Adv. & Appl. in Stat. **5**(2) (2005), 171-182

ITERATIVE MIXTURE ESTIMATION FOR MODELS WITH LENGTH-BIASED COMPONENT

OLCAY AKMAN

Department of Mathematics, Illinois State University
Normal, IL 61790, U. S. A.

Abstract

We estimate the mixing proportion in a discrete mixture of two lifetime pdf's, where one component is the length-biased version of the parent distribution using the EM algorithm. We compare our results with the estimates obtained by solving the MLE equations numerically.

1. Introduction

The concept of using weighted distributions in lifetime studies originates from the need of estimation using sampling distributions that are different from the parent populations. In many situations, experimenters do not work with a truly random sample from the population that they are interested in, either by design, or because experimental conditions make random selection from the target population impossible. Even if a random sample can be obtained, the experimenter may choose not to use it, since a carefully chosen biased sample may turn out to be more informative (e.g., Bayarri and DeGroot [2]).The usual statistical analysis assumes that a random sample from the original distribution is obtained; however, since the observations do not have equal chances of entering the sample, the resulting sampling distribution does not follow the assumed original distribution. Feller's [4]

2000 Mathematics Subject Classification: 62, 65C60.

Key words and phrases: inverse Gaussian, life-time distribution, Monte Carlo simulation, EM algorithm.

Received February 16, 2005

Waiting time for the bus paradox is perhaps the most delightful example of how a biased sample arises. If we consider interarrival times of buses at a bus stop as sampling units, then the arrival of an individual at the bus stop is equivalent to choosing an interarrival interval. Clearly the probability of a certain interval being chosen is proportional to its length. That is, if X be the interarrival time random variable with pdf $f(x)$, and Y is the length of the interarrival time point that contains the arbitrary point, t, of one's arrival, with cdf $G(y)$, then from Feller's result we have

$$G(y) = \mu^{-1}\int_0^y xf(x)dx, \tag{1.1}$$

when $\mu = E(X) < \infty$. Intuitively, an interval of length y has y times the original likelihood of covering the arbitrary point t, hence we *modify* the original density $f(x)$ to

$$g(y) = c \cdot yf(y),$$

where c is the normalizing constant. Such statistical models that incorporate the restrictions under which the observations were obtained are called *weighted models*, and the random samples obtained under sampling restrictions are called *selection samples*. In this article we focus on a special case of weighted distributions that arise when observations are selected with a probability proportional to their "length", namely the length-biased distributions. The length-biased distributions occur naturally for some sampling plans in reliability, biometry, wildlife studies, and survival analysis, among others. More specifically, a distribution function G_F defined on R^+ is called *length-biased distribution* corresponding to a df F (also defined on R^+), if

$$G_F(y) = \mu_F^{-1}\left\{\int_0^y x \cdot dF(x)\right\}, \quad \forall y \in R^+, \tag{1.2}$$

where $\mu_F = \int_0^\infty x \cdot dF(x)$. Note that the Radon-Nikodym derivative of G_F with respect to F is given by

$$\frac{dG(x)}{dF(x)} = \frac{x}{\mu_F}. \tag{1.3}$$

Hence, the length-biased pdf can be written as

$$g(x) = \frac{x \cdot f(x)}{\mu_F}, \tag{1.4}$$

where μ_F is assumed to be finite and nonzero. Most recently Akman [1] examined lifetime estimation in the presence of a length-biased sampling plan. Here we consider a random variable, X_p, whose probability density function is a discrete mixture of the assumed (original) pdf and its length-biased version, that is,

$$f_p(x) = (1 - p) \cdot f(x) + p \cdot g(x), \quad 0 \le p \le 1. \tag{1.5}$$

Using (1.4), we observe that

$$f_p(x) = f(x)\left(1 - p + p\,\frac{x}{E(X)}\right). \tag{1.6}$$

We call a pdf defined in (1.6) an *original length-biased* (OL) *mixture pdf*. Estimation of the parameter p, which we call *contamination parameter,* is important due to the fact that its value provides an indication of the presence of length-biasedness in the sampling plan.

This article deals with estimation of the mixing proportion p and is organized as follows: In Section 2, we consider estimation of p using the EM algorithm. Section 3 contains an application where we estimate the mixing proportion in a mixture of inverse Gaussian pdf and its length-biased counterpart.

2. Estimation via EM algorithm

In this section, we consider the development of the EM algorithm to estimate p, where the mixing pdf's are the original and length-biased versions of an assumed model.

Let Y be a random vector corresponding to the observed data y with pdf $g(y; \Theta)$, where $\Theta = (\theta_1, \ldots, \theta_d)'$ is a vector of unknown parameters with parameter space Ω. In the context of the EM algorithm, the observed data vector y is viewed as incomplete and is considered as an observable function of the complete data. The notion of *incomplete data* includes the conventional sense of missing data, but it also applies to cases where the complete data represent what would be available from some hypothetical experiment. Within this framework we let x denote the complete-data vector, and we let z denote the vector containing the additional data, commonly referred to as the unobservable or missing data. If we let $g_c(x; \Theta)$ be the pdf of the random vector X, then the complete data log likelihood function is given by

$$\log L_c(\Theta) = \log g_c(x; \Theta). \tag{2.1}$$

The EM algorithm approaches the problem of solving the incomplete data likelihood equation $\partial L(\Theta)/\partial\Theta = 0$ indirectly by proceeding iteratively in terms of the complete data log likelihood function, $\log_c L(\Theta)$. Since it is unobservable, we replace $\log_c L(\Theta)$ by its conditional expectation given y, using the current fit for Θ. More specifically, let $\Theta^{(0)}$ be some initial value for Θ. Then in the first iteration, the E-step requires the calculation of

$$Q(\Theta; \Theta^{(0)}) = E_{\Theta^{(0)}}\{\log_c(\Theta) \mid y\}. \tag{2.2}$$

On the other hand, the M-step requires the maximization of $Q(\Theta; \Theta^{(0)})$ with respect to Θ over Ω. In other words, we choose $\Theta^{(1)}$ such that

$$Q(\Theta^{(1)}; \Theta^{(0)}) \geq Q(\Theta; \Theta^{(0)}) \tag{2.3}$$

for all $\Theta \in \Omega$. The E- and M-steps are carried out iteratively until the difference

$$L(\Theta^{(k+1)}) - L(\Theta^{(k)}) \tag{2.4}$$

changes by an arbitrarily small amount in the case of convergence of the sequence of likelihood values $\{L(\Theta^{(k)})\}$. Good overviews of the EM algorithm are provided by Little and Rubin [5] and Tanner [9]. Redner and Walker [7] studied the properties of the algorithm in terms of maximum likelihood estimation in mixture densities, while McLachlan and Krishnan [6] provide a book-length treatment of EM where particularly an estimation of mixing proportions for a general mixture of densities is discussed.

In our case, we consider the EM algorithm in likelihood estimation for mixture models, where

$$f_\phi(\mathbf{X}) = \sum_{i=1}^{m} \pi_i \cdot f_i(\mathbf{X};\ \theta),$$

with $\sum \pi_i = 1$. The likelihood equation $\partial L(\phi)/\partial\phi = 0$ can be so manipulated that the likelihood estimate $\hat{\phi}$ satisfies

$$\hat{\pi}_i = \sum_{j=1}^{n} a_{ij}/n, \quad (i = 1, \ldots, m)$$

and

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\partial \log f_i(x_j;\ \hat{\theta})/\partial\hat{\theta} = 0.$$

In case of finite mixture models, we let the vector of indicator variables $z_j = (z_{1j}, \ldots, z_{mj})'$ be defined by

$$z_{ij} = \begin{cases} 1 & x_j \in G_i, \\ 0 & x_j \notin G_i, \end{cases}$$

where $z_1, \ldots, z_n \overset{iid}{\sim} M_m(1, \pi)$.

Under the necessary assumptions, x_j given z_j has log density

$$\sum_{i=1}^{m} z_{ij} \log f_i(x_j;\ \theta), \quad (j = 1, \ldots, n).$$

Hence, the log-likelihood for the complete data X and Z is given by

$$L(\phi) = \sum_i \sum_j z_{ij}\{\log \pi_i + \log f_i(x_j;\ \theta)\}.$$

Then the EM algorithm is applied to the mixture model by treating Z as missing data.

In particular, we have

$$f(x) = \sum_i \pi_i f_i(x \mid \theta_i) = \sum_i \pi_i f_i(x).$$

The log-likelihood is given by

$$L = \sum_j \log f(x_j)$$

$$= \sum_i \sum_j a_{ij} (\log(\pi_i f_i(x_j)) - \log(a_{ij})),$$

where $a_{ij} = \dfrac{\pi_i f_i(x_j)}{f(x_j)}$. Using Lagrange multipliers, we obtain

$$a_{ij}^* = \frac{\pi_i f_i(x_j)}{f(x_j)}$$

$$\pi_i^* = \frac{1}{n} \sum_{j=1}^{n} a_{ij}$$

$$\phi_i^* = \arg \max_{x \in \mathcal{D}} \sum_j a_{ij} \log f(x_j).$$

In particular, for the OL mixtures of the form

$$f(\mathbf{X};\ \psi) = \sum_{i=1}^{2} p_i \cdot f_i(x), \tag{2.5}$$

the loglikelihood is

$$\mathcal{L} = \sum_{j=1}^{n} \log f(x_j \mid \psi). \tag{2.6}$$

Substituting (2.5) in (2.6), we obtain

$$\mathcal{L} = \sum_{j=1}^{n} \log\left(\sum_{i=1}^{2} p_i \cdot f_i(x)\right). \tag{2.7}$$

For each j, we maximize

$$\sum_{i=1}^{2} a_{ij} \{\log(p_i f_i(x_j)) - \log(a_{ij})\}$$

subject to $\sum_{i=1}^{2} a_{ij} = 1$. That is, using Lagrange multipliers, we maximize

$$\sum_{i=1}^{2} a_{ij}\{\log(p_i f_i(x_j)) - \log(a_{ij})\} + \lambda\left(\sum_{i=1}^{2} a_{ij} - 1\right);$$

the result is

$$a_{ij} = \frac{p_i \cdot f_i(x_j)}{\sum_{i=1}^{2} p_i \cdot f_i(x_j)}.$$

Similarly, maximizing with respect to p_i gives

$$p_i = \frac{\sum_{j=1}^{n} a_{ij}}{n}.$$

Hence, we have maximized

$$f_\Theta = \sum_j (a_{1j} \log f_\theta(x_j) + a_{2j} \log f_\theta(x_j))$$

$$= \sum_j (a_{1j} \log f_\theta(x_j) + a_{2j} \log\left(\frac{x_j}{E(X)} f_\theta(x_j)\right)$$

$$= \sum_j \log f_\theta(x_j) - \sum_j a_{2j} \log(E(X))$$

$$+ \text{(term with no } \theta).$$

We can summarize the maximization process for OL mixtures by the following algorithm:

1. Initial guess for a_{ij};

2. Calculate p_i and θ;

3. Update a_{ij};

4. Go to step 2 until $| a_{ij}^{(k+1)} - a_{ij}^{(k)} | < \varepsilon$.

3. An Example

In this section, we perform a comparative study for the EM algorithm-based estimator using a mixture model of an inverse Gaussian and a length-biased inverse Gaussian densities. Inverse Gaussian is one of the most commonly used lifetime distributions applicable to a wide variety of real life problems. Its distributional properties and relevant applications can be found in Chhikara and Folks [3] and Seshadri [8].

Let X be an inverse Gaussian random variable with pdf

$$f(x \mid \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp^{-\frac{\lambda}{2\mu^2}\frac{(x-\mu)^2}{x}}, \quad x > 0, \lambda, \mu > 0. \tag{3.1}$$

The OL mixture of inverse Gaussian and length-biased inverse Gaussian (MIG) is given by

$$f(x \mid \mu, \lambda, p) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp^{-\frac{\lambda}{2\mu^2}\frac{(x-\mu)^2}{x}} \left(1 - p + p\frac{x}{\mu}\right), \quad 0 \le p \le 1. \tag{3.2}$$

Using a random sample $X_1, X_2, \ldots, X_n$ with pdf (3.2), we obtain the normal equations for μ, λ, and p as

$$\frac{\partial L}{\partial \mu} = \frac{\lambda}{\mu^3}\sum_i x_i - \frac{\lambda n}{\mu^2} - \sum_i \frac{p x_i}{\mu^2(1 - p + \mu p x_i)} \tag{3.3}$$

$$\frac{\partial L}{\partial \lambda} = \frac{n}{2\lambda} - \frac{1}{2\mu^2}\sum_i x_i + \frac{n}{\mu}\frac{1}{2}\sum_i \frac{1}{(x_i)} \tag{3.4}$$

$$\frac{\partial L}{\partial p} = \sum_i \frac{(x_i - \mu)/\mu}{(1 - p + p x_i/\mu)}. \tag{3.5}$$

For example, for fixed p, we obtain the MLE for triples (μ, λ, p) such as $(\overline{X}, (\widetilde{X}^{-1} - 1/\overline{X}), 0)$, $(\sqrt{\overline{X}\widetilde{X}}, 2(\widetilde{X}^{-1} - 1/\sqrt{\overline{X}\widetilde{X}})^{-1}, \frac{1}{2})$, and $(\widetilde{X}, \widetilde{X}^2/(\overline{X} - \widetilde{X}), 1)$.

In general, we can obtain unique solutions for (μ, λ) for fixed p that maximize the joint likelihood function obtained from (3.2). However, for unknown p, we consider the nonparametric and the EM algorithm for the estimation of mixing proportions since equation (3.5) results in a polynomial of order n.

3.1. Simulation study

The EM estimators of the mixing parameter for a mixture of inverse Gaussian and length-biased inverse Gaussian distributions were compared through a simulation study of 5,000 simulations of sample size 100 each. Inverse Gaussian samples with $(\mu/\lambda < 1), (\mu/\lambda = 1)$ and $(\mu/\lambda > 1)$ were generated. In the tables given below, the first column contains the values obtained using the EM algorithm. The second (search method) column contains estimates of p obtained by substituting values in $\frac{1}{1000}$ increments for $p \in (0, 1)$ until a solution for (μ, λ) is achieved in MLE equations. The MSE for each estimated value is presented in parentheses within the same cell. The computations were performed by using IMSL with FORTRAN. The following tables summarize the means and the MSE's (given in parentheses) of the estimators for different parameter values.

Estimates of p for Mixture IG

Table 1. $\left(\frac{\mu}{\lambda} < 1\right)$

	EM alg.	Search
$p = 0$	$1.761E01(3.47E02)$	$1.149E01(2.12E02)$
$p = 0.25$	$2.810E01(3.04E02)$	$2.269E01(3.11E02)$
$p = 0.50$	$5.293E01(2.84E02)$	$5.137E01(1.86E02)$
$p = 0.75$	$7.771E01(2.62E02)$	$7.603E01(2.19E02)$
$p = 1$	$9.626E01(3.09E02)$	$9.063E01(3.01E02)$

Table 2. $\left(\frac{\mu}{\lambda} = 1\right)$

	EM alg.	Search
$p = 0$	$1.298E01(3.66E02)$	$1.222E01(2.51E02)$
$p = 0.25$	$2.739E01(3.49E02)$	$2.912E01(3.36E02)$
$p = 0.50$	$5.919E01(2.61E02)$	$5.532E01(1.32E02)$
$p = 0.75$	$7.107E01(2.73E02)$	$7.204E01(2.84E02)$
$p = 1$	$9.006E01(3.55E02)$	$9.945E01(3.39E02)$

Table 3. $\left(\frac{\mu}{\lambda} > 1\right)$

	EM alg.	Search
$p = 0$	$2.194E01(4.45E02)$	$1.769E01(2.85E02)$
$p = 0.25$	$3.477E01(3.67E02)$	$2.253E01(3.58E02)$
$p = 0.50$	$6.347E01(2.86E02)$	$5.721E01(2.44E02)$
$p = 0.75$	$7.259E01(2.98E02)$	$7.554E01(2.99E02)$
$p = 1$	$8.875E01(4.78E02)$	$9.543E01(3.92E02)$

The graphs below depict the distribution of $\hat{p}$ obtained using the search method for mixture inverse Gaussian model simulated under different values of p.

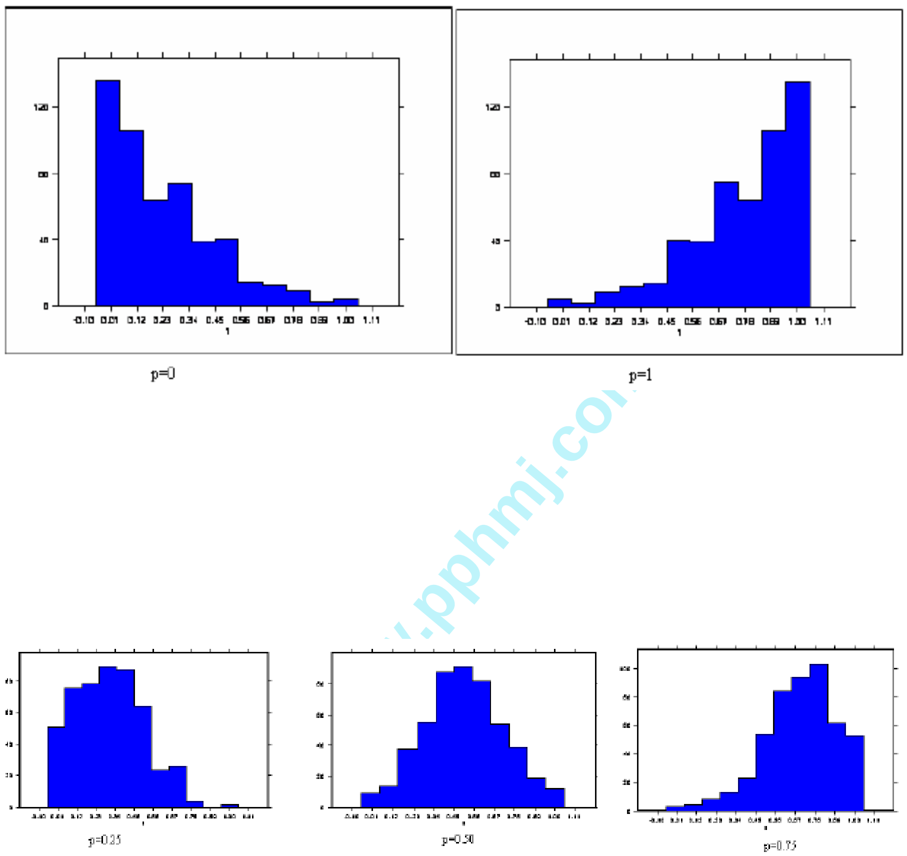

Figure 1. Distribution of $\hat{p}$ for mixture inverse Gaussian model.

3.2. Remarks

From the simulation results, we observe that overall EM estimation of the mixing parameter performs better since it has a smaller MSE than that of the search method. Additionally it is important to note that the computations the estimates obtained via the search method should be considered for reference only, and not for comparison purposes, as the search method is computationally cumbersome and impractical.

References

[1] O. Akman, On lifetime estimation in the presence of length-biased sampling plan, Int. J. Reliability, Quality and Safety Eng. 11(2) (2004), 187-194.

[2] M. J. Bayarri and M. H. DeGroot, Information in selection models, Probability and Bayesian Statistics, R. Viertl, ed., Plenum Press, New York, 1987.

[3] R. S. Chhikara and J. Leroy Folks, The Inverse Gaussian Distribution: Theory, Methodology, and Applications, Marcel Dekker, New York, 1989.

[4] W. Feller, An Introduction to Probability Theory and its Applications, Vol. 2, 2nd ed., John Wiley & Sons, Inc., New York, 1971.

[5] P. J. A. Little and D. B. Rubin, Statistical Analysis with Missing Data, Wiley, New York, 1987.

[6] G. McLachlan and T. Krishnan, The EM Algorithm and Extensions, Wiley, New York, 1997.

[7] R. A. Redner and H. F. Walker, Mixture densities, maximum likelihood and the EM algorithm, SIAM Rev. 26(2) (1984), 195-239.

[8] V. Seshadri, The Inverse Gaussian Distribution: A Case Study in Exponential Families, Oxford Science Publications, Oxford University Press, 1994.

[9] M. A. Tanner, Tools for Statistical Inference: Observed Data and Data Augmentation Methods, 3rd ed., Springer-Verlag, New York, 1996.

■