

**AN ADAPTIVE MOMENT ESTIMATOR  
OF A PARAMETER OF A DISTRIBUTION  
CONSTRUCTED FROM OBSERVATIONS WITH ADMIXTURE**

UDC 519.21

N. LODATKO AND R. MAĬBORODA

**ABSTRACT.** We consider the problem of estimating an unknown parameter from observations with an admixture. The concentration of the admixture is varying with observations and assumed to be known, while its distribution is unknown. We study moment estimators and prove that they are consistent and asymptotically normal. We use an adaptive technique that allows us to determine estimators whose asymptotic variance is minimal among moment estimators.

1. INTRODUCTION

We consider the problem of the estimation from observations with an admixture and continue investigations of [6] in this paper. Now we consider the case where the distribution of the main component of the mixture is known up to a parameter, while the distribution of the admixture is completely unknown. The problem is to estimate the unknown parameter of the main component. The concentration of the components in the mixture is varying with observations and assumed to be known.

We follow the generalized method of moments where we use weighted function moments instead of usual empirical moments to construct an estimator. The function moments are studied in [3, 4]. The consistency and asymptotic normality of the estimators are proved in Section 3. Since the test function used in the construction of the moment estimator is arbitrary, there arises the question of the optimality of this function. The test function that corresponds to the minimal asymptotic variance is found in Section 4. Unfortunately this function depends on the unknown parameter and probability density (also unknown) of the admixture.

For this reason we propose an adaptive approach consisting of two steps. The first step is to estimate the parameter from a rough initial estimator (say, from the non-optimal moment estimator) and the probability density of the admixture. We substitute these estimators into the expression for the optimal test function at the second step. As a result we obtain a random test function and apply it to the adaptive moment estimator. As a rule, the moment equation for this test function has no analytic solution, so we approximate the theoretical moment equation by its differential in a neighborhood of the initial estimator.

It is known that this adaptive scheme leads to an approximation of the maximal likelihood estimators (Wald estimators) that are asymptotically efficient in regular problems in the case of homogeneous observations without an admixture (see [1, Section 2.26]).

---

2000 *Mathematics Subject Classification.* Primary 62G07; Secondary 62G20.

*Key words and phrases.* Method of moments, adaptive estimator, a mixture with varying concentrations, consistency, asymptotic normality, asymptotic variance.

The general theory of adaptive estimation [8, 10] provides conditions that guarantee the asymptotic efficiency of the adaptive estimators irrespective of the quality of estimators obtained at the first step. These conditions do not hold in the scheme of observations with an admixture, unfortunately. Nevertheless we show in Section 5 (Theorem 5.1) that, under certain conditions, the adaptive estimator has the same asymptotic variance as the best moment estimator obtained in Section 4. Assumptions of Theorem 5.1 are similar to some extent to those for the efficiency of the penalization function method of [9]. These assumptions hold, for example, for the truncated exponential distribution if the probability density of the admixture is smooth enough.

## 2. SETTING OF THE PROBLEM

We consider the observations with an admixture; that is, the data is of the form

$$\Xi_N = \{\xi_{j:N}, j = 1, \dots, N\},$$

where, for a fixed  $N$ ,  $\xi_{j:N}$  are independent random variables with the distribution function

$$\mathbb{P}\{\xi_{j:N} < x\} = w_{j:N}H_1(x, \vartheta) + (1 - w_{j:N})H_2(x).$$

Here  $w_{j:N}$  is the concentration of the main component of the mixture at the moment when the observation  $j$  is made,  $H_1(x, \vartheta)$  is the distribution function of the main component,  $\vartheta \in \Theta \in \mathbb{R}$  is the unknown parameter,  $H_2$  is the distribution function of the admixture ( $H_2$  is assumed to be unknown). The problem is to estimate the parameter  $\vartheta$  from the observations  $\Xi_N$ .

In what follows we use the same symbol  $H_i$  to denote the distributions (measures) generated by the distribution function  $H_i$ . For convenience, we introduce random variables  $\eta_1$  and  $\eta_2$  whose distribution functions are  $H_1(\cdot, \vartheta)$  and  $H_2$ , respectively. For triangular matrices  $a = (a_{j:N}, j = 1, \dots, N; N = 1, 2, \dots)$ , we put  $\langle a \rangle_N = N^{-1} \sum_{j=1}^N a_{j:N}$ ,  $\langle ab \rangle_N = N^{-1} \sum_{j=1}^N a_{j:N}b_{j:N}$ , and so on. If the limit  $\lim_{N \rightarrow \infty} \langle a \rangle_N$  exists, we denote it by  $\langle a \rangle$ .

We assume throughout the paper that

$$(1) \quad \Delta_N = \langle (w)^2 \rangle_N - (\langle w \rangle_N)^2 > c > 0$$

for some  $c$  and all  $N$ . (Note that the condition  $\Delta_N = 0$  is equivalent to the condition that  $w_{j:N}$  does not depend on  $j$ . Thus condition (1) means that the varying concentration of the main component of the mixture does not become a constant as  $N \rightarrow \infty$ .)

One can use a modified method of moments to estimate  $\vartheta$ . This procedure requires a “test function”  $g: \mathbb{R} \rightarrow \mathbb{R}$ . Consider the function moment of the main component that corresponds to the value  $\vartheta = t$  of the unknown parameter:

$$G_g(t) := G(t) := \mathbf{E}_t g(\eta_1) = \int g(x) H_1(dx, t).$$

According to [3], the sample weighted function moment

$$\hat{g}_N = \frac{1}{N} \sum_{j=1}^N a_{j:N} g(\xi_{j:N})$$

is a nice estimator for  $G(\vartheta)$ , where

$$(2) \quad a_{j:N} = a_{j:N}^1 = \frac{1}{\Delta_N} [(1 - \langle w \rangle_N)w_{j:N} + \langle (w)^2 \rangle_N - \langle w \rangle_N]$$

is the minimax family of weight coefficients for the estimation of the distribution function of the main component (see [7]).

Equating the theoretical moment corresponding to the value  $t$  of the parameter to the corresponding empirical moment we get the generalized moment estimator

$$(3) \quad \hat{\vartheta}_N(g) := \hat{\vartheta}_N := G^{-1}(\hat{g}_N)$$

where  $G^{-1}$  is the inverse function to  $G$ . (The existence of the inverse function  $G^{-1}$  is one of the assumptions of the method of moments for a given test function  $g$ .)

**Examples. 1.** Let  $H_1$  be the exponential distribution with the density

$$h_1(x, \vartheta) = \vartheta e^{-\vartheta x} \mathbb{1}\{x > 0\},$$

$\vartheta \in \Theta = (0, +\infty)$ . Then  $\mathbb{E}_t \eta_1 = 1/t$ . One can take  $g(x) = x$  as a test function in this case. Then the corresponding estimator is

$$\hat{\vartheta}_N = \left( \frac{1}{N} \sum_{j=1}^N a_{j:N} \xi_{j:N} \right)^{-1}.$$

**2.** Let  $H_1$  be the “truncated exponential” distribution with the density

$$(4) \quad h_1(x, \vartheta) = \frac{\vartheta e^{-\vartheta x}}{1 - e^{-\vartheta T}} \mathbb{1}\{x \in (0, T)\}.$$

Here  $\vartheta \in \Theta = (0, +\infty)$  is the unknown parameter, while  $T$  is assumed to be known. The distribution of the admixture for this case can be assumed to have a support on  $[0, T]$ , since all observations outside this interval do not belong to the main component and one can be neglected.

Let the test function be  $g(x) = x$ . Then

$$G(\vartheta) = \frac{1 - e^{\vartheta T} + \vartheta T}{\vartheta(1 - e^{\vartheta T})}.$$

The latter function is increasing and continuous; thus the inverse function exists and estimator (3) is well defined.

### 3. ASYMPTOTIC BEHAVIOR OF MOMENT ESTIMATORS

We study the behavior of moment estimators as the size of a sample increases.

**Theorem 3.1** (consistency). *Let*

- (i) *the moments  $\mathbb{E}|g(\eta_i)| < \infty$  exist for  $i = 1, 2$ ;*
- (ii) *condition (1) hold;*
- (iii) *the function  $G^{-1}$  be well defined and be continuous at the point  $G(\vartheta)$ .*

*Then  $\hat{\vartheta}_N \rightarrow \vartheta$  in probability.*

*Proof.* Theorem 3.3.1 of [5] and assumptions (i)–(ii) imply that  $\hat{g}_N \rightarrow \mathbb{E}g(\eta_1) = G(\vartheta)$ . Now Theorem 3.1 follows, since  $G^{-1}$  is continuous.  $\square$

**Theorem 3.2.** *Let*

- (i) *the moments  $\mathbb{E}(g(\eta_i))^2 < \infty$  exist for  $i = 1, 2$ ;*
- (ii) *condition (1) hold;*
- (iii) *the derivative  $G'(t) = dG(t)/dt$  exist, be continuous and uniformly separated from 0 for all  $t \in \Theta$  (that is, either  $G'(t) > c > 0$  for all  $t \in \Theta$ , or  $G'(t) < c < 0$  for all  $t \in \Theta$ ).*

Then the distribution of  $\sqrt{N}(\hat{\vartheta}_N - \vartheta)/s_{\vartheta,N}$  weakly converges to the standard normal distribution. Here  $s_{\vartheta,N}^2 := \sigma_{\vartheta,N}^2/(G'(\vartheta))^2$  and

$$(5) \quad \begin{aligned} \sigma_{\vartheta,N}^2 &:= \sigma_{\vartheta,N}^2(g) \\ &:= \langle (a)^2 w \rangle_N \mathbf{E}(g(\eta_1))^2 + \langle (a)^2 (1-w) \rangle_N \mathbf{E}(g(\eta_2))^2 \\ &\quad - \left[ \langle (a)^2 (w)^2 \rangle_N (\mathbf{E}g(\eta_1))^2 \right. \\ &\quad \left. + 2\langle (a)^2 w(1-w) \rangle_N \mathbf{E}g(\eta_1) \mathbf{E}g(\eta_2) + \langle (a)^2 (1-w)^2 \rangle_N (\mathbf{E}g(\eta_2))^2 \right]. \end{aligned}$$

**Corollary 3.1.** *Let assumptions (i) and (iii) of Theorem 3.2 hold. If*

(ii') *the limits  $\langle (w)^k \rangle$  exist for  $k = 1, 2, 3, 4$ , and*

$$\Delta = \langle (w)^2 \rangle - (\langle w \rangle)^2 \neq 0,$$

*then  $\sqrt{N}(\hat{\vartheta}_N - \vartheta)$  weakly converges to the normal distribution with zero mean and variance  $s_{\vartheta}^2(g) := s_{\vartheta}^2 := \sigma_{\vartheta}^2/(G'(\vartheta))^2$  where  $\sigma_{\vartheta}^2$  is defined by (5) with  $\langle \cdot \rangle$  instead of  $\langle \cdot \rangle_N$ .*

*Remark.* The coefficients in equality (5) that depend on  $a$  can be expressed in terms of  $\langle (w)^k \rangle$ ,  $k = 1, 2, 3, 4$ , with the help of relation (2).

*Proof of Theorem 3.2.* Assumption (iii) implies assumption (iii) of Theorem 3.1. Thus the estimator  $\hat{\vartheta}_N$  is consistent. Assumptions (i) and (ii) and Theorem 3.3.2 of [5] imply that the distributions of the random variables  $\sqrt{N}(\hat{g}_N - G(\vartheta))/\sigma_{\vartheta,N}$  weakly converge to the standard normal distribution. Using Theorem 3A of [1, Chapter 1, Section 5] we complete the proof of the theorem in view of assumption (iii).  $\square$

*Proof of Corollary 3.1.* Assumption (ii') implies condition (1) and the convergence

$$\sigma_{\vartheta,N}^2 \rightarrow \sigma_{\vartheta}^2.$$

Now the corollary follows from the Slutsky theorem.  $\square$

#### 4. OPTIMAL TEST FUNCTION

According to Corollary 3.1, the best function  $g$  among test functions is determined by the minimum of the asymptotic variance (in other words, in terms of the limit variance) of the estimator  $s_{\vartheta}^2(g)$ . Denote by  $g^*$  the best (optimal) test function. We evaluate  $g^*$  in the case where  $\vartheta$  and  $H_2$  are known. We also assume that the densities of the distributions exist:

$$h_1^{\vartheta}(x) = \frac{\partial H_1(x, \vartheta)}{x}, \quad h_2^{\vartheta}(x) = \frac{\partial H_2(x)}{x}.$$

Some extra assumptions on  $h_1^{\vartheta}$  and  $h_2^{\vartheta}$  will be imposed in the course of the construction of the optimal test function  $g^*$ .

Note that if one uses the test function

$$(6) \quad g(x) = \alpha g^0(x) + \beta$$

instead of  $g^0(x)$ , where  $\alpha$  and  $\beta$  are arbitrary real numbers, then the moment estimator does not in fact change, since the changes of  $\hat{g}_N$  are balanced by the corresponding changes of  $G^{-1}$ . Thus  $s_{\vartheta}^2(g) = s_{\vartheta}^2(g^0)$ . This result can be obtained explicitly by using equality (5).

Therefore one can restrict the search of the optimal test function to the class of appropriately normalized and centered functions  $g$ . We choose the constants  $\alpha$  and  $\beta$  such that

$$(7) \quad G(\vartheta) = \int g(x) h_1^{\vartheta}(x) dx = 0 \quad (\text{that is, } \mathbf{E}g(\eta_1) = 0)$$

and

$$(8) \quad G'(\vartheta) = \frac{\partial}{\partial \vartheta} \int g(x) h_1^\vartheta(x) dx = \int g(x) \dot{h}_1^\vartheta(x) dx = 1.$$

Here  $\dot{h}_1^\vartheta(x) = \partial h_1^\vartheta(x) / \partial \vartheta$ . We add one more condition to the set of conditions imposed on the distributions of the components of the mixture; namely, we assume that the latter derivative exists and that one can interchange the operations of integration and differentiation.

In other words, in order that the function  $g$  satisfying (7)–(8) be obtained with the help of the linear transformation (6) from an arbitrary (nonconstant) function  $g^0$ , it is necessary that the three functions 1,  $h_1^\vartheta$ , and  $\dot{h}_1^\vartheta$  be noncoplanar; that is, the equality  $h_1^\vartheta - \alpha \dot{h}_1^\vartheta - \beta = 0$  does not hold almost surely for any pair of real numbers  $\alpha$  and  $\beta$ .

If this is the case, the problem reduces to the minimization of the functional

$$(9) \quad \begin{aligned} \sigma_\vartheta^2(g) = & \langle (a)^2 w \rangle \int (g(x))^2 h_1^\vartheta(x) dx \\ & + \langle (a)^2 (1-w) \rangle \int (g(x))^2 h_2(x) dx - \langle (a)^2 (1-w)^2 \rangle \left( \int g(x) h_2(x) dx \right)^2 \end{aligned}$$

given

$$(10) \quad \begin{cases} \int g(x) h_1^\vartheta(x) dx = 0, \\ \int g(x) \dot{h}_1^\vartheta(x) dx = 1. \end{cases}$$

Note that  $\sigma_\vartheta^2(g)$  is a nonnegative definite quadratic form in the linear space of corresponding functions  $g$ . This quadratic form is nonnegative definite, indeed, because  $\sigma_\vartheta^2(g)$  is the limit variance of  $\hat{g}_N$ . Conditions (10) define an affine subspace of this space. Thus the minimum of  $\sigma_\vartheta^2(g)$  is always attained; however it can be nonunique. One can evaluate the minimum by using the Lagrange multipliers method.

Put

$$(11) \quad \begin{aligned} I &= \int g(x) h_2(x) dx, \\ z(x) &= \langle (a)^2 w \rangle h_1^\vartheta(x) + \langle (a)^2 (1-w) \rangle h_2(x), \\ \gamma &= \langle (a)^2 (1-w)^2 \rangle. \end{aligned}$$

The Lagrange function is given by

$$\mathcal{L}(g) = \int (g(x))^2 z(x) dx - \gamma \left( \int g(x) h_2(x) dx \right)^2 + \lambda_1 \int g(x) h_1^\vartheta(x) dx + \lambda_2 \int g(x) \dot{h}_1^\vartheta(x) dx,$$

where  $\lambda_1$  and  $\lambda_2$  are Lagrange multipliers. The differential  $\mathcal{L}$  is such that

$$\delta \mathcal{L}(g) = \int \left( 2g(x) z(x) + \lambda_1 \dot{h}_1^\vartheta(x) + \lambda_2 \ddot{h}_1^\vartheta(x) - 2\gamma I h_2(x) \right) \delta(x) dx.$$

We have  $\delta \mathcal{L}(g) = 0$  at any critical point for all possible increments of the argument  $\delta(x)$ . Thus the extremum is attained at the point

$$(12) \quad g^*(x; \vartheta, h_2) := g^*(x) := \frac{\lambda_1 h_1^\vartheta(x) + \lambda_2 \dot{h}_1^\vartheta(x) + \gamma I h_2(x)}{\langle (a)^2 w \rangle h_1^\vartheta(x) + \langle (a)^2 (1-w) \rangle h_2(x)},$$

where  $\lambda_1$ ,  $\lambda_2$ , and  $I$  can be found from the system of equations (corresponding to conditions (10)–(11)):

$$(13) \quad \begin{cases} \lambda_1 \int \frac{(h_1^\vartheta(x))^2}{z(x)} dx + \lambda_2 \int \frac{h_1^\vartheta(x)h_1^\vartheta(x)}{z(x)} dx + I\gamma \int \frac{h_2(x)h_1^\vartheta(x)}{z(x)} dx = 0, \\ \lambda_1 \int \frac{h_1^\vartheta(x)h_1^\vartheta(x)}{z(x)} dx + \lambda_2 \int \frac{(h_1^\vartheta(x))^2}{z(x)} dx + I\gamma \int \frac{h_2(x)h_1^\vartheta(x)}{z(x)} dx = 1, \\ \lambda_1 \int \frac{h_1^\vartheta(x)h_2(x)}{z(x)} dx + \lambda_2 \int \frac{h_1^\vartheta(x)h_2(x)}{z(x)} dx + I\gamma \int \frac{(h_2(x))^2}{z(x)} dx = I. \end{cases}$$

If a function  $g^*$  satisfies the assumptions of Theorem 3.2, then it is an optimal test function for the method of the moment estimator  $\vartheta$ .

## 5. ADAPTIVE ESTIMATOR

If  $\vartheta$  and  $h_2$  are unknown, then one cannot use the optimal test function (12) to construct an estimator. The solution of this problem can be obtained in the framework of the adaptive approach.

To construct an adaptive estimator we first estimate  $\vartheta$  and  $h_2$  from rough “initial” estimators  $\tilde{\vartheta}_N$  and  $\tilde{h}_{2,N}$ . Then we substitute these estimators into (12). The function  $g_N^*(x) := g^*(x; \tilde{\vartheta}_N, \tilde{h}_{2,N})$  obtained in this way is used as a test function to construct the moment estimator, namely  $\vartheta_N^* = G_{g_N^*}^{-1}(\hat{g}_N^*)$ , where  $\hat{g}_N^* = N^{-1} \sum_{j=1}^N a_{j:N} g_N^*(\xi_{j:N})$ .

There are two problems when following this approach.

First, it is not clear at all whether or not the inverse function  $G_{g_N^*}^{-1}$  exists for the test function  $g_N^*$  constructed in this way. Moreover, it is not easy to evaluate  $G_{g_N^*}^{-1}$  provided it exists.

Second, since  $g_N^*$  is a random function depending on the data, one cannot apply Theorem 3.2 to the corresponding estimator, and thus an additional analysis is required to study the asymptotic behavior of the estimator.

The first problem can be solved by applying the Newton approximation method for solving the equation

$$(14) \quad G_{g_N^*}(t) = \hat{g}_N^*$$

with the initial approximation  $\tilde{\vartheta}_N$ . Indeed, expanding the function  $G_{g_N^*}(t)$  in the Taylor series in a neighborhood of  $\tilde{\vartheta}_N$  we approximate equation (14) by the following equation:

$$(15) \quad G_{g_N^*}(\tilde{\vartheta}_N) + G'_{g_N^*}(\tilde{\vartheta}_N)(t - \tilde{\vartheta}_N) = \hat{g}_N^*.$$

Since the function  $g_N^*$  satisfies conditions (7)–(8) with  $\vartheta = \tilde{\vartheta}_N$ , we get  $G_{g_N^*}(\tilde{\vartheta}_N) = 0$  and  $G'_{g_N^*}(\tilde{\vartheta}_N) = 1$ . Thus a solution of (15) is of the form

$$(16) \quad \check{\vartheta}_N = \tilde{\vartheta}_N + \hat{g}_N^*.$$

We refer to this solution as an adaptive generalized moment estimator for  $\vartheta$ .

Now we obtain conditions such that the estimator  $\check{\vartheta}_N$  is asymptotically normal and has the optimal asymptotic variance  $s_{\check{\vartheta}}^2(g^*)$ . We state the corresponding assertion for a somewhat more general case of the adaptive estimation than is needed for the above problem.

Let  $\mathcal{A}$  be a measurable space and  $g: \mathbb{R} \times \mathcal{A} \rightarrow \mathbb{R}$  a fixed nonrandom measurable function. Put

$$G(\vartheta, \alpha) = \mathbb{E}_\vartheta g(\eta_1, \alpha) = \int g(x, \alpha) h_1^\vartheta(x) dx,$$

$$\hat{g}(\alpha) = \frac{1}{N} \sum_{j=1}^N a_{j:N} g(\xi_{j:N}, \alpha).$$

Assume that an initial estimator  $\tilde{\vartheta}_N$  is constructed for  $\vartheta$  and that a sequence  $\{\alpha_N\}$  of random elements (estimators) of  $\mathcal{A}$  approximates the nonrandom element  $\alpha_\infty$ . We assume that  $\alpha_N \in \mathcal{A}_N$  where  $\mathcal{A}_N$  is a sequence of nonrandom subsets of  $\mathcal{A}$ . Consider the estimator

$$(17) \quad \check{\vartheta}_N = \tilde{\vartheta}_N + \frac{\hat{g}_N(\alpha_N) - G(\tilde{\vartheta}_N, \alpha_N)}{G'(\tilde{\vartheta}_N, \alpha_N)},$$

where  $G'(t, \alpha) = \partial G(t, \alpha)/\partial t$ .

*Remark.* If  $\hat{h}_{2,N}$  is an estimator for  $h_2$ ,

$$g(x, \alpha) = g^*(x, \vartheta_N, \hat{h}_{2,N}),$$

and if  $\alpha_N = (\tilde{\vartheta}_N, h_2)$ , then equality (17) becomes equality (16).

Denote by  $\mathcal{F}_N$  the family of all sets  $\{x \in \mathbb{R}: g(x, \alpha) - g(x, \alpha_\infty) < C\}$  considered for all possible  $\alpha \in \mathcal{A}_N$  and  $C \in \mathbb{R}$ . Let  $\nu_N(l) = \nu(l, \mathcal{F}_N)$  be the growth function of the family  $\mathcal{F}_N$ . (The definition of growth functions is given in Section 6 below.)

Let

$$\begin{aligned} \sigma_\infty^2 &:= \lim_{N \rightarrow \infty} \text{Var} \left( \frac{1}{\sqrt{N}} \sum_{j=1}^N a_{j:N} g(\xi_j, \alpha_\infty) \right) \\ &= \langle (a)^2 w \rangle \mathbf{E}(g(\eta_1, \alpha_\infty))^2 + \langle (a)^2 (1-w) \rangle \mathbf{E}(g(\eta_2, \alpha_\infty))^2 \\ &\quad - \left[ \langle (a)^2 (w)^2 \rangle (\mathbf{E} g(\eta_1, \alpha_\infty))^2 \right. \\ &\quad \left. + 2 \langle (a)^2 w(1-w) \rangle \mathbf{E} g(\eta_1, \alpha_\infty) \mathbf{E} g(\eta_2, \alpha_\infty) + \langle (a)^2 (1-w)^2 \rangle (\mathbf{E} g(\eta_2, \alpha_\infty))^2 \right]. \end{aligned}$$

**Theorem 5.1.** *Let*

- (i) *the moments  $\mathbf{E}(g(\eta_i, \alpha_\infty))^2 < \infty$  exist for  $i = 1, 2$ ;*
- (ii) *the limits  $\langle (w)^k \rangle$  exist for  $k = 1, 2, 3, 4$ , and*

$$\Delta = \langle (w)^2 \rangle - (\langle w \rangle)^2 \neq 0;$$

- (iii)  *$\sup_N \mathbf{P}\{\sqrt{N}(\tilde{\vartheta}_N - \vartheta) > c\} \rightarrow 0$  as  $c \rightarrow \infty$ ;*
- (iv)  *$G'(\vartheta, \alpha_\infty) \neq 0$ , where  $G'(t, \alpha)$  is a continuous function with respect to  $t$  in some neighborhood of  $\vartheta$  for all  $\alpha \in \mathcal{A}$ , and  $G'(t_N, \alpha_N) \rightarrow G'(\vartheta, \alpha_\infty)$  in probability as  $N \rightarrow \infty$  for all random sequences  $t_N$  such that  $t_N \rightarrow \vartheta$  in probability;*
- (v) *for some nonrandom sequence  $\delta_N \rightarrow 0$ ,*

$$\frac{1}{\delta_N} \sup_{x \in \mathbb{R}} |g(x, \alpha_N) - g(x, \alpha_\infty)| \rightarrow 0$$

*in probability as  $N \rightarrow \infty$ ;*

- (vi)  *$\ln \nu_N(2N) = o(\delta^{-2})$ .*

*Then the distribution of  $\sqrt{N}(\check{\vartheta}_N - \vartheta)$  weakly converges to the normal distribution with zero mean and variance  $\sigma_\infty^2 / (G'(\vartheta, \alpha_\infty))$ .*

In fact, we claim in the latter theorem that the asymptotic variance of the adaptive estimator  $\check{\vartheta}_N$  defined by (17) is the same as that of the moment estimator  $g(\cdot, \alpha_\infty)$ . The proof of Theorem 5.1 is provided in Section 6.

Consider an application of Theorem 5.1 to the data considered in Example 2 (where  $H_1$  is a truncated exponential distribution). One can take the moment estimator with the test function  $g(x) = x$  as an initial estimator  $\tilde{\vartheta}_N$  for  $\vartheta$ . We take a weighted histogram as an estimator for  $h_2$ . More precisely, let  $K_N$  be a number of subintervals partitioning

the interval  $[0, T)$  and put  $t_k = kT/K_N$ ,  $k = 0, \dots, K_N$ , where  $A_k = [t_{k-1}, t_k)$  is the subinterval  $k$  of the partition. Then the weighted histogram  $\hat{h}_{2,N}$  is defined by

$$(18) \quad \hat{h}_{2,N}(x) = \frac{K_N}{NT} \sum_{j=1}^N \sum_{k=1}^{K_N} a_{j:N}^2 \mathbb{1}\{\xi_{j:N} \in A_k\},$$

where  $a_{j:N}^2$  are minimax weight coefficients for the estimation of the second component (admixture):

$$(19) \quad a_{j:N}^2 = (\langle (w)^2 \rangle_N - \langle w \rangle_N w_{j:N}) / \Delta_N.$$

**Corollary 5.1.** *Let*

- (i) *the main component have a truncated exponential distribution (4);*
- (ii) *assumption (ii) of Theorem 5.1 hold;*
- (iii) *the distribution of the admixture possess the density  $h_2$  that is a continuously differentiable function on  $[0, T]$ ;*
- (iv)  *$K_N = CN^\beta$  for some  $C > 0$  and  $0 < \beta < 1/4$ ;*
- (v) *the initial estimator  $\tilde{\vartheta}_N$  be the moment estimator with the test function  $g(x) = x$ .*

*Then Theorem 5.1 holds.*

If the support of the distribution  $H_1$  is infinite (such as the exponential distribution considered in Example 1), then one can use the histogram on an interval that expands as the size of the sample grows, say on the interval  $[0, T_N)$ . If  $T_N$  tends to infinity slowly enough in order to fulfill assumptions (v) and (vi) of Theorem 5.1 and, at the same time, quickly enough to guarantee that  $\mathbb{P}\{\sup_j \xi_{j:N} > T_N\} \rightarrow 0$  as  $N \rightarrow \infty$ , then Theorem 5.1 holds. Consider Example 1, for instance: if the tail of the distribution  $H_2$  is subexponential (that is,  $H_2([x, +\infty)) < Ce^{-\alpha x}$  for some constants  $C$  and  $\alpha > 0$ ), then one can put  $T_N = C(\ln N)^{\ln N}$ .

## 6. PROOFS

We need some auxiliary results and definitions for the proof of Theorem 5.1.

Let  $\mathcal{F}$  be some family of subsets of  $\mathbb{R}$  and let  $X = (x^1, \dots, x^l) \in \mathbb{R}^l$ . Every set  $A \in \mathcal{F}$  generates a subsequence  $X^A$  of the sequence  $X$  consisting of those elements of  $X$  that belong to  $A$ . The number of different subsequences  $X^A$  generated by the sets  $A \in \mathcal{F}$  on  $X$  is denoted by  $\delta^{\mathcal{F}}(X)$ . The function  $\nu^{\mathcal{F}}(l) = \max_{X \in \mathbb{R}^l} \delta^{\mathcal{F}}(X)$  is called the growth function of the family  $\mathcal{F}$ .

Let  $\mathcal{G}$  be a family of functions  $g: \mathbb{R} \rightarrow \mathbb{R}$  and let

$$\mathcal{F}_{\mathcal{G}} = \{\{x \in \mathbb{R}: g(x) < c\} \text{ for all } c \in \mathbb{R} \text{ and } g \in \mathcal{G}\}.$$

The growth function of  $\mathcal{F}_{\mathcal{G}}$ :  $\nu^{\mathcal{G}}(l) := \nu^{\mathcal{F}_{\mathcal{G}}}(l)$  is called the growth function of  $\mathcal{G}$ .

The measure

$$\hat{H}_{i,N}(A) = \frac{1}{N} \sum_{j=1}^N a_{j:N}^i \mathbb{1}\{\xi_{j:N} \in A\}$$

is called the weighted empirical measure with coefficients  $a_{j:N}^i$ .

**Lemma 6.1.** *Let  $\mathcal{F}$  be some class of subsets of  $\mathbb{R}$  with the growth function  $\nu^{\mathcal{F}}(l)$ . Assume that condition (1) holds and that  $a^i$  is defined by (2) or by (19). Then*

$$\mathbb{P} \left\{ \sup_{A \in \mathcal{F}} |\hat{H}_{i,N}(A) - H_i(A)| > \lambda \right\} \leq C \nu^{\mathcal{F}}(2N) \exp(-\alpha \lambda^2 N)$$



for some constants  $\lambda_0, C$ , and  $\alpha$  such that  $\lambda > \lambda_0/N$ . Moreover the constants  $\lambda_0 > 0$ ,  $C < \infty$ , and  $\alpha > 0$  depend on the concentrations  $w_{j:N}$  only (that is, they do not depend on  $H_i, \mathcal{F}$ , and  $\lambda$ ).

This lemma is a trivial consequence of the Vapnik–Chervonenkis inequality for mixtures with varying concentrations (Theorem 2.4.2 in [7]).

**Lemma 6.2.** *Let  $\mathcal{G}$  be some class of functions  $f: \mathbb{R} \rightarrow \mathbb{R}$  and let  $K = \sup_{x \in \mathbb{R}, f \in \mathcal{G}} |f(x)|$ . Then*

$$\sup_{f \in \mathcal{G}} \left| \frac{1}{N} \sum_{j=1}^N f(\xi_{j:N}) a_{j:N}^i - \int f(x) H_i(dx) \right| \leq 2K \sup_{A \in \mathcal{F}} |\hat{H}_{i,N}(A) - H_i(A)|.$$

The proof of this result is similar to that of Theorem 13.1 in [2].

*Proof of Theorem 5.1.* Note that

$$\begin{aligned} \sqrt{N}(\check{\vartheta}_N - \vartheta) &= \sqrt{N}(\tilde{\vartheta}_N - \vartheta) + \frac{\sqrt{N}(\hat{g}_N(\alpha_N) - G(\vartheta, \alpha_N))}{G'(\tilde{\vartheta}_N, \alpha_N)} \\ &\quad + \frac{\sqrt{N}(G(\vartheta, \alpha_N) - G(\tilde{\vartheta}_N, \alpha_N))}{G'(\tilde{\vartheta}_N, \alpha_N)}. \end{aligned}$$

Since  $G(\vartheta, \alpha_N) - G(\tilde{\vartheta}_N, \alpha_N) = G'(\zeta_N, \alpha_N)(\vartheta - \tilde{\vartheta}_N)$  where  $\zeta_N$  is a point between  $\vartheta$  and  $\tilde{\vartheta}_N$ , we get

$$\sqrt{N}(\check{\vartheta}_N - \vartheta) = J_1 + J_2/G'(\tilde{\vartheta}_N, \alpha_N) + J_3,$$

where

$$\begin{aligned} J_1 &= \sqrt{N}(\tilde{\vartheta}_N - \vartheta) \left( 1 - \frac{G'(\zeta_N, \alpha_N)}{G'(\tilde{\vartheta}_N, \alpha_N)} \right), \\ J_2 &= \sqrt{N}(\hat{g}_N(\alpha_N) - G(\vartheta, \alpha_N) - \hat{g}_N(\alpha_\infty) + G(\vartheta, \alpha_\infty)), \\ J_3 &= (\hat{g}_N(\alpha_\infty) - G(\vartheta, \alpha_\infty))/G'(\vartheta_N, \alpha_N). \end{aligned}$$

According to assumptions (iii) and (iv),  $J_1 \rightarrow 0$  in probability. By Theorem 3.3.2 in [5],  $J_3$  weakly converges to the normal distribution with zero mean and variance  $\sigma_\infty^2/(G'(\vartheta, \alpha_\infty))$  in view of assumptions (i) and (ii).

It remains to show that  $J_2 \rightarrow 0$  in probability. Let

$$f(x; \alpha) = g(x, \alpha_\infty) - g(x, \alpha)$$

and consider the class of functions  $\mathcal{G}_N := \{f(\cdot; \alpha), \alpha \in \mathcal{A}_N\}$ .

Then

$$\begin{aligned} \hat{g}_N(\alpha_\infty) - \hat{g}_N(\alpha) &= \frac{1}{N} \sum_{j=1}^N a_{j:N} (g(\xi_{j:N}, \alpha_\infty) - g(\xi_{j:N}, \alpha)) = \frac{1}{N} \sum_{j=1}^N a_{j:N} f(\xi_{j:N}), \\ G(\vartheta, \alpha_\infty) - G(\vartheta, \alpha) &= \int f(x, \alpha) H_1(dx, \alpha). \end{aligned}$$

Thus, by Lemma 6.2,

$$\begin{aligned} p_N &:= \mathbb{P}\{|J_2| > \lambda\} \\ &\leq \mathbb{P}\left\{ \sup_x |g(x, \alpha_N) - g(x, \alpha_\infty)| > \delta_N \right\} + \mathbb{P}\left\{ \sup_{A \in \mathcal{F}_\mathcal{G}} |\hat{H}_{1,N}(A) - H_1(A)| > \frac{\lambda \delta^{-1}}{2\sqrt{N}} \right\} \end{aligned}$$

for all  $\lambda$ .

Taking into account assumption (v), the first term on the right hand side approaches 0. The second term does not exceed

$$C\nu_N(2N) \exp(-\alpha\lambda^2\delta_N^{-2}/4) \leq C \exp(-\alpha\lambda^2\delta_N^{-2}/4 + \ln \nu_N(2N)) \rightarrow 0$$

by Lemma 6.1 and assumption (vi).

Theorem 5.1 is proved.  $\square$

The following estimate of the rate of convergence of the weighted histogram  $\hat{h}_{2,N}$  defined by (18) to the true density is needed for the proof of Corollary 5.1.

**Lemma 6.3.** *Let the distributions of the components be concentrated on the interval  $[0, T]$ . Assume that*

- (i)  $h_2$  is a continuously differentiable function on  $[0, T]$ ;
- (ii) assumption (ii) of Theorem 5.1 holds;
- (iii)  $\delta_N = C_1N^{-\beta}$  and  $K_N = C_2N^\gamma$  for some constants  $0 < C_1, C_2, \beta, \gamma < \infty$ .

If  $\beta < \gamma$  and  $\beta + \gamma < 1/2$ , then

$$\frac{1}{\delta_N} \sup_{x \in [0, T]} |\hat{h}_{2,N}(x) - h_2(x)| \rightarrow 0$$

in probability.

*Proof.* Let  $\bar{h}_k = K_N H_2(A_k)/T$ ,  $\bar{h}_{2,N}(x) = \mathbf{E} \hat{h}_{2,N} = \sum_{k=1}^{K_N} \bar{h}_k \mathbb{1}\{x \in A_k\}$ , and

$$h' = \sup_{x \in [0, T]} dh_2(x)/dx.$$

Fix an arbitrary  $\lambda > 0$ . Then the statement of the lemma is equivalent to the following relation:

$$J_N := \mathbf{P} \left\{ \sup_{x \in [0, T]} |\hat{h}_{2,N}(x) - h_2(x)| > \lambda \delta_N \right\} \rightarrow 0$$

as  $N \rightarrow \infty$ . It is easy to see that  $J_N \leq J_N^1 + J_N^2$ , where

$$J_N^1 = \mathbf{P} \left\{ \sup_{x \in [0, T]} |\hat{h}_{2,N}(x) - \bar{h}_{2,N}(x)| > \lambda \delta_N / 2 \right\},$$

$$J_N^2 = \mathbb{1} \left\{ \sup_{x \in [0, T]} |\bar{h}_{2,N}(x) - h_2(x)| > \lambda \delta_N / 2 \right\}.$$

Let  $x \in A_k$ . We have

$$\begin{aligned} |\bar{h}_{2,N}(x) - h_2(x)| &\leq \frac{K_N}{T} \int_{t_{k-1}}^{t_k} |h_2(x) - h_2(t)| dt \leq \frac{K_N}{T} h' \int_{t_{k-1}}^{t_k} |x - t| dt \\ &\leq \frac{h'T}{2K_N} = \frac{Th'}{2C_2} N^{-\gamma} < C_1 N^{-\beta} \end{aligned}$$

for sufficiently large  $N$ . Thus  $J_2^N = 0$  for sufficiently large  $N$ .

By Lemma 6.1,

$$\begin{aligned} J_1^N &\leq \sum_{k=1}^{K_N} \mathbf{P} \left\{ \frac{K_N}{T} |\hat{H}_{2,N}(A_k) - H_2(A_k)| \geq \lambda \delta_N / 2 \right\} \\ &\leq CK_N \exp \left( -\alpha \left( \frac{T\lambda\delta_N}{2K_N} \right)^2 N \right) = CN^\gamma \exp \left( -\alpha \left( \frac{T\lambda}{2} \right)^2 N^{1-2\beta-2\gamma} \right) \rightarrow 0 \end{aligned}$$

as  $N \rightarrow \infty$ , whence  $J_1 \rightarrow 0$ .

The lemma is proved.  $\square$

*Proof of Corollary 5.1.* We check the assumptions of Theorem 5.1. Let

$$t \in \mathcal{T} := [\vartheta - \varepsilon, \vartheta + \varepsilon],$$

where  $\vartheta$  is the true value of the parameter and  $\varepsilon$  is an arbitrary positive number such that  $\vartheta - \varepsilon > 0$ . Then

$$\mathbb{P}\{\tilde{\vartheta}_N \in \mathcal{T}\} \rightarrow 1$$

as  $N \rightarrow \infty$ . Hence one can assume that  $\Theta = \mathcal{T}$ .

Then the random variables  $g^*(\eta_i, h_2(\eta_i), t)$  are bounded for all  $t \in \mathcal{T}$ , whence assumption (i) follows. Assumption (iii) follows from Theorem 3.2, while assumption (ii) holds, since  $G'$  is continuous and  $\tilde{\vartheta}_N \rightarrow \vartheta$  and  $\hat{h}_{2,N} \rightarrow h_2$ .

Put  $\delta_N = N^{-\gamma}$  where  $\gamma = 3\beta/2$  for  $\beta < 1/6$  and  $\gamma = 1/4$  for  $1/6 \leq \beta \leq 1/4$ . Assumption (iv) follows from Lemma 6.3 and Theorem 3.2, since  $g^*(x, h, t)$  is smooth with respect to  $h$  and  $t$  for  $h > 0$  and  $t \in \mathcal{T}$ .

Now we estimate  $\nu_N(2N)$ . If  $t = \text{const}$  and  $h = \text{const}$ , then  $g^*(x, h, t)$  as a function of  $x \in (0, +\infty)$  has at most  $C$  intervals of monotonicity, where  $C$  is a fixed number that does not depend on  $h$  and  $t$ . Thus  $g^*(x, \hat{h}_{2,N}, t)$  has at most  $CK_N$  intervals of monotonicity and

$$\nu_N(2N) \leq (2N)^{CK_N}.$$

Hence

$$\ln \nu_N(2N) \leq CN^\gamma \ln(2N) \leq \delta_N^{-2} = (C_1)^{-2} N^{2\beta},$$

since  $\gamma < 2\beta$ .

The corollary is proved.  $\square$

## 7. CONCLUDING REMARKS

The asymptotic behavior of moment estimators constructed from observations with an admixture is studied in the paper. We constructed an adaptive estimator whose asymptotic variance is equal to the minimal asymptotic variance among moment estimators. The behavior of the estimator requires an additional analysis if the size of a sample is small.

## BIBLIOGRAPHY

1. A. A. Borovkov, *Mathematical Statistics*, "Nauka", Moscow, 1984; English transl., Gordon and Breach Science Publishers, Amsterdam, 1998. MR782295 (86i:62001)
2. V. N. Vapnik and A. Ya. Chervonenkis, *Theory of Pattern Recognition. Statistical Problems of Learning*, "Nauka", Moscow, 1974. (Russian) MR0474638 (57:14274)
3. O. O. Kubaychuk, *Estimation of moments by observations from mixtures with varying concentrations*, Theory Stoch. Process. **8(24)** (2002), no. 3–4, 226–232. MR2027394 (2005g:62065)
4. O. O. Kubaychuk, *Estimation of moments from mixtures with the use of improved weighted empirical distribution function*, Visnyk Kyiv Univ. Matematyka. Mekhanika **9–10** (2003), 48–52. (Ukrainian)
5. O. O. Kubaychuk, *Statistical analysis of characteristics of random variables with the use of observations of a mixture*, Ph.D. Dissertation, Kyiv National University, Kyiv, 2004. (Ukrainian)
6. A. Lodatko and R. Maïboroda, *Estimation of the probability density by observations with an admixture*, Teor. Imovir. Mat. Stat. **73** (2005), 88–96; English transl. in Theory Probab. Math. Statist. **73** (2006), 99–108. MR2213844 (2007b:62043)
7. R. E. Maïboroda, *Statistical Analysis of Mixtures. A Course of Lectures*, Kyiv National University, Kyiv, 2003. (Ukrainian)
8. P. J. Bickel, *On adaptive estimation*, Ann. Statist. **10** (1982), no. 3, 647–671. MR0663424 (84a:62045)

9. X. Shen, *On methods of sieves and penalization*, Ann. Statist. **25** (1997), no. 6, 2555–2591. MR1604416 (2000a:62074)
10. C. Stone, *Adaptive maximum likelihood estimation of a location parameter*, Ann. Statist. **3** (1975), 267–284. MR0362669 (50:15109)

DEPARTMENT OF PROBABILITY THEORY AND MATHEMATICAL STATISTICS, FACULTY FOR MECHANICS AND MATHEMATICS, NATIONAL TARAS SHEVCHENKO UNIVERSITY, ACADEMICIAN GLUSHKOV AVENUE, 6, KYIV 03127, UKRAINE

*E-mail address:* lodatko@yandex.ru

DEPARTMENT OF PROBABILITY THEORY AND MATHEMATICAL STATISTICS, FACULTY FOR MECHANICS AND MATHEMATICS, NATIONAL TARAS SHEVCHENKO UNIVERSITY, ACADEMICIAN GLUSHKOV AVENUE, 6, KYIV 03127, UKRAINE

*E-mail address:* mre@univ.kiev.ua

Received 19/SEP/2005

Translated by S. KVASKO