MIXTURES OF DISTRIBUTIONS

ELIAS A. PARENT, JR.

MIXTURES OF DISTRIBUTIONS

* * * * *

ELIAS A. PARENT, JR.

MIXTURES OF DISTRIBUTIONS

by

Elias A. Parent, Jr.

Lieutenant, Supply Corps

United States Navy

Submitted in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE
with major in
Mathematics

United States Naval Postgraduate School
Monterey, California

1962

MIXTURES OF DISTRIBUTIONS

by

Elias A. Parent, Jr.

This work is accepted as fulfilling
the thesis requirements for the degree of
MASTER OF SCIENCE
with major in
Mathematics
from the
United States Naval Postgraduate School

ABSTRACT

If we assume that a population of elements is made up of several subgroups, each subgroup with its own underlying distribution, and the several subgroups mixed together according to certain proportions, we would have an instance of a mixture of distributions; i.e., the underlying distribution for the entire population would be a mixture of the distributions for each subgroup.

A study is made of the more recent developments in the theory of mixtures of distributions. The problem of identifiability in mixtures is considered in some detail. The special cases of linear mixtures and the distribution of sums of independent random variables are also considered. Finally, the problems encountered in estimation of parameters in mixtures are discussed.

TABLE OF CONTENTS

MIXTURES OF DISTRIBUTIONS

1. Introduction.

There exists a considerable body of literature relative to the theory of mixtures of probability distributions, and several results have been published relating to the statistical estimation of parameters when the underlying distribution has been assumed to be a mixture of distributions. There seems to be a growing interest in this problem, and one is certainly justified in studying it in its general form inasmuch as the general theory includes as a special case the classical statistical assumption of a single underlying distribution function for the population under study.

By way of introduction, we will consider some specific examples to show how mixtures of distributions come up quite naturally in statistical investigations.

Historically, the problem seems to have been studied first by Karl Pearson about 1894 [8].[1] He noticed that data (measurements) taken on various collections of biological specimens did not agree too well with the Gaussian distribution when plotted in histogram form. It was quite apparent in many instances that a definite bimodality existed where one would have expected unimodality. Pearson postulated that the underlying density function was of the following form

[1]Numbers in square brackets refer to bibliography.

$$f(x) = \frac{\alpha}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{x-\mu_1}{\sigma_1}\right]^2 + \frac{1-\alpha}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{x-\mu_2}{\sigma_2}\right]^2$$

and he tried to estimate the parameters α, μ_1, μ_2, σ_1, and σ_2 using the method of moments. He was led to an equation of ninth degree and had considerable difficulty calculating the roots of the polynomial. Pearson called this a problem of "dissection." His aim was to "dissect" the mixture of these two normal density functions into its components and then try to infer what could have caused such a mixture.

As a second example of mixtures of distributions, we draw on a familiar problem in life testing or reliability theory. It has been observed that in life tests of electron tubes the initial failure rate is relatively high and decreases as the population under test ages. In general, the failure rate becomes constant for a time and then increases with age. Such a behavior suggests that the population might be a mixture of several subpopulations and that the underlying distribution function might be a linear sum of several distribution functions.

Of further interest in devices such as electron tubes is the phenomenon that devices fail for different reasons and such a population of elements could be classified according to cause of failure. Then, assuming the underlying population is composed of such a mixture, one might try to estimate, from a sample of failures classed as to cause of failure, the proportion which will fail due to each cause in order to determine an allocation of research effort in improving the device.

In statistical decision theory, as developed by Wald, we find, for example, in the case of a stochastic process where the random variables are assumed to be identically and independently distributed according to $F(x;\theta)$, that θ is also assumed to be a random variable with its own probability law $G(\theta)$. Under this assumption, the random variables are in reality assumed to be distributed according to

$$H(x) = \int F(x;\theta)dG(\theta).$$

A special case of the mixture problem may be viewed as follows: suppose we assume that the population under investigation has an underlying distribution of known form $F(x;\theta)$ and that the parameter is also a random variable with distribution $G(\theta)$. If we further postulate that $G(\theta_0) = Pr[\theta=\theta_0]=1$, then the underlying distribution is

$$\int F(x;\theta)dG(\theta) = F(x;\theta_0).$$

What we have done here is tantamount to assuming that the underlying distribution is of a specified form, with θ a fixed value not subject to variation (in a probabilistic sense), and this amounts to assuming that the distribution is, say, normal with mean μ_0 and standard deviation σ_0, or exponential with parameter θ_0.

In this paper we propose to discuss the theory of mixtures of distribution from a far less general point of view than is found in the literature. We will illustrate the theoretical results with specific examples. We first discuss the general theory and state some useful results. We then consider the problem of identifiability and using some of the results in this area determine specifically which of the more standard distributions are identifiable. We then look at a special class of mixing distributions and determine some algebraic properties of the induced class of mixtures. A result analogous to the classical reproductive property of certain distributions is presented for a certain class of mixtures in Section 7. We then take up the problem of estimation of parameters in mixtures of distributions.

2. Theory.

By way of notation we let $\mathcal{J} = \left\{ F(x;\alpha) : \alpha \in E^m \right\}$ denote a family of one-dimensional distribution functions indexed by a real m-dimensional vector α, where E^m denotes Euclidean m-space. Although this development is restricted to one-dimensional distribution functions, the extension to n-dimensional distribution functions may be obtained in the usual manner. Let x be a point in E^1 and let B denote the σ-field of Borel sets in E^1. Define $Sx = \left\{ w : w \leq x \right\}$ and let μ be any probability measure on B. Then the function

$$F(x) = \mu(Sx).$$

is the distribution function corresponding to μ. Conversely, if F(x) is any distribution function in E^1, there is a unique probability measure μ on B such that

$$F(x) = \mu(S_x).$$

We denote the operation of Lebesque-Stieltjes integration relative to the measure μ by

$$\int_{E^1} f(x)d\mu = \int_{-\infty}^{\infty} f(x)dF(x).$$

However, all the results that follow may be read with integration in the Riemann-Stieltjes sense with little or no modification to the hypotheses of the theorems.

To illustrate this notation, we might consider the family of exponential distribution functions (d.f.'s)

$$\mathcal{H} = \left\{F(x;\alpha) = 1-\exp[-\alpha x] : \alpha > 0 \quad x \geq 0\right\}.$$

In this case α is one-dimensional and restricted to positive values. Each value of α determines one specific d.f. in the family and $\mathcal{H}$ consists of all such d.f.'s.

Definition 1. If G is a d.f. defined over E^m, then

$$H(x) = \int F(x;\alpha)dG(\alpha)$$

is called a mixture of the family $\mathcal{H} = \left\{F(x;\alpha)\right\}$, and more specifically a G-mixture of $\mathcal{H}$.

Definition 2. A G-mixture of $\mathcal{H}$, say H, will be called identifiable if, for any d.f. G^* we have

$$H(x) = \int F(x;\alpha)dG(\alpha) = \int F(x;\alpha)dG^*(\alpha)$$

implies that $G=G^*$.

If we consider a class $\mathcal{G} = \{G\}$ of mixing distributions and let $\mathcal{H}$ be a family of mixtures of the form $H = \int F dG$ where $F \in \mathcal{F}$ and $G \in \mathcal{G}$, then $\mathcal{H}$ will be called identifiable if every member H of $\mathcal{H}$ is identifiable. The mixing distribution G may be either discrete, continuous or a combination of both. In general, the cases which are useful in statistics are those where G is either entirely discrete or continuous; and in what follows, we have these cases in mind.

Definitions 1 and 2 really form the basis of this discussion inasmuch as they delineate the two general areas of interest in the theory of mixtures of distributions. From the mathematical-probabilistic point of view, properties of the mixtures H are studied when special properties are attributed to the class $\mathcal{F}$ or the class of mixing distributions $\mathcal{G}$, or both. The question of identifiability must be answered before meaningful statements (statistical) can be made relative to the parameter α. Proofs of the results cited in what follows may be found in the indicated references. Proofs will be given when it is thought useful and in those cases where theorems have been modified or extended.

3. General Results.

If we let $\mathfrak{D}$ denote the space of all probability distribution functions, we may consider the definition of a mixture to be a transformation of an element $F \in \mathfrak{D}$, relative to another member of $\mathfrak{D}$, say G, under the mapping $H = \int F dG$. To be useful in probability and statistical theories, it would be desirable to have the range of such a transformation be a subset of $\mathfrak{D}$. Robbins [12] proves, in general, that this is indeed the case, and we have

Theorem 1. Let $\mathfrak{F} = \{F(x;\alpha) : \alpha \in E^m\}$ be a family of n-dimensional d.f.'s and let G be a d.f. defined in E^m. Then the function $H(x) = \int_{-\infty}^{\infty} F(x;\alpha)dG(\alpha)$ is a distribution function in E^n.

As noted in the introduction, when a certain form is assumed for the underlying probability distribution in a statistical investigation, the idea embodied in definition 1 is really occurring. Such as assumption amounts to specifying a mixing distribution G relative to a family $\mathfrak{F}$. When one assumes that the underlying distribution is normal with mean μ_0, and standard deviation σ_0, one is choosing from the class of all mixing distributions a d.f. G which concentrates all its mass at a single point (μ_0, σ_0) in E^2, and we have

$$H(x) = \int \Phi(\mu,\sigma)dG = \Phi(\mu_0,\sigma_0)$$

where $\Phi(\mu,\sigma)$ is a generic element from the family of normal d.f.'s. Theorem 1 assures us that under more general conditions (i.e., more general mixing distributions) the closure property holds.

The characteristic function (c.f.) corresponding to any d.f. is defined by

$$\psi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

where F is defined in E^1. As a result of the properties of the Fourier integral, it is known that there is a one-to-one correspondence between distribution functions and characteristic functions.

We next present some theorems concerning the structure of the characteristic function, moments, and density function of a mixture.

Theorem 2. If H is a G-mixture of $\mathcal{J} = \{F(x;\alpha)\}$ and $\psi(t)$, $\psi(t;\alpha)$ are the c.f.'s of H and $F(x;\alpha)$, respectively, then $H(x) = \int F(x;\alpha)dG(\alpha)$ if, and only if, $\psi(t) = \int \psi(t;\alpha)dG(\alpha)$.

Proof: Suppose $H(x) = \int F(x;\alpha)dG(\alpha)$. Then, since $|e^{itx}| \leq 1$ we can use theorem 5 from Robbins [12] to ensure the following steps are valid:

$$\psi(t) = \int_{-\infty}^{\infty} e^{itx} dH(x) = \int_{-\infty}^{\infty} e^{itx} d_x \left\{ \int_{-\infty}^{\infty} F(x;\alpha)dG(\alpha) \right\}$$

$$= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} e^{itx} d_x F(x;\alpha) \right\} dG(\alpha)$$

$$= \int \psi(t;\alpha)dG(\alpha).$$

If $\psi(t) = \int \psi(t;\alpha)dG(\alpha)$ then, using the same theorem

$$\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} e^{itx} dF(x;\alpha)\right\} dG(\alpha)$$

$$= \int_{-\infty}^{\infty} e^{itx} d_x \left\{\int_{-\infty}^{\infty} F(x;\alpha) dG(\alpha)\right\}$$

$$\int_{-\infty}^{\infty} e^{itx} dH(x) = \psi(t)$$

but this shows that $H(x) = \int F(x;\alpha)dG(\alpha)$ on all but sets of measure zero.

Theorem 3. If $H(x) = \int F(x;\alpha)dG(\alpha)$ then any existing moment of H is a G-mixture of the family of moments (of the same order) of $\mathcal{F}$.

Proof: Let m_r be the $r^{\underline{th}}$ moment of H and $m_r(\alpha)$ the $r^{\underline{th}}$ moment of $F(x;\alpha)$ and assume m_r exists. Then

$$m_r = \int_{-\infty}^{\infty} x^r dH(x) = \int_{-\infty}^{\infty} x^r d_x \left\{\int_{-\infty}^{\infty} F(x;\alpha)dG(\alpha)\right\}$$

$$= \int_{-\infty}^{\infty}\left\{x^r d_x F(x;\alpha)\right\} dG(\alpha) = \int_{-\infty}^{\infty} m_r(\alpha)dG(\alpha).$$

Theorem 4. Let $H(x) = \int F(x;\alpha)dG(\alpha)$ and suppose $F(x;\alpha)$ is absolutely continuous. Let $f(x;\alpha) = \frac{\partial F(x;\alpha)}{\partial x}$.

Then the density function $h(x) = \frac{\partial}{\partial x} H(x)$ is given by $\int f(x;\alpha)dG(\alpha)$.

4. Identifiability.

Suppose we consider the case where the underlying distribution is a mixture of two binomial distributions. We assume the probability of success in the first population is p_1 and in the second, p_2 and that each population is well mixed with the other to form the total population. We assume that the proportion of elements from the first population is α where $0<\alpha<1$. The probability of success from such a mixture is $\alpha p_1+(1-\alpha)p_2 = p$; and if n independent trials are made, we have

$$\Pr[k \text{ successes}] = \binom{n}{k} p^k (1-p)^{n-k}$$

where the distribution is again binomial. As will be shown later, such a mixture is not identifiable. Using a sample from this mixture, we could estimate the parameter p, but not the parameters p_1, p_2, and α. The sampling scheme can be reformulated in some cases and estimators constructed for the individual population parameters (see Blischke [1]); however, it is not immediately obvious how this could be done in all cases of mixtures.

This leads us to the study of what properties a family $\mathcal{H} \doteq \{F(x;\alpha)\}$ must possess to lead to identifiable mixtures. We let D stand for an Abelian semigroup under addition and use D(I) to mean the integers, $D(I_+)$ the positive integers, and r and R to denote the rationals and reals, respectively.

Definition 3. A family of distribution functions $\mathcal{H} = \{F(x;\alpha);\ \alpha \in D\}$ is called additively closed if for each $\alpha, \beta \in D$ we have

$$F(x;\alpha) * F(x;\beta) = F(x;\alpha+\beta)$$

where * denotes convolution.

Additively closed families of distributions occur quite frequently in applications inasmuch as the normal, binomial, Poisson, gamma, and other distributions have the property. Of course, in random sampling the importance lies in the fact that the distribution function of the random variable $Z = X+Y$, where X and Y are independent random variables, is equal to the convolution of the distribution functions of X and Y.

Theorem 5. If F, G, and H are distribution functions in E^1 and $\psi_1(t)$, $\psi_2(t)$, and $\psi(t)$ the corresponding ch. fcn's, then $H(x) = F(x)*G(x)$ iff $\psi(t) = \psi_1(t)\psi_2(t)$. (Robbins [12]).

One of the uses of theorem 5 is the determination of families which are additively closed. As an example, we consider the family of normal distribution functions $\{F(x;\mu,\sigma)\}$. The corresponding class of characteristic functions is $\{e^{it\mu - \frac{1}{2}t^2\sigma^2}\}$. Then

$$H(x) = F(x;\mu_1,\sigma_1) * F(x;\mu_2,\sigma_2)$$

has the characteristic function

$$\psi(t) = e^{it\mu_1 - \frac{1}{2}t^2\sigma_1^2} e^{it\mu_2 - \frac{1}{2}t^2\sigma_2^2} = e^{it(\mu_1+\mu_2) - \frac{1}{2}t^2(\sigma_1^2+\sigma_2^2)}.$$

which is again the characteristic function of a normal distribution function. So $H(x;\mu_1+\mu_2,\sigma_1^2+\sigma_2^2) = F(x;\mu_1,\sigma_1) * F(x;\mu_2,\sigma_2)$ and the class of normal d.f.'s is additively closed.

Teicher [15] determined that the class of mixtures of a one-parameter family of additively-closed distributions is identifiable, and he gave conditions under which a class of scale or translation parameter mixtures is identifiable. We summarize these results in what follows.

Theorem 6. If m = 1 and D is $D(I_+)$, $D(r_+)$, or $D(R_+)$, the class of mixtures $\left\{\int_D F(x;\alpha)dG(\alpha)\right\}$ of an additively closed family $\left\{F(x;\alpha): \alpha \varepsilon D\right\}$ is identifiable.

The class of scale parameter mixtures consists of mixtures of the form $\left\{\int_0^\infty F(x\alpha)dG(\alpha)\right\}$ and the class of translation parameter mixtures are those of the form $\left\{\int_0^\infty F(x-\alpha)dG(\alpha)\right\}$.

Theorem 7. Let F be a d.f. which generates a family $\left\{F(x;\alpha)\right\}$ via a scale change such that $F(0^+) = 0$. If the Fourier transform of $\overline{F}(y) = F(e^y)$ is not identically zero in some non-degenerate real interval, the class of scale parameter mixtures is identifiable.

Proof: Let $x = e^y$ and $\alpha = e^{-\beta}$. Let $H(x) = F*G_1 = F*G_2$. Letting $\overline{F}(y) = F(e^y)$, $\overline{G}(\beta) = 1-G(e^{-\beta})$, we have for $-\infty < y, \beta < \infty$

$$\overline{F}*\overline{G} = \int_{-\infty}^{\infty} \overline{F}(y-\beta)d\overline{G}(\beta) =$$

$$\int_{-\infty}^{\infty} F(e^y e^{-\beta})d(1-G(e^{-\beta})) = \int_{0}^{\infty} F(x\alpha)dG(\alpha) = H(x).$$

Hence, $F*G_1 = F*G_2 \Rightarrow \overline{F}*\overline{G}_1 = \overline{F}*\overline{G}_2 = H(x)$; and since $\overline{F}$ and G_i $i = 1, 2$ are d.f.'s, we have, using theorem 5, $\psi_{\overline{F}}\cdot\psi_{\overline{G}_1} = \psi_{\overline{F}}\cdot\psi_{\overline{G}_2}$; and since $\psi_{\overline{F}}(t) = \int_{-\infty}^{\infty} e^{itx}d\overline{F}(x)$ is not identically zero (except possibly on a set of measure zero), then $\psi_{\overline{G}_1} = \psi_{\overline{G}_2}$ and $\overline{G}_1 = \overline{G}_2 \Rightarrow 1-G_1(e^{-\beta}) = 1-G_2(e^{-\beta}) \Rightarrow G_1(\alpha) = G_2(\alpha)$ and the class of scale parameter mixtures is identifiable.

Theorem 8. Let F be a d.f. which generates a family $\{F(x;\alpha)\}$ via a location change such that $F(0^+) = 0$. If the Fourier transform of $F(x)$ is not identically zero in some non-degenerate real interval, the class of translation parameter mixtures is identifiable.

Proof: The proof is essentially the same as in theorem 7. Note that we assume the mixing distribution is on the translation parameter only.

When we consider the class of mixtures of a specific family of d.f.'s, we can divide the class of mixing d.f.'s, $\mathcal{D}$, into two mutually exclusive classes. Let $\mathcal{D}_1$ denote the subclass of $\mathcal{G}$ which is composed of only those d.f.'s which concentrate all their mass at a single point in E^m, where $\alpha \varepsilon E^m$. Elements of this set are called degenerate m-dimensional d.f.'s. The remaining elements in $\mathcal{G}$, i.e., $\mathcal{G} - \mathcal{D}_1$, yield, in the strict sense of the word, mixtures of $\mathcal{F}$.

Now suppose we had a class $\mathcal{F}$ and the induced class of mixtures $\mathcal{H}$ was identifiable. Let $G^* \varepsilon \mathcal{G} - \mathcal{D}_1$ and let $H(x) = \int F(x;\alpha)dG^*(\alpha)$. If H is in the class $\mathcal{F}$, say $H(x) = F(x;\alpha^*)$, then the d.f. $G \varepsilon \mathcal{D}_1$, which concentrates all its mass at α^*, yields $H(x) = \int F(x;\alpha)dG(\alpha) = F(x;\alpha^*)$. But this means $G = G^*$, since $\mathcal{H}$ is identifiable, and clearly this is impossible. So we have

Theorem 9. Let $\mathcal{H}$ be identifiable with respect to $\mathcal{F}$. Then, no non-degenerate mixture of $\mathcal{F}$ is an element of $\mathcal{F}$.

This result establishes a necessary condition for identifiability. If we can find a non-degenerate mixture of a class such that the resulting mixture is again a member of the class, we know the class of mixtures is not identifiable.

Theorem 10. Let H_i be a G_i-mixture of $\mathcal{F} = \{F(x;\alpha)\}$, $i = 1,2$. Then, $H_1 * H_2(x) = \int F(x;\alpha)d(G_1 * G_2)(\alpha)$ if, and only if, $\mathcal{F}$ is additively closed.

Proof: Suppose $H_1 * H_2 = \int Fd(G_1 * G_2)$. Let $H = H_1 * H_2$ and $G = G_1 * G_2$, and suppose $\mathcal{F}$ is not additively closed; i.e., $\exists\ \alpha_0, \beta_0$ such that $F(x;\alpha_0) * F(x;\beta_0) \neq F(x;\alpha_0+\beta_0)$. Let $\mu_{G_1}(\alpha_0) = \mu_{G_2}(\beta_0) = 1$, then $H_1(x) = F(x,\alpha_0)$ and $H_2(x) = F(x,\beta_0)$. $\psi_{G_1}(t) = e^{it\alpha_0}$, $\psi_{G_2}(t) = e^{it\beta_0}$, and by theorem 2 $\psi_G = e^{it(\alpha_0+\beta_0)}$. Hence, $\int F(x;\alpha)dG = F(x;\alpha_0+\beta_0)$ but $H(x) = H_1 * H_2(x) = F(x;\alpha_0) * F(x,\beta_0) \neq F(x;\alpha_0+\beta_0)$ is clearly a contradiction, and $\mathcal{F}$ must be additively closed.

Suppose $\mathcal{F}$ is additively closed. Let $\psi(t)$, $\psi_1(t)$, $\psi_2(t)$, and $\psi(t;\alpha)$ denote the c.f.'s of H, H_1, H_2, and $F(x;\alpha)$, respectively. Using theorems 2 and 5, we have

$$\psi(t) = \psi_1(t)\psi_2(t) = \int\psi(t;\alpha)dG_1(\alpha)\cdot\int\psi(t;\beta)dG_2(\beta)$$

$$= \iint\psi(t;\alpha+\beta)dG_1(\alpha)dG_2(\beta)$$

$$= \iint\psi(t;\gamma)dG_1(\gamma-\beta)dG_2(\beta)$$

$$= \int\psi(t;\gamma)dG(\gamma)$$

and this implies that H is a $G = G_1 * G_2$-mixture of $\mathcal{F} = \{F(x;\alpha)\}$.

We note that in our statement of this theorem, in order to ensure $\mathcal{F}$ is additively closed, we have required that $H_1 * H_2(x) = \int F(x;\alpha)d(G_1 * G_2)(\alpha)$ hold for the entire mixing classes $\mathcal{G}_1$ and $\mathcal{G}_2$, far less stringent conditions are necessary, as shown by Teicher [17], and we state this result in

Theorem 11. If for some $r \geq 1$ and all G_1, G_2 having exactly r points of positive mass, the convolution of a G_1-mixture of $\mathcal{F}$ with a G_2-mixture of $\mathcal{F}$ is a $(G_1 * G_2)$-mixture of $\mathcal{F}$, then $\mathcal{F}$ is additively closed.

5. Additively Closed and Identifiable Distributions.

Using some of the foregoing results, we will determine which of the more standard distributions form an additively closed class and which are identifiable.

The Poisson distribution is given by

$$P(x;\lambda) = \sum_{k=0}^{x} e^{-\lambda} \frac{\lambda^k}{k!}, \quad \lambda > 0.$$

The characteristic function for the Poisson is

$$\psi(t) = \int e^{itx}\, dP(x) = \sum_{x=0}^{\infty} e^{itx} e^{-\lambda} \frac{\lambda^x}{x!} = e^{\lambda(e^{it}-1)}.$$

Letting $P_1(x;\lambda_1) * P_2(x;\lambda_2) = H(x)$, we find, using theorem 5, that

$$H(x) = \sum_{k=0}^{x} e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^k}{k!}$$

which is Poisson with parameter $\lambda_1+\lambda_2$, and hence the Poisson family is additively closed. By the same argument we can

[illegible]

[illegible]

Binomial: $B(x;n,p) = \sum_{k=0}^{x} \binom{n}{k} p^{k}(1-p)^{n-k}$, with respect to n.

Cauchy: $C(x;\alpha,\beta)=\int_{-\infty}^{x} \frac{1}{\pi\alpha\left\{1 + \left(\frac{x-\beta}{\alpha}\right)^{2}\right\}}\,dx$, with respect to α and β.

$$\psi(t) = e^{it\beta-\alpha|t|}$$

Chi-square: $\chi^{2}(x;n)=\int_{0}^{x} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}\,dx$, with respect to n.

$$\psi(t) = \frac{1}{(1-2it)^{\frac{n}{2}}}$$

Negative binomial: $B^{-}(x;r,p) = \sum_{k=0}^{x} \binom{r+k-1}{k} p^{r}(1-p)^{k}$, with respect to r.

$$\psi(t) = \left[\frac{1}{1-(1-p)e^{it}}\right]^{r}$$

Gamma: $G(x;\lambda,r)=\int_{0}^{x} \frac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} e^{-\lambda x}\,dx$, with respect to r.

$$\psi(t) = \left[1 - \frac{it}{\lambda}\right]^{-r}$$

Normal: $\Phi(x;\mu,\sigma)=\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} dx$, with respect to μ and σ.

$$\psi(t) = e^{it\mu-\frac{1}{2}t^2\sigma^2}$$

In view of theorem 5, we can also examine the products of characteristic functions of two members of a given class of distribution functions to determine that the class is not additively closed. For example, in the exponential class

$$F(x;\lambda)=\int_0^x \lambda\, e^{-\lambda x}\, dx \text{ and } \psi(t) = \frac{\lambda}{\lambda-it} .$$

So $\frac{\lambda_1}{\lambda_1-it} \cdot \frac{\lambda_2}{\lambda_2-it} = \frac{\lambda_1\lambda_2}{\lambda_1\lambda_2-t^2-it(\lambda_1+\lambda_2)}$ and this is not the the characteristic function for the exponential with parameter $\lambda_1+\lambda_2$. Using the same argument, we see that the Bernoulli, geometric, and Uniform classes are not additively closed.

By using theorem 6, we note that since $\lambda \varepsilon D(R_+)$ the Poisson family is identifiable. Similarly, the binomial, chi-square, gamma, negative binomial, and Cauchy families are identifiable. By using theorem 9, we note that if we can find a non-degenerate mixture of a certain class that is again a member of that class, then we can conclude the class is not identifiable. For example, if we consider a mixture of two Bernoulli distributions of the following form

$$H(x) = \alpha B(x;p_1) + (1-\alpha)B(x;p_2)$$

the characteristic function of $H(x)$ would be

$$\psi(t) = \alpha\left[p_1e^{it}+1-p_1\right] + (1-\alpha)\left[p_2e^{it}+1-p_2\right]$$

$$= \left[\alpha p_1+(1-\alpha)p_2\right]e^{it}+\alpha(1-p_1)+(1-\alpha)(1-p_2)$$

$$= pe^{it}+(1-p)$$

and $H(x)$ is again Bernoulli with $p = \alpha p_1 + (1-\alpha)p_2$ as a parameter. Hence, the class of Bernoulli distributions is not identifiable.

We will now observe that the property of additivity is not necessary to ensure identifiability. The exponential distribution is not additively closed, but in

$$F(x;\nu) = \int_0^x \nu e^{-\nu x}\, dx = 1-e^{-\nu x}$$

ν is a scale parameter; and as in Teicher [15], theorem 7 shows that $G(\nu)$-mixtures of $\left\{F(x;\nu)\right\}$ are identifiable.

Since the normal family is additively closed with respect to each parameter (singly), we may use theorem 6 again to conclude the family is identifiable for $G(\mu)$ and $G(\sigma)$-mixtures. For a discussion of mixtures on both parameters, see Teicher [17].

The foregoing results are summarized in the following table.

ADDITIVE CLOSURE AND IDENTIFIABILITY SUMMARY

Class	Additively Closed	Identifiable	Remarks
Bernoulli	No	No	
Binomial	Yes (n)	Yes	1.
Cauchy	Yes (α, β)	Yes	2.
Chi-square	Yes (n)	Yes	
Exponential	No	Yes	
Gamma	Yes (r)	Yes	
Geometric	No		
Neg. Binomial	Yes (r)	Yes	
Normal	Yes (μ, σ)	Yes	2.
Poisson	Yes (λ)	Yes	
Uniform	No	No	3.

---

1. The binomial family is not identifiable when the mixing distribution is over the parameter p.

2. The normal family is identifiable when the mixing distribution is on the mean only or on the standard deviation only. (Also see Teicher.)

3. For special conditions when the uniform family is identifiable, see Teicher [illegible].

6. Linear Mixtures.

We consider a special case in mixtures of distribution and apply the theory which has been developed.

Definition 4. $H(x) = \int F(x;\alpha)\,dG(\alpha)$ is called a linear mixture if G assigns positive weight to a finite or a countable number of points, we denote the class of mixing distributions by $\mathcal{L}$.

As an example, consider the finite linear mixture when $G \in \mathcal{L}$ is of the form

$$G(\alpha) = \begin{cases} 0 & \text{if } \alpha < \alpha_1 \\ \frac{1}{3} & \text{if } \alpha_1 \leq \alpha \leq \alpha_2 \\ 1 & \text{if } \alpha_2 \leq \alpha \end{cases}$$

The mixture relative to a class $\mathcal{J} = \{F(x;\alpha)\}$ is

$$H(x) = \frac{1}{3} F(x;\alpha_1) + \frac{2}{3} F(x;\alpha_2) \ .$$

We first consider the case of finite linear mixtures; i.e., $H(x) = \sum_{i=1}^{n} a_i F(x;\alpha_i)$. Clearly, H is again a distribution function; and letting $\psi(t;\alpha)$ be the characteristic function for $F(x;\alpha)$, we have

$$\psi(t) = \int \psi(t\ ;\alpha)\, dG(\alpha) = \sum_{i=1}^{n} a_i\ \psi(t;\alpha_i)$$

as the characteristic function of $H(x)$. Also, the moments of H are given as functions of the moments of $F(x;\alpha)$ by

$$m_r = \sum_{i=1}^{n} a_i\ m_r(\alpha_i)$$

where $m_r(\alpha_i)$ is the $r^{\underline{th}}$ moment of $F(x;\alpha_i)$. These results follow easily from previous theorems or can be verified directly.

We now consider the class $\mathcal{L}$ and determine some of its algebraic properties.

$$G \varepsilon \mathcal{L} \Rightarrow \quad \psi(t) = \int e^{itx}\, dG(x) = \sum_{k=1}^{\infty} p_k e^{itx_k}$$

By theorem 5 if G_1, $G_2 \varepsilon \mathcal{L}$, then $H = G_1 * G_2$ will have a characteristic function of the form

$$\psi(t) = \left(\sum_{k=1}^{\infty} p_k e^{itx_k}\right)\left(\sum_{k=1}^{\infty} q_k e^{ity_k}\right) = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} r_{kj} e^{itz_{kj}}$$

where $r_{kj} = p_k q_j$ and $z_{kj} = x_k + y_j$.

But this is precisely the form of characteristic functions of distributions in $\mathcal{L}$. Clearly, $(G_1 * G_2) * G_3 = G_1 * (G_2 * G_3)$ and $G_1 * G_2 = G_2 * G_1$. So considering $\mathcal{L}$ as an algebraic system with convolution as the binary composition defined in $\mathcal{L}$, we have

Theorem 12. Under the operation of convolution, $\mathcal{L}$ is an Abelian semi-group.

We also note that $I(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 \le x \end{cases}$ is of the required form to be a distribution function in $\mathcal{L}$. $I(x)$ has characteristic function

$$\psi(t) = \int e^{itx}\, dI(x) = 1 \cdot e^{it0} = 1$$

and by Theorem 5, $G \in \mathcal{L} \Rightarrow G * I = G$ and hence $I(x)$ is the identity element in $\mathcal{L}$.

7. Distribution of Sums of Independent Random Variables.

It is well known that some distribution functions enjoy a certain reproductive property; namely, that the distribution function of the sum, S_n, of n independent random variables, each having a distribution from the same family, is again distributed according to that family. For example, if X_i, $i = 1, 2, \ldots, n$ are independent and distributed according to Gaussian distributions, say $N(\mu_i, \sigma_i)$, then $S_n = X_1 + X_2 + \ldots + X_n$ will be distributed according to

$$N\left(\sum_{i=1}^{n} \mu_i, \ \sum_{i=1}^{n} \sigma_i^2\right).$$

Definition 6. A family of distribution functions $\mathcal{F} = \{F(x;\alpha)\}$ is called reproductive if $F(x;\alpha) * F(x;\beta) = F(x;g(\alpha,\beta))$.

We note that this is only a generalization of the property of being additively closed where $g(\alpha,\beta) = \alpha + \beta$, and we have the trivial result that additively closed families are reproductive families. Of more interest, perhaps, is the following result.

Theorem 13. Let $\mathcal{F}$ be an additively closed class of distribution functions and let $\mathcal{H}$ be the induced class obtained by mixing over $\mathcal{L}$. Then, $\mathcal{H}$ is a reproductive class.

Proof: Let $H = H_1 * H_2$, $L = L_1 * L_2$ and denote by $\psi(t)$, $\psi_1(t)$, $\psi_2(t)$, and $\psi(t;\alpha)$ the characteristic functions of H, H_1, H_2, and $F(x;\alpha)$. By theorem 2 we have

$$\psi_1(t) = \int \psi(t;\alpha)\,dL_1(\alpha) \qquad \psi_2(t) = \int \psi(t;\alpha)\,dL_2(\alpha)$$

and by theorem 5

$$\begin{aligned}\psi(t) &= \psi_1(t)\cdot\psi_2(t) = \int \psi(t;\alpha)\,dL_1(\alpha)\cdot\int \psi(t;\alpha)\,dL_2(\alpha)\\ &= \iint \psi(t;\alpha+\beta)\,dL_1(\alpha)\,dL_2(\beta)\\ &= \iint \psi(t;\gamma)\,dL_1(\gamma-\beta)\,dL_2(\beta)\\ &= \int \psi(t;\gamma)\,dL(\gamma)\end{aligned}$$

and this implies $H = \int F\,dL$. But $L = L_1 * L_2$ and since $\mathcal{L}$ is closed under the operation of convolution, we have $H \varepsilon \mathcal{H}$.

As an example of the foregoing theorem, suppose we assume the underlying distribution of a population to be a mixture of two normal distributions. Then,

$$H(x) = \alpha N_1(\mu_1, \sigma_1^2) + (1-\alpha)N_2(\mu_2, \sigma_2^2) \qquad 0<\alpha<1$$

is the d.f. Suppose we have a sample of size n and want to determine the d.f. of $S_n = X_1 + X_2 + \ldots + X_n$. We note that S_n

with each [illegible] distribution according to a linear mixture of normal distributions and compute

$$F_{S_n} = N*N*\dots*N \qquad \text{(n-fold convolution)} .$$

The characteristic function of $N(x)$ is given by $\psi_N(t) = \alpha\psi_1(t)+(1-\alpha)\psi_2(t)$ where $\psi_i(t)$ is the characteristic function of $N(\mu_i,\sigma_i^2)$, $i = 1, 2$. Using theorem 5, we get for the characteristic function of F_{S_n}, say $\psi(t)$,

$$\psi(t) = \left[\psi_N(t)\right]^n = \left[\alpha\psi_1(t) + (1-\alpha)\psi_2(t)\right]^n$$

$$= \sum_{k=0}^{n} \binom{n}{k} \alpha^k \psi_1^k(t)(1-\alpha)^{n-k}\psi_2^{n-k}(t)$$

$$= \sum_{k=0}^{n} \beta_k \psi_1^k(t)\psi_2^{n-k}(t) \qquad \text{where}$$

$$\beta_k = \binom{n}{k}\alpha^k(1-\alpha)^{n-k} \qquad k = 0, 1, \dots, n.$$

The distribution function corresponding to $\psi_1^k(t)\psi_2^{n-k}(t)$ is

$$N_k(k\mu_1+(n-k)\mu_2,\ k\sigma_1^2+(n-k)\sigma_2^2) \qquad k = 0, 1, \dots, n.$$

Hence, $F_{S_n} = \beta_0 N_0 + \beta_1 N_1 + \dots + \beta_n N_n$, where N_k is the normal distribution function. This result is easily extended to the case where the underlying distribution function is a mixture of more than two normal distributions by the multinomial theorem.

5. Problem of estimation in mixtures of distributions.

In this most common situation, the problem of point estimation of the parameters in a mixture of distributions is very difficult. The traditional methods, such as the method of moments and maximum likelihood estimation lead to equations seemingly impossible to solve.

Karl Pearson [8] in 1894 studied the problem of estimation of the parameters (which he called dissection) in a mixture of two normal distributions. Using the method of moments, he reduced the problem to finding the roots of a ninth degree polynomial. Using this same example, we will derive the maximum likelihood equations which must be solved to obtain estimators using this technique. Letting

$$f_i(x) = \frac{1}{\sqrt{2\pi}\,\sigma_i} e^{-\frac{1}{2}\left[\frac{x-\mu_i}{\sigma_i}\right]^2} \qquad i = 1, 2$$

we have
$$f(x) = \alpha f_1(x) + (1-\alpha) f_2(x) ,$$

and
$$L(\alpha, \mu_1, \sigma_1, \mu_2, \sigma_2) = \prod_{i=1}^{n} f(x_i)$$
is the likelihood equation for a sample of size n.

$$\log L = \sum_{i=1}^{n} \log \left\{ \alpha f_1(x_i) + (1-\alpha) f_2(x_i) \right\}$$

Taking derivatives and setting equal to zero, we get

$$(1) \qquad \frac{\partial \log L}{\partial \alpha} = \sum_{i=1}^{n} \frac{f_1(x_i) - f_2(x_i)}{\alpha f_1(x_i) + (1-\alpha) f_2(x_i)} = 0$$

$$(2) \quad \frac{\partial \log L}{\partial \mu_1} = \sum_{i=1}^{n} \frac{\alpha \frac{\partial}{\partial \mu_1} f_1(x_i)}{\alpha f_1(x_i) + (1-\alpha) f_2(x_i)} = 0$$

$$(3) \quad \frac{\partial \log L}{\partial \sigma_1} = \sum_{i=1}^{n} \frac{\alpha \frac{\partial}{\partial \sigma_1} f_1(x_i)}{\alpha f_1(x_i) + (1-\alpha) f_2(x_i)} = 0$$

$$(4) \quad \frac{\partial \log L}{\partial \mu_2} = \sum_{i=1}^{n} \frac{(1-\alpha) \frac{\partial}{\partial \mu_2} f_2(x_i)}{\alpha f_1(x_i) + (1-\alpha) f_2(x_i)} = 0$$

$$(5) \quad \frac{\partial \log L}{\partial \sigma_2} = \sum_{i=1}^{n} \frac{(1-\alpha) \frac{\partial}{\partial \sigma_2} f_2(x_i)}{\alpha f_1(x_i) + (1-\alpha) f_2(x_i)} = 0$$

When one considers that each function $f_1(x_i)$ and $f_2(x_i)$ involves the parameters μ_1, σ_1, and μ_2, σ_2, respectively, in each of the equations, the difficulty of finding a solution for the above set of equations in the form $\hat{\alpha} = \alpha(x)$, $\hat{\mu}_1 = \mu_1(x)$, $\hat{\mu}_2 = \mu_2(x)$, $\hat{\sigma}_1 = \sigma_1(x)$, and $\hat{\sigma}_2 = \sigma_2(x)$ where $x = (x_1, x_2, \ldots, x_n)$, readily becomes apparent.

Needless to say, when we consider more complex mixing distribution functions, the estimation problem becomes increasingly difficult.

Rao [9] considered the problem of estimating the parameters in a linear mixture of two normal distributions when it is also assumed the standard deviations are equal. Using the k-statistics devised by Fisher, Rao equated these to their expected values and solved the resulting equations for the parameters in question. Making the assumption of equality of standard deviations and using the k-statistics, he was able to reduce the problem to a third degree equation, rather than ninth degree as found by Pearson. We give here only a summary of the estimators. The derivation may be found in the reference. Let the density be given by

$$f(x) = \alpha f(x,\mu_1,\sigma) + (1-\alpha) f(x,\mu_2,\sigma).$$

The estimators are given by

$$\hat{\alpha} = \frac{d_2}{d_2 - d_1}$$

$$\hat{\mu}_1 = k_1 + d_1$$

$$\hat{\mu}_2 = k_1 + d_2$$

$$\hat{\sigma} = k_2 + y$$

where we must compute

$$S_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad S_2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - S_1)^2$$

$$S_3 = \frac{1}{n}\sum_{i=1}^{n} (x_i - S_1)^3 \qquad S_4 = \frac{1}{n}\sum_{i=1}^{n} (x_i - S_1)^4$$

then $\quad k_1 = S_1 \qquad k_2 = \frac{n}{n-1} S_2$

$$k_3 = \frac{n^2}{(n-1)(n-2)} S_3 \qquad k_4 = \frac{n^2}{(n-1)(n-2)(n-3)} (n+1) S_4 - 3(n-1) S_2^2$$

If the presence of corrections when grouped data with class interval h, the quantities $\frac{1}{12} h^2$ and $\frac{1}{120} h^4$ must be subtracted from the expressions for k_2 and k_4, respectively. Then, y is the negative root of the cubic

$$x^3 + \frac{1}{2} k_4 x + \frac{1}{2} k_3^2 = 0$$

and d_1 is the negative root of the quadratic

$$x^2 + \frac{k_3}{y} x + y = 0$$

and finally, $d_2 = -\frac{k_3}{y} - d_1$. In this same paper, Rao gives estimators for the same parameters in terms of a modified version of the method of maximum likelihood.

To further illustrate the difficulties inherent in estimation in mixtures of distributions, we summarize some of the results of Rider [10] in applying the method of moments to a linear mixture of two exponential distributions.

If we let $x_1, x_2, \ldots, x_n$ be a random sample from a population with an underlying distribution $F(x)$ indexed by k parameters, we have the theoretical population moments given by

$$\mu_r = \int_{-\infty}^{\infty} x^r \, dF(x) \qquad r = 1, 2, \ldots ,$$

and the sample moments given by

$$m_r = \frac{1}{n} \sum_{i=1}^{n} x_i^r \qquad r = 1, 2, \ldots .$$

and let μ_r be the moments of the theoretical dis-
tribution function $f(x)$; i.e., $\mu_r = \mu_r(\theta_1, \theta_2, \dots, \theta_k)$.
If we form the set of equations

$$m_r = \mu_r(\theta_1, \theta_2, \dots, \theta_k) \qquad r = 1, 2, \dots, k$$

and solve for

$$\hat{\theta}_i = \theta_i(m_1, m_2, \dots, m_k) \qquad i = 1, 2, \dots, k$$

then $\hat{\theta}_i$ is called the estimator of θ_i obtained by the method of moments.

Letting $f(x) = \frac{\alpha}{\theta_1} e^{-x/\theta_1} + \frac{(1-\alpha)}{\theta_2} e^{-x/\theta_2}$ where $0<\alpha<1$ and $\theta_1>0$, $\theta_2>0$, $x>0$ be the linear mixture, we have $\mu_r = r!\theta^r$ in the exponential distribution function and for the mixture

$$\alpha\theta_1 + (1-\alpha)\theta_2 = m_1$$

$$\alpha\theta_1^2 + (1-\alpha)\theta_2^2 = \frac{1}{2} m_2$$

$$\alpha\theta_1^3 + (1-\alpha)\theta_2^3 = \frac{1}{6} m_3$$

These equations may be solved to give

$$\hat{\theta}_1 = \frac{-\left[2(m_3 - 3m_1 m_2)\right] + \sqrt{4(m_3 - 3m_1 m_2)^2 - 24(2m_1^2 - m_2)(3m_2^2 - 2m_1 m_3)}}{12(2m_1^2 - m_2)}$$

$$\hat{\theta}_2 = \frac{-\left[2(m_3-3m_1m_2)\right] - \sqrt{4(m_3-3m_1m_2)^2-24(2m_1^2-m_2)(3m_2^2-2m_1m_3)}}{12(2m_1^2-m_2)}$$

$$\hat{\alpha} = \frac{m_1 - \hat{\theta}_2}{\hat{\theta}_1 - \hat{\theta}_2}$$

as estimators. These estimators have deficiencies, and we summarize Rider's results in the following statements:

(a) $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\alpha}$ may turn out to be negative or complex numbers, contrary to hypothesis.

(b) If $\theta_1 \neq \theta_2$, the estimators are consistent and

$$\left.\begin{aligned} &\Pr\left[\hat{\theta}_1 > 0\right] \rightarrow 1 \\ &\Pr\left[\hat{\theta}_2 > 0\right] \rightarrow 1 \\ &\Pr\left[0 \le \hat{\alpha} \le 1\right] \rightarrow 1 \end{aligned}\right\} \quad \text{as } n \rightarrow \infty .$$

(c) If $\theta_1 = \theta_2$, the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ have no constant limits in probability, and their imaginary parts do not become arbitrarily small as $n \rightarrow \infty$. Also, the estimators are not consistent.

(d) If α is known, the estimators are consistent, even when $\theta_1 = \theta_2$. However, the probability that $\hat{\theta}_1$ and $\hat{\theta}_2$ are real does not approach 1 as $n \rightarrow \infty$, although the imaginary parts do converge to zero in probability.

(e) In the case where α is known and $\theta_1 \neq \theta_2$, consistent estimators may be derived for θ_1 and θ_2, provided it is known that $\theta_1 > \theta_2$ or $\theta_2 > \theta_1$. If the relative magnitude of the [illegible] cannot be specified.

(f) Consistent estimates may be expected to be close to the true values when the sample size is large; however, even in this case, the method is not useful.

(g) Rider observes, also, that the estimators have so many shortcomings they should not be used in practice.

We see that the problem of estimation in mixtures of distributions is difficult, even in the case where the mixture is linear and consists of two distributions. We can well imagine that more general mixtures would present even more difficult analytical problems. Basically, we have

$$H(x) = \int F(x;\theta)dG(\theta)$$

and observations on the random variable X are available to estimate the d.f. H(x). Assuming we know the form of F(x;θ), the problem becomes one of estimating the form of G(θ), given the means of estimating H(x). Robbins [11] proposes this problem and indicates conceptually, at least, how this problem might be approached.

From a practical point of view it might be worthwhile to sample empirically from a mixed distribution and consider different (intuitively satisfying) estimators of the parameters and observe their performance.

9. Graphical Method of Estimation.

The following practical approach to the estimation of parameters in a linear mixture of two normal distributions is due to A. Hald [illegible]. The method is applicable when the sample is large and the histogram exhibits definite bimodality.

We first consider a justification for the method. Assume the data has been grouped and let

- n - sample size
- Δt - interval size in which data has been grouped
- t_j - midpoint of j^{th} interval
- a_j - number of observations in the j^{th} interval.

The theoretical density function is

$$f(x) = \frac{\alpha}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left[\frac{x-\mu_1}{\sigma_1}\right]^2} + \frac{1-\alpha}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left[\frac{x-\mu_2}{\sigma_2}\right]^2}$$

and assuming μ_1 and μ_2 are sufficiently far apart and σ_1 and σ_2 small enough to guarantee bimodality, we might have $f(x)$ looking like Figure 1. Each component of the mixture

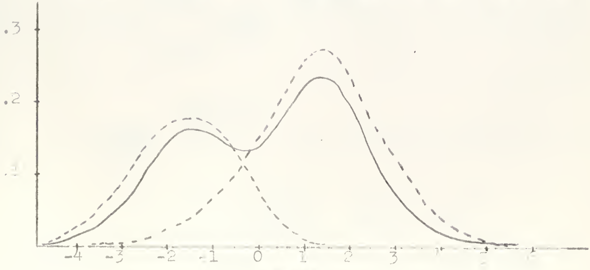


Figure 1

is [illegible]. The fundamental assumption is that the component on the right contributes very little to the left portion of the mixture f(x); and, similarly, the left component of the mixture contributes very little to the right side of the mixture. So, the left portion of the mixture may be approximated by

$$f(x) \cong \frac{\alpha}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left[\frac{x-\mu_1}{\sigma_1}\right]^2}$$

and the right side by

$$f(x) \cong \frac{(1-\alpha)}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left[\frac{x-\mu_2}{\sigma_2}\right]^2} .$$

In either case

$$\frac{a_j}{n} \cong Pr\left[t_j-\frac{\Delta t}{2} \le X \le t_j+\frac{\Delta t}{2}\right]$$

and $$a_j \cong nk \int_{t_j-\frac{\Delta t}{2}}^{t_j+\frac{\Delta t}{2}} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} dx \qquad (k = \alpha,\ 1-\alpha) .$$

So, $a_j \cong \frac{nk\Delta t}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left[\frac{t_j-\mu}{\sigma}\right]^2}$, and taking logarithms (base 10)

$$\log a_j \cong \log\left(\frac{nk\Delta t}{\sqrt{2\pi}\,\sigma}\right) - \frac{.4343}{2\sigma^2}(t_j-\mu)^2$$

[illegible] symmetry at $t = \mu$.

Thus, if we plot [illegible] versus [illegible], or plot the data on semi-logarithmic paper, the right and left-hand portions of data should fall along a parabola with axes of symmetry at μ_1 and μ_2. Using the data as plotted on semi-logarithmic paper, we estimate μ_1, μ_2, and the proportion of observations belonging to each component of the mixture. The cumulative frequencies for each component can then be computed and plotted on normal probability paper and σ_1 and σ_2 can be estimated. Hald gives an example of this method on pages 155-157.

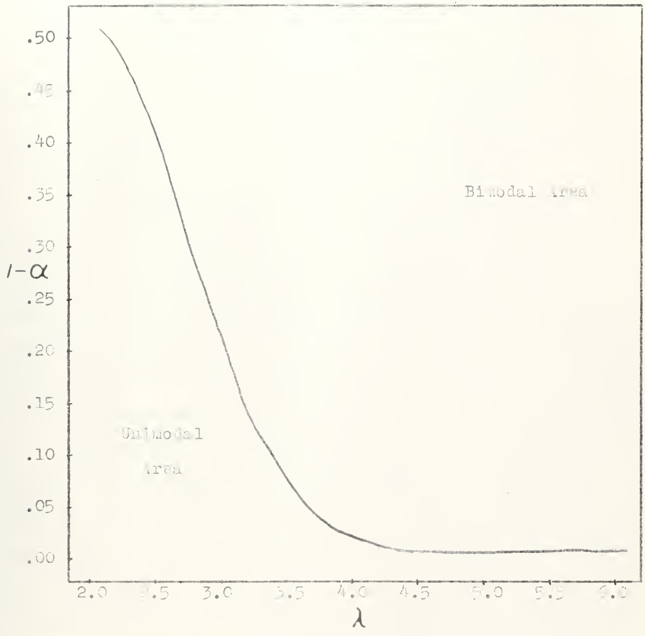
As noted above, this graphical procedure is useful when the histogram plotted from the data exhibits definite bimodality. Tvery [18] has investigated the special case

$$f(x) = \frac{\alpha}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left[\frac{x-\theta}{\sigma}\right]^2} + \frac{(1-\alpha)}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left[\frac{x-(\theta+k)}{\sigma}\right]^2} ;$$

i.e., equal variances, and the two means are separated by an amount k. Tvery was able to partition the parameter space for λ and α (where $\lambda = k/\sigma$) into those regions where f(x) is unimodal or bimodal. Figure 2 on page 36 is taken from his unpublished notes.

10. Related Results and Observations.

Blischke [1] discusses the case of a linear mixture of two binomials. We saw earlier that such a mixture on p is



.50
.45
.40
.35
.30
1-α
.25
.20
.15
.10
.05
.00
Bimodal Area
Unimodal
Area
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
λ

Figure 2

and [illegible] by [illegible] [illegible] but the estimators [illegible] satisfactory result. Neither proposes [illegible] distribution [illegible] of their estimators. A small empirical sampling study is presented. The paper illustrates clearly the difficulties [illegible] estimation of parameters in mixtures.

Burrau [illegible] [illegible] the problem of "dissecting" a linear mixture of two normal distributions by computing Thiele's half-invariants for this mixture, and Strömgren [14] in a related paper gives a convenient way of solving Burrau's equations. He illustrates the method and discusses the properties of both the moment and half-invariant methods of estimation since the methods turn out to be equivalent.

In making a series of observations of a random variable, it may happen that a small number of observations differ considerably from the others. These outliers (spurious or aberrant observations) can distort the results of a statistical investigation, in particular when the number of observations is small. A decision problem then arises as to the proper disposition of these observations. The outlier problem is concerned with making such decisions. Many times these outliers are results of unforeseen circumstances or errors and cannot be considered as truly belonging to the assumed underlying distribution. So, we see that