DEMONSTRATIO MATHEMATICA
Vol. IX No 4 1976

Krystyna Królikowska

# ESTIMATION OF THE PARAMETERS OF ANY FINITE MIXTURE OF GEOMETRIC DISTRIBUTIONS

The estimation of parameters for a mixture of two distributions has been the subject of numerous papers originated by the paper of K. Pearson [1] dealing with mixtures of normal distributions. The method of moments developed in that paper led to an equation of ninth degree, similarly as in [2] for a mixture of two Laplace distributions. An application of the method of greatest reliability to the estimation of parameters of a mixture would lead to a very involved computation. The first paper [3] dealing with estimating of parameters for a mixture of more than two distributions concerned Bernoulli distributions and made use of factorial moments. In papers [4] and [5] the authors have estimated the parameters of an arbitrary finite mixture of random variable distributions of continuous type.

The method of estimation proposed in the present paper is applied to a mixture of $k$ distributions of discrete (geometric) type. For any $k$ this method leads to an algebraic equation of $k$-th degree with given number coefficients and to two systems of linear equations: in this way we obtain estimators for $2k-1$ unknown parameters.

We consider a mixture of an arbitrary finite number $k \geqslant 2$ of geometric distributions

(1) $P(X_i=j) = (1-p_i)p_i^j, \quad 0<p_i<1, \quad i=1,2,\ldots,k, \quad j=0,1,\ldots$

and we assume that

(2) $$p_1 < p_2 < \cdots < p_k.$$

The distribution of the mixture is defined by the formula

(3) $$P(X=j) = \sum_{i=1}^{k} \alpha_i (1-p_i) p_i^j, \quad \alpha_i > 0, \quad \sum_{i=1}^{k} \alpha_i = 1.$$

From a population having the distribution (3) we take a random sample of $n$ elements

(4) $$x_1, x_2, \ldots, x_n,$$

where $x_l = 0,1,2,\ldots$ for $l = 1,2,\ldots,n$.

On the basis of this sample we wish to estimate the parameters $p_1, p_2, \ldots, p_k$ and $\alpha_1, \alpha_2, \ldots, \alpha_{k-1}$, i.e. together $2k-1$ parameters as $\alpha_k = 1 - \sum_{i=1}^{k-1} \alpha_i$.

Let $m_{[r]}$ denote factorial moments of order $r$, $r=1,2,\ldots$, of the random variable with distribution (1). These moments can be expressed by the formulas

(5) $$m_{[r]} = \sum_{j=0}^{\infty} j^{[r]} p_i^j (1-p_i) = r! \left(\frac{p_i}{1-p_i}\right)^r,$$

where $j^{[r]} = j(j-1)(j-2)\ldots(j-r+1)$.

L e m m a 1. Let

(6) $$\mu_i = \frac{p_i}{1-p_i}, \qquad i = 1,2,\ldots,k,$$

(7) $$M_r = \frac{\bar{m}_{[r]}}{r!},$$

where $\bar{m}_{[r]}$ is the factorial moment of order $r$ for the mixture (3). Then the numbers $\alpha_1, \alpha_2, \ldots, \alpha_k$ are a solution of equations

(8) $$\alpha_1 \mu_1^r + \alpha_2 \mu_2^r + \ldots + \alpha_k \mu_k^r = M_r, \qquad r=1,2,\ldots,2k.$$

P r o o f. From formulas (5) and (6) we obtain

$$\bar{m}_{[r]} = r!\mu_i^r, \qquad r = 1,2,\ldots \tag{9}$$

Hence the factorial moments $\bar{m}_{[r]}$ of the mixture with distribution (3) can be expressed by the formulas

$$\bar{m}_{[r]} = \sum_{i=1}^{k} \alpha_i r!\mu_i^r, \qquad r = 1,2,\ldots \tag{10}$$

Using (7) we show that $\alpha_1,\alpha_2,\ldots,\alpha_k$ are a solution of the system (8). Observe that in view of (2) and (6) the numbers $\mu_i$, $i=1,\ldots,k$, satisfy the inequality

$$0 < \mu_1 < \mu_2 < \ldots < \mu_k. \tag{11}$$

L e m m a 2. The numbers $\mu_1,\mu_2,\ldots,\mu_k$ are the roots of the equation

$$\mu^k + a_1\mu^{k-1} + \ldots + a_{k-1}\mu + a_k = 0 \tag{12}$$

whose coefficients are the only solution of the system of equations

$$a_k M_i + a_{k-1}M_{i+1} + \ldots + a_1 M_{i+k-1} + M_{i+k} = 0, \quad i=1,2,\ldots,k. \tag{13}$$

P r o o f. Let us multiply first $k$ equations of the system (8) by the coefficients $a_k, a_{k-1},\ldots,a_1$, respectively, and add them side by side. We then obtain

$$\mu_1\alpha_1(a_k+a_{k-1}\mu_1+\ldots+a_1\mu_1^{k-1}) + \mu_2\alpha_2(a_k+a_{k-1}\mu_2+\ldots a_1\mu_2^{k-1}) + \ldots$$

$$\ldots + \alpha_k\mu_k(a_k + a_{k-1}\mu_k + \ldots + a_1\mu_k^{k-1}) = a_k M_1 + a_{k-1}M_2+\ldots+a_1 M_k.$$

Hence in view of (12) we obtain the first equation of the system (13). Further if we delete the first equation from the system (8) and consider next $k$ equations, then applying the

the above procedure we obtain the second equation of the system (13). Repeating this k-times we obtain the system of $k$ linear equations (13). Hence the coefficients of equation (12) satisfy the system (13). We are going to show that the coefficients of equation (12) are the only solutions of the system (13). To this aim it suffices to show that the matrix $M$ of the system (13) is non-singular. We represent the matrix $M$ as follows

$$M = \begin{bmatrix} M_1 M_2 & \dots M_k \\ M_2 M_3 & \dots M_{k+1} \\ \dots & \dots \\ M_k M_{k+1} & \dots M_{2k-1} \end{bmatrix} = \begin{bmatrix} \mu_1 \mu_2 \dots \mu_k \\ \mu_1^2 \mu_2^2 \dots \mu_k^2 \\ \dots \\ \mu_1^k \mu_2^k \dots \mu_k^k \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 & \dots \\ 0 & \alpha_2 & \dots \\ \dots \\ 0 & 0 & \dots \alpha_k \end{bmatrix} \cdot \begin{bmatrix} 1 \mu_1 \dots \mu_1^{k-1} \\ 1 \mu_2 \dots \mu_2^{k-1} \\ \dots \\ 1 \mu_k \dots \mu_k^{k-1} \end{bmatrix}.$$

Since the determinants of the first and third matrix are the Vandermonde determinants, in view of (11) we infer that the determinant of $M$ is not zero.

L e m m a 3. The coefficients $a_i$ of equation (12) are rational functions of the ordinary moments $m_i$ $(i=1,2,\dots,r)$, of the distributions (1).

P r o o f. From Lemma 2 and Cramer's formulas it follows that the coefficients $a_1,\dots,a_k$ are rational functions of the factorial moments $\overline{m}_{[r]}$ of the mixture (3). In view of (9) and (10) these coefficients can be expressed rationally by the factorial moments $m_{[r]}$ of (1). It is known that the factorial moments of order $r$ for (1) are rational functions of the ordinary moments $m_i$ $(i=1,2,\dots,r)$ provided they exist. Namely, we have the following formula

$$(14) \qquad m_{[r]} = \sum_{k=0}^{\infty} k(k-1)\dots(k-r+1)p_k =$$

$$= \sum_{k=0}^{\infty} \left( S_0^r k^r p_k + S_1^r k^{r-1} p_k + \dots + S_{r-1}^r k p_k \right) = S_0^r m_1 + S_1^r m_2 + \dots + S_{r-1}^r m_1 ,$$

where $S_0^r = 1$, $S_i^r$, $(i=1,2,\dots,r-1)$, are Stirling numbers of the first kind. Hence we have shown that $a_1,\dots,a_k$ are rational functions of the ordinary moments for (1).

L e m m a 4. If $\xi_1^{(n)}, \xi_2^{(n)},\dots,\xi_k^{(n)}$ are consistent estimators of the parameters $\xi_1,\dots,\xi_k$, then the solutions of the equation

$$\xi_k^{(n)}x^k + \xi_{k-1}^{(n)}x^{k-1} + \dots + \xi_1^{(n)}x + \xi_0^{(n)} = 0$$

are consistent estimators of the solutions of the equation

$$\xi_k x^k + \xi_{k-1}x^{k-1} + \dots + \xi_1 x + \xi_0 = 0.$$

P r o o f. Let us assume the following notation

$$w_k^{(n)}(x) = \xi_k^{(n)}x^k + \xi_{k-1}^{(n)}x^{k-1} + \dots + \xi_1^{(n)}x + \xi_0^{(n)},$$

$$w_k(x) = \xi_k x^k + \xi_{k-1}x^{k-1} + \dots + \xi_1 x + \xi_0.$$

From the hypothesis that $\xi_i^{(n)} \xrightarrow[n\to 0]{P} \xi_i$, $i=0,1,\dots,k$, that is

$$\bigwedge_{\xi>0} \lim_{n\to\infty} P\left(\left|\xi_i^{(n)} - \xi_i\right| > \xi_i\right) = 0$$

and from Slucki's theorem [6] it follows that for a fixed $x$ we have

$$w_k^{(n)}(x) \xrightarrow[n\to\infty]{P} w_k(x).$$

Let $x_0$ be a root of the polynomial $w_k(x)$. Hence we have $w_k^{(n)}(x_0) \xrightarrow[n\to\infty]{D} 0$, that is

$$\bigwedge_{\varepsilon>0} \bigwedge_{\eta>0} \bigvee_{N>0} \bigwedge_{n\geqslant N} P\left(\left|w_k^{(n)}(x_0)\right|>\varepsilon\right) < \eta.$$

Now we are going to show that

$$\text{(a)}\quad \bigwedge_{n\geqslant N}\ \bigwedge_{\delta_n>0}\ \bigvee_{x_o^{(n)}\in I} w_k^{(n)}(x_o^n) = 0, \text{ where } I=(x_o-\delta_n, x_o+\delta_n)$$

up to sets of measure 0. Suppose that (a) does not hold, i.e.

$$\bigvee_{n\geqslant N}\ \bigwedge_{\delta_n>0}\ \bigwedge_{x_o^{(n)}\in I} w_k^{(n)}\left(x_o^{(n)}\right) \neq 0.$$

This implies $\left|w_k^{(n)}\left(x_o^{(n)}\right)\right| \neq 0$. However, this contradicts the continuity of polynomials. Hence (a) holds.

We must prove yet that

$$x_o^{(n)} \xrightarrow[n\to\infty]{P} x_o$$

that is

$$\text{(b)}\qquad \bigwedge_{\varepsilon_1>0}\ \bigwedge_{\eta_1>0}\ \bigvee_{N_1>0}\ \bigwedge_{n\geqslant N_1} P\left(\left|x_o^{(n)}-x_o\right|>\varepsilon_1\right)<\eta_1 .$$

Suppose that (b) does not hold, i.e.

$$\bigvee_{\varepsilon_1>0}\ \bigvee_{\eta_1>0}\ \bigwedge_{N_1>0}\ \bigvee_{n\geqslant N_1} P\left(\left|x_o^{(n)}-x_o\right|>\varepsilon_1\right)>\eta_1 .$$

This means that with some positive probability we have $\left|x_o^{(n)}-x_o\right|>\varepsilon_1$ for arbitrarily large $n$. Hence in the sequence of polynomials $w_k^{(n)}(x)$ there exists a polynomial with an arbitrarily large index $n$ which has no zero in some neighbourhood $(x_o-\varepsilon_1, x_o+\varepsilon_1)$. Thus we have

$$P\left(\left|w_k^{(n)}(x)\right|>\varepsilon\right)>\eta$$

which contradicts (a). Since $x_0$ is arbitrary and the polynomials $w_k^{(n)}(x)$ and $w_k(x)$ have at most $k$ zeros, it follows that if $x_i^{(n)}$, $(i=1,2,\dots,k)$, are zeros of $w_k^{(n)}(x)$, then $x_i^{(n)}$, $(i=1,2,\dots,k)$, are stochastically convergent to the zeros $x_i$, $(i=1,2,\dots,k)$, of the polynomial $w_k(x)$. This ends the proof.

T h e o r e m. If we take as estimators for the factorial moments $\bar{m}_{[r]}$ of the mixture of distributions (3) the factorial moments of the sample (4) in the form

$$\hat{\bar{m}}_{[r]} = \sum_{i=1}^{r} S_{r-i}^{r}\, \hat{\bar{m}}_i, \qquad (i=1,2,\dots,2k), \tag{15}$$

where $\hat{\bar{m}}_i = \frac{1}{n}\sum_{l=1}^{n} x_l^i$, and

$$\max_{1\leqslant l\leqslant n} (x_l) > 2k-1, \tag{16}$$

then we have:

$1^0$ $\hat{M}_r = \frac{1}{r!}\hat{\bar{m}}_{[r]}$ are consistent and non-based estimators of the parameters $M_r$ defined by formula (7).

$2^0$ If in place of the coefficients $M_r$ of the system (13) we take their estimators $\hat{M}_r$, then the estimators $a_1,a_2,\dots,a_k$ being the solution of the system (13) are consistent estimators for the parameters $\hat{a}_1,\hat{a}_2,\dots,\hat{a}_k$.

$3^0$ The solutions $\hat{\mu}_1,\hat{\mu}_2,\dots,\hat{\mu}_k$ of the equation

$$\mu^k + \hat{a}_1\mu^{k-1} + \dots + \hat{a}_{k-1}\mu + \hat{a}_k = 0 \tag{17}$$

are consistent estimators for the parameters (6).

$4^0$ The estimators

$$\hat{p}_i = \frac{\hat{\mu}_i}{1+\hat{\mu}_i}, \qquad i=1,2,\dots,k, \tag{18}$$

are consistent estimators for the parameters (2).

$5^{\circ}$ The system of equations

$$(19)\quad \alpha_1\left(\hat{\mu}_1^i-\hat{\mu}_k^i\right)+\alpha_2\left(\hat{\mu}_2^i-\hat{\mu}_k^i\right)+\ldots+\alpha_{k-1}\left(\hat{\mu}_{k-1}^i-\hat{\mu}_k^i\right)=\hat{M}_i-\mu_k^i,$$

$$i=1,\ldots,k-1,$$

has as solution the consistent estimators $\hat{\alpha}_1,\hat{\alpha}_2,\ldots,\hat{\alpha}_{k-1}$ of the parameters $\alpha_1,\alpha_2,\ldots,\alpha_{k-1}$.

P r o o f. $1^{\circ}$ In view of (16) there exists exactly one solution $\hat{M}_1,\hat{M}_2,\ldots,\hat{M}_{2k}$ of the system of equations $\hat{M}_r=\frac{1}{r!}\hat{\hat{m}}_{[r]}$. It is known that the estimators $\hat{\hat{m}}_{[r]}$ are consistent and non-based estimators of the factorial moments $\hat{\hat{m}}_{[r]}$, of the mixture of distributions (3). By Slucki's theorem we infer that $1^{\circ}$ holds.

$2^{\circ}$ Using lemma 3 and the first point of theorem we obtain $2^{\circ}$.

$3^{\circ}$ If $\hat{\mu}_1,\hat{\mu}_2,\ldots,\hat{\mu}_k$ are the solutions of equation (17), then from lemma 4 it follows that they are consistent estimators for the parameters (11). Hence we see that in fact we are solving the equation (17) by an approximate method. We cannot claim that the obtained solutions are all positive and distinct but, in view of (11) taking into account that the estimators are consistent and the sample (4) is taken from the mixture (3), we see that the event complementary to

$$(20)\qquad 0<\hat{\mu}_1<\hat{\mu}_2\ \ldots\ <\hat{\mu}_k$$

becomes practically impossible as $n\to\infty$.

$4^{\circ}$ From $3^{\circ}$ and Slucki's theorem it follows that the estimators $\hat{p}_i$ defined by (18) are consistent estimators for the parameters (2) and satisfy the inequalities $\hat{p}_1<\hat{p}_2<\ldots<\hat{p}_k$.

$5^{\circ}$ Making use of Lemma 1 and the condition $\alpha_k=-\sum_{i=1}^{k-1}\alpha_i$ we obtain the system of linear equations (19). The determinant of this system is, in view of (20), different from zero. Hence there exists exactly one solution $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{k-1}$ of the system (19), which we take as estimators of the parameters $\alpha_1, \alpha_2, \ldots, \alpha_{k-1}$, and the parameter $\alpha_k$ is estimated by $\hat{\alpha}_k = 1 - \sum_{i=1}^{k-1} \alpha_i$. From $3^0$ and Slucki's theorem it follows that they are consistent estimators of the parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$.

It may happen that some $\hat{\alpha}_i$ are negative, equal to 0 or even greater than 1, but since the sample (4) is taken from the population with distribution (3), the event complementary to

$$\left[ \left( 0 < \hat{\alpha}_i < 1 \right) \cap \left( \sum_{i=1}^{k-1} \hat{\alpha}_i < 1 \right) \right]$$

is for a large sample practically impossible. In this way we have obtained estimators for the whole group of 2k parameters

$$p_1, p_2, \ldots, p_k; \quad \alpha_1, \alpha_2, \ldots, \alpha_k$$

of the mixture of k geometric distributions, where k is an arbitrary number.

## BIBLIOGRAPHY

[1] K. Pearson: Contributions to the mathematical theory of evolution, Philos. Roy.Soc. London Ser. A. 185 (1894) 71-110.

[2] E. Kącki, W. Krysicki: Die Parameterschätzung einer Mischung von zwei Laplaceschen Verteilungen im allgemeinen Fall, Comment. Math. 11 (1967) 23-31.

[3] W.R. Blischke: Estimating the parameters of mixtures of binomial distributions, J. Amer. Statist. Assoc. 59 (1964) 510-520.

[4] A.B. Kabir: Estimation of parameters of a finite mixture of distributions, J. Roy Statist. Soc. Ser. B, 30 (1968) 472-482.

[5] W. K r y s i c k i: Estimation of the parameters of the mixture of an arbitrary number of exponential distributions. Demonstratio Math. 4 (1972) 175-183.

[6] M. F i s z: Rachunek prawdopodobieństwa i statystyka matematyczna, (Probability theory and mathematical statistics), Warszawa 1967.

[7] M.G. K e n d a l l, A. S t u a r t: The advanced theory of statistics v.1. London 1963.

INSTITUTE OF MATHEMATICS, TECHNICAL UNIVERSITY, ŁÓDŹ

Received August 9, 1975.