Теорія Ймовір. та Матем. Статист.
Вип. 93, 2015

Theor. Probability and Math. Statist.
No. 93, 2016, Pages 123–135
http://dx.doi.org/10.1090/tpms/998
Article electronically published on February 7, 2017

ADAPTIVE TEST ON MEANS HOMOGENEITY BY OBSERVATIONS FROM A MIXTURE

UDC 519.21

R. E. MAĬBORODA AND O. V. SUGAKOVA

Abstract. We consider the problem of testing the homogeneity of two components of a mixture with varying mixing probabilities and construct an adaptive test that minimizes the asymptotic probability of error of the second kind for local alternatives.

1. Introduction

Finite mixture models are often used to fit statistical data in biology and sociology (see, for example, [11, 9]). One of the problems arising when analyzing such data is to test the homogeneity of certain probability characteristics (means, variances, distribution functions, etc.) for different components of a mixture. In the current paper, we are concerned with the problem of testing the homogeneity of certain functional moments of two components of a mixture. We assume that the concentrations of components (mixing probabilities) of a mixture are known but may vary with observations. This problem is studied in the paper [10] where we propose a test based on the minimax estimators of the corresponding moments.

A natural generalization of this approach is to use a linear combination of observations with some weights $\mathbf{b}$ as the test statistics. Different weights $\mathbf{b}$ generate tests of different powers (for a given significance level). To choose optimal weights $\mathbf{b}$ we investigate the powers of tests at local alternatives and choose $\mathbf{b}^{\mathrm{opt}}$ such that the corresponding asymptotic power is maximal. Since the optimal weights $\mathbf{b}^{\mathrm{opt}}$ depend on unknown distributions of components, the resulting adaptive test uses estimators $\hat{\mathbf{b}}$ for $\mathbf{b}^{\mathrm{opt}}$. The main result of the paper is that the test based on $\hat{\mathbf{b}}$ is of the same asymptotic power as the test based on the optimal weights $\mathbf{b}^{\mathrm{opt}}$.

The results of the paper [10] are generalized in [8] for solving the problem of testing the homogeneity of moments for more than two components. Minimax estimators for moments are used in [8] to construct the tests. A test for equality of distributions of two components based on wavelet estimators of the probability density is considered in [7]. A test for the homogeneity of mixtures based on censored data is proposed in [5]. Other statistical problems related to mixtures with varying concentrations are considered in the papers [3, 6].

The paper is organized as follows. Section 2 contains the setting of the problem. Some auxiliary definitions and results concerning the estimation of functional moments by using the observations obtained from a mixture with varying concentrations are provided in Section 3. A general description of tests with arbitrary nonrandom weight coefficients $\mathbf{b}$

Key words and phrases. Adaptive algorithms, local alternatives, models of mixtures with varying concentrations, test for homogeneity of two means.

This paper was prepared following the talk at the International conference "Probability, Reliability and Stochastic Optimization (PRESTO-2015)" held in Kyiv, Ukraine, April 7–10, 2015.

©2017 American Mathematical Society

is given in Section 4. An investigation of the asymptotic power of such tests at local alternatives is presented in Section 5. An adaptive test with estimated optimal coefficients is constructed and the main result about its optimality is given in Section 6. Section 7 is devoted to a description of results obtained with the help of computer simulation; here we compare the powers of adaptive and nonadaptive tests for samples of a fixed size. The proofs are collected in the Appendix.

2. Setting of the problem

The statistical data in the model of mixtures with varying concentrations is a set $X_n = (\xi_{1;n}, \dots, \xi_{n;n})$ of observations $\xi_{j;n}$ that are independent random elements of some probability space $\mathcal{X}$ with the distribution

$$\mathsf{P}\{\xi_{j;n} \in A\} = \Psi_{j;n}(A) = \sum_{i=1}^{M} p^i_{j;n} F_i(A), \tag{1}$$

where F_i denotes the distribution of the ith component of the mixture, $p^i_{j;n}$ is the concentration of the ith component at the moment when the jth observation is collected ($p^i_{j;n}$ is called the mixing probability in [4]). In the current paper, we assume that the concentrations of components are known but the distributions are not known.

Let $g\colon \mathcal{X} \to \mathbb{R}$ be some measurable function such that the corresponding functional moments of components

$$\bar{g}_i = \int_{\mathcal{X}} g(x)\, F_i(dx), \qquad i = 1, \dots, M,$$

are finite. The problem of testing the equality of functional moments for two components of a mixture is considered in [10]. Without loss of generality we assume that we are dealing with a test of the hypothesis $H_0\colon \bar{g}_1 = \bar{g}_2$. The test proposed in [10] is based on the statistics being a weighted mean of the form

$$S_n = S_n(\mathbf{b}) = \frac{1}{n} \sum_{j=1}^{n} b_{j;n} g(\xi_{j;n}), \tag{2}$$

where $b_{j;n} = b^{\text{naive}}_{j;n}$ are "naive" coefficients that equal the difference between minimax coefficients (3) used to estimate $\bar{g}_1$ and $\bar{g}_2$. A Student type statistic S is compared in the test with a threshold that guarantees a given asymptotic significance level α of the test.

Along with $b^{\text{naive}}_{j;n}$, other weight coefficients can be used in statistics (2) for which the following natural condition is satisfied:

$$\text{given } H_0, \quad \mathsf{E}\, S_n(\mathbf{b}) = 0.$$

In the rest of this paper, we are dealing with the problem of choosing the optimal coefficients.

3. Estimation of moments in the model of mixture with varying coefficients

We recall some basic approaches to estimating the moments by using observations obtained from a mixture with varying concentrations.

In what follows we denote by $\langle \cdot \rangle_n$ the averaging over all observations for triangle matrices $\mathbf{b}$, $\mathbf{p}^i = (p^i_{j:n}, j = 1, \dots, n; n = 1, 2, \dots)$. The operations written inside the triangle brackets are understood in the coordinatewise sense. For example,

$$\langle \mathbf{b} \rangle_n = \frac{1}{n} \sum_{j=1}^{n} b_{j:n}, \qquad \left\langle \mathbf{p}^i \mathbf{p}^k \right\rangle_n = \frac{1}{n} \sum_{j=1}^{n} p^i_{j:n} p^k_{j:n}.$$

Let $\langle \mathbf{b} \rangle = \lim_{n\to\infty} \langle \mathbf{b} \rangle_n$ provided the limit exists. Consider the matrix

$$\Gamma_n = \left(\langle \mathbf{p}^i \mathbf{p}^k \rangle_n \right)_{i,k=1}^M$$

that can be treated as the Gram matrix for the set of vectors $(\mathbf{p}^1, \dots, \mathbf{p}^M)$ in the scalar product $\langle \mathbf{p}^i, \mathbf{p}^k \rangle_n = \langle \mathbf{p}^i \mathbf{p}^k \rangle_n$. Correspondingly, $\Gamma = \lim_{n\to\infty} \Gamma_n$.

Let $\det \Gamma_n \neq 0$. The minimax coefficients for the estimation of the distribution of ith component of a mixture are defined by

$$a^i_{j;n} = \frac{1}{\det \Gamma_n} \sum_{m=1}^M (-1)^{m+i} \gamma_{mi;n} p^m_{j;n}, \tag{3}$$

where $\gamma_{mi;n}$ is the minor (m, i) for the matrix Γ_n (see [4, Section 2.1] where we proved that these coefficients are minimax). The corresponding estimator for $\bar{g}_i$ is defined by

$$\hat{g}_{i;n} = \frac{1}{n} \sum_{j=1}^n a^i_{j;n} g(\xi_{j;n}). \tag{4}$$

Asymptotic properties of these estimators are described in [4, Section 2]. In particular, it is shown in [4] that these estimators are consistent under rather wide assumptions.

4. Constructing a nonadaptive test

Consider the statistic $S_n(\mathbf{b})$ defined by equality (2). Assume that $H_0 : \bar{g}_1 = \bar{g}_2$. Then $\mathsf{E}\, S_n(\mathbf{b}) = 0$ if

$$\left\langle \mathbf{b} \left(\mathbf{p}^1 + \mathbf{p}^2 \right) \right\rangle_n = 0, \qquad \left\langle \mathbf{b}\mathbf{p}^i \right\rangle_n = 0 \quad \text{for } i = 2, \dots, M. \tag{5}$$

In addition, if $\langle \mathbf{b}\mathbf{p}^1 \rangle_n \neq 0$, then $\mathsf{E}\, S_n(\mathbf{b}) = \langle \mathbf{b}\mathbf{p}^1 \rangle_n (\bar{g}_1 - \bar{g}_2) \neq 0$ for $\bar{g}_1 \neq \bar{g}_2$. Therefore, small values of $S_n(\mathbf{b})$ are in favor of the main (zero) hypothesis, while large deviations from zero are in favor of the alternative. To choose the critical value for the test that guarantees a given asymptotic significance level, we consider a Student type transformation of the statistic $S_n(\mathbf{b})$. Note that its variance equals

$$\operatorname{Var} S_n(\mathbf{b}) = \frac{1}{n} \left[\sum_{m=1}^M \left\langle (\mathbf{b})^2 \mathbf{p}^m \right\rangle_n \overline{g^2}_m - \sum_{m_1, m_2 = 1}^M \left\langle (\mathbf{b})^2 \mathbf{p}^{m_1} \mathbf{p}^{m_2} \right\rangle_n \bar{g}_{m_1} \bar{g}_{m_2} \right],$$

where

$$\overline{g^2}_m = \int (g(x))^2 \, F_m(dx).$$

The estimator $\hat{g}_{i;n}$ for $\bar{g}_i$ is defined by (4). A natural estimator for $\overline{g^2}_m$ is

$$\widehat{g^2}_{m;n} = \frac{1}{n} \sum_{j=1}^n a^m_{j;n} (g(\xi_{j;n}))^2. \tag{6}$$

Substituting these estimators to the expression for $\operatorname{Var} S(\mathbf{b})$ in place of the corresponding moments we get the following estimator:

$$D_n(\mathbf{b}) = D_n = \frac{1}{n} \left[\sum_{m=1}^M \left\langle (\mathbf{b})^2 \mathbf{p}^m \right\rangle_n \widehat{g^2}_{m;n} - \sum_{m_1, m_2 = 1}^M \left\langle (\mathbf{b})^2 \mathbf{p}^{m_1} \mathbf{p}^{m_2} \right\rangle_n \widehat{g}_{m_1;n} \widehat{g}_{m_2;n} \right].$$

Set

$$D_\infty(\mathbf{b}) = \lim_{n\to\infty} n \operatorname{Var} S_n(\mathbf{b}) = \sum_{m=1}^M \left\langle (\mathbf{b})^2 \mathbf{p}^m \right\rangle \overline{g^2}_m - \sum_{m_1, m_2 = 1}^M \left\langle (\mathbf{b})^2 \mathbf{p}^{m_1} \mathbf{p}^{m_2} \right\rangle \bar{g}_{m_1} \bar{g}_{m_2}$$

provided the limits exist.

Lemma 4.1. *Let*

1. $\det\Gamma \neq 0$;
2. *the limits* $\langle(\mathbf{b})^2\mathbf{p}^{m_1}\mathbf{p}^{m_2}\rangle$, $m_1, m_2 = 1, \dots, M$, *exist;*
3. $\overline{g^2}_m < \infty$, $m = 1, \dots, M$.
4. $D_\infty(\mathbf{b}) \neq 0$

Then $D_n(\mathbf{b})/\operatorname{Var} S_n(\mathbf{b}) \to 1$ *in probability as* $n \to \infty$.

Consider the Student type statistic

$$T_n(\mathbf{b}) = T_n = \frac{S_n(\mathbf{b})}{\sqrt{D_n(\mathbf{b})}}. \tag{7}$$

Then the test for the hypothesis H_0 based on $T_n(\mathbf{b})$ is written as follows:

$$\pi_{\mathbf{b}}(X_n) = \mathbb{1}\{|T_n(\mathbf{b})| > \lambda_{\alpha/2}\},$$

where $\lambda_{\alpha/2}$ is the quantile of level $1-\alpha/2$ for the standard normal distribution. In other words, the test $\pi_{\mathbf{b}}(X_n)$ accepts the hypothesis H_0 if $|T_n(\mathbf{b})| \le \lambda_{\alpha/2}$, and rejects it if $|T_n(\mathbf{b})| > \lambda_{\alpha/2}$.

Using Lemma 4.1 and the asymptotic normality of $S_n(\mathbf{b})$ (see [10]) we obtain the following result.

Theorem 4.1. *Let conditions 1–4 of Lemma* 4.1 *hold and, moreover,*

$$\sup_{j,n} |b_{j;n}| < \infty.$$

Then:

1. *under hypothesis* H_0,
$$T_n(\mathbf{b}) \Rightarrow N(0,1) \quad \text{as } n \to \infty;$$
2. *the asymptotic significance level of the test* $\pi_{\mathbf{b}}(X_n)$ *equals*
$$\lim_{n\to\infty} \mathsf{P}_{H_0}\{\pi_{\mathbf{b}}(X_n) = 1\} = \alpha.$$

5. Asymptotic study of nonadaptive test under near alternatives

When studying the asymptotic behavior of tests under near alternatives we restrict the consideration to the case of $\mathcal{X} = \mathbb{R}$, $g(x) = x$, and where the distributions F_i, $i = 2, \dots, M$, of all components, except the first one, do not depend on the size of a sample. The distribution of the first component

$$F_{1;n}(x) = F_{1;\infty}(x - v_n) \tag{8}$$

is fixed up to a shift parameter $v_n = v/\sqrt{n}$. Here, $F_{1;\infty}$ is a fixed probability distribution such that

$$\int x\, F_{1;\infty}(dx) = \int x\, F_2(dx), \tag{9}$$

that is, the hypothesis H_0 holds for a mixture with $F_1 = F_{1;\infty}$.

From the practical purposes, the latter assumption is not crucial. Indeed, if the distribution of the sample $X_n = (\xi_{1;n}, \dots, \xi_{n;n})$ of a fixed size n is given by (1), then the set $(\eta_1, \dots, \eta_n)$, $\eta_{j;n} = g(\xi_{j;n})$, is also a sample from a mixture with varying concentrations and the distribution of the first component can be represented in the form of (8) with a distribution function $F_{1;\infty}$ that satisfies condition (9) and

$$v = \sqrt{n} \int g(x)(F_1(dx) - F_2(dx)).$$

In the framework of model (1), the hypothesis $H_{1;n}$ is that the distribution function $F_1(x) = F_{1;n}$ is given by (8) and equality (9) holds.

Put

$$\bar{g}_1 = \int g(x)x\, F_{1;\infty}(dx).$$

The following result describes the asymptotic behavior of the probability of error of the second kind for the test $\pi_{\mathbf{b}}$ at alternatives $H_{1;n}$.

Theorem 5.1. *Let*

1. $\det\Gamma \neq 0$;
2. $\langle(\mathbf{b})^2\mathbf{p}^{m_1}\mathbf{p}^{m_2}\rangle$ *exist for all* $m_1, m_2 = 1, \dots, M$;
3. $\overline{g^2}_i < \infty$ *for all* $i = 1, \dots, M$;
4. $D_\infty(\mathbf{b}) \neq 0$;
5. $\sup_{j,n} |b_{j;n}| < \infty$.

Then

$$\beta(\mathbf{b}) \stackrel{\text{def}}{=} \lim_{n\to\infty} \mathsf{P}_{H_{1;n}}\{\pi_{\mathbf{b}}(X_n) = 0\} = \mathsf{P}\{|R(\mathbf{b}) + \zeta| < \lambda_{\alpha/2}\},$$

where ζ *is a standard normal random variable,*

$$R(\mathbf{b}) = \frac{v\left\langle \mathbf{bp}^1\right\rangle}{\sqrt{D_\infty(\mathbf{b})}}. \tag{10}$$

6. Adaptive test

For a given significance level, the test with a minimal probability of error of the second kind is optimal. It follows from the above asymptotic analysis that, given the alternative $H_{1;n}$, the tests $\pi_{\mathbf{b}}$ have the minimal asymptotic probability of error of the second kind if $R(\mathbf{b})$ is maximal. Moreover, the weights $\mathbf{b}$ have to satisfy condition (5). Since $R(\mathbf{b})$ is not changed if $\mathbf{b}$ is multiplied by a constant, one can introduce an additional normalization condition $\langle\mathbf{bp}^1\rangle = 1$. Then the problem of maximization of $R(\mathbf{b})$ reduces to the problem of minimization of $D_\infty(\mathbf{b})$ provided that

$$\left\langle \mathbf{bp}^1\right\rangle = 1, \qquad \left\langle \mathbf{bp}^2\right\rangle = -1, \qquad \left\langle \mathbf{bp}^i\right\rangle = 0, \qquad i = 3, \dots, M.$$

The latter conditions are equivalent to conditions (5) if the normalization condition holds. We change this problem by its “pre-limit” analog, that is, we minimize $D_n(\mathbf{b})$ under the conditions

$$\left\langle \mathbf{bp}^1\right\rangle_n = 1, \qquad \left\langle \mathbf{bp}^2\right\rangle_n = -1, \qquad \left\langle \mathbf{bp}^i\right\rangle_n = 0, \qquad i = 3, \dots, M.$$

Using the Lagrange multipliers method, we obtain a solution of this conditional optimization problem, namely

$$\mathbf{b} = \mathbf{b}^{\text{opt}} = \left(b_{j;n}^{\text{opt}}(\mathbf{G})\right)_{j=1}^{n},$$

where

$$b^{\rm opt}_{j;n}(\mathbf{G}) = \frac{1}{d_{j;n}(\mathbf{G})}\sum_{m=1}^{M}\lambda_m(\mathbf{G})p^m_{j;n},$$

$$\mathbf{G} = \left(\bar g_1,\dots,\bar g_M,\overline{g^2}_1,\dots,\overline{g^2}_M\right),$$

$$d_{j;n}(\mathbf{G},\mathbf{p}) = d_{j;n}(\mathbf{G}) = \operatorname{Var}(\xi_{j;n}) = \sum_{m=1}^{M} p^m_{j;n}\overline{g^2}_m - \sum_{m_1,m_2=1}^{M} p^{m_1}_{j;n}p^{m_2}_{j;n}\bar g_{m_1}\bar g_{m_2},$$

$$\lambda(\mathbf{G}) = (\lambda_1(\mathbf{G}),\dots,\lambda_M(\mathbf{G}))^T = \Gamma_n^{-1}(\mathbf{G})(1,-1,0,\dots,0)^T,$$

$$\Gamma_n(\mathbf{G}) = (\Gamma_{k,l;n}(\mathbf{G}))_{k,l=1}^M,\ \Gamma_{k,l;n}(\mathbf{G}) = \frac{1}{n}\sum_{j=1}^{n}\frac{p^k_{j;n}p^l_{j;n}}{d_{j;n}(\mathbf{G})},$$

$$\Gamma(\mathbf{G}) = \lim_{n\to\infty}\Gamma_n(\mathbf{G}).$$

It is clear that the existence of a solution requires the necessary condition $\det\Gamma_n(\mathbf{G}) \neq 0$. If $\det\Gamma(\mathbf{G}) \neq 0$, then $\det\Gamma_n(\mathbf{G}) \neq 0$ for sufficiently large n.

Theorem 6.1. *Let*

1. $\overline{g^2}_i < \infty$ *for all* $i = 1,\dots,M$*;*
2. *the limit in the definition of* $\Gamma(\mathbf{G})$ *exists and is finite,* $\det\Gamma(\mathbf{G}) \neq 0$*;*
3. $\langle(\mathbf{b}^{\rm opt})^2\mathbf{p}^{m_1}\mathbf{p}^{m_2}\rangle$ *exist for all* $m_1, m_2 = 1,\dots,M$*;*
4. $\sup_{j,n}|b^{\rm opt}_{j;n}| < \infty$.

Then the minimal value of $\beta(\mathbf{b})$ *is attained at* $\mathbf{b} = \mathbf{b}^{\rm opt}$*.*

Since $\mathbf{b}^{\rm opt}$ depends on unknown parameters of the model $\mathbf{G}$, these weight coefficients cannot be used explicitly for constructing a test. The adaptive approach suggests to estimate $\mathbf{G}$ by an estimator, say, by

$$\hat{\mathbf{G}} = \left(\hat g_{1;n},\dots,\hat g_{M;n},\widehat{g^2}_{1;n},\dots,\widehat{g^2}_{M;n}\right),$$

and then substitute this estimator in place of the true value in the expression for $\mathbf{G}^{\rm opt}$. As a result, we obtain estimators for the optimal weight coefficients $\hat{\mathbf{b}} = \mathbf{b}^{\rm opt}(\hat{\mathbf{G}})$. The adaptive test $\hat\pi(X_n)$ is based on the statistic

$$\hat T_n = \frac{S(\hat{\mathbf{b}})}{\sqrt{D_n(\hat{\mathbf{b}})}}.$$

The test $\hat\pi(X_n)$ accepts the zero hypothesis H_0 if $|\hat T_n| \le \lambda_{\alpha/2}$ and rejects it otherwise, that is, if $|\hat T_n| > \lambda_{\alpha/2}$.

Theorem 6.2. *Let all the assumptions of Theorem* 6.1 *hold. Assume additionally that* $\sigma^2_m = \overline{g^2_m} - (\overline{g_m})^2 > 0$ *for all* $m = 1,\dots,M$*. Then*

$$\beta(\hat{\mathbf{b}}) \stackrel{\rm def}{=} \lim_{n\to\infty}\mathsf{P}_{H_{1;n}}\{\hat\pi(X_n) = 0\} = \beta(\mathbf{b}^{\rm opt}(\mathbf{G})).$$

Therefore, the asymptotic probability of error of the second kind for the adaptive test is the same as that for the optimal nonadaptive test.

Remark. One can use not only weighted means with minimax coefficients to estimate $\mathbf{G}$. Instead, one can use other $\sqrt{n}$-consistent estimators. Since the minimax weight coefficients may be negative for some j, the corresponding weight estimators may have unpleasant properties if the size of a sample is small. In particular, it is quite possible that the estimator for $d_{j;n} = \operatorname{Var} \xi_{j;n}$ is negative in this case. For this reason, it is worthwhile to use improved estimators for moments described in [2] to construct estimators for the optimal weight coefficients.

7. Results of computer simulation

We compare the quality of test algorithms for samples of a finite size based on results of a computer simulation. We simulate two component mixtures with varying concentrations in every experiment. The concentrations of the first component $p^i_{j;n} = p_j$ are generated as independent uniformly distributed on the interval $[0, 1]$ random variables. Correspondingly, $p^2_{j;n} = 1 - p_j$. The following approach is followed when generating the distributions of the components of a mixture. We choose a basic distribution function F_0 with zero expectation and unit variance. Then the distribution function of a component m is given by $F_m(x) = F_0((x - \mu_m)/\sigma_m)$, where μ_m is the expectation and σ^2_m is the variance of the component m. To compare the probabilities of error of the first kind α we generate the data with the distribution that corresponds to the zero hypothesis, that is, to the case of $\mu_1 = \mu_2 = 0$. To compare the probabilities of error of the second kind, β, we generate the data with the distribution corresponding to the alternative $H_{1;n}\colon \mu_1 = v/\sqrt{n}$, $\mu_2 = 0$ (here n means the number of observations). Therefore, the distribution of the data in different experiments is different and is determined by the parameters F_0, σ^2_1, σ^2_2 and v.

In every experiment and for every size n, we generate 10,000 samples with a given distribution corresponding to the zero hypothesis H_0 and 10,000 samples with the distribution corresponding to the alternative $H_{1;n}$. For every sample, the hypothesis H_0 about the homogeneity of the means, $\mu_1 = \mu_2$, is checked by two different tests, namely by the naive test with weight coefficients $\mathbf{b}^{\text{naive}} = \mathbf{a}^1 - \mathbf{a}^2$ and adaptive test with coefficients $\hat{\mathbf{b}}$ proposed in the preceding section. Then the corresponding frequencies of errors of the first kind, α^{naive} and α^{adapt}, and those of the second kind, β^{naive} and β^{adapt}, are calculated for the naive and adaptive tests for different sample sizes n. These frequencies are shown in the tables. The lower row of every table contains asymptotic values of the corresponding probabilities $\alpha = 0.05$ (for errors of the first kind) and $\beta(\mathbf{b})$ (for errors of the second kind).

Finally, we use the improved moment estimators proposed in [2] to estimate unknown parameters $\mathbf{G}$.

Experiment 1. Here F_0 is the standard normal distribution with $\sigma^2_1 = 0.1$, $\sigma^2_2 = 1$, and $v = 10$. When choosing the parameters we take into account a relation between the probability of errors of the second kind for the adaptive and naive tests: the probability of errors of the second kind for the adaptive test is considerably smaller than those for the naive test if the variances of components are sufficiently different. So we could see if the adaptive test really performs better than the naive one in the case where the theory predicts so. The results of this experiment are shown in Table 1.

The results show that the approximation of the probability of error of the first kind by the asymptotic values is not precise but, nevertheless, is sufficiently good for practical needs if the sample size exceeds 500 for both tests. The asymptotic values provide more accurate approximations for the probabilities of errors of both kinds in the case of the naive test than in the case of the adaptive test. However, the differences between both cases are not essential. At the same time, the probability of error of the second kind for the adaptive test is smaller than that for the naive test for all sample sizes. For example, the adaptive test shows an almost double advantage for sample sizes exceeding $n = 500$ (compare with the asymptotic relation $\beta^{\text{naive}}/\beta^{\text{adapt}} = 2.795$).

Table 1. Results of computer simulation: Experiment 1

n	β^{naive}	β^{adapt}	α^{naive}	α^{adapt}
100	0.1715	0.1376	0.0412	0.0344
250	0.2047	0.1501	0.0458	0.039
500	0.1388	0.0776	0.0449	0.0414
1000	0.1325	0.0686	0.0461	0.0422
5000	0.1173	0.0492	0.0486	0.0422
10000	0.1052	0.0411	0.0477	0.047
∞	0.104869	0.037512	0.05	0.05

Table 2. Results of computer simulation: Experiment 2

n	β^{naive}	β^{adapt}	α^{naive}	α^{adapt}
100	0.1838	0.1878	0.0556	0.0562
250	0.1057	0.1088	0.055	0.0561
500	0.1283	0.1287	0.0474	0.0478
1000	0.0799	0.0809	0.0513	0.0512
5000	0.0807	0.0804	0.0495	0.0493
10000	0.0903	0.0912	0.0478	0.0479
∞	0.0790173	0.0790173	0.05	0.05

Table 3. Results of computer simulation: Experiment 3

n	β^{naive}	β^{adapt}	α^{naive}	α^{adapt}
100	0.1872	0.1413	0.0398	0.0369
250	0.1632	0.0931	0.0412	0.0386
500	0.1807	0.0939	0.0457	0.040944
1000	0.1315	0.0586	0.051	0.0435
5000	0.1375	0.0616	0.047	0.044
10000	0.1114	0.0402	0.0452	0.0414
∞	0.104869	0.037512	0.05	0.05

Experiment 2. Here F_0 is the standard normal distribution with $\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.5$, and $v = 10$. The variances of both components are chosen identical and thus the adaptive test has no advantages over the naive test. Since the adaptive test requires a more sophisticated estimation technique, one may expect that the asymptotic theory will exhibit the advantages of the naive test. The aim of the experiment is to check this idea.

The results of the experiment are shown in Table 2.

The results confirm that the asymptotic approximation is of a lower accuracy for the adaptive test than for the naive one. However the difference is not essential from the point of view of applications.

Experiment 3. Here F_0 is the centered and normalized χ^2 distribution with six degrees of freedom such that the expectation is zero, while the variance is unit. Also, $\sigma_1^2 = 0.1$, $\sigma_2^2 = 1$, and $v = 10$. The asymptotic results for this experiment are similar to those in Experiment 1. However, the distributions of components are not symmetric in Experiment 3 and have heavier right tails as compared to the standard normal distribution. The aim of the experiment is to check whether these features of the distribution influence the asymptotic values.

The results of the experiment are shown in Table 3.

As the results of the experiment show, the accuracy of asymptotic values is lower only for small sample sizes. The adaptive test has an advantage over the naive test for all sample sizes as far as the error of the second kind is concerned.

It is worth mentioning that the test show a conservative behavior in the sense that the empirical frequency of errors is smaller than the true probability in all cases, while the probability of error of the first kind for both tests differs sufficiently from the theoretical value 0.05 for small sample sizes.

8. Concluding remarks

The techniques of the adaptive estimation allow one to improve essentially the errors of the second kind in the test of homogeneity of means based on observations obtained from a mixture with varying concentrations. One can suggest to use the adaptive test if the sample size n is larger than 500, especially in the case of a big difference between the variances of components of a mixture. Even if the sample size is smaller than 500 and variances are identical, the adaptive test does not increase the probabilities of errors essentially as compared to the naive test. It is also true that an asymmetric distribution or moderate heavy tails do not lead to significantly worse results of the adaptive test.

Appendix

Proof of Lemma 4.1. By [4, Theorem 3.1.1], conditions 1 and 3 imply that $\widehat{g^2}_{m;n} \to \overline{g^2}_m$ and $\hat{g}_{m;n} \to \hat{g}_m$ in probability. Further, condition 2 yields $\langle(\mathbf{b})^2\mathbf{p}^m\rangle_n \to \langle(\mathbf{b})^2\mathbf{p}^m\rangle$ and $\langle(\mathbf{b})^2\mathbf{p}^{m_1}\mathbf{p}^{m_2}\rangle_n \to \langle(\mathbf{b})^2\mathbf{p}^{m_1}\mathbf{p}^{m_2}\rangle$. Thus,

$$\frac{D_n(\mathbf{b})}{\operatorname{Var} S_n(\mathbf{b})} = \frac{nD_n(\mathbf{b})}{n \operatorname{Var} S_n(\mathbf{b})} \to \frac{D_\infty(\mathbf{b})}{D_\infty(\mathbf{b})} = 1. \qquad \square$$

For further proofs, we need a special construction of the data under consideration.

Consider a family of jointly independent random variables,

$$\eta_j^m \quad \text{and} \quad \kappa_{j;n}, \qquad j = 1, 2, \ldots, \quad m = 1, \ldots, M,$$

whose distributions are $\eta_j^1 \sim F_{1,\infty}$, $\eta_j^m \sim F_m$, $m = 2, \ldots, M$, and $\mathsf{P}\{\kappa_{j;n} = m\} = p_{j;n}^m$. Put $\delta_{j;n}^m = \mathbb{1}\{\kappa_{j;n} = m\}$.

The random variables

$$\tilde{\xi}_{j;n} = \sum_{m=1}^{M} \delta_{j;n}^m \eta_j^m \tag{11}$$

have distribution (1) corresponding to zero hypothesis H_0, while the distribution of the data $X_n = (\xi_{1;1}, \ldots, \xi_{n;n})$, where

$$\xi_{j;n} = \tilde{\xi}_{j;n} + \delta_{j;n}^1 v_n, \tag{12}$$

corresponds to the local alternative $H_{1;n}$.

Therefore, we may assume that the data are given by (12) when calculating the probabilities related to the tests under consideration provided an alternative hypothesis holds. Then model (12) corresponds to the main hypothesis with $v_n = 0$. In what follows we use the structure of the data described above.

Proof of Theorem 5.1. Now we assume that the alternative hypothesis holds in contrast to the case of Lemma 4.1.

Put

$$\tilde{S}_n(\mathbf{b}) = \frac{1}{n} \sum_{j=1}^{n} b_{j;n} \tilde{\xi}_{j;n}.$$

By [4, Theorem 3.1.2],

$$\sqrt{n}\tilde{S}_n(\mathbf{b}) \Rightarrow N(0, D_\infty(\mathbf{b})).$$

Then

$$\sqrt{n}(S_n(\mathbf{b}) - \tilde{S}_n(\mathbf{b})) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \delta^1_{j;n} b_{j;n} \frac{v}{\sqrt{n}} = v\langle \delta^1 \mathbf{b}\rangle_n \to v\langle \mathbf{bp}^1\rangle$$

by the law of large numbers (see [1, Theorem 3, Section 3, Chapter 8]).

Hence,

$$\sqrt{n}S_n(\mathbf{b}) \Rightarrow N\left(\left\langle \mathbf{bp}^1\right\rangle, D_\infty(\mathbf{b})\right). \tag{13}$$

Since

$$\hat{g}_{m;n} - \frac{1}{n}\sum_{j=1}^{n} b_{j;n}\tilde{\xi}_{j;n} = v_n \frac{1}{n}\sum_{j=1}^{n} b_{j;n}\delta^1_{j;n} \to 0$$

in probability, we conclude that $\hat{g}_{m;n} \to \bar{g}_m$. Similarly, $\widehat{g^2}_{m;n} \to \overline{g^2}_m$ in probability. Therefore, $D_n(\mathbf{b})/D_\infty(\mathbf{b}) \to 1$ as $n \to \infty$. This together with (13) implies

$$T_n(\mathbf{b}) \Rightarrow N\left(\frac{\langle \mathbf{bp}^1\rangle}{\sqrt{D_\infty(\mathbf{b})}}, 1\right). \tag{14}$$

Since $\{\pi_{\mathbf{b}}(X_n) = 0\} = \{|T_n(\mathbf{b})| < \lambda_{\alpha/2}\}$, relation (14) completes the proof. □

Proof of Theorem 6.2. We show that

$$n\big(D_n(\hat{\mathbf{b}}) - D_n(\mathbf{b}^{\mathrm{opt}})\big) \to 0 \tag{15}$$

and

$$\Delta_n(\hat{\mathbf{G}}) \stackrel{\mathrm{def}}{=} \sqrt{n}\big(S_n(\mathbf{b}) - S_n(\hat{\mathbf{b}})\big) = \frac{1}{n}\sum_{j=1}^{n}\big(b^{\mathrm{opt}}_{j;n}(\mathbf{G}) - b^{\mathrm{opt}}_{j;n}(\hat{\mathbf{G}})\big)\xi_{j;n} \to 0 \tag{16}$$

in probability as $n \to \infty$. This yields the statement of the theorem.

Thus, we prove relation (15). Put $\sigma^2_{\min} = \min_m \sigma^2_m$. Then

$$\begin{aligned} d_{j;n}(\mathbf{G}) &= \operatorname{Var}(\xi_{j;n}) = \mathsf{E}\operatorname{Var}(\xi_{j;n} \mid \kappa_{j;n}) + \operatorname{Var}\mathsf{E}(\xi_{j;n} \mid \kappa_{j;n}) \\ &\geq \mathsf{E}\operatorname{Var}(\xi_{j;n} \mid \kappa_{j;n}) = \sum_{m=1}^{M} p^m_{j;n}\sigma^2_m \geq \sigma^2_{\min}. \end{aligned} \tag{17}$$

Moreover,

$$\begin{aligned} &\sup_{\mathbf{p}}\big|d_{j;n}(\mathbf{G},\mathbf{p}) - d_{j;n}(\hat{\mathbf{G}},\mathbf{p})\big| \\ &\quad\leq \sup_{\mathbf{p}}\left|\sum_{m=1}^{M} p^m_{j;n}\big(\overline{g^2_m} - \widehat{g^2}_{m;n}\big)\right| + \sup_{\mathbf{p}}\left|\sum_{m_1,m_2=1}^{M} p^{m_1}_{j;n}p^{m_2}_{j;n}(\bar{g}_{m_1}\bar{g}_{m_2} - \hat{g}_{m_1;n}\hat{g}_{m_2;n})\right| \\ &\quad\leq C\left(|\mathbf{G}|\cdot|\mathbf{G} - \hat{\mathbf{G}}| + |\mathbf{G} - \hat{\mathbf{G}}|^2\right), \end{aligned}$$

where C is a constant that does not depend on $\mathbf{G}$, $\hat{\mathbf{G}}$, and $\mathbf{p}$.

Under the assumptions of the theorem, $\varepsilon_n \stackrel{\mathrm{def}}{=} |\hat{\mathbf{G}} - \mathbf{G}| \to 0$ in probability as $n \to \infty$.

Next,

$$\begin{aligned} \left|\Gamma_{kl;n}(\mathbf{G}) - \Gamma_{kl;n}(\hat{\mathbf{G}})\right| &\leq \frac{1}{n}\sum_{j=1}^{n} \frac{p^k_{j;n}p^l_{j;n}|d_{j;n}(\mathbf{G}) - d_{j;n}(\hat{\mathbf{G}})|}{d_{j;n}(\mathbf{G})d_{j;n}(\hat{\mathbf{G}})} \\ &\leq \frac{1}{n}\sum_{j=1}^{n} \frac{C(|\mathbf{G}|\varepsilon_n + \varepsilon_n^2)}{\sigma^2_{\min}(\sigma^2_{\min} - C(|\mathbf{G}|\varepsilon_n + \varepsilon_n^2))} \to 0 \end{aligned}$$

as $n \to \infty$. Hence, $\Gamma(\hat{\mathbf{G}}) \to \Gamma(\mathbf{G})$ in probability. Taking into account $\det \Gamma(\mathbf{G}) \neq 0$, we obtain

$$\Gamma^{-1}(\hat{\mathbf{G}}) \to \Gamma^{-1}(\mathbf{G}) \quad \text{and} \quad \lambda(\hat{\Gamma}) \to \lambda(\Gamma)$$

in probability as $n \to \infty$. This implies relation (15).

Now we prove (16). By construction of $\mathbf{b}^{\mathrm{opt}}$,

$$\left\langle \mathbf{b}^{\mathrm{opt}}(\tilde{\mathbf{G}})\mathbf{p}^1 \right\rangle_n = -\left\langle \mathbf{b}^{\mathrm{opt}}(\tilde{\mathbf{G}})\mathbf{p}^2 \right\rangle_n, \qquad \left\langle \mathbf{b}^{\mathrm{opt}}(\tilde{\mathbf{G}})\mathbf{p}^m \right\rangle_n = 0, \qquad m = 3, \dots, M,$$

for all $\tilde{\mathbf{G}} = (\tilde{G}_1, \dots, \tilde{G}_{2M})$. Thus,

$$\mu(\tilde{\mathbf{G}}) = \mathsf{E}\sqrt{n}\Delta_n(\tilde{\mathbf{G}}) = v\big(\langle \mathbf{b}^{\mathrm{opt}}(\mathbf{G})\mathbf{p}^1 \rangle_n - \langle \mathbf{b}^{\mathrm{opt}}(\tilde{\mathbf{G}})\mathbf{p}^1 \rangle_n\big).$$

Since $\hat{\mathbf{G}} \to \mathbf{G}$ in probability as $n \to \infty$, we proceed as in the proof of relation (15) and obtain

$$\mu(\tilde{\mathbf{G}}) \to 0 \quad \text{in probability as } n \to \infty. \tag{18}$$

Put $\xi'_{j;n} = \xi_{j;n} - \mathsf{E}\,\xi_{j;n}$,

$$\tilde{\Delta}_n(\tilde{\mathbf{G}}) \stackrel{\mathrm{def}}{=} \Delta_n(\tilde{\mathbf{G}}) - \mu(\tilde{\mathbf{G}}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(b^{\mathrm{opt}}_{j;n}(\mathbf{G}) - b^{\mathrm{opt}}_{j;n}(\tilde{\mathbf{G}}) \right) \xi'_{j;n}.$$

Let $\varepsilon > 0$. Set $G^k = (G^k_1, \dots, G^k_{2M})$, $k = 0, 1$, where $G^0_i = G_i - \varepsilon$, $G^1_i = G_i + \varepsilon$,

$$K^\varepsilon = \bigotimes_{i=1}^{2M} \left[G^0_i, G^1_i\right].$$

In what follows we choose ε to be sufficiently small and n_0 to be sufficiently large in order to have the property that $\det \Gamma_n(\tilde{\mathbf{G}}) \neq 0$ for all $\tilde{\mathbf{G}} \in K^\varepsilon$ and all $n > n_0$.

Now we show that

$$P_\varepsilon(\lambda) = \sup_{n > n_0} \mathsf{P}\left\{ \sup_{\tilde{\mathbf{G}} \in K^\varepsilon} |\tilde{\Delta}_n(\tilde{\mathbf{G}})| > \lambda \right\} \to 0 \quad \text{as } \varepsilon \to 0 \tag{19}$$

for all $\lambda > 0$. To prove relation (19), we need a Sobolev type inequality which provides an upper bound for the uniform norm of a function expressed in terms of L_2-norms of its derivatives.

For all $\alpha = (\alpha_1, \dots, \alpha_{2M}) \in \{0,1\}^{2M}$ and an arbitrary $u = (u_1, \dots, u_{2M}) \in K^\varepsilon$, let

$$K^\varepsilon(\alpha) = \bigotimes_{i:\alpha_i = 1} \left[G^0_i, G^1_i\right],$$

let $u|_\alpha$ be a vector of $\mathbb{R}^{2M}$ whose ith coordinate equals G^0_i if $\alpha_i = 0$ or equals u_i if $\alpha_i = 1$, and

$$D^\alpha = \prod_{i:\ \alpha_i = 1} \frac{\partial}{\partial u_i}, \qquad (du)^\alpha = \prod_{i:\ \alpha_i = 1} du_i.$$

By [4, Lemma 7.1.1],

$$\sup_{u \in K^\varepsilon} |f(u)| \le f(\mathbf{G}^0) + \|f\|_H \bar{V}(\varepsilon) \tag{20}$$

for any $2M$ times differentiable function $f\colon K^\varepsilon \to \mathbb{R}$, where

$$\bar{V}(\varepsilon) = \left(\sum_{\alpha \in \{0,1\}^{2M}, \alpha \neq 0} V_\alpha(\varepsilon) \right)^{1/2},$$

$V_\alpha(\varepsilon) = \prod_{i:\alpha_i=1}(G_i^1 - G_i^0)$,

$$\|f\|_H = \left(\sum_{\alpha\in\{0,1\},\alpha\neq 0} \int_{K^\varepsilon(\alpha)} (D^\alpha f(u|_\alpha))^2 (du)^\alpha \right)^{1/2}.$$

Now we apply inequality (20) with $f(u) = \tilde{\Delta}(u)$ to prove relation (19).

It is clear that

$$J_1(\varepsilon) \overset{\text{def}}{=} \mathsf{E}\big(\tilde{\Delta}_n\left(G^0\right)\big)^2 = \operatorname{Var}\tilde{\Delta}_n(G^0)$$
$$= \frac{1}{n}\sum_{j=1}^n \big(b_{j;n}^{\text{opt}}(\mathbf{G}) - b_{j;n}^{\text{opt}}(\mathbf{G})^0\big)^2 \operatorname{Var}\xi_{j;n} \tag{21}$$
$$\le \sigma_{\max}^2 \sup_{n>n_0} \left(b_{j;n}^{\text{opt}}(\mathbf{G}) - b_{j;n}^{\text{opt}}\left(\mathbf{G}^0\right)\right)^2 \to 0$$

as $\varepsilon \to 0$.

Further, $\mathsf{E}\,\|\tilde{\Delta}_n\|_H^2 \le \sum_{\alpha\in 0,1^{2M},\alpha\neq 0} J_\alpha(\varepsilon)$, where

$$J_\alpha(\varepsilon) = \mathsf{E}\int_{K^\varepsilon(\alpha)} \big(D^\alpha \tilde{\Delta}_n(u|_\alpha)\big)^2 (du)^\alpha$$
$$= \frac{1}{n^2}\int_{K^\varepsilon(\alpha)} \mathsf{E}\left(\sum_{j=1}^n D^\alpha\big(b_{j;n}^{\text{opt}}(u|_\alpha)\xi'_{j;n}\big)\right)^2 (du)^\alpha \tag{22}$$
$$\le \max_{m=1,\dots,M} \overline{g_m^2} \sup_{j=1,\dots,n;n>n_0;u\in K^\varepsilon(\alpha)} \big(D^\alpha\big(b_{j;n}^{\text{opt}}(u)\big)\big)^2.$$

It is easy to see that

$$\sup_{j=1,\dots,n;n>n_0;u\in K^\varepsilon} D^\alpha\big(b_{j;n}^{\text{opt}}(u)\big)^2 < C \tag{23}$$

for sufficiently small ε_0 and $0 < \varepsilon < \varepsilon_0$, where C is a constant that does not depend on ε.

Taking into account (20)–(23) we obtain

$$P_\varepsilon(\lambda) \le \frac{2(J_1(\varepsilon) + C\bar{V}_\varepsilon)}{\lambda^2} \to 0$$

as $\varepsilon \to 0$ by the Chebyshev inequality. The latter result proves relation (19).

Note that

$$\mathsf{P}\big\{|\tilde{\Delta}_n(\hat{\mathbf{G}}_n)| > \lambda\big\} \le \mathsf{P}\{\hat{\mathbf{G}} \notin K^\varepsilon\} + \mathsf{P}\left\{\sup_{\tilde{\mathbf{G}}\in K^\varepsilon} \big|\tilde{\Delta}_n(\tilde{\mathbf{G}})\big| > \lambda\right\}.$$

In view of (19), the second term is as small as we wish by choosing a small enough ε. Fix ε. Then the first term approaches 0 as $n \to \infty$, since $\hat{\mathbf{G}}$ is a consistent estimator of $\mathbf{G}$. Hence, (18) implies (16).

Then

$$J_n(\varepsilon) \overset{\text{def}}{=} \mathsf{P}\left\{\left|\frac{S_n(\mathbf{b}^{\text{opt}})}{\sqrt{nD_n(\mathbf{b}^{\text{opt}})}} - \frac{S_n(\hat{\mathbf{b}})}{\sqrt{nD_n(\hat{\mathbf{b}})}}\right| \ge \varepsilon\right\} \to 0 \tag{24}$$

as $n \to \infty$ for all $\varepsilon > 0$ (this follows from (15) and (16)).

Since

$$\beta(\mathbf{b}^{\text{opt}}) = \lim_{n\to\infty} \mathsf{P}\left\{|S_n(\mathbf{b}^{\text{opt}})|/\sqrt{nD_n(\mathbf{b}^{\text{opt}})} < \lambda_{\alpha/2}\right\}$$

is a continuous function of $\lambda_{\alpha/2}$, we derive from (24) that

$$\beta(\hat{\mathbf{b}}) = \lim_{n\to\infty} \mathsf{P}\left\{\left|S_n(\hat{\mathbf{b}})\right|/\sqrt{nD_n(\hat{\mathbf{b}})} < \lambda_{\alpha/2}\right\} = \beta\left(\mathbf{b}^{\mathrm{opt}}\right).$$

The theorem is proved. □

Bibliography

1. A. A. Borovkov, *Probability Theory*, "Nauka", Moscow, 1986; English transl., Gordon and Breach Science Publishers, Amsterdam, 1998. MR1711261
2. O. O. Kubaĭchuk and R. E. Maĭboroda, *Improved estimates for moments based on observations from a mixture*, Teor. Ĭmovīr. Mat. Stat. **70** (2004), 74–81; English transl. in Theory Probab. Math. Statist. **70** (2005), 83–92. MR2109826
3. O. V. Doronin, *Adaptive estimation in a semiparametric mixture model*, Teor. Ĭmovīr. Mat. Stat. **91** (2014), 26–37; English transl. in Theory Probab. Math. Statist. **91** (2015), 29–41. MR3364121
4. R. E. Maĭboroda and O. V. Sugakova, *Estimation and Classification by Observations of a Mixture*, "Kyiv University", Kyiv, 2008. (Ukrainian)
5. A. Yu. Ryzhov, *A test of the hypothesis about the homogeneity of components of a mixture with varying concentrations by using censored data*, Teor. Ĭmovīr. Mat. Stat. **72** (2005), 129–139; English transl. in Theory Probab. Math. Statist. **72** (2006) 145–155.
6. A. Shcherbina, *Estimation of the mean value in a model of mixtures with varying concentrations*, Teor. Ĭmovīr. Mat. Stat. **84** (2011), 142–154; English transl. in Theory Probab. Math. Statist **84** (2012) 151–164. MR2857425
7. F. Autin and Ch. Pouet, *Test on the components of mixture densities*, Stat. Risk Model. **28** (2011), no. 4, 389–410. MR2877572
8. A. Doronin and R. Maiboroda, *Testing hypotheses on moments by observations from a mixture with varying concentrations*, Mod. Stoch. Theory Appl. **1** (2014), no. 2, 195–209. MR3316487
9. G. J. McLachlan and D. Peel, *Finite Mixture Models*, Wiley-Interscience, New York, 2000. MR1789474
10. R. Maiboroda and O. Sugakova, *Statistics of mixtures with varying concentrations with application to DNA microarray data analysis*, J. Nonparametr. Stat. **24** (2012), no. 1, 201–215. MR2885834
11. D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distribution*, Wiley, New York, 1985. MR838090

Department of Probability Theory, Statistics, and Actuarial Mathematics, Faculty for Mechanics and Mathematics, National Taras Shevchenko University, Academician Glushkov Avenue, 6, Kyiv 03127, Ukraine

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

Department of Higher Mathematics and Theoretical Radiophysics, Faculty for Radiophysics, National Taras Shevchenko University, Academician Glushkov Avenue, 6, Kyiv 03127, Ukraine

E-mail address: `sugak@univ.kiev.ua`

Translated by N. N. SEMENOV