

ESTIMATION OF THE PARAMETERS OF FINITE
LOCATION AND SCALE MIXTURES

BY

JAVAD BEHBOODIAN

TECHNICAL REPORT NO. 28

OCTOBER 1976

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)

OFFICE OF NAVAL RESEARCH

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



ESTIMATION OF THE PARAMETERS OF FINITE
LOCATION AND SCALE MIXTURES

by

Javad Behboodian

Stanford University, Stanford, California
and Pahlavi University, Shiraz, Iran

Technical Report No. 28

October 1976

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)

OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

ESTIMATION OF THE PARAMETERS OF FINITE
LOCATION AND SCALE MIXTURES

by

Javad Behboodian

Stanford University, Stanford, California
and Pahlavi University, Shiraz, Iran

I. INTRODUCTION

Let $\mathcal{F} = \{F(x-u): u \in A\}$ be a location parameter family of distributions, where A is an appropriate subset of the real line such as an interval or a countable collection of real numbers. A distribution $M(x)$ of a random variable X is called a finite location mixture on $\mathcal{F}$ if and only if there exists a positive integer k , positive constants $p_1, p_2, \dots, p_k$ with sum equal to one, and distinct distributions $F(x-u_1), F(x-u_2), \dots, F(x-u_k)$ belonging to $\mathcal{F}$, such that

$$M(x) = \sum_{i=1}^k p_i F(x-u_i) . \quad (1)$$

Suppose that U is a discrete random variable defined on A with the probability mass function

$$P(U=u_i) = p_i, \quad i=1,2,\dots,k . \quad (2)$$

If A includes zero, then the distribution $F(s)$ of a random variable, say S , generates the family $\mathcal{F}$. Now it can be shown easily, by using conditional distribution or characteristic function, that X is distributed as $S+U$ or briefly $X \stackrel{d}{=} S+U$, where S and U are independent. We assume

that the distribution of the generator random variable S is completely known and the distribution of U depends on the unknown parameters p_i 's and u_i 's.

Similarly a distribution $N(y)$ of a random variable Y is called a finite scale mixture if

$$N(y) = \sum_{i=1}^k p_i G(y/v_i) . \quad (3)$$

Here the distinct distributions $G(y/v_1), G(y/v_2), \dots, G(y/v_n)$ belong to a scale family of distributions $\mathcal{G} = \{G(y/v): v \in B\}$, where B is an appropriate subset of $(0, \infty)$; p_i 's are defined as before.

Suppose that V is a discrete positive random variable defined on B with the probability mass function

$$P(V=v_i) = p_i, \quad i=1,2,\dots,k . \quad (4)$$

If B includes 1, then the distribution $G(t)$ of a random variable, say T , generates the family $\mathcal{G}$. It is easy to show that $Y \stackrel{d}{=} TV$, where T and V are two independent random variables.

From $X \stackrel{d}{=} S+U$ we have $e^X \stackrel{d}{=} e^S e^U$, i.e., a finite location mixture can be reduced to a finite scale mixture by the exponential transformation $y = e^x$. Similarly, from $Y \stackrel{d}{=} TV$ we have $\log|Y| \stackrel{d}{=} \log|T| + \log V$, i.e., a finite scale mixture can be reduced to a finite location mixture by the logarithmic transformation $x = \log|y|$.

Representation of these finite mixtures in terms of random variables is quite useful for generating random samples and for analyzing their properties.

A finite mixture of normal distributions with the same known variance is an example of a finite location mixture, and a finite mixture of negative exponential distributions is an example of a scale mixture. The point estimation of the parameters in a mixture of two such distributions has received some attention in the literature. However, when we have a mixture of more than two of these distributions, or more generally a finite location or scale mixture, estimation becomes more difficult.

In this article we first show, by a simple argument, that finite location and finite scale mixtures are identifiable. Next, we suggest a method of moments for estimating the unknown parameters. Finally, as an example, we apply the method to a finite mixture of negative exponential distributions.

2. IDENTIFIABILITY OF FINITE LOCATION AND FINITE SCALE MIXTURES

Definition. A finite location mixture given by (1) is identifiable with respect to $\mathcal{F}$ if it has a unique representation as far as the mixed distributions, their number, and the mixing proportions are concerned. In other words,

$$M(x) = \sum_{i=1}^k p_i F(x-u_i) = \sum_{j=1}^{k'} p'_j F(x-u'_j) \quad (5)$$

implies $k = k'$ and for each i there is some j such that $p_i = p'_j$ and $u_i = u'_j$.

In terms of random variables, we can interpret this definition in the following way: if $X \stackrel{d}{=} S+U$, for the independent random variables S and U ,

and if $X \stackrel{d}{=} S+U'$, for the independent random variables S and U' , then the distribution of X is identifiable with respect to $\mathcal{I}$ if and only if $U \stackrel{d}{=} U'$.

In general, estimation problem for a mixture makes sense if the mixture is identifiable. We can easily show that a finite location mixture is identifiable by using the above interpretation. For this purpose, we denote the characteristic functions of S , U , and U' respectively by $\varphi_S(t)$, $\varphi_U(t)$, and $\varphi_{U'}(t)$. If $S+U \stackrel{d}{=} S+U'$, then

$$\varphi_S(t)\varphi_U(t) = \varphi_S(t)\varphi_{U'}(t). \quad (6)$$

It is known that a characteristic function is equal to 1 at $t = 0$, and it is a continuous function. Thus for some $\epsilon > 0$, $\varphi_S(t) \neq 0$ and

$$\varphi_U(t) = \varphi_{U'}(t) \quad (7)$$

for all t in the interval $(-\epsilon, \epsilon)$. Since by (2) all the moments of U exist, it follows from (7) that

$$E(U^n) = E(U'^n) \quad (8)$$

for all non-negative integers n . It is also easy to show that

$$\sup_n \frac{\sqrt[n]{|E(U^n)|}}{n} < \max(|u_1|, |u_2|, \dots, |u_k|) < \infty. \quad (9)$$

Now, using (8) and (9), by the moment problem (see [4] page 182), we conclude that $U \stackrel{d}{=} U'$.

The identifiability of a finite scale mixture follows immediately from the identifiability of a finite location mixture by using the logarithmic transformation we referred to in Section 1.

The identifiability of a finite location mixture has also been demonstrated by Yakowitz and Spragins [8] by using a general characterization theorem regarding the identifiability of finite mixtures.

3. ESTIMATION OF PARAMETERS

To estimate the parameters of a finite location or scale mixture we apply a moment technique similar to that used by Bliscke [3], and Rennie [5]. However, to simplify the procedure, we often use the relations $X \stackrel{d}{=} S+U$ and $Y \stackrel{d}{=} TV$.

Consider a random sample $X_1, X_2, \dots, X_n$ from the location mixture (1), and denote the r th moments of X , S , and U respectively by a_r , b_r , and c_r with $d_0 = b_0 = c_0 = 1$. We assume that all the moments in question exist. Since the distribution $M(x)$ depends on $2k-1$ unknown parameters, we use the first $2k-1$ sample moments of X , i.e.,

$$\tilde{a}_r = \sum_{i=1}^n X_i^r / n, \quad r=1,2,\dots,2k-1, \quad (10)$$

to estimate the unknown parameters $u_1, u_2, \dots, u_k, p_1, p_2, \dots, p_{k-1}$.

Taking r th moment from both sides of $X \stackrel{d}{=} S+U$ and using the independence of S and U , we have

$$a_r = \sum_{j=0}^r \binom{r}{j} b_{r-j} c_j, \quad r=0,1,\dots,2k-1. \quad (11)$$

It is clear that the matrix $B = (b_{ij})$ with

$$\begin{aligned} b_{ij} &= \binom{i}{j} b_{ij}, \quad i \geq j, i, j = 0, 1, \dots, 2k-1, \\ &= 0, \quad \text{otherwise,} \end{aligned} \quad (12)$$

is a non-singular matrix. Now, using the vectors $a = (a_0, a_1, \dots, a_{2k-1})'$, $c = (c_0, c_1, \dots, c_{2k-1})'$, and the matrix B defined by (12), the relations (11) can be written as

$$a = Bc \quad \text{or} \quad c = B^{-1}a. \quad (13)$$

Thus, from (13) and (10), we can find the moment estimates of U in terms of the moments of S , which are known, and the moment estimates of X as

$$\tilde{c} = B^{-1}\tilde{a}. \quad (14)$$

It follows from (2) that the r th moment of U is

$$c_r = \sum_{i=1}^k p_i u_i^r. \quad (15)$$

Using (15), with $r=0, 1, \dots, k-1$, we obtain

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ u_1 & u_2 & \dots & u_k \\ \vdots & \vdots & \ddots & \vdots \\ u_1^{k-1} & u_2^{k-1} & \dots & u_k^{k-1} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{k-1} \end{bmatrix}. \quad (16)$$

The parameters $u_1, u_2, \dots, u_k$ can be assumed to be the zeros of the polynomial

$$P(x) = \prod_{i=1}^k (x - u_i) = x^k - \alpha_1 x^{k-1} - \dots - \alpha_{k-1} x - \alpha_k. \quad (17)$$

We now find the coefficients $\alpha_1, \alpha_2, \dots, \alpha_k$ in terms of $c_1, c_2, \dots, c_{2k-1}$.

It is obvious that $U^\ell P(U)$ is a zero random variable for any non-negative integer ℓ . Hence, we have $E(U^\ell P(U)) = 0$ or

$$\alpha_k c_\ell - \alpha_{k-1} c_{\ell+1} + \dots + \alpha_1 c_{\ell+k-1} = c_{\ell+k} \quad \text{for } \ell=0, 1, \dots, k-1. \quad (18)$$

From (18) we have the matrix equation

$$\begin{bmatrix} c_0 & c_1 & \dots & c_{k-1} \\ c_1 & c_2 & \dots & c_k \\ \vdots & \vdots & \ddots & \vdots \\ c_{k-1} & c_k & \dots & c_{2k-2} \end{bmatrix} \begin{bmatrix} \alpha_k \\ \alpha_{k-1} \\ \vdots \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} c_k \\ c_{k+1} \\ \vdots \\ c_{2k-1} \end{bmatrix}. \quad (19)$$

The matrix in (19), by using (15), can be expressed as WDW' where $W = (w_{ij})$, with $w_{ij} = u_j^i$ for $i=0, 1, \dots, k-1$, $j=1, 2, \dots, k$, is the Vandermonde matrix and D is the diagonal matrix with $p_1, p_2, \dots, p_k$ on the main diagonal. Since $p_i > 0$ and $u_i \neq u_j$ for $i \neq j$, we conclude that all these matrices are non-singular. Therefore, by solving (19) we can find the coefficients of the polynomial (15) and then the roots of $P(x) = 0$, i.e., u_i 's in terms of c_i 's. To identify these roots, we can assume, without loss of generality, that $u_1 < u_2 < \dots < u_k$. Substituting the moment estimates of c_i 's in (19), we find $\tilde{u}_i$'s, i.e., the moment estimates of u_i 's. It should be noted that there is no guarantee that we have distinct real $\tilde{u}_i$'s.

However, since the sample moments a_i 's converge to a_i 's with probability one, the chance of having imaginary or equal $\tilde{u}_i$'s tends to zero as sample size n tends to infinity.

Replacing the moment estimates of c_i 's and u_i 's in (16), we obtain the moment estimates of p_i 's.

In practice we may not know the distribution of S ; however, it is just enough to have some prior knowledge about the first $2k-1$ moments of S . For example, suppose that a certain material is produced by k machines whose products flow automatically into a common shipping room. Let X be some characteristic of the material and assume that X has the same distribution for all machines when they are new. Thus, we can say the distribution of X and S are identical at the beginning, and we can find the sample moments of S from some observed values of X . After a period the machines become old and there would be some different shifts in X from machine to machine. Now, X for the products in the shipping room has a finite location of mixtures. Observing a random sample and using the prior knowledge about S , we can estimate the unknown parameters by the above procedure.

To estimate the parameters of a finite scale mixture, we first find the moment estimates of V by using the relation $Y \stackrel{d}{=} TV$ and then we apply the above procedure. Actually, since T and V are independent we have

$$E(V^r) = E(Y^r)/E(T^r) . \quad (20)$$

If $E(T^r) = 0$ for some of the required r 's, we cannot use (20).

Here we transform the scale mixture to a location mixture by using the logarithmic transformation we referred to in Section 1. An example of this case is the normal mixture

$$M(x) = \sum_{i=1}^k p_i N(0, \sigma_i^2) , \quad (21)$$

in which T , the standard normal variable, has zero odd moments.

4. A FINITE MIXTURE OF NEGATIVE EXPONENTIAL DISTRIBUTIONS

As an example we apply the above method to find the moment estimates of a finite mixture of negative exponential distributions with probability density function

$$f(y) = \sum_{i=1}^k p_i f_i(y) , \quad (22)$$

where

$$f_i(y) = \exp(-y/v_i)/v_i . \quad (23)$$

This distribution is often used for the analysis of the completed length of service [2], and it has also some application in reliability and life testing [1]. For the case $k=2$, the moment estimation of the parameters has already received some attention, for example, by Rider [5] and Tallis and Light [6]. But for $k > 2$ the problem becomes more involved and there is no practical method to estimate the parameters.

The function (22) is the density of the random variable $Y \stackrel{d}{=} TV$, where T has the probability density function $\exp(-t)$ for $t \geq 0$ and V has the probability mass function given by (4). We know that

$$E(T^r) = \Gamma(r+1) = r! \quad (24)$$

for any non-negative integer r . Now, consider a random sample $Y_1, Y_2, \dots, Y_n$ from a distribution with density (22). Taking

$$c_r = E(V^r) = E(Y^r)/E(T^r) = E(Y^r)/r! , \quad (25)$$

Y_i for X_i , and v_i for u_i , we can find the moment estimates by using (15)-(19).

It is not difficult to show that the sampling variability of the sample moments of the density (23) becomes larger and larger as we use higher moments. This behavior of sample moments reduces the efficiency of the estimates. We may modify the method by taking r th fractional moment from both sides of $Y \stackrel{d}{=} TV$. To estimate the parameters, for v_i we take $w_i = v_i^{1/(2k-1)}$ and for r we use $1/(2k-1), 2/(2k-1), \dots, (2k-2)/(2k-1), 1$. Application of fractional moments may be also useful when T does not have integral moments.

References

- [1] Barlow, R. E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing. Holt, Rinehart and Winston, Inc., New York.
- [2] Bartholomew, D. J. (1973). Stochastic Models for Social Sciences, 2nd ed. John Wiley and Sons, New York.
- [3] Blischke, W. R. (1962). Estimating the parameters of mixtures of binomial distributions. JASA, 59, 510-528.
- [4] Breiman, L. (1968). Probability. Addison-Wesley Publishing Company, Reading, Mass.
- [5] Rennie, R. R. (1974). An identification algorithm for finite mixtures of distributions with common central moments. Sāṅkhyā, 36, 315-320.
- [6] Rider, P. L. (1961). The method of moments applied to a mixture of two exponential distributions. Ann. Math. Stat., 32, 143-147.
- [7] Tallis, G. M. and Light, R. (1968). The use of fractional moments for estimating the parameters of a mixed exponential distributions. Technometrics, 10, 161-175.
- [8] Yakowitz, S. and Spragins, J. (1968). On the identifiability of finite mixtures. Ann. Math. Stat., 40, 1728-1735.

TECHNICAL REPORTS

OFFICE OF NAVAL RESEARCH CONTRACT N00014-67-A-0112-0030 (NR-042-034)

1. "Confidence Limits for the Expected Value of an Arbitrary Bounded Random Variable with a Continuous Distribution Function," T. W. Anderson, October 1, 1969.
2. "Efficient Estimation of Regression Coefficients in Time Series," T. W. Anderson, October 1, 1970.
3. "Determining the Appropriate Sample Size for Confidence Limits for a Proportion," T. W. Anderson and H. Burstein, October 15, 1970.
4. "Some General Results on Time-Ordered Classification," D. V. Hinkley, July 30, 1971.
5. "Tests for Randomness of Directions against Equatorial and Bimodal Alternatives," T. W. Anderson and M. A. Stephens, August 30, 1971.
6. "Estimation of Covariance Matrices with Linear Structure and Moving Average Processes of Finite Order," T. W. Anderson, October 29, 1971.
7. "The Stationarity of an Estimated Autoregressive Process," T. W. Anderson, November 15, 1971.
8. "On the Inverse of Some Covariance Matrices of Toeplitz Type," Raul Pedro Mentz, July 12, 1972.
9. "An Asymptotic Expansion of the Distribution of "Studentized" Classification Statistics," T. W. Anderson, September 10, 1972.
10. "Asymptotic Evaluation of the Probabilities of Misclassification by Linear Discriminant Functions," T. W. Anderson, September 28, 1972.
11. "Population Mixing Models and Clustering Algorithms," Stanley L. Sclove, February 1, 1973.
12. "Asymptotic Properties and Computation of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," John James Miller, November 21, 1973.
13. "Maximum Likelihood Estimation in the Birth-and-Death Process," Niels Keiding, November 28, 1973.
14. "Random Orthogonal Set Functions and Stochastic Models for the Gravity Potential of the Earth," Steffen L. Lauritzen, December 27, 1973.
15. "Maximum Likelihood Estimation of Parameters of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure," T. W. Anderson, December, 1973.
16. "Note on a Case-Study in Box-Jenkins Seasonal Forecasting of Time Series," Steffen L. Lauritzen, April, 1974.

TECHNICAL REPORTS (continued)

17. "General Exponential Models for Discrete Observations,"
Steffen L. Lauritzen, May, 1974.
18. "On the Interrelationships among Sufficiency, Total Sufficiency and
Some Related Concepts," Steffen L. Lauritzen, June, 1974.
19. "Statistical Inference for Multiply Truncated Power Series Distributions,"
T. Cacoulllos, September 30, 1974.

Office of Naval Research Contract N00014-75-C-0442 (NR-042-034)

20. "Estimation by Maximum Likelihood in Autoregressive Moving Average Models
in the Time and Frequency Domains," T. W. Anderson, June 1975.
21. "Asymptotic Properties of Some Estimators in Moving Average Models,"
Raul Pedro Mentz, September 8, 1975.
22. "On a Spectral Estimate Obtained by an Autoregressive Model Fitting,"
Mituaki Huzii, February 1976.
23. "Estimating Means when Some Observations are Classified by Linear
Discriminant Function," Chien-Pai Han, April 1976.
24. "Panels and Time Series Analysis: Markov Chains and Autoregressive
Processes," T. W. Anderson, July 1976.
25. "Repeated Measurements on Autoregressive Processes," T. W. Anderson,
September 1976.
26. "The Recurrence Classification of Risk and Storage Processes,"
J. Michael Harrison and Sidney I. Resnick, September 1976.
27. "The Generalized Variance of a Stationary Autoregressive Process,"
T. W. Anderson and Raul P. Mentz, October 1976.
28. "Estimation of the Parameters of Finite Location and Scale Mixtures,"
Javad Behboodian, October 1976.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 28	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ESTIMATION OF THE PARAMETERS OF FINITE LOCATION AND SCALE MIXTURES		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Javad Behboodian		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0442
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, California 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-034)
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 436 Arlington, Virginia 22217		12. REPORT DATE October 1976
		13. NUMBER OF PAGES 11
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Location and scale mixtures, identifiability, estimation of parameters.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A finite location mixture is represented as the sum of two independent random variables, and a finite scale mixture is represented as the product of two independent random variables. Using these representations, it is shown how to generate random samples from these mixtures, how to prove their identifiabilities, and how to estimate the unknown parameters.		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)