Journal of the Korean Data & Information Science Society
2014, **25**(1), 255–261

http://dx.doi.org/10.7465/jkdi.2014.25.1.255
한국데이터정보과학회지

A maximum likelihood estimation method for a mixture of shifted binomial distributions

Changhyuck Oh[1]

[1]Department of Statistics, Yeungnam University

Received 16 December 2013, revised 8 January 2014, accepted 13 January 2014

Abstract

Many studies have estimated a mixture of binomial distributions. This paper considers an extension, a mixture of shifted binomial distributions, and the estimation of the distribution. The range of each component binomial distribution is first evaluated and then for each possible value of shifted parameters, the EM algorithm is employed to estimate those parameters. From a set of possible value of shifted parameters and corresponding estimated parameters of the distribution, the likelihood of given data is determined. The simulation results verify the performance of the proposed method.

Keywords: EM algorithm, likelihood, mixture, shifted binomial.

1. Introduction

Mixtures of distributions are of great interest in many disciplines for theory and applications (see McLachlan and Peel, 2001). Bonnini *et al.* (2012) consider a mixture of binomial and uniform distributions and Domenico (2003), a mixture of uniform and shifted binomial distributions. Lee and Oh (2006) and Oh (2006) investigate a method for estimating parameters of a mixture of the shifted Poisson distributions. Many studies have estimated a mixture of binomial distributions and its applications (Wasilewski, 1988; Blischke, 1964; Johnson *et al.*, 2005; Liu, 2006; Park, 2013). This paper considers a method for estimating parameters of a mixture of shifted binomial distributions, namely a mixture of generalized binomial distributions, by using the EM algorithm. The EM algorithm, introduced in Dempster *et al.* (1977), has become a widely popular technique (see McLachlan and Krishnan, 2008). When the EM algorithm is applied to a mixture of distributions of some specific type, obtained data are expected to be incomplete. That is, the data set contains no information on components that generate each data value. The E-step of EM algorithm estimates values of components to get estimated complete data set. The shifted binomial distribution, considered by Čekanavičius (2009), has three parameters, and its probability function is given by

$$f(x;m,\theta,k)=\binom{m}{x-k}\theta_i^{x-k}(1-\theta)^{m-x+k},\ x=k,\ k+1,...,\ k+m, \tag{1.1}$$

[1] Professor, Department of Statistics, Yeungnam University, Gyeongsan 712-749, Korea.
E-mail: choh@yu.ac.kr

where θ is a real value between 0 and 1 and k is an integer. Here m is assumed to be a fixed positive integer. If k is 0, then the probability function in (1) reduces to an ordinary binomial probability function. The mean and variance of a random variable with the probability function (1.1) are $m\theta + k$ and $m\theta(1-\theta)$, respectively. On the other hand, the probability function of a mixture of shifted binomial distributions is given by

$$f(x;\Phi) = \sum_{i=1}^{g} \pi_i f(x; m_i, \theta_i, k_i) \tag{1.2}$$

where g is the number of mixture components and π_i, $i = 1, ..., g$ is the weight for component i with constraints $0 < \pi_i < 1$ and $\pi_1 + \cdots + \pi_g = 1$. It is assumed that all m_i are known and equal to m. Here the notation $\Phi = (\pi_1, ..., \pi_{g-1};\ \theta_1, ..., \theta_g;\ k_1, ..., k_g)$ is adopted for all parameters.

Because the probability function of a mixture of shifted binomial distributions is represented as a weighted linear combination of binomial probability functions, it is not easy to obtain maximum likelihood estimates directly from the likelihood function by, for example, differentiating it. Therefore, an iterative algorithm such as the EM algorithm is generally applied.

Oh (2006) investigates a method for estimating parameters of a mixture of shifted Poisson distributions. The present paper adopts this methodology to estimate parameters of a mixture of shifted binomial distributions.

For a mixture of shifted binomials, this paper examines a method for estimating parameters when no parameters are known. As in Oh (2006), for each possible value of the shift parameter k based on observed data, other parameters are first estimated using the EM algorithm, and then a value of k is selected using estimates of corresponding parameters with the highest likelihood values.

2. The procedure of estimation

It is assumed that n observations, denoted by $\boldsymbol{x} = (x_1, ..., x_n)$ are obtained from the mixture of shifted binomial distributions given in (1.2). Observed data have no information on components, and therefore it is not possible to determine which component generates each x_j, $j = 1, ..., n$. In this sense, the data are regarded as incomplete ones carrying no information on components and $f(x;\Phi)$ is referred as a probability function for incomplete data.

For the given observation $\boldsymbol{x}$, the log-likelihood function for Φ is given by

$$L_{\boldsymbol{x}}(\Phi) = \sum_{j=1}^{n} \log f(x_j;\Phi). \tag{2.1}$$

The maximum likelihood estimate $\widehat{\Phi}_{\boldsymbol{x}}$ of Φ can be taken by finding the value of Φ that maximizes the function $L_{\boldsymbol{x}}(\Phi)$. However, it is well known that for a mixture of shifted binomial distributions, finding $\widehat{\Phi}_{\boldsymbol{x}}$ by differentiating the log-likelihood function is not feasible because of the structural form of $L_{\boldsymbol{x}}(\Phi)$. Therefore, by considering $\boldsymbol{x}$ as part of complete data, the paper proposes a procedure based on the EM algorithm for finding local maximum likelihood estimates. It is assumed that each observation x_j, $j = 1, ..., n$, has no information on components. The component vector $z_j = (z_{j1},\ z_{j2}, ..., z_{jg})$ is considered to denote information on the component probability function generating x_j. The elements $z_{j1}, z_{j2}, ..., z_{jg}$ of the vector z_j have a value 1 for components generating x_j and 0 for others.

If observations and component information are both observed, the data (x_1, z_1), ..., $(x_{n,} z_n)$ are referred as complete data, and the log-likelihood for complete data is given by

$$L_{\boldsymbol{x},\boldsymbol{z}}(\Phi) = \sum_{i=1}^{g}\sum_{j=1}^{n} z_{ji}\left\{\log \pi_i + \log f_i(x_j; m, \theta_i, k_i)\right\}. \tag{2.2}$$

For complete data, maximum likelihood estimates of parameters can be obtained easily from equation (2.2):

$$\widehat{\pi}_i = \frac{\sum\limits_{j=1}^{n} z_{ji}}{n}, \tag{2.3}$$

$$\widehat{\theta}_i = \frac{\sum\limits_{j=1}^{n} z_{ji}(x_j - \widehat{k}_i)}{\widehat{m}\sum\limits_{j=1}^{n} z_{ji}},\ i = 1, 2, ..., g.$$

Here, without loss of generality it is assumed that there is at least one observation from each component distribution.

Because a situation of no observed component information is assumed, the estimate in (2.3) cannot be directly obtained. To overcome this difficulty, component information is estimated first. In the $p+1$-th iteration, the E step of the EM algorithm obtain the estimate $\widehat{z}_{ji}$ of components z_{ji}. Then the estimated values $(x_1, z_1^{(p)}), ..., (x_n, z_n^{(p)})$ for complete data are supplied to (2.3) for the estimates $\widehat{\pi}_i$ and $\widehat{\theta}_i$ of π_i and θ_i, respectively.

Consider a set of possible shift parameters with given observations, namely

$$\mathbb{K} = \left\{K = (k_1, ..., k_g) : x_{(1)} - m = k_1 < \cdots < k_g = x_{(n)} + m\right\}, \tag{2.4}$$

for the estimation of the component $z_j, j = 1, ..., n$, where $x_{(1)} = \min(x_1, ..., x_n)$ and $x_{(n)} = \max(x_1, ..., x_n)$. Let the number of all possible elements of the set $\mathbb{K}$ be V and use the notation $\mathbb{K} = \left\{K^{(1)}, ..., K^{(V)}\right\}$. For a fixed shift parameter $K^{(v)} = (k_1^{(v)}, ..., k_g^{(v)})$, Oh (2006) proposes a procedure for estimating $\Phi_S = (\pi_1, ..., \pi_g;\ \theta_1, ..., \theta_g)$ for a mixture of shifted Poisson distributions. That is, for the fixed shift parameter $K^{(v)}$, let the initial value $\Phi_S^{(p,v)} = (\pi_1^{(p,v)}, ..., \pi_g^{(p,v)};\ \theta_1^{(p,v)}, ..., \theta_g^{(p,v)})$ be given. If $k_g^{(v)} < x_j < k_g^{(v)} + m$, then

$$z_{ji}^{(p+1,v)} = \frac{\pi_i^{(p,v)} f_i(x_j; \theta_i^{(p,v)}, k_i^{(v)})}{\sum_{h=1}^{g} \pi_h^{(p,v)} f_h(x_j; \theta_h^{(p,v)}, k_i^{(v)})}, i = 1, 2, ..., g. \tag{2.5}$$

For the estimates of complete data $(x_1, z_1^{(p+1,v)}), ..., (x_n, z_n^{(p+1,v)})$ in the $p+1$-th iteration, they are denoted for the parameters by

$$\pi_i^{(p+1,v)} = \frac{\sum_{j=1}^{n} z_{ji}^{(p+1,v)}}{n},$$
$$\theta_i^{(p+1,v)} = \frac{\sum_{j=1}^{n} z_{ji}^{(p+1,v)}(x_j - k_i^{(v)})}{m\sum_{j=1}^{n} z_{ji}^{(p+1,v)}}, i = 1, ..., g. \tag{2.6}$$

Let $\pi_i^{(p,v)} \leftarrow \pi_i^{(p+1,v)}$ and $\theta_i^{(p,v)} \leftarrow \theta_i^{(p+1,v)}$, and repeat (2.5), and (2.6). If the estimates converge, then let $\Phi_S^{(\infty,v)}$ be $\Phi_S^{(v)}$, and call it the estimate of Φ_S given shift parameter $K^{(v)} = (k_1^{(v)}, ..., k_g^{(v)})$. Estimates of Φ corresponding $K^{(v)}$ are denoted by $\Phi^{(v)} = (\pi_1^{(v)}, ..., \pi_g^{(v)};\ \theta_1^{(v)}, ..., \theta_g^{(v)};\ k_1^{(v)}, ..., k_g^{(v)})$. For $\Phi^{(v)}$ obtained from all $K^{(v)}$ in the set of shifted parameters $\mathbb{K}$, estimates of the parameter $\Phi = (\pi_1, ..., \pi;\ \theta_1, ..., \theta_g;\ k_1, ..., k_g)$ are given as

$$\widehat{\Phi} = \text{argmax}\left\{L_{\boldsymbol{x}}(\Phi^{(v)})|v = 1, 2, ..., V\right\}. \tag{2.7}$$

In the estimation of the shift parameter, Φ_S is estimated for all possible combination of shift parameters for given data.

If the observation $\boldsymbol{x}$ is given, then the procedure for the estimation of parameters can be summarized as follows:

Procedure for estimation of the parameters

Set the initial value: $v \leftarrow 1$.
Step 1: Set the value of the shift parameter $K^{(v)} = (k_1^{(v)}, ..., k_g^{(v)}) \in \mathbb{K}$.
Sub-step 1-1: Set $p \leftarrow 1$.
Set the initial value $\Phi_S^{(p,v)} = (\pi_1^{(p,v)}, ..., \pi_g^{(p,v)}; \theta_1^{(p,v)}, ..., \theta_g^{(p,v)})$ for Φ_S.
Sub-step 1-2: For a given parameter $\Phi_S^{(p,v)}$, obtain estimates $z_1^{(p+1,v)}, ..., z_n^{(p+1,v)}$ with (2.4) or (2.5). With estimates of complete data $(x_1,\ z_1^{(p+1,v)})$, ..., $(x_n, z_n^{(p+1,v)})$ and the method in (2.6), obtain the estimate $\Phi_S^{(p+1,v)}$.
Sub-step 1-3: If the condition for convergence is not satisfied, then go to sub-step 1-2 by setting $p \leftarrow p+1$ and $\Phi_S^{(p,v)} \leftarrow \Phi_S^{(p+1,v)}$. If satisfied, then denote the estimate of the parameter as $\widehat{\Phi}(v)$. If $v < V$, then set $v \leftarrow v+1$ and go to step 1. If not, go to step 2.
Step 2: Estimate the given parameter based on (2.7).

In sub-step 1-3, the condition for convergence is $\left|L_{\boldsymbol{x}}(\Phi^{(p+1)}) - L_{\boldsymbol{x}}(\Phi^{(p)})\right| < \delta$. That is, increment of the log-likelihood is some range of values. Here $\delta > 0$ is an upper bound for convergence.

3. Simulations

Monte Carlo simulations are conducted to evaluate the performance of the proposed method. Here a mixture of binomial distributions with g=2 is considered. For each simulation, the number of repeats and sample size are set to 1000 and $n = g \times 50$, respectively.

Table 3.1 is for g=2, $(\theta_1, \theta_2) = (0.5, 0.5)$, (k_1, k_2)=(0, 3), (0, 4), (0, 5), and (π_1, π_2)=(.7, .3), (.6, .4), (.5, .5) and the values are sample means and standard errors (parentheses) for 1000 simulations. In each simulation n values of a mixture of shifted binomials are generated and the procedure is applied to each data set to obtain the maximum likelihood estimate. The sample means of estimated values for shift parameters are quite close to true values. On the other hand, for each value of (π_1, π_2), as the value of k_2 increases, the sample means approach true values, whereas the standard errors remain stable. This is expected in that when two shift parameters are far apart, two components of a mixture distribution can be easily identified. For each value of (k_1, k_2), there is a trend in which the value of π_1 approaches 0.5 and the sample means of estimates of k_1, k_2, π_1, π_2, θ_1, θ_2 approach true parameter values. For (k_1, k_2) =(0, 4), for example, the sample means of estimates of π_1 =.7, .6, and .5 are 0.693, 0.692, and 0.501, respectively, and their standard errors are quite similar. This corresponds to the heuristic in that when ratios of two components of a mixture distribution are similar, there is good estimation performance. For each (π_1, π_2), the sample means and standard errors approach true values of the parameters as the difference between two shift parameters increases. Overall, k_1 tends to be under-estimated, whereas k_2, over-estimated. The sample means of estimates of π_1 and π_2 are all close to true values. The standard errors of π_1 and π_2 in each case are the same because their sum is 1 and the variance of the sample mean is $n\pi_1\pi_2$.

Table 3.1 Sample means and standard errors of 1000 simulations with g=2, (θ_1, θ_2)=(0.5, 0.5)

(π_1, π_2)	(k_1, k_2)	$\widehat{\pi}_1$	$\widehat{\pi}_2$	$\widehat{k}_1$	$\widehat{k}_2$	$\widehat{\theta}_1$	$\widehat{\theta}_2$
0.7, 0.3	0, 3	0.697	0.303	-0.260	4.052	0.515	0.436
		(0.164)	(0.164)	(1.909)	(2.916)	(0.191)	(0.294)
	0, 4	0.702	0.298	0.043	4.141	0.493	0.501
		(0.089)	(0.089)	(1.589)	(2.408)	(0.157)	(0.237)
	0, 5	0.702	0.298	(0.007)	4.860	0.498	0.518
		(0.059)	(0.059)	(1.607)	(2.073)	(0.156)	(0.203)
	0, 6	0.698	0.302	0.152	5.763	0.485	0.528
		(0.052)	(0.052)	(1.520)	(2.037)	(0.148)	(0.197)
	0, 7	0.701	0.299	0.051	6.737	0.495	0.528
		(0.047)	(0.047)	(1.473)	(1.956)	(0.146)	(0.193)
0.6, 0.4	0, 3	0.617	0.383	-0.141	4.174	0.505	0.416
		(0.156)	(0.156)	(1.976)	(2.711)	(0.199)	(0.268)
	0, 4	0.598	0.402	-0.047	4.086	0.500	0.499
		(0.090)	(0.090)	(1.723)	(2.156)	(0.170)	(0.211)
	0, 5	0.600	0.400	0.079	4.750	0.491	0.528
		(0.064)	(0.064)	(1.590)	(1.896)	(0.154)	(0.185)
	0, 6	0.600	0.400	0.106	5.846	0.490	0.517
		(0.054)	(0.054)	(1.566)	(1.844)	(0.154)	(0.179)
	0, 7	0.599	0.401	0.045	6.861	0.496	0.514
		(0.050)	(0.050)	(1.524)	(1.837)	(0.150)	(0.180)
0.5, 0.5	0, 3	0.521	0.479	-0.031	4.037	0.491	0.425
		(0.167)	(0.167)	(1.992)	(2.567)	(0.213)	(0.247)
	0, 4	0.509	0.491	0.146	4.110	0.482	0.497
		(0.093)	(0.093)	(1.765)	(2.048)	(0.177)	(0.200)
	0, 5	0.503	0.497	0.232	5.033	0.476	0.501
		(0.064)	(0.064)	(1.673)	(1.789)	(0.164)	(0.175)
	0, 6	0.500	0.500	0.160	5.909	0.483	0.511
		(0.056)	(0.056)	(1.641)	(1.724)	(0.161)	(0.168)
	0, 7	0.499	0.501	0.170	6.913	0.483	0.510
		(0.051)	(0.051)	(1.630)	(1.635)	(0.159)	(0.160)

Table 3.1. Continued.

(π_1, π_2)	(k_1, k_2)	$\widehat{\pi}_1$	$\widehat{\pi}_2$	$\widehat{k}_1$	$\widehat{k}_2$	$\widehat{\theta}_1$	$\widehat{\theta}_2$
0.7, 0.3	0, 3	0.697 (0.059)	0.303 (0.059)	-0.391 (1.635)	3.524 (2.213)	0.439 (0.162)	0.553 (0.220)
	0, 4	0.698 (0.053)	0.302 (0.053)	-0.392 (1.594)	4.314 (2.083)	0.438 (0.159)	0.570 (0.206)
	0, 5	0.697 (0.049)	0.303 (0.049)	-0.358 (1.541)	5.377 (1.951)	0.436 (0.154)	0.563 (0.193)
	0, 6	0.702 (0.046)	0.298 (0.046)	-0.374 (1.491)	6.511 (2.035)	0.439 (0.149)	0.548 (0.202)
	0, 7	0.702 (0.043)	0.298 (0.043)	-0.525 (1.577)	7.706 (2.043)	0.453 (0.158)	0.529 (0.203)
0.6, 0.4	0, 3	0.603 (0.063)	0.397 (0.063)	-0.332 (1.637)	3.637 (2.145)	0.434 (0.163)	0.540 (0.210)
	0, 4	0.600 (0.056)	0.400 (0.056)	-0.338 (1.625)	4.339 (1.875)	0.433 (0.161)	0.566 (0.185)
	0, 5	0.598 (0.052)	0.402 (0.052)	-0.404 (1.585)	5.515 (1.887)	0.441 (0.157)	0.550 (0.187)
	0, 6	0.601 (0.051)	0.399 (0.051)	-0.497 (1.681)	6.394 (1.839)	0.450 (0.167)	0.559 (0.182)
	0, 7	0.601 (0.053)	0.399 (0.053)	-0.454 (1.610)	7.564 (1.880)	0.446 (0.161)	0.543 (0.186)
0.5, 0.5	0, 3	0.503 (0.065)	0.497 (0.065)	-0.416 (1.799)	3.509 (1.927)	0.440 (0.176)	0.552 (0.189)
	0, 4	0.501 (0.056)	0.499 (0.056)	-0.371 (1.715)	4.513 (1.796)	0.437 (0.170)	0.551 (0.176)
	0, 5	0.500 (0.054)	0.500 (0.054)	-0.453 (1.732)	5.427 (1.728)	0.444 (0.171)	0.558 (0.171)
	0, 6	0.500 (0.053)	0.500 (0.053)	-0.514 (1.774)	6.428 (1.684)	0.452 (0.175)	0.558 (0.168)
	0, 7	0.500 (0.050)	0.500 (0.050)	-0.459 (1.702)	7.593 (1.706)	0.447 (0.168)	0.541 (0.170)

4. Discussion

Liu *et al*. (2006) introduce a method for estimating parameters of a mixture of binomial distributions by using the EM algorithm. However, in the estimation of a mixture of shifted binomial distributions, the method cannot be applied directly. This paper proposes a method for employing the EM algorithm to estimate a mixture of shifted binomial distributions. The mixture ratio and the component probability of success are first determined for each component binomial distribution for each possible value of the shift parameter, and then the value of shift parameters providing the highest log-likelihood value is selected. The simulation results show that the proposed method provides reasonable performance in terms of sample means and standard errors. The satisfactory results for estimated values for the success probability of component binomial distributions indicate a need to extend the proposed method for even better performance. In addition, future research should investigate large-sample properties. The proposed method can be considered a generalization of Liu *et al*. (2006) in that a mixture of binomial distribution is studied.

References

Blischke, W. (1964). Estimating the parameters of mixture of binomial distributions, *Journal of the American Statistical Association*, **59**, 510-528.

Bonnini, S., Piccolo, D., Salmaso, L. and Solmi, F. (2012). Permutation inference for a class of mixture models. *Communications in Statistics-Theory and Methods*, **41**, 2879-2895.

Čekanavičius, V., Peköz, E. A., Röllin, A. and Shwartz, M. (2009). A three- parameter binomial approximation. Avalable from `http://arxiv.org/abs/0906.2855`.

Copsey, K. and Webb, A. (2003). Bayesian gamma mixture model approach to radar target recognition. *IEEE Transactions on Aerospace and Electronic Systems*, **39**, 1201-1217.

Dempster, A. P., Laird, N. M. and Rubin, D. R. (1977). Maximum likelihood from incomplete data. *Journal of the Royal Statistical Society B*, **39**, 1-38.

Domenico, P. (2003). On the moments of a mixture of uniform and shifted binomial random variables. *Quaderni di Statistica*, **5**, 1-20.

Johnson, N. L., Kemp, A. W. and Kotz, S. (2005). *Univariate discrete distributions*, 3rd Ed., Wiley-Interscience, New York.

Lee, H. J. and Oh, C. (2006). Estimation in mixture of shifted Poisson distributions with known shift parameters. *Journal of the Korean Data & Information Science Society*, **17**, 785-794.

Liu, Z., Almhana, J., Choulakian, V. and McGorman, R. (2006). Online EM algorithm for mixture with application to internet traffic modeling. *Computational Statistics & Data Analysis*, **50**, 1052-1071.

McLachlan, G. J. and Krishnan, T. (2008). *The EM algorithm and extensions*, 2nd ed., Wiley, Hoboken, NJ.

McLachlan, G. J. and Peel, D. (2001). *Finite mixture models*, John Wiley & Sons, Inc., New York.

Oh, C. (2006). Estimation in mixture of shifted Poisson distributions. *Journal of the Korean Data & Information Science Society*, **17**, 1209-1217.

Park, C. (2013). Simple principle component analysis using Lasso. *Journal of the Korean Data & Information Science Society*, **24**, 533-541.

Skipper, M. (2012). A Pólya approximation to the Poisson-binomial law. *Journal of Apply Probability*, **49**, 745-757.

Wasilewski, M. J. (1988). Estimation of the parametes of the mixture k>2 log- normal distributions. *Trabajos de estadistica*, **3**, 167-175,