

MIXTURE DECOMPOSITION FOR DISTRIBUTIONS FROM THE EXPONENTIAL FAMILY USING A GENERALIZED METHOD OF MOMENTS

S. T. Sum and B.J. Oommen*
*School of Computer Science
Carleton University
Ottawa, Canada, K1S 5B6*

ABSTRACT

A finite mixture distribution consists of the superposition of a finite number of component probability densities, and is typically used to model a population composed of two or more subpopulations. Mixture models find utility in situations where there is a difficulty in directly observing the underlying components of the population of interest. This paper examines the method of moments as a general estimation technique for estimating the parameters of the component distributions and their mixing proportions. It is shown that the same basic solution can be applied to any continuous or discrete density from the exponential family with a known common shape parameter. Results of an empirical study of the method are also presented.

I. INTRODUCTION

Consider the random vector \underline{X} whose probability distribution can be modeled according to the probability density function (PDF)

$$f(\underline{x}; \underline{\Theta}, \underline{\alpha}) = \sum_{i=1}^c \alpha_i f_i(\underline{x}; \underline{\theta}_i) \quad \underline{x} \in \mathbb{R}^d, \quad (1.1)$$

where the constraints on α_i ($i=1, \dots, c$), are

$$\alpha_i > 0 \text{ and } \sum_{i=1}^c \alpha_i = 1.$$

The density function (1.1) defines a finite mixture distribution consisting of a superposition of c component densities, $f_i(\underline{x}; \underline{\theta}_i)$.

In addition to its constituent functions, $f_i(\cdot)$, the mixture density is characterized by two vectors, namely, $\underline{\Theta}$ and $\underline{\alpha}$. The former denotes the c -component parameter vector $\underline{\Theta} = (\underline{\theta}_1, \dots, \underline{\theta}_c)$ whose elements represent the parameter vectors $\underline{\theta}_i = (\theta_{i1}, \dots, \theta_{ir})$ of the

* Both authors partially supported by Natural Sciences and Engineering Research Council of Canada. A preliminary version of this paper will be presented at ISCIS IX, the 1994 International Symposium on Computer and Information Sciences, Antalya, Turkey, November 1994.

individual densities for $i = 1, \dots, c$. The vector, $\underline{\alpha} = (\alpha_1, \dots, \alpha_c)$, comprises what are alternatively known as the mixing weights, mixing parameters, or mixing proportions of the component densities forming the mixture.

A mixture distribution describes a population which comprises two or more subpopulations combined in fixed proportions. Probability models of this type are utilized in situations where the sample observations can only be drawn from the whole population and not separately from the individual component populations. Such cases typically occur due to a difficulty in distinguishing the underlying subpopulations forming the mixed population [18]. Consequently, mixture models are employed in fields which include pattern recognition, biology, chemistry, geology, medicine, and actuarial science [27].

An example taken from the discipline of biology will serve to illustrate the type of situation in which mixture models arise [12]. Biologists are often faced with the task of inferring population parameters on the basis of samples collected during the course of field experiments. A particular characteristic of the species of interest is measured for each individual in the sample. An important consideration when taking these measurements is the age group or sex of the individual, since the distribution of many characteristics is dependent on these. However, the age or sex of a species is not always readily determined when measurements are performed in the natural environment. Hence, observations must be drawn from the mixed population, where the component populations are the different age groups or sexes of the species.

Ultimately, the goal in studying distributions of the foregoing nature is to decompose the mixture population into its individual subpopulations. That is, we wish to determine the component densities, $f_i(\underline{x}|\theta_i)$, as well as the proportions, α_i , in which the components are present in the mixture. Most commonly, the only information available to aid in achieving such a goal is a sample of measurements representative of the population in question. If one can assume a functional form for the component densities, then the problem becomes one of estimating the parameters of the individual distributions and their mixing weights on the basis of the given sample. The probability model defined by these parameter estimates becomes our best estimate of the “true” mixture population.

An aspect of particular importance when dealing with the parameter estimation of mixture distributions is the concept of identifiability. A given estimation method cannot reasonably be expected to yield reliable estimators if the mixture in question is not identifiable. That is, there must exist a set of parameters, $\{\underline{\Theta}, \underline{\alpha}\}$, which uniquely characterize the mixture distribution in order for the method to provide meaningful results. More formally [11], a class \mathbb{F} of finite mixture densities is termed **identifiable** if and only if for any two $f(\underline{x}; \underline{\Theta}, \underline{\alpha})$, $f(\underline{x}; \underline{\Theta}^*, \underline{\alpha}^*) \in \mathbb{F}$, the equality

$$\sum_{i=1}^c \alpha_i f_i(\underline{x}; \underline{\theta}_i) = \sum_{j=1}^{c^*} \alpha_j^* f_j(\underline{x}; \underline{\theta}_j^*)$$

implies that $c = c^*$, and that the summations can be ordered such that $\alpha_i = \alpha_j^*$, and $\underline{\theta}_i = \underline{\theta}_j^*$ for $i, j = 1, \dots, c$. Several authors, including Yakowitz and Spragins [29], and Teicher [26], have extensively investigated the difficulty of identifiability. However, commonly-used distributions do not pose a problem of non-identifiability except in certain simple or degenerate cases. We shall assume that the mixtures treated in this paper are identifiable.

A general analytic technique used for the parameter estimation of mixtures is known as the method of moments (MM). This method was introduced a century ago in 1894 by Karl Pearson [19] in order to resolve a mixture of two univariate normal densities. Normal mixtures are particularly difficult to resolve using the MM due to the algebraic complexity of the problem. Even for Pearson's relatively simple mixture, calculation of the parameter estimates requires determining the appropriate negative root of a ninth-degree polynomial arrived at by the laborious reduction of a system of five simultaneous equations. Charlier and Wicksell [5] as well as Cohen [7] reported somewhat simplified derivations of this solution.

Since Pearson's time, numerous researchers from a variety of different fields have investigated mixture distributions from both theoretical and practical perspectives (see the bibliographies of [21,27]). A majority of the literature concerns mixtures of normal distributions, which is partly a consequence of the relative importance of this distribution, and partly due to the greater difficulty presented by such mixtures. Often constraints on specific parameters of a mixture are imposed in order to simplify the estimation problem. For example, in mixtures of normal distributions, one might assume that the means or variances are equal or known. Alternatively, one might be concerned solely with estimating the mixing proportions in a mixture of known component densities [5,7,8,11,27].

Other continuous as well as discrete mixtures can also be decomposed using the MM. In particular, many different mixtures of densities from the exponential family can readily be resolved using this method. The estimation of the parameters of these distributions is often treated for the two-component case in the published work. Studies of continuous and discrete mixtures by Rider [22,23], John [13,14], and Cohen [7], are typical of this. Several authors have generalized the method to c -component mixtures of a particular form. In this vein, we cite Blischke's [3] work on mixed binomial distributions, Everitt and Hand's [11] research on Poisson mixtures, and Lingappaiah's [17] study of

mixed exponential densities. Titterton *et. al.* [27] reported results for several general mixtures of continuous and discrete functional forms.

Despite the considerable interest manifested in mixture models over the years, there exist surprisingly few sources of literature that address the problem of general mixture distributions in a comprehensive manner. In treatments of the method of moments, a significant number of papers have confined their discussion to a single type of mixture distribution. Moreover, many authors have focused exclusively on two-component mixtures. Typically, expressions have been derived specifically for mixtures of two densities of a particular functional form. There has been curiously little reported on the generalization of the solution to mixtures of any number of component densities of an arbitrary form.

This paper presents a comprehensive study of the method of moments as an estimation technique for finite mixture distributions. Given the number of component densities and their functional form, as well as a sample of observations from the population of interest, the procedure for estimating the parameters of the component densities and their mixing proportions is shown. A generic solution of this method is derived in detail for general c -component mixture densities from the exponential family. An empirical study using simulated mixtures of these densities is performed to investigate the estimators produced by the method.

Section 2 describes the various probability density functions collectively called the exponential family. Several oft-encountered probability distributions from this family are presented. These include the exponential, gamma, Weibull, Poisson, binomial, and geometric densities. In Section 3, the general estimation scheme known as the method of moments is discussed in the context of mixture distributions. The method is based on a system of estimating equations obtained by equating sample moments to their corresponding theoretical moments. Section 3 proceeds to detail the solution of general mixtures of an arbitrary number of component distributions from the exponential family. It is shown that the parameters of many continuous and discrete distributions from this family can be estimated in an identical fashion.

The simulation workbench utilized for the experimental component of the paper is described in Section 4. In particular, the manner in which random variables and the simulated mixtures are generated is given. The chi-square (χ^2) metric and the Expected Mutual Information Measure (EMIM) for evaluating the goodness of fit of the models to the simulated data are defined. Results of the MM decomposition of two-component mixtures of the aforementioned densities are discussed.

II. EXPONENTIAL FAMILY OF DISTRIBUTIONS

A large number of probability densities belong to a family of distributions known as the exponential family. These functions may be either continuous or discrete and form an important class of commonly used distributions both in theory and in practice. A detailed mathematical treatise on the exponential family is found in Barndorff-Nielsen [1]. A more informal discussion is given by Dobson [10].

Consider the random variable X whose PDF is a function of the parameter θ . Then a probability density $f(x; \theta)$ is said to belong to the exponential family if $f(x; \theta)$ has the form

$$f(x; \theta) = g(x) \exp[d(\theta)t(x) + h(d(\theta))] \quad (2.1)$$

where $g(x)$ is nonnegative and is independent of the parameter θ . In many instances it is more convenient to express a probability density from the exponential family in the above form where the reparametrization, $\theta \rightarrow d(\theta)$, is made. The parameter $d(\theta)$ is called the natural parameter of the density. This form of the PDF is equivalent to one known as the Koopman-Darmois or the Pittman-Koopman form [15].

Equation (2.1) is easily generalized to the r -parameter case where h and d are now functions of the parameter vector $\underline{\theta} = (\theta_1, \dots, \theta_r)$. Thus, $f(x; \underline{\theta})$ is given by

$$f(x; \underline{\theta}) = g(x) \exp \left[\sum_{j=1}^s d_j(\underline{\theta}) t_j(x) + h(d_1(\underline{\theta}), \dots, d_s(\underline{\theta})) \right], \quad 1 \leq r \leq s. \quad (2.2)$$

Examples of probability functions which belong to the exponential family include the exponential, gamma, Weibull, Poisson, binomial, and geometric distributions. Table 2.1 lists their standard and natural parametrizations.

Distribution	Probability Density Function	
	Standard Parametrization	Natural Parametrization
Exponential	$\frac{1}{b} \exp(-x/b)$	$\exp[-x/b - \ln(b)]$
Gamma	$\frac{b^{-a} x^{a-1} \exp(-x/b)}{\Gamma(a)}$	$\exp[-x/b + (a-1)\ln(x) - a\ln(b) - \ln(\Gamma(a))]$
Weibull	$ab^{-a} x^{a-1} \exp[-(x/b)^a]$	$\exp[-(x/b)^a + (a-1)\ln(x) + \ln(a) - a\ln(b)]$
Poisson	$\frac{\lambda^x \exp(-\lambda)}{x!}$	$\frac{1}{x!} \exp[x\ln(\lambda) - \lambda]$
Binomial	$\binom{m}{x} p^x (1-p)^{m-x}$	$\binom{m}{x} \exp[x \ln(p/(1-p)) + m \ln(1-p)]$
Geometric	$(1-p)^x p$	$\exp[x\ln(1-p) + \ln(p)]$

Table 2.1: Natural and standard parametric forms of several densities from the exponential family.

With the exception of the binomial, mixtures of any of these densities are usually identifiable [27]. A necessary and sufficient condition for the identifiability of binomial mixtures is that $m \geq 2c-1$, where m is the number of independent Bernoulli trials [3,26].

III. ESTIMATION USING THE METHOD OF MOMENTS

3.1 Method Of Moments

It can be shown that the theoretical moments of a distribution are reasonably approximated by their observational counterparts when the number of sample values is sufficiently large [24]. The method of moments utilizes this fact to estimate the parameters of a distribution. Theoretical moments, which are functions of the unknown parameters, are equated to their corresponding sample moments. Estimates of the parameters are obtained by setting up as many equations as there are parameters, and solving these equations in terms of the sample moments.

Suppose that there are r parameters $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_r)$ to be estimated. Then a system of r equations is created as follows:

$$\mu_k(\hat{\underline{\theta}}) = m_k(\mathbf{X}^n) \quad k= 1, 2, \dots, r \quad (3.1)$$

where, $\mu_k(\hat{\underline{\theta}})$ is the k th population moment,

$m_k(\mathbf{X}^n)$ is the k th sample moment,

$\hat{\underline{\theta}} = (\theta_1, \dots, \theta_r)$ is the estimated parameter vector,

and, $\mathbf{X}^n = \{x_1, \dots, x_n\}$ is the sample data.

Expressing the above system of equations as

$$\underline{\mu}(\hat{\underline{\theta}}) = \underline{m}(\mathbf{X}^n), \quad (3.2)$$

the estimated parameters are obtained by inversion

$$\hat{\underline{\theta}} = \underline{\mu}^{-1} [\underline{m}(\mathbf{X}^n)]. \quad (3.3)$$

Of course, the moment equations may not always be invertible and, in such cases, one must resort to a different estimation method. Fortunately, for mixtures of most of the common parametric densities of theoretical and practical interest, this is not a concern. The moment estimates for several such mixture densities are derived below.

Before deriving the MM solutions given by equation (3.3), it should be noted that these may be indeterminate. That is, the moment estimates may not necessarily fall in the permissible range of values for the parameters. In particular, imaginary estimates of real-valued parameters may result, and estimates of mixing proportions may not be in the interval (0,1). No general theoretical conditions exist under which estimation by the method of moments can be assured to yield determinate solutions for the general mixture problem [12].

3.2 Mixtures of Continuous Distributions

The general form of the single continuous distribution from the exponential family is given in natural parametrization by

$$f(x; a, b) = g(x) \exp \left[\sum_{j=1}^s d_j(a, b) t_j(x) + h(d_1(a, b), \dots, d_s(a, b)) \right], \quad 1 \leq s \leq 2, \quad (3.4)$$

where (3.4) is obtained from (2.2) for $\underline{\theta} = (a, b)$, and $a > 0$ and $b > 0$ are called the shape and scale parameters respectively.

We will now present the fundamental result of our paper which is valid for all continuous distributions from the exponential family of the form of (3.4).

Theorem 1

Consider a general mixture of c continuous distributions of a given parametric form from the exponential family with a known common shape parameter, a , and component scale parameters, b_i ($i = 1, \dots, c$). Every mixture of this family can be decomposed using the same generic solution.

Proof:

A general mixture of c continuous distributions with common shape parameter, a , and component scale parameters, b_i ($i = 1, \dots, c$) has the PDF

$$f(x; a, \underline{b}, \underline{\alpha}) = \sum_{i=1}^c \alpha_i f_i(x; a, b_i), \quad x > 0, \quad (3.5)$$

where α_i ,

$$0 < \alpha_i < 1 \quad \text{and} \quad \sum_{i=1}^c \alpha_i = 1,$$

is the i th component mixing proportion, and, $f_i(\cdot)$ is the i th component density function shown in (3.4). Assuming that the shape parameter is known, there are $2c - 1$ parameters to be estimated for such a mixture, specifically, $\underline{\alpha} = (\alpha_1, \dots, \alpha_{c-1})$ and $\underline{b} = (b_1, \dots, b_c)$.

The parameters of compound distributions of this type can be estimated using non-central moments. The k th population moment about the origin is defined as

$$\mu'_k = \int_{-\infty}^{\infty} x^k f(x) dx, \quad k = 0, 1, 2, \dots$$

The corresponding sample moment about the origin is given by

$$m'_k = \frac{1}{n} \sum_{j=1}^n x_j^k, \quad k = 0, 1, 2, \dots$$

where x_1, x_2, \dots, x_n are the n sample values drawn from the population of interest.

Following the method of moments for parameter estimation, the first $2c-1$ sample moments are equated to their theoretical counterparts to yield the system of equations

$$m_k' = \sum_{i=1}^c \alpha_i \int_0^{\infty} x^k f_i(x) dx \quad k = 0, 1, 2, \dots, 2c-1. \quad (3.6)$$

Evaluating the right-hand side of (3.6) leads to a system of equations linear in the mixing proportions, α_i , of the form

$$\phi_k = \sum_{i=1}^c \alpha_i b_i^k, \quad k = 0, 1, \dots, 2c-1, \quad (3.7)$$

where ϕ_k is a function of the k th order sample moment, m_k' , and of the shape parameter, a .

The c scales, b_i ($i = 1, \dots, c$), are taken to be distinct. Note that the form of equation (3.7) is characteristic of the exponential family.

After some algebraic manipulations, the system in (3.7) can be reduced to a single c -degree polynomial of the form

$$b^c + \beta_{c-1} b^{c-1} + \dots + \beta_1 b + \beta_0 = 0 \quad (3.8)$$

where β_i ($i = 0, \dots, c-1$) are the constants such that the component scale parameters b_1, \dots, b_c are the ordered roots of (3.8). Thus, we need to determine the constants β_i such that the system (3.7) can be equivalently represented by the polynomial (3.8).

Now taking the sum of the product of β_k and the k th equation in (3.7) for $k = 0, 1, \dots, c-1$, and then adding the c th equation results in the expression

$$\sum_{k=0}^{c-1} \beta_k \phi_k + \phi_c = \sum_{k=0}^{c-1} \beta_k \left(\sum_{i=1}^c \alpha_i b_i^k \right) + \sum_{i=1}^c \alpha_i b_i^c.$$

Rearranging, we have

$$\sum_{k=0}^{c-1} \beta_k \phi_k + \phi_c = \sum_{i=1}^c \alpha_i \left(\sum_{k=0}^{c-1} \beta_k b_i^k + b_i^c \right).$$

Applying (3.8), this reduces to

$$-\phi_c = \sum_{k=0}^{c-1} \beta_k \phi_k.$$

In similar fashion, taking the sum of the product of β_k and the $(k+1)$ st equation in (3.7) for $k = 0, 1, \dots, c-1$, and then adding the $(c+1)$ st equation leads to

$$-\phi_{c+1} = \sum_{k=0}^{c-1} \beta_k \phi_{k+1}.$$

Proceeding in this manner, a system of c linear equations is created such that

$$-\phi_{c+h} = \sum_{k=0}^{c-1} \beta_k \phi_{k+h} \quad h= 0,1,...,c-1,$$

$$\text{or} \quad \tilde{\Phi} \underline{\beta} = \underline{\phi}, \quad (3.9)$$

$$\text{where} \quad \tilde{\Phi} = \begin{bmatrix} \phi_0 & \phi_1 & \dots & \phi_{c-1} \\ \phi_1 & \phi_2 & \dots & \phi_c \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{c-1} & \phi_c & \dots & \phi_{2c-2} \end{bmatrix}, \quad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{c-1} \end{bmatrix}, \quad \underline{\phi} = \begin{bmatrix} \phi_c \\ \phi_{c+1} \\ \vdots \\ \phi_{2c-1} \end{bmatrix}.$$

Since the moment equations are independent, $\tilde{\Phi}$ is non-singular [2], and hence, the coefficients β_k ($k=0,...,c-1$) may be determined by inverting (3.9):

$$\underline{\beta} = \tilde{\Phi}^{-1} \underline{\phi}. \quad (3.10)$$

Subsequently, the c coefficients β_k can be substituted into equation (3.8) and the estimates of the parameters b_i ($i= 1,...,c$) determined as the roots of the polynomial. A number of root-finding routines (such as Laguerre's [20]) can be used to accomplish this.

Once the estimates $\hat{b}_1, \dots, \hat{b}_c$ have been determined, the mixing proportions are obtained by solving the first c equations in (3.7), i.e.,

$$\phi_k = \sum_{i=1}^c \alpha_i b_i^k, \quad k= 0,1,...,c-1.$$

This system may be expressed as

$$\tilde{B} \underline{\alpha} = \underline{\phi}, \quad (3.11)$$

$$\text{where} \quad \tilde{B} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ b_1 & b_2 & \dots & b_c \\ b_1^2 & b_2^2 & \dots & b_c^2 \\ \vdots & \vdots & \ddots & \vdots \\ b_1^{c-1} & b_2^{c-1} & \dots & b_c^{c-1} \end{bmatrix}, \quad \underline{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{c-1} \end{bmatrix}, \quad \underline{\phi} = \begin{bmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_{c-1} \end{bmatrix}.$$

The matrix \tilde{B} is the well-known Vandermonde matrix which can be solved by Rybicki's algorithm [20]. Hence, the result. ◆◆◆

To highlight the main thrust of this result we recapitulate the procedure informally. For any mixture with a known common shape parameter, the first $2c-1$ sample moments about the origin are first calculated. Then, using the calculated moments in (3.10) the constants β_k ($k=0,...,c-1$) are determined. Next, these constants are utilized in a c -degree polynomial and the c ordered roots corresponding to the parameters $0 < b_1 < b_2 < \dots < b_c$

are found. Finally, the estimated mixing proportions $\hat{\alpha}_i$ ($i= 1,...,c$) are determined using the scale parameter estimates \hat{b}_i ($i=1,...,c$) in (3.11).

We shall now use the above theorem for certain specific distributions. In the interest of brevity, we shall only consider the exponential, gamma, and Weibull mixtures in detail. Although they follow the same general form described above, we have included them here to assist a "practitioner" who would encounter the specific distributions in a real-life scenario. However, the solution, in its generality, is valid for all members of the continuous exponential family.

3.2.1 Specific Continuous Mixtures

For the mixed **exponential** PDF we have the moment equations,

$$m'_k = k! \cdot \sum_{i=1}^c \alpha_i b_i^k, \quad k= 0,1,2,...,2c-1.$$

In this case, the form for ϕ_k to be used in (3.7) is :

$$\phi_k = (m'_k / k!).$$

Hence, the parameter vectors \underline{b} and $\underline{\alpha}$ can be estimated by solving equations (3.8-3.11).

The mixture of c **gamma** distributions with known common shape parameter, a , yields the estimating equations,

$$m'_k = a(a+1) \cdots (a+k-1) \cdot \sum_{i=1}^c \alpha_i b_i^k, \quad k= 0,1,2,...,2c-1.$$

In this case ϕ_k has the functional form :

$$\phi_k = m'_k \{a(a+1) \cdots (a+k-1)\}^{-1}$$

which leads to the familiar system of equations (3.7).

Thus, the mixing proportions $\underline{\alpha} = (\alpha_1, ..., \alpha_{c-1})$ and the scales $\underline{b} = (b_1, ..., b_c)$ can be estimated as shown in Section 3.2 by solving (3.8-3.11). Note that the exponential random variable is simply the gamma random variable with shape parameter $a=1$. Consequently, using the previous expression for ϕ_k with $a=1$ yields the analogous results for the exponential mixture. Further, if the shape parameter, a , is an integer then the gamma distribution becomes the a -Erlang density.

For the c -component **Weibull** mixture with known common shape parameter, a , the moment equations are:

$$m'_k = \Gamma(1 + \frac{k}{a}) \sum_{i=1}^c \alpha_i b_i^k, \quad k = 0, 1, 2, \dots, 2c-1.$$

Defining $\phi_k = m'_k \left\{ \Gamma(1 + \frac{k}{a}) \right\}^{-1}$, we obtain the system of equations of (3.7).

Employing the same process detailed earlier, estimates $\hat{\underline{\alpha}} = (\hat{\alpha}_1, \dots, \hat{\alpha}_{c-1})$ and $\hat{\underline{b}} = (\hat{b}_1, \dots, \hat{b}_c)$ can be determined. Again, note that with $a=1$ the gamma function in the moment equations becomes a function of a positive integer, yielding $\Gamma(1 + k) = k!$. As a result, ϕ_k reduces to the same expression as that obtained for the exponential mixture.

The following example clarifies the estimation process.

Example 3.1 : Mixture of Two Gamma Densities

As an illustration of the estimation technique, consider a two-component gamma mixture with known shape parameter, a :

$$f(x; a, b_1, b_2, \alpha) = \alpha \frac{x^{a-1} \exp(-x/b_1)}{b_1^a \Gamma(a)} + (1-\alpha) \frac{x^{a-1} \exp(-x/b_2)}{b_2^a \Gamma(a)},$$

where $x > 0$ and $0 < b_1 < b_2^1$. The objective is to estimate the three parameters b_1 , b_2 , and α . Hence, the first three moments are required:

$$\begin{aligned} \phi_1 &= \alpha b_1 + (1-\alpha) b_2, \\ \phi_2 &= \alpha b_1^2 + (1-\alpha) b_2^2, \\ \phi_3 &= \alpha b_1^3 + (1-\alpha) b_2^3, \end{aligned}$$

where we have defined $\phi_k = m'_k \{a(a+1)\dots(a+k-1)\}^{-1}$ for $k = 1, 2, 3$. Rearranging we have

$$\alpha(b_1 - b_2) = \phi_1 - b_2, \quad (3.12a)$$

$$\alpha(b_1^2 - b_2^2) = \phi_2 - b_2^2, \quad (3.12b)$$

$$\alpha(b_1^3 - b_2^3) = \phi_3 - b_2^3. \quad (3.12c)$$

We now seek to reduce the above system to a quadratic equation of the form of (3.8). Replacing $\alpha(b_1 - b_2)$ by $(\phi_1 - b_2)$ in the last two equations of (3.12) yields

$$\phi_1(b_1 + b_2) - b_1 b_2 = \phi_2,$$

$$\phi_1[(b_1 + b_2)^2 - b_1 b_2] - b_1 b_2(b_1 + b_2) = \phi_3.$$

¹We assume here, with no loss of generality that $b_1 < b_2$. The only constraint is that b_1 and b_2 be distinct.

Then substituting $b_1 b_2 = \phi_1(b_1 + b_2) - \phi_2$ from the first equation into the second, and solving for $(b_1 + b_2)$ we obtain

$$(b_1 + b_2) = \frac{\phi_3 - \phi_1 \phi_2}{\phi_2 - \phi_1^2}.$$

Hence,

$$b_1 b_2 = \frac{\phi_1 \phi_3 - \phi_2^2}{\phi_2 - \phi_1^2}.$$

Letting $\beta_0 = b_1 b_2$ and $\beta_1 = -(b_1 + b_2)$ we then see that

$$b^2 + \beta_1 b + \beta_0 = 0,$$

which is the quadratic of interest.

Thus, the two scale parameter estimates are given by

$$\hat{b}_{1,2} = \frac{-\beta_1 \mp \sqrt{\beta_1^2 - 4\beta_0}}{2}. \quad (3.13)$$

In the two-component case the mixing proportion is readily obtained using (3.12a):

$$\hat{\alpha} = \frac{\phi_1 - \hat{b}_2}{\hat{b}_1 - \hat{b}_2}. \quad (3.14)$$

◆◆◆

Similar derivations can be obtained for the specific two-component cases of other distributions as well. Explicit expressions for the estimates are analogous to those of equations (3.13) and (3.14). Other researchers who have reported such results for continuous two-component mixtures include Rider [22,23], and John [14]. Lingappaiah [17] investigated mixtures of c exponential densities.

3.3 Mixtures of Discrete Distributions

For discrete distributions, we shall now derive a result analogous to the one derived earlier for the case of continuous distributions. The analytic formulation essentially follows the case of continuous distributions as in Section 3.2, but is distinct because we utilize factorial rather than the "traditional" central or non-central moments. It is included here in the interest of completeness.

Theorem 2

Consider a general mixture of c *discrete* distributions of a given parametric form from the exponential family with component parameters, p_i ($i= 1,...,c$). Every mixture of this family can be decomposed using the same generic solution.

Proof:

The general PMF of a mixture of c discrete distributions with unknown component parameters, $0 \leq p_i \leq 1$, and mixing proportions α_i satisfying :

$$0 < \alpha_i < 1, \quad \text{and} \quad \sum_{i=1}^c \alpha_i = 1, \text{ is given by}$$

$$P(x; \underline{p}, \underline{\alpha}) = \text{Prob}[X_j = x] = \sum_{i=1}^c \alpha_i P_i(x; p_i) \quad x = 0, 1, \dots \quad (3.15)$$

where X_1, \dots, X_n are n independent, identically distributed (IID) random variables each with PMF $P_i(\cdot)$. Such a mixture is characterized by a total of $2c-1$ parameters, namely, $\underline{\alpha} = (\alpha_1, \dots, \alpha_{c-1})$ and $\underline{p} = (p_1, \dots, p_c)$.

The use of factorial rather than regular moments for parameter estimation of discrete mixtures greatly simplifies the moment equations. The k^{th} order theoretical factorial moment about the origin is defined as [16]

$$\mu'_{[k]} = \sum_{j=1}^{\infty} x_j(x_j-1)\dots(x_j - k + 1) f(x_j) \quad k=1, 2, \dots$$

where $\mu'_{[0]}=1$. The observational analog of the latter is given by :

$$m'_{[k]} = \frac{1}{n} \sum_{j=1}^n x_j(x_j-1)\dots(x_j - k + 1) \quad k= 1, 2, \dots ; \quad m'_{[0]}=1.$$

For mixtures of discrete distributions, a function of the population factorial moment, $\mu'_{[k]}$, and of some known distribution parameter, ρ , common to all components,

can be expressed as a linear equation in the mixing proportions. That is,

$$\phi_k = \sum_{i=1}^c \alpha_i p_i^k \quad k=0, 1, \dots, 2c-1 \quad (3.16)$$

Thus, parameter estimation is effected by substituting in the left hand side of (3.16) the first $2c-1$ sample moments for their theoretical counterparts to create a system of equations analogous to that of (3.7). The c parameters $p_i > 0$ ($i=1, \dots, c$) are assumed to be distinct.

Following the same process as described earlier for continuous distributions, the estimators $\hat{p}_1, \dots, \hat{p}_c$ can be obtained as the roots of a polynomial of the form

$$p^c + \beta_{c-1}p^{c-1} + \dots + \beta_1p + \beta_0 = 0. \quad (3.17)$$

The constants β_i ($i=0, \dots, c-1$) are determined from equation (3.9) using the factorial moment functions, ϕ_k , defined above for $k=0, \dots, 2c-1$. The estimated mixing proportions, $\hat{\alpha}_1, \dots, \hat{\alpha}_c$, result from solving an equation analogous to (3.11) with the estimators \hat{p}_i ($i=1, \dots, c$) and moment functions, ϕ_k , for $k=0, \dots, c-1$. Hence, the result. $\blacklozenge \blacklozenge \blacklozenge$

As was done for the continuous densities, we shall now use the above theorem for certain specific discrete distributions². Again, in the interest of brevity, we shall only consider the Poisson, binomial, and geometric mixtures explicitly, noting that the solution, in its generality is valid for all discrete members of the exponential family.

3.3.1 Specific Discrete Mixtures

For general mixed **Poisson** distributions, the factorial moment equations are found by equating the first $2c-1$ observed moments to their expected values:

$$\begin{aligned} m'_{[k]} &= E[m'_{[k]}] = \sum_{x=0}^{\infty} x(x-1)\dots(x-k+1) \sum_{i=1}^c \alpha_i \frac{\lambda_i^x \exp(-\lambda_i)}{x!} \\ &= \sum_{i=1}^c \alpha_i \exp(-\lambda_i) \lambda_i^k \sum_{x=0}^{\infty} \frac{1}{x!} \frac{d^k \lambda_i^x}{d\lambda_i^k} \\ &= \sum_{i=1}^c \alpha_i \exp(-\lambda_i) \lambda_i^k \frac{d^k \exp(\lambda_i)}{d\lambda_i^k} \\ &= \sum_{i=1}^c \alpha_i \exp(-\lambda_i) \lambda_i^k \exp(\lambda_i) \\ &= \sum_{i=1}^c \alpha_i \lambda_i^k \quad k=0, 1, 2, \dots, 2c-1. \end{aligned}$$

In this case the functional form for ϕ_k is :

$$\phi_k = m'_{[k]},$$

² As in Section 3.2.1 these specific cases are included to assist a "practitioner" who encounters these distributions in real-life.

which yields the resulting system of equations of the form of (3.16). Consequently, the parameter estimates given by $\hat{\underline{\alpha}} = (\hat{\alpha}_1, \dots, \hat{\alpha}_{c-1})$ and $\hat{\underline{\lambda}} = (\hat{\lambda}_1, \dots, \hat{\lambda}_c)$ follow.

For general mixtures of c **binomial** densities, the number of trials, m , is assumed to be known and common to all component distributions. Hence, binomial mixtures involve the estimation of $2c-1$ parameters, namely, the individual probabilities of success, p_1, \dots, p_c and the mixing proportions $\alpha_1, \dots, \alpha_{c-1}$.

In the case of the binomial mixture, the theoretical factorial moment is given by

$$\mu'_{[k]} = m(m-1)\cdots(m-k+1) \cdot \sum_{i=1}^c \alpha_i p_i^k, \quad k = 1, 2, \dots$$

Applying the moments method for the first $2c-1$ moments we obtain

$$\phi_k = \sum_{i=1}^c \alpha_i p_i^k, \quad k = 0, 1, \dots, 2c-1,$$

where $\phi_k = \mu'_{[k]} \cdot \{m(m-1)\cdots(m-k+1)\}^{-1}$. This, again leads to the familiar system of equations (3.16) needed to estimate the mixing proportions, $(\alpha_1, \dots, \alpha_{c-1})$, and the probabilities, (p_1, \dots, p_c) .

Negative binomial mixtures are similarly resolved using the k th moment function, ϕ_k , given by

$$\phi_k = \sum_{i=1}^c \alpha_i \left(\frac{p_i}{q_i} \right)^k, \quad k = 0, 1, 2, \dots, 2c-1,$$

where $\phi_k = \mu'_{[k]} \cdot \{m(m+1)\cdots(m+k-1)\}^{-1}$ and $q_i = 1 - p_i$ when the negative binomial density is denoted by

$$\binom{m+x-1}{x} p^m (1-p)^x, \quad x = 0, 1, \dots$$

For **geometric** mixtures, factorial moments can be shown to have the form

$$\mu'_{[k]} = k! \cdot \sum_{i=1}^c \alpha_i \left(\frac{q_i}{p_i} \right)^k, \quad k = 1, 2, \dots$$

where again $q_i = 1 - p_i$. As usual, constructing the $2c-1$ moment equations leads to :

$$\phi_k = \sum_{i=1}^c \alpha_i \left(\frac{q_i}{p_i} \right)^k \quad k = 0, 1, 2, \dots, 2c-1$$

with $\phi_k = m_{[k]} / k!$. Progressing as described previously, this system ultimately yields the desired estimates of the component probabilities, $\{p_i\}$ and the mixing proportions, $\{\alpha_i\}$.

The following example for binomial densities clarifies the estimation process.

Example 3.2 : Mixture of Two Binomial Densities

Another approach for constructing the estimators is derived by Rider [23] for a mixture of two binomial densities. This solution is identical to the one outlined above except that in this case, traditional non-central moments are utilized instead of the factorial moments. The resulting coefficients of the estimating polynomial are given by slightly more complicated expressions than those obtained with the factorial moments. If, instead, the factorial moments were utilized in the computations, the expressions involved would be identical to those obtained in Section 3.2.1 for the gamma distribution. In this example, we have rather opted to use the non-central moments to demonstrate the power of the estimation strategy.

The two-component binomial mixture is given by

$$P(x; p_1, p_2, \alpha) = \binom{m}{x} [\alpha p_1^x (1-p_1)^{m-x} + (1-\alpha) p_2^x (1-p_2)^{m-x}], \quad x=0, 1, \dots, m,$$

where $0 \leq p_1 < p_2$.

Setting up the first three moment equations we have

$$m_1 = \alpha m p_1 + (1 - \alpha) m p_2,$$

$$m_2 = \alpha [m p_1 + m(m-1) p_1^2] + (1 - \alpha) [m p_2 + m(m-1) p_2^2],$$

$$m_3 = \alpha [m p_1 + 3m(m-1) p_1^2 + m(m-1)(m-2) p_1^3] + (1 - \alpha) [m p_2 + 3m(m-1) p_2^2 + m(m-1)(m-2) p_2^3].$$

Rearranging and manipulating the above system, we solve for p_1 and p_2 by explicitly deriving expressions for their product and sum. This yields:

$$p_1 p_2 = \frac{m_1 [1 + (m-1)(p_1 + p_2)] - m_2}{m(m-1)} \quad (3.18a)$$

and

$$(p_1 + p_2) = \frac{m(m_3 - 3m_2 + 2m_1) - (m-2)m_1(m_2 - m_1)}{(m-2)[m(m_2 - m_1) - (m-1)m_1^2]}. \quad (3.18b)$$

Whence, the component probability estimates are obtained from

$$\hat{p}_{1,2} = \frac{-\beta_1 \pm \sqrt{\beta_1^2 - 4\beta_0}}{2}. \quad (3.19)$$

Finally, the mixing proportion estimate is given by

$$\hat{\alpha} = \frac{m_1 - mp_2}{m(p_1 - p_2)}. \quad (3.20)$$

The details of the manipulation are found in [25]. ◆◆◆

Solving the above estimation problem with factorial moments is evidently less tedious. Setting $c=2$ in the generic solution, estimates of the parameters are given by equations of the form (3.13) and (3.14). The derivations for other specific two-component cases entail similar expressions.

A number of other investigators have also studied the parameter estimation of discrete distributions by the method of moments. In particular, Blischke [2], Cohen [6], and Rider [23] investigated two-component binomial mixtures. Blische [3,4] later reported a solution for general c -component binomial mixtures, while Everitt and Hand [11] gave the results for mixtures of c Poisson distributions.

3.4 Non-Integer Moments for the Exponential Family

From the above discussion, we see that both continuous and discrete mixtures can be similarly decomposed using the same generic solution. This is generally true for any mixture of densities from the exponential family in which each density is characterized by one unknown parameter, and perhaps another known parameter common to all components. The set of estimating equations given by ϕ_k ($k=0,1,\dots,2c-1$) actually constitutes a subset of a more general system

$$\phi_k = \sum_{i=1}^c A_i \cdot [f(b_i)]^k, \quad A_i \neq 0,$$

where $f(b_i)$ are strictly monotonic functions of the component parameters b_i . Indeed, if X is a random variable from the exponential family and if $Y = X^r$ for $r > 0$, we have the following result which is the generalization of Theorems 1 and 2 for mixtures of Y .

Theorem 3

Let $Y = X^r$ where $r > 0$. If X is a random variable from the exponential family, then the results of Theorems 1 and 2 hold for mixtures of distributions of the random variable Y .

Proof:

If $Y = X^r$, then the moment functions for Y are given by

$$\phi_k = \sum_{i=1}^c \alpha_i b_i^{rk},$$

or

$$\phi_k = \sum_{i=1}^c \alpha_i \xi_i^k,$$

where $\xi_i = b_i^r$ for some real constant $r > 0$. This is exactly the form of system of equations (3.7), and hence, the result. ◆◆◆

IV. EMPIRICAL STUDY

An empirical study of the parameter estimation of mixture distributions was undertaken. Included in this investigation were several randomly generated continuous and discrete distributions from the exponential family. In each case, the method of moments was implemented and applied to these mixtures. The details of this study follow.

4.1 Experimental Setup

For this study, mixtures of six different distributions from the exponential family were generated. These included three continuous density functions: exponential, gamma, and Weibull; and three discrete density functions: Poisson, binomial, and geometric. The generation of the random variables was based on two general procedures, namely, the transformation and the rejection methods. Specific algorithms for simulating distributions of each of the random variables are found in [9] and [20].

The generation of each mixture random variable, X_k ($k = 1, \dots, n$), consists of a two-step process. First, the component from which the sample value is to be drawn is selected according to a discrete PMF described by the probability vector $\underline{\alpha} = (\alpha_1, \dots, \alpha_c)$, where α_i is the i th mixing weight. Next, the random variable from the selected component density is generated. This process is repeated n times to generate the complete sample, X^n .

The estimators produced by the method of moments were evaluated by examining the mean and variance of their sampling distributions. These statistics were compared with the true parameters and mixing weights of the component densities.

The goodness of fit of the estimated distribution to the true population distribution was also evaluated by the use of two tests, namely, the chi-square (χ^2) metric and the Expected Mutual Information Measure (EMIM)³ [28] described below.

(a) The χ^2 metric

The χ^2 metric is typically utilized for comparing a “true” distribution to a given set of data. In the present study, the objective is to compare a known distribution, $f(x; \underline{\Theta}, \underline{\alpha})$, to an estimated distribution, $f(x; \hat{\underline{\Theta}}, \hat{\underline{\alpha}})$. The true PDF denotes the simulated population prescribed by the parameter vectors $\underline{\Theta}$, and $\underline{\alpha}$. The estimated PDF represents the distribution characterized by the estimators $\hat{\underline{\Theta}}$ and $\hat{\underline{\alpha}}$. These estimated parameter vectors are determined on the basis of sample data, $X^n = \{x_1, \dots, x_n\}$, drawn from the population. Hence, if the population space is subdivided into K bins, we can make use of a χ^2 -based metric defined as

$$I_\chi = \sum_{k=1}^K \frac{(p_k - \hat{p}_k)^2}{p_k},$$

where p_k is the true probability of an outcome falling in the k th bin
and \hat{p}_k is the estimated probability of an outcome falling in the k th bin.

(b) Expected Mutual Information Measure (EMIM)

Another way of expressing the goodness of fit of an estimated distribution to the “true” population distribution is through the EMIM metric [28]. This method measures an effective distance between the estimated and “true” densities. For continuous distributions, the EMIM metric is defined as

$$I[f(x; \underline{\Theta}), f(x; \hat{\underline{\Theta}})] = \int_{\mathbf{x}} f(x; \underline{\Theta}) \cdot \log \frac{f(x; \underline{\Theta})}{f(x; \hat{\underline{\Theta}})} dx,$$

where $f(x; \underline{\Theta})$ and $f(x; \hat{\underline{\Theta}})$ are the “true” and estimated mixture densities, respectively. Similarly, for discrete densities, $P(x; \underline{\Theta})$ and $P(x; \hat{\underline{\Theta}})$,

$$I[P(x; \underline{\Theta}), P(x; \hat{\underline{\Theta}})] = \sum_{\mathbf{x}} P(x; \underline{\Theta}) \cdot \log \frac{P(x; \underline{\Theta})}{P(x; \hat{\underline{\Theta}})}.$$

³ Although the EMIM metric is not used much in the statistical literature, we have included it here as a measure of goodness of fit since it is extensively used in information theory.

It can be shown that, based on the properties of PDFs (or PMFs),
 $I[f(x; \underline{\Theta}), f(x; \hat{\underline{\Theta}})] \geq 0$,
 where equality holds only when $f(x; \underline{\Theta}) = f(x; \hat{\underline{\Theta}})$ for *all* x .
 Hundreds of experiments were conducted to verify the results derived here. The details of these experimental results are found in [25] but are omitted here in the interest of brevity. In the forthcoming sections, we shall briefly report some of the results that we have obtained.

4.2 Simulation Results

4.2.1 Mixtures of Continuous Distributions

The parameter estimation of two-component exponential, gamma, and Weibull mixtures was investigated using the method of moments. For both gamma and Weibull mixtures, it was assumed that the shape parameter, a , was known and common to both components. Thus, the two scale parameters b_1 and b_2 , and the mixing weight, α , were to be estimated in all three cases.

Table 4.1 lists the mean and variance of the parameter estimates as well as the true parameter values of the continuous distributions. The mean and variance were based on ten samples of 1000 observations.

From the table, we see that the mean estimates of the exponential mixture parameters were relatively close to the true parameter values. We are unable to explain the unusually large variance in the estimate of the second scale parameter. The cause may be imputed to the small sampling proportion of the second population. In any case, the true and estimated mixture distributions are extremely close, as can be seen in Figure 4.1.

Distribution	Parameter	Mean Parameter Estimates	Variance of Parameter Estimates	True Parameter Values
Exponential	b_1	0.4393	2.903×10^{-3}	0.45
	b_2	3.3316	1.2378	3.29
	α	0.8804	2.658×10^{-3}	0.90
Gamma $a = 2$	b_1	0.1937	8.445×10^{-5}	0.20
	b_2	0.7120	1.281×10^{-2}	0.80
	α	0.8571	2.658×10^{-3}	0.90
Weibull $a = 2$	b_1	0.6863	5.368×10^{-5}	0.69
	b_2	1.6950	9.581×10^{-3}	1.69
	α	0.8979	1.782×10^{-4}	0.90

Table 4.1: The scale parameters and mixing proportion of two-component exponential, gamma, and Weibull mixtures.

The results given in Table 4.1 for gamma mixtures are for component densities with a shape parameter of $a = 2$. It should be noted that although the final estimates were reasonably close to the true parameters, the method of moments was somewhat unreliable in resolving gamma mixtures. On several occasions, indeterminate estimates were produced by the method. That is, the component density parameter estimates, \hat{b}_1, \hat{b}_2 , were found to be negative. The results obtained in these cases have not been reported.

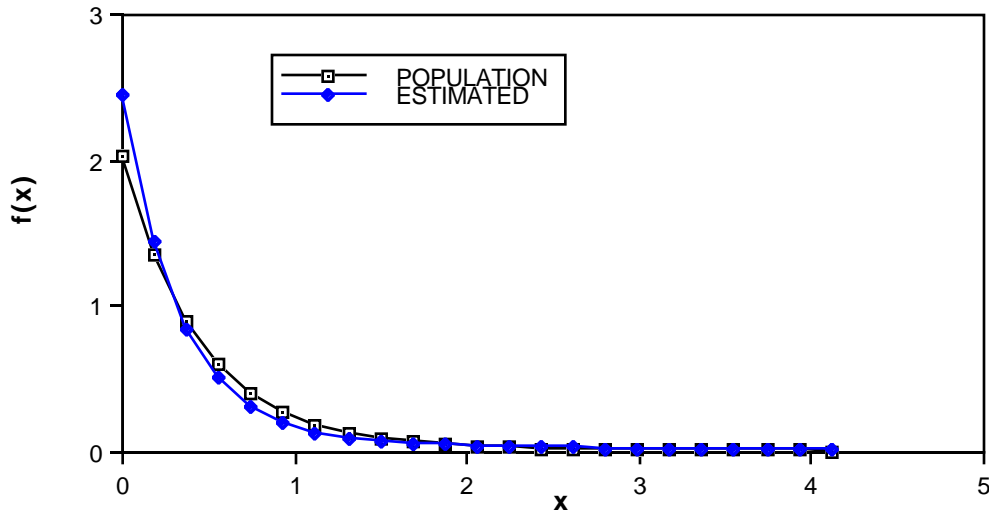


Figure 4.1: The population and estimated mixture distributions of two exponential PDFs with the parameters tabulated in Table 4.1.

The investigation of a Weibull mixture with shape parameter $a = 2$, scale parameters $(b_1, b_2) = (0.69, 1.69)$, and mixing weight $\alpha = 0.90$, yielded the estimates shown in Table 4.1. The resulting estimates were quite good for all three parameters. The largest deviation was for the estimate of the second scale parameter which had a mean value of 1.6950 and a variance of only 9.581×10^{-3} .

The χ^2 and EMIM metrics, as averaged for 10 samples of 10 different sizes ranging from 100 to 1000 observations, were calculated for all three types of mixtures. These were used to determine the influence of the number of sample observations on the moment estimates. Figure 4.2 illustrates a plot of the two metrics as a function of sample size for the exponential mixture. From this figure, the increased accuracy of the moment estimates based on larger samples is evident. The χ^2 value decreased by about 85%, while the EMIM value fell by roughly 92% in increasing the sample size from $n = 100$ to 1000.

The average metrics manifested similar behaviour to that of the exponential distribution for both the gamma and Weibull mixtures. In both cases, the goodness of fit improved significantly with increasing sample size from 100 to 1000 observations. In the first case, the approximately 75% drop in χ^2 value was noticeably smaller and less rapid than for the exponential mixture with a comparable degree of mixing, whereas in the second case, the roughly 94% decrease in χ^2 was greater (cf. Figs 4.4 and 4.6 in [25]).

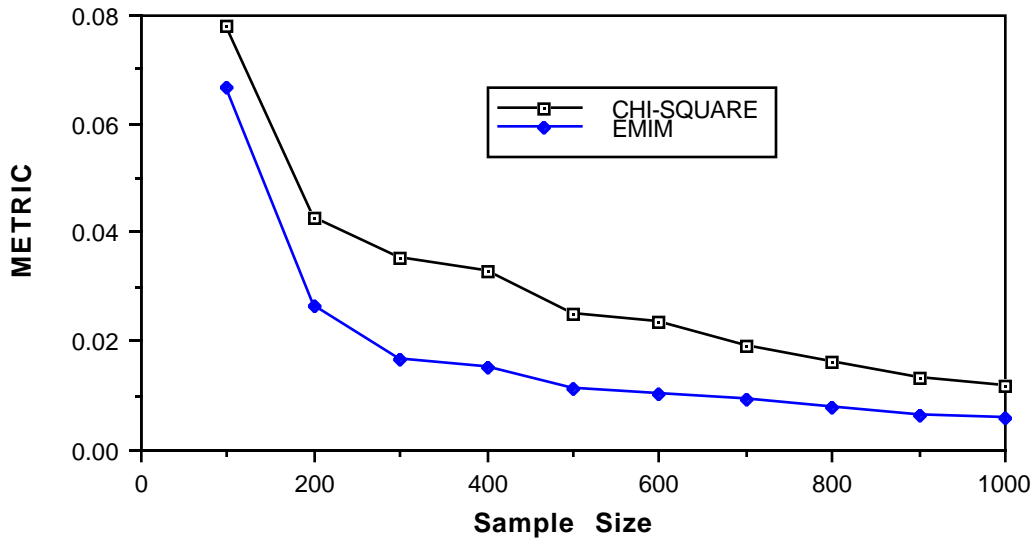


Figure 4.2: The average χ^2 and EMIM metrics as a function of sample size for the exponential distribution of Table 4.1.

4.2.2 Mixtures of Discrete Distributions

Mixtures of two-component Poisson, binomial, and geometric distributions were also simulated in order to study mixture decomposition using the method of moments. In the case of the binomial distribution, the number of independent Bernoulli trials, m , was assumed known and common to both components. Hence, in all cases, there were three mixture parameter estimators to be determined.

Similar difficulties with indeterminate estimates as those encountered with gamma mixtures arose with Poisson and geometric distributions. In the former case, a number of samples yielded negative values for the component density estimates. These estimates were excluded from the results, which were otherwise reasonably good. Figure 4.3 illustrates a histogram representing the true and estimated Poisson distributions. In the latter case, the method of moments manifested a particular difficulty in resolving the mixtures. Estimates

of both the probability of occurrence of an event, p_i , and the mixing proportion, α_i , were frequently outside the range (0,1) for sample sizes less than 500 observations. Consequently, it was necessary to draw larger samples from the population for geometric mixtures than for other types of distributions to consistently obtain determinate estimates. Specifically, the results given in Table 4.2 for these mixtures are for samples of 1400 observations, whereas 1000 observations were taken for Poisson and binomial mixtures. The mean and variance of the sampling distribution in each case was determined from ten samples.

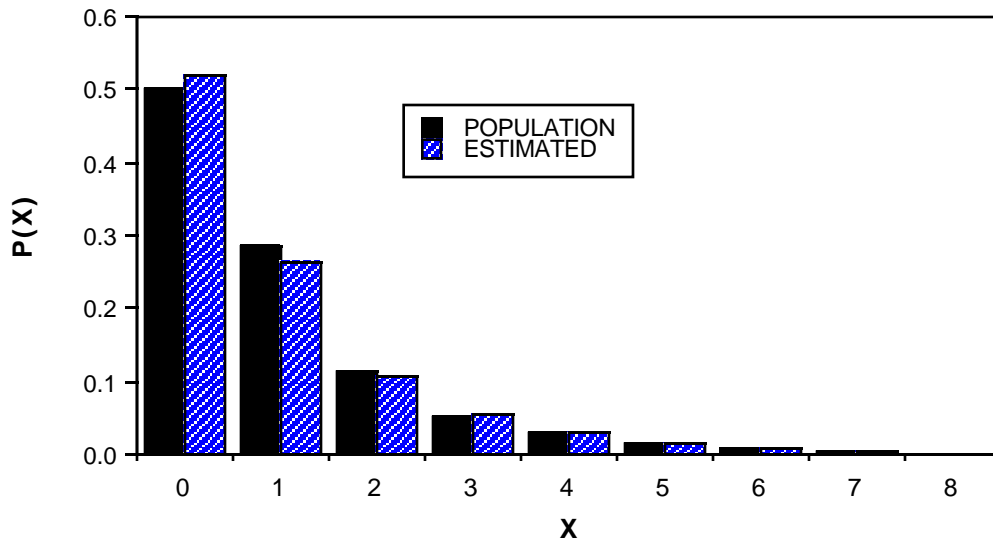


Figure 4.3: The population and estimated mixture distributions of two Poisson PMFs with the parameters tabulated in Table 4.2.

Distribution	Parameter	Mean Parameter Estimates	Variance of Parameter Estimates	True Parameter Values
Poisson	λ_1	0.4583	6.315×10^{-3}	0.50
	λ_2	2.4138	0.1050	2.50
	α	0.7776	3.825×10^{-3}	0.80
Binomial m = 10	p1	0.3970	4.575×10^{-4}	0.40
	p2	0.6957	1.992×10^{-4}	0.70
	α	0.3910	2.819×10^{-3}	0.40
Geometric	p1	0.7873	1.598×10^{-3}	0.8
	p2	0.1824	1.900×10^{-3}	0.2
	α	0.9079	8.035×10^{-4}	0.9

Table 4.2: The distribution parameters and mixing proportion of two-component Poisson, binomial, and geometric mixtures.

The number of Bernoulli trials per given binomial experiment was set at $m = 10$ for both components. The results of Table 4.2 reveal a close agreement between the moment estimates and the prescribed parameters. The estimate of the mixing proportion differed by just over 2% from the population value, while the component probability estimates differed by less than 1% from their theoretical counterparts. The variances of the estimators were in the order of 10^{-3} or less.

As was carried out for the continuous mixtures, the average EMIM metric was calculated for a range of sample sizes to evaluate the goodness of fit of the estimated mixed PMFs. Sample sizes ranged from 100 to 1000 observations for the Poisson and binomial distributions, and from 500 to 1400 observations for the geometric mixture. Each point consisted of an average of ten values.

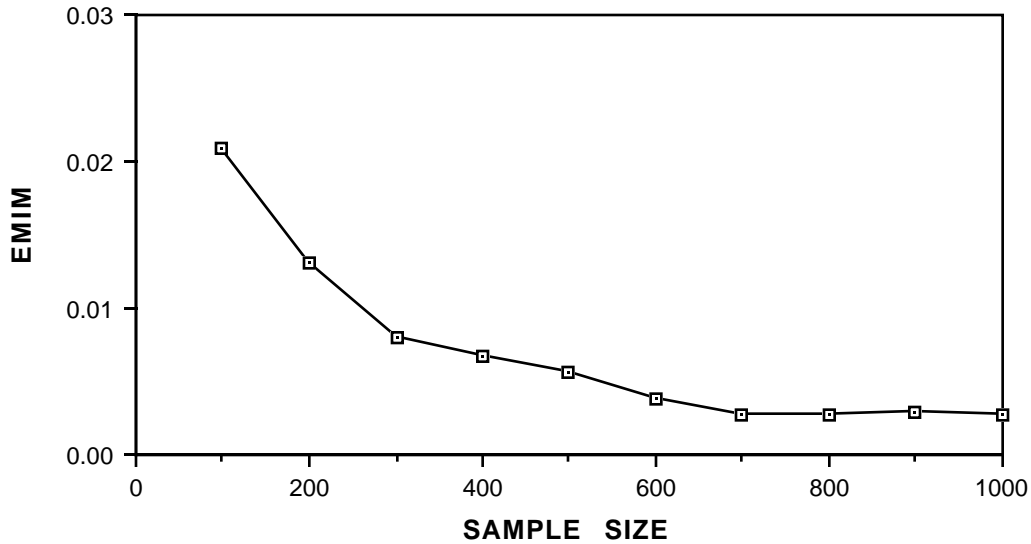


Figure 4.4: The average EMIM metric as a function of sample size for the Poisson distribution of Table 4.2.

For the Poisson distribution, the fit improved rapidly in increasing the sample size from $n = 100$ to $n = 300$. The EMIM value was observed to drop by about 62% in this short interval. In contrast, the metric remained more or less constant for sample sizes $n = 700$ to $n = 1000$. This behaviour of the EMIM is observed in Figure 4.4.

The plot of the EMIM metric as a function of sample size for the binomial mixture revealed an equally rapid decrease of 70% in the average EMIM value in the interval $n = 100$ to $n = 400$. Samples of 500 or more observations afforded only modest improvements in fit over smaller samples (cf. Fig. 4.10 in [25]). Similarly, the greatest improvement in fit for the geometric densities occurred in the initial increase of 400 observations where the

EMIM value decreased by about 50%. A comparatively smaller change was measured for larger sample sizes. As mentioned earlier, more extensive experimental results are found in [25], (see, for example, Fig. 4.12 in [25]), but are omitted here in the interest of brevity.

V. CONCLUSIONS

The focus of this paper was the parameter estimation of mixture distributions using the method of moments. This general estimation method for the exponential family was derived in detail and its application to a number of continuous and discrete distributions from the exponential family was shown. In particular, estimators for exponential, gamma, Weibull, Poisson, binomial, and geometric mixtures were constructed. The results reported in the literature are, generally speaking, special cases of the results derived here for the general exponential family.

For certain continuous distributions, the shape parameter, a , was assumed known. This assumption is reasonable since in most cases of practical interest, one has some idea of the general form of the individual distributions involved. Moreover, the component distributions are typically of the same form, and hence, the shape parameter can be assumed identical. The question of solving the problem when the shape parameters are not the same remains open. However, we believe that a **single** generic solution will not be applicable in such scenarios.

A number of experiments were performed on simulated mixtures to investigate the method of moments empirically. Overall, estimates obtained by this method agreed quite well with the true values. Certain types of mixtures were more problematic than others in that the method was prone to yielding undefined estimates. The goodness of fit of the estimated mixtures to the true distributions was quantified through the use of the χ^2 and EMIM metrics. The calculation of these quantities demonstrated that the parameter estimates improved significantly with increasing sample size.

REFERENCES

- [1] O. Barndorff-Nielsen, Identifiability of Mixtures of Exponential Families, *J. Math. Anal. Appl.*, 12:115-121 (1965).
- [2] W.R. Blischke, Moment Estimators for the Parameters of a Mixture of Two Binomial Distributions, *Ann. Math. Statist.*, 33:444-454 (1962).
- [3] W.R. Blischke, Estimating the Parameters of Mixtures of Binomial Distributions, *J. Amer. Statist. Assoc.*, 59:510-528 (1964).
- [4] W.R. Blischke, Mixtures of Discrete Distributions in *Classical and Contagious Discrete Distributions* (Ed. G.P. Patil), Pergamon Press, New York, 351-372 (1965).
- [5] C.V.L. Charlier, S.D. Wicksell, On the Dissection of Frequency Functions, *Arkiv f. Matematik Astron. och Fysik.*, Bd. 18, No. 6 (1924).

- [6] A.C. Cohen, Estimation in Mixtures of Discrete Distributions in *Classical and Contagious Discrete Distributions* (Ed. G.P. Patil), Pergamon Press, New York, 373-378.
- [7] A.C. Cohen, Estimation in Mixtures of Two Normal Distributions, *Technometrics*, 9:15-28 (1967).
- [8] N.E. Day, Estimating the Components of a Mixture of Normal Distributions, *Biometrika*, 56:463-474 (1969).
- [9] L. Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986.
- [10] A.J. Dobson, *An Introduction to Generalized Linear Models*, Chapman and Hall, London, 1990.
- [11] B.S. Everitt, D.J. Hand, *Finite Mixture Distributions*, Chapman and Hall, London, 1981.
- [12] B.S. Everitt, Mixture Distributions in *Encyclopedia of Statistical Sciences*, vol. 5, (Eds. S. Kotz and N.L. Johnson), New York, Wiley, 1985, 559-569.
- [13] S. John, On Identifying the Population of Origin of Each Observation in a Mixture of Observations from Two Normal Populations, *Technometrics*, 12:553-563 (1970).
- [14] S. John, On Identifying the Population of Origin of Each Observation in a Mixture of Observations from Two Gamma Populations, *Technometrics*, 12:565-568 (1970).
- [15] O. Kempthorne, L. Folks, *Probability, Statistics, and Data Analysis*, Iowa State University Press, Ames, 1971.
- [16] M.G. Kendall, A. Stuart, *The Advanced Theory of Statistics*, Vols.1 and 2, Charles Griffin and Company Ltd, London, 1967.
- [17] G. Lingappaiah, On the Mixture of Exponential Distributions, *Metron*, 33:403-411 (1975).
- [18] P.D.M. Macdonald, Estimation of Finite Distribution Mixtures in *Applied Statistics*, (Ed. R.P. Gupta), North-Holland, Amsterdam, 1975, 231-245.
- [19] K. Pearson, Contribution to the Mathematical Theory of Evolution, *Phil. Trans. Roy. Soc. A*, 185:71-110 (1894).
- [20] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, New York, 1986.
- [21] R.A. Redner, H.F. Walker, Mixture Densities, Maximum Likelihood, and the EM Algorithm, *SIAM Rev.*, 26:195-239 (1984).
- [22] P.R. Rider, The Method of Moments Applied to a Mixture of Two Exponential Distributions, *Ann. Math. Statist.*, 32:143-47 (1961).
- [23] P.R. Rider, Estimating the Parameters of Mixed Poisson, Binomial, and Weibull Distributions by the Method of Moments, *Bull. Inst. Internat. Stat. (Part 2)*, 33:225-232 (1961).
- [24] H.W. Sorenson, *Parameter Estimation: Principles and Problems*, Marcel Dekker, New York, 1980.
- [25] S.T. Sum, *Estimation Methods for Finite Mixture Distributions*, M. Sc. Thesis, School of Computer Science, Carleton University, Ottawa, 1993.
- [26] H. Teicher, Identifiability of Mixtures, *Ann. Math. Statist.*, 32:244-248 (1961).
- [27] D.M. Titterton, A.F.M. Smith, U.E. Makov, *Statistical Analysis of Finite Mixture Distributions*, John Wiley and Sons, Chichester, 1985.
- [28] R. S. Valiveti and B. J. Oommen, "Determining Stochastic Dependence for Normally Distributed Vectors Using the Chi-Squared Metric", *Pattern Recognition*, Vol. 26, No. 6, 1993, pp.975-987.
- [29] S.J. Yakowitz, J.D. Spragins, On the Identifiability of Finite Mixtures, *Ann. Math. Statist.*, 39:209-214 (1968).