On Adaptive Estimation Using a Prior Guess

YURI G. DMITRIEV AND PETER F. TARASSENKO
Department of Applied Mathematics and Cybernetics,
Tomsk State University, Tomsk, Russia
e-mail: dmit@mail.tsu.ru, ptara@mail.tsu.ru

Abstract

Statistical estimators of a linear functional of an unknown distribution are considering based on combined estimator in the form of weighted sum of non-parametric estimator and a prior guess about the value of this functional. The optimal (in terms of mean square error) weighting coefficient is subject of adaptive estimation itself. A series of k–adaptive estimators are constructed by using the prior guess recursively k times. Examples of combined estimators and results of numerical calculations are provided, that illustrates how the difference between prior guess and unknown value of functional affects the limit distributions of estimators and their probabilistic characteristics.

Keywords: linear functional, prior guess, a priori information, combined estimator, nonparametric estimator, k - adaptive combined estimator.

Introduction

The term 'prior guess' has been probably first introduced by Ferguson [11] and used later in various contexts. There are many papers in the literature devoted to the estimation of the probability characteristics with using additional information (prior guess). Combined statistical estimators adapting a prior guess and their properties have been considered in [2], [8], [9], [10], [17]. Estimators of the mean were proposed in [1], [3], [13], [18]. Estimators of the variance of finite samples have been considered in [4] and [19]. Estimators of conditional quantile have been developed in [19]. In [16] this problem was considered for dependent data. A new class of M–estimators with auxiliary information has been introduced in [14]. Missing data case presented in [7], censored data case has been considered in [15]. Problems of adaptive classification and optimization are considered in [5].

In this paper we consider the case when there exists an assumption on the value of estimated functional. The assumed value we will refer to as a prior guess. We propose k–adaptive combined estimators that use prior guess recursively k times. Asymptotic distributions of the estimators have been obtained, that allow to study the influence of a prior guess to the estimation accuracy.

The obtained asymptotic results extend the results presented in the paper [10].

1 Structure of estimator utilizing a prior guess

Let $X_1, ..., X_n$ be independent observations of size n over a random variable X with unknown distribution function F on R^1. Following to [10], consider the problem of statistical estimation of a linear functional on a certain class of distributions $\mathcal{F}$,

$$J(F) = M_F[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x)dF(x), \quad F \in \mathcal{F}, \tag{1}$$

using a prior guess J_a as a possible value of $J(F)$, specified by researcher. The real function φ is known. Nonparametric estimator of the functional is

$$\hat{J} = J(F_n) = n^{-1}\sum\nolimits_{i=1}^{n} \varphi(X_i),$$

where $F_n(x) = n^{-1}\sum_{i=1}^{n} c(x - X_i)$ is empirical distribution function, $c(t) = \{0 : t < 0; 1, t \geq 0\}$. Following to [8], [9], [10], [17], consider the combined estimator utilizing simultaneously $\hat{J}$ and prior guess J_a in the form

$$\hat{J}(\lambda) = (1-\lambda)\hat{J} + \lambda J_a = \hat{J} - \lambda(\hat{J} - J_a), \tag{2}$$

where the weighting coefficient λ is selected from the minimum of mean square error (MSE) $S_F(\lambda) = M_F[\hat{J}(\lambda) - J]^2$. Optimal value of λ is given by

$$\lambda^* = \lambda^*(F) = (1 + n\Delta^2/\sigma^2)^{-1} = (1 + b_n^2(F))^{-1}, \tag{3}$$

where $\sigma^2 = \sigma^2(F) = D_F(\varphi(X))$ is the variance of $\varphi(X)$, $\Delta = \Delta(F) = J(F) - J_a$ is the value of displacement of the prior guess from the true value $J(F)$, and $b_n(F) = \sqrt{n}\Delta(F)/\sigma(F)$ is the normalized displacement.

The minimal value of MSE is given by the expression $nS_F(\lambda^*) = \sigma^2(1 - \lambda^*)$. The weighting factor λ^* varies between $0 < \lambda_n^* \leq 1$, and shows contribution of each estimator to the combined estimator (2) and their influence to the optimal MSE. Particularly, if $\Delta_F = 0$, we have $\lambda^* = 1$, and prior guess J_a should be taken as the estimator of the functional $J(F)$. When $\Delta_F \neq 0$, which usually happens in practice, $\lambda^* < 1$, and $\lambda^* \to 0$ with the growth of sample size ($n \to \infty$), so the influence of a prior guess and the advantage in the estimation accuracy decrease. More conclusions can be obtained if we assume that Δ decreases simultaneously with growth of n such that for each fixed $F \in \mathcal{F}$ there exists a limit $\lim b_n(F) = b$. Then $\lim nS_F(\lambda^*) = \sigma^2 b^2/(1 + b^2)$.

Practical usage of the combined estimator (2) is complicated because optimal coefficient λ^* is not possible to calculate due to distribution function F is unknown.

Construction of statistical estimators for λ^* leads to adaptive estimation of the functional (1). However, the weighting coefficient becomes non-optimal, and the question arises, under what conditions the adaptive estimator is more preferable by MSE as compared to the estimator $\hat{J}$. We consider this issue in the following sections.

2 Adaptive estimators and their asymptotic properties

We construct adaptive estimators by the method of substitution and consequent use of a prior guess. Let substitute unknown F with F_n in (3) and let use a prior guess

σ_a instead of σ. Then we do have the first estimator for λ^*:

$$\hat{\lambda}_1 = (1 + n\hat{\Delta}^2/\sigma_a^2)^{-1} = (1 + \hat{b}_n^2)^{-1},$$

where $\hat{\Delta} = \hat{J} - J_a$ is estimator of displacement Δ, $\hat{b}_n = \sqrt{n}\hat{\Delta}/\sigma_a$ is estimator of normalized displacement. By substitution λ with $\hat{\lambda}_1$ in (2), we obtain the first adaptive combined estimator $\hat{J}_1 = \hat{J} - \hat{\lambda}_1(\hat{J} - J_a)$. Using $\hat{J}_1$ in estimation of displacement Δ, we obtain $\hat{\Delta}_1 = \hat{J}_1 - J_a$ and $\hat{b}_{1,n} = \sqrt{n}\hat{\Delta}_1/\sigma_a$. Then the second estimator will be given by $\hat{\lambda}_2 = (1 + \hat{b}_{1,n}^2)^{-1}$ and $\hat{J}_2 = \hat{J} - \hat{\lambda}_2(\hat{J} - J_a)$. After repeating this procedure k times consecutively, we obtain the following expressions for the estimator

$$\hat{J}_k = \hat{J} - \hat{\lambda}_k(\hat{J} - J_a) = J_a + (1 - \hat{\lambda}_k)(\hat{J} - J_a), \quad (4)$$
$$\hat{\lambda}_k = \left(1 + n\hat{\Delta}_{k-1}^2/\sigma_a^2\right)^{-1} = \left(1 + \hat{b}_{k-1,n}^2\right)^{-1},$$

where $\hat{b}_{k,n} = \sqrt{n}\hat{\Delta}_k/\sigma_a$, $\hat{\Delta}_{k-1} = \hat{J}_{k-1} - J_a$, $\hat{\Delta}_0 = \hat{\Delta} = \hat{J} - J_a$, $\hat{b}_{0,n} = \hat{b}_n$.

Let us refer to $\hat{J}_k$ as k–adaptive estimator with parameter σ_a. We emphasize here that the prior guess J_a has been used at each step of estimation of Δ, but unknown value σ is replaced by the specified value σ_a. Let us note that in [10] the sample estimate $\hat{\sigma}^2$ was used instead of σ^2.

Consider asymptotic behavior of $\hat{J}_k$. Let

$$\hat{\xi}_k = \frac{\sqrt{n}(\hat{J}_k - J)}{\sigma}.$$

Denote

$$\eta_n = \frac{\sqrt{n}(\hat{J} - J)}{\sigma}, \qquad \tau = \frac{\sigma}{\sigma_a}.$$

Then we can write

$$\hat{b}_n = (\eta_n + b_n)\tau, \quad \hat{b}_{k,n} = q_k(\hat{b}_n) = q_k((\eta_n + b_n)\tau),$$
$$\hat{\lambda}_k = \left[1 + q_{k-1}^2((\eta_n + b_n)\tau)\right]^{-1},$$
$$\hat{\xi}_k = \frac{\sqrt{n}(\hat{J}_k - J)}{\sigma} = -b_n + q_k((\eta_n + b_n)\tau)/\tau,$$

where $q_k(x) = xq(q_{k-1}(x))$, $k \in \{1, 2, 3, \ldots\}$, $q(x) = x^2/(1 + x^2)$, $q_0(x) = x$.

Theorem 1. *Let $\sigma^2 < \infty$ for each $F \in \mathcal{F}$ and sequence b_n converges to non-random value b as $n \to \infty$. Then for each k the random sequence $\hat{\xi}_k$ converges in distribution to the random variable*

$$\xi_k = -b + q_k((\eta + b)\tau)/\tau \quad \text{if} \quad |b| < \infty, \quad 0 < \tau < \infty.$$

$$P\{\xi_k < x\} = \Phi(q_k^{-1}((x + b)\tau)\tau^{-1} - b), -\infty < x < \infty,$$

where η is the standard normal random variable with distribution function $\Phi(x)$, $q_k^{-1}(x)$ is inverse function.

Proof. Since functions $q_k(x)$ are continuous and monotonically increasing, then the statement of the theorem follows from convergency of η_n to η in distribution by the central limit theorem and the continuity theorem ([6], Chapter 6).

Corollary 1. *Under the conditions of the theorem 1, the following statements hold true.*

1. $\xi_k = \eta$ *if* $|b| = \infty$, $0 < \tau < \infty$.

2. $\xi_k \to \eta$ *in distribution as* $\tau \to \infty$, $|b| < \infty$.

3. If $\tau \to 0$ *and* $|b| < \infty$ *then the distribution of* ξ_k *converges to degenerate distribution at point* $-b$ *(formally,* $\xi_k \to -b$*).*

Proof. The first statement follows from the representation

$$\hat{\xi}_k = \eta_n - \frac{\eta_n + b_n}{1 + q_{k-1}^2((\eta_n + b_n)\tau)}, \tag{5}$$

where the second term converges weakly to zero as $|b_n| \to \infty$ due to the proposition 5 from lemma 1 [10]. The second and third statements of the corollary follows from the limit form of representation (5), convergency of $q_{k-1}^2(x)$ to infinity as $x \to \infty$, and convergency of $q_{k-1}(x)$ to zero as $x \to 0$.

3 Examples of k–adaptive combined estimators and numerical results

In this section we provide some examples of estimators, their asymptotic properties, and results of numeric calculations. Consider the k–adaptive combined estimators (4) $\hat{J}_k$ under $k \in \{1, 2, \ldots\}$.

$$\hat{J}_1 = \hat{J} - \left[1 + \hat{b}_n^2\right]^{-1} (\hat{J} - J_a),$$
$$\hat{J}_2 = \hat{J} - \left[1 + \frac{\hat{b}_n^3}{1 + \hat{b}_n^2}\right] (\hat{J} - J_a).$$

According to lemma 1 [10] where the expression for $q_\infty(x)$ is derived, the limit estimator (obtained after using the prior guess infinite number of times, $k = \infty$), can be written as

$$\hat{J}_\infty = \begin{cases} \hat{J} - \left[1 + \frac{\left(\hat{b}_n - \sqrt{\hat{b}_n^2 - 4}\right)^2}{4}\right]^{-1} (\hat{J} - J_a), & \hat{b}_n \le -2, \\ J_a, & |\hat{b}_n| < 2, \\ \hat{J} - \left[1 + \frac{\left(\hat{b}_n + \sqrt{\hat{b}_n^2 - 4}\right)^2}{4}\right]^{-1} (\hat{J} - J_a), & \hat{b}_n \ge 2. \end{cases}$$

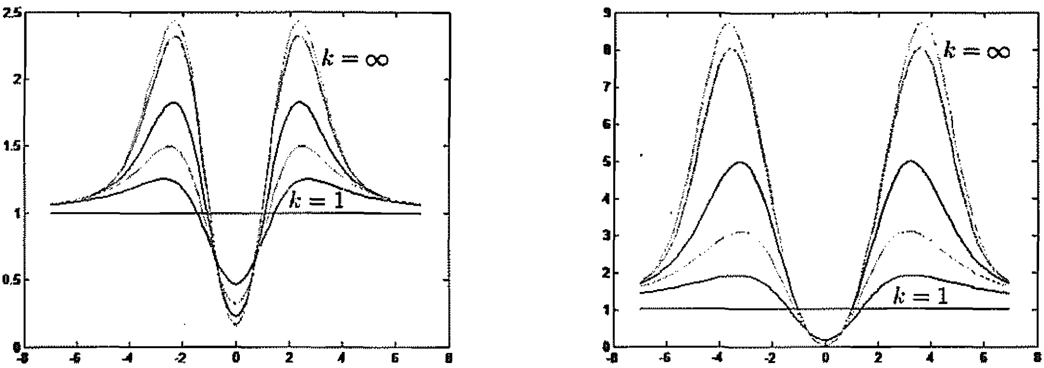

Figure 1: Dependence of the MSE $S\xi_k$ on normalized displacement b and $k \in \{1, 2, 4, 16, \infty\}$ for $\tau = 1.0$ (left plot) and $\tau = 0.5$ (right plot).

Using the theorem 1 we can compute moments of random variable ξ_k. Most interesting is the second moment, which due to (5) can be written in the form

$$M\xi_k^2 = S\xi_k = M\left[\eta - \frac{\eta + b}{1 + q_{k-1}^2((\eta + b)\tau)}\right]^2.$$

Figures 1 and 2 present the plots of $S\xi_k$ in dependence of k, b and τ. At the left plot of figure 1 the case of $\tau = 1$ is considered. It shows that there exist range of values of $|b|$ where $S\xi_k < 1$. Outside the range the combined estimators lose on MSE to regular empirical estimator represented by random variable ξ_0 with $S\xi_0 = 1$. The mentioned intervals and maximal loss are presented in the table 1 in numbers.

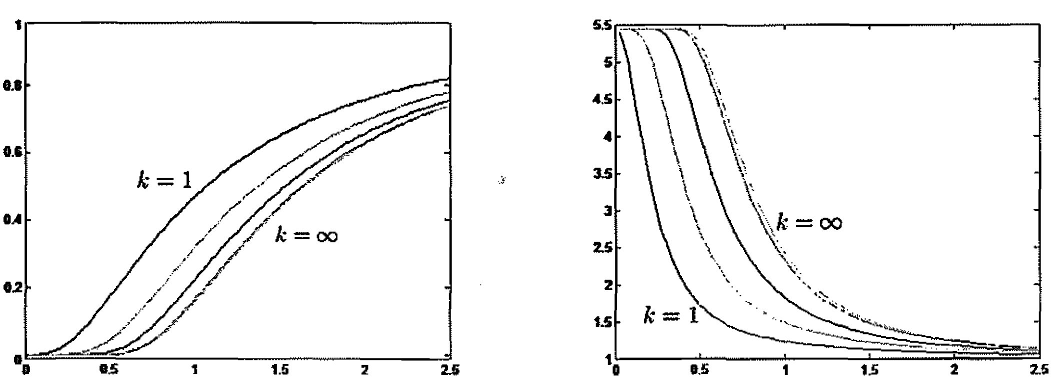

Figure 2: Dependence of the MSE $S\xi_k$ on τ and $k \in \{1, 2, 4, 16, \infty\}$ for normalized displacement $b = 0$ (left plot) and $b = 2.33$ (right plot).

When τ decreases, the maximum of $S\xi_k$ grows and minimum decreases down to zero (see for examples the right plot at the figure 1 and both plots at the figure 2). The inverse behavior is observing when τ increases, in that case $S\xi_k$ tends to

Table 1: Extremal points of $S\xi_k$ under $\tau = 1$ and points of its intersection with level one are presented with accuracy ± 0.07.

k	1	2	4	16	∞
$\max_b S\xi_k$	1.25	1.49	1.82	2.31	2.43
$\arg\max_b S\xi_k$	± 2.66	± 2.52	± 2.38	± 2.38	± 2.24
$b:\ S\xi_k < 1$	$\lvert b\rvert < 1.40$	$\lvert b\rvert < 1.26$	$\lvert b\rvert < 1.12$	$\lvert b\rvert < 0.98$	$\lvert b\rvert < 0.98$

$S\xi_0 = 1$ for all b and τ. In the case of $b = 0$ (left plot at the figure 2) the value of $S\xi_k < S\xi_0 = 1$ for all $0 < \tau < \infty$, and this advantage is increasing with growth of k.

Acknowledgement

The authors thank Tomsk State University for organizational and financial assistance as the article is prepared within the Tomsk State University Competitiveness Improvement Program.

References

[1] Abu-Dayyeh W.A., Ahmed M.S., Ahmed R.A., Muttlak H.A. (2003) Some estimators of a finite population mean using auxiliary information. *Applied Mathematics and Computation.* Vol. 139, pp. 287–298.

[2] Albers C.J., Schaafsma W. (2003) Estimating a density by adapting an initial guess. *Computational Statistics and Data Analysis*, Vol. 42, pp. 27 – 36.

[3] Al-Omari Amer I. (2012) Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling. *Statistics and Probability Letters.* Vol. 82, pp. 1883–1890.

[4] Arcos A., Rueda, M., Martinez M.D., Gonzalez S., Roman Y. (2005) Incorporating the auxiliary information available in variance estimation. *Applied Mathematics and Computation.* Vol. 160, pp. 387–399.

[5] Baklizi A. (2005) Preliminary test estimation in the two parameter exponential distribution with time censored data. *Applied Mathematics and Computation.* Vol. 163, pp. 639–643.

[6] Borovkov A.A. (1998) *Mathematical statistics.* Gordon and Breach Science Publishers, Amsterdam.

[7] Bravo F. (2010) Efficient M-estimators with auxiliary information. *Journal of Statistical Planning and Inference.* Vol. 140, pp. 3326–3342.

[8] Dmitriev Yu.G., Skripin, S.V. (2012) On a combined assessment of the probability of failure-free operation for the full sample. *Tomsk State University Journal of Control and Computer Science*, Vol. 21, issue 4, pp. 32–38.

[9] Dmitriev Yu.G., Tarasenko P.F. (1992) The use of a priori information in the statistical processing of experimental data. *Russian Physics Journal*, September 1992, Vol. 35, Issue 9, pp. 888–893.

[10] Dmitriev Yu.G., Tarassenko P.F., Ustinov Y.K. (2014) On estimation of linear functional by utilizing a prior guess. *Communications in Computer and Information Science. A. Dudin et al. (Eds.): ITMM 2014*. Vol. 487, pp. 82–90.

[11] Ferguson T.S. (1973) A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*, Vol. 1, Issue 2, pp. 209-230.

[12] Han F., Ling Q.-H. (2008) A new approach for function approximation incorporating adaptive particle swarm optimization and a priori information. *Applied Mathematics and Computation*. Vol. 205, pp. 792–798.

[13] Haq A., Shabbir J. (2014) An improved estimator of finite population mean when using two auxiliary attributes. *Applied Mathematics and Computation*. Vol. 241, pp. 14–24.

[14] Liang H.-Y., Jacobo de Una-Alvarez. (2011) Conditional quantile estimation with auxiliary information for left-truncated and dependent data. *Journal of Statistical Planning and Inference*. Vol. 141, pp. 3475–3488.

[15] Liu X., Liu P., Zhou Y. (2011) Distribution estimation with auxiliary information for missing data. *Journal of Statistical Planning and Inference*. Vol. 141, pp. 711–724.

[16] Qin Y.S., Wu Y. (2001) An estimator of a conditional quantile in the presence of auxiliary information. *Journal of Statistical Planning and Inference*. Vol. 99, pp. 59–70.

[17] Tarima S.S., Dmitriev Yu.G. (2009) Statistical estimation with possibly incorrect model assumptions. *Tomsk State University Journal of Control and Computer Science*, Vol. 8, issue 4, pp. 87–99.

[18] Vishwakarma G.K., Singh H.P. (2012) A general procedure for estimating the mean using double sampling for stratification and multi-auxiliary information. *Journal of Statistical Planning and Inference*. Vol. 142, pp. 1252–1261.

[19] Yadav S.K., Kadilar C. (2014) A two parameter variance estimator using auxiliary information. *Applied Mathematics and Computation*. Vol. 226, pp. 117–122.