Теорія Ймовір. та Матем. Статист.
Вип. 86, 2012

Theor. Probability and Math. Statist.
No. 86, 2013, Pages 205–217
S 0094-9000(2013)00899-0
Article electronically published on August 20, 2013

ESTIMATION OF THE PARAMETERS OF THE BINOMIAL DISTRIBUTION IN A MODEL OF MIXTURE

UDC 519.21

A. SHCHERBINA

Abstract. A model for observations sampled from a two component mixture is considered. Each object is associated with a certain numerical characteristic that may assume two values, namely zero (failure) or one (success) with identical probabilities for all objects in every class. Probabilities for these values are constant for all objects from the same component. The total numbers of objects of the first and second classes in groups as well as their characteristics are known. We study the problem of estimation of success probabilities for both components. We solve the problem by using the maximum likelihood method. We prove that the estimator is consistent and asymptotically normal. We apply the results obtained in the paper to a problem in genetics. An explicit form of the estimator and the asymptotic dispersion matrix is presented.

1. Introduction

The problem of estimation of parameters of a two component mixture model from observations belonging to different samples (groups) is considered in the current paper. We assume that the total number of objects belonging to each component (class) of the mixture is known for all groups. Various problems of this kind naturally occur when analyzing the data in sociology as well as in medicine and biology.

As an example, consider a survey whose aim is to study a dependence between students' performance assessments and cheating during quizzes or tests. Students often "help" one another by sharing answers during quizzes or tests, allowing others to look at their papers. This behavior is considered by themselves as a kind of cheating and thus the question on whether or not they follow this approach is sensitive. In order to establish an attitude concerning this kind of behavior, an anonymous survey is performed among the students in different groups (in different classes). As a result, the numbers of those students who are sharing the answers during quizzes or tests have become known in every group (the first component of the mixture) as well as the numbers of those students who do not follow this approach (the second component of the mixture). For every student, his/her performance assessment is also known. This characteristic could be binary in the simplest case, namely it equals 1 if a student passes the test and equals 0 otherwise. The main aim of the survey is to estimate the average performance assessment (in other words, the success probability) for each component of the mixture.

The statistical analysis of mixtures has a long history starting with the papers by Newcomb [9] and Pearson [10]. The reader may consult the book by McLachan and Pell [8] for a discussion of the current state of the art. The problem of estimating the characteristics of mixtures from several samples is studied in the book by Titterington, Smith,

2010 *Mathematics Subject Classification.* Primary 62F10, Secondary 62P10.
Key words and phrases. Estimation in a model of mixture, parametric estimation, genetic studies.

and Makov [12]. A generalization of the approach presented in [12], called the model of mixtures with varying concentrations, is considered by Maĭboroda and Sugakova [4].

A feature of the problem considered in the current paper is that the observable objects are taken from small populations (groups) without repetition and the total numbers of components in these groups are known and fixed. Therefore, there exists a dependence between the objects which allows one to estimate unknown parameters of the model more precisely.

We consider the problem of estimation of the success probability for components of a mixture by using the maximum likelihood method. We prove that the estimators are consistent and asymptotically normal. In contrast to the usual approach in estimating the parameters of mixtures with varying concentrations that ignores a dependence between the objects, our estimators are consistent even for the case where the concentrations of components in a mixture are constant for all groups of objects.

2. The setting of the problem

Let a sample consist of K groups of objects. The sizes of the groups are denoted by $N_1, \dots, N_K$. The objects are distributed among two classes; a group i contains N_{i1} objects of the first class and N_{i2} objects of the second class. The total number $N_{i1} + N_{i2}$ of objects in a group i is denoted by N_i. The pairs (N_{i1}, N_{i2}) are independent identically distributed random vectors with the distribution

$$G(n_1, n_2) = \mathsf{P}\left(N_{i1} = n_1, N_{i2} = n_2\right), \qquad n_1, n_2 \in \mathbb{N}_0.$$

Since the size of each group is positive, $G(0,0) = 0$. We assume that there are groups that contain objects of both classes. In other words, there are numbers $n_1 > 0$ and $n_2 > 0$ such that $G(n_1, n_2) > 0$.

Let C_{ij} be the number of the class that contains object j in a group i. These numbers are varying in a sample. If N_{i1} and N_{i2} are fixed, then the vector $(C_{i1}, \dots, C_{iN_i})$ is uniformly distributed, that is, $(C_{i1}, \dots, C_{iN_i})$ assumes all possible values with equal probabilities.

A certain numerical characteristic X is associated with every object. This characteristic may assume value zero (treated as a failure) or one (treated as a success). We denote the characteristics of objects in a group i by $X_{i1}, \dots, X_{iN_i}$. These characteristics are random variables whose distributions depend on a class containing a given object, namely

$$\mathsf{P}_q\left(X_i = 1 \mid C_i = m\right) = q_m \in [0,1], \qquad i = 1, \dots, N,\ m = 1, 2,$$

where q_m is the success probability for an object belonging to a class i. Put $q = (q_1, q_2)$.

The problem is to estimate the unknown parameter q from the data

$$\{N_{i1}, N_{i2}, X_{ij},\ i = 1, \dots, K,\ j = 1, \dots, N_i\}.$$

3. Probabilities related to the model

Consider a group i. Let

$$X_i = \sum_{j=1}^{N_i} X_{ij}.$$

We show that the statistic $S_i = (X_i, N_{i1}, N_{i2})$ is sufficient for the estimation of the parameter q by using the observations of a group i. This is clear in the case where the probability of the random event $\{X_{ij} = x_j, j = 1, \dots, N_i\}$ depends only on the sum $\sum_{j=1}^{N_i} x_i$.

Consider two sets of numbers, $x'_j \in \{0,1\}$ and $x''_j \in \{0,1\}$, $j = 1, \dots, N_i$, that sum up to the same number, namely

$$\sum_{j=1}^{N_i} x'_j = \sum_{j=1}^{N_i} x''_j.$$

Since the number of zeros in both sets as well as the number of units are equal, there exists a permutation σ such that $x''_j = x'_{\sigma(j)}$, $j = 1, \dots, N_i$. Thus,

$$\begin{aligned}\mathsf{P}_q\left(X_{ij} = x'_j, j = 1, \dots, N_i\right) &= \sum_{c_j \in \{1,2\}} \mathsf{P}_q\left(X_{ij} = x'_j, C_{ij} = c_j, j = 1, \dots, N_i\right) \\ &= \sum_{c_j \in \{1,2\}} \mathsf{P}_q\left(X_{ij} = x'_{\sigma(j)}, C_{ij} = c_{\sigma(j)}, j = 1, \dots, N_i\right) \\ &= \mathsf{P}_q\left(X_{ij} = x''_j, j = 1, \dots, N_i\right).\end{aligned}$$

Here we used the property that the vector $(C_{i1}, \dots, C_{iN_i})$ is uniformly distributed.

Let the unknown parameter be equal to $t = (t_1, t_2)$ and let $s = (x, n_1, n_2)$. We introduce the following notation:

$$\begin{aligned}f(s,t) = f(x, n_1, n_2, t) &= \mathsf{P}_t\left(X_i = x, N_{i1} = n_1, N_{i2} = n_2\right), \\ g(s,t) = g(x, n_1, n_2, t) &= \mathsf{P}_t\left(X_i = x \mid N_{i1} = n_1, N_{i2} = n_2\right).\end{aligned}$$

Thus,

$$\begin{aligned}g(s,t) &= \sum_{k=0}^{x} \mathsf{P}_t\left(\sum_{j=1}^{n} X_{ij}\, \mathbf{1}_{\{C_{ij}=1\}} = k\right) \mathsf{P}_t\left(\sum_{j=1}^{n} X_{ij}\, \mathbf{1}_{\{C_{ij}=2\}} = x - k\right) \\ &= \sum_{k=0}^{x} C_{n_1}^{k} t_1^k (1-t_1)^{n_1-k} C_{n_2}^{x-k} t_2^{x-k} (1-t_2)^{n_2-x+k}\, \mathbf{1}_{\{x-n_2 \le k \le n_1\}},\end{aligned}$$

where $n = n_1 + n_2$.

Also,

$$f(s,t) = \mathsf{P}_t\left(X = x \mid N_1 = n_1, N_2 = n_2\right) G(n_1, n_2) = g(s,t) G(n_1, n_2).$$

4. Maximum likelihood estimator

Let $S = (S_1, \dots, S_K)$. Then the logarithmic likelihood function is equal to

$$L(S,t) = \sum_{i=1}^{K} \ln f(S_i, t) = \sum_{i=1}^{K} l(S_i, t), \tag{1}$$

where $l(S_i, t) = \ln f(S_i, t)$.

Hence the maximum likelihood estimator is a value of the parameter t that maximizes the expression on the right-hand side of (1), namely

$$\hat{q} = \operatorname*{argmax}_{t \in [0,1]^2} L(S,t).$$

5. The consistency

Let q be the true value of the unknown parameter. We distinguish between the following two cases when studying the asymptotic behavior of the maximum likelihood estimator:

1) the sizes of the classes are the same for all groups, that is, $N_{11} = N_{12}$ almost surely;
2) there are groups with different sizes of the classes, that is, $\mathsf{P}_q\left(N_{11} \neq N_{12}\right) > 0$.

The problem is not identifiable in the first case, since the likelihood function is symmetric, that is,

$$L(S,(t_1,t_2)) = L(S,(t_2,t_1)).$$

This means that a solution can, in fact, be found up to a permutation of the coordinates. For definiteness, let $q_1 < q_2$. In this case, we search the maximum of the likelihood function among points $t = (t_1,t_2) \in [0,1]^2$ such that $t_1 \le t_2$:

$$\hat{q}^* = \operatorname*{argmax}_{0\le t_1\le t_2\le 1} L(S,t).$$

A solution of the problem is unique occasionally in the second case listed above.

Theorem 1. *Assume that the expectations of the sizes of groups are finite, that is,*

$$\mathsf{E}_q N_1 < \infty.$$

If $N_{11} = N_{12}$ almost surely, then the maximum likelihood estimator $\hat{q}^*$ is strongly consistent. If $\mathsf{P}_q(N_{11} \ne N_{12}) > 0$, then the maximum likelihood estimator $\hat{q}$ is strongly consistent.

In what follows we use the Kullback–Leibler distance between two distributions $f(s,t)$ and $f(s,q)$:

$$\rho(t) = \sum_s f(s,q) \ln \frac{f(s,q)}{f(s,t)} = \mathsf{E}_q l(S_1,q) - \mathsf{E}_q l(S_1,t).$$

The following proposition is needed for the proof.

Proposition 1 ([1, Theorem 16.2])**.** *Assume that*

(1) *the parametric set Θ is compact,*
(2) *the function $\rho(t)$ attains its minimal value at a unique point $t = q$,*
(3) *the function $f(s,t)$ is differentiable with respect to t, and*

$$\sum_s \ln \frac{f^\Theta(s)}{f(s,q)} f(s,q) < \infty, \qquad f^\Theta(s) = \sup_{t\in\Theta} f(s,t)$$

for all $t \in \Theta$.

Then the maximum likelihood estimator $\hat{q}$ is strongly consistent.

According to Proposition 1, we need to investigate the behavior of the function ρ in a vicinity of its minimum in order to prove the consistency of the maximum likelihood estimator. This can be done with the help of the following auxiliary result.

Lemma 1 ([1, Lemma 16.1])**.** *Let f and g be two probability densities with respect to the measure μ. Then*

$$\int f(x) \ln f(x)\, \mu(dx) \ge \int f(x) \ln g(x)\, \mu(dx)$$

if both integrals are finite. The inequality becomes an equality if and only if $f = g$ almost everywhere with respect to the measure μ.

In the case under consideration, μ is a counting measure in the set of points (x,n_1,n_2) such that $n_1,n_2 \ge 0$ and $0 \le x \le n_1+n_2$. Thus, the function ρ attains its minimum at the point q. If the function ρ has another point of minimum, say t, then

$$f(x,n_1,n_2,t) = f(x,n_1,n_2,q)$$

for all $n_1,n_2 \ge 0$ and $0 \le x \le n_1+n_2$.

The latter equality is meaningful if $G(n_1,n_2) > 0$. For such sample sizes n_1 and n_2, the above equality can be rewritten as follows:

$$g(x,n_1,n_2,t) = g(x,n_1,n_2,q). \tag{2}$$

For the proof of Theorem 1 we also need the following two results.

Lemma 2. *Let $N_{11} = N_{12}$ almost surely. Then the function ρ attains its minimum at the points (q_1, q_2) and (q_2, q_1).*

Lemma 3. *Let $\mathsf{P}_q(N_{11} = N_{12}) < 1$ almost surely. Then the function ρ attains its minimum at a unique point (q_1, q_2).*

Proof of Theorem 1. If $N_{11} = N_{12}$ almost surely, then we choose

$$\Theta = \{(q_1, q_2) \in [0,1]^2 \mid q_1 \le q_2\}.$$

By Lemma 2, the function $\rho(t)$ attains its minimum at points (q_1, q_2) and (q_2, q_1). Since $q_1 \le q_2$, the point of minimum of the function $\rho(t)$ is unique in the set Θ and is equal to q.

If $\mathsf{P}_q(N_{11} = N_{12}) < 1$, then we choose $\Theta = [0,1]^2$. By Lemma 3, the function $\rho(t)$ attains its minimum at a unique point q.

Thus, assumptions (1) and (2) of Proposition 1 are satisfied. It is also obvious that $f(s,t)$ is a differentiable function with respect to t if s is fixed. Using the inequality $x \ln x > -1$, we get

$$\sum_s \ln \frac{f^\Theta(s)}{f(s,q)} f(s,q) = \sum_s \ln \frac{\sup_{t\in\Theta} g(s,t)}{g(s,q)} f(s,q) \le -\sum_s f(s,q) \ln g(s,q)$$

$$= -\sum_{n_1,n_2\ge 0} G(n_1,n_2) \sum_{x=0}^{n_1+n_2} g(s,q) \ln g(s,q) \le \sum_{n_1,n_2\ge 0} G(n_1,n_2) n = \mathsf{E}_q N_1.$$

Therefore, all the assumptions of Proposition 1 are satisfied and thus the maximum likelihood estimator is strongly consistent. □

6. The asymptotic normality

Theorem 2. *Assume that the second moments of the sizes are finite for all groups, that is, $\mathsf{E}_q N_1^2 < \infty$, and that the true value of the parameter q belongs to the square $(0,1)^2$. Then*

1. *if the sizes of the classes coincide in all groups, that is, $N_{11} = N_{12}$ almost surely, and if $q_1 < q_2$, then the maximum likelihood estimator $\hat{q}^*$ is asymptotically normal;*
2. *if there are groups with different sizes of the classes, that is, $\mathsf{P}_q(N_{11} \ne N_{12}) > 0$, and if either $q_1 \ne q_2$ or there is no number $C > 0$ such that $N_{11} = CN_{12}$ almost surely, then the maximum likelihood estimator $\hat{q}$ is asymptotically normal.*

The proof of Theorem 2 uses the following assertion.

Proposition 2 ([1, Theorem 14.4])**.** *Assume that*

1. *the set of parameters Θ is compact;*
2. *the function $l(s,t)$ is twice continuously differentiable with respect to t,*

$$\sup_{t\in\Theta} \left| l''_{t_i t_j}(s,t) \right| < \gamma(s), \qquad i,j = 1,2,\ \mathsf{E}_q \gamma(S_1) < \infty;$$

3. *the matrix $I = \mathsf{E}_q \frac{\partial^2}{\partial t^2} l(S_1, q)$ exists and is such that $\det I \ne 0$;*
4. *the equation $\mathsf{E}_q \frac{\partial}{\partial t} l(S_1, t) = 0$ has a unique root $t = q$.*

Then the maximum likelihood estimator $\hat{q}$ is asymptotically normal, that is,

$$\sqrt{K}(\hat{q}_K - q) \to \mathcal{N}(0, I^{-1}).$$

Proof of Theorem 2. Consider the same set of parameters Θ as in the proof of Theorem 1. Then the maximum likelihood estimator is consistent and thus the assumptions of Proposition 2 need to be checked in a certain vicinity of the true value q.

In the case under consideration, the function $l(s,t)$ is twice continuously differentiable. Fix an arbitrary $\varepsilon > 0$ such that $q \in (\varepsilon, 1-\varepsilon)^2$. We check assumption (2) of Proposition 2 for $t \in [\varepsilon, 1-\varepsilon]^2$. Write

$$l''_{t_it_j}(s,t) = \frac{f''_{t_it_j}(s,t)}{f(s,t)} - \frac{f'_{t_i}(s,t)}{f(s,t)}\frac{f'_{t_j}(s,t)}{f(s,t)} = \frac{g''_{t_it_j}(s,t)}{g(s,t)} - \frac{f'_{t_i}(s,t)}{g(s,t)}\frac{f'_{t_j}(s,t)}{g(s,t)}$$

and estimate the ratios on the right-hand side:

$$\left|\frac{g'_{t_1}(s,t)}{g(s,t)}\right| \le \frac{\sum_{k=0}^{x}\left|\frac{k}{t_1} - \frac{n_1-k}{1-t_1}\right| C_{n_1}^k t_1^k (1-t_1)^{n_1-k} C_{n_2}^{x-k} t_2^{x-k}(1-t_2)^{n_2-x+k}\,\mathbf{1}_{\{x-n_2\le k\le n_1\}}}{\sum_{k=0}^{x} C_{n_1}^k t_1^k (1-t_1)^{n_1-k} C_{n_2}^{x-k} t_2^{x-k}(1-t_2)^{n_2-x+k}\,\mathbf{1}_{\{x-n_2\le k\le n_1\}}} \le \frac{2n}{\varepsilon}.$$

Similarly, we obtain

$$\left|\frac{g'_{t_2}(s,t)}{g(s,t)}\right| \le \frac{2n}{\varepsilon}, \qquad \left|\frac{g''_{t_it_j}(s,t)}{g(s,t)}\right| \le \frac{4n^2}{\varepsilon^2}, \quad i,j = 1,2.$$

Hence,

$$\left|l''_{t_it_j}(s,t)\right| \le \frac{4n^2}{\varepsilon^2} + \frac{2n}{\varepsilon}\frac{2n}{\varepsilon} = \frac{8n^2}{\varepsilon^2}.$$

Since the second moment of the size of a group is finite, assumption (2) of Proposition 2 follows.

Consider the matrix $I = \mathsf{E}_q \frac{\partial^2}{\partial t^2} l(S_1, q)$. We are going to show that I is a nonsingular matrix under the assumptions of Theorem 2. We expand the entry (i,j) of the matrix I as follows:

$$\begin{aligned} I_{ij} &= \mathsf{E}_q\left[\frac{f''_{t_it_j}(S_1,q)}{f(S_1,q)} - \frac{f'_{t_i}(S_1,q)f'_{t_j}(S_1,q)}{f^2(S_1,q)}\right] = \sum_s\left[f''_{t_it_j}(s,q) - \frac{f'_{t_i}(s,q)f'_{t_j}(s,q)}{f(s,q)}\right] \\ &= \frac{\partial^2}{\partial t_i \partial t_j}\sum_s f(s,q) - \sum_s \frac{f'_{t_i}(s,q)f'_{t_j}(s,q)}{f(s,q)} = -\sum_s \frac{f'_{t_i}(s,q)f'_{t_j}(s,q)}{f(s,q)}. \end{aligned} \tag{3}$$

The determinant of the matrix I is equal to $I_{11}I_{22} - I_{12}^2$ and is always nonnegative, since

$$\left(\sum_s \frac{g'_{t_1}(s,q)g'_{t_2}(s,q)}{g(s,q)}\right)^2 \le \left(\sum_s \frac{g'_{t_1}(s,q)g'_{t_1}(s,q)}{g(s,q)}\right)\left(\sum_s \frac{g'_{t_2}(s,q)g'_{t_2}(s,q)}{g(s,q)}\right)$$

by the Cauchy–Bunyakovskiĭ inequality. The latter inequality becomes an equality if and only if the functions

$$\frac{\partial}{\partial t_1} g(s,q) \quad \text{and} \quad \frac{\partial}{\partial t_2} g(s,q)$$

are proportional. Now we prove that they are in fact not proportional under the assumptions of Theorem 2.

Choose some integer numbers $n_1, n_2 > 0$ such that $G(n_1, n_2) > 0$. Consider the ratio of the functions $\frac{\partial}{\partial t_1} g(s,q)$ and $\frac{\partial}{\partial t_2} g(s,q)$ for $s' = (0, n_1, n_2)$ and $s'' = (n_1 + n_2, n_1, n_2)$:

$$\frac{g'_{t_1}(s',q)}{g'_{t_2}(s',q)} = \frac{-n_1(1-q_1)^{n_1-1}(1-q_2)^{n_2}}{-n_2(1-q_1)^{n_1}(1-q_2)^{n_2-1}} = \frac{n_1(1-q_2)}{n_2(1-q_1)},$$

$$\frac{g'_{t_1}(s'',q)}{g'_{t_2}(s'',q)} = \frac{-n_1 q_1^{n_1-1} q_2^{n_2}}{-n_2 q_1^{n_1} q_2^{n_2-1}} = \frac{n_1 q_2}{n_2 q_1}.$$

Thus, the functions $\frac{\partial}{\partial t_1} g(s,q)$ and $\frac{\partial}{\partial t_2} g(s,q)$ are not proportional if $q_1 \neq q_2$. Otherwise, if $q_1 = q_2$, then we show that $\frac{\partial}{\partial t_1} g(s,q)$ and $\frac{\partial}{\partial t_2} g(s,q)$ are proportional if and only if there exists a number $C > 0$ such that $N_1 = CN_2$ almost surely.

Consider the mapping $s(t) = \mathsf{E}_q \frac{\partial}{\partial t} l(S_1, t)$; it equals zero at the point q:

$$s(q) = \mathsf{E}_q \frac{\partial}{\partial t} l(S_1, q) = \sum_s f(s,q) \frac{\frac{\partial}{\partial t} f(s,q)}{f(s,q)} = \frac{\partial}{\partial t} \sum_s f(s,q) = \frac{\partial}{\partial t} 1 = 0.$$

Since I is a nonsingular matrix, a solution of the equation $s(t) = 0$ is unique in a certain vicinity of the point q.

To check assumption (4) of Proposition 2 we write

$$\operatorname{Cov}_q \frac{\partial}{\partial t} l(S_1, q) = \mathsf{E}_q \frac{\partial}{\partial t} l(S_1, q) \left(\frac{\partial}{\partial t} l(S_1, q) \right)^T - s(q) s(q)^T = -I.$$

Thus, all the assumptions of Proposition 2 are satisfied and thus the maximum likelihood estimator is asymptotically normal. □

7. An application

Consider a particular case, where each group consists of two objects, one of them belongs to the first class and the other one belongs to the second class, that is, $G(1,1) = 1$.

Such a model is useful for modelling the phenomenon called genomic imprinting. By this genetic phenomenon, certain genes are expressed in a manner that depends on their parent origins: imprinted alleles are silenced such that the genes are either expressed only from the nonimprinted allele inherited from the mother or, in other instances, from the nonimprinted allele inherited from the father.

Let an expression of a certain gene be studied for a homozygote. Assume that the phenotype of the homozygote allows one to guess for a given (say, i-th) organism whether the expression of a gene is processed from a single chromosome ($X_i = 1$), or from both chromosomes ($X_i = 2$), or from none of chromosomes ($X_i = 0$). Each organism can be viewed as a separate group containing two objects (chromosomes): one of them belongs to the first component (inherited from the father), while the other one belongs to the second component (inherited from the mother).

Let $X_{ij} = 1$ if the expression of a gene occurs for a j-th chromosome of an i-th organism, and let $X_{ij} = 0$ otherwise. Then $X_i = X_{i1} + X_{i2}$. Despite the values of X_{i1} and X_{i2} are not observed, X_i is a sufficient statistic for the construction of the estimators of the probabilities q_k (q_k is the probability of the event that the expression of a gene belonging to a k-th component occurs).

We study this problem with the help of the maximum likelihood method. We introduce the following notation:

$$f_i(t) = f(i, 1, 1, t) = g(i, 1, 1, t), \qquad i = 0, 1, 2.$$

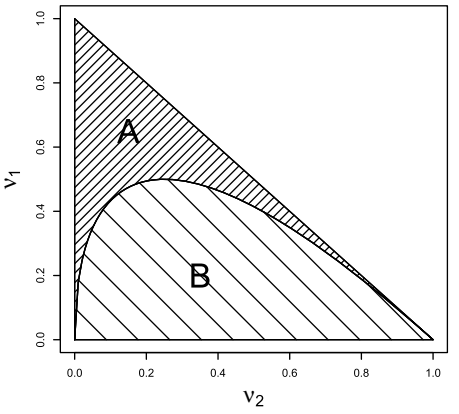

FIGURE 1. The set of values of the vector (ν_1, ν_2) for which equations (4) and (5) have a solution

Then the likelihood function (1) can be rewritten as

$$L(S,t) = \sum_{i=1}^{K} \ln f_{X_i}(t) = K \sum_{l=0}^{2} \nu_l \ln f_l(t),$$

where ν_l is the frequency of the value l in the sample $X_1, \ldots, X_K$.

Since ν_l and $f_l(t)$ are probability distributions in the set $\{0,1,2\}$, Lemma 1 implies that the maximum is attained if $\nu_l = f_l(t)$. For $l = 1, 2$ we obtain the following equations:

$$\nu_1 = f_1(t) = t_1(1-t_2) + (1-t_1)t_2, \tag{4}$$

$$\nu_2 = f_2(t) = t_1 t_2. \tag{5}$$

We express t_2 via t_1 by using equation (4). Then we substitute this expression with equation (5) which transforms it to the following form:

$$t_1^2 - (\nu_1 + 2\nu_2)t_1 + \nu_2 = 0.$$

The latter quadratic equation has the solutions $t = \mu/2 \pm \sqrt{\mu^2/4 - \nu_2}$ if $\nu_2 \le \mu^2/4$, where μ denotes the sampling mean, that is,

$$\mu = \frac{1}{K} \sum_{i=1}^{K} x_i = \nu_1 + 2\nu_2.$$

Thus, we obtain the following estimator:

$$\hat{q}^* = \left(\frac{\mu}{2} - \sqrt{\frac{\mu^2}{4} - \nu_2}, \frac{\mu}{2} + \sqrt{\frac{\mu^2}{4} - \nu_2} \right).$$

Introduce the following two sets:

$$A = \left\{ (\nu_1, \nu_2) \mid 0 \le \nu_1, \nu_2 \le 1, \nu_1 + \nu_2 \le 1, \nu_2 \le (\nu_1 + 2\nu_2)^2/4 \right\},$$
$$B = \left\{ (\nu_1, \nu_2) \mid 0 \le \nu_1, \nu_2 \le 1, \nu_1 + \nu_2 \le 1, \nu_2 > (\nu_1 + 2\nu_2)^2/4 \right\}.$$

The sets A and B are depicted in Figure 1. Since equations (4) and (5) have a solution if and only if $(\nu_1, \nu_2) \in A$, we conclude that

$$A = \left\{ (f_1(t), f_2(t)) \mid t \in [0,1]^2 \right\}. \tag{6}$$

Next we consider the case of $(\nu_1, \nu_2) \in B$. We rewrite the likelihood function (1) as a function of the arguments ν_1, ν_2, $f_1(t)$, and $f_2(t)$:

$$L(S,t) = K\left[\nu_0(1 - f_1(t) - f_2(t)) + \nu_1 \ln f_1(t) + \nu_2 \ln f_2(t)\right].$$

Taking into account equality (6), we rewrite the problem of minimization of the likelihood function as follows:

$$\begin{cases} F(p_1,p_2) := \nu_0 \ln(1 - p_1 - p_2) + \nu_1 \ln p_1 + \nu_2 \ln p_2 \to \max, \\ (p_1,p_2) \in A. \end{cases}$$

Now we show that the function $F(p_1,p_2)$ is concave with respect to p_1 and p_2. The matrix of its partial derivatives of the second order is given by

$$\frac{\partial^2 F(p_1,p_2)}{\partial p_1 \partial p_2} = \begin{pmatrix} -\dfrac{\nu_0}{(1-p_1-p_2)^2} - \dfrac{\nu_1}{p_1^2} & -\dfrac{\nu_0}{(1-p_1-p_2)^2} \\ -\dfrac{\nu_0}{(1-p_1-p_2)^2} & -\dfrac{\nu_0}{(1-p_1-p_2)^2} - \dfrac{\nu_2}{p_2^2} \end{pmatrix}.$$

Since the first principal minor is negative and the determinant is positive, the matrix is negative semi-definite by Sylvester's criterion. Thus the maximum of the function $F(p_1,p_2)$ is attained at the curvilinear boundary of the set A if $(\nu_1,\nu_2) \in B$. One can easily check that $(p_1,p_2) = (f_1(s,s), f_2(s,s))$ for some $s \in [0,1]$, where (p_1,p_2) is an arbitrary point of this boundary. Therefore, the function $L(S,t)$ attains its maximum if $t_1 = t_2$ and we obtain the following problem:

$$\begin{cases} G(s) = \nu_0 \ln(1-s)^2 + \nu_1 \ln 2s(1-s) + \nu_2 \ln s^2 \to \max, \\ s \in [0,1]. \end{cases}$$

The derivative of the function $G(s)$ is equal to

$$G'(s) = -\frac{2\nu_0}{1-s} + \frac{\nu_1}{s} - \frac{\nu_1}{1-s} + \frac{2\nu_2}{s}.$$

It can easily be proved that $s = \mu/2$ is a unique root of the equation $G'(s) = 0$. Hence the maximum likelihood estimator equals $\hat{q}^* = (\mu/2, \mu/2)$ if $\nu_2 > \mu^2/4$.

Theorem 1 implies that the estimator $\hat{q}^*$ is consistent if $q_1 \le q_2$. Theorem 2 implies that this estimator is asymptotically normal if $0 < q_1 < q_2 < 1$. According to equality (3) the information matrix is equal to

$$I = \frac{1}{q_1 + q_2 - 2q_1q_2} \begin{pmatrix} \dfrac{q_1 + q_2^2 - 2q_1q_2}{q_1 - q_1^2} & 1 \\ 1 & \dfrac{q_1^2 + q_2 - 2q_1q_2}{q_2 - q_2^2} \end{pmatrix}.$$

8. Concluding remarks

We investigate the problem of estimation for the two component model. The maximum likelihood estimator is constructed and studied for this model. For a particular case, we provide an explicit form of the estimator and asymptotic dispersion matrix.

In a forthcoming paper, we plan to apply the technique developed in this paper to the case of mixtures with a general number of components. If $\{X_{ij}, j = 1, \dots, N_i\}$ are observations with an arbitrary distribution, then one can consider the data

$$\{Y_{ij} = \mathbf{1}_{\{X_{ij} < x\}},\ j = 1, \dots, N_i\}$$

for an arbitrary real number x. For the latter data $\{Y_{ij}\}$, one can apply the maximum likelihood method. The maximum likelihood estimators are the values of distribution functions at the point x. This means that one can estimate the distribution function of the components of a mixture and hence every characteristic of the components.

9. Proof of the lemmas

Proof Lemma 2. It is easy to prove that

$$g(x, n_1, n_1, (q_1, q_2)) = g(x, n_1, n_1, (q_2, q_1))$$

for all $n_1 > 0$ and $0 \le x \le 2n_1$. Thus, the pair (q_2, q_1) also minimizes the function ρ.

Our aim is to show that the function ρ attains its minimum only at these points.

Consider the case of $q \in (0,1)^2$. Assume that the function ρ attains its minimum at the point $t = (t_1, t_2)$. The assumption of Lemma 2 implies that there exists a positive integer number n such that $G(n, n) > 0$. We apply condition (2) to the triples $(0, n, n)$ and $(2n, n, n)$:

$$\begin{gathered}(1-t_1)^n(1-t_2)^n = (1-q_1)^n(1-q_2)^n,\\ t_1^n t_2^n = q_1^n q_2^n,\end{gathered}$$

whence

$$\begin{gathered}1 - t_1 - t_2 + t_1 t_2 = 1 - q_1 - q_2 + q_1 q_2,\\ t_1 t_2 = q_1 q_2.\end{gathered}$$

This means that

$$\begin{gathered}t_1 + t_2 = q_1 + q_2,\\ t_1 t_2 = q_1 q_2.\end{gathered}$$

This system of two equations possesses two solutions (q_1, q_2) and (q_2, q_1) by Viet's formulas.

Now we consider the case where the parameter q belongs to the boundary of the square $[0,1]^2$. Let, for example, $q_1 = 0$. Then equalities (2) for the triples $(0, n, n)$ and $(2n, n, n)$ become of the following form:

$$\begin{gathered}(1-t_1)^n(1-t_2)^n = (1-q_2)^n,\\ t_1^n t_2^n = 0.\end{gathered}$$

Now we conclude that both pairs $(0, q_2)$ and $(q_2, 0)$ are solutions.

The proof for other cases is similar. □

Proof of Lemma 3. Assume that there is a second point $t = (t_1, t_2)$ where the function ρ attains its minimum. Consider the case of $q \in (0,1)^2$.

Below are particular cases of condition (2) corresponding to n_1 and n_2 involved in the statement of Theorem 1 and corresponding to $x = 0$, 1, $n_1 + n_2 - 1$, and $n_1 + n_2$, respectively:

$$(1-t_1)^{n_1}(1-t_2)^{n_2} = (1-q_1)^{n_1}(1-q_2)^{n_2}, \tag{7}$$

$$\begin{aligned}n_1 t_1 (1-t_1)^{n_1-1}(1-t_2)^{n_2} + n_2(1-t_1)^{n_1} t_2 (1-t_2)^{n_2-1}\\ = n_1 q_1 (1-q_1)^{n_1-1}(1-q_2)^{n_2} + n_2(1-q_1)^{n_1} q_2 (1-q_2)^{n_2-1},\end{aligned} \tag{8}$$

$$\begin{aligned}n_1 t_1^{n_1-1}(1-t_1)t_2^{n_2} + n_2 t_1^{n_1} t_2^{n_2-1}(1-t_2)\\ = n_1 q_1^{n_1-1}(1-q_1)q_2^{n_2} + n_2 q_1^{n_1} q_2^{n_2-1}(1-q_2),\end{aligned} \tag{9}$$

$$t_1^{n_1} t_2^{n_2} = q_1^{n_1} q_2^{n_2}. \tag{10}$$

Dividing equality (8) by (7) and equality (9) by (10) we obtain

$$n_1\frac{t_1}{1-t_1}+n_2\frac{t_2}{1-t_2}=n_1\frac{q_1}{1-q_1}+n_2\frac{q_2}{1-q_2},$$
$$n_1\frac{1-t_1}{t_1}+n_2\frac{1-t_2}{t_2}=n_1\frac{1-q_1}{q_1}+n_2\frac{1-q_2}{q_2}.$$

Let

$$\alpha=\frac{n_1}{n_2},\qquad p_i=\frac{1-t_i}{t_i},\quad r_i=\frac{1-q_i}{q_i},\quad i=1,2.$$

Then $\alpha\neq 1$ and

$$\frac{\alpha}{p_1}+\frac{1}{p_2}=\frac{\alpha}{r_1}+\frac{1}{r_2},$$
$$\alpha p_1+p_2=\alpha r_1+r_2. \tag{11}$$

Multiplying the latter two equalities we get

$$\alpha^2+\alpha\left(\frac{p_1}{p_2}+\frac{p_2}{p_1}\right)+1=\alpha^2+\alpha\left(\frac{r_1}{r_2}+\frac{r_2}{r_1}\right)+1$$

or

$$\frac{p_1}{p_2}+\frac{p_2}{p_1}=\frac{r_1}{r_2}+\frac{r_2}{r_1}.$$

If $r_1=r_2$, then we have a unique solution $p_1=p_2$. It also follows from (11) that $p_1=p_2=r_1$. Thus, the function ρ possesses a unique point of maximum $t=q$.

Let $r_1\neq r_2$. One can show that the latter equation considered with respect to the unknown p_1/p_2 has two roots, namely r_1/r_2 and r_2/r_1. Hence $p_2=p_1r_2/r_1$, whence $p_2=p_1r_1/r_2$. Dividing (7) be (10) we conclude that

$$p_1^{\alpha}p_2=r_1^{\alpha}r_2.$$

Substituting p_1r_2/r_1 instead of p_2 in the latter equation, we obtain $p_1=r_1$ or $p_1=r_1^{\frac{\alpha-1}{\alpha+1}}r_2^{\frac{2}{\alpha+1}}$. In the first case, we have $p_2=r_2$, whence $t=q$.

In the second case, we get $p_2=r_1^{\frac{2\alpha}{\alpha+1}}r_2^{\frac{1-\alpha}{\alpha+1}}$. Now our aim is to show that this is not a solution. Substituting $r_1^{\frac{2\alpha}{\alpha+1}}r_2^{\frac{1-\alpha}{\alpha+1}}$ instead of p_2 in the first equation in (11), we obtain

$$\alpha r_1^{\frac{\alpha-1}{\alpha+1}}r_2^{\frac{2}{\alpha+1}}+r_1^{\frac{2\alpha}{\alpha+1}}r_2^{\frac{1-\alpha}{\alpha+1}}=\alpha r_1+r_2.$$

Letting $v=(r_1/r_2)^{1/(\alpha+1)}$, we get $v\neq 1$. Dividing the preceding equality by r_2 yields

$$\alpha v^{\alpha-1}+v^{2\alpha}=\alpha v^{\alpha+1}+1$$

or

$$\frac{v^{\alpha}-v^{-\alpha}}{\alpha}=v-v^{-1}. \tag{12}$$

Consider the function

$$F(v,\alpha)=\left(v^{\alpha}-v^{-\alpha}\right)/\alpha-(v-v^{-1}).$$

It is obvious that $F(1,\alpha)=0$. Differentiating the defining equality for $F(v,\alpha)$ with respect to the first argument, we obtain

$$F'_v(w,\alpha)=\frac{w^{\alpha}+w^{-\alpha}-w-w^{-1}}{w}=0$$

for some number $w\neq 1$.

Also, $F'_v(w,1)=0$. Hence there exists a number $\beta\neq 1$ such that

$$F''_{v\alpha}(w,\beta)=\ln w\frac{w^{\beta}-w^{-\beta}}{w}=0.$$

However, this is impossible and thus the function ρ has a unique point of maximum $t = q$.

Now we consider the case where the parameter q belongs to the boundary of the square $[0,1]^2$. Let, for example, $q_1 = 0$ and $q_2 > 0$. Then equalities (7), (8), and (10) become of the following form:

$$(1-t_1)^{n_1}(1-t_2)^{n_2} = (1-q_2)^{n_2},$$
$$n_1 t_1 (1-t_1)^{n_1-1}(1-t_2)^{n_2} + n_2(1-t_1)^{n_1} t_2 (1-t_2)^{n_2-1} = n_2 q_2 (1-q_2)^{n_2-1}.$$

If $t_1 = 0$, then the first equality implies that $t_2 = q_2$. Now we show that the equality $t_2 = 0$ is not possible. The preceding equalities imply

$$(1-t_1)^{n_1} = (1-q_2)^{n_2},$$
$$n_1 t_1 (1-t_1)^{n_1-1} = n_2 q_2 (1-q_2)^{n_2-1}. \tag{13}$$

Letting $\alpha = n_1/n_2$ and dividing the second equality by the first one we obtain

$$(1-t_1)^{\alpha} = 1-q_2,$$
$$\alpha \frac{t_1}{1-t_1} = \frac{q_2}{1-q_2}.$$

Excluding q_2 from both equalities, we get

$$1-(1-t_1)^{\alpha} = \alpha t_1 (1-t_1)^{\alpha-1}.$$

Consider the function $F(t_1,\alpha) = (1-t_1)^{\alpha} + \alpha t_1 (1-t_1)^{\alpha-1}$. Then $F(0,\alpha) = 1$ and its derivative with respect to t_1 is equal to

$$\begin{aligned} F'_{t_1}(v,\alpha) &= -\alpha(1-t_1)^{\alpha-1} + \alpha(1-t_1)^{\alpha-1} - \alpha(\alpha-1)t_1(1-t_1)^{\alpha-2} \\ &= -\alpha(\alpha-1)t_1(1-t_1)^{\alpha-2}. \end{aligned}$$

The derivative does not change the sign for $t_1 \in (0,1]$, thus the equality $F(t_1,\alpha) = 1$ is only possible if $t_1 = 0$. On the other hand, equality (13) does not hold in this case.

Other cases are considered similarly. □

Bibliography

1. A. A. Borovkov, *Mathematical Statistics*, "Nauka", Novosibirsk, 1997; English transl., Gordon and Breach, New York, 1998. MR1712750 (2000f:62003)
2. R. E. Maĭboroda, *Estimation of the distributions of the components of mixtures having varying concentrations*, Ukr. Matem. Zh. **48** (1996), no. 4, 562–566; English transl. in Ukrainian Math. J. **48** (1996), no. 4, 618–622. MR1417019 (97j:62055)
3. R. E. Maĭboroda, *Statistical Analysis of Mixtures*, Kyiv University Press, Kyiv, 2003. (Ukrainian).
4. R. E. Maĭboroda and O. V. Sugakova, *An estimation and classification by observations from a mixture*, Kyiv University Press, Kyiv, 2008 (Ukrainian).
5. A. M. Shscherbina, *Estimation of the mean value in a model of mixtures with varying concentrations*, Teor. Imovir. Matem. Statyst. **84** (2011), 142–154; English transl. in Theor. Probability and Math. Statist. **84** (2012), 151–164.
6. A. M. Shscherbina, *A comparison of estimators of the mean values for mixtures with varying concentrations by using the generated data*, Visnyk Kyiv Nats. Univer. Ser. Mathematics. Mechanics **25** (2011), 43–47 (Ukrainian).
7. O. O. Kubaychuk, *Estimation of moments by observations from mixtures with varying concentrations*, Theory Stoch. Process. **8(24)** (2002), no. 3–4, 226–232. MR2027394 (2005g:62065)
8. G. J. McLachan and D. Pell, *Finite Mixture Models*, Wiley, NY, 2000. MR1789474 (2002b:62025)
9. S. Newcomb, *A generalized theory of the combination of observations so as to obtain the best result*, Amer. J. Math. **8** (1886), no. 4, 343–366. MR1505430
10. K. Pearson, *Contribution to the mathematical theory of evolution*, Trans. Roy. Soc. A. **185** (1894), 71–110.

11. A. Shcherbina and R. Maĭboroda, *Merging data from anonymous and open surveys: two-population problems*, Proceedings of the Baltic–Nordic–Ukrainian Summer School on Survey Statistics, "TViMS", Kyiv, 2009.
12. D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York, 1985. MR838090 (87j:62033)

Department of Probability Theory, Statistics, and Actuarial Mathematics, Faculty for Mechanics and Mathematics, National Taras Shevchenko University, Academician Glushkov Avenue, 4E, Kiev 03127, Ukraine

E-mail address: artshcherbina@gmail.com

Received 20/MAY/2011
Translated by S. V. KVASKO