SEQUENTIAL METHODS IN STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 16
PWN-POLISH SCIENTIFIC PUBLISHERS
WARSAW 1985

ABOUT A FIXED PRECISION ESTIMATION OF THE PARAMETER OF UNIFORM DENSITY

ANDRZEJ SIEROCIŃSKI

Institute of Mathematics, Technical University of Warsaw, Warsaw, Poland

Introduction

Let the function

$$f_\theta(x) = \begin{cases} \theta^{-1} & \text{for} \quad x = [0, \theta], \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $\theta > 0$, be a uniform density. The problem is to determine n, the size of a random sample from $f_\theta(x)$, such that an estimator $\hat{\theta}_n$ based on this sample satisfies the condition

$$P_\theta\{|\theta - \hat{\theta}_n| < d\} \geqslant 1-\gamma, \tag{2}$$

where $d > 0$ and $\gamma \in (0, 1)$ are specified in advance. In other words, we shall determine a confidence interval for the parameter θ with a fixed width $2d$, uniformly with respect to θ. The problem of a fixed precision estimation of θ seems to be very simple, but this is not quite true.

In most problems of estimation, estimators based on samples of fixed sizes have precisions which depend on unknown parameters and estimators with prescribed precision are not available without resort to sequential sampling in two or more stages, as in Stein's procedure for the estimation of the mean of a normal distribution with unknown variance. Since the parameter θ in the uniform distribution $U(0, \theta)$ is the scale parameter (i.e., $f_\theta(x) = \frac{1}{\theta} f\left(\frac{x}{\theta}\right)$, where f is a density function), it follows from the Blum and Rosenblatt lemma ([3]) (see also [7]) that it is impossible for a fixed, nonrandom sample size to construct a fixed precision confidence interval for θ. Two-sample procedures solving this problem are given in [2] and [4]. Another fixed precision sequential estimation procedure is presented in [5].

In this paper we shall consider some class of stopping rules for a sequence of independent random variables, distributed identically, uniformly on $(0, \theta)$ (i.i.d. $U(0, \theta)$ r.v's) and we shall show that in this class of stopping rules there exists no optimal stopping rule (optimal in the sense of minimal expectation, uniformly with respect to θ) which satisfies condition (2).

Result

Consider the statistical space $(R_+^1, B_+^1, \{P_\theta, \theta > 0\})$, where R_+^1 is a positive half-line, B_+^1 is the family of Borel subsets of R_+^1 and P_θ is a uniform distribution with probability density function $f_\theta(x)$ defined in (1). The problem consists in fixed-precision estimating θ. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. $U(0, \theta)$ r.v's and let $d > 0$ and $\gamma \in (0, 1)$ be given. There are many functions of the random sample that could be used as the estimators of θ, but we shall use the largest observation in the random sample of size n, because the largest order statistic is the minimal sufficient statistic for θ. The unbiased estimator of θ is equal to $\frac{n+1}{n}\max(X_1, \ldots, X_n)$, but neither $\frac{n+1}{n}\max(X_1, \ldots, X_n)$ nor $\max(X_1, \ldots, X_n)$ is uniformly better for all values of θ (in the sense of minimizing the probability $P_\theta\{|\theta - \theta_n| > d\}$). Since we always have $\max(X_1, \ldots, X_n) \leqslant \theta$, we usually use

$$\hat{\theta}_n = \max(X_1, \ldots, X_n). \tag{3}$$

The class C of stopping rules under consideration is described as follows:

DEFINITION 1. A stopping rule σ *belongs to* C if and only if there exists a nondecreasing sequence of positive reals $(A_n, n \geqslant 1)$ tending to infinity and such that

$$\sigma = \inf\{n \geqslant 1: \hat{\theta}_n \leqslant A_n\}, \qquad P_\theta\text{-a.s.} \tag{4}$$

for all $\theta > 0$.

Let us notice that the assumption $A_n \to \infty$ is equivalent to the assumption that

$$P_\theta\{\sigma < \infty\} = 1 \tag{5}$$

for each $\theta > 0$ and, moreover, that the monotonicity assumption is not important. Indeed, if a stopping time of form (4) is given, where a sequence $(A_n, n \geqslant 1)$ is not nondecreasing, we may replace that sequence by the sequence $B_n = \max(A_1, \ldots, A_n)$, which is nondecreasing, and determine the same stopping rule. In view of the equivalence between a stopping rule $\sigma \in C$ and a sequence $(A_n, n \geqslant 1)$ in Definition 1 they are identified $(\sigma \equiv (A_n, n \geqslant 1))$.

Our interest is in finding a random sample size required for a fixed precision estimation of the parameter θ; for this reason we define a new class D of stopping rules.

DEFINITION 2. A stopping rule σ belongs to D if and only if $\sigma \in C$ and

$$P_\theta\{\theta-\hat{\theta}_\sigma > d\} \leqslant \gamma \tag{6}$$

for each $\theta > 0$.

LEMMA 1. *The class D of stopping rules is nonempty.*

Proof. It is easy to see that the stopping rule $\sigma_0 \equiv (A_n^0,\, n \geqslant 1)$, where

$$A_n^0 = \frac{d\sqrt[n]{A}}{\sqrt[n]{n^2}-\sqrt[n]{A}}, \qquad n \geqslant 1 \text{ and } A = \frac{6\gamma}{\pi^2}, \tag{7}$$

belongs to C. Moreover, from (4) we have $\hat{\theta}_{\sigma_0} \leqslant A_{\sigma_0}^0$; then

$$\begin{aligned} P_\theta\{\theta-\hat{\theta}_{\sigma_0} > d\} &= \sum_{n=1}^{\infty} P_\theta\{\theta-\hat{\theta}_n > d,\, \sigma_0 = n\} \\ &\leqslant \sum_{n=1}^{\infty} P_\theta\left\{\theta-\hat{\theta}_n > d\frac{\hat{\theta}_n}{A_n^0},\, \sigma_0 = n\right\} \leqslant \sum_{n=1}^{\infty} P_\theta\left\{\theta-\hat{\theta}_n > d\frac{\hat{\theta}_n}{A_n^0}\right\} \\ &= \sum_{n=1}^{\infty} P_\theta\left\{\hat{\theta}_n < \frac{A_n^0}{A_n^0+d}\theta\right\} = \sum_{n=1}^{\infty} [A_n^0/(A_n^0+d)]^n = \sum_{n=1}^{\infty}\left(\frac{A}{n^2}\right) = \gamma \end{aligned}$$

for all $\theta > 0$, and the lemma follows. ∎

LEMMA 2. *For all natural numbers n ($n \in N$)*

$$A_n \leqslant \frac{d\sqrt[n]{\gamma}}{1-\sqrt[n]{\gamma}}, \tag{8}$$

provided $\sigma \equiv (A_n,\, n \geqslant 1) \in D$.

Proof. Let $m \in N$ be fixed. The condition $\sigma \in D$ yields (6) for all $\theta > 0$. If we put $\theta = A_m + d$, then

$$P_\theta\{\theta-\hat{\theta}_n > d,\, \sigma = n\} = \begin{cases} P_\theta\{\sigma = n\} & \text{for } n \leqslant m, \\ 0 & \text{for } n > m, \end{cases}$$

and so

$$\begin{aligned} \gamma \geqslant P_\theta\{\theta-\hat{\theta}_\sigma > d\} &= \sum_{n=1}^{\infty} P_\theta\{\theta-\hat{\theta}_n > d,\, \sigma = n\} = \sum_{n=1}^{\infty} P_\theta\{\sigma = n\} \\ &= P_\theta\{\sigma \leqslant m\} \geqslant P_\theta\{\hat{\theta}_m \leqslant A_m\} = \left(\frac{A_m}{A_m+d}\right)^m. \end{aligned}$$ ∎

It is easy to see that the stopping rule

$$t \equiv \left(\frac{d\sqrt[n]{\gamma}}{1-\sqrt[n]{\gamma}},\ n \geqslant 1\right) \qquad \text{belongs to } C, \tag{9}$$

but it will be shown below that $t \notin D$ (Theorem 1).

LEMMA 3. *Let the stopping rules* $\sigma \equiv (A_n, n \geqslant 1)$ *and* $\sigma' \equiv (A'_n, n \geqslant 1)$ *from the class* C *be given. If* $A'_n \leqslant A_n$ *for all* $n \in N$, *then*

(i) $P_\theta\{\sigma \leqslant \sigma'\} = 1$;

(ii) $E_\theta \sigma \leqslant E_\theta \sigma'$;

(iii) $P_\theta\{\theta - \hat{\theta}_{\sigma'} > d\} \leqslant P_\theta\{\theta - \hat{\theta}_\sigma > d\}$

for all $\theta > 0$.

Proof. Let us notice that for all $n \in N$ the following inclusions are satisfied:

$$\{\sigma' = n\} \subset \{\hat{\theta}_n \leqslant A'_n\} \subset \{\hat{\theta}_n \leqslant A_n\} \subset \{\sigma \leqslant n\}$$

and so

$$P_\theta\{\sigma \leqslant \sigma'\} = P_\theta\Big(\bigcup_{n=1}^{\infty} \{\sigma \leqslant n, \sigma' = n\}\Big) = P_\theta\Big(\bigcup_{n=1}^{\infty} \{\sigma' = n\}\Big) = 1$$

for all $\theta > 0$. (ii) and (iii) follow immediately from (i). ∎

LEMMA 4. *If* $\tau' \equiv (B_n, n \geqslant 1) \in C$ *and for some* $k \in N$ *and all* $\theta \leqslant B_k + d$ *we have* $P_\theta\{\theta - \hat{\theta}_{\tau'} > d\} \leqslant \gamma$, *then there exists a stopping rule* $\tau'' \equiv (A_n, n \geqslant 1)$ *from the class* D *such that*

$$A_i = B_i \qquad \textit{for all} \quad i \leqslant k. \tag{10}$$

Proof. For $i \leqslant k$ we put $A_i = B_i$. Let us notice that, for all $\theta \leqslant B_k + d$, $P_\theta\{\theta - \hat{\theta}_{\tau'} > d, \tau' > k\} = 0$, and so $P_\theta\{\theta - \hat{\theta}_{\tau'} > d\}$ does not depend on $(B_j, j \geqslant k+1)$.

Consider the case $\theta > B_k + d$ and a stopping rule s_r defined by a sequence $(A_1, \ldots, A_{k+r}, B_{k+r+1}, \ldots)$ where

$$A_k = A_{k+1} = \ldots = A_{k+r}, \qquad r \in N. \tag{11}$$

Simple calculations give us

$$P_\theta\{\theta - \hat{\theta}_{s_r} > d, s_r \leqslant k\} = \sum_{j=1}^{k} P_\theta\{\theta - \hat{\theta}_j > d, s_r = j\}$$

$$= \sum_{j=1}^{k} P_\theta\{\hat{\theta}_j < \theta - d, \hat{\theta}_1 > A_1, \ldots, \hat{\theta}_{j-1} > A_{j-1}, \hat{\theta}_j \leqslant A_j\}$$

$$= \sum_{j=1}^{k} \theta^{-j} \int_{\{(x_1,\ldots,x_j) \in C_j(A_1,\ldots,A_j)\}} \!\!\cdots\!\! \int dx_1 \cdot \ldots \cdot dx_j = \sum_{j=1}^{k} \theta^{-j} |C_j| \downarrow 0$$

as $\theta \to \infty$, where $C_j(A_1, \ldots, A_j)$ denotes a subset of $\bigtimes_{i=1}^{j} [0, A_j]$ and $|C_j|$ its Lebesgue's measure.

It is obvious that for $\theta = A_k + d$ we have

$$P_\theta\{\theta-\hat{\theta}_{s_r} > d\} = P_{A_k+d}\{A_k+d-\hat{\theta}_{s_r} > d,\, s_r \leqslant k\} = \sum_{j=1}^{k} \frac{|C_j|}{(A_k+d)^j} \leqslant \gamma;$$

then for all $\theta > A_k + d$

$$\sum_{j=1}^{k} \frac{|C_j|}{\theta^j} < \sum_{j=1}^{k} \frac{|C_j|}{(A_k+d)^j} \cdot \frac{(A_k+d)}{\theta} \leqslant \frac{\gamma(A_k+d)}{\theta}.$$

By analogy we obtain

$$\begin{aligned} &P_\theta\{\theta-\hat{\theta}_{s_r} > d,\, s_r > k\} \\ &= \sum_{j=k+r+1}^{\infty} P_\theta\{\hat{\theta}_j < \theta-d,\, \hat{\theta}_1 > A_1, \ldots, \hat{\theta}_k > A_k,\, \hat{\theta}_{k+r+1} > B_{k+r+1}, \ldots \\ &\qquad\qquad \ldots, \hat{\theta}_{j-1} > B_{j-1},\, \hat{\theta}_j \leqslant B_j\} \\ &\leqslant \sum_{j=k+r+1}^{\infty} P_\theta\{\hat{\theta}_j < \theta-d, \text{ there exists } i \leqslant k, \text{ such that } X_i > A_k\} \\ &\leqslant \sum_{j=k+r+1}^{\infty} \left(\frac{\theta-d}{\theta}\right)^{j-1} \binom{k}{1} \frac{\theta-d-A_k}{\theta} = k\frac{\theta-d-A_k}{d}\left(\frac{\theta-d}{\theta}\right)^{k+r} \end{aligned}$$

This implies of course that

$$P_\theta\{\theta-\hat{\theta}_{s_r} > d\} \leqslant \frac{\gamma(A_k+d)}{\theta} + k\frac{\theta-d-A_k}{d}\left(\frac{\theta-d}{\theta}\right)^{k+r}$$

for all $\theta > A_k + d$. Consider the function

$$h(\theta) = \frac{\gamma(A_k+d)}{\theta} + k\frac{\theta-d-A_k}{d}\left(\frac{\theta-d}{\theta}\right)^{k+r}$$

It is obvious that there exist $r \in N$ and real $\delta > 0$ such that $h(\theta) < \gamma$ for all $\theta \in (A_k+d,\, A_k+d+\delta]$.

Now we put

$$A_{k+r+1} = A_k + \delta = A_{k+r} + \delta \tag{12}$$

and $\gamma' = h(A_{k+r+1}+d) < \gamma$. From the proof of Lemma 1 it follows that the sequence $(A_n^0,\, n \geqslant 1)$ defined by (7) satisfies the condition

$$\sum_{i=1}^{\infty} \left(\frac{A_i^0}{A_i^0+d}\right)^i = \gamma.$$

Finally, we define the sequence (A_i) for $i \geqslant k+r+2$. Namely, we put

$$A_{k+r+2} = A_{k+r+3} = \ldots = A_{m-1} = A_{k+r+1} \tag{13}$$

and

$$A_j = A_j^0 \qquad \text{for all} \quad j \geqslant m, \tag{14}$$

where

$$m = \inf\left\{n > k+r+1 \colon \sum_{i=n}^{\infty} \left(\frac{A_i^0}{A_i^0+d}\right)^i < \gamma-\gamma',\ A_n^0 \geqslant A_{k+r+1}\right\}.$$

We shall show that the stopping rule $\tau'' = (A_n, n \geqslant 1)$ given by (10)–(14) satisfies (6), so that $\tau'' \in D$. Indeed, for $\theta \leqslant A_{k+r+1}+d$ this follows from the fact that $P_\theta\{\theta-\hat{\theta}_{\tau''} > d\}$ does not depend on $(A_j, j > k+r+1)$. Moreover, in the case $\theta > A_{k+r+1}+d$ we have

$$\begin{aligned} &P_\theta\{\theta-\hat{\theta}_{\tau''} > d\} \\ &\quad = P_\theta\{\theta-\hat{\theta}_{\tau''} > d,\ \tau'' \leqslant k+r+1\} + P_\theta\{\theta-\hat{\theta}_{\tau''} > d,\ \tau'' > k+r+1\} \\ &\quad \leqslant \gamma' + P_\theta\left\{\theta-\hat{\theta}_{\tau''} > d\frac{\hat{\theta}_{\tau''}}{A_{\tau''}},\ \tau'' > k+r+1\right\} \\ &\quad = \gamma' + \sum_{j=m}^{\infty} P_\theta\left\{\hat{\theta}_j \leqslant \frac{A_j}{A_j+d}\cdot\theta\right\} = \sum_{j=m}^{\infty}\left(\frac{A_j}{A_j+d}\right)^j + \gamma' \leqslant \gamma. \end{aligned}$$ ■

THEOREM 1. *There exists no stopping rule $t^* \in D$ satisfying*

$$E_\theta t^* \leqslant E_\theta \sigma \tag{15}$$

for all $\sigma \in D$ and all $\theta > 0$.

Proof. Assume that there exists a stopping rule $t^* \equiv (A_i^*, i \geqslant 1) \in D$ such that (15) holds for every $\sigma \in D$ and all $\theta > 0$. From Lemmas 2 and 4 it follows that $A_1^* = d\gamma/(1-\gamma)$. Namely, if $\sigma \equiv (A_i, i \geqslant 1)$ and $A_1 < A_1^*$, then $E_\theta t^* = 1 < E_\theta \sigma$ for all $\theta \in (A_1, A_1^*)$.

Now, assume that $A_2^* > A_1^*$ and consider $\theta \leqslant A_2^*+d$. Then

$$P_\theta\{\theta-\hat{\theta}_{t^*} > d\} = \begin{cases} \dfrac{\theta-d}{\theta} & \text{for} \quad \theta \leqslant A_1^*+d, \\[2ex] \dfrac{A_1^*}{\theta}+\dfrac{(\theta-d)(\theta-d-A_1^*)}{\theta^2} & \text{for} \quad \theta \in (A_1^*+d,\ A_2^*+d]. \end{cases}$$

Since $\psi[(A_1^*+d)^+] = \gamma$ and $\psi'[(A_1^*+d)^+] > 0$, where

$$\psi(\theta) = \frac{A_1^*}{\theta}+\frac{(\theta-d)(\theta-d-A_1^*)}{\theta^2} \qquad \text{for} \quad \theta \in (A_1^*+d,\ A_2^*+d],$$

it follows that $A_2^* = A_1^*$.

By similar arguments to that used above we obtain

$$A_3^* = \frac{d\sqrt{\gamma}}{1-\sqrt{\gamma}}.$$

Let us consider a stopping rule $t_1 \equiv (A'_i,\ i \geqslant 1) \in D$ such that $A'_1 = 0$ and $A'_2 = A^*_3$. The existence of t_1 follows from Lemma 4. If we put $\theta = A^*_3$ then simple calculations give us

$$\begin{aligned} E_\theta t^* &= 1 \cdot P_\theta\{\hat{\theta}_1 \leqslant A^*_1\} + 3 \cdot P_\theta\{\hat{\theta}_1 > A^*_1,\ \hat{\theta}_3 \leqslant A^*_3\} \\ &= \frac{d \cdot \gamma}{1-\gamma}\,\frac{1-\sqrt{\gamma}}{d \cdot \sqrt{\gamma}} + 3\left(1 - \frac{\sqrt{\gamma}}{1+\sqrt{\gamma}}\right) = \frac{3+\sqrt{\gamma}}{1+\sqrt{\gamma}} \\ &> 2 = 2 \cdot P_\theta\{t_1 = 2\} = E_\theta t_1 . \end{aligned}$$

The last inequality contradicts assumption (15) and the theorem follows. ∎

Consider again inequality (8) in Lemma 2. It is impossible to improve this inequality. Indeed, if a stopping rule $s_k \equiv (A^k_n,\ n \geqslant 1)$, where $A^k_1 = \ldots \ldots = A^k_{k-1} = 0$ and $A^k_n = d\sqrt[n]{\gamma}/(1-\sqrt[n]{\gamma})$ for $n \geqslant k$, satisfies the assumption of Lemma 4, then there exists a stopping rule $w_k \equiv (B^k_n,\ n \geqslant 1) \in D$ such that $B^k_k = A^k_k$. We can say that the stopping rule t defined by (9) is an envelope of class D (see also Lemma 3). Now consider a stopping rule

$$\tau \equiv \left(-\frac{nd}{\ln \gamma},\ n \geqslant 1\right) \in C. \tag{16}$$

The stopping rule τ and the sequence $(\hat{\theta}_n,\ n \geqslant 1)$ satisfy the assumptions of the Anscombe theorem (see [1]). From this theorem it follows that

$$\lim_{d \to 0} P_\theta\{\theta - \hat{\theta}_\tau > d\} = \gamma$$

for all $\theta > 0$. Since the inequality

$$-\frac{nd}{\ln \gamma} > \frac{d\sqrt[n]{\gamma}}{1-\sqrt[n]{\gamma}}$$

holds for all $n \in N$ and $\gamma \in (0, 1)$, from Lemma 3 if follows that the stopping rule t is asymptotically consistent, i.e.,

$$\lim_{d \to 0} P_\theta\{\theta - \hat{\theta}_t > d\} \leqslant \gamma$$

for all $\theta > 0$, but $t \notin D$ (Theorem 1).

At the end of our considerations notice that stopping rules from class C have the following interesting property:

Assume that $\theta \leqslant \theta_0$; *then a stopping rule* $\sigma \equiv (A_n,\ n \geqslant 1) \in C$ *is bounded from above by*

$$S = \inf\{n \geqslant 1:\ A_n \geqslant \theta_0\}.$$

References

[1] F. J. Anscombe, *Large sample theory of sequential estimation*, Proc. Cambridge Philos. Soc. **48** (1952), 600–607.

[2] A. Birnbaum and W. C. Healy, *Estimates with prescribed variance based on two-stage sampling*, Ann. Math. Statist. **31** (1960), 662–676.

[3] J. R. Blum and J. Rosenblatt, *Fixed precision estimation in the class of IFR distributions*, Ann. Inst. Statist. Math. **21** (1969), 211–213.

[4] F. A. Graybill and T. L. Connell, *Sample size required to estimate the parameter in the uniform density within d unit of the true value*, J. Amer. Statist. Assoc. **59** (1964), 550–556.

[5] A. Sierociński, *On fixed precision estimation of the maximal value of a random variable*, Demonstratio Math. **13** (1980), 995–1003.

[6] —, *Estymacja sekwencyjna parametru θ w rozkladach o nośniku w przedziale* $[0, \theta]$, Doctoral Dissertation, Technical University of Warsaw, Warsaw 1981, in Polish.

[7] S. Zacks, *The Theory of Statistical Inference*, Wiley, New York 1971.

Presented to the semester
Sequential Methods in Statistics
September 7–December 11, 1981