Inverse Problems 9 (1993) 39–56. Printed in the UK

High-resolution inversion of the discrete Poisson and binomial transformations

Charles Byrne†, Dominique Haughton‡ and Tom Jiang§

† Department of Mathematics, University of Massachusetts-Lowell, Lowell, MA 01854, USA

‡ Department of Mathematical Sciences, Bentley College, Waltham, MA 02154, USA

§ Stryker Biotech, 1 Apple Hill, 594 Worcester Road, Natick, MA 01760, USA

Received 19 July 1991, in final form 22 September 1992

Abstract. The discrete Poisson transformation involved in finite Poisson mixtures arises in many applications, for example in relating observed fluctuations in photon counts to unobserved fluctuations in the number of molecules present, in the analysis of egg counts in zoology, of plankton abundances in fisheries research, of death notice frequencies or of comet frequencies to name a few. The binomial transformation, in addition to its many applications, serves to approximate the Poisson transformation.

Procedures such as the maximum likelihood method (MLM) and the moment method (MM) are frequently used to invert these transforms. For the limited data case we consider here, the low signal-to-noise ratio in the data necessitates some modification of existing methods. We propose a filtering procedure based on regularized Fourier transform estimation, followed by high-resolution eigenvector-based spectral estimation.

Examples are given which compare this approach with MM and MLM. We find improved performance in the estimation of parameters for the discrete model, particularly for the cases of low data set size and reduced signal-to-noise ratios.

1. Introduction

The Poisson transformation of the non-negative function $F(\omega)$, $\omega \geqslant 0$ is the infinite sequence $\{f_n\}$ given by

$$f_n = \int_0^\infty F(\omega) e^{-\omega} \omega^n / n! \, d\omega \qquad n = 0, 1, 2, \ldots . \tag{1}$$

A problem arising in a wide variety of applications is the estimation of the function $F(\omega)$ from noisy estimates of finitely many of the $\{f_n\}$. Often the $F(\omega)$ is a probability density function (PDF), so has integral equal to one. In that case $\{f_n\}$ is also a probability distribution. A random variable X governed by the distribution $\{f_n\}$ would be said to have a *compound Poisson distribution*, with compounding density $F(\omega)$. Typically one would observe T independent instances of X and would use the observed frequencies of occurrence of the integers n $(n = 0, 1, \ldots, N)$ as estimates of the f_n. From such data one attempts to recover $F(\omega)$.

In this paper, we consider the case of discrete $F(\omega)$; here we have $F(\omega)$ supported on a discrete set $\{\omega_j\}$, and having the form

$$F(\omega) = \sum_j F_j \delta(\omega - \omega_j) \tag{2}$$

0266-5611/93/010039+18$07.50

where $\delta(\omega)$ is the Dirac delta and the F_j are positive. We assume that the support of $F(\omega)$ is finite, so that

$$f_n = \sum_{j=1}^{J} F_j e^{-\omega_j} \omega_j^n / n! \qquad n = 0, 1, 2, \ldots. \tag{3}$$

The goal is to estimate J, the $\{\omega_j\}$ and the $\{F_j\}$.

The binomial transform associates with the non-negative function $B(\theta)$, $0 \leqslant \theta \leqslant 1$, the finite sequence

$$b_n^M = b_n = \binom{M}{n} \int_0^1 B(\theta)\theta^n (1-\theta)^{M-n} \, d\theta. \tag{4}$$

Here M is predetermined and fixed; we shall use the notation b_n instead of b_n^M, the dependence on M being understood.

As in the case of the Poisson mixture, we consider the case of finitely supported $B(\theta)$: we assume $B(\theta)$ has the form

$$B(\theta) = \sum_{j=1}^{J} B_j \delta(\theta - \theta_j) \tag{5}$$

for some integer J, positive $\{B_j\}$ and $\{\theta_j\}$ in $[0, 1]$. Our objective will be to estimate J, the $\{\theta_j\}$ and the $\{B_j\}$ from noisy estimates of the $\{b_n\}$.

Once the $\{\omega_j\}$ are determined, the coefficients $\{F_j\}$ are found by constrained maximum likelihood (see section 4 for details). Other types of mixtures are also considered in the literature; we refer the reader to [1, 2] and the references given there.

In [3] Qian uses the discrete Poisson mixture model (3) to relate observed fluctuations in photon counts to unobserved fluctuations in the number of fluorescent molecules present. In this case the ω_j are integral multiples of a fixed known mean intensity ($\omega_j = j\lambda$) and the F_j are sought.

The problem of estimating $F(\omega)$ in the case of a compound Poisson mixture has been considered by Byrne *et al* in [4] and by Bertero and Pike in [5]. In [5], the singular value analysis of the linear operator in equation (1) and the signal-to-noise ratio are used to estimate the function $F(\omega)$ (see also Bertero [6]). Assume it known that the data were generated from a finite Poisson mixture as in equation (3); in this case, using the approach in [5, 6] to estimate the F_j and $\omega_j, j = 1, \ldots, J$ will not take advantage of the prior information we have on the process which generated the data. The techniques we propose are designed to specifically address the finite Poisson mixture problem. The Fourier transform plays a major role in our approach; high-resolution methods (maximum entropy, Capon's minimum variance method, MUSIC and other eigenvector methods) typically involve exploitation of algebraic properties of certain matrices formed from the data. The key questions we try to answer are as follows. Which matrices should be used and how? How sensitive is the estimation to noise in the data?

Titterington *et al* [2] and Everitt and Hand [1] treat the subject of finite mixtures from a statistical perspective and provide a thorough overview of that literature.

A number of methods have been proposed to solve the problem of inverting the Poisson and binomial transforms, for both the continuous and the discretely supported cases (see [1, 2, 5, 6]). Two approaches, the *method of moments* (MM) and the *maximum likelihood method* (MLM) are discussed in some detail in [1, 2].

The method of moments consists of two separate steps: first one estimates from the data several of the moments

$$v_m = \int_0^\infty F(\omega)\omega^m \, d\omega \qquad \text{(Poisson)} \tag{6}$$

or

$$v_m = \int_0^1 B(\theta)\theta^m \, d\theta \qquad \text{(binomial)} \tag{7}$$

second, the estimates $\{V_m\}$ replace the $\{v_m\}$ in the so-called moment equations and the parameters are determined from the solution to these equations, in a manner to be discussed below. In our approach we use the first step, but not the second.

As we shall see, the estimated moments $\{V_m\}$ can be quite inaccurate, particularly for large m. Proceeding to the second step as if they were exact estimates of the v_m is risky. We shall instead employ a filtering process based on regularized Fourier transform estimation, followed by high-resolution eigenvector-based spectral estimation.

The maximum likelihood method is a general approach to parameter estimation. Its desirable properties are asymptotic, that is, they hold in the limit, as the size of the data set goes to infinity. For moderately sized data sets, the signal-to-noise ratio in the data can be low; the MLM does not adequately deal with the presence of noise in the data. In addition, the MLM requires a computationally expensive iterative maximization scheme, which is usually slow to converge and dependent on starting values. Our approach is non-iterative, and explicitly filters the data to reduce the effects of noise.

Comparisons will be made between our approach and the MM and MLM, using as examples the various binomial and Poisson mixture problems set forth in Everitt and Hand [1].

2. The data

2.1. Finite Poisson mixture probability

We assume that we have T independent samples $\{X_t\}$, $t = 1, \ldots, T$, of a non-negative integer-valued random variable X with $\text{Prob}(X = n) = f_n$, $n = 0, 1, 2, \ldots$.

From the $\{X_t\}$ we calculate Y_n = proportion of the $\{X_t\}$ having the value n, for $n = 0, 1, \ldots, N$, where N is the highest value of n observed. Each Y_n is then an estimate of f_n, with mean value (expectation) $E(Y_n) = f_n$ and variance $\text{var}(Y_n) = f_n(1 - f_n)/T$. We define the relative error in Y_n as

$$\text{rel error}(Y_n) = [\text{var}(Y_n)]^{1/2} / \text{mean}(Y_n).$$

It follows that the rel error(Y_n) is $\sqrt{(1/f_n - 1)}/\sqrt{T}$. We see then that the relative error in Y_n is greater when f_n is smaller. Because the relative error varies inversely with $\sqrt{T}$, going from $T = 100$ to $T = 10\,000$ reduces the relative error by only a factor of 10.

Writing $Y_n = f_n + (Y_n - f_n)$ we have Y_n = signal plus noise, where f_n is the (non-random) signal component and $Y_n - f_n$ is the random additive noise. The mean of the noise is zero, since $E(Y_n - f_n) = 0$, but the variance is that of Y_n itself. We see therefore that the noise is related to the signal; we do not say 'correlated with the signal' since the signal is deterministic.

2.2. Finite binomial mixture probability

Assume that the data $\{X_t\}$ are T independent samples of a random variable X taking values in $\{0, 1, \ldots, M\}$, with $\text{Prob}(X = n) = b_n$. As in the Poisson case, we define Y_n to be that proportion of the T samples whose value is n, for $n = 0, 1, \ldots, M$. The Y_n then serve as estimates of the b_n.

The mean of Y_n is $E(Y_n) = b_n$ and the variance is $b_n(1 - b_n)/T$. As in the Poisson case the relative error is $\sqrt{(1/b_n - 1)}/\sqrt{T}$.

With $Y_n = b_n + (Y_n - b_n)$ = signal + noise, we see that the noise $(Y_n - b_n)$ has mean zero, but variance the same as that of Y_n, so the noise is once again related to the deterministic signal, with the relative error increasing as f_n decreases.

In both the Poisson and the binomial case each noise term is related to the signal term for that n. Hence the noise is non-additive white noise.

Defining the signal-to-noise ratio (SNR) as the reciprocal of the relative error, we have that SNR decreases as the signal decreases, so the SNR is related to signal level and varies with n.

3. Estimation techniques

3.1. Maximum likelihood estimation: finite Poisson mixture

We discuss the maximum likelihood (MLM) estimation of the parameters of the finite Poisson mixture probability; the binomial case is essentially the same.

Given the T independent samples $\{X_t\}$, the likelihood function is defined to be

$$\mathrm{L}(J, \{\omega_j\}, \{F_j\}) = \prod_{t=1}^{T} \sum_{j=1}^{J} F_j \mathrm{e}^{-\omega_j} \omega_j^{X_t} / X_t! \tag{8}$$

and the log likelihood function as

$$\mathrm{LL}(J, \{\omega_j\}, \{F_j\}) = \sum_{t=1}^{T} \log \left(\sum_{j=1}^{J} F_j \mathrm{e}^{-\omega_j} \omega_j^{X_t} / X_t! \right). \tag{9}$$

The maximum likelihood method is to select those J, $\{\omega_j\}$ and $\{F_j\}$ for which LL is maximized.

We can rewrite LL as follows

$$\mathrm{LL}(J, \{\omega_j\}, \{F_j\}) = T \sum_{n=0}^{N} Y_n \log \left(\sum_{j=1}^{J} F_j \mathrm{e}^{-\omega_j} \omega_j^{n} / n! \right).$$

The Kullback–Leibler distance between non-negative sequences $\{g_n\}$ and $\{h_n\}$ is defined as

$$\mathrm{KL}(\{g_n\},\{h_n\}) = \sum_n [g_n \log(g_n/h_n) + h_n - g_n] \tag{10}$$

it is non-negative and equals zero if and only if $g_n = h_n$ for each n [7].

We see then that maximizing LL is equivalent to minimizing $\mathrm{KL}(\{Y_n\},\{f_n\})$, with f_n defined by (3). The ML estimate of the parameters attempts to fit the noisy data $\{Y_n\}$ as closely as possible to the $\{f_n\}$ modelled by (3).

Typically an upper bound on J is imposed or some sort of penalty for high J is added to LL; information-theoretic criteria for selecting J are also used [8, 9].

Actually maximizing LL as a function of the parameters requires iterative schemes. One such scheme is the 'expectation maximization' (EM) algorithm (see [10]).

3.2. Moment method: finite Poisson and binomial mixtures

The first step of the moment method (MM) is to estimate the moments, given by

$$v_m = \sum_{j=1}^{J} F_j \omega_j^m \qquad m = 0, 1, 2, \ldots \tag{11}$$

for the Poisson case, and for the binomial case by

$$v_m = \sum_{j=1}^{J} B_j \theta_j^m \qquad m = 0, 1, \ldots, M. \tag{12}$$

For the Poisson case one uses the 'factorial moment estimators' (see [1, 2])

$$V_m = T^{-1} \sum_{t=1}^{T} X_t(X_t - 1) \times \cdots \times (X_t - m + 1) \tag{13}$$

with the sum actually only over those $X_t \geqslant m$. For the binomial case one uses

$$V_m = \frac{1}{T} \sum_{t=1}^{T} \frac{X_t(X_t - 1) \times \cdots \times (X_t - m + 1)}{M(M-1) \times \cdots \times (M - m + 1)} \tag{14}$$

where, again, the sum is actually over those $X_t \geqslant m$.

The second step of the MM uses the solution $\beta = (\beta_0, \beta_1, \ldots, \beta_{J-1})'$ of the linear system $\mathbf{Z}\beta = \varphi$, with $\mathbf{Z}$ the J by J matrix whose entries are $\mathbf{Z}_{m,n} = v_{m+n-2}$, $m, n = 1, \ldots, J$, and φ the J by 1 vector whose entries are $\varphi_m = -v_{J+m-1}$, $m = 1, 2, \ldots J$. Substituting the estimates V_m for the v_m as needed, we obtain an estimated solution β. The true β has the property that the polynomial

$$\beta(z) = \beta_0 + \beta_1 z + \cdots + \beta_{J-1} z^{J-1} + z^J \tag{15}$$

has for its zeros the values $z = \omega_j$ in the Poisson case and $z = \theta_j$ in the binomial case.

From the estimate of β we form an estimate of $\beta(z)$ and estimate the parameters from the zeros of this polynomial. This is the moment method.

In practice the zeros may not lead to feasible estimates: the zeros could be non-real or negative or not in $[0,1]$ for the binomial case. This is due, in part, to inaccuracies in the V_m as estimates of the v_m, but also to the way in which the method relies on inaccurate estimates; the equations $\mathbf{Z}\beta = \varphi$ are solved as if they held exactly, even though the V_m are used in place of the v_m.

To find J components using MM it is necessary to have estimates of v_m for $m = 0, 1, \ldots, 2J-1$. If J is also to be estimated it will be necessary to employ more than $2J$ moment estimates. As we shall see, the V_m can be quite inaccurate, particularly for larger m; in the next section we illustrate this point with an example of a mixture of $J = 4$ Poisson distributions in which V_9 is less than 50% of v_9. If the higher moment estimates are used at all their sizable noise content must be dealt with.

3.3. *Inaccuracies in the V_m as estimates of the v_m*

3.3.1. Poisson case. From (1), (3) and (11) it follows that

$$v_m = \sum_{n=m}^{\infty} \frac{n!}{(n-m)!} f_n \qquad m = 0, 1, 2, \ldots. \tag{16}$$

We can rewrite (13) as

$$\begin{aligned} V_m &= \sum_{n=m}^{N} n(n-1) \times \cdots \times (n-m+1) Y_n \\ &= \sum_{n=m}^{N} \frac{n!}{(n-m)!} Y_n \qquad m = 0, 1, \ldots, N. \end{aligned} \tag{17}$$

Therefore, not only do the V_m have Y_n in place of f_n, but the infinite series (16) is truncated at $n = N$. The value N is determined by the data, not by examining the convergence of the infinite series (16); in an example we consider later, the next term omitted $(n = N + 1)$ is larger than the Nth term, so the V_m is less than half of v_m in that case.

The V_m place most weight on the Y_n with higher n. In the Poisson case it is typical for the Y_n for n near N to be very small, and therefore highly unreliable. The V_m for m smaller are the more reliable, since they use the larger Y_n with n nearer to zero.

The moment method requires $\{V_m, \ m = 0, 1, \ldots, 2J-1\}$. The reliability of the V_m for m near $2J-1$ depends on the ω_j; if all the ω_j are small, then the V_m for larger m are bad estimates.

3.3.2. Binomial case. From (4), (5) and (12) and the identities

$$\sum_{n=m}^{M} \alpha_n \left[\binom{M}{n} \theta^n (1-\theta)^{M-n} \right] = \theta^m \qquad 0 \leqslant m \leqslant M \tag{18}$$

where

$$\alpha_n = \frac{n(n-1)\times\cdots\times(n-m+1)}{M(M-1)\times\cdots\times(M-m+1)} \qquad m \leqslant n \leqslant M \tag{19}$$

it follows that for $m = 0, 1, \ldots, M$

$$v_m = \sum_{n=m}^{M} \frac{n(n-1)\times\cdots\times(n-m+1)}{M(M-1)\times\cdots\times(M-m+1)} b_n. \tag{20}$$

We can rewrite (14) as follows

$$V_m = \frac{1}{T}\sum_{n=m}^{M}(TY_n)\frac{n(n-1)\times\cdots\times(n-m+1)}{M(M-1)\times\cdots\times(M-m+1)} \tag{21}$$

or

$$V_m = \sum_{n=m}^{M} \frac{n(n-1)\times\cdots\times(n-m+1)}{M(M-1)\times\cdots\times(M-m+1)} Y_n. \tag{22}$$

So we see that V_m estimates v_m simply by replacing each b_n with the estimate Y_n.

How small the Y_n are, and consequently how reliable, depends on the locations of the θ_j. For m near M the V_m rely on the Y_n with n near M. If all the θ_j are near zero, these Y_n are small, hence unreliable. If all of the θ_j are near one these Y_n for n near M become more reliable. In general these Y_n with n larger are more reliable with respect to information about θ_j near one and less reliable about the θ_j near zero.

3.3.3. Example. Consider a Poisson mixture with $J = 4$, $F_1 = F_2 = F_3 = 0.2$, $F_4 = 0.4$, $\omega_1 = 0.5$, $\omega_2 = 1.5$, $\omega_3 = 3.0$ and $\omega_4 = 5.0$ (see [1, p 101]). Then

$$v_9 = \sum_{n=9}^{\infty} n(n-1)\cdots(n-8)f_n$$
$$= 9!\, f_9 + 10!\, f_{10} + \cdots.$$

Even though f_{10} is about half of f_9 in this case $10!\, f_{10}$ is five times $9!\, f_9$. In simulations with $T = 200$ (as in [1]) we obtained estimates $Y_9 = 0.0150$ and $Y_{10} = 0$ for f_9 and f_{10} respectively. Consequently the sum for the estimator V_9 stopped at $9!\, Y_9$; the next omitted term is five times larger than $9!\, f_9$ in the true infinite series. The estimate V_9 differs from the true v_9 by at least 50%.

3.4. Data reversal for the binomial mixture

The moment estimates V_m emphasize the counts Y_n for n nearer to M. When the binomial parameter θ is near zero the b_n for n small are larger than for n near M. Observed count data, Y_n, will be similarly skewed, with those Y_n for n nearer to M being nearly zero; this is particularly evident when the sample size T is small. As a result the calculated moments V_m weight more heavily precisely the more unreliable Y_n, the smaller ones with the most relative error.

When there are several θ_j in the mixture the moments will contain more accurate information about the θ_j near one, and less accurate information about the θ_j near zero. We can avoid this imbalance in estimate accuracy through a two step process that estimates the two groups of θ_j separately.

In the first step we process the Y_n as described above, letting Y_n serve as our estimate of b_n in the moment calculations. The θ_j so estimated should be more accurate if they are in $(1/2, 1)$, and less so in $(0, 1/2)$. We accept those in $(1/2, 1)$ but not in $(0, 1/2)$, and proceed to step two.

In step two we let b_{M-n} be estimated by Y_n. In effect we switch the definitions of 'success' and 'failure' in the binomial trials. Previously a value of $Y_6 = 14$ meant that out of T repetitions of M trials we had 6 successes 14 times. Now we view these 6 as failures and say that there were $M - 6$ successes. Each θ_j is converted into $1 - \theta_j$ by this switch in interpretation of the data, so those θ_j in $(0, 1/2)$ now show up as values $1 - \theta_j$ in $(1/2, 1)$. Thus we estimate the θ_j in $(0, 1/2)$ and $(1/2, 1)$ separately.

3.5. Fourier transform methods

The functions $F(\omega)$ and $B(\theta)$ can be viewed as power spectra having Fourier transforms

$$f(x) = \int_0^\infty F(\omega)e^{ix\omega}\, d\omega \tag{23}$$

and

$$b(x) = \int_0^1 B(\theta)e^{ix\theta}\, d\theta. \tag{24}$$

One approach to recovering the $F(\omega)$ or $B(\theta)$ is to estimate values of $f(x)$ or $b(x)$ from the data and then to employ high-resolution power spectral estimates.

Expanding $e^{ix\omega}$ in a Taylor series we get

$$f(x) = \sum_{m=0}^{\infty} (ix)^m \int_0^\infty F(\omega)\omega^m\, d\omega/m! \tag{25}$$

and

$$b(x) = \sum_{m=0}^{\infty} (ix)^m \int_0^1 B(\theta)\theta^m\, d\theta/m!. \tag{26}$$

From (1) and (4) we see that we have estimates of

$$f_n = \int_0^\infty F(\omega)e^{-\omega}\omega^n\, d\omega/n!$$

and

$$\frac{b_n}{\binom{M}{n}} = \int_0^1 B(\theta)(1-\theta)^{M-n}\theta^n\, d\theta.$$

In each case there is an unwanted term in the integrand, if we are to use (25) or (26).

From (6), (7), (25) and (26) we have

$$f(x) = \sum_{n=0}^{\infty}(\mathrm{i}x)^n v_n / n! \tag{27}$$

and

$$b(x) = \sum_{n=0}^{\infty}(\mathrm{i}x)^n v_n / n!. \tag{28}$$

So we can estimate $f(x)$ or $b(x)$ from the moment estimates V_m.

For the binomial case we have V_m, $m = 0, \ldots, M$, while for the Poisson case, for $m = 0, 1, \ldots, N$. In either case we would be using truncated power series to estimate $f(x)$ or $b(x)$. The rate of convergence of these series may pose a problem. If the θ_j in the binomial case are well distributed within $(0, 1)$ we will have nearly M usable V_m; on the other hand, it is more common in the Poisson case to have fewer usable V_m.

We shall consider next the approximation of a Poisson mixture by a binomial; then $\theta_j = \omega_j / M$ and the θ_j are clustered near zero, with few usable V_m.

3.6. Poisson mixture approximated by binomial mixture

We can approximate a finite Poisson mixture by a finite binomial mixture, for suitably chosen M. The binomial parameters are then found as above and transformed to obtain the Poisson parameters.

From (4) and (5) we have

$$b_n = b_n^M = \binom{M}{n} \sum_{j=1}^{J} [B_j (1 - \theta_j)^{-n}] (1 - \theta_j)^M \theta_j^n \tag{29}$$

or

$$b_n \approx \binom{M}{n} M^{-n} \sum_{j=1}^{J} [B_j (1 - \theta_j)^{-n}] \mathrm{e}^{-M\theta_j} (M\theta_j)^n. \tag{30}$$

With $\theta_j = \omega_j / M$ we then have

$$b_n \approx \frac{\binom{M}{n} n!}{M^n} \sum_{j=1}^{J} [B_j (1 - \omega_j / M)^{-n}] \mathrm{e}^{-\omega_j} \omega_j^n / n!. \tag{31}$$

If M is large enough and ω_j is small enough so that $(1 - \omega_j / M)^{-n}$ is nearly one, at least for $n = 0, 1, \ldots N$ we have (compare (3))

$$b_n \approx f_n = \sum_{j=1}^{J} B_j \mathrm{e}^{-\omega_j} \omega_j^n / n! \tag{32}$$

note that for $0 \leqslant n \leqslant N$ and $M \gg N$ we have $\binom{M}{n} n! M^{-n} \approx 1$.

Therefore, given estimates of F_n for $n = 0, 1, \ldots, N$ we select $M > N$ and view the f_n as b_n^M for that M. The ω_j become $\theta_j = \omega_j / M$, so the θ_j are now clustered near zero, not spread throughout $(0, 1)$. If we have an *a priori* upper bound L on the support of $F(\omega)$ then the θ_j lie in $(0, L/M)$. We reverse the data, viewed as binomial, so that the interval of support becomes $(A, B) = (1 - L/M, 1)$. The values $A = 1 - L/M$, $B = 1$ are then used in the Fourier transform estimation step we discuss next.

3.7. *Estimating $f(x)$ and $b(x)$ from moments*

We begin with the problem of estimating $f(x)$ from finitely many moments v_m, $m = 0, 1, \ldots, N$. The case of $b(x)$ will be similar.

First, we seek a linear estimator of $F(\omega)$ of the form

$$\hat{F}(\omega) = \sum_{n=0}^{N} a_n \omega^n \qquad 0 \leqslant \omega \leqslant L \tag{33}$$

where we assume $F(\omega) = 0$ for ω outside $[A, B]$. Matching moments we get, for $m = 0, 1, \ldots, N$

$$v_m = \int_A^B \hat{F}(\omega) \omega^m \, d\omega = \sum_{n=0}^{N} a_n \int_A^B \omega^n \omega^m \, d\omega \tag{34}$$

or

$$v_m = \sum_{n=0}^{N} \mathbf{T}_{m,n} a_n \qquad m = 0, 1, \ldots, N. \tag{35}$$

The entries $\mathbf{T}_{m,n}$ are

$$\mathbf{T}_{m,n} = \int_A^B \omega^n \omega^m \, d\omega = \frac{B^{m+n+1} - A^{m+n+1}}{m+n+1}. \tag{36}$$

The matrix $\mathbf{T} = [\mathbf{T}_{m,n}]$ is positive definite but typically ill-conditioned. To regularize the solution of the system we can increase the main diagonal by adding $\epsilon \mathbf{I}$ to $\mathbf{T}$ before inversion (where ϵ is a small positive quantity). Using $\epsilon > 0$ leads to a solution $\hat{F}(\omega)$ that does not exactly match the moments. When the inaccurate V_m are used in place of v_m relaxing the moment matching criterion is good and acknowledges the errors in V_m. This estimation procedure therefore permits prior information about the support of $F(\omega)$ to be included in the filtering of the noisy V_m (see [11, 12]).

Once we have $\hat{F}(\omega)$ we estimate $f(x)$ for suitable values of x. The $\hat{F}(\omega)$, being a linear estimate of $F(\omega)$, may not resolve the various discrete components, so may not be an adequate reconstruction of $F(\omega)$ for our purposes. We use $\hat{F}(\omega)$ only to estimate $f(x)$, then apply high-resolution methods. The point of view we adopt is that $\hat{F}(\omega)$ is a smoothed version of $F(\omega)$, capable of providing decent estimates of $f(x)$ for x nearer to zero (low-frequency information) but lacking high-frequency information. The role of high-resolution estimation is to extract the parameter estimates by supposing a discrete $F(\omega)$ and looking for its points of support.

The estimation of $f(x)$ is as follows

$$\hat{f}(x) = \int_A^B \hat{F}(\omega) e^{ix\omega} \, d\omega \tag{37}$$

or

$$\hat{f}(x) = \sum_{n=0}^{N} a_n \int_A^B e^{ix\omega} \omega^n \, d\omega \qquad \hat{f}(0) = V_0. \tag{38}$$

Let

$$S_n = \int_A^B e^{ix\omega} \omega^n \, d\omega. \tag{39}$$

Then from integration by parts we have, for $x \neq 0$

$$S_n = \frac{1}{ix}(B^n e^{iBx} - A^n e^{iAx}) - \frac{n}{ix} S_{n-1}. \tag{40}$$

Starting with $S_0 = \frac{1}{ix}(e^{iBx} - e^{iAx})$, we obtain the S_n recursively. So we have

$$\hat{f}(x) = \sum_{n=0}^{N} a_n S_n \qquad \hat{f}(0) = V_0. \tag{41}$$

The formula for estimating $b(x)$ is exactly the same. In the binomial case $N = M$, but in either case it is not necessary to use all the available moment estimates V_m; higher indexed V_m may be dropped, so that the N in (33) may be replaced by a smaller value.

We use formula (41) for $\hat{f}(x)$ (or $\hat{b}(x)$) to estimate $f(x)$ (or $b(x)$) at certain values of x. In the Poisson case (the binomial is identical), if the support of $F(\omega)$ is known to be within $[A, B]$ then the Nyquist sampling rate is $\Delta = 2\pi/(B - A)$. We calculate $f(k\Delta)$ for $k = 1, 2, \ldots, K$ from (41) and take $\hat{f}(0) = V_0 = 1$. Then $\hat{f}(-k\Delta) = f(k\Delta)$ since $\hat{F}(\omega)$ is real-valued. These estimates $\hat{f}(k\Delta)$, $|k| \leqslant K$ are then used as input to high-resolution estimates for $F(\omega)$; in this paper we consider eigenvector-based methods [13].

We shall want the estimates $\hat{f}(k\Delta)$, $|k| \leqslant K$ to be non-negative definite, which will happen whenever $\hat{F}(\omega)$ is non-negative, although this is not necessary. If $\hat{F}(\omega)$ is not non-negative, then it may still happen that for K sufficiently small the values $\{\hat{f}(k\Delta), |k| \leqslant K\}$ can be made part of a non-negative definite sequence. This is all we need.

The binomial mixture is handled similarly. With no prior knowledge that the $\{\theta_j\}$ are clustered within a subregion of $(0, 1)$ we let $A = 0$, $B = 1$ and proceed as above.

If the binomial problem arises as an approximation to a Poisson mixture problem, then, as we saw earlier, the data reversal is performed, so that the new $B(\theta)$ then has support $(A, B) = (1 - L/M, 1)$. These values of A and B are then used in (36) and (39).

3.8. Eigenvector-based parameter estimation

Once we have $\hat{f}(k\Delta)$, $|k| \leqslant K$, estimates of the Fourier transform of the non-negative function $F(\omega)$, we estimate the non-negative definite Toeplitz matrix $\mathbf{R}$ whose entries are $R_{m,n} = f((m-n)\Delta)$, $m, n = 1, \ldots, K+1$.

The method is based on the fact that the $K+1$ by $K+1$ matrix $\mathbf{R}$ has rank J and can be expressed as

$$\mathbf{R} = \sum_{j=1}^{J} F_j e_j e_j^{+} \tag{42}$$

with $e(\omega) = (1, \exp(\mathrm{i}\Delta\omega), \ldots, \exp(\mathrm{i}K\Delta\omega))^{+}$ and $e_j = e(\omega_j)$.

Let $u_1, \ldots, u_{K+1}$ be orthonormal eigenvectors of $\mathbf{R}$ corresponding to the ordered eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{K+1} \geqslant 0$. Then (see [13])

$$\lambda_J > \lambda_{J+1} = \cdots = \lambda_{K+1} = 0$$

and

$$u_k^{+} e_j = 0 \qquad k \geqslant J+1 \qquad j = 1, \ldots, J.$$

The function

$$U(\omega) = \sum_{k=J+1}^{K+1} |u_k^{+} e(\omega)|^2 \tag{43}$$

is then zero at precisely the values $\omega = \omega_j$.

In practice we do not know $\mathbf{R}$ exactly; the entries $R_{m,n} = f((m-n)\Delta)$ are replaced by the estimated values, $\hat{f}((m-n)\Delta)$. The condition we need is that the estimated matrix $\hat{\mathbf{R}}$ so obtained be non-negative definite. In that case there should be J essentially non-zero eigenvalues and $K+1-J$ essentially zero ones. We then calculate the u_k for $k \geqslant J+1$ and proceed as before.

4. Examples and comparison of methods

In this section we investigate the performance of our techniques in both the binomial and Poisson problem, for small and large sample sizes, and we compare our results with those obtained by other methods.

4.1. Binomial problem

4.1.1. Four-binomial mixture; 200 observations. Here $J = 4$, $T = 200$, $M = 30$ and we use the data given in [1, pp 92–3]. The true amplitudes of the components are $B_1 = B_2 = B_3 = 0.2$, $B_4 = 0.4$. The moments are calculated according to equation (22), and their Fourier transform approximately evaluated according to equation (28), at values of x a distance of $2\pi/\ell$ apart, where $\ell = 15$. With this choice of ℓ, the series in equation (28) converges very fast. In figure 1 we draw the graph of the inverse of the estimator in equation (43) and expect peaks approximately at the θ_j's,

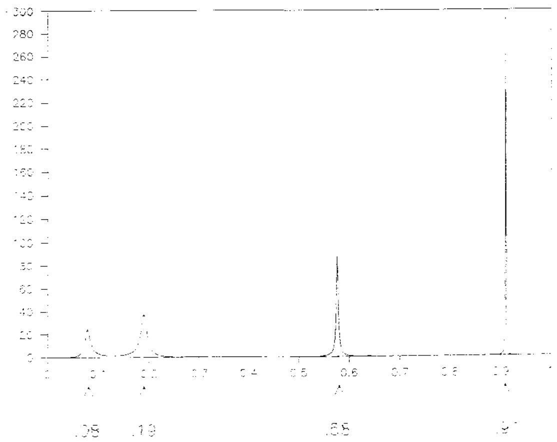

Figure 1. Graph of the inverse of the estimator in equation (43) for the four-binomial mixture and 200 observations.

Table 1. Results for the four-binomial mixture with 200 observations.

True Value	MM	ML (True starting values)	ML (MM starting values)	Our Method
$\theta_1 = 0.1$	0.067	0.09049	0.09088	0.08
$\theta_2 = 0.2$	0.177	0.21884	0.21968	0.19
$\theta_3 = 0.6$	0.569	0.60456	0.60518	0.58
$\theta_4 = 0.9$	0.909	0.91217	0.91232	0.91

$j = 1, \ldots, 4$. Table 1 gives results of applying the MM, ML methods with initial values for the parameters selected to be (i) the true values, (ii) the MM values, and our method (see [1, pp 92–3].

We note that the Toeplitz matrix, in equation (42), very clearly has four dominant eigenvalues in this example, so that we obtain a reliable estimator for J. In order to use the MM or ML methods, one must assume a value for J. Our technique improves on the moment method; the ML method works well here because the θ_j's are well separated but it evidently attempts to fit the noise in the data since it tends to move away from the true starting values. We also note that reversing the data leads to the same results.

4.1.2. Two-binomial mixture; 20 000 observations. In this example we generated 20 000 observations from a mixture of two binomials, with $B_1 = 0.2$, $B_2 = 0.8$, $M = 30$ and $\theta_1 = 0.2$, $\theta_2 = 0.7$. Figure 2 gives the graph of the inverse of the estimator in equation (43), obtained by the same process as in the previous example, also with $\ell = 15$. We observe two very clear peaks, at 0.201 and 0.701—very accurate estimates of θ_1 and θ_2.

4.2. Poisson problem

4.2.1. Death notices data from The Times of London. Here the data give the numbers of death notices of women 80 years of age and over, appearing in The Times, on each day for three consecutive years (see [1, p 99]). In this example the sample size T is 1096. In [1], the MM method was applied assuming that $J = 2$, yielding $\omega_1 = 1.102$ and $\omega_2 = 2.581$. In [2, p 90] the ML method was used (with the EM algorithm, and the MM estimators as starting values), again assuming that $J = 2$,

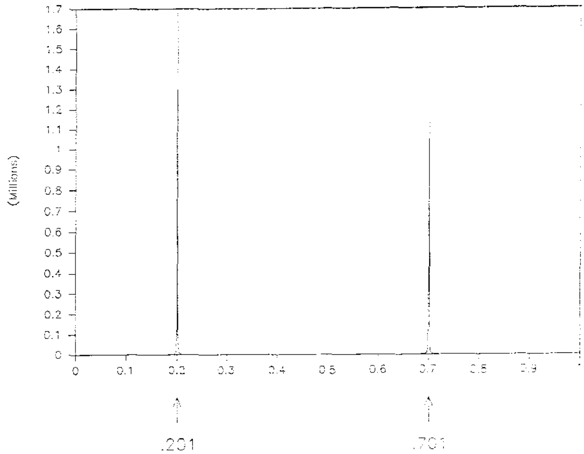

Figure 2. Graph of the inverse of the estimator in equation (43) for the two-binomial mixture and 20 000 observations.

yielding $\omega_1 = 1.255$ and $\omega_2 = 2.66$. A two-component Poisson mixture fits the data much better than just one Poisson component (see [1, p 90]). In order to apply our methodology to this data set, we reverse the data and view the reversed data as approximately generated by a mixture of binomials with $M = 9$. Using a bound of 7 on the possible Poisson components, we are lead to search for binomial components for the reversed data over the interval $(1 - 7/9, 1)$ of length $7/9$. We then use the estimator of the Fourier transform of the binomial moments given in equation (41) and condition the $\mathbf{T}$ matrix in equation (36) (taken to be of size 6×6) by adding to it 10^{-6} times the identity matrix. This procedure yields three binomial components, using a Toeplitz matrix of size 6×6; the corresponding Poisson components are found to be $\omega_1 = 3.132$, $\omega_2 = 2.124$ and $\omega_3 = 1.17$. The amplitudes F_1, F_2 and F_3 of the three components are calculated by constrained maximization of the likelihood function of the data, using a program based on grid searches (see [14, pp 712–6], for details on grid methods). The amplitudes are found to be $F_1 = 0.26495$, $F_2 = 0.4935$ and $F_3 = 0.24155$, yielding a log-likelihood function of -1990.23. In order to decide whether a model with three components or a model with two components is most appropriate, we calculate the likelihood of all submodels with two components, and use an information criterion (see [8, 9]) to help determine whether the increase in likelihood justifies the choice of a larger model with three components. An appropriate information criterion here is

$$\text{BIC} = \text{log-likelihood } -(1/2)\times \text{ dimension of model } \times \log T$$

where T is the number of observations (see [8, 9]). To choose the best model we maximize BIC. The results are shown in tables 2 and 3.

Table 2. Results for the three component model.

Poisson components	Amplitudes	log-likelihood	BIC
$\omega_1 = 3.132$	$F_1 = 0.26495$		
$\omega_2 = 2.124$	$F_2 = 0.4935$	-1990.23	-2000.73
$\omega_3 = 1.17$	$F_3 = 0.24155$		

Table 3. Results for various two component models.

Poisson components	Amplitudes	log-likelihood	BIC
$\omega_1 = 3.132$	$F_1 = 0.08745$		
$\omega_2 = 2.124$	$F_2 = 0.91255$	−1998.96	−2006.0
$\omega_1 = 2.124$	$F_1 = 0.87995$		
$\omega_2 = 1.17$	$F_2 = 0.12005$	−1996.9	−2003.9
$\omega_1 = 3.132$	$F_1 = 0.52951$		
$\omega_2 = 1.17$	$F_2 = 0.47049$	−1998.8	−2005.8

The BIC criterion selects the model with three components, a reasonable option if one keeps in mind that none of the amplitudes F_1, F_2 or F_3 seems to be close to zero. In [1, 2] the possibility of a three-component Poisson mixture to fit this data set was not explored. Figure 3 gives a graph of the inverse of the estimator in equation (43).

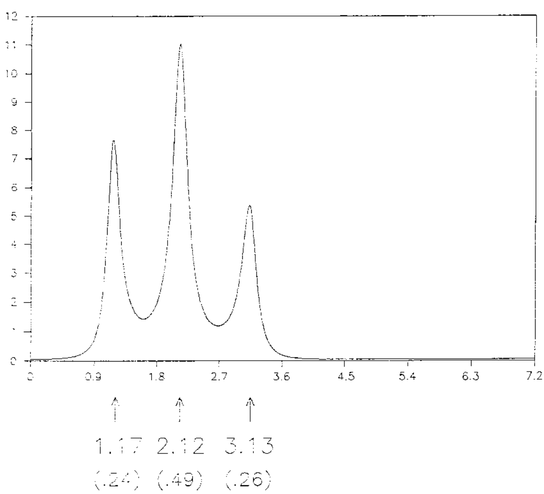

Figure 3. Graph of the inverse of the estimator in (43) for death data from The Times (1096 observations). Amplitudes are shown in parentheses.

4.2.2. Three-Poisson mixture; 500 observations. Using data from [1, p 100] generated from a mixture of three Poisson components with $\omega_1 = 0.5$, $\omega_2 = 3$, $\omega_3 = 6$ and $F_1 = F_2 = 0.3$ and $F_3 = 0.4$, we apply the same methodology as in section 4.2.1, with $M = 15$ for the approximating binomial mixture. The technique yields four components, but one of the components has a very small amplitude of 4.5×10^{-5}. One can obtain three components by taking a smaller Toeplitz matrix in equation (42), but too small a Toeplitz matrix risks distorting the estimated values for ω_1, ω_2 and ω_3.

The results are shown in table 4.

Our results improve on the moment method (see figure 4 for a graph of the inverse of the estimator). Note that the relative strength of a component cannot be inferred from the relative height of its peak.

Table 4. Results for the three-Poisson mixture with 500 observations (including MM results from [1, p 100], for comparative purposes).

True	MM	Our method	Amplitudes
$\omega_1 = 0.5$	0.143	0.705	0.2255
$\omega_2 = 3$	2.921	2.625	0.4431
$\omega_3 = 6$	7.165	6.555	0.3314
		1.65	4.5×10^{-5}

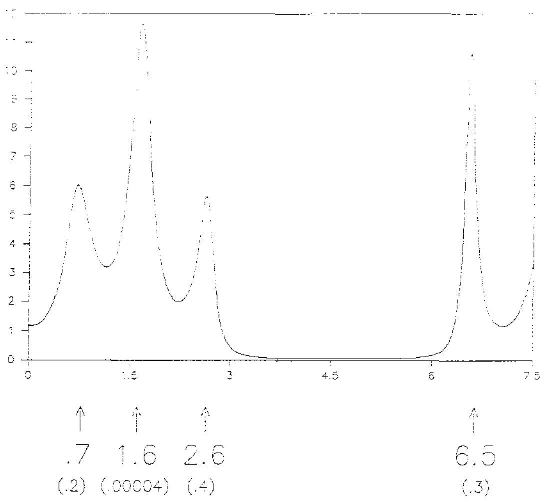

Figure 4. Graph of the inverse of the estimator in (43) for the three-Poisson mixture with 500 observations. Amplitudes are shown in parentheses.

4.2.3. Three-Poisson mixture; 20 000 observations. In this example we generated 20 000 observations from a Poisson mixture with the same parameters as in section 4.2.2. We obtain $\omega_1 = 0.195$, $\omega_2 = 2.34$, $\omega_3 = 6.375$ and $\omega_4 = 1.32$, with amplitudes $F_1 = 0.1915$, $F_2 = 0.382$, $F_3 = 0.3986$ and the small $F_4 = 0.0269$ (see figure 5 for a graph of the inverse of the estimator). With so many observations one would actually get better results by using (25) to estimate the Fourier transform of $F(\omega)$, in (23), from estimates of the $\{f_n\}$, and sampling the estimator at points $x = k\Delta + \mathrm{i}$, where k is an integer, and $\Delta = 2\pi/\ell$, with ℓ equal to a bound of the Poisson components; we add i to counteract the effect of the $\mathrm{e}^{-\omega}$ term in equation (1) (see figure 6 for an illustration of the results). This latter technique fails dramatically as soon as the sample size drops; this very fact is one of the motivating factors in using binomial approximations to Poisson mixtures for low sample sizes. The benefits of this approach become particularly apparent in the next example, in which the MM fails, and the ML depends very much on initial values and requires many iterations, while our method gives very reasonable results, even with a low sample size of 200.

4.2.4. Four-Poisson mixture; 200 observations. In this example we apply the same methods as in the examples given in sections 4.2.1–3, with $M = 9$ for the approximating binomials, to data from [1, p 101]. Here we have used a bound of $\ell = 6$ for the possible Poisson components. The results are shown in table 5. Figure 7 shows a graph of the inverse of the estimator. Our results improve on both the MM and ML methods.

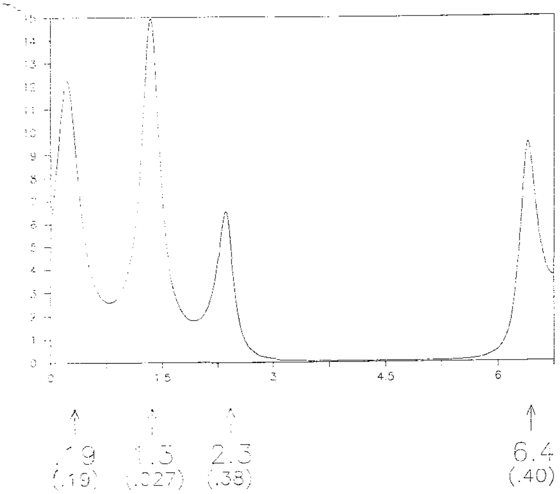

Figure 5. Moments and Fourier transform for the three-Poisson mixture with 20 000 observations. Amplitudes are shown in parentheses.

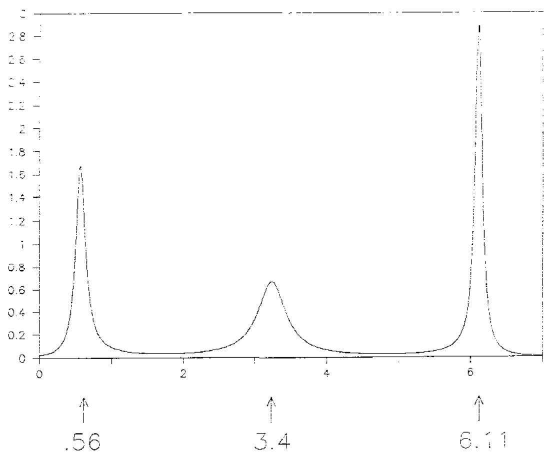

Figure 6. Direct Fourier transform for the three-Poisson mixture with 20 000 observations.

Table 5. Results for the four-Poisson mixture with 200 observations.

True	ML/EM True starting values	ML/EM Starting values (1,2,4.5,6)	MM	Our method
$\omega_1 = 0.5$	0.75	0.02	Fails due	0.612
$\omega_2 = 1.5$	1.34	1.31	to complex	1.584
$\omega_3 = 3.0$	3.37	4.10	roots	2.457
$\omega_4 = 5.0$	5.08	4.87		5.724

5. Conclusions

When the data set size T is small the observed counts provide noisy estimates of the moments, which cannot be accepted as exact, as the moment equations method would have us do. By using the estimated moments to obtain a regularized estimate of the Fourier transform of $F(\omega)$ or $B(\theta)$ we effectively filter the noisy moment data. High-resolution eigenvector-based methods can then be used to obtain reliable estimates of the number of components and their parameters.

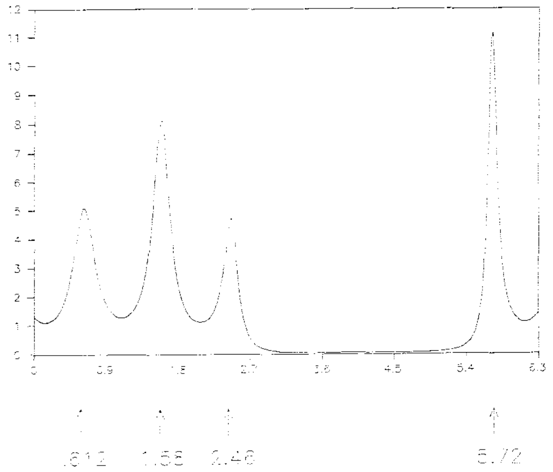

Figure 7. Graph of the inverse of the estimator for the four-Poisson mixture with 200 observations.

The maximum likelihood method attempts to fit the noisy observed frequencies to the model of the probabilities. For low values of T the ML estimates are not as good as those obtained via the Fourier transform approach presented here.

References

[1] Everitt B S and Hand D J 1981 *Finite Mixture Distributions* (London: Chapman and Hall)

[2] Titterington D M, Smith A E M and Makov U E 1985 *Statistical Analysis of Finite Mixture Distributions* (New York: Wiley)

[3] Qian H 1990 Inverse Poisson transformation and shot noise filtering *Rev. Sci. Instrum.* **61** 2088–91

[4] Byrne C L, Levine B M and Dainty J C 1984 Stable estimation of the probability density function of intensity from photon frequency counts *J. Opt. Soc. Am.* A **1** 11

[5] Bertero M and Pike E R 1986 Intensity fluctuation distributions from photon counting distributions: a singular-system analysis of Poisson transform inversion *Inverse Problems* **2** 259–69

[6] Bertero M 1992 Sampling theory, resolution limits and inversion methods *Inverse Problems in Scattering and Imaging* ed M Bertero and E R Pike (Bristol: Adam Hilger) pp 71–94

[7] Kullback S and Leibler R A 1951 On information and sufficiency *Ann. Math. Statist.* **22** 79–86

[8] Schwarz G 1978 Estimating the dimension of a model *Ann. Statist.* **6** 461–4

[9] Haughton D M A 1988 On the choice of a model to fit data from an exponential family *Ann. Statist.* **16** 342–55

[10] Dempster A, Laird N and Rubin D 1977 Maximum likelihood from incomplete data via the EM algorithm *J. R. Statist. Soc.* B **39** 1–38

[11] Byrne C L and Fitzgerald R M 1984 Spectral estimators that extend the maximum entropy and maximum likelihood methods *SIAM J. Appl. Math.* **44** 425

[12] Byrne C L, Fitzgerald R, Fiddy M, Darling A and Hall T 1983 Image restoration and resolution enhancement *J. Opt. Soc. Am.* A **73** 1481–7

[13] Johnson D and DeGraff S 1982 Improving the resolution of bearing in passive SONAR arrays by eigenvector analysis *IEEE Trans. ASSP* **30** 638–47

[14] Quandt R E 1983 Computational problems and methods *Handbook of Econometrics* vol I, ed Z Griliches and M D Intriligator (Amsterdam: North-Holland)