POISSON MIXTURES, ASYMPTOTIC VARIANCE, AND ALTERNANT DETERMINANTS

K. O. BOWMAN*
Computational Sciences and Engineering Division
Oak Ridge National Laboratory
P.O.Box 2008, 5700, MS-6085
Oak Ridge, Tennessee 37831-6085, U.S.A.
bowmanko@ornl.gov

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

June 10, 2004

Abstract

This paper explains a certain duality property occurring in the maximum likelihood covariance determinant. A special case of this determinant turns out to be a dichotomized alternant, the dichotomy arising from **derivatives** and **differences** involving the mixture components. Some errors in Bowman and Shenton (2003) are corrected.

Some keywords: Alternant determinants, alternant identities, covariance matrix, factorization, maximum likelihood estimators, orthogonal system.

1 Introduction

Continuing the study of Bowman and Shenton (2003) we consider the Poisson mixture probability function

$$P(x, \underline{\theta}, \underline{\pi}) = \sum_{i=1}^{s} \pi_i p(x, \theta_i) \quad (x = 0, 1, \cdots) \tag{1}$$

where for $i = 1, 2, \cdots, s$

$$p(x, \theta_i) = \frac{e^{-\theta_i}\theta_i^x}{x!}, \quad 0 < \pi_i < 1, \ \sum \pi_i = 1, \ \theta_i > 0,$$

$$\frac{\partial}{\partial \theta_i} p(x, \theta_i) = \frac{x - \theta_i}{\theta_i} p(x, \theta_i).$$

The maximum likelihood estimators are $\hat{\theta}_1$,$\hat{\theta}_2$, $\cdots$, $\hat{\theta}_s$ and $\hat{\pi}_1$, $\hat{\pi}_2$, $\cdots$, $\hat{\pi}_{s-1}$. Everett and Hand (1981) have provided iterative schemes to determine the estimators associated with a given appropriate data base.

Parseval type expansions are set up for log derivatives of the probability function $P(x, \underline{\theta}, \underline{\pi})$. For example, for 3 components (5 parameters)

$$\begin{aligned}
\left(\frac{x-\theta_1}{\theta_1}\right) p(x, \theta_1) &\sim \{A_0 q_0(x) + A_1 q_1(x) + \cdots\} P(x, \underline{\theta}, \underline{\pi}) \\
\left(\frac{x-\theta_2}{\theta_2}\right) p(x, \theta_2) &\sim \{B_0 q_0(x) + B_1 q_1(x) + \cdots\} P(x, \underline{\theta}, \underline{\pi}) \\
\left(\frac{x-\theta_3}{\theta_3}\right) p(x, \theta_3) &\sim \{C_0 q_0(x) + C_1 q_1(x) + \cdots\} P(x, \underline{\theta}, \underline{\pi}) \\
\{p(x, \theta_1) - p(x, \theta_3)\} &\sim \{A_0^* q_0(x) + A_1^* q_1(x) + \cdots\} P(x, \underline{\theta}, \underline{\pi}) \\
\{p(x, \theta_2) - p(x, \theta_3)\} &\sim \{B_0^* q_0(x) + B_1^* q_1(x) + \cdots\} P(x, \underline{\theta}, \underline{\pi}).
\end{aligned} \tag{2}$$

The upper segment of these equations refers to the θ's, the lower segment to the π's. We use the orthogonal system $\{q_r(x)\}$ associated with probability function $P(x, \underline{\theta}, \underline{\pi})$, so that

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

In terms of the central moment (μ_s) of the probability function (1)

$$q_r(x) = (-1)^r \begin{vmatrix} 1 & X & X^2 & \cdots & X^r \\ \mu_0 & \mu_1 & \mu_2 & \cdots & \mu_r \\ \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_{r+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_{r-1} & \mu_r & \mu_{r+1} & \cdots & \mu_{2r-1} \end{vmatrix} \div W_{r-1}$$

so that the coefficient of X^r is unity, and

$$W_r = \begin{vmatrix} \mu_0 & \mu_1 & \mu_2 & \cdots & \mu_r \\ \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_{r+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_r & \mu_{r+1} & \mu_{r+2} & \cdots & \mu_{2r} \end{vmatrix}.$$

As shown in Bowman and Shenton (2003), a simplified version of the covariance determinant arises when, for example, there are 5 parameters; in this case equations (2) are truncated out the sixth term in the series, and we study the form

$$\Delta_5 = \frac{\pi_1^2\pi_2^2\pi_3^2}{\phi_1\phi_2\phi_3\phi_4\phi_5} \begin{vmatrix} A_1\phi_1 & A_2\phi_2 & \cdots & A_5\phi_5 \\ B_1\phi_1 & B_2\phi_2 & \cdots & B_5\phi_5 \\ C_1\phi_1 & C_2\phi_2 & \cdots & C_5\phi_5 \\ A_1^*\phi_1 & A_2^*\phi_2 & \cdots & A_5^*\phi_5 \\ B_1^*\phi_1 & B_2^*\phi_2 & \cdots & B_5^*\phi_5 \end{vmatrix}^2 = \frac{(\pi_1\pi_2\pi_3)^2}{\phi_1\phi_2\phi_3\phi_4\phi_5}\Delta_5^{*2}.$$

Here $\phi_s = W_s/W_{s-1}$.

We have demonstrated the following results for the approximate asymptotic variances to order n^{-1}, n being the sample size. The 5 parameter case is

$$Var_1(\hat{\theta}_1) \sim \frac{W_5/W_4}{\pi_1^2\{(\theta_1-\theta_2)(\theta_1-\theta_3)\}^4},$$
$$Var_1(\hat{\theta}_2) \sim \frac{W_5/W_4}{\pi_2^2\{(\theta_2-\theta_1)(\theta_2-\theta_3)\}^4},$$
$$Var_1(\hat{\theta}_3) \sim \frac{W_5/W_4}{\pi_3^2\{(\theta_3-\theta_1)(\theta_3-\theta_2)\}^4},$$

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

Here we shall explain the origin of the unexpected terms in $Var_1(\hat{\pi_1})$ and $Var_1(\hat{\pi_2})$ and the interesting forms for these asymptotic variances where there are up to 5 components (9 parameters); there is also a general conjecture incorrectly stated in the 2003 paper.

The asymptotic variance may only be of moderate interest, but the associated dichotomized alternants may be new displaying a remarkable duality between expressions such as (2θ), (3θ),$\cdots$, and $\theta_1 + \theta_2$, $\theta_1^2 + \theta_1\theta_2 + \theta_2^2$, $\theta_1^3 + \theta_1^2\theta_2 + \theta_1\theta_2^2 + \theta_2^3$. Thus the duality in general is expressed by

$$(r\theta_1^{r-1}) \longleftrightarrow \theta_1^{r-1} + \theta_1^{r-2}\theta_2 + \cdots + \theta_2^{r-1} = \frac{\theta_1^r - \theta_2^r}{\theta_1 - \theta_2}$$

or

$$(-\theta_1^r) \longleftrightarrow \frac{\theta_1^r - \theta_j^r}{\theta_1 - \theta_j} \quad (j = 2, 3, \cdots). \tag{3}$$

2 Stirling number and logarithmic derivatives

In Bowman and Shenton (2003) we showed that

$$\begin{aligned}\nu_s &= \sum_{x=0}^{\infty} \left(\frac{x - \theta_i}{\theta_i}\right) \frac{e^{-\theta_i}\theta_i^x}{x!}(x - \mu_1')^s \\ &= \sum_{r=0}^{s}(-1)^r \binom{s}{r} (\mu_1')^r \frac{\partial}{\partial\theta_i}\left\{\mathcal{S}_{s-r}^{(1)} + \mathcal{S}_{s-r}^{(2)}\theta_i^2 + \cdots + \mathcal{S}_{s-r}^{(s-r)}\theta_i^{s-r}\right\}\end{aligned}$$

in terms of Stirling number of the second kind, and μ_1' the mean of the mixture. We can arrange the expression as

$$\nu_s = H_1^{(s)} + H_2^{(s)}(\underline{2\theta_i}) + H_3^{(s)}(\underline{3\theta_i^2}) + \cdots + H_s^{(s)}(\underline{s\theta_i^{s-1}})$$

where the H's do not explicitly involve the θ's. Although this expression was cited in the earlier paper, it was not used in the subsequent development. It is much simpler in form and makes determinant reductionism less obscure. Note that in view of (3), we have

$$\nu_s^* = \sum_{x=0}^{\infty} \{p(x,\theta_1) - p(x,\theta_3)\} (x - \mu_1')^s$$
$$= H_1^{(s)} + H_2^{(s)} \left(\frac{\theta_3^2 - \theta_1^2}{\theta_3 - \theta_1}\right) + H_3^{(s)} \left(\frac{\theta_3^3 - \theta_1^3}{\theta_3 - \theta_1}\right) + H_4^{(s)} \left(\frac{\theta_3^4 - \theta_1^4}{\theta_3 - \theta_1}\right) + \cdots$$

where $H_r^{(s)} = 1$, if $r = s$, $H_r^{(s)} = 0$, if $r > s$.

Note that in the upper segment of the covariance determinant derivatives of the probability function occur, whereas in the lower segment differences of probability functions are found.

3 Variance of $\hat{\pi}$ and 4 components (7 parameters)

The basic determinant is

$$\left|\begin{array}{ccccc|c}
1 & 2\theta_1 & 3\theta_1^2 & \cdots & 7\theta_1^6 & (\theta_1) \\
1 & 2\theta_2 & 3\theta_2^2 & \cdots & 7\theta_2^6 & (\theta_2) \\
1 & 2\theta_3 & 3\theta_3^2 & \cdots & 7\theta_3^6 & (\theta_3) \\
1 & 2\theta_4 & 3\theta_4^2 & \cdots & 7\theta_4^6 & (\theta_4) \\
1 & S_1(1,4) & S_2(1,4) & \cdots & S_6(1,4) & \theta_1 - \theta_4, \pi_1 \\
1 & S_1(2,4) & S_2(2,4) & \cdots & S_6(2,4) & \theta_2 - \theta_4, \pi_2 \\
1 & S_1(3,4) & S_2(3,4) & \cdots & S_6(3,4) & \theta_3 - \theta_4, \pi_3
\end{array}\right. .$$

where $S_r(i,j) = (\theta_i^{r+1} - \theta_j^{r+1})/(\theta_i - \theta_j)$, remembering that the factors $(\theta_1 - \theta_4)$, $(\theta_2 - \theta_4)$, and $(\theta_3 - \theta_4)$ are omitted at this stage. The symbols in the 8th column indicate the derivative or difference associated with the corresponding row. Using Maple, we have

$$Var_1(\hat{\theta}_i) \sim \frac{W_7/W_6}{\pi_i^2\{\prod_{r=1}^{4*}(\theta_i - \theta_r)^4\}} \qquad (i = 1, 2, 3, 4)$$

which is readily generalized and proved for the case of 2 components up to 5 components. Moreover

$$Var_1(\hat{\pi}_3) \sim \frac{4(W_7/W_6)F_2^2(\theta_3; \theta_1, \theta_2, \theta_4)}{\prod_{i=1}^{4*}(\theta_3 - \theta_i)^6}$$

where $\prod_i^*$ means a zero term is omitted, and

$$\begin{aligned} F_2(\theta_3) &= 3\theta_3^2 - 2\theta_3(\theta_1 + \theta_2 + \theta_4) + \theta_1\theta_2 + \theta_1\theta_4 + \theta_2\theta_4. \\ &= \frac{\partial}{\partial\theta_3}\{(\theta_3 - \theta_1)(\theta_3 - \theta_2)(\theta_3 - \theta_4)\} \end{aligned}$$

For 3 components (5 parameters), $Var_1(\hat{\pi}_1)$ involves the factor

$$F_1(\theta_1) = 2\theta_1 - \theta_2 - \theta_3 = \frac{\partial}{\partial\theta_1}(\theta_1 - \theta_2)(\theta_1 - \theta_3)$$

and for $Var_1(\hat{\pi}_2)$

$$F_1(\theta_2) = 2\theta_2 - \theta_1 - \theta_3 = \frac{\partial}{\partial\theta_2}(\theta_2 - \theta_1)(\theta_2 - \theta_3)$$

For s components $(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_s; \hat{\pi}_1, \hat{\pi}_2, \cdots, \hat{\pi}_{s-1})$, the asymptotic variance $Var_1(\hat{\pi}_1)$ has the factor

$$F_{s-2}(\theta_1) = \frac{\partial}{\partial\theta_1}\left\{\prod_{r=2}^{s}(\theta_1 - \theta_r)\right\}. \quad (s \geq 3)$$

Similarly $Var_1(\hat{\pi}_2)$ has the factor

$$F_{s-2}(\theta_2) = \frac{\partial}{\partial\theta_2}\{(\theta_2 - \theta_1)(\theta_2 - \theta_3)\cdots(\theta_2 - \theta_s)\},$$

with similar expressions for the remaining proportions. These formulas has been verified up to and including 5 components.

4 Setting up identities

Related to $Var_1(\hat{\pi}_3)$ and 7 parameters mixture we have proved

$$D_6 = \begin{vmatrix} 1 & 2\theta_1 & 3\theta_1^2 & 4\theta_1^3 & 5\theta_1^4 & 6\theta_1^5 \\ 1 & 2\theta_2 & 3\theta_2^2 & 4\theta_2^3 & 5\theta_2^4 & 6\theta_2^5 \\ 1 & 2\theta_3 & 3\theta_3^2 & 4\theta_3^3 & 5\theta_3^4 & 6\theta_3^5 \\ 1 & 2\theta_4 & 3\theta_4^2 & 4\theta_4^3 & 5\theta_4^4 & 6\theta_4^5 \\ 1 & S_1(1,4) & S_2(1,4) & S_3(1,4) & S_4(1,4) & S_5(1,4) \\ 1 & S_1(2,4) & S_2(2,4) & S_3(2,4) & S_4(2,4) & S_5(2,4) \end{vmatrix}_{(6\times6)} \quad (4)$$

$$= 2F_2(\theta_3)\left\{(\theta_3-\theta_1)(\theta_3-\theta_2)(\theta_3-\theta_4)(\theta_1-\theta_2)^4(\theta_2-\theta_4)^3(\theta_1-\theta_4)^3\right\}$$

where

$$F_2(\theta_3) = \frac{\partial}{\partial\theta_3}\{(\theta_3-\theta_1)(\theta_3-\theta_2)(\theta_3-\theta_4)\}$$

Now it is clear that (4) could be expanded by elements in the third row since this row is distinct because it is function of θ_3 only. Moreover the polynomial will be of degree 5. But as far as differences are concerned θ_3 occurs only in $(\theta_3-\theta_1)(\theta_3-\theta_2)(\theta_3-\theta_4)$ so that this leads to expect the appearance of a terms in θ_3 of degree two; actually $F_2(\theta_3)$. Next the term $(\theta_3-\theta_1)(\theta_3-\theta_2)(\theta_3-\theta_4) = \theta_3$ springs from the variance asymptotic term. θ_3^6 (Product of difference)^8, is square root relating to θ_3 itself. Lastly the excluded differences are $(\theta_1-\theta_4)(\theta_2-\theta_4)(\theta_1-\theta_2)$; the first two of these are related to the 5th and 6th rows of equation (4) for which $(\theta_1-\theta_4)$ and $(\theta_2-\theta_4)$ are excluded factors.

Now consider a general case; see expression (10) of Bowman and Shenton (2003) which has an error. We have

$$D_{2s-2} = \begin{vmatrix} 1 & 2\theta_1 & 3\theta_1^2 & \cdots & (2s-2)\theta_1^{2s-3} \\ 1 & 2\theta_2 & 3\theta_2^2 & \cdots & (2s-2)\theta_2^{2s-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 2\theta_s & 3\theta_s^2 & \cdots & (2s-2)\theta_s^{2s-3} \\ 1 & S_1(1,s) & S_2(1,s) & \cdots & S_{2s-3}(1,s) \\ 1 & S_1(2,s) & S_2(2,s) & \cdots & S_{2s-3}(2,s) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & S_1(s-2,s) & S_2(s-2,s) & \cdots & S_{2s-3}(s-2,s) \end{vmatrix}_{(2s-2)\times(2s-2)}$$

of order $(2s-2)$ by $(2s-2)$, and relating to a mixture of s components ($2s-1$ parameters), and numerator determinant D_{2s-2} of $Var_1(\hat{\pi}_{s-1})$.

In factorial form

$$D_{2s-2} = \frac{2A_s^4(\underline{\theta})F_{s-2}(\theta_{s-1})}{B_s^3(\underline{\theta})C_s(\underline{\theta})}$$

where

$$A_s(\underline{\theta}) = \prod_{r=1}^{s-1} \prod_{t=r+1}^{s} (\theta_r - \theta_t),$$
$$B_s(\underline{\theta}) = \prod_{r=1}^{s*} (\theta_{s-1} - \theta_r),$$
$$C_s(\underline{\theta}) = \prod_{r=1}^{s-2} (\theta_r - \theta_s),$$

the result having been checked for $s = 3$, 4, and 5. (As an assist, the reader may keep in mind that it is general case relates to $Var_1(\hat{\pi}_{s-1})$).

5 Further remarks

The approximate variances for the estimators $\hat{\theta}_1$, $\hat{\theta}_2$, $\cdots$, $\hat{\pi}_1$, $\hat{\pi}_2$, $\cdots$, at least suggest a mathematical interpretation of the descriptive "closeness" of components, and also suggest that the closeness factor affects the proportions $(\hat{\pi}_1, \hat{\pi}_2, \cdots)$ more than the components $(\hat{\theta}_1, \hat{\theta}_2, \cdots)$; singularities for the former being order six, for the latter of order four. So much for the statistical context.

However the study introduces a new type of alternant determinant consisting of two parts which depends on a duality phenomenon relating $2x$ to $x + y$, $3x^2$ to $x^2 + xy + y^2$, $4x^3$ to $x^3 + x^2y + xy^2 + y^3$ and so on. There are identities for various alternant forms, generally expressible as products of powers of differences such as $\theta_1 - \theta_2$, etc. However, there is one surprising exception relating to the first order variance of a $\hat{\pi}$. Here the product form contains a factor such is

$$\frac{\partial}{\partial \theta_1} \{(\theta_1 - \theta_2)(\theta_1 - \theta_3) \cdots (\theta_1 - \theta_s)\}$$

which relates to the gradient of a polynomial form with real coefficients, so that several stationary points will be involved resulting in an irregularity in the variance of a proportion.

There is an extensive literature on alternants. According to Muir (1882), *Alternating Functions* were introduced by Cauchy (1812). About this time it was known that

$b^2 - ac$ was the *determinant* of $ax^2 + 2bxy + cy^2$, and similar functions of higher degree. Muir's *Historical and Bibliographical Summary* of the History of determinants is well worth a read. So, our claims to newness with respect to dichotomized determinant is somewhat muted.

There is interest in testing out the Factor functions (Maple, Mathematika) to implement a simplification of high order alternants.

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

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.

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

Muir, Thomas (1882). *A Treatise on the Theory of Determinants*, MacMillan & Co., London.

Muir, Sir Thomas (1930). *Contributions to the History of Determinants*, Blackie & Son Ltd., London and Glasgow.