TECHNOMETRICS VOL. 9, NO. 1 FEBRUARY 1967

Estimation in Mixtures of Two Normal Distributions*

A. CLIFFORD COHEN

The University of Georgia

This paper is concerned primarily with the method of moments in dissecting a mixture of two normal distributions. In the general case, with two means, two standard deviations, and a proportionality factor to be estimated, the first five sample moments are required, and it becomes necessary to find a particular solution of a ninth degree polynomial equation that was originally derived by Karl Pearson [10]. A procedure which circumvents solution of the nonic equation and thereby considerably reduces the total computational effort otherwise required, is presented. Estimates obtained in the simpler special case in which the two standard deviations are assumed to be equal, are employed as first approximations in an iterative method for simultaneously solving the basic system of moment equations applicable in the more general case in which the two standard deviations are unequal. Conditional maximum likelihood and conditional minimum chi-square estimation subject to having the first four sample moments equated to corresponding population moments, are also considered. An illustrative example is included.

1. INTRODUCTION

Distributions which result from the mixing of two or more component distributions are designated as "compound" or "mixed" distributions. They may be further described by designating distribution types of the individual components. Such distributions arise in a wide variety of practical situations ranging from distributions of wind velocities to distributions of physical dimensions of various mass produced items. Compound normal distributions were studied as early as 1894 by Karl Pearson [10] and later by Charlier [3] and by Charlier and Wicksell [4]. More recently, compound Poisson and compound exponential distributions have been studied by Rider [12] and by the writer [5, 6, 7, 8]. Compound binomial distributions have been studied by Blischke [1].

This paper is concerned with estimation of the parameters θ_1 , σ_1 , θ_2 , σ_2 , and α of the compound normal distribution with density

$$f(x) = \alpha f_1(x) + (1 - \alpha)f_2(x), \tag{1}$$

where

$$\left.\begin{aligned} f_1(x) &= (2\pi\sigma_1^2)^{-\frac{1}{2}} \exp\left[-(x - \theta_1)/\sigma_1\right]^2/2, \\ f_2(x) &= (2\pi\sigma_2^2)^{-\frac{1}{2}} \exp\left[-(x - \theta_2)/\sigma_2\right]^2/2. \end{aligned}\right\} \tag{2}$$

Received June 1965. Revised November 1965.

* Research sponsored by the Aero-Astrodynamics Laboratory of the Marshall Space Flight Center, National Aeronautics and Space Administration; Contract NAS8-11175.

Karl Pearson [10] derived estimators for the parameters of this distribution by equating sample moments to corresponding population (theoretical) moments. The evaluation of his estimators involved the solution of a ninth degree polynomial equation. Before the advent of modern electronic computers, this was considered a rather formidable obstacle to the application of his results. Charlier and Wicksell [4] succeeded in considerably simplifying Pearson's results. The present effort represents an attempt to further simplify these estimators in order that they might be more readily available for use in appropriate practical applications. A procedure is presented whereby the direct solution of Pearson's nonic equation can be circumvented through an iterative process which involves solving a cubic equation for a unique negative root. In addition to considering the most general case which all five parameters of (1) must be estimated from the data available in a given sample, some of the more important special cases in which one or more of the parameters are known in advance of making sample observations, are also examined.

2. Estimation in the General Case

Except for a few changes in notation, the presentation of this section is essentially that of Charlier and Wicksell [4].

The kth moment of $f(x)$ taken about the origin may be written as

$$\mu_k' = \alpha \int_{-\infty}^{\infty} x^k f_1(x)\, dx + (1 - \alpha) \int_{-\infty}^{\infty} x^k f_2(x)\, dx, \tag{3}$$

where $f_1(x)$ and $f_2(x)$ are density functions of the two component distributions as given in equations (2). The kth central moment becomes

$$\mu_k = \alpha \int_{-\infty}^{\infty} (x - \mu_1')^k f_1(x)\, dx + (1 - \alpha) \int_{-\infty}^{\infty} (x - \mu_1')^k f_2(x)\, dx. \tag{4}$$

If we let

$$m_1 = \theta_1 - \mu_1' \quad \text{and} \quad m_2 = \theta_2 - \mu_1', \tag{5}$$

the first non-central and the first five central moments of (1) follow as

$$\left.\begin{aligned}
\mu_1' &= \alpha\theta_1 + (1 - \alpha)\theta_2, \\
\mu_1 &= \alpha m_1 + (1 - \alpha) m_2 = 0, \\
\mu_2 &= \alpha(\sigma_1^2 + m_1^2) + (1 - \alpha)(\sigma_2^2 + m_2^2), \\
\mu_3 &= \alpha m_1(3\sigma_1^2 + m_1^2) + (1 - \alpha) m_2(3\sigma_2^2 + m_2^2), \\
\mu_4 &= \alpha[3\sigma_1^4 + 6m_1^2\sigma_1^2 + m_1^4] + (1 - \alpha)[3\sigma_2^4 + 6m_2^2\sigma_2^2 + m_2^4], \\
\mu_5 &= \alpha m_1[15\sigma_1^4 + 10m_1^2\sigma_1^2 + m_1^4] + (1 - \alpha) m_2[15\sigma_2^4 + 10m_2^2\sigma_2^2 + m_2^4].
\end{aligned}\right\} \tag{6}$$

Without any loss of generality let us suppose that $\theta_1 \le \theta_2$. Accordingly $\theta_1 \le \mu_1' \le \theta_2$, and $m_1 \le 0 \le m_2$.

Upon equating population moments to corresponding sample moments, it follows from (6) that

$$\alpha[\theta_1 - \bar{x}] + (1 - \alpha)[\theta_2 - \bar{x}] = 0, \tag{7}$$

and further that

$$\left.\begin{aligned} \alpha m_1 + (1-\alpha)m_2 &= 0,\\ \alpha[\sigma_1^2 + m_1^2 - \nu_2] + (1-\alpha)[\sigma_2^2 + m_2^2 - \nu_2] &= 0,\\ \alpha[3\sigma_1^2 m_1 + m_1^3 - \nu_3] + (1-\alpha)[3\sigma_2^2 m_2 + m_2^3 - \nu_3] &= 0,\\ \alpha[3\sigma_1^4 + 6m_1^2\sigma_1^2 + m_1^4 - \nu_4] + (1-\alpha)[3\sigma_2^4 + 6m_2^2\sigma_2^2 + m_2^4 - \nu_4] &= 0,\\ \alpha[15\sigma_1^4 m_1 + 10\sigma_1^2 m_1^3 + m_1^5 - \nu_5] + (1-\alpha)[15\sigma_2^4 m_2 + 10\sigma_2^2 m_2^3 + m_2^5 - \nu_5] &= 0, \end{aligned}\right\} \quad (8)$$

where $\bar{x}$ is the sample mean and ν_i, $(i = 2, 3, 4, \cdots)$ is the ith central moment of the sample. Equations (8) accordingly constitute a system of five equations to be solved simultaneously for estimates of the five parameters $\alpha, m_1, \sigma_1, m_2, \sigma_2$.

We eliminate α between the first and the subsequent equations of (8) in turn and thereby reduce this system to the following four equations in the four unknowns $\sigma_1, m_1, \sigma_2, m_2$:

$$\left.\begin{aligned} m_1^{-1}(\sigma_1^2 + m_1^2 - \nu_2) &= m_2^{-1}(\sigma_2^2 + m_2^2 - \nu_2) = \beta,\\ m_1^{-1}(3\sigma_1^2 m_1 + m_1^3 - \nu_3) &= m_2^{-1}(3\sigma_2^2 m_2 + m_2^3 - \nu_3),\\ m_1^{-1}(3\sigma_1^4 + 6m_1^2\sigma_1^2 + m_1^4 - \nu_4) &= m_2^{-1}(3\sigma_2^4 + 6m_2^2\sigma_2^2 + m_2^4 - \nu_4),\\ m_1^{-1}(15\sigma_1^4 m_1 + 10\sigma_1^2 m_1^3 + m_1^5 - \nu_5) &= m_2^{-1}(15\sigma_2^4 m_2 + 10\sigma_2^2 m_2^3 + m_2^5 - \nu_5). \end{aligned}\right\} \quad (9)$$

In view of the presence of m_1 and m_2 in the denominators of (9), we recognize that these equations and those subsequently derived from them, are not valid in the symmetric case to be dealt with later in which $m_1 = m_2 = 0$. In all subsequent results presented in this section is it understood that $m_1 < 0$ and $m_2 > 0$.

With the introduction of β in the first equation of (9) it follows that

$$\left.\begin{aligned} \sigma_1^2 &= m_1\beta + \nu_2 - m_1^2,\\ \sigma_2^2 &= m_2\beta + \nu_2 - m_2^2. \end{aligned}\right\} \quad (10)$$

On replacing σ_1^2 and σ_2^2 in the second, third and fourth equations of (9) with the values given in (10) above, it follows after considerable algebraic manipulations that

$$\left.\begin{aligned} &m_1m_2[3\beta - 2(m_1 + m_2)] = -\nu_3,\\ &m_1m_2[3\beta^2 - 2(m_1^2 + m_1m_2 + m_2^2)] = -k_4,\\ &m_1m_2[15(m_1 + m_2)\beta^2 - 20(m_1^2 + m_1m_2 + m_2^2)\beta\\ &\qquad\qquad + 6(m_1 + m_2)(m_1^2 + m_2^2)] = -k_5, \end{aligned}\right\} \quad (11)$$

where k_4 and k_5 are respectively the fourth and fifth order sample cumulants or semi-invariants; i.e.

$$\left.\begin{aligned} k_4 &= \nu_4 - 3\nu_2^2,\\ k_5 &= \nu_5 - 10\nu_2\nu_3. \end{aligned}\right\} \quad (12)$$

When referring to population (theoretical) cumulants, we employ the Greek kappa thus:

$$\kappa_4 = \mu_4 - 3\mu_2^2 , \qquad \kappa_5 = \mu_5 - 10\mu_2\mu_3 .$$

Equations (11) accordingly constitute a system of three equations in the three unknowns, β, m_1 and m_2 .

In order to further simplify the system of equations (11), let

$$r = m_1 + m_2 , \quad \text{and} \quad v = m_1 m_2 . \tag{13}$$

When these transformations are introduced into equations (11), the system becomes

$$\left.\begin{aligned} &3\beta v - 2rv = -\nu_3 , \\ &3\beta^2 v - 2v(r^2 - v) = -k_4 , \\ &15vr\beta^2 - 20v(r^2 - v)\beta + 6vr(r^2 - 2v) = -k_5 . \end{aligned}\right\} \tag{14}$$

On making the further transformation

$$w = rv, \tag{15}$$

the system of equations (14) becomes

$$\left.\begin{aligned} &3\beta v - 2w = -\nu_3, \\ &3\beta^2 v^2 - 2w^2 + 2v^3 = -k_4 v, \\ &15v^2 w\beta^2 - 20vw^2\beta + 20v^4\beta + 6w^3 - 12v^3 w = -k_5 v^2. \end{aligned}\right\} \tag{16}$$

We now eliminate β between the first and the second and between the first and the third equations of (16), and our system becomes

$$\left.\begin{aligned} &2w^2 - 6v^3 - 3vk_4 + 4w\nu_3 - \nu_3^2 = 0, \\ &5w(2w - \nu_3)^2 - 20w^2(2w - \nu_3) + 20v^3(2w - \nu_3) + 18w^3 - 36wv^3 = -3k_5 v^2. \end{aligned}\right\} \tag{17}$$

By introducing the further transformation

$$z = w + \nu_3 , \tag{18}$$

the two equations of (17) become

$$\left.\begin{aligned} &2z^2 = 6v^3 + 3k_4 v + 3v_3^2 , \\ &2z^2(z - 3\nu_3) + z(\nu_3^2 - 4v^3) + 3\nu_3^3 + 24\nu_3 v^3 = 3k_5 v^2. \end{aligned}\right\} \tag{19}$$

When the expression for $2z^2$ from the first of the above equations is substituted into the second of those equations, we obtain

$$z = (-6\nu_3 v^3 + 3k_5 v^2 + 9\nu_3 k_4 + 6\nu_3^3)/(2v^3 + 3k_4 v + 4\nu_3^2), \tag{20}$$

and when this value is reinserted into the first equation of (19), we obtain a polynomial equation of the ninth degree in v, which for convenience we write as follows

$$a_9 v^9 + a_8 v^8 + a_7 v^7 + a_6 v^6 + a_5 v^5 + a_4 v^4 + a_3 v^3 + a_2 v^2 + a_1 v + a_0 = 0, \tag{21}$$

where

$$\left.\begin{aligned} a_9 &= 24, & a_4 &= 444k_4\nu_3^2 - 18k_5^2\,, \\ a_8 &= 0, & a_3 &= 288\nu_3^4 - 108\nu_3k_4k_5 + 27k_4^3\,, \\ a_7 &= 84k_4\,, & a_2 &= -(63\nu_3^2k_4^2 + 72\nu_3^3k_5), \\ a_6 &= 36\nu_3^2\,, & a_1 &= -96\nu_3^4k_4\,, \\ a_5 &= 90k_4^2 + 72k_5\nu_3 & a_0 &= -24\nu_3^6\,. \end{aligned}\right\} \tag{22}$$

This is the well known nonic which was first given in 1894 by Pearson [10].

Since here $m_1 < 0$ and $m_2 > 0$, then $v = m_1m_2 < 0$, and the required estimate v^* is to be found as a negative real root of (21). Throughout this paper asterisks (*) are employed to distinguish estimates from the parameters being estimated. Prior to the advent of electronic computers, the task of solving this equation would have been considered a formidable assignment. Today, however, modern computers are available to perform the otherwise burdensome calculations involved. The ninth degree polynomial can be evaluated for any desired value of v in the vicinity of v^* either by straight-forward substitution or by synthetic division. Standard iterative procedures will quickly lead to the required value of v^*. Once v^* is determined with the desired degree of accuracy, the estimate w^* follows from (20) and (18) as

$$w^* = \frac{-8\nu_3v^{*3} + 3k_5v^{*2} + 6\nu_3k_4v^* + 2\nu_3^3}{2v^{*3} + 3k_4v^* + 4\nu_3^2}.$$

From (15), we have $r^* = w^*/v^*$, and accordingly

$$r^* = \frac{-8\nu_3v^{*3} + 3k_5v^{*2} + 6\nu_3k_4v^* + 2\nu_3^3}{v^*(2v^{*3} + 3k_4v^* + 4\nu_3^2)}. \tag{23}$$

It follows from the defining relations (13), that estimates of m_1 and m_2 are the roots of the quadratic equation

$$Y^2 - r^*Y + v^* = 0. \tag{24}$$

Thus, we have

$$\left.\begin{aligned} m_1^* &= \tfrac{1}{2}[r^* - \sqrt{r^{*2} - 4v^*}], \\ m_2^* &= \tfrac{1}{2}[r^* + \sqrt{r^{*2} - 4v^*}]. \end{aligned}\right\} \tag{25}$$

Using the first equation of (14), we estimate β as

$$\beta^* = \tfrac{1}{3}[2r^* - \nu_3/v^*], \tag{26}$$

and from (10), we have

$$\left.\begin{aligned} \sigma_1^{*2} &= m_1^*\beta^* + \nu_2 - m_1^{*2}, \\ \sigma_2^{*2} &= m_2^*\beta^* + \nu_2 - m_2^{*2}. \end{aligned}\right\} \tag{27}$$

Finally, from (5) and from the second equation of (6), we have

$$\left.\begin{aligned} \theta_1^* &= m_1^* + \bar{x}, \\ \theta_2^* &= m_2^* + \bar{x}, \\ \alpha^* &= m_2^*/(m_2^* - m_1^*). \end{aligned}\right\} \tag{28}$$

Attention is again invited to the fact that the preceding results are valid only if $\theta_1 \neq \theta_2$. The symmetric case in which $\theta_1 = \theta_2$ and thus $m_1 = m_2 = r = v = w = 0$ is treated separately in Section 4.

Unfortunately, for some combinations of sample data the nonic equation (21) may have more than one negative root and accordingly we must choose between two or more sets of estimates. This lack of uniqueness bothered Pearson and he suggested choosing the set of estimates which resulted in the closest agreement between the sixth central moment of the sample and the corresponding moment of the "fitted" compound curve.

2.1 *On Circumventing the Nonic Estimating Equation.* In the event that r is known, we need only consider the first two equations of (14), and when β is eliminated between them, we have the following cubic equation in v

$$6v^3 - 2r^2v^2 + (3k_4 - 4r\nu_3)v + \nu_3^2 = 0. \tag{29}$$

Using Descarte's rule of signs, we find that this equation has one negative root plus either two positive or a pair of imaginary roots. The negative root is the required estimate v^*.

Using this value for v^* and the known value for r, the required estimates follow from (25), (26), (27) and (28) as before.

Even though r is not known a priori, we might assume a value and employ the foregoing results to determine approximations to the required estimates which in turn can be substituted into the final equation of (6), to approximate the fifth central moment, μ_5.

Let $r_{(i)}$ designate the ith approximation to r^* and let $\mu_{5(i)}$ designate the ith approximation to μ_5^* based on $r_{(i)}$. It should be relatively easy to find approximations $r_{(i)}$ and $r_{(i+1)}$ such that the sample moment ν_5 is in the interval $(\mu_{5(i)}, \mu_{5(i+1)})$. As shown in section 3, a satisfactory initial approximation to r can usually be found by assuming $\sigma_1 = \sigma_2$ and solving the appropriate estimating equation for this special case. Once the interval between $r_{(i)}$ and $r_{(i+1)}$ has been narrowed sufficiently, the required estimate r^* can be obtained by simple linear interpolation as indicated below.

r	μ_5
$r_{(i)}$	$\mu_{5(i)}$
r^*	ν_5
$r_{(i+1)}$	$\mu_{5(i+1)}$

With r specified, the well known method of Horner which utilizes synthetic division procedures, is quite effective in solving (29) for v. Any standard iterative method, however, might be employed. A "trial and error" procedure based on linear interpolation with direct substitution in (29) though perhaps not very economical of computational effort, will generally give satisfactory results.

Various special cases in which one or more of the parameters are known or in some way restricted are sometimes of interest. With fewer parameters to be estimated, the number of sample moments involved is correspondingly reduced and the estimating equations are accordingly simpler and easier to solve. Some of the more important special cases are considered in Sections 3 and 4.

2.2 *Conditional Maximum Likelihood and Conditional Minimum Chi-Square Estimation.* In order to eliminate the effect of sampling errors resulting from direct use of the fifth order moment, let us consider a conditional maximum likelihood procedure. The first four sample moments are equated to corresponding population moments, and subject to this condition, r is determined so as to maximize

$$L'(r) = \prod_{i=1}^{n} f(x_i \mid \bar{x}, \nu_2, \nu_3, \nu_4).$$

Since derivatives of $L'(r)$ are somewhat unwieldy, the value of r which maximizes $L'(r)$ can conveniently be determined in most practical applications by actually calculating $L'(r)$ for several values of r in the vicinity of its maximum and employing either finite difference or graphical techniques.

In the case of grouped data, the conditional likelihood function might be expressed as

$$L'(r) = \prod_{i=1}^{k} p_i^{n_i},$$

where k is the number of groups or classes into which the sample has been divided, n_i is the number of sample observations in the ith class, x_i is the upper boundary of the ith class, and

$$p_i = \int_{x_{i-1}}^{x_i} f(x \mid \bar{x}, \nu_2, \nu_3, \nu_4)\, dx.$$

In practice it is sometimes more convenient to minimize chi-square than to maximize $L'(r)$. Kendall [9, Vol. II p. 55–56] has shown that with grouped data, the two methods are equivalent to the order $n^{-1/2}$. We are therefore free to choose the method requiring the least computational effort. In the method of minimum chi-square, we seek the value of r which results in the minimum value for

$$\chi^2(r \mid \bar{x}, \nu_2, \nu_3, \nu_4) = \sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i},$$

where $e_i = np_i$ is the expected number of observations in the ith class subject to the restriction that the first four sample moments are equated to corresponding population moments. As previously indicated, n_i is the number of sample observations in the ith class.

In practice, chi-square can be calculated for several values of r in the vicinity of its minimum and the desired value of r can be determined either graphically or by employing finite difference techniques.

Unrestricted maximum likelihood estimates would, of course, be preferable to any moment estimates, but with five parameters to be estimated, the unrestricted maximum likelihood estimating equations become quite intractable. The conditional maximum likelihood procedures suggested here are believed to represent a feasible compromise between the need for estimating equations that are easy to solve and the need for reliable estimating procedures.

3. Estimation When $\sigma_1 = \sigma_2 = \sigma$

Here, we need only the four parameters θ_1, θ_2, α, and σ where as in the general case $\theta_1 < \theta_2$. Accordingly only the first four moments and or cumulants are required. Charlier and Wicksell [4], and Rao [11] among others have previously considered this special case. With $\sigma_1 = \sigma_2 = \sigma$, equations (10) in the general five-parameter case become

$$\sigma^2 = m_1\beta + \nu_2 - m_1^2 = m_2\beta + \nu_2 - m_2^2 . \tag{30}$$

From this it follows that

$$\left.\begin{aligned} \beta &= m_1 + m_2 = r, \\ \sigma^2 &= m_1 m_2 + \nu_2 = v + \nu_2 , \end{aligned}\right\} \tag{31}$$

where r and v are as defined in (13).

The first two equations of (14) which are applicable here now become

$$3rv - 2rv = -\nu_3 ,$$

$$3r^2v - 2v(r^2 - v) = -k_4 ,$$

and subsequently

$$\left.\begin{aligned} rv &= -\nu_3 , \\ r^2v + 2v^2 &= -k_4 . \end{aligned}\right\} \tag{32}$$

From the first of the above equations

$$r = -\nu_3/v. \tag{33}$$

When this value is substituted into the second equation of (32), and both sides are multiplied by v, the applicable estimating equation becomes

$$2v^3 + k_4 v + \nu_3^2 = 0. \tag{34}$$

From Descarte's rule of signs it follows that unless $\nu_3^2 = 0$, equation (34) has a single negative root, which is the required estimate v^* regardless of whether k_4 is positive or negative. The other two roots are of no interest to us here. It is relatively easy to solve (34) for v^* using standard iterative procedures, and from (33) it follows that

$$r^* = -\nu_3/v^*. \tag{35}$$

With r^* and v^* thus determined, the estimates m_1^*, m_2^*, θ_1^*, θ_2^* and α^* are given by (25) and (28) as in the general case, while σ^{*2} follows from the second equation of (31) as

$$\sigma^{*2} = v^* + \nu_2 . \tag{36}$$

3.1 *A first approximation to r in the general case.* In view of the relative ease with which estimates can be calculated when $\sigma_1 = \sigma_2$, we are thus provided with a simple procedure for obtaining initial approximations to r in the general case. With σ_1 assumed equal to σ_2, we obtain $v_{(0)}$ as the negative root of (34), and calculate $r_{(0)} = -\nu_3/v_{(0)}$, from (33). Unless the disparity between σ_1 and σ_2 is quite large, the resulting value, $r_{(0)}$, provides a satisfactory starting point from which we can iterate to the final estimate r^* in the general case.

4. Estimation in the Symmetric Cases

The compound normal distribution is symmetrical if i) $\alpha = 1/2$ with $\sigma_1 = \sigma_2 = \sigma$ and if ii) $\theta_1 = \theta_2$. In the former instance the compound distribution has been shown to be bimodal when $\theta_2 - \theta_1 > 2\sigma$, but otherwise unimodal (cf. [5]). In the second case, the resultant distribution is always unimodal. In the limiting (trivial) case in which $\theta_1 = \theta_2 = \mu$ and $\sigma_1 = \sigma_2 = \sigma$, the resultant distribution degenerates into a single normal distribution (μ, σ).

4.1 *Symmetric case with* $\alpha = 1/2$ *and* $\sigma_1 = \sigma_2$, $(\theta_1 \neq \theta_2)$. In this case, we need only estimate the three parameters σ, θ_1 and θ_2. From the second equation of (6), with $\alpha = 1/2$, it follows that $m_1 = -m_2$ and from the first equation of (31), $r = \beta = 0$. Consequently, from the second equation of (32), we have

$$2v^2 + k_4 = 0. \tag{37}$$

With the vanishing of the odd central population moments, equation (37) might have been obtained as a special case of (34) with ν_3 replaced by zero, which in this instance is the appropriate estimate of μ_3.

It follows from (37) that

$$v^* = -\sqrt{-k_4/2}. \tag{38}$$

With v^* given by (38) and with $r = 0$, the required estimates follow from (25), (28), and (36) as

$$\left.\begin{aligned} m_2^* &= \sqrt{-v^*}, \qquad m_1^* = -m_2^*, \\ \theta_1^* &= m_1^* + \bar{x}, \qquad \theta_2^* = m_2^* + \bar{x}, \\ \sigma^{*2} &= v^* + \nu_2 . \end{aligned}\right\} \tag{39}$$

4.2 *Symmetric case with* $\theta_1 = \theta_2 = \theta$. Since estimating equations (9) involve division by m_1 and m_2, and since it follows that $m_1 = m_2 = 0$ when $\theta_1 = \theta_2$, neither equations (9) nor subsequent equations derived from them are applicable here. We estimate θ from (7) as

$$\theta^* = \bar{x}. \tag{40}$$

With the vanishing of the odd central moments, estimation of the three remaining parameters, α, σ_1 and σ_2 necessitates use of the second, fourth, and sixth central moments. Applicable estimating equations accordingly become

$$\left.\begin{aligned} \alpha(\sigma_1^2 - \nu_2) + (1 - \alpha)(\sigma_2^2 - \nu_2) &= 0, \\ \alpha(3\sigma_1^4 - \nu_4) + (1 - \alpha)(3\sigma_2^4 - \nu_4) &= 0, \\ \alpha(15\sigma_1^6 - \nu_6) + (1 - \alpha)(15\sigma_2^6 - \nu_6) &= 0. \end{aligned}\right\} \tag{41}$$

The first two of these equations follow directly from (8), and the third follows from (4) with $k = 6$ when μ_6 is estimated as ν_6.

From the first equation of (41), we have

$$\alpha = \frac{\nu_2 - \sigma_2^2}{\sigma_1^2 - \sigma_2^2}, \quad \text{and} \quad (1 - \alpha) = \frac{\sigma_1^2 - \nu_2}{\sigma_1^2 - \sigma_2^2}. \tag{42}$$

On substituting these values for α and $(1 - \alpha)$ into the second and third equations of (41), after considerable algebraic manipulations we obtain

$$\left.\begin{aligned}(\sigma_1^2 - \nu_2)(\sigma_2^2 - \nu_2) &= -k_4/3,\\ (\sigma_1^2 - \nu_2)(\sigma_2^2 - \nu_2)(\sigma_1^2 + \sigma_2^2 - 2\nu_2) &= -k_6/15,\end{aligned}\right\} \tag{43}$$

where k_4 and k_6 are the fourth and sixth order sample cumulants respectively; i.e.

$$\left.\begin{aligned}k_4 &= \nu_4 - 3\nu_2^2 ,\\ k_6 &= \nu_6 - 15\nu_4\nu_2 - 10\nu_3^2 + 30\nu_2^3.\end{aligned}\right\} \tag{44}$$

Let

$$t_1 = \sigma_1^2 - \nu_2 \quad \text{and} \quad t_2 = \sigma_2^2 - \nu_2 , \tag{45}$$

and equations (43) become

$$\left.\begin{aligned}t_1 t_2 &= -k_4/3,\\ t_1 t_2(t_1 + t_2) &= -k_6/15.\end{aligned}\right\} \tag{46}$$

It follows from (46) that t_1^* and t_2^* are roots of the quadratic equation

$$Y^2 - (k_6/5k_4)Y - (k_4/3) = 0. \tag{47}$$

Accordingly

$$\left.\begin{aligned}t_1^* &= \tfrac{1}{2}[(k_6/5k_4) - \sqrt{(k_6/5k_4)^2 + (4k_4/3)}],\\ t_2^* &= \tfrac{1}{2}[(k_6/5k_4) + \sqrt{(k_6/5k_4)^2 + (4k_4/3)}].\end{aligned}\right\} \tag{48}$$

With t_1^* and t_2^* thus determined, it follows from (45) and from (41) that

$$\sigma_1^{*2} = t_1^* + \nu_2 , \qquad \sigma_2^{*2} = t_2^* + \nu_2 , \qquad \alpha^* = -t_2^*/(t_1^* - t_2^*), \tag{49}$$

and of course $\theta^* = \bar{x}$ as given in (40).

5. Determining Which Case is Applicable

In the absence of a priori information concerning whether or not one of the symmetric special cases is applicable in lieu of the general case, the following criteria provide a basis for resolving this issue.

i) If $\mu_3 = 0$ and if $\kappa_4 < 0$, the compound distribution is symmetric with $\sigma_1 = \sigma_2$ and with $\alpha = 1/2$.

ii) If $\mu_3 = 0$ and if $\kappa_4 > 0$, the compound distribution is symmetric with $\theta_1 = \theta_2$.

iii) If $\mu_3 = 0$ and $\kappa_4 = 0$, then $\theta_1 = \theta_2$, $\sigma_1 = \sigma_2$, and the "resultant" distribution is in fact a single normal distribution.

Of course the third central moment is zero in all symmetrical distributions, and the converse likewise is true. Therefore, the foregoing statements can be verified by examining applicable expressions for the fourth cumulant, which is defined as $\kappa_4 = \mu_4 - 3\mu_2^2$. Using expressions from equations (6) for μ_4 and μ_2, the fourth population cumulant in the most general case follows as

$$\kappa_4 = 3\alpha(1 - \alpha)[(\sigma_1^2 - \sigma_2^2) + (m_1^2 - m_2^2)]^2 - 2[\alpha m_1^4 + (1 - \alpha)m_2^4]. \tag{50}$$

When $\alpha = 1/2$ *and* $\sigma_1 = \sigma_2$, this implies that $-m_1 = m_2$. With these values substituted into (50), we have

$$\kappa_4 = -[m_1^4 + m_2^4] < 0. \tag{51}$$

When $\theta_1 = \theta_2$, this implies that $m_1 = m_2 = 0$, and in this instance (50) becomes

$$\kappa_4 = 3\alpha(1 - \alpha)(\sigma_1^2 - \sigma_2^2)^2 \geq 0. \tag{52}$$

When $\theta_1 = \theta_2$ *and* $\sigma_1 = \sigma_2$, it follows from (52) that in this limiting case $\kappa_4 = 0$, since (52) is applicable in all cases where $\theta_1 = \theta_2$.

In practical applications our classification problem is reduced to that of utilizing the sample statistics ν_3 and $k_4 = \nu_4 - 3\nu_2^2$ in choosing the most acceptable hypothesis from among the following alternatives: $H_{0:\mu_3=0,\kappa_4=0}$, $H_{1:\mu_3=0,\kappa_4<0}$, $H_{2:\mu_3=0,\kappa_4>0}$, $H_{3:\mu_3\neq0}$.

6. An Illustrative Example

To illustrate the practical application of computational procedures developed in this paper, we consider a mixed sample obtained by combining two separate component samples consisting of 334 and 672 observations respectively from normal populations. For these two component samples, $\bar{x}_1 = 47.716$, $s_1^2 = 33.5663$, $N_1 = 334$, $\bar{x}_2 = 57.607$, $s_2^2 = 9.1790$ and $N_2 = 672$. For the resultant mixed sample, $n = N_1 + N_2 = 1006$, $\bar{x} = 54.32306$, $\nu_2 = s^2 = 38.9753$, $\nu_3 = -233.876$, $\nu_4 = 5365.13$, $\nu_5 = -67{,}821$ and $k_4 = 807.91$. In an effort to compensate for errors due to grouping, Sheppard's corrections (c.f. Kendall, [9, Vol. I, p. 71]) have been applied both in computing s_1^2 and s_2^2 for the separate components and in computing moments of the resultant mixed sample.

We obtain a first approximation to v by substituting the values given above for k_4 and ν_3 into equation (34) and solving for the negative root. By thus employing equation (34), we implicitly assume $\sigma_1 = \sigma_2$ for this first approximation. In subsequent approximations, we abandon this restriction and accordingly replace equation (34) with (29). For the first approximation our estimating equation, after making proper substitutions and simplifying, becomes

$$v^3 + 403.955v + 27{,}349 = 0.$$

With the aid of a desk calculator, straightforward substitution quickly yields as a solution, the first approximation $v_{(1)} = -25.7$. A first approximation to r then follows from (35) as

$$r_{(1)} = -(-233.876)/(-25.7) = -9.10.$$

Using these values, first approximations to the basic parameters follow from (36), and (25) and (28) as $\sigma_{1(1)}^2 = \sigma_{2(1)}^2 = 14.61$, $m_{1(1)} = -11.36$, $m_{2(1)} = 2.26$, and $\alpha_{(1)} = 0.166$. On substituting these values into the last equation of (6), we subsequently calculate $\mu_{5(1)} = -62{,}396$ which is to be compared with the sample value, $\nu_5 = -67{,}821$.

As a second approximation to r, we let $r_{(2)} = -4.50$ since further examination of the sample data with due regard for the shape of the histogram indicates that a value in the vicinity of -4 or -5 should be a good choice. This time we employ equation (29) rather than (34) in determining our new approximation to v. With $r = r_{(2)} = -4.50$, and with k_4 and ν_3 as previously given, equation (29) becomes

$$v^3 - 6.75v^2 - 297.7v + 9{,}116 = 0.$$

On solving for the negative root of this equation, we find as our second approximation $v_{(2)} = -23.13$. Using the above values for $r_{(2)}$ and $v_{(2)}$, equations (25), (26), (27), and (28) yield as second approximations to the remaining parameters of interest, $m_{1(2)} = -7.56$, $m_{2(2)} = 3.06$, $\alpha_{(2)} = 0.288$, $\beta_{(2)} = -6.3705$, $\sigma^2_{1(2)} = 29.983$, and $\sigma^2_{2(2)} = 10.118$. When these values are substituted into the last equation of (6), we have $\mu_{5(2)} = -68{,}186$. For our next approximation to r, we interpolate linearly as indicated below.

r	μ_5
-9.10	-624×10^2
-4.82	-678×10^2
-4.50	-682×10^2

With $r_{(3)} = -4.82$, we subsequently calculate the remaining third approximations just as the second approximations were calculated except that this time we retain additional significant digits. We accordingly obtain $v_{(3)} = -23.521$, $\beta_{(3)} = -6.52777$, $m_{1(3)} = -7.826$, $m_{2(3)} = 3.006$, $\alpha_{(3)} = 0.2775$, $\sigma^2_{1(3)} = 28.8156$, $\sigma^2_{2(3)} = 10.3167$, and finally $\mu_{5(3)} = -67{,}863$.

Extrapolation using the values of μ_5 calculated above with $r_{(2)} = -4.50$ and $r_{(3)} = -4.82$ yields $r_{(4)} = -4.87$ as the next and final five-moment estimate of r. Corresponding to this value for r, final five-moment estimates of other

TABLE 1

Summary of Estimates for Various Trial Values of r

Parameters Estimated	r					
	-4.87	-4.82	-4	-3	-2.5	-2
v	-23.575	-23.521	-22.481	-20.098	-20.001	-19.022
m_1	-7.867	-7.826	-7.146	-6.312	-5.894	-5.475
m_2	2.997	3.006	3.146	3.312	3.394	3.475
β	-6.5535	-6.5278	-6.1344	-5.7287	-5.5644	-5.4317
σ_1^2	28.6397	28.8156	31.7464	35.2933	37.0327	38.7382
σ_1	5.352	5.368	5.635	5.941	6.086	6.224
σ_2^2	10.3513	10.3167	9.7791	9.0327	8.5705	8.0246
σ_2	3.218	3.212	3.127	3.006	2.928	2.833
θ_1	46.456	46.497	47.177	48.011	48.429	48.848
θ_2	57.320	57.329	57.469	57.635	57.717	57.798
α	0.2759	0.2775	0.3057	0.3441	0.3654	0.3883
μ_5	$-67{,}821$	$-67{,}863$	$-68{,}657$	$-69{,}499$	$-69{,}835$	$-70{,}256$
χ^2	3.20	3.06	1.58	0.80	0.74	0.98

parameters of interest become $\theta_1^* = 46.456$, $\sigma_1^* = 5.352$, $\theta_2^* = 57.320$, $\sigma_2^* = 3.218$ and $\alpha^* = 0.2759$.

To calculate conditional minimum chi-square estimates, we require values of χ^2 for several values of r in the vicinity of the minimum. Since estimates of the basic parameters are already available for $r = -4.82$ and $r = -4.87$, we calculate expected frequencies and in turn χ^2 for these values of r and subsequently make the same calculations for $r = -4, -3, -2.5$ and -2, utilizing equations (29), (25), (26), (27) and (28) as previously described. Results of the pertinent calculations are summarized in Table 1 which follows.

When the values of χ^2 from Table 1 are plotted against corresponding values of r, it is readily observed that the minimum value ($\chi^2 = 0.72$) occurs when $r^{**} = -2.65$. With this value for r in (29), corresponding estimates for the remaining parameters of interest are computed as before. Accordingly as final conditional minimum chi-square estimates, we find $\theta_1^{**} = 48.304$, $\sigma_1^{**} = 6.042$, $\theta_2^{**} = 57.692$, $\sigma_2^{**} = 2.951$ and $\alpha^{**} = 0.3589$.

For comparison, the expected frequencies based both on the five-moment estimates and the conditional minimum chi-square estimates are shown in Table 2 along with the observed frequencies.

Agreement between observed frequencies and expected frequencies based on either set of estimates is satisfactory. However, in view of the large sampling errors inherent in the fifth sample moment, it should come as no surprise to find that in this example, χ^2 for the five-moment estimates is substantially larger than that based on the conditional minimum chi-square estimates.

TABLE 2

Observed and Expected Frequencies for 1006 Observations from a Compound Normal Distribution

		Expected Frequencies	
Class Boundaries	Observed Frequencies	Based on Five-Moment Estimates	Based on Cond. Min. χ^2 Estimates
27.5–31.5	1	.7	.9
31.5–35.5	5	4.9	5.2
35.5–39.5	20	21.2	20.1
39.5–43.5	52	53.7	50.8
43.5–47.5	86	80.5	84.6
47.5–51.5	98	94.1	103.3
51.5–55.5	200	217.9	201.6
55.5–59.5	363	349.5	353.8
59.5–63.5	164	163.3	167.8
63.5–67.5	15	19.5	17.5
67.5–71.5	2	.6	.5
χ^2		3.20	0.72
d.f.		3	3
P		0.362	0.868

For comparison, both sets of estimates calculated from the mixed sample are shown below along with corresponding estimates based on the individual components.

Comparison of Estimates

Parameters	Component Estimates	Moment Estimates	Min. χ^2 Estimates
θ_1	47.716	46.456	48.304
θ_2	57.607	57.320	57.692
σ_1	5.794	5.352	6.042
σ_2	3.030	3.218	2.951
α	0.3320	0.2759	0.3589

In concluding, it is deemed appropriate to emphasize that the methods presented in this paper are recommended only with large samples. Furthermore, it is desirable that moments be calculated from the raw ungrouped data when possible. If the ungrouped data are not available, then at least grouping intervals should be relatively narrow in order to minimize errors in the higher moments from this source. When moments can only be computed from grouped data, it will usually be advisable to apply Sheppard's corrections.

References

1. Blischke, W. R., 1955. Estimating the parameters of mixtures of binomial distributions. *Journal of the Am. Stat. Association, 59,* 510–528.
2. Brown, George M., 1933. On sampling from compound populations. *Annals of Mathematical Statistics, 4,* 288–342.
3. Charlier, C. V. L., 1906. Researches into the theory of probability. *Meddelanden frau Lunds Astron. Observ.*, Sec. 2, Bd. 1.
4. Charlier, C. V. L. and Wicksell, S. D., 1924. On the dissection of frequency functions. *Arkiv för Matematik, Astronomi och Fysik,* Bd. 18, No. 6.
5. Cohen, A. Clifford, Jr., 1953. On some conditions under which a compound normal distribution is unimodal. *Technical Report No. 7, Contract DA-01-099 ORD-288,* Univ. of Georgia.
6. Cohen, A. Clifford, Jr., 1963. Estimation in mixtures of discrete distributions. *Proceedings of the International Symposium on the Classical and Contagious Discrete Distributions,* Montreal, Pergamon Press, 373–378.
7. Cohen, A. Clifford, Jr., 1964. Estimation in mixtures of two Poisson distributions. *Aero-Astrodynamics, Research and Development, Research Review No. 1, NASA TM X-53189,* 104–107.
8. Cohen, A. Clifford, Jr., 1965. Estimation in mixtures of Poisson and mixtures of exponential distributions. NASA Technical Memorandum, *NASA TM X-53245,* George C. Marshall Space Flight Center, Huntsville, Alabama.
9. Kendall, M. G., 1948. *The Advanced Theory of Statistics,* Vols. I and II. Chas. Griffin and Co., Ltd., London.
10. Pearson, Karl, 1894. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society, 185,* 71–110.
11. Rao, C. Radhakrishna, 1952. *Advanced Statistical Methods in Biometric Research,* John Wiley and Sons, New York, 300–304.
12. Rider, Paul R., 1962. Estimating the parameters of mixed Poisson, binomial and Weibull distributions by the method of moments. *Bulletin de l'Institut International de Statistique, 39.*