Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family

LOREN COBB, PETER KOPPSTEIN, and NENG HSIN CHEN*

Multimodal generalizations of the normal, gamma, inverse gamma, and beta distributions are introduced within a unified framework. These multimodal distributions, belonging to the exponential family, require fewer parameters than corresponding mixture densities and have unique maximum likelihood estimators. Simple moment recursion relations, which make maximum likelihood estimation feasible, also yield easily computed estimators that themselves are shown to be consistent and asymptotically normal. Lastly, a statistic for bimodality, based on Cardan's discriminant for a cubic shape polynomial, is introduced.

KEY WORDS: Bimodality; Catastrophe theory; Parameter estimation; Pearson system; Polynomial exponential distributions; Shape polynomial.

1. INTRODUCTION

The model generally used in the analysis of multimodal densities is a mixture of normals, or possibly of other unimodal densities. There is a class of alternatives, however, that may be appropriate when a mixture assumption is not required or justified. Four major types of nonmixture multimodal probability densities within this class are presented here, each of which can arise as the stationary probability density function of a nonlinear diffusion process. Many common unimodal families (e.g., normal, gamma, beta) are represented as special cases of these types. This class of probability densities is expressed in the general form

$$f_k(x) = \xi(\boldsymbol{\beta})\exp\left[-\int^x \{g(s)/v(s)\}ds\right], \qquad (1.1)$$

where $g(x) = \beta_0 + \beta_1 x + \cdots + \beta_k x^k$, $k > 0$, and $v(x)$ has one of the following principal forms (other forms are, of course, possible):

Type **N**: $v(x) = 1 \qquad -\infty < x < \infty.$

Type **G**: $v(x) = x \qquad 0 < x < \infty.$

Type **I**: $v(x) = x^2 \qquad 0 < x < \infty.$

Type **B**: $v(x) = x(1 - x) \qquad 0 < x < 1.$

The integral in (1.1) is to be understood as an indefinite integral, and the domain of $f_k(x)$ is the open interval on which $v(x)$ is positive. The normalization function $\xi:\mathbf{R}^{k+1} \to \mathbf{R}$ is chosen so that the integral of f over its domain is unity. In this article, the terms mode and antimode will be reserved, respectively, for local maxima and minima of the density function at which the density's derivative vanishes. Modes are thus distinguished from poles and nonmodal local maxima on the boundaries of the domain of the density function.

The densities described by (1.1) are a generalization of the Pearson system. On differentiation with respect to x, (1.1) yields

$$f'(x)/f(x) = -g(x)/v(x), \qquad (1.2)$$

which contains Pearson's differential equation as a special case. In the Pearson system (Ord 1972, Johnson and Kotz 1970), the degree k of the polynomial g is one and the degree of v is at most two. In this article we are principally concerned with the multimodal forms that appear when the degree of g exceeds one. The polynomial g will be called the *shape polynomial* for the density f.

The capacity for multimodality in the class described by (1.1) is illustrated in Figure 1, which shows a sequence of densities of Type N, with $g(x) = 10x^3 - \beta x - .1$, for various values of β.

The maximum number of modes possible in a given family is determined by the degree of its shape polynomial, k. From (1.2) it may be seen that the critical points of the density (i.e., those points x such that $f'(x) = 0$) are exactly the roots of $g(x)$. Whether such a point is a mode or an antimode (a relative minimum) depends on the sign of $g''(x) - \{g'(x)\}^2$. At the roots of $v(x)$ the density either has a zero ($f(x) \to 0$) or a pole ($f(x) \to \infty$), depending on the coefficients of $g(x)$. The only exceptions to this occur at points that are roots of both $g(x)$ and $v(x)$: these are degenerate boundary points for the density (Cobb 1981b).

The generalized family of Pearson distributions may also be characterized in terms of nonlinear diffusion processes (see, e.g., Wong 1964). Let $2\mu(x) = g(x) - v'(x)$,

* Loren Cobb is Associate Professor, Department of Biometry, Medical University of South Carolina, Charleston, SC 29425. Peter Koppstein is Instructor, Department of Political Science, Washington University, St. Louis, MO 63130. Neng Hsin Chen is a Postdoctoral Fellow in the Department of Biometry, Medical University of South Carolina. The research of the first author was supported by the National Institute of Mental Health (MH-14594-04) and the National Science Foundation (ISP80-11451). The third author was supported by the National Heart, Blood, and Lung Institute (T32-HL07380). The authors wish to thank Brad Crain, Hurshell Hunt, and I.R. Savage for their suggestions.

© Journal of the American Statistical Association
March 1983, Volume 78, Number 381
Theory and Methods Section

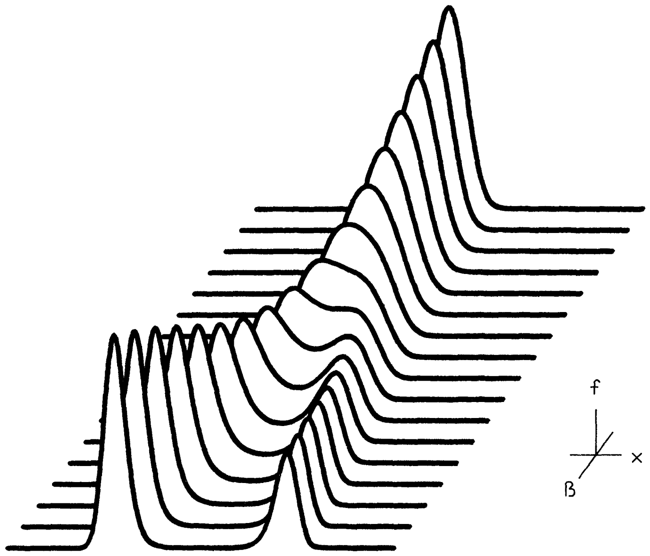

Figure 1. A sequence of Type **N** *densities with cubic shape polynomial $g(x) = 10x^3 - \beta x - .3$, for various values of β.*

and $\sigma^2(x) = v(x)$. Then $f(x)$ is the stationary density of a stochastic process x_t that is governed by the stochastic differential equation (Soong 1973)

$$dx_t = -\mu(x_t)dt + \sigma(x_t)dw_t, \tag{1.3}$$

where w_t is a standard Wiener process. Consider the deterministic version of this system, namely $dx/dt = -\mu(x)$. It has equilibria at the solutions of $g(x) - v'(x) = 0$. In the Type **N** cases ($v(x) = 1$) these equilibria are exactly the modes and antimodes of the corresponding probability density function. In the other types the modes and antimodes are shifted away from the equilibria of the deterministic system (Cobb and Watson 1980). In these four cases, modes correspond to stable equilibria, while antimodes correspond to unstable equilibria. Thus multimodality in these models is generally a result of multiple stable states in a dynamical system, rather than of heterogeneous populations as in the usual interpretation of mixture densities. Note, however, that bimodal stationary densities, for example, can arise when there is but one corresponding stable equilibrium, as discussed at the end of the following section.

The estimation problem for these multimodal densities can be stated this way: given the type and degree of the density, estimate the coefficient vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)$. If it is assumed that the underlying model is the nonlinear stochastic system (1.3), as, for example, in elementary catastrophe theory (Poston and Stewart 1978), then these estimates lead indirectly to an identification of the deterministic component of the system.

2. THE PRINCIPAL TYPES

Each distinct specification of the function $v(x)$ in (1.1) leads to a distinct family of distributions, each family being indexed by the degree of the polynomial $g(x)$. To simplify the notation, let $N_k(x)$ refer to the density of the Type **N** family of degree k for permissible k and similarly for G_k, I_k, and B_k.

The $\mathbf{N}_k$ densities have as their principal member the normal density, $\mathbf{N}_1$. The bimodal density $\mathbf{N}_3$ (Figure 1) was first discussed by Fisher (1922) but has received only occasional attention since that time (e.g., O'Toole 1933, Aroian 1948, Matz 1978). The relevance of $\mathbf{N}_3$ and indeed $\mathbf{G}_3$ and $\mathbf{I}_3$ to statistical analyses of the cusp catastrophe model (Cobb 1978, 1981a,b, Cobb and Watson 1980, Koppstein 1980) has recently spurred interest in the whole generalized Pearson family. The general form for an $\mathbf{N}_k$ density is

$$N_k(x) = \xi \exp[\theta_1 x + \theta_2 x^2 + \cdots + \theta_{k+1}x^{k+1}], \tag{2.1}$$

where $\theta_j = -\beta_{j-1}/j$. $\mathbf{N}_k$ has finite moments of all orders if k is odd and $\theta_{k+1} < 0$.

The $\mathbf{G}_k$ densities have as their principal member the gamma density, $\mathbf{G}_1$, and include the exponential and Rayleigh densities. The general form for the $\mathbf{G}_k$ density is

$$G_k(x) = \xi x^{\alpha-1}\exp[\theta_1 x + \cdots + \theta_k x^k], \tag{2.2}$$

where $\alpha = 1 - \beta_0$ and $\theta_j = -\beta_j/j$. $\mathbf{G}_k$ has moments of all orders if $\alpha > 0$ and $\theta_k < 0$.

The $\mathbf{I}_k$ densities have as their principal member the inverse gamma density, $\mathbf{I}_1$ (Pearson Type V). The $\mathbf{I}_2$ density is a generalized inverse Gaussian density. The general form for the $\mathbf{I}_k$ density is

$$I_k(x) = \xi x^{\alpha}\exp[\gamma x^{-1} + \theta_1 x + \cdots + \theta_{k-1}x^{k-1}], \tag{2.3}$$

where $\alpha = -\beta_1$, $\gamma = \beta_0$, and $\theta_j = -\beta_{j+1}/j$. The $\mathbf{I}_k$ densities have finite moments all orders if $\gamma < 0$ and $\theta_{k-1} < 0$.

The $\mathbf{B}_k$ densities have as their principal member the beta density, $\mathbf{B}_1$. The $\mathbf{B}_3$ density has been used in population genetics (e.g., Ludwig 1974) to describe the frequency of a gene with heterozygotic advantage, such as the gene for sickle-cell anemia. The $\mathbf{B}_3$ density is particularly interesting because it can adopt the W shape shown in Figure 2, which exhibits a central mode surrounded by two antimodes and two poles. Qualitatively, the $\mathbf{B}_3$ density has much in common with the S_B system (Johnson and Kotz 1970, p. 25). The general form for the $\mathbf{B}_k$ density is

$$B_k(x) = \xi x^{\alpha-1}(1 - x)^{\gamma-1}\exp[\theta_1 x + \cdots + \theta_{k-1}x_{k-1}] \tag{2.4}$$

where $\alpha = 1 - \beta_0$, $\gamma = 1 + \sum_{i=0}^{k} \beta_i$, and $\theta_j = \sum_{i=j+1}^{k} \beta_i/j$, for $j = 1, \ldots, k - 1$. $\mathbf{B}_k$ has finite moments of all orders if $\alpha > 0$ and $\gamma > 0$.

The four classes of distributions identified above may together be referred to as the multimodal Pearson system: the restriction that $v(x)$ be a polynomial of degree at most two is preserved, but the degree of the polynomial $g(x)$ is arbitrary. Further generalizations are, of course, possible. We mention in particular the closely related class

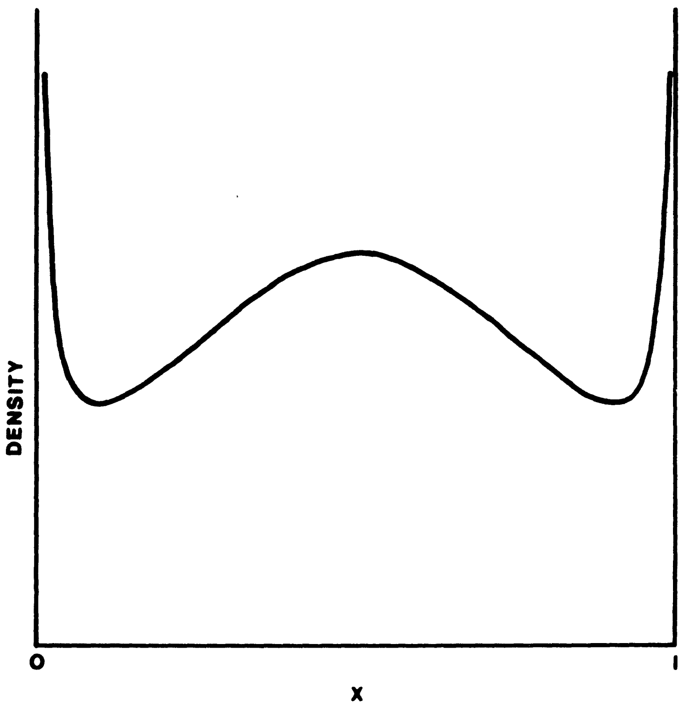

Figure 2. A Type **B** *density with cubic shape polynomial chosen so that the density has two poles, two antimodes, and one mode. The shape polynomial is* $g(x) = -14(x - .1)(x - .5)(x - .9)$.

of distributions on (0, 1) defined as above but with $v(x) = x^2(1 - x)^2$. This class of distributions stands in relation to Type **B** as Type **I** stands to Type **G**, and thus should perhaps be included in our enumeration of principal types; certainly the discussion that follows applies equally to this class as well. We shall, however, simply remark here that this class arises in the study of logistic growth and is noteworthy because the stationary densities of the related stochastic differential equations may exhibit bimodality even when the deterministic dynamic has only one stable equilibrium (Lefever 1981).

Finally, we observe that not all distributions defined by (1.1) have finite moments of all orders. For example, $g(x) = (1 + r)x$ and $v(x) = r + x^2$ yields Student's t density with r degrees of freedom.

3. ESTIMATION

3.1 Maximum Likelihood Estimation (MLE)

Since the densities $\mathbf{N}_k$, $\mathbf{G}_k$, $\mathbf{I}_k$, and $\mathbf{B}_k$ belong to the well-known exponential family, we shall be brief. If $(X_1, \ldots, X_n)$ is a random sample of a random variable with one of these densities, then the minimal sufficient statistic for $\boldsymbol{\beta}$ is

Type **N**: $(\sum X, \sum X^2, \ldots, \sum X^{k+1})$.
Type **G**: $(\sum \ln(X), \sum X, \ldots, \sum X^k)$.
Type **I**: $(\sum X^{-1}, \sum \ln(X), \sum X, \ldots, \sum X^{k-1})$.
Type **B**: $(\sum \ln(X), \sum \ln(1 - X), \sum X, \ldots, \sum X^{k-1})$.

It is not difficult to show that the Hessian of the negative log-likelihood function is a positive definite matrix. Thus the unique MLE's can in principle be readily computed. The numerical integrations involved, however, can be quite formidable. Nevertheless, as O'Toole's paper (1933b) suggests, the simplest quadrature methods may be expected to yield good results. Further, as we show in Section 3.2, simple moment recursion formulas enable a trivial calculation of consistent estimators. These recursion formulas also enable straightforward calculation of the Hessian once the numerical integrations required for calculation of the gradient vector of the log-likelihood function have been performed. For $\mathbf{N}_3$, for example, only three integrations are required to calculate the gradient (and Hessian).

3.2 Consistent Estimators From Moment Recursion Relations

Pearson's method of parameter estimation depends on the existence of a linear system of equations relating the $k + 1$ parameters to the first $k + 1$ moments of the density. If such a system can be found, then sample moment estimates are inserted and the system is solved for the parameters. The direct application of this method to the multimodal exponential families discussed here fails because of the lack of a general formula relating the first $k + 1$ moments to the parameters. However, a formula relating $2k$ moments to the parameters can be found, based on the following theorem.

Theorem 1. Let X be a random variable with probability density function $f(x)$ of Type **N**, **G**, **I**, or **B**, with $k > 0$. For any $j \geq 0$,

$$E[X^j g(X)] = E[\{X^j v(X)\}'],$$

where (′) denotes differentiation.

Proof. Use (1.2) and integration by parts. Let the domain of f be denoted by (a, b). Then

$$\begin{aligned} E[X^j g(X)] &= \int_a^b x^j g(x) f(x) dx \\ &= \int_a^b x^j \{-v(x) f'(x)/f(x)\} f(x) dx \\ &= -\int_a^b x^j v(x) f'(x) dx. \end{aligned}$$

Now integrate this expression by parts:

$$\begin{aligned} &-\int_a^b x^j v(x) f'(x) dx \\ &\qquad = -x^j v(x) f(x) \Big|_a^b + \int_a^b \{x^j v(x)\}' f(x) dx. \end{aligned}$$

Note that $x^j v(x) f(x) \to 0$ as $x \to a$ and as $x \to b$ for each of the principal densities (2.1–2.4).

Remark. This theorem applies to any density in the class (1.1) for which the first term in the integration by parts vanishes, even if not all moments are finite. In the case of Student's t, for example, it implies that $(r - j - 1)\mu_{j+1} = rj\mu_{j-1}$, where r denotes the degrees of freedom.

The moment recursion relations and the estimators derivable from them are direct consequences of Theorem 1:

Corollary 1. For each of the principal types of densities in (1.1) there is a recursion relation for the noncentral moments μ_i:

$$\text{Type } \mathbf{N}\text{:} \quad \sum_{i=0}^{k} \beta_i \mu_{i+m} = m\mu_{m-1}. \tag{3.1}$$

$$\text{Type } \mathbf{G}\text{:} \quad \sum_{i=0}^{k} \beta_i \mu_{i+m} = (m + 1)\mu_m. \tag{3.2}$$

$$\text{Type } \mathbf{I}\text{:} \quad \sum_{i=0}^{k} \beta_i \mu_{i+m} = (m + 2)\mu_{m+1}. \tag{3.3}$$

$$\text{Type } \mathbf{B}\text{:} \quad \sum_{i=0}^{k} \beta_i \mu_{i+m} = (m + 1)\mu_m - (m + 2)\mu_{m+1}. \tag{3.4}$$

These moment relations have long been well known in the special case $k = 1$.

In 1948, Aroian used the recursion formula for $\mathbf{N}_3$ to obtain parameter estimates for the quartic exponential distribution. The following corollary generalizes his procedure.

Corollary 2. Let $\mathbf{M}$ be the $(k + 1) \times (k + 1)$ matrix of moments for the random variable X: $[\mathbf{M}]_{ij} = \mu_{i+j-2}$. Then $\mathbf{M}\boldsymbol{\beta} = \boldsymbol{\alpha}$, where $\alpha_j = E[\{X^{j-1}v(X)\}']$.

This corollary provides a relationship between moments and parameters that is useful for estimation. Simply use $\hat{\boldsymbol{\beta}} = \hat{\mathbf{M}}^{-1}\hat{\boldsymbol{\alpha}}$, where the entries of $\hat{\mathbf{M}}$ and $\hat{\boldsymbol{\alpha}}$ are the ordinary sample moments. The entries of $\hat{\boldsymbol{\alpha}}$ depend on the type of density: in the case of Type N, for example, $\alpha_j = (j - 1)\,\mu_{j-2}$. The following lemma is needed:

Lemma 1. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables. Let $[\hat{\mathbf{M}}]_{ij} = \sum_{k=1}^{n} X_k^{i+j-2}/n$. Then $\hat{\mathbf{M}}$ is positive definite with probability one.

Proof. Let $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_k)$ be an arbitrary nonzero vector. Note that $n\boldsymbol{\gamma}'\hat{\mathbf{M}}\boldsymbol{\gamma} = \sum_{i=1}^{n} (\gamma_0 + \gamma_1 X_i + \cdots + \gamma_k X_i^k)^2$. But, since X has a continuous density, we have $\text{Prob}\{\gamma_0 + \cdots + \gamma_k X^k = 0\} = 0$ for $i = 1, \ldots, n$. The result follows immediately.

The bias and relative efficiency of the estimator $\hat{\boldsymbol{\beta}} = \hat{\mathbf{M}}^{-1}\hat{\boldsymbol{\alpha}}$ are not as yet known, but it can be shown that $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normal.

Theorem 2. The estimator $\hat{\boldsymbol{\beta}} = \hat{\mathbf{M}}^{-1}\hat{\boldsymbol{\alpha}}$ is consistent, and $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically multivariate normal with covariance matrix $\mathbf{V}$, such that

$$[\mathbf{MVM}]_{ij} = E\{(\hat{\alpha}_i - [\hat{\mathbf{M}}\boldsymbol{\beta}]_i)(\hat{\alpha}_j - [\hat{\mathbf{M}}\boldsymbol{\beta}]_j)\}.$$

Proof. Consistency: it has already been established that $\hat{\mathbf{M}}$ is invertible (w.p.1). The function that takes an invertible matrix into its inverse is differentiable with respect to each of its entries, and $\hat{\mathbf{M}} \xrightarrow{P} \mathbf{M}$, so $\hat{\mathbf{M}}^{-1} \xrightarrow{P} \mathbf{M}^{-1}$. Furthermore, $\hat{\boldsymbol{\alpha}} \xrightarrow{P} \boldsymbol{\alpha}$, therefore $\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}$.

Normality: we have $\sqrt{n}(\hat{\mathbf{M}} - \mathbf{M}) = \mathbf{O}_p(1)$ and $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{o}_p(1)$. Consider the identity

$$\sqrt{n}\mathbf{M}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sqrt{n}(\hat{\boldsymbol{\alpha}} - \hat{\mathbf{M}}\boldsymbol{\beta}) - \sqrt{n}(\hat{\mathbf{M}} - \mathbf{M})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Each entry of the second term on the right side is $\mathbf{O}_p'(1)\mathbf{o}_p(1) = o_p(1)$, where here (′) denotes matrix transposition. Thus $\sqrt{n}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \mathbf{M}^{-1}(\hat{\boldsymbol{\alpha}} - \hat{\mathbf{M}}\boldsymbol{\beta})] \xrightarrow{P} 0$. The vector $\sqrt{n}\mathbf{M}^{-1}(\hat{\boldsymbol{\alpha}} - \hat{\mathbf{M}}\boldsymbol{\beta})$ can be written as $\sum_{i=1}^{n} \mathbf{h}(X_i)/\sqrt{n}$, where $\mathbf{h}(x)$ is a vector of polynomials in x. Note that $E[\mathbf{h}(X)] = \mathbf{0}$. Let $[\mathbf{V}]_{ij} = E[h_i(X)h_j(X)]$. Then $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically $N(\mathbf{0}, \mathbf{V})$, by the multivariate Central Limit Theorem.

The $(k + 1) \times (k + 1)$ asymptotic covariance matrix $\mathbf{V}$ of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ can be written in the form $\mathbf{V} = \mathbf{M}^{-1}\mathbf{BGB}'\mathbf{M}^{-1}$, where $\mathbf{G}$ is the $(2k) \times (2k)$ covariance matrix with $[\mathbf{G}]_{ij} = \text{cov}(X^i, X^j)$ for $i, j = 1, \ldots, 2k$, and $\mathbf{B}$ is a $(k + 1) \times (2k)$ matrix that depends on the type and order of the density. The pattern of the matrix $\mathbf{B}$ for each of the principal types, $\mathbf{N}_k$, $\mathbf{G}_k$, $\mathbf{I}_k$, and $\mathbf{B}_k$, is established by the form of $\mathbf{B}$ for $k = 3$, as follows:

$$\mathbf{N}_3\text{: } \mathbf{B} = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & 0 & 0 & 0 \\ \beta_0 & \beta_1 & \beta_2 & \beta_3 & 0 & 0 \\ -2 & \beta_0 & \beta_1 & \beta_2 & \beta_3 & 0 \\ 0 & -3 & \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{bmatrix}$$

$$\mathbf{G}_3\text{: } \mathbf{B} = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & 0 & 0 & 0 \\ \beta_0 - 2 & \beta_1 & \beta_2 & \beta_3 & 0 & 0 \\ 0 & \beta_0 - 3 & \beta_1 & \beta_2 & \beta_3 & 0 \\ 0 & 0 & \beta_0 - 4 & \beta_1 & \beta_2 & \beta_3 \end{bmatrix}$$

$$\mathbf{I}_3\text{: } \mathbf{B} = \begin{bmatrix} \beta_1 - 2 & \beta_2 & \beta_3 & 0 & 0 & 0 \\ \beta_0 & \beta_1 - 3 & \beta_2 & \beta_3 & 0 & 0 \\ 0 & \beta_0 & \beta_1 - 4 & \beta_2 & \beta_3 & 0 \\ 0 & 0 & \beta_0 & \beta_1 - 5 & \beta_2 & \beta_3 \end{bmatrix}$$

$$\mathbf{B}_3\text{: } \mathbf{B} = \begin{bmatrix} \beta_1 + 2 & \beta_2 & \beta_3 & 0 & 0 & 0 \\ \beta_0 - 2 & \beta_1 + 3 & \beta_2 & \beta_3 & 0 & 0 \\ 0 & \beta_0 - 3 & \beta_1 + 4 & \beta_2 & \beta_3 & 0 \\ 0 & 0 & \beta_0 - 4 & \beta_1 + 5 & \beta_2 & \beta_3 \end{bmatrix}$$

It is not difficult to show that $\mathbf{V}$ has full rank for each of the principal types.

3.3 Approximation Theory

The estimators derived in the previous section can be given an additional justification within the framework of approximation theory. In this context the task is to find a polynomial $\hat{g}(x)$ that comes as close as possible to an *unknown* shape function, $g(x) = -v(x)f'(x)/f(x)$, as defined in (1.2). We show that the estimator derived in the previous section provides a polynomial of fixed degree that is closest to $g(x)$ in a natural sense (Cheney 1966).

Consider the space $\mathbf{L}(X)$ of functions $h\colon \mathbf{R} \longrightarrow \mathbf{R}$ for which $E[h^2(X)] < \infty$, where X is a random variable with density $f(x)$ in the class (1.1). A natural norm for $L(X)$ is $|h| = \sqrt{E}[h^2(X)]$. Thus in this space the squared distance between any two functions h_1 and h_2 is

$$|h_1 - h_2|^2 = E[(h_1(X) - h_2(X))^2].$$

The approximation problem is to find a polynomial $\alpha_0 + \cdots + \alpha_k x^k$ that is as close as possible to $g(x)$ in the sense of this norm.

Let $Q(\alpha) = |\alpha_0 + \cdots + \alpha_k x^k - g(x)|^2$. This quadratic criterion has a global minimum at the point, say $\boldsymbol{\beta}$, at which the gradient of Q is zero. This calculation yields

$$E[\beta_0 X^j + \cdots + \beta_k X^{j+k}] = E[X^j g(X)], \qquad j = 0, 1, \ldots, k.$$

An application of Theorem 1 to the right side produces

$$\sum_{i=0}^{k} \beta_i E[X^{i+j}] = E[\{X^j v(X)\}'], \quad j = 0, 1, \ldots, k,$$

which are exactly the same as the moment relations (3.1–3.4) from which the estimators were derived. Thus the estimated $\hat{g}(x) = \beta_0 + \beta_1 + \cdots + \beta_k x^k$ is the closest polynomial of degree k to the unknown $g(x)$ in the space $\mathbf{L}(X)$, given a specified form for $v(x)$.

4. BIMODAL DENSITIES

Among all the distributions of the four principal types as described by (1.1), the relevant ones for bimodal data are those of order three, the minimum order necessary for bimodality. Obtaining consistent estimates for the four coefficients is as easy as solving four simultaneous linear equations, as provided by Corollary 2.

However, as Figure 3 suggests, the Aroian estimates for the $\mathbf{N}_3$ density for small sample size may be noticeably inferior to the maximum likelihood estimates. Further, as suggested by the characterization given in Section 3.3, the Aroian estimates may be quite misleading if the actual distribution in fact has more than two modes. Figure 3 displays the histogram for the crude birth rates of 59 countries (Weinstein 1976, p. 88). Parameter estimates are given in Table 1. The density $\mathbf{N}_3$ is also contrasted with $\mathbf{G}_3$ in Figure 3, since birth rates are always positive.

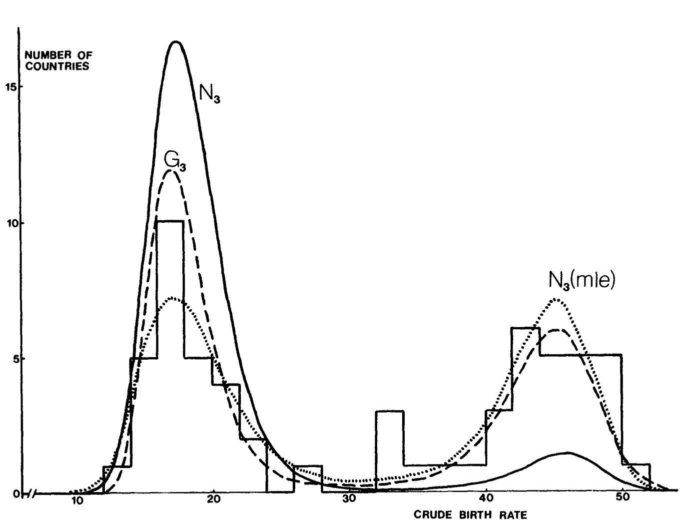

Figure 3. A comparison of the $\mathbf{N}_3$ and $\mathbf{G}_3$ densities as fitted to the data for annual crude birth rates of 59 countries (Weinstein 1966).

Whether an exponential family in the class (1.1) is multimodal depends on the number of roots possessed by the shape polynomial. If the shape polynomial is cubic, then it is possible to construct a statistic that is negative if there are three distinct roots and positive if there is only one real root. This construction was first described by the 16th century mathematician Cardan, for whom it is named. Let $g(x) = b_0 + b_1 x + b_2 x^2 + b_3 x^3$, and let $\lambda = -b_2/(3b_3)$. Then

$$\delta = [g(\lambda)]^2/4 + (b_1 + b_2\lambda)^3/(27b_3)$$

is Cardan's Discriminant, which will serve as our statistic for bimodality. In the case of the Type $\mathbf{N}_3$ density this statistic is particularly useful. If we let

$$\sigma = b_3^{-.25},$$

$$\alpha = -\sigma g(\lambda), \text{ and}$$

$$\beta = -\sigma^2(b_1 + b_2\lambda),$$

then $\delta = (\alpha/2)^2 - (\beta/3)^3$, and the density can be reparameterized as

$$N_3(x) = \xi\exp[\alpha z + \beta z^2/2 - z^4/4], \quad \text{where } z = (x - \lambda)/\sigma. \qquad (4.1)$$

Thus λ is a location parameter and σ is a scale parameter, and the modes and antimode of the density are at the solutions to $\alpha + \beta z - z^3 = 0$. If $\delta < 0$ the density is bimodal, and if $\delta \geq 0$ the density is unimodal. The parameters α (asymmetry) and β (bifurcation) are invariant with respect to changes in location and scale (as is δ), and they have the following approximate interpretations:

asymmetry: if $\delta > 0$ then α is a measure of *skewness*, while if $\delta < 0$ then α indicates the relative heights of the two modes.

bifurcation: if $\delta > 0$ then β is a measure of *kurtosis*, while if $\delta < 0$ then β indicates the relative separation of the two modes.

The relationship between α and β and the modes and antimodes of the $\mathbf{N}_3$ family is shown in Figure 4. The

Table 1. Parameter Estimation for Fitting the Quartic Exponential Distribution ($\mathbf{N}_3$) to the Data on Annual Crude Birth Rates Displayed in Figure 3

	MLE	*S.E. of MLE*	*Aroian Estimates*
λ	31.65	.46	32.47
σ	7.83	.37	7.42
α	−.007	.078	−.64
β	3.28	.30	3.78
δ	−1.3		−1.9

NOTE: S.E. signifies estimated standard error of the MLE and δ denotes Cardan's discriminant. The standard errors were estimated using the Hessian matrix of the log likelihood function. The parameters are as defined in (4.1).

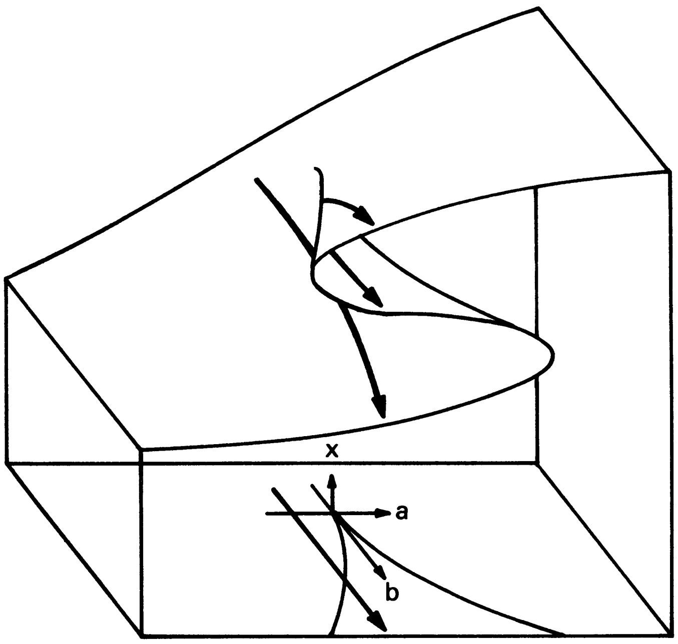

Figure 4. The location of the roots of a cubic shape polynomial, $g(x) = x^3 - bx - a$, graphed in the vertical dimension as a function of the parameters a and b. These roots are the modes and antimodes of the corresponding density. A trajectory parallel to the b-axis is shown together with its image in the surface above. The densities of Figure 1 follow this trajectory.

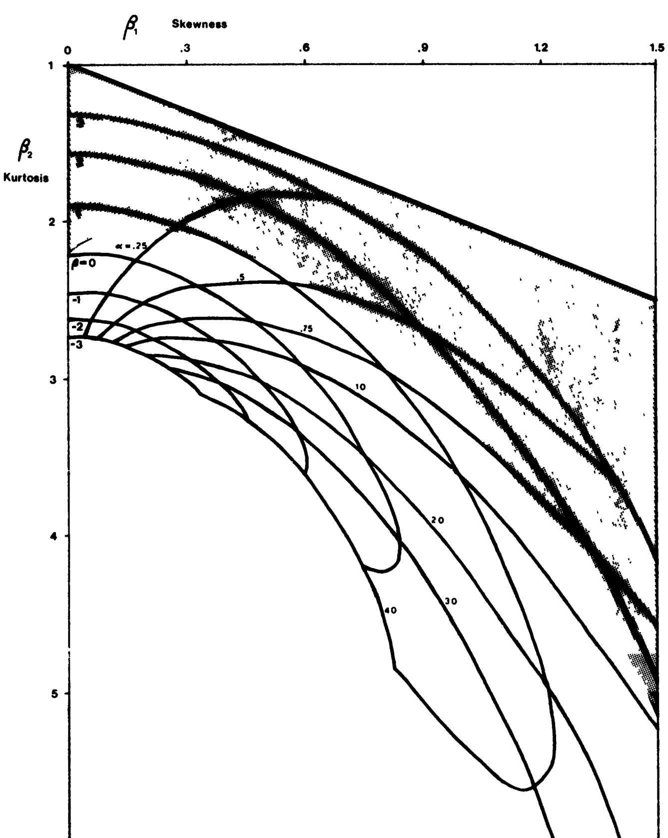

Figure 5. The asymmetry-bifurcation coordinate system of the Type N_3 family mapped into the Pearson $\beta_1 - \beta_2$ coordinate system. The bimodal region (shaded) corresponds roughly to the U-shaped Pearson Type I family. The map has a singularity at $(\beta_1, \beta_2) = (0, 3)$, which is the normal density. (Cf. Johnson and Kotz 1970, p. 14).

relationship between these parameters and Pearson's β_1 and β_2 is shown in Figure 5.

In the cases $\mathbf{G}_3$, $\mathbf{I}_3$, and $\mathbf{B}_3$, Cardan's discriminant is not quite as useful. This is because the interpretation depends on how many of the real roots are actually located within the domain of the density. In addition, even when $\mathbf{B}_3$ has three distinct real roots within its domain it may still be unimodal: this possibility is the case depicted in Figure 2.

An approximate standard error for δ can be calculated by the usual methods, based on the covariance matrix of the estimators for the coefficients of the shape polynomial. This covariance matrix depends on the type of the density and on which method of estimation was used. In each case a test for bimodality can be constructed.

6. CONCLUSIONS

There is a single moment relationship, expressed in Theorem 1, that is valid for a very large class of probability density functions. This class is a generalization of the Pearson system, and it includes many types of multimodal exponential distributions. Consistent estimates may be obtained simply by solving a linear system of moment relations. If maximum likelihood is to be used, then these estimates may serve as the initial vector for the Newton-Raphson iterative procedure.

Except when the mixture assumption is justifiable, the multimodal densities described above are preferable to the class of mixture densities in several respects. The typical mixture density with j modes requires $3j - 1$ parameters, whereas the equivalent multimodal exponential family requires only $2j$ parameters, for which the maximum likelihood method yields unique estimates. In the case $j = 2$, Cardan's discriminant can be used as an indicator of bimodality.

[*Received November 1978. Revised May 1982.*]

REFERENCES

AROIAN, LEO A. (1948), "The Fourth Degree Exponential Distribution," *Annals of Mathematical Statistics*, 19, 589–592.

CHENEY, E.W. (1966), *Introduction to Approximation Theory*, New York: McGraw-Hill.

COBB, LOREN (1978), "Stochastic Catastrophe Models and Multimodal Distributions," *Behavioral Science*, 23, 360–374.

——— (1981a), "Estimation Theory for the Cusp Catastrophe Model," in *1980 Proceedings of the Section of Survey Research Methods*, Washington, D.C.: American Statistical Association, 772–776.

——— (1981b), "The Multimodal Exponential Families of Statistical Catastrophe Theory," in *Statistical Distributions in Scientific Work* (Vol. 4), ed. C. Taillie, G.P. Patil, and B. Baldessari, Dordrecht, Holland: Reidel Press.

COBB, LOREN, and WATSON, WILLIAM B. (1980), "Statistical Catastrophe Theory: An Overview," *Mathematical Modelling*, 1, 311–317.

FISHER, RONALD A. (1922), "On the Mathematical Foundations of Theoretical Statistics," *Philosophical Transactions of the Royal Society of London*, Ser. A, 222, 309–368.

JOHNSON, NORMAN L., and KOTZ, SAMUEL (1970), *Continuous Univariate Distributions* (Vol. 2), New York: Houghton-Mifflin.

KOPPSTEIN, PETER (1980), "The Quartic Exponental Distribution and Bimodal Regression," unpublished manuscript.

LEFEVER, RENÉ (1981), "Noise Induced Transitions in Biological Systems," in *Stochastic Nonlinear Systems*, ed. L. Arnold and R. Lefever, New York: Springer Verlag.

LUDWIG, DONALD (1974), *Stochastic Population Theories*, New York: Springer-Verlag.

MATZ, A.W. (1978), "Maximum Likelihood Parameter Estimation for the Quartic Exponential Distribution," *Technometrics*, 20, 475–484.

ORD, J. KEITH (1972), *Families of Frequency Distributions*, New York: Hafner.

O'TOOLE, A.L. (1933a), "On the System of Curves for Which the Method of Moments is the Best Method of Fitting," *Annals of Mathematical Statistics*, 4, 1–29.

——— (1933b), "A Method of Determining the Constants in the Bimodal Fourth Degree Exponential Function," *Annals of Mathematical Statistics*, 4, 80–93.

POSTON, T., and STEWART, I. (1978), *Catastrophe Theory and Its Applications*, London: Pitman.

SOONG, T.T. (1973), *Random Differential Equations in Science and Engineering*, New York: Academic Press.

WEINSTEIN, JAY A. (1976), *Demographic Transition and Social Change*, Morristown, N.J.: General Learning Press.

WONG, EUGENE (1964), "The Construction of a Class of Stationary Markoff Processes," in *Sixteenth Symposium in Applied Mathematics—Stochastic Processes in Mathematical Physics and Engineering*, ed. Richard Bellman, Providence, R.I.: American Mathematical Society, 264–276.