MIXTURES, HYBRID MIXTURES, CANONICAL FORMS, AND MATUSITA'S DISTANCE

K. O. BOWMAN*
Computational Sciences and Engineering Division
Oak Ridge National Laboratory
P.O.Box 2008, 4500N, MS-6237
Oak Ridge, Tennessee 37831-6237, U.S.A.
bowmanko@ornl.gov

L. R. SHENTON
Department of Statistics
University of Georgia
Athens, Georgia 30602, U.S.A.

September 18, 2003

Abstract

First of all we introduce an orthogonal system associated with the mixture probability function. Then for maximum likelihood estimators, Parseval expansions are introduced related to logarithmic derivatives of the mixture probability function. In this way approximants are set up for the maximum likelihood covariance determinant.

For a hybrid mixture of Poisson-binomial, three parameters are involved and the covariance determinant approximant turns out to be a quartic for a ratio r, this relating to the means of the two components in the mixture. We prove explicitly that the zeros of this quartic are strictly complex; a mathematical proof is needed since computational values do not state the nature of the zeros. A study of a hybrid mixture of Poisson-negative binomial, two components, shows that the covariance approximant is a quartic, again having strictly complex zeros. In addition we briefly consider mixtures of gamma components, firstly with scale varying with fixed shape, and secondly, shape varying with fixed scale. Another mixture mentioned consists of two lognormal densities.

Lastly it is well known that estimation procedures for mixtures are sensitive to sample size, because of the closeness of the individual components. From a statistical point of view this characteristic is almost self-evident. There is therefore interest in a distance concept of probability function differences dual to Matusita (1955) who also introduces the dual notion of affinity between two probability functions.

Key words and phrases; affinity, covariance determinant, covariance matrix, distance, maximum likelihood, orthogonal system, Parseval expansion, persymmetric determinant, quartic zeros.

1 Introduction

In a previous paper (Bowman and Shenton, 2003) we have used approximants to the maximum likelihood covariance determinant to throw light on the asymptotic variances of maximum likelihood estimators of the parameters in Poisson mixture and binomial mixture distributions. For these distributions, the probability functions are

$$\text{Poisson}: \quad P_p(x, \underline{\theta}, \underline{\pi}) = \sum_{r=1}^{s} \pi_r \frac{e^{-\theta_r}\theta_r^x}{x!} \quad (\sum \pi_r = 1, 0 < \pi_r < 1, \theta_r > 0) \tag{1}$$

$$\text{Binomial}: \quad P_b(x; n, \underline{\theta}) = \sum_{r=1}^{s} \pi_r \binom{n}{x} \theta_r^x (1-\theta_r)^{n-x}. \quad (0 < \theta_r < 1, n = 1, 2, \cdots)$$

We give an example for the Poisson mixture.

For $s = 2$ in (1), the maximum likelihood estimators are $\hat{\theta_1}$, $\hat{\theta_2}$, and $\hat{\pi_1}$.

Then for the first order asymptotic variances,

$$\begin{cases} Var_1(\hat{\theta_1}) \sim \{W_3/W_2\} / \{\pi_1^2(\theta_1 - \theta_2)^4\}, \\ Var_1(\hat{\theta_2}) \sim \{W_3/W_2\} / \{\pi_2^2(\theta_1 - \theta_2)^4\}, \\ Var_1(\hat{\pi_1}) \sim \{4W_3/W_2\} / \{(\theta_1 - \theta_2)^6\}, \end{cases}$$

where W_s is the persymmetric determinant

$$W_s = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_s \\ \mu_1 & \mu_2 & \cdots & \mu_{s+1} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_s & \mu_{s+1} & \cdots & \mu_{2s} \end{vmatrix}.$$

What happens when we consider a two component mixture of a Poisson and a binomial distribution? An account of discrete mixtures is given in Johnson, Kotz and Kemp (1993). The essential quantity is the denominator of the variance matrix in determinant form. In this paper we give the canonical form for the two component hybrid mixture (Poisson-binomial), and their covariance denominator complex zeros.

The question arises as to the characteristics of a two component mixture that ensure strictly complex zeros associated with the maximum likelihood covariance denominator. We have studied a mixture of (a) Poisson and negative binomial distributions, (b) gamma components, one in which the shape varies, and the other in which the scale varies, (c) log-normal components. It appears that zeros associated with the maximum likelihood covariance denominator are only strictly complex for two cases, first the hybrid Poisson-binomial, and second the hybrid Poisson-negative binomial distribution.

It is clear that the closeness of components in a mixture will create an estimation problem; for example take the case of a Poisson mixture with parameters $\theta_1 = 1$, and $\theta_2 = 1.01$. The hybrid cases of Poisson-binomial, and Poisson-negative binomial have support for $x = 0, 1, 2, \cdots$; is this infinite range the clue? Note that our Parseval expansions relate to derivatives and differences; thus for the two component Poisson, we have expansions associated with $\frac{\partial \ln P}{\partial \theta_1}$, $\frac{\partial \ln P}{\partial \theta_2}$, and $P_1(x) - P_2(x)$; derivetives and differences. These derivatives and differences also occur in the maximum likelihood covariance matrix.

The notion of differences has been associated with distance and affinity by Matusita (1955), who looks at the distance $\sqrt{p_1(x)} - \sqrt{p_2(x)}$ between two probability functions. If the Matusita's distance is zero then there is high affinity. This aspect of mixtures is mentioned in the study.

2 A Poisson-Binomial Mixture Distribution

2.1 Basic formulas

$$P(x; \underline{\theta}, \underline{\pi}) = \frac{\pi_1 e^{-\theta_1} \theta_1^x}{x!} + \pi_2 \binom{n}{x} \theta_2^x (1 - \theta_2)^{n-x}.$$

Mean, $\mu_1' = \pi_1 \theta_1 + (1 - \pi_1) n \theta_2 \quad (\theta_1 > 0, 0 < \theta_2 < 1, n = 2, 3, \cdots, \pi_1 + \pi_2 = 1)$.
Factorial Moments: Poisson θ^s, Binomial $n^{(s)} \theta^s$, $s = 0, 1, \cdots$. For the mixture of Poisson-binomial, factorial moments are

$$\mu_{[s]}' = \pi_1 \theta_1^s + (1 - \pi_1) n^{(s)} \theta_2^s \quad (0 < \pi_1 < 1).$$

Maximum Likelihood Estimators are $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\pi}_1$,

$$\frac{1}{\pi_1}\frac{\partial \ln P}{\partial \theta_1} = \frac{(x-\theta_1)}{\theta_1}\frac{e^{-\theta_1}\theta_1^x/x!}{P},$$

$$\frac{1}{\pi_2}\frac{\partial \ln P}{\partial \theta_2} = \frac{(x-n\theta_2)}{\theta_2(1-\theta_2)}\frac{\begin{pmatrix} n \\ x \end{pmatrix}\theta_2^x(1-\theta_2)^{n-x}}{P},$$

$$\frac{\partial \ln P}{\partial \pi_1} = \frac{1}{P}\left\{\frac{e^{-\theta_1}\theta_1^x}{x!} - \begin{pmatrix} n \\ x \end{pmatrix}\theta_2^x(1-\theta_2)^{n-x}\right\}.$$

Parseval Expansions are

$$\frac{(x-\theta_1)}{\theta_1}\cdot\frac{e^{-\theta_1}\theta_1^x}{x!} \sim \{A_0q_0(x) + A_1q_1(x) + \cdots\}\,P,$$

$$\frac{(x-n\theta_2)}{\theta_2(1-\theta_2)}\begin{pmatrix} n \\ x \end{pmatrix}\theta_2^x(1-\theta_2)^{n-x} \sim \{B_0q_0(x) + B_1q_1(x) + \cdots\}\,P,$$

$$\frac{e^{-\theta_1}\theta_1^x}{x!} - \begin{pmatrix} n \\ x \end{pmatrix}\theta_2^x(1-\theta_2)^{n-x} \sim \{C_0q_0(x) + C_1q_1(x) + \cdots\}\,P$$

where $P \equiv P(x;\underline{\theta},\underline{\pi})$.

We use an orthogonal system

$$\sum_{x=0}^{\infty} q_r(x)q_s(x)P(x;\underline{\theta},\underline{\pi}) = \phi_r\delta_{r,s} \qquad (r,s=0,1,\cdots)$$

and

$$q_r(x) = [1, X, X^2, \cdots, X^r]/W_{r-1} \qquad (X = x-\mu_1', r=1,2,\cdots;, W_0 = 1)$$

$$W_r = [\mu_0, \mu_1, \cdots, \mu_r]$$

i.e. $q_0 = 1$, $q_1 = X = x - \mu_1'$. μ_1' is the mean of the mixture random variate.

$$q_2(x) = [1, X, X^2]/W_1, \qquad W_1 = [u_0, \mu_1] = \begin{vmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{vmatrix}.$$

See Bowman and Shenton, 2003, equations (4) and (5).

2.2 Covariance approximation

$$\Delta = \frac{\pi_1^2 \pi_2^2}{\phi_1 \phi_2 \phi_3} \begin{vmatrix} A_1\phi_1 & A_2\phi_2 & A_3\phi_3 \\ B_1\phi_1 & B_2\phi_2 & B_3\phi_3 \\ C_1\phi_1 & C_2\phi_2 & C_3\phi_3 \end{vmatrix}^2 .$$

$$\phi_1 = W_1 = \mu_2, \qquad \phi_2 = W_2/W_1, \qquad \phi_3 = W_3/W_2.$$

We have

$$A_1\phi_1 = 1,$$

$$A_2\phi_2 = \begin{vmatrix} 0 & 1 & 1+2\lambda_1 \\ \mu_0 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \end{vmatrix} \div W_1 = d_3[0, 1, \alpha_2]/W_1$$

and $\alpha_2 = 1 + 2(\theta_1 - \mu_1') = 1 + 2\lambda_1$, $\lambda_1 = \theta_1 - \mu_1'$.

$$A_3\phi_3 = d_4[0, 1, \alpha_2, \alpha_3]/W_2$$

where

$$\begin{aligned}
\alpha_3 &= \sum_{x=0}^{\infty} \left(\frac{x-\theta_1}{\theta_1}\right) \frac{e^{-\theta_1}\theta_1^x}{x!}(x-\mu_1')^3 \\
&= \sum_{x=0}^{\infty} (x-\mu_1')^3 \frac{\partial}{\partial\theta_1} P_1(x;\theta_1) \qquad (P_1 \equiv e^{-\theta_1}\theta_1^x/x!) \\
&= \frac{\partial}{\partial\theta_1}\nu_3' - 3\mu_1' \frac{\partial}{\partial\theta_1}\nu_2' + 3\mu_1'^2 \frac{\partial}{\partial\theta_1}\nu_1' \\
&= \frac{\partial}{\partial\theta_1}(\nu_{[3]} + 3\nu_{[2]} + \nu_{[1]}) - 3\mu_1' \frac{\partial}{\partial\theta_1}(\nu_{[2]} + \nu_{[1]}) + 3\mu_1'^2 \frac{\partial}{\partial\theta_1}\nu_{[1]}
\end{aligned}$$

using Stirling numbers of the Second Kind, and factorial moments related to $P_1(x, \theta_1)$. Hence

$$\alpha_3 = 3\theta_1^2 + 6\theta_1 + 1 - 3\mu_1'(2\theta_1 + 1) + 3\mu_1'^2.$$

Similarly for $B_1\phi_1$, $B_2\phi_2$ and $B_3\phi_3$.

$$B_1\phi_1 = n = \beta_1,$$

$$B_2\phi_2 = d_3[0, \beta_1, \beta_2]/W_1,$$

$$B_3\phi_3 = d_4[0, \beta_1, \beta_2, \beta_3]/W_2$$

where

$$\beta_2 = n\left\{1 + 2[(n-1)\theta_2 - \mu_1']\right\},$$
$$\beta_3 = n\left\{3(n-1)^{(2)}\theta_2^2 + 6(n-1)\theta_2 + 1 - 3\mu_1'[2(n-1)\theta_2 + 1] + 3\mu_1'^2\right\}.$$

For $C_1\phi_1$, $C_2\phi_2$ and $C_3\phi_3$.

$$C_1\phi_1 = \sum_{x=0}^{\infty}\left\{\frac{e^{-\theta_1}\theta_1^x}{x!} - \binom{n}{x}\theta_2^x(1-\theta_2)^{n-x}\right\}(x - \mu_1') = \theta_1 - n\theta_2 = \gamma_1,$$
$$C_2\phi_2 = d_3[0,\gamma_1,\gamma_2]/W_1,$$
$$C_3\phi_3 = d_4[0,\gamma_1,\gamma_2,\gamma_3]/W_2$$

where

$$\begin{aligned}\gamma_2 &= \theta_1^2 + \theta_1 - (n^{(2)}\theta_2^2 + n\theta_2) - 2\mu_1'(\theta_1 - n\theta_2)\\ &= \theta_1^2 - n^{(2)}\theta_2^2 + \theta_1 - n\theta_2 - 2\mu_1'(\theta_1 - n\theta_2)\\ &= (\theta_1 - n\theta_2)(\theta_1 + n\theta_2 + 1 - 2\mu_1') + n\theta_2^2.\end{aligned}$$

Note that $\theta_1 - n\theta_2$ is not a factor unless $n\theta_2^2 = 0$.

$$\begin{aligned}\gamma_3 &= \sum_{x=0}^{\infty}\left\{\frac{e^{-\theta_1}\theta_1^x}{x!} - \binom{n}{x}\theta_2^x(1-\theta_2)^{n-x}\right\}(x - \mu_1')^3\\ &= \theta_1 + 3\theta_1^2 + \theta_1^3 - (n\theta_2 + 3n^{(2)}\theta_2^2 + n^{(3)}\theta_2^3) - 3\mu_1'(\theta_1 + \theta_1^2 - n\theta_2 - n^{(2)}\theta_2^2) + 3\mu_1'^2(\theta_1 - n\theta_2).\end{aligned}$$

Again $\theta_1 - n\theta_2$ is not a factor unless $\gamma_3 = 0$ where

$$\begin{aligned}\gamma_3 &= 3\theta_2^2(n^2 - n^{(2)}) + \theta_2^3(n^3 - n^{(3)}) - 3\mu_1'\theta_2^2(n^2 - n^{(2)})\\ &= n\theta_2^2(3n\theta_2 - 2\theta_2 + 3 - 3\mu_1')\\ &\neq 0 \qquad (n = 1, 2, \cdots)\end{aligned}$$

We have now $\Delta_3 = \frac{\pi_1^2\pi_2^2}{\phi_1\phi_2\phi_3}\cdot(\Delta_3^*)^2$, where

$$\Delta_3^* = \begin{vmatrix} 1 & d_3[0,\alpha_1,\alpha_2]/W_1 & d_4[0,\alpha_1,\alpha_2,\alpha_3]/W_2 \\ n & d_3[0,\beta_1,\beta_2]/W_1 & d_4[0,\beta_1,\beta_2,\beta_3]/W_2 \\ \theta_1 - n\theta_2 & d_3[0,\gamma_1,\gamma_2]/W_1 & d_4[0,\gamma_1,\gamma_2,\gamma_3]/W_2 \end{vmatrix}$$

a determinant whose elements are determinants. Inspection reveals no factors, and there is support for this statement using the function [factor] from MAPLE system.

2.3 Zeros of Δ_3^* and the covariance denominator

By linear reductionism by columns,

$$\Delta_3^* = n \begin{vmatrix} 1 & 2\theta_1 & 3\theta_1^2 \\ 1 & 2(n-1)\theta_2 & 3(n-1)^{(2)}\theta_2^2 \\ \theta_1 - n\theta_2 & \theta_1^2 - n^{(2)}\theta_2^2 & \theta_1^3 - n^{(3)}\theta_2^3 \end{vmatrix}$$

and by expansion

$$\Delta_3^* = n\theta_2^4[r^4 - 4(n-1)r^3 + 6(n-1)^2r^2 - 4n^{(3)}r + (n-1)n^{(3)}] \tag{2}$$

where $r = \theta_1/\theta_2$. Defining $X = r - (n-1)$,

$$\Delta_3^* = n\theta_2^4[X^4 + 4(n-1)X + 3(n-1)^2].$$

The quartic in X has zeros depending on the real root of the cubic

$$u^3 - 12(n-1)^2u - 16(n-1)^2 = 0,$$

for which the discriminant $\Delta = -64(n-1)^4(n^2-2n)$. There is therefore one real root, which is

$$u_1 = 4(n-1)\cos\left(\frac{1}{3}\arctan\sqrt{n^2-2n}\right) = 4(n-1)\cos(\theta/3)$$
$$(n = 2, 3, \cdots, 0 < \theta < \pi/2, \theta = \theta(n)).$$

The formula in NBS (1964, p17, 3.8.3) is incorrect. The sign in the final quartic (the product of two quadratics) should be $\pm$, $\mp$ and not $\pm$, $\pm$. Using the formula CRC Tables (1996)

$$\begin{aligned} r_1 &= n-1+\sqrt{(n-1)}\left\{-\sqrt{\phi_1} + \mathbf{i}\sqrt{\phi_1 - \sqrt{4\phi_1^2-3}}\right\}, \\ r_2 &= n-1+\sqrt{(n-1)}\left\{-\sqrt{\phi_1} - \mathbf{i}\sqrt{\phi_1 - \sqrt{4\phi_1^2-3}}\right\}, \\ r_3 &= n-1+\sqrt{(n-1)}\left\{\sqrt{\phi_1} + \mathbf{i}\sqrt{\phi_1 + \sqrt{4\phi_1^2-3}}\right\}, \\ r_4 &= n-1+\sqrt{(n-1)}\left\{\sqrt{\phi_1} - \mathbf{i}\sqrt{\phi_1 + \sqrt{4\phi_1^2-3}}\right\}. \end{aligned} \tag{3}$$

$$(n = 2, 3, \cdots, \phi_1 = \cos\frac{\theta}{3}, \tan\theta = \sqrt{n(n-2)}).$$

Clearly $\sqrt{4\phi_1^2 - 3} < \phi_1$ for $n = 3, 4, \cdots$, $0 < \theta < \pi/2$, so that the four zeros are complex of the form $a_j + ib_j$, $j = 1, 2, 3, 4$, and $b_j \neq 0$. When $n = 1$ the quartic in r reduces to $\theta_2^4(r^4)$, and when $n = 2$ the zeros are 0, 0, $2 \pm i\sqrt{2}$. The zeros of the quartic in (2) are complex and involve irrationals such as $\sqrt{a + \sqrt{b}}$ where a and b are exact terms; the irrationals can not be identified using finite arithmetic on a computer. As a check (3) insert $r - (n-1) = X$ in (2) using MAPLE. Clearly then the associated Δ_3^* can not be zero under the specified domain of the parameters.

2.4 Asymptotic variance approximants

We find

$$\begin{aligned} Var_1(\hat{\theta}_1) &\sim \frac{\begin{vmatrix} n & d_3[0, \beta_1, \beta_2]/W_1 \\ \theta_1 - n\theta_2 & d_3[0, \gamma_1, \gamma_2]/W_1 \end{vmatrix}^2 \phi_3}{\pi_1^2(\Delta_3^*/n)^2}, \\ &= \frac{(W_3/W_2)[\gamma_2 - (\theta_1 - n\theta_2)\beta_2/n]^2}{\pi_1^2(\Delta_3^*/n)^2} \end{aligned}$$

and

$$\begin{aligned} \gamma_2 - (\theta_1 - n\theta_2)\beta_2/n &= (\theta_1 - n\theta_2)(\theta_1 + n\theta_2 + 1 - 2\mu_1') + n\theta_2^2 - (\theta_1 - n\theta_2)(1 + 2[(n-1)\theta_2 - \mu_1']) \\ &= (\theta_1 - n\theta_2)(\theta_1 - n\theta_2 + 2\theta_2) + n\theta_2^2 \end{aligned}$$

so that

$$Var_1(\hat{\theta}_1) \sim \frac{[(\theta_1 - n\theta_2)(\theta_1 - n\theta_2 + 2\theta_2) + n\theta_2^2]^2 W_3/W_2}{\pi_1^2(\Delta_3^*/n)^2},$$

or

$$Var_1(\hat{\theta}_1) \sim \frac{[(\theta_1 - n\theta_2)(\theta_1 - n\theta_2 + 2\theta_2) + n\theta_2^2]^2 W_3/W_2}{\pi_1^2\{[\theta_1 - (n-1)\theta_2]^4 + 4(n-1)[\theta_1 - (n-1)\theta_2]\theta_2^3 + 3(n-1)^2\theta_2^4\}^2}.$$

Similarly

$$Var_1(\hat{\theta}_2) \sim \frac{\begin{vmatrix} 1 & d_3[0, \alpha_1, \alpha_2]/W_1 \\ \theta_1 - n\theta_2 & d_3[0, \gamma_1, \gamma_2]/W_1 \end{vmatrix}^2 \phi_3}{\pi_2^2(\Delta_3^*)^2},$$

$$= \frac{(W_3/W_2)[\gamma_2 - (\theta_1 - n\theta_2)\alpha_2]^2}{\pi_2^2(\Delta_3^*)^2}$$

and

$$\begin{aligned}\gamma_2 - (\theta_1 - n\theta_2)\alpha_2 &= (\theta_1 - n\theta_2)[\theta_1 + n\theta_2 + 1 - 2\mu_1' - 1 - 2(\theta_1 - \mu_1')] + n\theta_2^2 \\ &= (\theta_1 - n\theta_2)(n\theta_2 - \theta_1) + n\theta_2^2.\end{aligned}$$

Thus,

$$Var_1(\hat{\theta_2}) \sim \frac{[(\theta_1 - n\theta_2)(n\theta_2 - \theta_1) + n\theta_2^2]^2 W_3/W_2}{\pi_2^2(\Delta_3^*)^2},$$

or

$$Var_1(n\hat{\theta_2}) \sim \frac{[(\theta_1 - n\theta_2)^2 - n\theta_2^2]^2 W_3/W_2}{\pi_2^2\{[\theta_1 - (n-1)\theta_2]^4 + 4(n-1)[\theta_1 - (n-1)\theta_2]\theta_2^3 + 3(n-1)^2\theta_2^4\}^2}.$$

Similarly,

$$Var_1(\hat{\pi_1}) \sim \frac{4(W_3/W_2)[(n-1)\theta_2 - \theta_1]^2}{(\Delta_3^*/n)^2}.$$

or

$$Var_1(\hat{\pi_1}) \sim \frac{4(W_3/W_2)[\theta_1 - (n-1)\theta_2]^2}{\{[\theta_1 - (n-1)\theta_2]^4 + 4(n-1)[\theta_1 - (n-1)\theta_2]\theta_2^3 + 3(n-1)^2\theta_2^4\}^2}.$$

Comments:

- Compare with the 3pPoisson $(\hat{\theta_1}, \hat{\theta_2}, \hat{\pi_1})$, in which, for example,

  $$Var_1(\hat{\pi_1}) \sim \frac{4(W_3/W_2)}{(\theta_1 - \theta_2)^6}$$

  and

  $$Var_1(\hat{\theta_1}) \sim \frac{W_3/W_2}{\pi_1^2(\theta_1 - \theta_2)^4}, \qquad Var_1(\hat{\theta_2}) \sim \frac{W_3/W_2}{\pi_2^2(\theta_1 - \theta_2)^4},$$

- It appears that in this case of a hybrid mixture (Poisson and binomial), variances of maximum likelihood estimators are not likely to be sensitive to the difference of the means, i.e. $(\theta_1 - n\theta_2)$. In other words, in the case of two components, affinity between components is a major estimation problem.

3 Poisson-Negative Binomial Mixture Distribution

Probability density:

$$P(x;\underline{\theta},\underline{\pi}) = \frac{\pi_1 e^{-\theta_1}\theta_1^x}{x!} + \pi_2 \frac{\Gamma(k+x)}{\Gamma(k)} \frac{\theta_2^x}{(1+\theta_2)^{x+k}} \quad (k>0, \theta_2>0)$$

Probability generating function of Poisson: $e^{\theta_1(t-1)}$ $(\theta_1 > 0)$

Probability generating function of negative binomial: $(\theta_2 + 1 - \theta_2 t)^{-k}$ $(\theta_2 > 0)$

Factorial Moments: Poisson θ_1^s, negative binomial $[\Gamma(k+s)/\Gamma(k)]\theta_2^s$, $(s = 0, 1, \cdots)$.

Means are, for Poisson θ_1; for negative binomial $k\theta_2$.

The unexpurgated version of Δ_3^* in this case, under row by row reductionism reduces to, apart from a constant,

$$\Delta_3^* = \begin{vmatrix} 1 & 2\theta_1 & 3\theta_1^2 \\ 1 & 2(k+1)\theta_2 & 3(k+2)^{(2)}\theta_2^2 \\ \theta_1 - k\theta_2 & \theta_1^2 - (k+1)^{(2)}\theta_2^2 & \theta_1^3 - (k+2)^{(3)}\theta_2^3 \end{vmatrix},$$

having used factorial moments. Then defining $r = \theta_1/\theta_2$

$$\begin{aligned} \Delta_3^* &= \theta_2^4 \begin{vmatrix} 1 & 2r & 3r^2 \\ 1 & 2(k+1) & 3(k+2)^{(2)} \\ r-k & r^2-(k+1)^{(2)} & r^3-(k+1)^{(3)} \end{vmatrix} \\ &= \theta_2^4\{r^4 - 4(k+1)r^3 + 6(k+1)^{(2)}r^2 - 4(k+2)^{(3)}r + (k+1)(k+2)^{(3)}\} \\ &= \theta_2^4\{(r-k-1)^4 + 4(k+1)(r-k-1) + 3(k+1)^2\}. \end{aligned}$$

For the zeros of this quartic we need the real root of the cubic

$$u^3 - 12(k+1)^2u - 16(k+1)^2 = 0.$$

The discriminant is

$$\Delta = 64(k+1)^4 - 64(k+1)^6 = -64(k+1)^4(k^2+2k) \quad (k>0).$$

Hence there is one real root, and we have the irreducible case. We use

$$p = r\cos\theta, \quad q = r\sin\theta$$

where

$$p = 8(k+1)^2, \qquad q = 8(k+1)^2\sqrt{k^2+2k},$$

yielding $\tan\theta = \sqrt{k^2+2k}$, and $r^2 = 64(k+1)^4(k+1)^2$, the solution to the cubic being

$$u_1 = 4(k+1)\cos\left(\frac{1}{3}\arctan\sqrt{k^2+2k}\right).$$

The zeros are found from the two quadratic

$$v^2 \pm \sqrt{u_1}v + \frac{u_1}{2} \mp \left[\left(\frac{u_1}{2}\right)^2 - a_0\right]^{1/2} = 0$$

(see CRC, 1996) where $a_0 = 3(k+1)^2$, yielding,

$$r_{1,2} = (k+1) + \sqrt{(k+1)}\left\{-\psi \pm \mathbf{i}\sqrt{\psi - \sqrt{4\psi^2-3}}\right\},$$

$$r_{3,4} = (k+1) + \sqrt{(k+1)}\left\{\psi \pm \mathbf{i}\sqrt{\psi + \sqrt{4\psi^2-3}}\right\}$$

where $k > 0$, and

$$\psi = \cos\left(\frac{1}{3}\arccos\frac{1}{k+1}\right).$$

Here the roots of the quartic equations are strictly complex, so that the quartic can not be zero under the stated conditions.

4 Mixture of Gamma Distributions

4.1 Gamma distributions with known shape parameter

The mixture, in the 5p case $(a_1, a_2, a_3, \pi_1, \pi_2)$ is

$$P(x; \underline{a}, \underline{\pi}) = \sum_{r=1}^{3} \frac{\pi_r e^{-x/a_r}(x/a_r)^{\rho-1}}{a_r\Gamma(\rho)} = \sum_{r=1}^{3}\pi_r P_r(x) \qquad (x > 0, \rho > 0, a_1, a_2, a_3 > 0).$$

This 5p case is chosen since it relates to the significant characteristics of the general case. Now

$$\frac{\partial P_r(x; a_r)}{\partial a_r} = \frac{(x - \rho a_r)}{a_r^2}P_r(x; a_r) \qquad (r = 1, 2, 3)$$

and

$$\Delta_5 = \frac{\pi_1^2\pi_2^2\pi_3^2}{\phi_1\phi_2\phi_3\phi_4\phi_5}(\Delta_5^*)^2$$

where

$$\Delta_5^* = \begin{vmatrix} 1 & d_3[0,\alpha_1^{(1)},\alpha_2^{(1)}]/W_1 & \cdots & d_6[0,\alpha_1^{(1)},\alpha_2^{(1)},\cdots,\alpha_5^{(1)}]/W_4 \\ 1 & d_3[0,\alpha_1^{(2)},\alpha_2^{(2)}]/W_1 & \cdots & d_6[0,\alpha_1^{(2)},\alpha_2^{(2)},\cdots,\alpha_5^{(2)}]/W_4 \\ 1 & d_3[0,\alpha_1^{(3)},\alpha_2^{(3)}]/W_1 & \cdots & d_6[0,\alpha_1^{(3)},\alpha_2^{(3)},\cdots,\alpha_5^{(3)}]/W_4 \\ a_1-a_3 & d_3[0,\gamma_1^{(1)},\gamma_2^{(1)}]/W_1 & \cdots & d_6[0,\gamma_1^{(1)},\gamma_2^{(1)},\cdots,\gamma_5^{(1)}]/W_4 \\ a_2-a_3 & d_3[0,\gamma_1^{(2)},\gamma_2^{(2)}]/W_1 & \cdots & d_6[0,\gamma_1^{(2)},\gamma_2^{(2)},\cdots,\gamma_5^{(2)}]/W_4 \end{vmatrix} \tag{4}$$

in the notation of §2, along with,

$$\alpha_s^{(r)} = \sum_{t=0}^{s}(-1)^t \begin{pmatrix} s \\ t \end{pmatrix} (\mu_1')^t(\rho+s-t-1)^{(s-t)}\frac{\partial}{\partial a_r}a_r^{s-t} \qquad (r=1,2,\cdots,s) \tag{5}$$

since the non-central moments of the regular gamma density $\{e^{-x/a}(x/a)^{\rho-1}/[a\Gamma(\rho)]\}$ are $\nu_r' = a^r(\rho+r-1)^{(r)}$. Moreover

$$\begin{aligned} \gamma_s^{(r)} &= \int_0^\infty \left\{\frac{e^{-x/a_1}(x/a_1)^{\rho-1}}{a_1\Gamma(\rho)} - \frac{e^{-x/a_3}(x/a_3)^{\rho-1}}{a_3\Gamma(\rho)}\right\}(x-\mu_1')^{(s)}dx \\ &= (\rho+s-1)^{(s)}(a_1^s-a_3^s) - \begin{pmatrix} s \\ 1 \end{pmatrix}\mu_1'(\rho+s-2)^{(s-1)}(a_1^{s-1}-a_3^{s-1}) \\ &+ \begin{pmatrix} s \\ 2 \end{pmatrix}\mu_1'^2(\rho+s-3)^{(s-2)}(a_1^{s-2}-a_3^{s-2}) + \cdots + (-1)^s\begin{pmatrix} s \\ s \end{pmatrix}\mu_1'^s(a_1^0-a_3^0) \end{aligned} \tag{6}$$

Comparing (4) and (5) we note symbolic isomorphism

$$\frac{a_1^s-a_3^s}{a_1-a_3} \longrightarrow s a_i^{s-1}, \qquad \frac{a_1^{s-1}-a_3^{s-1}}{a_1-a_3} \longrightarrow (s-1)a_i^{s-2}$$

and so on. Hence reduction of determinants in (6) by columns, and noting the common factor in each column, we have

$$\Delta_5^* = (\rho+1)^{(2)}(\rho+2)^{(3)}(\rho+3)^{(4)}(\rho+4)^{(5)} \times \begin{vmatrix} 1 & 2a_1 & 3a_1^2 & 4a_1^3 & 5a_1^4 \\ 1 & 2a_2 & 3a_2^2 & 4a_2^3 & 5a_2^4 \\ 1 & 2a_3 & 3a_3^2 & 4a_3^3 & 5a_3^4 \\ \mathcal{S}_1(1,3) & \mathcal{S}_2(1,3) & \mathcal{S}_3(1,3) & \mathcal{S}_4(1,3) & \mathcal{S}_5(1,3) \\ \mathcal{S}_1(2,3) & \mathcal{S}_2(2,3) & \mathcal{S}_3(2,3) & \mathcal{S}_4(2,3) & \mathcal{S}_5(2,3) \end{vmatrix}$$

The asymptotic variance approximations:

$$Var_1(\hat{a_1}) \sim \{W_5/W_4\}/\{\pi_1^2[(a_1 - a_2)(a_1 - a_3)]^4[(\rho+4)^{(5)}]^2\}$$
$$Var_1(\hat{a_2}) \sim \{W_5/W_4\}/\{\pi_2^2[(a_2 - a_1)(a_2 - a_3)]^4[(\rho+4)^{(5)}]^2\}$$
$$Var_1(\hat{a_3}) \sim \{W_5/W_4\}/\{\pi_3^2[(a_3 - a_1)(a_3 - a_2)]^4[(\rho+4)^{(5)}]^2\}$$
$$Var_1(\hat{\pi_2}) \sim \{4(a_1 - 2a_2 + a_3)^2W_5/W_4\}/\{[(a_2 - a_1)(a_2 - a_3)]^6[(\rho+4)^{(5)}]^2\}$$
$$Var_1(\hat{\pi_1}) \sim \{4(a_2 - 2a_1 + a_3)^2W_5/W_4\}/\{[(a_1 - a_2)(a_1 - a_3)]^6[(\rho+4)^{(5)}]^2\}$$

4.2 Gamma distributions with known scale parameters

In this case

$$P(x; \underline{\rho}, \underline{\pi}) = \sum_{r=1}^{3} \frac{\pi_r e^{-x/a}(x/a)^{\rho_r - 1}}{a\Gamma(\rho_r)} = \sum_{r=1}^{3} \pi_r P_r(x) \qquad (x > 0, \rho_r > 0, a > 0)$$

and the estimators are $\hat{\rho_1}$, $\hat{\rho_2}$, $\hat{\rho_3}$, $\hat{\pi_1}$, $\hat{\pi_1}$. We find for the asymptotic variance approximants

$$Var_1(\hat{\rho_1}) \sim \{W_5/W_4\}/\{\pi_1^2[(\rho_1 - \rho_2)(\rho_1 - \rho_3)]^4 a^{10}\}$$
$$Var_1(\hat{\rho_2}) \sim \{W_5/W_4\}/\{\pi_2^2[(\rho_2 - \rho_1)(\rho_2 - \rho_3)]^4 a^{10}\}$$
$$Var_1(\hat{\rho_3}) \sim \{W_5/W_4\}/\{\pi_3^2[(\rho_3 - \rho_1)(\rho_3 - \rho_2)]^4 a^{10}\}$$
$$Var_1(\hat{\pi_2}) \sim \{4(\rho_1 - 2\rho_2 + \rho_3)^2W_5/W_4\}/\{[(\rho_2 - \rho_1)(\rho_2 - \rho_3)]^6 a^{10}\}$$
$$Var_1(\hat{\pi_1}) \sim \{4(\rho_2 - 2\rho_1 + \rho_3)^2W_5/W_4\}/\{[(\rho_1 - \rho_2)(\rho_1 - \rho_3)]^6 a^{10}\}$$

5 Log-Normal Distribution

Basic Density:

$$P(x;\theta) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}[\ln x - \theta]^2/\sigma^2} \qquad (\theta \ \ real, x > 0)$$

Mixture of Two Distributions

$$P(x; \underline{\theta}) = \pi_1 P(x; \theta_1) + \pi_2 P(x; \theta_2)$$

The rth non-central moments of a single component is

$$\mu'_r = e^{r\theta + \frac{1}{2}r^2\sigma^2}$$

MLE

$$Var_1(\hat{\theta_1}) \sim \frac{W_3/W_2}{\pi_1^2(\theta_1-\theta_2)^4(\exp(\sigma^9))}$$
$$Var_1(\hat{\pi_1}) \sim \frac{4W_3/W_2}{(\theta_2-\theta_1)^6(\exp(\sigma^9))}$$

Mixture of Three Distributions

$$P(x;\underline{\theta}) = \pi_1 P(x;\theta_1) + \pi_2 P(x;\theta_2) + \pi_3 P(x;\theta_3)$$

MLE

$$Var_1(\hat{\theta_1}) \sim \frac{W_5/W_4}{\pi_1^2(\theta_1-\theta_2)^4(\theta_1-\theta_3)^4(\exp(\sigma^{25}))}$$
$$Var_1(\hat{\pi_2}) \sim \frac{4W_5/W_4(\theta_1-2\theta_2+\theta_3)^2}{(\theta_2-\theta_1)^6(\theta_2-\theta_3)^6(\exp(\sigma^{25}))}$$

For an account of the log-normal distribution, see Johnson, Kotz, Balakrishnan, (1994).

6 Further Remarks and Matusita's Distance

We have now concluded Poisson mixtures, binomial mixtures, Poisson-binomial hybrid mixtures, and Poisson-negative binomial hybrid mixtures. The asymptotic variance approximants involve Δ^*, the determinant in the denominator. For Poisson mixtures Δ^* can be zero whenever two components are identical; the zeros, in the case of two components, are of order $(\theta_1 - \theta_2)$ for $\hat{\theta_1}$, $\hat{\theta_2}$, but of order six, $(\theta_1-\theta_2)^6$ for π proportion $\hat{\pi_1}$, estimates being by maximum likelihood. But for the two hybrid cases considered (Poisson-binomial, Poisson-negative binomial), the associated Δ^* is a quartic in r ($r = \theta_1/\theta_2$), and this quartic has 4 strictly complex roots, so that Δ^* can not be zero for the domain of the variates considered. It follows that in a certain sense Δ^* is a measure of closeness of the components.

Matusita (1955) and several subsequent papers introduced the notion of distribution closeness in the form of a distance. Ahmad (1985) gives a very interesting abbreviated account of Matusita's ideas. For the generalized distance, we have

$$\|D\|_r = \left\{ \sum_{x=0}^{\infty} \left[(p_1(x))^{1/r} - (p_2(x))^{1/r} \right]^r \right\}^{1/r}$$

for probability function $p_1(\cdot)$, $p_2(\cdot)$ with support $x = 0, 1, 2, \cdots$.

The simplest version occur when $r = 2$, for which

$$\|D\|_{r=2} = \left\{ \sum_{x=0}^{\infty} \left[\sqrt{p_1(x)} - \sqrt{p_2(x)} \right]^2 \right\}^{1/2}$$

or

$$\|D\|_{r=2} = \sqrt{2} \left\{ 1 - \sum_{x=0}^{\infty} \sqrt{p_1(x) p_2(x)} \right\}^{1/2}$$

for which the product term is described as the affinity. For $\|D\|$ small, the distance between the components is small, so that comparatively speaking the affinity is large. Here are some examples.

We have three examples of the Matusita's distance.

1. Mixtures of two Poisson distributions. θ_1, and θ_2 $(\theta_1 > \theta_2, \theta_2 > 0)$

$$\|D\| = \left\{ 2 \left(1 - e^{-\frac{1}{2}(\sqrt{\theta_1} - \sqrt{\theta_2})^2} \right) \right\}^{1/2}, \qquad (r \geq 1)$$

$\|D\| = 0$ if $\theta_1 = \theta_2$. The affinity is

$$\rho(\theta_1, \theta_2) = e^{-\frac{1}{2}(\sqrt{\theta_1} - \sqrt{\theta_2})^2}$$

2. Mixture of two binomial distributions. θ_1, θ_2, π_1 unknown, n known.

$$\begin{aligned} \|D\| &= \sqrt{2} \left\{ 1 - \left(\sqrt{\theta_1 \theta_2} + \sqrt{(1-\theta_1)(1-\theta_2)} \right)^n \right\}^{1/2} \\ &= \sqrt{2} \left\{ 1 - \left(\cos(\arccos \sqrt{\theta_1} - \arccos \sqrt{\theta_2}) \right)^n \right\}^{1/2} \qquad (0 < \theta_1, \theta_2 < 1) \end{aligned}$$

3. Hybrid mixture of Poisson-binomial distributions.

$$\|D\| = \sqrt{2} \left\{ 1 - \sum_{x=0}^{n} \left(\frac{e^{-\theta_1} \theta_1^x}{x!} \binom{n}{x} \theta_2^x (1-\theta_2)^{n-x} \right)^{1/2} \right\}^{1/2}$$

4. For two component negative binomial distribution with probability generating function $(1+\theta-\theta t)^{-k}$ for a single component.

$$\|D\| = \sqrt{2}\left\{1-\left(\cosh(\cosh^{-1}\sqrt{1+\theta_1}-\cosh^{-1}\sqrt{1+\theta_2})\right)^{-k}\right\}^{1/2} \qquad (\theta_1, \theta_2, k > 0)$$

Table 1. Matusita's distance

	θ_1	θ_2	n	Affinity	$\|D\|$
	1.0	1.5		0.9751	0.2233
	1.0	2.0		0.9178	0.4055
Poisson Mixture	1.0	3.0		0.7649	0.6856
	1.0	4.0		0.6065	0.8871
	1.0	5.0		0.4658	1.0336
	0.1	0.2	10	0.9039	0.4384
	0.1	0.3	10	0.7144	0.7558
Binomial Mixture	0.1	0.4	10	0.5098	0.9901
	0.1	0.5	10	0.3277	1.1596
	0.1	0.6	10	0.1855	1.2763
	1.0	0.1	10	0.9993	0.0383
	1.0	0.2	10	0.9075	0.4302
Poisson-binomial	1.0	0.3	10	0.7265	0.7395
Mixture	1.0	0.4	10	0.5294	0.9702
	1.0	0.5	10	0.3510	1.1393

Table 1. shows that the affinity decreases as the distance increases for each of the three mixtures. For a mixture of two binomials (common values n) the distance is zero if $\theta_1 = \theta_2$ since $\sqrt{\theta_1^2 + \sqrt{(1-\theta_1)^2}}$ equals unity. For the hybrid mixture (Poisson-binomial) there is no obvious solution to make $\|D\| = 0$ as long as n is fixed and an integer.

There are cases for which proving the distance zero might involve problems. Similarly for the variance approach (Δ_3^*), it may be difficult to prove the corresponding zeros are all complex, or real and complex.

As far as Matusita's distance is concerned, problems arise when three or more (preferably hybrid) components are involved. In this case, see Ahmad (1985).

It is also of interest to mention a study of Everitt and Hand (1981) where mixtures of Poisson, and mixtures of binomials are studied from the viewpoint of estimation procedures, using moment, and maximum likelihood methods. They gave simple iterative solutions in both cases, and amusingly enough take the case of four components, involving seven estimators. In their summary they do remark on the problem of closeness and its potential for creating the necessity for large samples.

It does seem that Matusita's distance is not a very sensitive measure of closeness of probability functions. However for maximum likelihood estimators and approximants to the covariance determinant using Parseval expansions and a related orthogonal system, the singularities of this covariance approximation may be as good a measure of closeness of probability functions, in other words a distance measure.

Acknowledgment

The first author's research was sponsored by the Applied Mathematical Sciences Research Program, Oak Ridge National Laboratory. It is operated by UT-Battelle, LLC for the U. S. Department of Energy, under contract number DE-AC05-00OR22725.

References

Abramowitz, M. and Stegun, I.A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards, Applied Mathematics Series 55.

Ahmad, I.A. (1985). Matusita's Distance, *Encyclopedia of Statistical Sciences*, Volume 5, Edited by S. Kotz, N.L. Johnson and C.B. Read, John Wiley & Sons, New york.

Bowman, K.O. and Shenton, L.R. (2003). Canonical dichotomized alternant determinants and Poisson mixtures, *Far East Journal of Theoretical Statistics*, 10(2), 87-109.

Everitt, B.S. and Hand, D.J. (1981). *Finite Mixture Distributions*, Chapman and Hall, London.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Volume 1, Second Edition, John Wiley & Sons, Inc. New York.

Johnson, N.L., Kotz, S. and Kemp, A.W. (1993). *Univariate Discrete Distributions*, Second Edition, John Wiley & Sons, Inc. New York.

Matusita, K. (1955). Decision Rules, Based on the Distance, for Problems of fit, two Samples, and Estimation, *AMS*, 26, 631-640.

Zwillinger, Daniel (1996). *Standard Mathematical Tables and Formulae*, 30th Edition, CRC Press, New York.