

Testing hypotheses on moments by observations from a mixture with varying concentrations

Alexey Doronin, Rostyslav Maiboroda

Kyiv National Taras Shevchenko University, Kyiv, Ukraine

al_doronin@ukr.net (A. Doronin), mre@univ.kiev.ua (R. Maiboroda)

Received: 22 December 2014, Revised: 21 January 2015, Accepted: 22 January 2015,
Published online: 4 February 2015

Abstract A mixture with varying concentrations is a modification of a finite mixture model in which the mixing probabilities (concentrations of mixture components) may be different for different observations. In the paper, we assume that the concentrations are known and the distributions of components are completely unknown. Nonparametric technique is proposed for testing hypotheses on functional moments of components.

Keywords Finite mixture model, functional moments, hypothesis testing, mixture with varying concentrations

2010 MSC 62G10, 62G20

1 Introduction

Finite mixture models (FMMs) arise naturally in statistical analysis of biological and sociological data [11, 13]. The model of mixture with varying concentrations (MVC) is a modification of the FMM where the mixing probabilities may be different for different observations. Namely, we consider a sample of subjects O_1, \dots, O_N where each subject belongs to one of subpopulations (mixture components) $\mathcal{P}_1, \dots, \mathcal{P}_M$. The true subpopulation to which the subject O_j belongs is unknown, but we know the probabilities $p_{j;N}^m = \mathbf{P}[O_j \in \mathcal{P}_m]$ (mixing probabilities, concentrations of \mathcal{P}_m in the mixture at the j th observation, $j = 1, \dots, N, m = 1, \dots, M$). For each subject O , a variable $\xi(O)$ is observed, which is considered as a random element in a measurable space \mathcal{X} equipped by a σ -algebra \mathfrak{F} . Let

$$F_m(A) = \mathbf{P}[\xi(O) \in A \mid O \in \mathcal{P}_m], \quad A \in \mathfrak{F},$$

© 2014 The Author(s). Published by VTeX. Open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.

be the distribution of $\xi(O)$ for subjects O that belong to the m th component. Then the unconditional distribution of $\xi_{j;N} = \xi(O_j)$ is

$$\mathbb{P}[\xi_{j;N} \in A] = \sum_{m=1}^M p_{j;N}^m F_m(A), \quad A \in \mathfrak{F}. \quad (1)$$

The observations $\xi_{j;N}$ are assumed to be independent for $j = 1, \dots, N$.

We consider the nonparametric MVC model where the concentrations $p_{j;N}^m$ are known but the component distributions F_m are completely unknown. Such models were applied to analyze gene expression level data [8] and data on sensitive questions in sociology [12]. An example of sociological data analysis based on MVC is presented in [9]. In this paper, we consider adherents of different political parties in Ukraine as subpopulations \mathcal{P}_i . Their concentrations are deduced from 2006 parliament election results in different regions of Ukraine. Individual voters are considered as subjects; their observed characteristics are taken from the Four-Wave Values Survey held in Ukraine in 2006. (Note that the political choices of the surveyed individuals were unknown. So, each subject must be considered as selected from mixture of different \mathcal{P}_i .) For example, one of the observed characteristics is the satisfaction of personal income (in points from 1 to 10).

A natural question in the analysis of such data is homogeneity testing for different components. For example, if $\mathfrak{X} = \mathbb{R}$, then we may ask if the means or variances (or both) of the distributions F_i and F_k are the same for some fixed i and k or if the variances of all the components are the same.

In [8], a test is proposed for the hypothesis of two-means homogeneity. In this paper, we generalize the approach from [8] to a much richer class of hypotheses, including different statements on means, variances, and other generalized functional moments of component distributions.

Hypotheses of equality of MVC component distributions, that is, $F_i \equiv F_k$, were considered in [6] (a Kolmogorov–Smirnov-type test is proposed) and [1] (tests based on wavelet density estimation). The technique of our paper also allows testing such hypotheses using the “grouped χ^2 ”-approach.

Parametric tests for different hypotheses on mixture components were also considered in [4, 5, 13].

The rest of the paper is organized as follows. We describe the considered hypotheses formally and discuss the test construction in Section 2. Section 3 contains auxiliary information on the functional moments estimation in MVC models. In Section 4, the test is described formally. Section 5 contains results of the test performance analysis by a simulation study and an example of real-life data analysis. Technical proofs are given in Appendix A.

2 Problem setting

In the rest of the paper, we use the following notation.

The zero vector from \mathbb{R}^k is denoted by \mathbb{O}_k . The unit $k \times k$ -matrix is denoted by $\mathbb{I}_{k \times k}$, and the $k \times m$ -zero matrix by $\mathbb{O}_{k \times m}$. Convergences in probability and in distribution are denoted \xrightarrow{P} and \xrightarrow{d} , respectively.

We consider the set of concentrations $p = (p_{j;N}^m, j = 1, \dots, N; m = 1, \dots, M; N = 1, \dots)$ as an infinite array, $p_{\cdot;N}^m = (p_{j;N}^m, j = 1, \dots, N; m = 1, \dots, M)$ as an $(N \times m)$ -matrix, and $p_{\cdot;N}^m = (p_{j;N}^m, j = 1, \dots, N) \in \mathbb{R}^d$ and $p_{j;N}^m = (p_{j;N}^m, m = 1, \dots, M)$ as column vectors. The same notation is used for arrays of similar structure, such as the array a introduced further.

Angle brackets with subscript N denote averaging of an array over all the observations, for example,

$$\langle a_{\cdot;N}^m \rangle_N = \frac{1}{N} \sum_{j=1}^N a_{j;N}^m.$$

Multiplication, summation, and other similar operations inside the angle brackets are applied to the arrays componentwise, so that

$$\langle a_{\cdot;N}^m p_{\cdot;N}^k \rangle_N = \frac{1}{N} \sum_{j=1}^N a_{j;N}^m p_{j;N}^k, \quad \langle (a_{\cdot;N}^m)^2 \rangle_N = \frac{1}{N} \sum_{j=1}^N (a_{j;N}^m)^2,$$

and so on.

Angle brackets without subscript mean the limit of the corresponding averages as $N \rightarrow \infty$ (assuming that this limit exists):

$$\langle p^m a^k \rangle = \lim_{N \rightarrow \infty} \langle p_{\cdot;N}^m a_{\cdot;N}^k \rangle_N.$$

We introduce formally random elements $\eta_m \in \mathfrak{X}$ with distributions F_m , $m = 1, \dots, M$.

Consider a set of $K \leq M$ measurable functions $g_k : \mathfrak{X} \rightarrow \mathbb{R}^{d_k}$, $k = 1, \dots, K$. Let \bar{g}_k^m be the (vector-valued) functional moment of the m th component with moment function g_k , that is,

$$\bar{g}_k^m := \mathbf{E}[g_k(\eta_m)] \in \mathbb{R}^{d_k}. \quad (2)$$

Fix a measurable function $T : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \dots \times \mathbb{R}^{d_K} \rightarrow \mathbb{R}^L$. For data described by the MVC model (1) we consider testing a null-hypothesis of the form

$$H_0 : T(\bar{g}_1^1, \dots, \bar{g}_K^K) = \mathbb{O}_L \quad (3)$$

against the general alternative $T(\bar{g}_1^1, \dots, \bar{g}_K^K) \neq \mathbb{O}_L$.

Example 1. Consider a three-component mixture ($M = 3$) with $\mathfrak{X} = \mathbb{R}$. We would like to test the hypothesis $H_0^\sigma : \text{Var } \eta_1 = \text{Var } \eta_2$ (i.e., the variances of the first and second components are the same). This hypothesis can be reformulated in the form (3) by letting $g_1(x) = g_2(x) = (x, x^2)^T$ and $T((y_{11}, y_{12})^T, (y_{21}, y_{22})^T) = (y_{12} - (y_{11})^2, y_{22} - (y_{21})^2)^T$.

Example 2. Let $\mathfrak{X} = \mathbb{R}$. Consider the hypothesis of mean homogeneity $H_0^\mu : \mathbf{E} \eta_1 = \dots = \mathbf{E} \eta_M$. Then the choice of $g_m(x) = x$, $T(y_1, \dots, y_M) = (y_1 - y_2, y_2 - y_3, \dots, y_{M-1} - y_M)^T$ reduces H_0^μ to the form (3).

Example 3. Let \mathfrak{X} be a finite discrete space: $\mathfrak{X} = \{x_1, \dots, x_r\}$. Consider the distribution homogeneity hypothesis $H_0^\equiv : F_1 \equiv F_2$. To present it in the form (3), we can use $g_i(x) = (\mathbb{1}\{x = x_i\}, k = 1, \dots, r-1)^T$ and $T(y_1, y_2) = y_1 - y_2$ ($y_i \in \mathbb{R}^{r-1}$

for $i = 1, 2$). In the case of continuous distributions, H_0^{\equiv} can be discretized by data grouping.

To test H_0 defined by (3), we adopt the following approach. Let there be some consistent estimators $\hat{g}_{k;N}^m$ for \bar{g}_k^m . Assume that T is continuous. Consider the statistic $\hat{T}_N = T(\hat{g}_{1;N}^1, \dots, \hat{g}_{K;N}^K)$. Then, under H_0 , $\hat{T}_N \approx \mathbb{O}_L$, and a far departure of \hat{T}_N from zero will evidence in favor of the alternative.

To measure this departure, we use a Mahalanobis-type distance. If $\sqrt{N}\hat{T}_N$ is asymptotically normal with a nonsingular asymptotic covariance matrix D , then, under H_0 , $N\hat{T}_N^T D^{-1} \hat{T}_N$ is asymptotically χ^2 -distributed. In fact, D depends on unknown component distributions F_i , so we replace it by its consistent estimator \hat{D}_N . The resulting statistic $\hat{s}_N = N\hat{T}_N^T \hat{D}_N^{-1} \hat{T}_N$ is a test statistic. The test rejects H_0 if $\hat{s}_N > Q^{\chi^2_L}(1 - \alpha)$, where α is the significance level, and $Q^G(\alpha)$ denotes the quantile of level α for distribution G .

Possible candidates for the role of estimators $\hat{g}_{k;N}^m$ and \hat{D}_N are considered in the next section.

3 Estimation of functional moments

Let us start with the nonparametric estimation of F_m by the weighted empirical distribution of the form

$$\hat{F}_{m;N}(A) = \frac{1}{N} \sum_{j=1}^N a_{j;N}^m \mathbb{1}\{\xi_{j;N} \in A\},$$

where $a_{j;N}^m$ are some nonrandom weights to be selected “in the best way.” Denote $e_m = (\mathbb{1}\{k = m\}, k = 1, \dots, M)^T$ and

$$\Gamma_N = \frac{1}{N} (p_{\cdot;N}^m)^T p_{\cdot;N} = \langle (p_{\cdot;N}^m p_{\cdot;N}^i)_N \rangle_{m,i=1}^M.$$

Assume that Γ_N is nonsingular. It is shown in [8] that, in this case, the weight array

$$a_{\cdot;N}^m = p_{\cdot;N} \Gamma_N^{-1} e_m$$

yields the unbiased estimator with minimal assured quadratic risk.

The *simple* estimator $\hat{g}_{i;N}^m$ for \bar{g}_i^m is defined as

$$\hat{g}_{i;N}^m = \int_{\mathfrak{X}} g_i(x) \hat{F}_{m;N}(dx) = \frac{1}{N} \sum_{j=1}^N a_{j;N}^m g_i(\xi_{j;N}).$$

We denote $\Gamma = \lim_{N \rightarrow \infty} \Gamma_N = \langle (p^i p^m) \rangle_{i,m=1}^M$. Let $h : \mathfrak{X} \rightarrow \mathbb{R}^d$ be any measurable function.

Theorem 1. ([9], Lemma 1) *Assume that:*

- (i) Γ exists, and $\det \Gamma \neq 0$;
- (ii) $\mathbf{E}[|h(\eta_m)|] < \infty$, $m = 1, \dots, M$.

Then $\hat{h}_N^m \xrightarrow{P} \mathbf{E}[h(\eta_m)]$ as $N \rightarrow \infty$ for all $m = 1, \dots, M$.

To formulate the asymptotic normality result for the simple moment estimators, we need some additional notation.

We consider the set of all moments \bar{g}_k^k , $k = 1, \dots, K$, as one long vector belonging to \mathbb{R}^d , $d := d_1 + \dots + d_K$:

$$\bar{g} := ((\bar{g}_1^1)^T, \dots, (\bar{g}_K^K)^T)^T \in \mathbb{R}^d. \quad (4a)$$

The corresponding estimators also form a long vector

$$\hat{g}_N := ((\hat{g}_{1;N}^1)^T, \dots, (\hat{g}_{K;N}^K)^T)^T \in \mathbb{R}^d. \quad (4b)$$

We denote the matrices of mixed second moments of $g_k(x)$, $k = 1, \dots, K$, and the corresponding estimators as

$$\bar{g}_{k,l}^m := \mathbf{E}[g_k(\eta_m)g_l(\eta_m)^T] \in \mathbb{R}^{d_k \times d_l}, \quad k, l = 1, \dots, K, \quad m = 1, \dots, M; \quad (5a)$$

$$\hat{g}_{k,l;N}^m := \frac{1}{N} \sum_{j=1}^N a_{j;N}^m g_k(\xi_{j;N}) g_l(\xi_{j;N})^T \in \mathbb{R}^{d_k \times d_l}, \quad k, l = 1, \dots, K. \quad (5b)$$

We consider the function T as a function of d -dimensional argument, that is, $T(y) := T(y^1, \dots, y^K)$. Then $\hat{T}_N := T(\hat{g}_N) = T(\hat{g}_{1;N}^1, \dots, \hat{g}_{K;N}^K)$.

Let us define the following matrices (assuming that the limits exist):

$$\alpha_{r,s;N} := (\alpha_{r,s;N}^{k,l})_{k,l=1,\dots,K} := ((a_{\cdot;N}^k a_{\cdot;N}^l \langle P_{\cdot;N}^r P_{\cdot;N}^s \rangle_N)_{k,l=1,\dots,K}) \in \mathbb{R}^{K \times K}; \quad (6a)$$

$$\alpha_{r,s} := (\alpha_{r,s}^{k,l})_{k,l=1,\dots,K} := \left(\lim_{N \rightarrow \infty} \alpha_{r,s;N}^{k,l} \right)_{k,l=1,\dots,K} \in \mathbb{R}^{K \times K}, \quad r, s = 1, \dots, M; \quad (6b)$$

$$\beta_{m;N} := (\beta_{m;N}^{k,l})_{k,l=1,\dots,K} := ((a_{\cdot;N}^k a_{\cdot;N}^l \langle P_{\cdot;N}^m \rangle_N)_{k,l=1,\dots,K}) \in \mathbb{R}^{K \times K}; \quad (7a)$$

$$\beta_m := (\beta_m^{k,l})_{k,l=1,\dots,K} := \left(\lim_{N \rightarrow \infty} \beta_{m;N}^{k,l} \right)_{k,l=1,\dots,K} \in \mathbb{R}^{K \times K}, \quad m = 1, \dots, M. \quad (7b)$$

Then the asymptotic covariance matrix of the normalized estimate $\sqrt{N}(\hat{g}_N - \bar{g})$ is Σ , where Σ consists of the blocks $\Sigma^{(k,l)}$:

$$\Sigma^{(k,l)} := \sum_{m=1}^M \beta_m^{k,l} \bar{g}_{k,l}^m - \sum_{r,s=1}^M \alpha_{r,s}^{k,l} \bar{g}_k^r (\bar{g}_l^s)^T \in \mathbb{R}^{d_k \times d_l}; \quad (8a)$$

$$\Sigma := (\Sigma^{(k,l)})_{k,l=1,\dots,K} \in \mathbb{R}^{d \times d}. \quad (8b)$$

Theorem 2. Assume that:

- (i) The functional moments \bar{g}_k^m , $\bar{g}_{k,l}^m$ exist and are finite for $k, l = 1, \dots, K$, $m = 1, \dots, M$.
- (ii) There exists $\delta > 0$ such that $\mathbf{E}[|g_k(\eta_m)|^{2+\delta}] < \infty$, $k = 1, \dots, K$, $m = 1, \dots, M$.
- (iii) There exist finite matrices Γ , Γ^{-1} , $\alpha_{r,s}$, and β_m for $r, s, m = 1, \dots, M$.

Then $\sqrt{N}(\hat{g}_N - \bar{g}) \xrightarrow{d} \zeta \simeq \mathcal{N}(\mathbb{O}_d, \Sigma)$, $N \rightarrow \infty$.

Thus, to construct a test for H_0 , we need a consistent estimator for Σ . The matrices $\alpha_{r,s;N}$ and $\beta_{m;N}$ are natural estimators for $\alpha_{r,s}$ and β_m . It is also natural to estimate $\bar{g}_{k,l}^m$ by $\hat{g}_{k,l;N}^m$ defined in (5b). In view of Theorem 1, these estimators are consistent under the assumptions of Theorem 2. But they can possess undesirable properties for moderate sample size. Indeed, note that $\hat{F}_{m;N}$ is not a probability distribution itself since the weights $a_{j;N}^m$ are negative for some j . Therefore, for example, the simple estimator of the second moment of some component can be negative, estimator (5b) for the positive semidefinite matrix $\bar{g}_{k,k}^m$ can be not positive semidefinite matrix, and so on. Due to the asymptotic normality result, this is not too troublesome for estimation of \bar{g} . But it causes serious difficulties when one uses an estimator of the asymptotic covariance matrix D based on $\hat{g}_{k,l;N}^m$ in order to calculate \hat{s}_N .

In [10], a technique is developed of $\hat{F}_{m;N}$ and \hat{h}_N^m improvement that allows one to derive estimators with more adequate finite sample properties if $\mathfrak{X} = \mathbb{R}$.

So, assume that $\xi(O) \in \mathbb{R}$ and consider the weighted empirical CDF

$$\hat{F}_{m;N}(x) = \frac{1}{N} \sum_{j=1}^N a_{j;N}^m \mathbb{1}\{\xi_{j;N} < x\}.$$

It is not a nondecreasing function, and it can attain values outside $[0, 1]$ since some $a_{j;N}^m$ are negative. The transform

$$\tilde{F}_{m;N}^+(x) = \sup_{y < x} \hat{F}_{m;N}(y)$$

yields a monotone function $\tilde{F}_{m;N}^+(x)$, but it still can be greater than 1 at some x . So, define

$$\hat{F}_{m;N}^+(x) = \min\{1, \tilde{F}_{m;N}^+(x)\}$$

as the improved estimator for $F_m(x)$. Note that this is an ‘‘improvement upward,’’ since $\tilde{F}_{m;N}^+(x) \geq \hat{F}_{m;N}(x)$. Similarly, a downward improved estimator can be defined as

$$\begin{aligned} \tilde{F}_{m;N}^-(x) &= \inf_{y \geq x} \hat{F}_{m;N}(y), \\ \hat{F}_{m;N}^-(x) &= \max\{0, \tilde{F}_{m;N}^-(x)\}. \end{aligned}$$

Any CDF that lies between $\hat{F}_{m;N}^-(x)$ and $\hat{F}_{m;N}^+(x)$ can be considered as an improved version of $\hat{F}_{m;N}(x)$. We will use only one such improvement, which combines $\hat{F}_{m;N}^-(x)$ and $\hat{F}_{m;N}^+(x)$:

$$\hat{F}_{m;N}^\pm(x) = \begin{cases} \hat{F}_{m;N}^+(x) & \text{if } \hat{F}_{m;N}^+(x, a) \leq 1/2, \\ \hat{F}_{m;N}^-(x) & \text{if } \hat{F}_{m;N}^-(x, a) \geq 1/2, \\ 1/2 & \text{otherwise.} \end{cases} \quad (9)$$

Note that all the three considered estimators $\hat{F}_{m;N}^*$ (* means any symbol from +, −, or ±) are piecewise constants on intervals between successive order statistics of

the data. Thus, they can be represented as

$$\hat{F}_{m;N}^*(x) = \frac{1}{N} \sum_{j=1}^N b_{j;N}^{m*} \mathbb{1}\{\xi_{j;N} < x\},$$

where $b_{j;N}^{m*}$ are some random weights that depend on the data.

The corresponding *improved* estimator for \bar{g}_i^m is

$$\hat{g}_{i;N}^{m*} = \int_{-\infty}^{+\infty} g_i(x) \hat{F}_{m;N}^*(dx) = \frac{1}{N} \sum_{j=1}^N b_{j;N}^{m*} g_i(\xi_{j;N}).$$

Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a measurable function.

Theorem 3. Assume that Γ exists and $\det \Gamma \neq 0$.

(I) If for some $c_- < c_+$, $c_- \leq \eta_m \leq c_+$ for all $m = 1, \dots, M$ and h has bounded variation on (c_-, c_+) , then $\hat{h}_N^{m*} \rightarrow \bar{h}^m$ a.s. as $N \rightarrow \infty$ for all $m = 1, \dots, M$ and $* \in \{+, -, \pm\}$.

(II) Assume that:

- (i) For some $\gamma > 0$, $\mathbb{E}[|h(\eta_m)|^{2+\gamma}] < \infty$.
- (ii) h is a continuous function of bounded variation on some interval $[c_-, c_+]$ and monotone on $(-\infty, c_-]$ and $[c_+, +\infty)$.

Then $\hat{h}_N^{m\pm} \rightarrow \bar{h}^m$ in probability.

4 Construction of the test

We first state an asymptotic normality result for \hat{T}_N . Denote

$$T'(y) := \left(\frac{\partial}{\partial y_1} T(y), \dots, \frac{\partial}{\partial y_d} T(y) \right) \in \mathbb{R}^{L \times d}.$$

Theorem 4. Assume that:

- (i) $T'(\bar{g})$ exist.
- (ii) The assumptions of Theorem 2 hold.
- (iii) The matrix $D = T'(\bar{g}) \Sigma (T'(\bar{g}))^T$ is nonsingular.

Then, under H_0 , $\sqrt{N} \hat{T}_N \xrightarrow{d} N(\mathbb{O}_L, D)$.

For the proof, see Appendix. Note that (iii) implies the nonsingularity of Σ .

Now, to estimate D , we can use

$$\hat{D}_N = T'(\tilde{g}_N) \tilde{\Sigma}_N (T'(\tilde{g}_N))^T,$$

where \tilde{g}_N is any consistent estimator for \bar{g} ,

$$\tilde{\Sigma}_N^{(k,l)} := \sum_{m=1}^M \beta_{m;N}^{k,l} \tilde{g}_{k,l;N}^m - \sum_{r,s=1}^M \alpha_{r,s;N}^{k,l} \tilde{g}_{k;N}^r (\tilde{g}_{l;N}^s)^T \in \mathbb{R}^{d_k \times d_l}; \quad (10a)$$

$$\tilde{\Sigma}_N := (\tilde{\Sigma}_N^{(k,l)})_{k,l=1,\dots,K} \in \mathbb{R}^{d \times d}, \quad (10b)$$

where $\tilde{g}_{k,l;N}^m$ is any consistent estimator for $\bar{g}_{k,l;N}^m$. For example, we can use

$$\tilde{g}_{k,l;N}^m = \hat{g}_{k,l;N}^{m\pm} = \frac{1}{N} \sum_{j=1}^N b_{j;N}^{m\pm} g_k(\xi_{j;N}) g_l(\xi_{j;N})^T$$

if $\mathfrak{X} = \mathbb{R}$ and the assumptions of Theorem 3 hold for all $h(x) = g_l^i(x) g_k^n(x)$, $i, k = 1, \dots, K$, $i = 1, \dots, d_l$, $n = 1, \dots, d_k$, $g_l(x) = (g_l^1(x), \dots, g_l^{d_l}(x))^T$.

Now let the test statistic be $\hat{s}_N = N(\hat{T}_N)^T \hat{D}_N^{-1} \hat{T}_N$. For a given significance level α , the test $\pi_{N,\alpha}$ accepts H_0 if $\hat{s}_N \leq Q^{\xi_L}(1 - \alpha)$ and rejects H_0 otherwise.

The p -level of the test (i.e., the attained significance level) can be calculated as $p = 1 - G(\hat{s}_N)$, where G means the CDF of χ_L^2 -distribution.

Theorem 5. *Let the assumptions of Theorem 4 hold. Moreover, assume the following:*

- (i) \tilde{g}_N and $\tilde{g}_{k,l;N}^m$ ($k, l = 1, \dots, K$, $m = 1, \dots, M$) are consistent estimators for \bar{g} and $\bar{g}_{k,l;N}^m$, respectively.
- (ii) T' is continuous in some neighborhood of \bar{g} .

Then $\lim_{N \rightarrow \infty} \mathbb{P}_{H_0} \{ \pi_{N,\alpha} \text{ rejects } H_0 \} = \alpha$.

Example 2 (Continued). Consider testing H_0^μ by the test $\pi_{N,\alpha}$ with $g_i(x) = x$ and $T(y_1, \dots, y_M) = (y_1 - y_2, y_2 - y_3, \dots, y_{M-1} - y_M)^T$. It is obvious that $T'(y)$ is a constant matrix of full rank. Assume that $\text{Var}[\eta_m] > 0$ for all $m = 1, \dots, M$ and $\det \Gamma \neq 0$. Then Σ is nonsingular, and so is D . Thus, in this case, assumptions (i) and (iv) of Theorem 2, (i) and (iii) of Theorem 4, and (ii) of Theorem 5 hold.

To ensure assumption (ii) of Theorem 2, we need $\mathbb{E}[|\eta_m|^{2+\delta}] < \infty$ for some $\delta > 0$ and all $m = 1, \dots, M$. In view of Theorem 1, this assumption also implies the consistency of \hat{g}_N and $\hat{g}_{kl;N}^m$. If one uses \hat{g}_N^\pm and $\hat{g}_{kl;N}^{m\pm}$ as estimators \tilde{g}_N and $\tilde{g}_{kl;N}^m$ in \hat{D}_N , then a more restrictive assumption $\mathbb{E}[|\eta_m|^{4+\delta}] < \infty$ is needed to ensure their consistency by Theorem 3.

5 Numerical results

5.1 Simulation study

To access the proposed test performance on samples of moderate size, we conducted a small simulation study. Three-component mixtures were analyzed ($M = 3$) with Gaussian components $F_m \sim N(\mu_m, \sigma_m^2)$. The concentrations were generated as $p_{j;N}^m = \zeta_{j;N}^m / s_{j;N}$, where $\zeta_{j;N}^m$ are independent, uniformly distributed on $[0, 1]$ random variables, and $s_{j;N} = \sum_{m=1}^M \zeta_{j;N}^m$. In all the experiments, 1000 samples were generated for each sample size $N = 50, 100, 250, 500, 750, 1000, 2000$, and 5000. Three modifications of $\pi_{N;\alpha}$ test were applied to each sample. In the first modification, (ss), simple estimators were used to calculate both \hat{T}_N and \hat{D}_N . In the second modification, (si), simple estimators were used in \hat{T}_N , and the improved ones were used in \hat{D}_N . In the last modification (ii), improved estimators were used in \hat{T}_N and \hat{D}_N . Note that the modification (ii) has no theoretical justification since, as far as we know, there are no results on the limit distribution of $\sqrt{N}(\hat{g}_N^\pm - \bar{g})$.

All tests were used with the nominal significance level $\alpha = 0.05$.

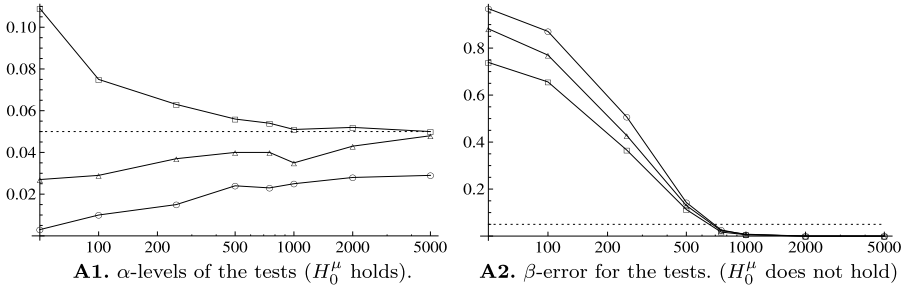


Fig. 1. Testing homogeneity of means (H_0^μ)

In the figures, frequencies of errors of the tests are presented. In the plots, \square corresponds to (ss), Δ to (si), and \circ to (ii) modification.

Experiment A1. In this experiment, we consider testing the mean homogeneity hypothesis H_0^μ . The means were taken $\mu_m = 0, m = 1, 2, 3$, so H_0^μ holds. To shadow the equality of means, different variances of components were taken, namely $\sigma_1^2 = 1, \sigma_2^2 = 4$, and $\sigma_3^2 = 9$. The resulting first-type error frequencies are presented on the left panel of Fig. 1. For the (ss) test, for small N , there were 1.4% cases of incorrect covariance matrix estimates (\hat{D}_N was not positive definite). Incorrect estimates were absent for large N .

Experiment A2. Here we also tested H_0^μ for components with the same variances as in A1. But $\mu_1 = 2$ and $\mu_2 = \mu_3 = 0$, so H_0^μ does not hold. The frequencies of the second-type error are presented on the right panel of Fig. 1. The percent of incorrect estimates \hat{D}_N is 1.6% for (ss) and small N .

Experiment B1. In this and the next experiment, we tested $H_0^\sigma: \sigma_1^2 = \sigma_2^2$. The data were generated with $\mu_1 = 0, \mu_2 = 3, \mu_3 = -2, \sigma_1^2 = \sigma_2^2 = 1$, and $\sigma_3^2 = 4$, so H_0^σ holds. The frequencies of the first-type error are presented on the left panel of Fig. 2. The percent of incorrect \hat{D}_N in (ss) varies from 19.4% for small N to 0% for large N .

Experiment B2. Now μ_m and σ_3^2 are the same as in B1, but $\sigma_1^2 = 1$ and $\sigma_2^2 = 4$, so H_0^σ does not hold. The frequencies of the second-type error are presented on the left panel of Fig. 2. The percent of incorrect \hat{D}_N in (ss) was 15.5% for small N and decreases to 0% for large N .

The presented results show reasonable agreement of the observed significance levels of the tests to their nominal level 0.05 when the sample sizes were larger then 500. The power of the tests increases to 1 as the sample sizes grow. It is interesting to note that the (ii) modification, although theoretically not justified, demonstrates the least first-type error and rather good power. From these results the (si) modification of the test seems the most prudent one.

5.2 Example of a sociological data analysis

Consider the data discussed in [9]. It consists of two parts. The first part is the data from the Four-Wave World Values Survey (FWWVS) held in Ukraine by the European Values Study Foundation (www.europeanvalues.nl) and World Values Survey

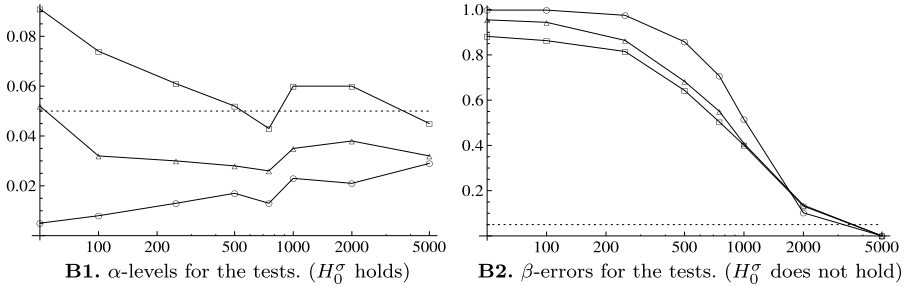


Fig. 2. Testing equality of variances (H_0^σ)

Table 1. Means (μ) and variances (σ^2) for the subjective income distribution on different political populations

	<i>PR</i>	<i>OC</i>	<i>Other</i>
μ	2.31733	2.65091	4.44504
μ^+	2.45799	2.64187	4.44504
σ^2	0.772514	4.85172	4.93788
σ^{2+}	2.09235	4.7639	4.93788

Association (www.worldvaluessurvey.org) in 2006. They contain answers of $N = 4006$ Ukrainian respondents on different questions about their social status and attitudes to different human values. We consider here the level of satisfaction of personal income (subjective income) as our variable of interest ξ , so $\xi_{j;N}$ is the subjective income of the j th respondent.

Our aim is to analyze differences in the distribution of ξ on populations of adherents of different political parties. Namely, we use the data on results of Ukrainian Parliament elections held in 2006. 46 parties took part in the elections. The voters could also vote against all or not to take part in the voting. We divided all the population of Ukrainian voters into three large groups (political subpopulations): \mathcal{P}_1 which contains adherents of the Party of Regions (*PR*, 32.14% of votes), \mathcal{P}_2 of Orange Coalition supporters (*OC* which consisted of “BJUT” and “NU” parties, 36.24%), and \mathcal{P}_3 of all others, including the persons who voted against all or did not take part in the pool (*Other*).

Political preferences of respondents are not available in the FWWVS data, so we used official results of the elections by 27 regions of Ukraine (see the site of Ukrainian Central Elections Commission www.cvk.gov.ua) to estimate the concentrations $p_{j;N}^m$ of the considered political subpopulations in the region where the j th respondent voted.

Means and variances of ξ over different subpopulations were estimated by the data (see Table 1). Different tests were performed to test their differences. The results are presented in the Table 2. Here μ_m means the expectation, and σ_m^2 means the variance of ξ over the m th subpopulation. Degrees of freedom for the limit χ^2 distribution are placed in the “df” column.

These results show that the hypothesis of homogeneity of all variances must be definitely rejected. The variances of ξ for *PR* and *OC* adherents are different, but the

Table 2. Test statistics and p -values for hypotheses on subjective income distribution

Hypotheses	ss	si	ii	df
$\mu_1 = \mu_2 = \mu_3$	11.776	10.8658	8.83978	2
p -value	0.00277252	0.0043704	0.0120356	
$\mu_1 = \mu_2$	2.15176	2.04539	0.621483	1
p -value	0.142407	0.152668	0.430497	
$\mu_1 = \mu_3$	10.7076	10.0351	8.75216	1
p -value	0.00106696	0.00153585	0.00309236	
$\mu_2 = \mu_3$	7.40835	7.10653	7.17837	1
p -value	0.00649218	0.00768036	0.00737877	
$\sigma_1^2 = \sigma_2^2 = \sigma_3^2$	15.8317	14.786	6.40963	2
p -value	0.000364914	0.000615547	0.0405664	
$\sigma_1^2 = \sigma_2^2$	14.7209	13.8844	5.95528	1
p -level	0.000124657	0.000194405	0.0146733	
$\sigma_1^2 = \sigma_3^2$	1.92166	1.77162	0.826778	1
p -level	0.165674	0.183182	0.363206	
$\sigma_2^2 = \sigma_3^2$	0.000741088	0.00072198	0.00294353	1
p -level	0.978282	0.978564	0.956733	

tests failed to observe significant differences in the pairs of variances *PR-Other* and *OC-Other*. For the means, all the tests agree that *PR* and *OC* has the same mean ξ , whereas the mean of *Other* is different from the common mean of *PR* and *OC*.

6 Concluding remarks

We developed a technique that allows one to construct testing procedures for different hypotheses on functional moments of mixtures with varying concentrations. This technique can be applied to test the homogeneity of means or variances (or both) of some components of the mixture. Performance of different modifications of the test procedure is compared in a small simulation study. The (ss) modification showed the worst first-type error and the highest power. The (ii) test has the best first-type error and the worst power. It seems that the (si) modification can be recommended as a golden mean.

Acknowledgement

The authors are thankful to the anonymous referee for fruitful comments.

A Appendix

Proof of Theorem 2. Note that $S_N = \sqrt{N}(\hat{g}_N - \bar{g}) = \sum_{j=1}^N \zeta_{j;N}$, where

$$\zeta_{j;N} = \frac{1}{\sqrt{N}}(a_{j;N}^i(g_1(\xi_{j;N}) - M[g_1(\xi_{j;N})]))^T, i = 1, \dots, K)^T.$$

We will apply the CLT with the Lyapunov condition (see Theorem 11 from Chapter 8 and Remark 4 in Section 4.8 in [2]) to S_N . It is readily seen that $\zeta_{j;N}, j = 1, \dots, N$, are independent for fixed N and $E[\zeta_{j;N}] = 0$.

Let $\Sigma_{j;N} = \text{Cov}(\zeta_{j;N})$. Then $\Sigma_{j;N}$ consists of the blocks

$$\begin{aligned} \Sigma_{j;N}^{(k,l)} &= \frac{1}{N} a_{j;N}^k a_{j;N}^l (\mathbb{E}[g_k(\xi_{j;N})(g_l(\xi_{j;N}))^T] - \mathbb{E}[g_k(\xi_{j;N})] \mathbb{E}[g_l(\xi_{j;N})]^T) \\ &= \frac{1}{N} a_{j;N}^k a_{j;N}^l \left(\sum_{m=1}^M p_{j;N} \bar{g}_{k,l}^m - \left(\sum_{m=1}^M p_{j;N} \bar{g}_k^m \right) \left(\sum_{m=1}^M p_{j;N} \bar{g}_l^m \right)^T \right). \end{aligned}$$

It is readily seen that $\sum_{j=1}^N \Sigma_{j;N}^{(k,l)} \rightarrow \Sigma^{(k,l)}$ as $N \rightarrow \infty$. So

$$\text{Cov } S_N \rightarrow \Sigma \quad \text{as } N \rightarrow \infty. \tag{11}$$

To apply the CLT, we only need to verify the Lyapunov condition

$$\sum_{j=1}^N \mathbb{E}[|\zeta_{j;N}|^{2+\delta}] \rightarrow 0 \quad \text{for some } \delta > 0. \tag{12}$$

Note that assumption (iii) implies

$$\sup_{1 \leq j \leq N, 1 < m \leq M, N > N_0} |a_{j;N}^m| < C_1$$

for some N_0 and $C_1 < \infty$; thus,

$$\sum_{j=1}^N \mathbb{E}[|\xi_{j;N}|^{2+\delta}] \leq \sum_{j=1}^N \frac{C_1^{2+\delta}}{N^{1+\delta/2}} \mathbb{E}[|g(\xi_{j;N})|^{2+\delta}], \tag{13}$$

where $g(x) = (g_1(x)^T, \dots, g_K(x)^T)$. Since $|g(\xi_{j;N})|^2 = \sum_{k=1}^K |g_k(\xi_{j;N})|^2$ and, by the Hölder inequality,

$$|g(\xi_{j;N})|^{2+\delta} \leq K^{\delta/2} \sum_{k=1}^K |g_k(\xi_{j;N})|^{2+\delta},$$

we obtain

$$\begin{aligned} \mathbb{E}[|g(\xi_{j;N})|^{2+\delta}] &\leq K^{\delta/2} \sum_{k=1}^K \mathbb{E}[|g_k(\xi_{j;N})|^{2+\delta}] \\ &= K^{\delta/2} \sum_{k=1}^K \sum_{m=1}^M p_{j;N}^m \mathbb{E}[|g(\eta_m)|^{2+\delta}] < C_2 < \infty, \end{aligned}$$

where the constant C_2 does not depend on j and N . This, together with (13), yields (12). By the CLT we obtain $S_N \xrightarrow{d} N(\mathbb{O}, \Sigma)$. □

Proof of Theorem 3. Part (I). Since $\hat{F}_{m;N}^+(x)$ is piecewise constant and $F_m(x)$ is nondecreasing, the \sup_x of $|\hat{F}_{m;N}^+(x) - F_m(x)|$ can be achieved only at jump points of $\hat{F}_{m;N}^+(x)$. But $\hat{F}_{m;N}^+(x) \geq \hat{F}_{m;N}(x)$ for all x , and if x is a jump point, then

$$\hat{F}_{m;N}(x-) \leq \hat{F}_{m;N}^+(x-) \leq \hat{F}_{m;N}^+(x) \leq \hat{F}_{m;N}(x).$$

Therefore,

$$\sup_{x \in \mathbb{R}} |\hat{F}_{m;N}^+(x) - F_m(x)| \leq \sup_{x \in \mathbb{R}} |\hat{F}_{m;N}(x) - F_m(x)|.$$

Similarly,

$$\sup_{x \in \mathbb{R}} |\hat{F}_{m;N}^*(x) - F_m(x)| \leq \sup_{x \in \mathbb{R}} |\hat{F}_{m;N}(x) - F_m(x)|.$$

By the Glivenko–Cantelli-type theorem for weighted empirical distributions (which can be derived, e.g., as a corollary of Theorem 2.4.2 in [7])

$$\sup_{x \in \mathbb{R}} |\hat{F}_{m;N}(x) - F_m(x)| \rightarrow 0 \quad \text{a.s. as } N \rightarrow \infty$$

if $\sup_{j=1, \dots, N; N > N_0} |a_{j;N}^m| < \infty$. The latter condition is fulfilled since $\det \Gamma \neq 0$. Thus,

$$\sup_{x \in \mathbb{R}} |\hat{F}_{m;N}^*(x) - F_m(x)| \rightarrow 0 \quad \text{a.s. as } N \rightarrow \infty. \tag{14}$$

For any $h : \mathbb{R} \rightarrow \mathbb{R}$ and any interval $A \subseteq \mathbb{R}$, let $V_A(h)$ be the variation of h on A . Take $A = (c_-, c_+)$. Then, under the assumptions of the theorem,

$$\begin{aligned} |\hat{h}_N^{m*} - \bar{h}_m| &= \left| \int_A h(x) d(\hat{F}_{m;N}^*(x) - F_m(x)) \right| \\ &\leq \sup_{x \in A} |\hat{F}_{m;N}(x) - F_m(x)| \cdot V_A(h) \rightarrow 0 \quad \text{a.s. as } N \rightarrow \infty. \end{aligned}$$

Part (II). Note that if the assumptions of this part hold for some $A = (c_-, c_+)$, then they will also hold for any new c_-, c_+ such that $A \subset (c_-, c_+)$. Thus, we may assume that $F_m(c_-) < 1/4$ and $F_m(c_+) > 3/4$.

Consider the random event $B_N^1 = \{\hat{F}_{m;N}^\pm(x) = \tilde{F}_{m;N}^\pm(x) \text{ for all } x \leq c_-\}$. Then (14) implies $\mathbb{P}\{B_N^1\} \rightarrow 1$ as $N \rightarrow \infty$.

We bound

$$|\hat{h}_N^{m\pm} - \bar{h}^m| \leq J_1 + J_2 + J_3, \tag{15}$$

where

$$\begin{aligned} J_1 &= \left| \int_{-\infty}^{c_-} h(x) d(\hat{F}_{m;N}^\pm(x) - F_m(x)) \right|, \\ J_2 &= \left| \int_{c_-}^{c_+} h(x) d(\hat{F}_{m;N}^\pm(x) - F_m(x)) \right|, \\ J_3 &= \left| \int_{c_+}^{+\infty} h(x) d(\hat{F}_{m;N}^\pm(x) - F_m(x)) \right|. \end{aligned}$$

Then $J_2 \xrightarrow{P} 0$ as in Part (I).

Let us assume that the event B_N^1 occurred and bound

$$J_1 \leq \left| \int_{-\infty}^{c_-} h(x) d(\tilde{F}_{m;N}^+(x) - F_m(x)) \right|. \tag{16}$$

If $h(x)$ is bounded as $x \rightarrow -\infty$, then we can take $c_- = -\infty$ and obtain $J_1 = 0$. Consider the case of unbounded h . Since h is monotone, we have $h(x) \rightarrow +\infty$ or $h(x) \rightarrow +\infty$ as $x \rightarrow -\infty$. We will consider the first case; the reasoning for the second one is analogous. Thus, $h(x) \rightarrow +\infty$ as $x \rightarrow -\infty$, and we can take $h(x) > 0$ for $x < c_-$.

By the inequality (16) in [10],

$$\mathbb{P}\left[\sup_{t < x} |\tilde{F}_{m;N}^+(t) - F_m(t)| > \varepsilon\right] \leq C_1(\bar{F}^2(x)\varepsilon^{-4}N^{-2} + \bar{F}(x)\varepsilon^{-2}N^{-1}), \quad (17)$$

where $\bar{F}(x) = \sum_{m=1}^M F_m(x)$, $C_1 < \infty$.

Let us take x_0, \dots, x_n, \dots such that $h(x_j) = 2^j h(c_-)$. By assumption (ii) and the Chebyshev inequality,

$$\bar{F}(x) = \sum_{m=1}^M \mathbb{P}[\eta_m < x] \leq \sum_{m=1}^M h^{-2-\gamma}(x) \mathbb{E}[|h(\eta_m)|^{2+\gamma}],$$

and

$$\bar{F}(x_j) \leq D2^{-(2+\gamma)j}$$

for some $D < \infty$.

Let $\varepsilon_j = 2^{-(1+\gamma/4)j} N^{-1/4}$. Then by (17)

$$\mathbb{P}\left[\sup_{t < x_j} |\tilde{F}_{m;N}^+(t) - F_m(t)| > \varepsilon_j\right] \leq C_2(2^{-\gamma j} N^{-1} + 2^{-\gamma j/2} N^{-1/2})$$

for some $C_2 < \infty$. Denote $B_N^2 = \cap_j \{\sup_{t < x_j} |\tilde{F}_{m;N}^+(t) - F_m(t)| \leq \varepsilon_j\}$. Then

$$\mathbb{P}[B_N^2] \geq 1 - \sum_{j=1} C_2(2^{-\gamma j} N^{-1} + 2^{-\gamma j/2} N^{-1/2}) \geq 1 - C_3 N^{-1} - C_4 N^{-1/2} \rightarrow 1$$

as $N \rightarrow \infty$. Now, $J_1 = |\int_{-\infty}^{c_-} h(x)d(\tilde{F}_{m;N}^+(x) - F_m(x))|$. If B_N^1 and B_N^2 occur, then

$$\begin{aligned} J_1 &\leq \left| \sum_{j=0}^N \int_{x_{j+1}}^{x_j} |\tilde{F}_{m;N}^+(x) - F_m(x)| h(dx) \right| \\ &\leq \sum_{j=0}^N h(c_-) 2^{-(1+\gamma/4)j} N^{-1/4} (2^{j+1} - 2^j) \leq C_5 N^{-1/4}. \end{aligned}$$

Thus, $\mathbb{P}[J_1 \leq C_5 N^{-1/4}] \geq \mathbb{P}[B_N^1 \cap B_N^2] \rightarrow 1$ and $J_1 \xrightarrow{P} 0$.

Similarly, $J_3 \xrightarrow{P} 0$.

Combining these bounds with (15), we accomplish the proof. □

Proof of Theorem 4. This theorem is a simple corollary of Theorem 2 and the continuity theorem for weak convergence (Theorem 3B in Chapter 1 of [3]). □

Proof of Theorem 5. Since \tilde{g}_N and $\tilde{g}_{kl;N}^m$ are consistent, $\tilde{\Sigma}_N \xrightarrow{P} \Sigma$. Similarly, the continuity of T' and consistency of \tilde{g}_N imply $T'(\tilde{g}_N) \xrightarrow{P} T'(\bar{g})$. Then, with $\det D \neq 0$ in mind, we obtain $\hat{D}_N^{-1} \xrightarrow{P} D^{-1}$.

Denote $\tilde{s}_N = N\hat{T}_N^T D^{-1} \hat{T}_N$. By Theorem 4 and the continuity theorem, $\tilde{s}_N \xrightarrow{d} \chi_L^2$. By Theorem 4 $\sqrt{N}\hat{T}_N$ is stochastically bounded. Thus,

$$|\tilde{s}_N - \hat{s}_N| = |\sqrt{N}\hat{T}_N^T (D^{-1} - \hat{D}_N^{-1})(\sqrt{N}\hat{T}_N)| \xrightarrow{P} 0.$$

Therefore, \hat{s}_N converges in distribution to the same limit as \tilde{s}_N , that is, to χ_L^2 . \square

References

- [1] Autin, F., Pouet, Ch.: Test on the components of mixture densities. *Stat. Risk. Model.* **28**(4), 389–410 (2011). [MR2877572](#). doi:[10.1524/strm.2011.1065](#)
- [2] Borovkov, A.A.: *Probability Theory*. Gordon and Breach Science Publishers, Amsterdam (1998). [MR1711261](#)
- [3] Borovkov, A.A.: *Mathematical statistics*. Gordon and Breach Science Publishers, Amsterdam (1998). [MR1712750](#)
- [4] Johnson, N.L.: Some simple tests of mixtures with symmetrical components. *Commun. Stat.* **1**(1), 17–25 (1973). [MR0315824](#). doi:[10.1080/03610927308827004](#)
- [5] Liu, X., Pasarica, C., Shao, Y.: Testing homogeneity in gamma mixture model. *Scand. J. Stat.* **20**, 227–239 (2003). [MR1965104](#). doi:[10.1111/1467-9469.00328](#)
- [6] Maiboroda, R.E.: A test for the homogeneity of mixtures with varying concentrations. *Ukr. J. Math.* **52**(8), 1256–1263 (2000). [MR1819720](#). doi:[10.1023/A:1010305121413](#)
- [7] Maiboroda, R.: *Statistical Analysis of Mixtures*. Kyiv University Publishers, Kyiv (2003). (in Ukrainian)
- [8] Maiboroda, R., Sugakova, O.: Statistics of mixtures with varying concentrations with application to DNA microarray data analysis. *J. Nonparametr. Stat.* **24**(1), 201–205 (2012). [MR2885834](#). doi:[10.1080/10485252.2011.630076](#)
- [9] Maiboroda, R.E., Sugakova, O.V., Doronin, A.V.: Generalized estimating equations for mixtures with varying concentrations. *Can. J. Stat.* **41**(2), 217–236 (2013). [MR3061876](#). doi:[10.1002/cjs.11170](#)
- [10] Majboroda, R., Kubajchuk, O.: Improved estimates for moments by observations from mixtures. *Theory Probab. Math. Stat.* **70**, 83–92 (2005). doi:[10.1090/S0094-9000-05-00642-3](#)
- [11] McLachlan, G.J., Peel, D.: *Finite Mixture Models*. Wiley-Interscience, New York (2000). [MR1789474](#). doi:[10.1002/0471721182](#)
- [12] Shcherbina, A.: Estimation of the mean value in the model of mixtures with varying concentrations. *Theory Probab. Math. Stat.* **84**, 151–164 (2012). [MR2857425](#). doi:[10.1090/S0094-9000-2012-00866-1](#)
- [13] Titterton, D.M., Smith, A.F., Makov, U.E.: *Analysis of Finite Mixture Distributions*. Wiley, New York (1985)