Теорія Ймовір. та Матем. Статист.
Вип. 78, 2008

Theor. Probability and Math. Statist.
No. 78, 2009, Pages 147–156
S 0094-9000(09)00768-6
Article electronically published on August 4, 2009

ESTIMATION OF MEAN POSITIONS AND CONCENTRATIONS FROM OBSERVATIONS OF A TWO-COMPONENT MIXTURE OF SYMMETRIC DISTRIBUTIONS

UDC 519.21

R. MAĬBORODA

Abstract. A statistician observes a sample from a mixture of two symmetric distributions that differ from one another by a shift parameter. Estimators for mean position parameters and concentrations (mixing probabilities) for both components are constructed by the method of moments. Conditions for the consistence and asymptotic normality of these estimators are obtained. The asymptotic variance (dispersion coefficient) of the estimator of the concentration is found.

1. Introduction

Let N subjects be observed. We assume that each of them may belong to one of the two given populations. The number δ_j of the population containing the subject j is unknown. For the subject j, a numeric characteristic (variable) η_j is observed. Its distribution depends on the population containing the subject, namely $\mathsf{P}\{\eta_j < x\} = H_i(x)$ if $\delta_j = i$, $i = 1, 2$. The (a priori) probability that the subject j belongs to the first population is $p = \mathsf{P}\{\delta_j = 1\}$. Therefore $\mathsf{P}\{\delta_j = 2\} = 1 - p$. Thus the data $(\eta_1, \dots, \eta_N)$ forms a sample of independent identically distributed random variables with the distribution function

$$\mathsf{P}\{\eta_j < x\} = pH_1(x) + (1-p)H_2(x). \tag{1}$$

The data with distributions of this kind are called a sample from a finite (two-component) mixture. The populations from which the subjects are drawn are called the components of the mixture, H_i are the distributions of the components, and p and $1 - p$ are the mixing probabilities or concentrations of the components in the mixture.

There are many papers devoted to the problem of estimation of distributions and concentrations of components of finite mixtures starting with [6] and [7]. Surveys of modern methods of statistics of finite mixtures can be found in [5, 8]. Most of the literature deals with parametric models of distributions, since the nonparametric models of finite mixtures are, generally speaking, not identifiable (this means that the mixing probabilities may not be uniquely estimated for the nonparametric setting).

To the author's knowledge, the first paper treating the nonparametric estimation for a two-component mixture of independent identically distributed observations is [3]. The problem is identifiable in [3] due to the assumption that the observed characteristic is a vector whose dimension is at least 3 and the coordinates of these vectors (in other words, the variables characterizing the observed subjects) are independent for every component of the mixture.

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

Key words and phrases. Method of moments, a finite mixture of probability distributions, consistence, asymptotic normality, asymptotic variance.

We consider another nonparametric model of a two-component mixture that allows us to estimate uniquely its characteristics. We assume that the distributions H_i differ from one another by a shift parameter, namely

$$H_2(x) = H_1(x - a)$$

for some $a \in \mathbb{R}$, and that the distribution H_1 is symmetric about its median. The identifiability of the "parametric part" of this model is proved in [4]. Consistent estimators of the median of H_1 as well as the parameter a and probability p are also considered, and the asymptotic normality of these estimators is proved in [4].

Another method of constructing the estimators of these parameters is proposed in [2]. An estimator of the distribution function H_1 is constructed in [2].

The estimators proposed in the papers [2, 4] are rather complicated: their evaluation is based on a numerical minimization of certain functions that depend on a sample. In the simplest case considered in [4], the function mentioned above is the U statistic. The calculation of this function for given values of the parameters requires CN^2 elementary operations.

This paper is devoted to an application of the method of moments to the problem of estimation described above. The estimators obtained for this problem are written explicitly and their calculation requires CN elementary operations. Moreover, the estimators change "regularly" if the measurement scale is transformed linearly, namely the estimators of the location parameters are equivariant, while the estimator of p is invariant with respect to those transformations.

The paper is organized as follows. The model is introduced in Section 2; the moment estimators of the parameters are also constructed there. We study the invariance and equivariance properties of the estimators with respect to addition and multiplication by a constant in Section 3; the asymptotic behavior of estimators is also studied in Section 3. The quality of the estimators for a finite size of a sample is assessed in Section 4 via simulation. As an example, we apply our estimators to real life data in the same section. Some remarks concerning further investigation of our estimators are placed in Section 5.

Note that the methods of construction and investigation of the estimators proposed in this paper are standard. The main technical problem is to solve the system of five equations of the method of moments and to represent their solutions in a compact form. To perform complicated algebraic manipulations and to exhibit our algorithms, we used the software Mathematica.

2. Setting of the problem and construction of estimators

Consider a sample of independent identically distributed random variables $(\eta_1, \dots, \eta_N)$ described by the following model:

$$\eta_j = a_{\delta_j} + \varepsilon_j, \tag{2}$$

where a_i, $i = 1, 2$, are nonrandom real numbers, δ_j are random variables assuming the value 1 with probability p and the value 2 with probability $1 - p$, and ε_j are random variables with the distribution function $H(x) = \mathsf{P}\{\varepsilon \le x\}$. The random variables δ_j, ε_j, $j = 1, \dots, N$, are assumed to be jointly independent.

It is clear that the distribution of the sample $(\eta_1, \dots, \eta_N)$ is described by model (1) for

$$H_i(x) = H(x - a_i).$$

The number a_i can be treated as a "mean value" of the variable η for subjects belonging to the population i, while the ε_j are random deviations from the mean values. In what follows we assume that the distribution of random deviations is symmetric about zero;

that is, $H(x) = 1 - \lim_{t\downarrow 0} H(-x-t)$. Therefore a_i is equal to both the median and the expectation of H_i (provided the latter exists).

We construct estimators for a_i and p assuming that the distribution function H is unknown. In addition to the assumption that the distribution function H is symmetric, we require that certain moments of random variables ε_j be finite.

We follow the method of moments when constructing the estimators. Let $\mathsf{E}\,|\varepsilon_j|^5 < \infty$. Put $e_k = \mathsf{E}(\varepsilon_j)^k$ and let $y_k = \mathsf{E}(\eta_j)^k$ be the "theoretical moments" of order k for observations. Since the distribution of ε_j is symmetric, $e_1 = e_3 = e_5 = 0$. We obtain from (2) that

$$y_1 = pa_1 + (1-p)a_2, \tag{3a}$$

$$y_2 = pa_1^2 + (1-p)a_2^2 + e_2, \tag{3b}$$

$$y_3 = pa_1^3 + (1-p)a_2^3 + 3y_1e_2, \tag{3c}$$

$$y_4 = pa_1^4 + (1-p)a_2^4 + 6\left(pa_1^2 + (1-p)a_2^2\right)e_2 + e_4, \tag{3d}$$

$$y_5 = pa_1^5 + (1-p)a_2^5 + 10\left(pa_1^3 + (1-p)a_2^3\right)e_2 + 5y_1e_4. \tag{3e}$$

To obtain the moment estimators for a_i and p one needs to substitute the empirical moments

$$\hat{y}_k = \hat{y}_{k;N} = \frac{1}{N}\sum_{j=1}^{N}(\eta_j)^k$$

for the corresponding theoretical moments y_k in (3) and to solve this system of equations with respect to a_1,a_2, e_2, e_4, and p. There are some "redundant" solutions of this system that do not correspond to consistent estimators of unknown parameters. It turns out that, among the solutions of the system, one can uniquely determine those that correspond to the consistent estimators. Now we describe these solutions.

First we consider the first three equations of system (3) assuming that the parameter p is known. It is easy to exclude e_2 and a_2 from those equations and obtain an equation for a_1, namely

$$pa_1^3 - \frac{(\hat{y}_1 - pa_1)^3}{(1-p)^2} - \frac{3\hat{y}_1\left(\hat{y}_1^2 - \hat{y}_2 + p(a_1^2 - 2\hat{y}_1a_1 + \hat{y}_2)\right)}{1-p} = \hat{y}_3.$$

This is a cubic equation with respect to a_1, which in general has three roots. If $p \in (0, 1/2)$, then the root

$$\hat{a}_{1,n}(p) = \hat{a}_{1,n}(p;\hat{y}_1,\hat{y}_2,\hat{y}_3) = \hat{y}_1 + \frac{\sqrt[3]{(1-3p+2p^2)^2(3\hat{y}_1\hat{y}_2 - 2\hat{y}_1^3 - \hat{y}_3)}}{\sqrt[3]{p}(1-2p)} \tag{4}$$

is a solution we need.[1] It is straightforward to check that the equality

$$\hat{a}_{1,n}(p; y_1, y_2, y_3) = a_1$$

holds if one substitutes the empirical moments in (4) for the corresponding theoretical moments defined in (3).

Therefore one can use $\hat{a}_{1,n}(p)$ to estimate a_1 if $p \in (0, 1/2)$ is known. The estimator for a_2 can be derived from (3a):

$$\hat{a}_{2,n}(p) = \frac{1}{1-p}(\hat{y}_1 - p\hat{a}_{1,n}(p)). \tag{5}$$

[1]Here and in what follows $\sqrt[3]{x}$ for $x \in \mathbb{R}$ means (if the opposite is not specified explicitly) the real root, for example $\sqrt[3]{-8} = -2$.

Now let $p \in (0, 1/2)$ be unknown. Excluding the unknown moments of ε_j from (3b), (3d), and (3e) we get

$$0 = -\hat{y}_5 + pa_1^5 + (1-p)a_2^5 + 10\left(pa_1^3 + (1-p)a_2^3\right)\left(\hat{y}_2 - pa_1^2 + (1-p)a_2^2\right) + 5\hat{y}_1\left(\hat{y}_4 - pa_1^4 - (1-p)a_2^4 - 6\left(pa_1^2 + (1-p)a_2^2\right)\left(\hat{y}_2 - pa_1^2 + (1-p)a_2^2\right)\right).$$

Substituting $\hat{a}_{1,n}(p)$ and $\hat{a}_{2,n}(p)$ for a_1 and a_2 in the latter equation, respectively, we obtain an equation for the estimator of p. One can transform this equation to the form

$$U_N^5\left(1 - 12p + 12p^2\right)^3 = V_N^3 p^2\left(1 - 3p + 2p^2\right)^2, \tag{6}$$

where

$$U_N = 2\hat{y}_1^3 - 3\hat{y}_1\hat{y}_2 + \hat{y}_3, \tag{7}$$

$$V_N = 24\hat{y}_1^5 - 60\hat{y}_1^3\hat{y}_2 + 30\hat{y}_1\hat{y}_2^2 + 20\hat{y}_1^2\hat{y}_3 - 10\hat{y}_2\hat{y}_3 - 5\hat{y}_1\hat{y}_4 + \hat{y}_5. \tag{8}$$

Choosing a necessary root among the six solutions of this equation, we obtain the estimator for p:

$$\hat{p}_N = \hat{p}(C_N) = \begin{cases} P_1(C_N) & \text{if } C_N > 432, \\ \frac{1}{2} - \frac{\sqrt{2}}{3} & \text{if } C_N = 432, \\ P_2(C_N) & \text{if } C_N < 432, \end{cases} \tag{9}$$

where

$$C_N = \frac{V_N^3}{U_N^5},$$

$$A(C) = \sqrt[3]{-(-432 + C)^2 C + 12\sqrt{3}\sqrt{(432 - C)^3 C^2}},$$

$$B(C) = \frac{1}{3}\left(\pi - \arctan\left(\frac{12\sqrt{3}}{\sqrt{-432 + C}}\right)\right),$$

$$P_1(C) = \frac{1}{2} - \frac{\sqrt{6}}{12}\sqrt{4 + \frac{2C\left(-\cos(B(C)) + \sqrt{3}\sin(B(C))\right)}{\sqrt{(-432 + C)\,C}}},$$

$$P_2(C) = \frac{1}{2} - \frac{\sqrt{-(C + A(C))^2 + 432\,(C + 2A(C))}}{2\sqrt{3}\sqrt{(432 - C)\,A(C)}}.$$

It is straightforward to check that if $0 < p < 1/2$ and the theoretical moments are used in the definition of $\hat{p}$ instead of the corresponding empirical moments, then expression (9) is reduced to p.

Remark 1. The statistic C_N as well as the corresponding theoretical characteristic $C(p)$ defined below by (12) may assume any real value. The function $\hat{p}$ defined by (9) is decreasing and maps the real line to the interval $(0, 1/2)$. Thus the estimator $\hat{p}_N$ belongs to the interval $(0, 1/2)$.

Remark 2. It is surprising that the trigonometric functions appear in the formula for P_1. In fact, this follows from the complex form of the solution of equation (6), which is written as follows:

$$P_1(C) = \frac{1}{2} - \frac{\sqrt{6}}{12}\sqrt{4 + \frac{i\left(i + \sqrt{3}\right)C}{A(C)} - \frac{i\left(-i + \sqrt{3}\right)A(C)}{-432 + C}}.$$

The quadratic and cubic roots are treated as the main part of the analytic function $\sqrt[k]{x} = \exp(\ln(x)/k)$ when evaluating $A(C)$ and P_1, that is, for example, $\sqrt[3]{-8} = 1 + \sqrt{3}i$. (This convention is accepted in the software Mathematica.)

When the function $A(C)$ is used in (9) for the evaluation of $\hat{p}$, that is if $C < 432$, all the expressions under the root sign as well as the corresponding solutions are positive real numbers. Similarly, A and P_2 for $C < 432$ and B and P_1 for $C > 432$ are real-valued functions of a real argument. The complexity of calculation of these functions with a computer is comparable with that of the calculation of a cubic root.

Finally, we take

$$\hat{a}_{1;N} = \hat{a}_1(\hat{p}_N), \tag{10}$$

$$\hat{a}_{2;N} = \hat{a}_2(\hat{p}_N) \tag{11}$$

as the estimators of a_1 and a_2 if p is unknown.

3. Properties of estimators

Prior to studying the asymptotic behavior of the above estimators, we prove that they are invariant and equivariant. Recall that the statistic $S = S(Y)$ evaluated from a sample $Y = (\eta_1, \dots, \eta_N)$ is called invariant with respect to addition if

$$S(Y + z) = S(Y)$$

for all $z \in \mathbb{R}$ (here $Y + z = (\eta_1 + z, \dots, \eta_N + z)$). A statistic S is called equivariant with respect to addition if $S(Y + z) = S(Y) + z$ for all $z \in \mathbb{R}$. Similarly we define these properties with respect to multiplication.

Since $\hat{a}_{1,N}$ and $\hat{a}_{2,N}$ estimate the location parameters, it is natural to require that they are equivariant with respect to the shifts and changes of scale, that is, with respect to addition and multiplication. An estimator of the probability p should not depend on shifts and changes of scale; that is, this estimator should be invariant.

Theorem 3.1. *The estimators $\hat{a}_{1,N}$ and $\hat{a}_{2,N}$ defined by* (10) *and* (11)*, respectively, are equivariant with respect to addition and multiplication. The estimator $\hat{p}_N$ defined by* (9) *is invariant with respect to addition and multiplication.*

Proof. Let y'_k be the sample moment of order k for the sample $Y + z$. Then

$$y'_k = \sum_{i=0}^{k} \binom{k}{i} y_k z^{k-i},$$

where, as above, the y_k are sample moments of the sample Y. Substituting y'_k in the expression for C_N and canceling similar terms, we prove that the statistic C_N is invariant with respect to addition. Since $\hat{p}_N$ is a function of C_N, $\hat{p}_N$ also is an invariant estimator. Now we substitute y'_k into (4) and (5) and prove that $\hat{a}_{1,N}$ and $\hat{a}_{2,N}$ are equivariant.

The proof that the estimators are invariant and equivariant with respect to multiplication is analogous. □

Now we turn to the asymptotic behavior of the estimators. Recall that an estimator is called strongly consistent if it converges almost surely to the true value of the parameter as the sample size tends to infinity.

Theorem 3.2. *Let*

(i) $\mathsf{E}\,|\varepsilon_j|^5 < \infty$;
(ii) $a_1 \neq a_2$;
(iii) $0 < p < \frac{1}{2}$.

Then $\hat{a}_{i:N}$ are strongly consistent estimators of a_i, $i = 1, 2$, and $\hat{p}_N$ is a strongly consistent estimator of p.

Remark 3. The case of $p \in (1/2, 1)$ can be reduced to that considered in Theorem 3.2 by exchanging the components. Then p is exchanged with $1-p$, while a_1 is exchanged with a_2. In fact, the restriction $0 < p < \frac{1}{2}$ means that the members of the second component occur more often than those of the first component.

Remark 4. If $a_1 = a_2$, then the distributions of both components are the same and therefore p cannot be estimated at all. The problem is not identifiable in this case.

Remark 5. If $p = 1/2$, then one can estimate $y_1 = \frac{1}{2}(a_1 + a_2)$ by $\hat{y}_1$, but the parameters a_1 and a_2 cannot be estimated separately. Indeed, let the model for the data be described by equation (2) for $a_1 \neq a_2$ and $p = 1/2$. Consider the new data $\tilde{\eta}_j$ for which the location parameters coincide; that is,

$$\tilde{a}_1 = \tilde{a}_2 = y_1 = \frac{1}{2}(a_1 + a_2),$$

and the random deviations are such that $\tilde{\varepsilon}_j = a_{\delta_j} - y_1 + \varepsilon_j$. (Since $p = 1/2$, the distribution of $\tilde{\varepsilon}_j$ is symmetric.) It is clear that $\eta_j = \tilde{\eta}_j$. Thus one cannot decide from the observations which of the two parameters a_i or $\tilde{a}_i$ is the "true" one.

Hence the assumptions (ii) and (iii) of Theorem 3.2 are natural conditions for the identifiability of the problem.

Remark 6. If $p = 1/2$, then $U_N \to 0$ and $C_N \to \pm\infty$ as $N \to \infty$. Thus the estimator $\hat{p}_N$ for large N can be close to both 0 and 1/2. This is also evidence of the nonidentifiability: in each of the cases $p = 1/2$ and $p = 0$, the distribution of observations is symmetric (with respect to its own median) and it is not possible to separate these two cases without additional information.

If one estimates the parameters in the framework of the above model and if the distribution of the observations is symmetric, then it is reasonable to sieve the data by using the corresponding statistical tests.

Proof of Theorem 3.2. According to assumption (i), $\hat{y}_{k;N} \to y_k$ almost surely by the strong law of large numbers. Thus U_N and V_N approach U and V, respectively. The latter variables can be obtained by the substitution of theoretical moments instead of empirical moments in (7) and (8). Expressing these moments in terms of the parameters with the help of system (3), we get

$$U = p\left(1 - 3p + 2p^2\right)(b_1 - b_2)^3,$$
$$V = p\left(1 - 15p + 50p^2 - 60p^3 + 24p^4\right)(b_1 - b_2)^5.$$

It is easy to see that $U \neq 0$ under the assumptions of the theorem, whence

$$C_N \to V^3/U^5 = C(p) = \frac{\left(1 - 12p + 12p^2\right)^3}{p^2(1 - 3p + 2p^2)^2} \quad \text{almost surely.} \tag{12}$$

The function $\hat{p}(C)$ defined in (9) is the inverse to $C(p)$. This function is continuous for all $C \in \mathbb{R}$. Thus $\hat{p}(C_N) \to p$ almost surely.

The consistence of $\hat{a}_{i:N}$ can be proved similarly. □

Prior to stating the result on the asymptotic normality of the estimators, we study the asymptotic variance of $\hat{p}_N$. Put

$$b = (a_1 - a_2)/\sqrt{e_2}, \qquad \kappa_n = e_n/(e_2)^{n/2},$$

and

$$s^2 = \frac{1}{4b^{10}}(\alpha_0 + \alpha_1 p(1-p)), \tag{13}$$

where

$$\alpha_0 = 8100 - 8100\kappa_4 + 1125{\kappa_4}^2 + 1440\kappa_6 - 90\kappa_4\kappa_6 - 180\kappa_8 + 9\kappa_{10} + b^4\left(225 - 150\kappa_4 + 25\kappa_6\right) + b^2\left(2700 - 2250\kappa_4 + 150{\kappa_4}^2 + 390\kappa_6 - 30\kappa_8\right),$$

$$\alpha_1 = 4b^{10} + b^6\left(-225 + 225\kappa_4\right) + b^4\left(5400 - 4950\kappa_4 + 750\kappa_6\right) + b^2\left(-40500 + 37800\kappa_4 - 2025{\kappa_4}^2 - 7380\kappa_6 + 585\kappa_8\right).$$

Theorem 3.3. *Let* $\mathsf{E}\,\varepsilon_j^{10} < \infty$ *and let assumptions* (ii) *and* (iii) *of Theorem* 3.2 *hold. Then the normalized estimator* $\sqrt{N}(\hat{p}_N - p)$ *weakly converges as* $N \to \infty$ *to the normal law with zero mean and variance* s^2*.*

Proof. The proof follows from Theorem 2A of Section 4, Chapter 2 of [1], where the asymptotic normality of the moment estimators is studied. It is reasonable to treat $\hat{p}_N$ in our proof of the asymptotic normality as a solution of the equation

$$\frac{V_N}{(U_N)^{5/3}} = \frac{\left(1 - 12p + 12p^2\right)}{p^{2/3}(1 - 3p + 2p^2)^{2/3}}$$

(with respect to $p \in (0, 1/2)$). The right hand side of this equation is a strictly decreasing function of $p \in (0, 1/2)$, and its derivative is separated from 0.

In the evaluation of the asymptotic variance, Theorem 2A also was used together with the property that $\hat{p}_N$ is invariant with respect to addition and multiplication (see Theorem 3.1). This allows one to restrict consideration to the case of $a_1 = 0$, $a_2 = b$, and $e_2 = 1$, since all other cases are reduced to the latter case by a linear transformation of the measurement scale. □

Remark 7. The same Theorem 2A of [1] is helpful in the proof of the asymptotic normality of the estimators $\hat{a}_{i;N}$. It is not yet possible to write down their asymptotic variance in a closed form yet. Nevertheless it can be calculated in the software system Mathematica.

Remark 8. The parameter b plays the role of the signal/noise ratio in the problem considered. Indeed, the greater the distance $|a_1 - a_2|$ with respect to the mean square dispersion of random deviations ε_j, the more precise should be the estimators of the parameters of the model. However, the asymptotic variance of $\hat{p}_N$ defined in (13) tends to $p(1-p)$ as $b \to \infty$, so it does not tend to 0. This can be explained by an additional (besides ε_j) source of the randomness in the model, namely by the random distribution of subjects between the components of the mixture. If b is sufficiently large, then one can determine the number δ_j of the population containing the subject from an observation η_j. An efficient estimator of p constructed from observations δ_j is the relative frequency of the event $\delta_j = 1$ in the sample. This estimator is just the one that has the asymptotic variance $p(1-p)$.

Therefore, as $b \to \infty$, the asymptotic variance of the estimator $\hat{p}_N$ approaches the best possible variance.

Remark 9. If the distribution of ε_j is known, the expression for s^2 can be simplified substantially. For example, in the case of the normal distribution,

$$s^2 = \frac{270}{b^{10}} + \frac{75}{2b^6} + \left(1 + \frac{1350}{b^8} + \frac{450}{b^6} + \frac{225}{2b^4}\right)(1-p)\,p.$$

4. Results of simulation and an example of real life data analysis

We consider simulated samples in order to evaluate how well the asymptotic results explain the behavior of the estimators for a finite sample size. We simulated samples of sizes $N = 50$, 100, 250, 500, 750, 1000, and 5000 with distribution (2) and for random

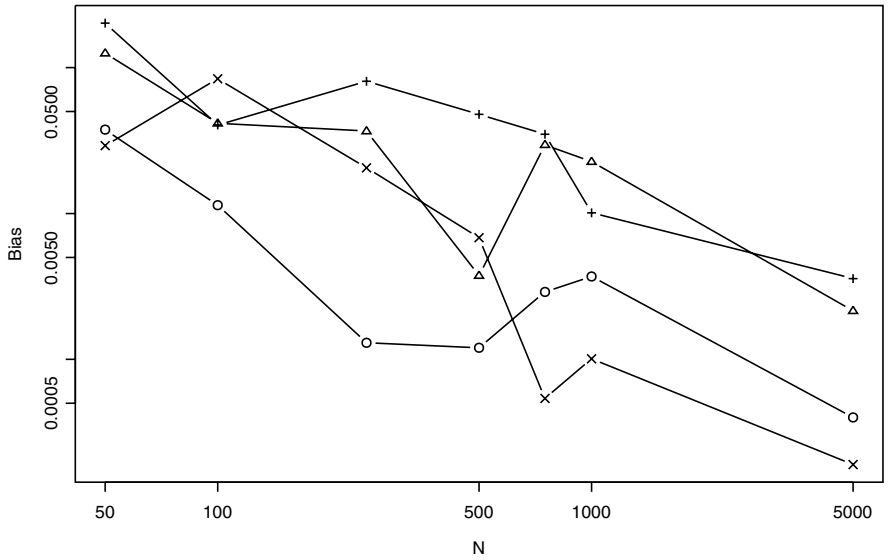

FIGURE 1. The bias of the estimators $\hat{a}_{1:N}$

deviations having the normal distribution (N), uniform distribution (U) in the interval $[-1, 1]$, and the Laplace distribution (L) with zero mean and scale parameter $\lambda = 1$. We take $p = 0.25$ and choose the location parameters of the components to obtain a given signal/noise ratio b.

For each of the sets of parameters we simulated 1000 samples and used them to evaluate the estimators $\hat{a}_{1;N}$ and $\hat{p}_N$. Using these samples, the sample means and variances are evaluated for the estimators.

The comparison of the sample variances $\hat{\sigma}^2$ of normalized estimators $\sqrt{N}(\hat{p}_N - p)$ with the asymptotic variance s^2 obtained from (13) is presented in the following table whose entries show the ratios $\hat{\sigma}^2/s^2$:

Distribution	b	s^2	N						
			50	100	250	500	750	1000	5000
N	2	4.6626	0.221	0.361	0.606	0.767	0.846	0.805	0.931
N	4	0.14024	1.24	1.04	1.04	0.96	0.966	1.02	0.937
U	2	0.21652	0.86	0.826	0.874	0.889	0.866	0.89	0.883
L	4	1.62388	0.254	0.325	0.509	0.663	0.708	0.819	0.911

As can be seen from the table, the asymptotic formula provides a good approximation for $N > 1000$ and can be used as a first approximation for $N > 500$. However, this result corresponds to the cases where the variance of $\hat{p}_N$ calculated by the asymptotic formula is not large. For example, if $b = 1$ and the distribution is normal, then $s^2 = 898.343$, while the simulated data shows $\hat{\sigma}^2 = 148.071$ if the size of the sample is $N = 5000$.

This phenomenon has a simple explanation. The behavior of the asymptotic variance is determined by the behavior of the function $C(p)$ in a neighborhood of the true value p. In doing so, one ignores the fact that the function $C^{-1}(\cdot) = \hat{p}(\cdot)$ defined by (9) assumes values in the interval $(0, 1/2)$. This means that $0 < \hat{p}_N < 1/2$. If $\mathsf{E}\,\hat{p}_N \approx p = 0.25$, then the variance of $\hat{p}_N$ does not exceed 0.0625 (recall that the simulated data shows $\mathsf{E}\,\hat{p}_N = 0.247662$). Thus the variance of $\sqrt{N}(\hat{p}_N - p)$ cannot exceed 312.5, which means that the value $s^2 = 898.343$ is too large. Note that $\hat{\sigma}^2 = 148.071$ does not exceed 312.5

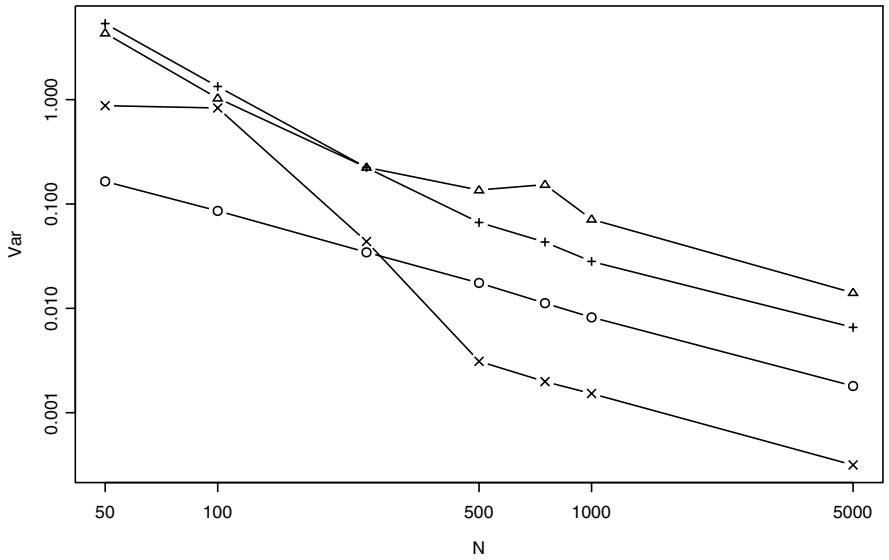

FIGURE 2. The variance of the estimators $\hat{a}_{1:N}$

essentially, and therefore the estimator of p for $b = 1$ does not provide an approximate idea about the true value of the parameter even for $N = 5000$.

The simulated data for the estimator $\hat{a}_{1,N}$ are presented in Figures 1 and 2. The absolute value of the bias and the variance of the estimators as functions of the sample size are depicted in these figures in the logarithmic scale for the cases of the normal distribution with $b = 4$ ($\circ$) or with $b = 2$ ($+$), uniform distribution with $b = 2$ ($\times$), and Laplace distribution with $b = 4$ ($\triangle$).

One of the classical examples of the data used to test algorithms for analyzing the mixtures is the data set consisting of measurements of eruption lengths of the Old Faithful geyser in the Yellowstone National Park, USA. The collections of examples of packages S+ and R include this data set (named "geyser"). The histogram constructed for this data clearly shows the mixture of two components. The results of the estimation of parameters for the model (2) can be found in [4]. The method used in [4] gives $\hat{a}_1 = 54$, $\hat{a}_2 = 80$, and $\hat{p} = 0.352$. The maximum likelihood estimators in [4] are $\hat{a}_1 = 54.61$, $\hat{a}_2 = 80.09$, and $\hat{p} = 0.361$ under the assumption that the components have normal distributions. The method-of-moments estimators constructed in this paper are $\hat{a}_1 = 56.1952$, $\hat{a}_2 = 83.3471$, and $\hat{p} = 0.406333$, respectively.

5. Concluding remarks

We constructed estimators for concentrations and location parameters of the distributions of components of a mixture and proved that they are consistent and asymptotically normal under the natural conditions of the identifiability and some (not restrictive) moment assumptions. The asymptotic variance of the estimator of the concentration is close to the best one for large values of the signal/noise ratio. However, if this ratio is low, then the estimators studied in this paper cannot be recommended for use, since their asymptotic variance turns out to be too large. The analysis of simulated data confirms this recommendation, although the real variance of the estimators of concentrations in samples of small sizes is often smaller than the one found by the asymptotic formula.

The question on whether or not better estimators exist for small values of the signal/noise ratio is still open.

Another interesting question concerns the nonparametric estimator of the distribution function of random deviations.

Acknowledgement

The author is grateful to A. Chubatyuk for a discussion of the results and for help in doing algebraic transformations with a computer.

Bibliography

1. A. A. Borovkov, *Mathematical Statistics*, Nauka, Moscow, 1984; English. transl., Gordon and Breach, Amsterdam, 1998. MR782295 (86i:62001); MR1712750 (2000f:62003)
2. L. Bordes, S. Mottelet, and P. Vandekerkhove, *Semiparametric estimation of a two-component mixture model*, Ann. Statist. **34** (2006), no. 3, 1204–1232. MR2278356 (2008e:62064)
3. P. Hall and X.-H. Zhou, *Nonparametric estimation of component distributions in a multivariate mixture*, Ann. Statist. **31** (2003), no. 1, 201–224. MR1962504 (2003m:62105)
4. D. R. Hunter, S. Wang, and T. R. Hettmansperger, *Inference for Mixtures of Symmetric Distributions*, Technical Report 04-01, Penn State University, Philadelphia, 2004.
5. G. J. McLachlan and D. Peel, *Finite Mixture Models*, Wiley, New York, 2000. MR1789474 (2002b:62025)
6. S. Newcomb, *A generalized theory of the combination of observations so as to obtain the best result*, Amer. J. Math. **8** (1886), 343–366. MR1505430
7. K. Pearson, *Contributions to the mathematical theory of evolution*, Phil. Trans. R. Soc. Lond. A **185** (1894), 71–110.
8. D. M. Titterington, A. F. Smith, and O. E. Makov, *Statistical Analysis of Finite Mixture Distributions*, Wiley, Chichester, 1985. MR0838090 (87j:62033)

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 22/MAR/2007

Translated by N. SEMENOV