ON OPTIMAL SEQUENTIAL ESTIMATION OF A SHIFT PARAMETER IN THE CLASS OF INVARIANT PROCEDURES

MATHEMATICS

1970

SovietRxiv

Abstract

Full Text

UDC 519.281

MATHEMATICS

A. I. Shalyt

ON OPTIMAL SEQUENTIAL ESTIMATION OF A SHIFT PARAMETER IN THE CLASS OF INVARIANT PROCEDURES

(Presented by Academician Yu. V. Linnik on 23 I 1970)

The possibilities afforded by sequential procedures in estimation theory consist, first, in enlarging the stock of parametric functions admitting unbiased estimates (see, for example, (^4, ^1)), and, second, for those functions for which unbiased estimates exist from a sample of fixed size, in some cases better sequential estimates can be indicated. The present note is devoted precisely to the latter aspect of the theory of sequential estimation.

In Sec. 1 a definition is given of optimality of a sequential estimate in a certain class; in essence it is analogous to Wald' s definition of optimality of a sequential test for distinguishing two hypotheses.

Optimal sequential estimates within the class of invariant procedures for the scheme with a shift parameter are indicated in Sec. 2. Of interest, in our opinion, is Theorem 4, which establishes a certain integral relation between the risk and the mean-sample-size function for optimal estimates.

1. Let a family of probability distributions $F_\theta(x)$, $\theta \in \Theta$, be given on (R^1, B). A sequential estimate is a decision function which is completely determined by a stopping rule (st. r.) and a terminal decision.

Denote by (R^∞, B^∞) the measurable space of all numerical sequences $\omega = \{x_1, x_2, ...\}$ and consider an increasing sequence of σ-algebras $F_1 \subset F_2 \subset \cdots \subset F_\infty \subset B^\infty$, where $F_n \subset B^n$. A stopping rule, consistent with the sequence $F = \{F_n\}$ (we denote all such rules by $\mathfrak{M}_F$), is a random variable (r.v.) $\tau = \tau(\omega)$ with values in the set of positive integers such that the event $\{\tau = n\} \in F_n$.

By a terminal decision d we shall mean a countable collection of functions

$$d = \{\bar\theta_1(x_1),\ \bar\theta_2(x_1, x_2),\ \dots,\ \bar\theta_n(x_1, \dots, x_n),\ \dots\};$$

its meaning is that, having stopped at the moment $\tau = n$, we use $\bar\theta_n$ as an estimate of the parameter θ.

Definition 1. We shall call the pair $[\tau, d]$ a **sequential estimate**.

The risk of the sequential estimate $[\tau, d]$, corresponding to the loss function $r(\bar{\theta}, \theta)$, is defined as follows:

$$R_\theta[\tau, d] = E_\theta r(\bar{\theta}_{\tau(\omega)}(\omega), \theta).$$

Definition 2. A sequential estimate $[\tau_0, d_0]$ is called **optimal in the class** $(\mathfrak{M}_F, D)$, where D is a certain set of terminal decisions, if: 1) $\tau_0 \in \mathfrak{M}_F$, $d_0 \in D$; 2) from the conditions $\tau \in \mathfrak{M}_F$ and $E_\theta \tau \leq E_\theta \tau_0$ for all $\theta \in \Theta$ it follows that, for every $d \in D$, one has

$$R_\theta[\tau_0, d_0] \leq R_\theta[\tau, d]$$

also for all $\theta \in \Theta$.

2. In the scheme of direct measurements the observations x_i have the form $x_i = \theta + \varepsilon_i$, where θ is the (unknown) value of the measured quantity, and ε_i are the observation errors, which are usually assumed to be independent identically distributed r.v.' s. Let $P(\varepsilon_i < x) = F(x)$; then $P(x_i < x) = P(\theta +$

$+\varepsilon_i < x) = F(x - \theta)$, and therefore the parameter θ is called the location parameter. We shall indicate an optimal sequential estimate of $\theta \in R^1$ within a certain class $(\mathfrak{M}_F, D)$ and $r(\tilde{\theta}, \theta) = (\tilde{\theta} - \theta)^2$.

In sequential estimation of the location parameter it is natural to use invariant stopping rules, i.e., to put $F_1 = \phi$ and, for $n \geq 2$, $F_n = \sigma(x_2 - x_1, \dots, x_n - x_1)$, where $\sigma(\xi)$ denotes the σ-algebra generated by the vector ξ. Note that when invariant stopping rules $\tau \in \mathfrak{M}_F$ are used, we never stop at the first step (with probability $\tau > 1$), and therefore the countable set of functions determining the terminal decision may be specified without the first component $\tilde{\theta}_1$. As D we take the set of those d which are formed by functions $\tilde{\theta}_n$ satisfying, for any $c \in R^1$, the condition

$$\tilde{\theta}_n(x_1 + c, \dots, x_n + c) = c + \tilde{\theta}_n(x_1, \dots, x_n).$$

With this choice of the class $(\mathfrak{M}_F, D)$, every sequential estimate $[\tau, d]$ belonging to this class differs from an unbiased estimate of the parameter θ only by an additive constant.

Put

$$t_n = t_n(x_1, \dots, x_n) = \frac{x_1 + \dots + x_n}{n} - E_0 \left(\frac{x_1 + \dots + x_n}{n} \middle| F_n \right),$$

$n \geq 2$, and introduce the terminal decision $t = \{t_2, t_3, \dots, t_n, \dots\}$.

Theorem 1. Suppose that on (R^1, B) a family of distributions $F(x - \theta)$, $\theta \in R^1$, is given; then every sequential estimate $[\tau_\varepsilon, t]$, where τ_ε satisfies the condition

$$E_0(t^2_{\tau_\varepsilon} + \varepsilon\tau_\varepsilon) = \inf_{\tau\in\mathfrak{M}_F} E_0(t^2_\tau + \varepsilon\tau), \qquad \varepsilon > 0, \tag{1}$$

is optimal in the class $(\mathfrak{M}_F, D)$.

It follows from the results of ([2]) that such a stopping rule τ_ε exists and $P_\theta(\tau_\varepsilon < \infty) = 1$. As a consequence of this general theorem we obtain two others, the formulations of which are more convenient.

Theorem 2. Suppose that the distribution $F(x)$ is such that, for every $n \geq 2$,

$$E_0(t^2_n \mid F_n) = \text{const} = D_n.$$

If $\tau_0 \in \mathfrak{M}_F$ and $P_0(\tau_0 = N) = 1 - \alpha$, $P_0(\tau_0 = N + 1) = \alpha$, where N is an integer and $0 \leq \alpha < 1$, then the sequential estimate $[\tau_0, t]$ is optimal in the class $(\mathfrak{M}_F, D)$; moreover $E_\theta\tau_0 = N + \alpha$ and the variance of the estimate

$$D_\theta t_{\tau_0} = (1-\alpha)D_N + \alpha D_{N+1}.$$

Examples: normal distribution

$$N(0,\sigma); \qquad t_n = (x_1 + \dots + x_n)/n = \bar{x}; \qquad E_0(\bar{x}^2 \mid F_n) = \sigma^2/n;$$

exponential distribution with density $p(x) = e^x$ for $x \geq 0$, $p(x) = 0$ for $x < 0$;

$$t_n = \min_{1\leq i\leq n} x_i - 1/n; \qquad E_0(t^2_n \mid F_n) = 1/n^2.$$

Theorem 3. Suppose $d_n = d_n(x_1, \dots, x_n) = E_0(t^2_n \mid F_n)$, and the distribution function $F(x)$ satisfies the following condition: for every $n \geq 2$ and $\varepsilon > 0$, from the inequality

$$E_0(d_{n+2} \mid F_{n+1}) - d_{n+1} < -\varepsilon$$

it follows that

$$E_0(d_{n+1} \mid F_n) - d_n < -\varepsilon.$$

Then the sequential estimate $[\tau_\varepsilon, t]$, where

$$\tau_\varepsilon = \{k : E_0(d_{k+1} \mid F_k) - d_k \geq -\varepsilon\},$$

is optimal in the class $(\mathfrak{M}_F, D)$.

Example: the uniform distribution with density $p(x) = 1$ for $|x| \leq 1/2$, $p(x) = 0$ for $x > 1/2$.

Denote $x_{(n)} = \min_{1\leq i\leq n} x_i$ and $x^{(n)} = \max_{1\leq i\leq n} x_i$; then

$$t_n = (x_{(n)} + x^{(n)})/2;$$

$$d_n = \frac{1}{12}(1+x_{(n)}-x^{(n)})^2; \qquad E_0(d_{n+1} \mid F_n) = \frac{1}{24}(1+x_{(n)}-x^{(n)})^2(1+x^{(n)}-x_{(n)})$$

and

$$\tau_\varepsilon = \min\{n : x^{(n)} - x_{(n)} \geq 1 - \sqrt[3]{24\varepsilon}\}.$$

It was already indicated in ([3]) that this sequential estimate is better than the estimate constructed from a sample of fixed size; however, nothing was known about its optimal character.

The theorem given below makes it possible to establish a lower bound for the variance of an arbitrary sequential estimate $[\tau, d]$ in the class $(\mathfrak{M}_F, D)$.

Theorem 4. Suppose that for every $\varepsilon > 0$ there exists a unique stopping rule τ_ε which realizes the infimum in (1). Then the mean number of observations $q(\varepsilon) = E_\theta\tau_\varepsilon$ of the optimal sequential estimate $[\tau_\varepsilon, t]$ and

the magnitude of its variance $D(\varepsilon) = E_\theta(t_{\tau_\varepsilon} - \theta)^2$ are related by

$$D(\varepsilon) = -\int_0^\varepsilon x\, dq(x) = \int_{E_\theta\tau_\varepsilon}^\infty q^{-1}(y)\, dy,$$

where the first integral is understood as a Stieltjes integral, and the second as a Riemann integral.

Corollary. For any sequential estimate $[\tau, \tilde{d}]$, where $\tau \in \mathfrak{M}_F$ and $\tilde{d} = \{\tilde{\theta}_2, \tilde{\theta}_3, \ldots, \tilde{\theta}_n, \ldots\} \in D$, the inequality

$$E_\theta(\tilde{\theta}_\tau - \theta)^2 \geqslant \int_{E_\theta\tau}^\infty q^{-1}(y)\, dy. \tag{2}$$

holds.

Example: the uniform distribution, $q(\varepsilon) = 1/\sqrt[3]{3\varepsilon}$, and from inequality (2) it follows that for every sequential estimate $[\tau, \tilde{d}]$

$$E_\theta(\bar{\theta}_\tau - \theta)^2 \geqslant \frac{1}{6}(E_\theta \tau)^2.$$

The author expresses gratitude to A. M. Kagan for his help in the course of the work.

Leningrad State University
named after A. A. Zhdanov

Received
5 I 1970

REFERENCES

[1] R. A. Zaidman, Yu. V. Linnik, I. V. Romanovskii, DAN, **185**, No. 6 (1969).
[2] I. S. Chzhou, G. Robbins, Collection of Translations: Mathematics, **9**, 3 (1965).
[3] A. I. Shalit, DAN, **189**, No. 1 (1969).
[4] M. H. De Groot, Ann. Math. Statist., **30**, 80 (1959).

Note: Figure translations are in progress. See original paper for figures.