The study of estimating the parameters of a mixture of two exponential components

by

S. D. L. Geeganage.

This thesis is submitted in partial fulfillment of the requirements for the Degree of Master of Philosophy of the Faculty of Applied Science, University of Sri Jayawardenepura, Nugegoda, Sri Lanka.

December, 1997
Registration Number : 16923

The work described in this thesis was carried out by me at the Department of Mathematics, University of Sri Jayawardenepura, Sri Lanka, under the supervision of Dr. Sunethra Weerakoon and Mr. P. Dias and, a report on this has not been submitted to any University for another degree.

December, 1997

S. D. L. Geeganage

Supervisors

Dr. Sunethra Weerakoon

Mr. P. Dias

ABSTRACT

The problem of estimating parameters of finite mixtures, is one of the oldest estimation problems. Due to the lack of a completely satisfactory solution, this problem still attracts a great deal of attention. Other than the mixtures of normal components, the most widely used mixture distributions are the mixtures of exponential components. The simplest is the mixture of two exponential components whose probability density function is given by,

$$f(x) = p\lambda_1 e^{-\lambda_1 x} + (1-p)\lambda_2 e^{-\lambda_2 x} \quad ; \quad x > 0$$
$$= 0 \quad ; \quad \text{otherwise}$$

for $\lambda_1, \lambda_2 > 0$ and $0 < p < 1$.

Mixtures of this type are frequently applied in life statistics and failure data. In this thesis, the problem of estimating parameters of a mixture of two exponential components is studied.

Our first effort, the use of the method of moments, did not give us satisfactory solutions. The simulation study has shown that the resulting estimates deviated drastically from the actual parameters. Next, the method of maximum likelihood

was applied with the following optimization techniques.

1. Nelder and Mead's method (unconstrained)
2. Newton - Raphson method (unconstrained)
3. Sequential Unconstrained Minimization Technique (SUMT) (constrained)

It was found that, SUMT is suitable to find maximum likelihood estimates of the parameters of the mixture. When the parameters of the components of the mixture are well- separated and the mixture contains not more than 70% of the observations from the component with larger mean, then the estimates seem to be accurate and precise.

ACKNOWLEDGEMENT

I wish to convey my sincere gratitude to my research supervisors Dr. Sunethra Weerakoon and Mr. P. Dias for the invaluable guidance and advice generously given to me in the process of this study, at the University of Sri Jayawardenepura. My thanks are extended to Prof. R. A. Dayananda, Dean, Faculty of Applied Science and to all the academic staff of the Department of Mathematics of the University of Sri Jayawardenepura for their guidance during this study.

I thank the Director, Cancer Hospital, Maharagama for the permission granted me to collect the data. Special thanks are due to Dr. (Mrs.) Chintha Weeraratne as well, for providing me with necessary information about cancers.

The invaluable and numerous help provided by Mr. Nimal Hettiarachchi of the University of Ruhuna, in the completion of this study is also appreciated greatfully.

The kind help given by all the staff members of the Department of Mathematics and the Computer Unit of the University of Ruhuna is greatly appreciated. Special thanks also go to Mrs. M. A. L. Kalyani, Mr. B. H. Saranapala and Mr. W. Senanayake of the Computer Unit and Mr. P. A. Jayantha of Department of Mathematics. I express my thanks to the Universities of Ruhuna and Sri Jayawardenepura for the facilities provided me, to pursue my graduate studies.

Finally, I wish to express my gratitude to all those who supported me, in various ways to achieve my objective in completing this research study and producing the thesis.

CONTENTS

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1 MIXTURE DISTRIBUTIONS

In the last century, the interest focused on studying the statistical distributions which can be expressed as the superposition of component distributions. Such superpositions are termed mixture distributions. Most research studies were concentrated on finite mixture distributions, *i.e.,* about the mixtures involving finite number of components. It is not surprising that the interest on infinite mixtures is so low, considering the fact that even the analyses of finite mixtures are still at the research stage.

Density functions of finite mixture distributions have received increasing attention in the statistical literature recently, because these finite mixture distributions have been involved in various applied fields, such as law enforcement/criminal justice, traffic engineering and bio medical studies [1]. All these models require some statistical analysis such as estimating parameters and testing hypotheses, in order to have some value in practice. Thus, this addresses the statistical problem of parameter estimation from data samples. A vital problem arises when data are not available for each marginal distribution separately, but only for the overall mixture distribution. Often such situations arise because, it is impossible to observe underlying variables which split the observations into groups and only the combined distribution can be studied. In these circumstances, the natural tendency of the statisticians is to estimate mixing proportions and parameters in the combined distributions. The problem of estimating the parameters in mixture distributions is one of the oldest estimation problems in the statistical literature. In 1894 *Pearson* [30], in 1906 *Charlier* [31], and in 1924 *Charlier* and *Wicksell* [32] were some of the earlier authors, who considered the estimation problems of mixture distributions. Due to the lack of a completely satisfactory solution, this problem still attracts a great deal of attention.

The most widely used finite mixture distributions are those involving normal components, since normal distributions often provide good approximations to the distributions of data in practice. The following two facts give heuristic justification for this.

1. When the sample size is large, the sampling distribution of the sample mean is normal, irrespective of the parent distribution (Central Limit Theorem).
2. Any linear combination of a normal mixture is normal.

However, there are circumstances when the normal distribution is inappropriate and hence it is necessary to concentrate on other distributions. In such situations, exponential distributions have appeared to have received much attention.

Very little attention has been devoted to practical applications of mixtures of continuous distributions which are not normal or exponential.

The mixtures of discrete components were also discussed in the literature. *Pearson* appears to be the first to study such distributions in detail. In 1915, he derived moment estimates for the parameters in a mixture of binomial distributions. According to the literature, extensive studies have taken place in the area of discrete mixtures of binomial and Poisson components ([2] , [3]).

Many studies of finite mixtures are devoted to the mixtures having same component distributions. There seem to be only a few applications involving mixtures, where the components are of different types. However, *Ashton* [4] applied such a mixture in studying the distribution of time gaps in road traffic flow. She used gamma distribution and displaced exponential distribution as the component distributions. Practical applications of mixtures of different types of components also seem rare in the literature.

In the study of mixture distributions, it is assumed that they are identifiable. Mixtures that are not identifiable can not be expressed uniquely as a function of components and mixing proportions. Identifiability is crucial, since it is not possible to estimate parameters for unidentifiable mixtures [5], [6] and [7]. *Teicher* [5] has shown that the mixtures of exponential components are identifiable.

1.1.1 Estimating parameters of mixture distributions

In estimating the parameters of mixtures, there is no single best estimation method. Many approaches have been devised and some standard methods are stated below.

1. **The method of moments**

   This method was introduced by *K. Pearson* in 1894 and *Rider* [10], *Muller* [33], *Gargantini* and *Henrici* [34], and

Kabir [11] were some statisticians who applied this method for mixtures.

2. **The method of maximum likelihood**

   Many statisticians applied this method in estimating the parameters of mixtures. Some of them were *Hasselblad* [12], *Day* [35], *Oppenheimer* [13], *Duda* and *Hart* [36], and *Hosmer* [37].

3. **The method of inversion and error minimization.**

   In 1968 *Kabir* [11], has applied this method to estimate parameters of exponential components.

4. **The method of Bayesian estimation**

   In 1976, *Titterington* [38] and in 1978, *Smith* and *Makov* [39] used this method in estimating parameters of a mixture of multivariate normal components.

5. **Difference methods**

   Choi and *Bulgren* [40] used the sum of squares of error function to measure the difference between the observed and theoretical cumulative distribution functions and, used it to estimate the parameters of mixtures in 1968, as a difference method.