium on Information Theory, Debrecen 343-357, also in Selected Papers of Alfred Renyi, Academiai Kiado, Budapest, 1976, Vol. 3, 560-576.

Wald, A. (1950). Statistical Decision Functions. New York: John Wiley.