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

FIXED PRECISION ESTIMATION OF A NORMAL MEAN

BENIAMIN GOŁDYS

Institute of Mathematics, Polish Academy of Sciences,
Warsaw, Poland

Introduction

The problem of fixed precision estimation of a normal mean has been well known since 1945, when Stein published his pioneering paper ([7]). From that time this topic was carefully investigated by many authors. In their papers it was usually required that the observed random variables be independent and identically distributed. The independence assumption turned out to be necessary in the sense shown in [2].

In this paper we make an attempt to relax these two assumptions. To attain this goal we shall make use of Zieliński's approach ([11]), which allows us to obtain the fixed precision estimation procedures for sequences of dependent random variables and thereby to avoid the limitations of the classical approach (presented by Blum and Rosenblatt in [2]).

Now we shall present the problem more formally.

Formalization

Let $(X_n)_{n=1}^{\infty}$ denote a sequence of normally distributed random variables. The distribution of the stochastic process $(X_n)_{n=1}^{\infty}$ will be denoted by P_θ with θ of the form $\theta = (\mu, \theta')$ where μ is assumed to be any real number and is called the *mean value* or *nearly mean value* (we shall explain this term later on) and θ' is a certain nuisance parameter belonging to the known parameter set U, so that we can write $\theta \in R \times U$. The parameter θ' can be either a real number or a finite- or infinite-dimensional vector.

Let us suppose that we can observe k ($k \geqslant 1$) independent copies of the process $(X_n)_{n=1}^{\infty}$, namely $(X_n^{(1)}), (X_n^{(2)}), \ldots, (X_n^{(k)})$. Usually in the theory of statistical inference k is taken equal to one. The σ-fields of events which are observable up to time n are denoted by

$$\mathscr{F}_n^{(k)} = \sigma(X_1^{(1)}, \ldots, X_n^{(1)}, \ldots, X_1^{(k)}, \ldots, X_n^{(k)}).$$

The problem of fixed precision estimation of the parameter μ consists in finding such a sequence of $\mathscr{F}_n^{(k)}$-measurable estimates $(\hat{\mu}_n)$ and a stopping rule N with the property

$$\{N = n\} \in \mathscr{F}_n^{(k)}$$

that the following conditions are fulfilled:

a) $P_\theta(N < \infty) = 1, \quad (\forall \theta \in R \times U)$

b) $P_\theta(|\hat{\mu}_N - \mu| \leqslant d) \geqslant \alpha, \quad (\forall \theta \in R \times U)$

where d and α are given constants, $d > 0$ and $0 < \alpha < 1$. We call the pair $((\hat{\mu}_n), N)$ a sampling plan. Of course, if no constraints are imposed on the set of nuisance parameters U, then this problem has no solution. A positive answer exists for some special cases only, and now we give a short review of them.

The case of independent identically distributed random variables

The distribution of the observation X_n is denoted by $N(\mu, \sigma^2)$ for every n.

1. *The first case.* The variance σ_0^2 is known.

This means that $\theta = \mu$ and $U = \{\sigma_0\}$. This problem is a trivial one. To obtain the $2d$-length α-confidence interval for μ it is enough to take $k = 1$ and to define the sampling plan as follows:

$$\hat{\mu}_n = \bar{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n},$$

$$n_0 = \inf\left\{n \geqslant 1;\ n \geqslant \frac{a^2 \sigma_0^2}{d^2}\right\}$$

where a is such a real number that $2\Phi(a) - 1 = \alpha$ and Φ is a standard normal distribution function. Then for every real μ

$$P_\mu(|\bar{X}_{n_0} - \mu| \leqslant d) \geqslant \alpha.$$

For a less trivial procedure and the proof of its optimality see [8].

2. *The second case.* The variance σ^2 is unknown and

$$\theta = (\mu, \sigma), \qquad U = \{\sigma;\ \sigma > 0\}.$$

This case is more complicated. The fixed length fixed confidence interval for μ cannot be achieved by means of any deterministic stopping rule.

THEOREM ([10]). *Given $d > 0$ and $\alpha \in (0, 1)$, there is no sampling plan $((\hat{\mu}_n), n_0)$ such that for every μ and σ*

$$P_{\mu,\sigma}(|\hat{\mu}_{n_0} - \mu| \leqslant d) \geqslant \alpha$$

provided that:

- *the number k of copies observed is equal to one,*
- *the stopping rule n_0 is a deterministic one,*
- *estimates $\hat{\mu}_n$ are functions of the sufficient statistics: $\hat{\mu}_n = \hat{\mu}_n(\bar{X}_n, S_n^2)$*

where

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

Let us remark that, since the sample mean is a minimax estimate of the mean value μ with respect to the risk function $R(\hat{\mu}_n, \mu) = P_{\mu,\sigma}(|\hat{\mu}_n - \mu| > d)$, it is always assumed that $\hat{\mu}_n = \bar{X}_n$. This implies that the problem of finding the proper sampling plan reduces to that of finding the stopping rule N. The first such stopping rule, the famous two-stage procedure, belongs to Stein ([8]). The story of searching for the best stopping rule is long and not finished yet. The more important works in this field are [3], [5], [6], [9].

Dependent observations

Let us assume now that the subsequent observations $X_1, X_2, \ldots$ can be dependent random variables. Once the independence assumption is relaxed, Blum and Rosenblatt have proved in [2] that fixed precision estimation is impossible if $k = 1$, even in the small class of m-dependent Gaussian sequences. We shall present their result.

Let $\ldots, X_{-1}, X_0, X_1, \ldots$ be a doubly infinite sequence of independent random variables with common normal distribution $N(0, 1)$. The observations $Y_1, Y_2, \ldots$ are of the form

$$Y_n = \mu + X_n + Z_{n,m}$$

where

$$Z_{n,m} = \begin{cases} 0 & \text{if } m = 0, \\ \dfrac{1}{\sqrt{m}}\displaystyle\sum_{j=1}^{m} X_{n-j} & \text{if } m > 0, \end{cases}$$

and $\theta = (\mu, m)$, $U = \{0, 1, 2, \ldots\}$.

THEOREM. *Let $2d > 0$ and $0 < \alpha < \frac{1}{3}$ be given. Under the condition that*

$$2\Phi(d) - 1 < 1 - 3\alpha$$

there exists no $\mathcal{F}_n^{(1)}$-measurable sampling plan terminating with probability one for all real μ and all $m = 0, 1, \ldots$ and leading to a $2d$-length $(1-\alpha)$-confidence interval for μ.

The intuitive reason for such a phenomenon is as follows: If m is large and $Z_{1,m} = z$, then the $Z_{n,m}$'s vary only slightly with n. So, by observing only one copy of the process (Y_n), it is impossible to distinguish between the case where m is large and the case where $m = 0$ and the mean value is equal to $\mu + z$.

In [11] a sampling plan for the Blum–Rosenblatt process is given provided that $k \geqslant 4$ copies of this process are observed. In fact, Zieliński considered a more general situation, namely that of a class of processes

$$X_n = \mu + \xi_n \tag{1}$$

where the ξ_n's are normally distributed with zero mean. Let us take $\theta = (\mu, (K_n)_{n=1}^{\infty})$ where K_n is a covariance matrix of the vector $(\xi_1, \xi_2, \ldots, \xi_n)$. Suppose that for some $\delta > 0$, K_n belongs to $U(\delta)$, where

$$U(\delta) = \left\{(K_n)_{n=1}^{\infty};\ \sum_{i=1}^{n} \sum_{j=1}^{n} E_{(\mu,(K_n))} \xi_i \xi_j = 0(n^{2-\delta}),\ n \to \infty\right\}.$$

Then the following theorem is true:

THEOREM. *Given $d > 0$ and $\alpha \in (0, 1)$ there exist such a positive integer k and such an $\mathscr{F}_n^{(k)}$-measurable sampling plan $((\hat{\mu}_n), N)$ that N terminates with probability one and for every $\theta \in R \times U(\delta)$*

$$P_\theta(|\hat{\mu}_N - \mu| \leqslant d) \geqslant \alpha.$$

It is easy to show that the Blum–Rosenblatt example can be covered by this theorem. Zieliński gave an explicit form of the sampling plan too. It will be presented in due course.

Now we shall consider a slightly more general class of processes. We would like to make a $2d$-length α-confidence interval for the "nearly" mean value μ of the process

$$X_n = \mu + \mu_n + \xi_n \tag{2}$$

where μ_n tends to zero as n tends to infinity, ξ_n are Gaussian random variables: $\xi_n \sim N(0, \sigma_n^2)$ and $\sigma_n \to 0$ as n tends to infinity. The unknown parameter θ is of the form

$$\theta = (\mu, (\mu_n), (\sigma_n)),$$

where the sequence $((\mu_n), (\sigma_n))$ belongs to a certain known set U.

Such a process is general enough to include many important cases:

a) Let us take the i.i.d. sequence (X_n) where $X_n \sim N(\mu, \sigma^2)$ and make the transformation

$$Y_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

It is easily seen that the process (Y_n) has the properties required for process (2) with $\mu_n = 0$ and $\sigma_n = \sigma n^{-1/2}$.

b) If (X_n) is the process from the Blum–Rosenblatt example, then $X_n \sim N(\mu, 2)$, and the same transformation as in the preceding example makes Y_n be of the form (2).

c) It is easy to show that processes with the properties of process (1) are contained in the class just defined.

d) Let (X_n) be a Robbins–Monro procedure for finding the zero of a regression function f (see for example [1]). Let μ be the zero of f,

$$f(\mu) = 0,$$

and assume that noises are normally distributed. If the assumptions of the convergence theorem for the Robbins–Monro procedure ([1]) are fulfilled, then X_n can be written in the form $X_n = \mu + \mu_n + \xi_n$, where all the properties of process (2) hold.

Now we shall define, after Zieliński, the sampling plan. The estimate $\hat{\mu}_n$ of the parameter μ is the sample mean over k copies:

$$\hat{\mu}_n = \frac{X_n^{(1)} + X_n^{(2)} + \ldots + X_n^{(k)}}{k}.$$

The stopping rule N is described as follows:

$$N = N(n_1, d, (c_n)) = \begin{cases} \inf\{n \geqslant n_1;\ c_n S_n \leqslant d\} & \text{if such } n \text{ exists,} \\ \infty & \text{otherwise,} \end{cases}$$

where

$$S_n^2 = \frac{1}{k-1} \sum_{i=1}^{k} (X_n^{(i)} - \hat{\mu}_n)^2$$

is a sample variance over k copies, n_1 is a fixed positive integer and (c_n) is a fixed sequence of positive reals.

The properties of the stopping rule N are described by the following two lemmas:

LEMMA 1. *If for each* $\theta \in R \times U$

$$\lim_{n \to \infty} c_n \sigma_n = 0, \tag{3}$$

then for every $n_1 \geqslant 1$, *for every* $\theta \in R \times U$ *and for every* $d > 0$

$$P_\theta\big(N(n_1, d, (c_n)) < \infty\big) = 1$$

provided $k \geqslant 2$.

LEMMA 2. *Let d be greater than zero. If a sequence (c_n) and the number of copies k are chosen in such a way that for every $\theta \in R \times U$*

$$\sum_{n=1}^{\infty} c_n \sigma_n \exp\left(-\frac{(k-1)d^2}{4c_n^2 \sigma_n^2}\right) < \infty, \tag{4}$$

then for each n_1 and $m = 1, 2, \ldots$

$$E_\theta[N(n_1, d, (c_n))]^m < \infty \qquad (\forall \theta \in R \times U).$$

The proofs are easy and are based on the fact that the random variable S_n^2 has chi-square distribution with $k-1$ degrees of freedom.

In the sequel we denote by $T(k, \delta)$ a random variable with the noncentral Student distribution with k degrees of freedom and the noncentrality parameter δ. For such a random variable the following lemma can easily be proved:

LEMMA 3. *If x is greater than zero, then*

$$P(|T(k, \delta)| > x) \leqslant [B_1(k) + B_2(k)|\delta|^k] x^{-k},$$

where $B_1(k)$ and $B_2(k)$ are some constants which do not depend on x.

Now we can formulate the main theorem.

THEOREM. *Let α be a fixed confidence level and let $2d$ be the prescribed length of the confidence interval. Let the number of copies k and the sequence (c_n) be such that for every $\theta \in R \times U$*

a) $\lim_{n \to \infty} c_n \sigma_n = 0$,

b) $\sum_{n=1}^{\infty} \left|\frac{\mu_n}{\sigma_n}\right|^{k-1} \frac{1}{c_n^{k-1}} < \infty$,

c) $\sum_{n=1}^{\infty} \frac{1}{c_n^{k-1}} < \infty$.

Then there exists an $n_1 = n_1(\alpha)$ such that $N = N(n_1, d, (c_n))$ has the following properties:

1) $P_\theta(N < \infty) = 1 \quad (\forall \theta \in R \times U)$,

2) $P_\theta(|\hat{\mu}_N - \mu| \leqslant d) \geqslant \alpha \quad (\forall \theta \in R \times U)$.

Additionally, if (c_n) fullfils the condition

d) $\limsup_{n \to \infty} (c_n \sigma_n \sqrt{\log n}) \leqslant \frac{1}{2} d \quad (\forall \theta \in R \times U)$,

then

3) $E_\theta N^m < \infty$ *for every* $m = 1, 2, \ldots$ *and every* $\theta \in R \times U$.

We shall sketch the proof of statement 2). The definition of the stopping rule N implies that

$$P_\theta(|\hat{\mu}_N-\mu|>d)\leqslant \sum_{n=n_1}^{\infty} P_\theta(|\hat{\mu}_n-\mu|>d,\ d>c_n S_n)$$
$$\leqslant \sum_{n=n_1}^{\infty} P_\theta\left(\frac{\sqrt{k}\,|\hat{\mu}_n-\mu|}{S_n}>c_n\sqrt{k}\right).$$

The random variable $\sqrt{k}(\hat{\mu}_n-\mu)/S_n$ has a $T(k-1,\ c_n\sigma_n^{-1}\sqrt{k})$ distribution, which is independent of μ. This statement and Lemma 3 yield

$$P_\theta(|\hat{\mu}_N-\mu|>d)\leqslant \frac{B_1(k)}{k^{(k-1)/2}}\sum_{n=n_1}^{\infty}\frac{1}{c_n^{k-1}}+\frac{B_2(k)}{k^{(k-1)/2}}\sum_{n=n_1}^{\infty}\frac{|\delta_n|^{k-1}}{c_n^{k-1}},$$

where $\delta_n=\mu_n\sigma_n^{-1}\sqrt{k}$. The series on the right-hand side of the above inequality are convergent by the assumptions b) and c). Now it is easily seen that there exists an $n_1=n_1(\alpha)$ such that

$$P_\theta(|\hat{\mu}_N-\mu|>d)<1-\alpha,$$

which ends the proof.

Statement 3) is a result of combining assumption c) of this theorem and assumption (4) of Lemma 2.

One can ask how to use this theorem. Let us suppose that two sequences (a_n) and (b_n) are given such that for every $\theta\in R\times U$

$$|\mu_n\sigma_n^{-1}|\leqslant b_n \qquad\text{and}\qquad \sigma_n=O(a_n).$$

In other words, one has to know that $\theta\in R\times U$ and

$$U=U\big((a_n),(b_n)\big)=\left\{((\mu_n),(\sigma_n));\ \mu_n\to 0,\ \sigma_n\to 0,\ \left|\frac{\mu_n}{\sigma_n}\right|\leqslant b_n,\ \sigma_n=O(a_n)\right\}.$$

If sequences (a_n) and (b_n) are given, then it is possible to compute the value of $n_1(\alpha)$, which defines the stopping rule allowing the construction of a $2d$-length α-confidence interval for μ.

Remark. It can happen that the sequence (c_n) fulfilling the conditions of the theorem does not exist. For example, if

$$\mu_n=0 \qquad\text{and}\qquad \sigma_n=O\left(\frac{1}{\log n}\right)$$

then, given k, there is no sequence (c_n) such that

$$\frac{c_n}{\log n}\to 0 \qquad\text{and}\qquad \sum_{n=1}^{\infty}\frac{1}{c_n^{k-1}}<\infty.$$

Let us consider the special case

$$\sigma_n = O\left(\frac{1}{\sqrt{n}}\right).$$

It is easily seen that this case covers all four examples governed by the general model (2).

Let us take $c_n = n^{\beta}$, $\beta > 0$. Conditions $c_n \sigma_n \to 0$ and $\sum_{n=1}^{\infty} c_n^{-(k-1)} < \infty$ yield the inequality

$$\frac{1}{k-1} < \beta < \frac{1}{2},$$

and this implies that the number of observed copies k has to be greater than 3: $k \geqslant 4$. Additionally we have

$$c_n \sigma_n \sqrt{\log n} \to 0,$$

which implies that all moments of the stopping rule N exist. Of course $k = 4$ is a minimal number of copies. It can be greater than four in order to fulfil condition c) of the theorem.

Can we decrease the number of copies? The change of the sequence (c_n) does not help since there is no such a sequence (c_n) that

$$\frac{c_n}{\sqrt{n}} \to 0 \quad \text{and} \quad \left(\sum_{n=1}^{\infty} \frac{1}{c_n} < \infty \text{ or } \sum_{n=1}^{\infty} \frac{1}{c_n^2} < \infty\right).$$

In order to decrease the number of copies one has to obtain sharper bounds in the proof of statement 2) of the theorem.

Let us go back to the Robbins–Monro procedure. It can be shown that $a_n = O(n^{-1/2})$ provided the assumptions of the Sacks theorem ([4]) on asymptotic normality hold. These assumptions ensure, too, that there exists a sequence (b_n) such that the conditions of our theorem are fulfilled but the explicit form of that sequence is not known as yet.

References

[1] J. R. Blum, *Approximation methods which converge with probability one*, Ann. Math. Statist. **25** (1954), 382–386.

[2] J. R. Blum and J. Rosenblatt, *On fixed precision estimation in time series*, Ann. Math. Statist. **40** (1969), 1021–1032.

[3] W. D. Ray, *Sequential confidence intervals for the mean of a normal distribution with unknown variance*, J. Roy. Statist. Soc. Ser. B **19** (1956), 133–143.

[4] J. Sacks, *Asymptotic distribution of stochastic approximation procedures*, Ann. Math. Statist. **29** (1958), 373–405.

[5] G. Simons, *On the cost of not knowing the variance when making a fixed-width confidence interval for the mean*, Ann. Math. Statist. **39** (1968), 1946–1952.

[6] N. Starr, *The performance of a sequential procedure for the fixed-width interval estimation of the mean*, Ann. Math. Statist. **37** (1966), 36–50.

[7] C. Stein, *A two-sample test for a linear hypothesis whose power is independent of the variance*, Ann. Math. Statist. **16** (1945), 243–258.

[8] C. Stein and A. Wald, *Sequential confidence intervals for the mean of a normal distribution with known variance*, Ann. Math. Statist. **18** (1947), 427–433.

[9] M. Woodroofe, *Second order approximations for sequential point and interval estimation*, Ann. Math. Statist. **48** (1977), 984–995.

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

[11] R. Zieliński, *Fixed precision estimate of mean of a Gaussian sequence with unknown covariance structure*, in *Math. Statist. and Prob. Theory, Proc. Sixth Int. Conf. Wisla (Poland)*, Lecture Notes in Statist. **2**, Springer-Verlag, Berlin 1978.

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