The Annals of Statistics
1975, Vol. 3, No. 1, 228–233

ON A LOWER BOUND FOR MOMENTS OF POINT ESTIMATORS[1]

BY R. Z. HASMINSKII AND I. A. IBRAGIMOV

Institute of Information Transmission, Moscow; Leningrad State University

We consider the problem of estimating an unknown parameter θ on the basis of independent identically distributed observations with a common density $f(x, \theta)$ and give some lower bounds for the accuracy of estimates of θ expressed in terms of the Hellinger distance

$$\rho(\theta; \theta') = \int_{\mathscr{X}} (f^{\frac{1}{2}}(x; \theta) - f^{\frac{1}{2}}(x; \theta'))^2 \, d\nu .$$

1. Introduction and results. Let $X_1, X_2 \cdots$ be a sequence of independent identically distributed random variables (observations) taking their values in a measurable space $(\mathscr{X}, \mathscr{R})$ with common distribution $\mathscr{P}_\theta$. We suppose that $\mathscr{P}_\theta$ depends upon an unknown parameter $\theta \in \Theta$ and that Θ is an open subset of R^k. Denote the n-fold Cartesian product space by $(\mathscr{X}^n, \mathscr{R}^n)$ and the n-fold product measure $\mathscr{P}_\theta \times \cdots \times \mathscr{P}_\theta$, by $\mathscr{P}_\theta^n$. We write P_θ instead of $\mathscr{P}_\theta^\infty$; $E_\theta(\cdot)$ denotes mathematical expectation relative to P_θ.

Suppose that there exists a measure ν on $\mathscr{R}$ such that all $\mathscr{P}_\theta$ are absolutely continuous relative to the measure ν and

$$\frac{d\mathscr{P}_\theta}{d\nu} = f(x; \theta) , \qquad x \in \mathscr{X}, \ \theta \in \Theta .$$

Let $\nu^n \equiv \nu \times \cdots \times \nu$ and

$$\frac{d\mathscr{P}_\theta^n}{d\nu^n} = f_n(x^n; \theta) = \prod_1^n f(x_j; \theta) , \qquad x^n = (x_1, \cdots, x_n) \in \mathscr{X}^n .$$

The Hellinger distance

$$\tilde\rho(\mathscr{P}_\theta, \mathscr{P}_{\theta'}) = (\int_{\mathscr{X}} |f^{\frac{1}{2}}(x; \theta) - f^{\frac{1}{2}}(x; \theta')|^2 \, d\nu)^{\frac{1}{2}}$$

between the measures $\mathscr{P}_\theta$ and $\mathscr{P}_{\theta'}$ induces the distance

$$\tilde\rho(\theta; \theta') = \tilde\rho(\mathscr{P}_\theta; \mathscr{P}_{\theta'})$$

between the parametric points θ and θ'. Let $|\theta - \theta'|$ be a distance between θ and θ' in R^k. Assuming certain regularity conditions it is proved in [2] that if for some $0 < \alpha \leqq \beta$,

$$(1) \qquad K_1(\theta)|\theta - \theta'|^\alpha \geqq \rho(\theta; \theta') \geqq K_2(\theta)|\theta - \theta'|^\beta , \qquad \rho(\theta; \theta') = \tilde\rho^2(\theta; \theta') ,$$

then there exist estimators t_n of θ such that

$$(2) \qquad \lim_{n\to\infty} n^{\lambda m} E_\theta |t_n - \theta|^m < \infty$$

for all $m > 0$ and $0 < \lambda < 1/\beta$. If $\alpha = \beta$ it is possible to let $\lambda = 1/\beta$.

Received December 1973.

[1] This paper was written when the second author was at the Department of Statistics, University of North Carolina at Chapel Hill.

In this paper we prove inequalities which are complementary to (2).

THEOREM 1. *Suppose that*

$$\rho(\theta; \theta') \leqq K(\theta)|\theta - \theta'|^{\alpha}, \qquad \alpha > 0,$$

and let T_n denote an estimate of θ satisfying

$$S_n^{(m)}(\theta; T_n) = S_n^{(m)}(\theta) = E_\theta |T_n - \theta|^m,$$
$$\bar{S}_n^{(m)}(\theta; T_n) = \bar{S}_n^{(m)}(\theta) = \inf_{|u|=1} S_n^{(m)}(\theta + (8K(\theta)n)^{-1/\alpha} u).$$

Then

$$\liminf_{n\to\infty} n^{m/\alpha}(S_n^{(m)}(\theta) + \bar{S}_n^{(m)}(\theta)) \geqq 2^{-4m-1}(8K(\theta))^{-m/\alpha}. \tag{3}$$

The analogous result also holds in a more general situation of sequential estimation. This estimation procedure is as follows. We are given: 1) a stopping time τ, a random variable with positive integer values such that the events $\{\tau = n\}$ are measurable with respect to the σ-algebra generated by $(X_1, \cdots, X_n)$ and $P_\theta\{\tau < \infty\} = 1$ (we shall suppose for the sake of brevity that the σ-algebra of events generated by $(X_1, \cdots, X_n)$ coincides with $\mathscr{B}^n$); 2) a sequence of statistics $T = \{T_n(X_1, \cdots, X_n)\}$. As an estimate of the parameter θ we use the random variable $T_\tau(X_1, \cdots, X_\tau)$. Following Ju. V. Linnik, we call the pair $d = [T, \tau]$ a sequential estimation plan.

Let $d = [T, \tau]$ be a sequential estimation plan. Define

$$S^{(m)}(\theta; d) = S^{(m)}(\theta) = E_\theta |T_\tau - \theta|^m,$$
$$\bar{S}^{(m)}(\theta; d) = \bar{S}^{(m)}(\theta) = \inf_{|u|=1} S^{(m)}(\theta + (200K(\theta)n)^{-\frac{1}{2}} u).$$

THEOREM 2. *Under the condition of Theorem* 1, *for all sequential estimation plans $d = [T, \tau]$ with $E_\theta \tau \leqq n$, $\theta \in \Theta$,*

$$\liminf_{n\to\infty} n^{m/\alpha}(S^{(m)}(\theta) + \bar{S}^{(m)}(\theta)) \geqq 2^{-4m-1}(200K(\theta))^{-m/\alpha}. \tag{4}$$

Both Theorems 1 and 2 are almost immediate consequences of the following:

THEOREM 3. *Let $d = [T, \tau]$ be a sequential estimation plan. Then for all $\theta, \theta' \in \Theta$ with $\rho(\theta; \theta') \cdot \max\{E_\theta \tau, E_{\theta'} \tau\} \leqq 200^{-1}$ and all $m \geqq 1$*

$$S^{(m)}(\theta) + S^{(m)}(\theta') \geqq 2^{-4m-1}|\theta - \theta'|^m. \tag{5}$$

If in addition $\tau \equiv n$, then for all $\theta, \theta' \in \Theta$ with $n\rho(\theta; \theta') \leqq \frac{1}{8}$ and all $m \geqq 1$

$$S^{(m)}(\theta) + S^{(m)}(\theta') \geqq 2^{-4m-1}|\theta - \theta'|^m. \tag{6}$$

2. Proof of Theorem 3. If $m > 1$ then

$$S^{(m)}(\theta) + S^{(m)}(\theta') \geqq (S^{(1)}(\theta)) + (S^{(1)}(\theta'))^m$$
$$\geqq 2^{m-1}(S^{(1)}(\theta) + S^{(1)}(\theta'))^m$$

so we need only prove the theorem for the case $m = 1$.

Let

$$S^{(1)}(\theta) = S(\theta), \qquad E_\theta T_\tau = M(\theta), \qquad M(\theta) - \theta = b(\theta).$$

Taking into account the measurability of $\{\tau = n\}$ relative to the σ-algebra generated by $X_1, \cdots, X_n$ we may consider $\{\tau = n\}$ as a subset of $\mathscr{X}^n$. Then, by the Schwarz inequality

$$
\begin{aligned}
&|M(\theta) - M(\theta')|^2 \\
&\quad = |E_\theta(T_\tau - \tfrac{1}{2}(M(\theta) + M(\theta'))) - E_{\theta'}(T_\tau - \tfrac{1}{2}(M(\theta) + M(\theta'))|^2 \\
&\quad = |\textstyle\sum_{n=1}^\infty \int_{\{\tau=n\}} (T_n - \tfrac{1}{2}(M(\theta) + M(\theta'))(f_n(x^n; \theta) - f_n(x^n; \theta'))\, d\nu^n|^2 \\
&\quad \leqq \textstyle\sum_1^\infty \int_{\{\tau=n\}} |T_n - \tfrac{1}{2}(M(\theta) + M(\theta'))|(f_n^{\frac{1}{2}}(x^n, \theta) + f_n^{\frac{1}{2}}(x^n; \theta'))^2\, d\nu^n \\
&\qquad \times \textstyle\sum_1^\infty \int_{\{\tau=n\}} |T_n - \tfrac{1}{2}(M(\theta) + M(\theta'))| \\
&\qquad \times (f_n^{\frac{1}{2}}(x^n; \theta) - f^{\frac{1}{2}}(x^n; \theta'))^2\, d\nu^n .
\end{aligned}
\tag{7}
$$

It is easy to see that

$$
\begin{aligned}
&\textstyle\int_{\{\tau=n\}} |T_n - \tfrac{1}{2}(M(\theta) + M(\theta'))|(f_n^{\frac{1}{2}}(x^n; \theta) + f^{\frac{1}{2}}(x^n; \theta'))^2\, d\nu^n \\
&\quad \leqq 2(\textstyle\int_{\{\tau=n\}} |T_n - \theta| f_n(x^n; \theta)\, d\nu^n + \int_{\{\tau=n\}} |T_n - \theta'| f_n(x^n; \theta)\, d\nu^n) \\
&\qquad + 2(|M(\theta) - \theta| \cdot P_\theta\{\tau = n\} + |M(\theta') - \theta'| \cdot P_{\theta'}\{\tau = n\}) \\
&\qquad + |M(\theta) - M(\theta')|(P_\theta\{\tau = n\} + P_{\theta'}\{\tau = n\})
\end{aligned}
$$

and so the first multiplier on the right side of (7) is less than

$$
4(S(\theta) + S(\theta')) + 2|M(\theta) - M(\theta')| . \tag{8}
$$

Upon setting $A_n = \{x^n : f_n(x^n; \theta) \geqq f_n(x^n; \theta')\}$, we observe that the second multiplier in (7) is less than

$$
\begin{aligned}
&\textstyle\sum_{n=1}^\infty (\int_{\{\tau=n\}A_n} |T_n - M(\theta)| f_n(x^n; \theta)\, d\nu^n \\
&\quad + \textstyle\int_{\{\tau=n\}\bar{A}_n} |T_n - M(\theta')| f_n(x^n; \theta')\, d\nu^n) \\
&\quad + |M(\theta) - M(\theta')| \textstyle\sum_1^n \int_{\{\tau=n\}} (f_n^{\frac{1}{2}}(x^n; \theta) - f^{\frac{1}{2}}(x^n; \theta'))^2\, d\nu^n \\
&\qquad \leqq 2(S(\theta) + S(\theta')) + 2|M(\theta) - M(\theta')| \\
&\qquad\quad \times \left[1 - E_\theta \left\{\textstyle\prod_{j=1}^\tau \left(\dfrac{f(X_j; \theta')}{f(X_j; \theta)}\right)^{\frac{1}{2}}\right\}\right].
\end{aligned}
\tag{9}
$$

To finish the proof we need the following:

LEMMA 1. *For any stopping rule* τ

$$
0 \leqq 1 - E_\theta \prod_1^\tau \left(\frac{f(X_j; \theta')}{f(X_j; \theta)}\right)^{\frac{1}{2}} \leqq 50 \cdot \rho(\theta; \theta') E_\theta \tau . \tag{10}
$$

If, in addition $\tau \equiv n$, *then*

$$
0 \leqq 1 - E_\theta \prod_1^n \left(\frac{f(X_j; \theta')}{f(X_j; \theta)}\right)^{\frac{1}{2}} \leqq 2n\rho(\theta; \theta') . \tag{11}
$$

We postpone the proof of the lemma until the next section. The inequality (10) of Lemma 1 together with (7)—(9) implies that if $50 \cdot \rho(\theta; \theta') \cdot \max\{E_\theta \tau, E_{\theta'} \tau\} \leqq \frac{1}{4}$ then

$$
\mu^2 \leqq (4\sigma + 2\mu)(2\sigma + \tfrac{1}{4}\mu) ,
$$

where $\mu = |M(\theta) - M(\theta')|^2$, $\sigma = S(\theta) + S(\theta')$, and hence that

$$\sigma \geqq \mu/16 . \tag{12}$$

Now if $|b(\theta)| + |b(\theta')| \leqq \frac{1}{2}|\theta - \theta'|$, then $|M(\theta) - M(\theta')| \geqq \frac{1}{2}|\theta - \theta'|$ and (12) implies (5). If $|b(\theta)| + |b(\theta')| \geqq \frac{1}{2}|\theta - \theta'|$, then

$$\sigma \geqq |b(\theta)| + |b(\theta')| \geqq \tfrac{1}{2}|\theta - \theta'|$$

and again (5) holds. The proof of (6) on the basis of (11) is the same.

3. **Proof of Lemma 1.** Consider first the simpler case $\tau \equiv n$. We have

$$1 - E_\theta \prod_1^n \left(\frac{f(X_j; \theta')}{f(X_j; \theta)}\right)^{\frac{1}{2}} = 1 - \left(\int_{\mathscr{X}} (f(x; \theta)f(x; \theta'))^{\frac{1}{2}}\, d\nu\right)^n$$
$$\leqq n\left(1 - \int_{\mathscr{X}} (f(x, \theta)f(x, \theta'))^{\frac{1}{2}}\, d\nu\right) = 2n\rho(\theta; \theta')$$

and (11) is proved.

To prove (10) we establish a few lemmas.

LEMMA 2. (Wald). *Let τ be a stopping time relative to a sequence of independent identically distributed random variables $\{\xi_j\}$ with $E\xi_j^2 < \infty$. Then*

$$E \sum_1^\tau \xi_j = E\xi_1 \cdot E\tau , \qquad \operatorname{Var}\left(\sum_1^\tau (\xi_j - E\xi_j)\right) = \operatorname{Var} \xi_1 \cdot E\tau . \tag{13}$$

For the proof see [1], page 350.

LEMMA 3. *Let $\eta \geqq 0$ and ξ be random variables. Then*

$$E\eta e^\xi \geqq E\eta \cdot \exp\left\{\frac{E\xi\eta}{E\eta}\right\} \geqq E\eta + E\xi\eta . \tag{14}$$

The first part of (14) is a consequence of Jensen's well-known inequality (see [3], page 159); the second part follows from the elementary inequality

$$e^y \geqq 1 + y , \qquad y \in R' .$$

Let $B = \{x : \frac{3}{2} \geqq (f(x; \theta')/f(x; \theta))^{\frac{1}{2}} \geqq \frac{2}{3}\}$. Denote by χ_j the indicator of the random event $X_j \in B$. Define random variables Z_j in the following way:

$$Z_j = \tfrac{1}{2} \ln \frac{f(X_j; \theta')}{f(X_j; \theta)} , \qquad X_j \in B$$
$$= 0 , \qquad X_j \bar{\in} B$$

LEMMA 4. *The following inequalities hold:*

$$P_\theta\{X_j \bar{\in} B\} = E_\theta(1 - \chi_j) \leqq 9\rho(\theta; \theta') , \tag{15}$$
$$P_{\theta'}\{X_j \bar{\in} B\} = E_{\theta'}(1 - \chi_j) \leqq 9\rho(\theta; \theta') .$$

PROOF. If $x \bar{\in} B$ then either $f^{\frac{1}{2}}(x; \theta') - f^{\frac{1}{2}}(x; \theta) > \frac{1}{2}f^{\frac{1}{2}}(x; \theta)$ or $f^{\frac{1}{2}}(x, \theta') - f^{\frac{1}{2}}(x; \theta) < -\frac{1}{3}f^{\frac{1}{2}}(x; \theta)$. In both cases

$$P_\theta\{X_j \bar{\in} B\} = \int_{\bar{B}} f(x; \theta)\, d\nu \leqq 9 \int_{\bar{B}} (f^{\frac{1}{2}}(x; \theta) - f^{\frac{1}{2}}(x; \theta'))^2\, d\nu \leqq 9\rho(\theta; \theta') .$$

The proof of the second part of (15) is the same.

LEMMA 5. *The following inequalities hold*:

$$|E_\theta Z_j| \leqq 23\rho(\theta; \theta'), \qquad E_\theta Z_j^2 \leqq 7\rho(\theta; \theta'), \qquad \mathrm{Var}_\theta Z_j \leqq 7\rho(\theta; \theta'). \tag{16}$$

PROOF. For $x \in B$

$$\frac{1}{2} \ln \frac{f(x; \theta')}{f(x; \theta)} = -(f^{\frac{1}{2}}(x; \theta) - f^{\frac{1}{2}}(x; \theta')) \cdot f^{-\frac{1}{2}}(x; \theta) + R,$$

where $|R| \leqq 3(f^{\frac{1}{2}}(x; \theta) - f^{\frac{1}{2}}(x; \theta'))^2 \cdot f^{-1}(x; \theta)$.

Using (15) we have

$$\begin{aligned} |E_\theta Z_j| &\leqq |\textstyle\int_B (f^{\frac{1}{2}}(x; \theta) - f^{\frac{1}{2}}(x; \theta'))f^{\frac{1}{2}}(x; \theta)\, d\nu| \\ &\quad + 3\rho(\theta; \theta') \leqq [1 - \textstyle\int_{\mathscr{X}} (f(x; \theta)f(x; \theta'))^{\frac{1}{2}}\, d\nu] \\ &\quad + 21\rho(\theta; \theta') = 23\rho(\theta; \theta'). \end{aligned}$$

Further, for $x \in B$ $|f^{\frac{1}{2}}(x; \theta) - f^{\frac{1}{2}}(x; \theta')| \leqq \frac{1}{2} f^{\frac{1}{2}}(x; \theta)$, so that

$$E_\theta Z_j^2 \leqq 2 \textstyle\int_B [(f^{\frac{1}{2}}(x; \theta) - f^{\frac{1}{2}}(x; \theta'))^2 + R^2 \cdot f(x; \theta)]\, d\nu \leqq 7\rho(\theta; \theta').$$

We are now ready to prove Lemma 1. We have

$$1 - E_\theta \prod_1^\tau \left(\frac{f(X_j; \theta')}{f(X_j; \theta)}\right)^{\frac{1}{2}} \leqq 1 - E_\theta (\textstyle\prod_1^\tau \chi_j \cdot \exp\{\sum_1^\tau Z_j\}). \tag{17}$$

By Lemmas 3—5

$$\begin{aligned} 1 &- E_\theta(\textstyle\prod_1^\tau \chi_j \cdot \exp\{\sum_1^\tau Z_j\}) \\ &\leqq 1 - E_\theta \textstyle\prod_1^\tau \chi_j - E_\theta[\prod_1^\tau \chi_j \sum_1^\tau Z_j] \\ &= E_\theta(1 - \textstyle\prod_1^\tau \chi_j) - E_\theta(\tau \prod_1^\tau \chi_j) \cdot E_\theta Z_1 \\ &\qquad + E_\theta(1 - \textstyle\prod_1^\tau \chi_j)(\sum_1^\tau Z_j - EZ_j) \\ &\leqq E_\theta \textstyle\sum_1^\tau (1 - \chi_j) + E_\theta \tau |E_\theta Z_1| + E_\theta^{\frac{1}{2}}(1 - \prod_1^\tau \chi_j)^2 \cdot \mathrm{Var}^{\frac{1}{2}} \sum_1^\tau Z_j \\ &= E_\theta \tau \cdot E_\theta(1 - \chi_j) + E_\theta \tau |E_\theta Z_1| + E_\theta^{\frac{1}{2}}(1 - \textstyle\prod_1^\tau \chi_j) \cdot E_\theta^{\frac{1}{2}} \tau \cdot \mathrm{Var}^{\frac{1}{2}} Z_1 \\ &\leqq 41\rho(\theta; \theta') E_\theta \tau. \end{aligned} \tag{18}$$

Thus Theorem 3 is proved.

4. Remarks. 1. It is easy to see that it will be sufficient to suppose that Θ is a subset of a normed space B. Theorems 1—3 are valid in this case if $|\theta|$ is the norm in B.

2. The requirement of absolute continuity of all measures $\mathscr{P}_\theta$ relative to some common measure ν is unnecessary. It is sufficient to take

$$f(x; \theta) = \frac{d\mathscr{P}_\theta}{d(\mathscr{P}_\theta + \mathscr{P}_{\theta'})}, \qquad f(x; \theta') = \frac{d\mathscr{P}_{\theta'}}{d(\mathscr{P}_\theta + \mathscr{P}_{\theta'})}$$

when points θ, θ' are being considered.

Acknowledgment. The authors thank the referee for suggestions for shortening a proof, and also for comments concerning English usage.

REFERENCES

[1] Doob, J. L. (1953). *Stochastic Processes*. Wiley, New York.

[2] Ibragimov, I. A. and Has'minskii, R. Z. (1973). On the moments of generalized Bayesian and maximum likelihood estimates. (in Russian). *Teor. Verojatnost. i Primen.* **18** 535–546.

[3] Loève, M. (1963). *Probability Theory*. Van Nostrand, Princeton.

Department of Mathematics
Leningrad University
10 Linia, 33
Leningrad, U.S.S.R.