Maximum Likelihood Estimates for Multinomial Probabilities via Geometric Programming

M. Mazumdar; T. R. Jefferson

Biometrika, Vol. 70, No. 1. (Apr., 1983), pp. 257-261.

Biometrika (1983), **70**, 1, *pp.* 257–61
Printed in Great Britain

Maximum likelihood estimates for multinomial probabilities via geometric programming

By M. MAZUMDAR and T. R. JEFFERSON
Department of Industrial Engineering, University of Pittsburgh, Pennsylvania, U.S.A.

SUMMARY

This paper illustrates how the duality theory of geometric programming and the special methods available for solution can be used to obtain maximum likelihood estimates for multinomial probabilities. Three examples are given.

Some key words: Duality; Nonlinear programming; Posynomial; Signomial program.

1. INTRODUCTION

The purpose of this paper is to illustrate the use of geometric programming to obtain maximum likelihood estimates for multinomial probabilities. Three examples are given. The first example (Rao, 1973, p. 370) pertains to estimation of gene frequencies from observed data on blood groups. The second example refers to a problem of estimating compounded probabilities that was considered by Fisher (1942). The third example is a problem of Samaniego & Jones (1981) relating to estimation for a class of multinomial distributions arising in reliability.

Geometric programming considers the following classes of optimization problems involving 'posynomials' in $t = (t_1, \ldots, t_m)$: minimize $g_0(t)$ subject to $g_k(t) \leqslant 1$ $(k = 1, \ldots, p)$ and positivity conditions $t_j > 0$ $(j = 1, \ldots, m)$, where

$$g_k(t) = \sum_{i \in [k]} c_i \prod_{j=1}^{m} t_j^{a_{ij}} \quad (k = 0, 1, \ldots, p),$$

$$[k] = m_k, m_k + 1, \ldots, n_k \quad (k = 0, 1, \ldots, p),$$

$$m_0 = 1, \quad m_1 = n_0 + 1, \ldots, m_p = n_{p-1} + 1, \quad n_p = n.$$

Here, the sets $[k]$ are a collection of disjoint index sets which form a sequential partition of the integers 1 to n, the a_{ij} are arbitrary real numbers and the coefficients c_j are positive. The name posynomial is derived from the fact that the functions have polynomial forms with positive coefficients. The dual of this program is given by: maximize

$$v(\delta) = \prod_{i=1}^{n} (c_i/\delta_i)^{\delta_i} \prod_{k=1}^{p} \lambda_k(\delta)^{\lambda_k(\delta)},$$

where

$$\lambda_k(\delta) = \sum_{i \in [k]} \delta_i \quad (k = 1, \ldots, p)$$

subject to the linear constraints

$$\sum_{i \in [0]} \delta_i = 1, \quad \sum_{i=1}^{n} a_{ij} \delta_i = 0 \quad (j = 1, \ldots, m), \quad \delta_i \geqslant 0 \quad (i = 1, \ldots, n).$$

The solutions to the primal and dual programs, denoted by t^* and δ^* respecitvely, are related by $g_0(t^*) = v(\delta^*)$ and

$$c_i \prod_{j=1}^{m} t_j^{a_{ij}} = \begin{cases} \delta_i^* v(\delta^*) & i \in [0], \\ \delta_i^*/\lambda_k(\delta^*) & i \in [k], \quad \lambda_k(\delta^*) > 0 \quad (k = 1, \ldots, p). \end{cases} \tag{1·1}$$

The quantity $n-(m+1)$ is termed the degree of difficulty of a geometric program and is a measure of the computational complexity of the dual program. The basic theory of geometric programming was developed by Duffin, Peterson & Zener (1967, Chapter 3). Ecker (1980) surveys the method, applications and solution algorithms for geometric programming.

One of the requirements for the application of geometric programming is that the constraints must be inequalities. The inequality ensures that for the problem a local minimum is a global minimum. If for example probabilities must sum to 1 $(p+q = 1)$ it does not matter if we search over a larger set of values $(p+q \leqslant 1)$ provided only that the optimum solution satisfies the equality. Modelling the replacement of equalities by inequalities requires some insight or failing that trial and error. From a well-known result in mathematical programming (Luenberger, 1973, p. 232), we know that at optimality every equality may be relaxed to an inequality of greater than or less than form according to the sign of Lagrange multiplier. In the geometric programming formulation, the λ_k's represent these multipliers. A constraint is binding if the multiplier λ_k is nonzero and if the constraint is inactive, the multiplier is zero.

An extension to geometric programming called signomial programming addresses the restriction that the coefficients c_j be positive. It has been shown by Duffin & Peterson (1972) that a program with negative coefficients may be transformed into a problem with posynomials but with constraints reversed from less than to greater than inequalities. The form is to minimize $g_0(t)$ subject to

$$g_k(t) \leqslant 1 \quad (k = 1, \ldots, p) \quad g_k(t) \geqslant 1 \quad (k = p+1, \ldots, r), \quad t_j \geqslant 0.$$

The last $r-p$ constraints are called reversed constraints. With this extension, while one can guarantee to improve any feasible solution, one can no longer guarantee that the local problem is a global optimum.

Computer codes for solving geometric programs with and without reversed constraints have been available for a number of years and are generally reliable. The algorithm used here was developed by Avriel, Dembo & Passy (1975). It relies on a log linear approximation of the posynomials and linear programming.

2. THE MAXIMUM LIKELIHOOD PROBLEMS

2·1. *Estimation of gene frequencies of blood groups*

The first example (Rao, 1973, p. 370) refers to the estimation of gene frequencies of blood groups for the ABO system. If the gene frequencies of A, B and O are denoted by p, q and r $(p+q+r = 1)$, then the probabilities of the four blood groups O, A, B and AB are respectively r^2, $p(p+2r)$, $q(q+2r)$ and $2pq$.

Let the observed relative frequency for the four classes O, A, B and AB be denoted by f_1, f_2, f_3 and f_4. The likelihood function is proportional to

$$\begin{aligned} L(p,q,r) &= (r^2)^{f_1}(p^2+2pr)^{f_2}(q^2+2qr)^{f_3}(2pq)^{f_4} \\ &= 2^{f_4} p^{f_2+f_4} q^{f_3+f_4}(1-p-q)^{2f_1}(2-p-2q)^{f_2}(2-q-2p)^{f_3}. \end{aligned}$$

The maximization of L with respect to p and q can be shown to be equivalent to the following geometric programming problem, if we set $t_1 = p$, $t_2 = q$ and $t_3 = 1-p-q = r$: minimize

$$t_1^{-(f_2+f_4)}\,t_2^{-(f_3+f_4)}\,t_3^{-2f_1}\,t_4^{-f_2}\,t_5^{-f_3},$$

subject to

$$t_1+t_2+t_3 \leqslant 1, \quad \tfrac{1}{2}t_1+t_2+\tfrac{1}{2}t_4 \leqslant 1, \quad t_1+\tfrac{1}{2}t_2+\tfrac{1}{2}t_5 \leqslant 1, \quad t_j > 0 \quad (j = 1,\ldots,5). \qquad (2{\cdot}1)$$

The dual program is to maximize

$$\prod_{i=1}^{10}\left(\frac{1}{\delta_i}\right)^{\delta_i}\prod_{k=1}^{3}\lambda_k^{\lambda_k},$$

where

$$\lambda_1 = \delta_2+\delta_3+\delta_4, \quad \lambda_2 = \delta_5+\delta_6+\delta_7, \quad \lambda_3 = \delta_8+\delta_9+\delta_{10}$$

subject to

$$\delta_i \geqslant 0 \quad (i = 1,\ldots,10), \quad \delta_1 = 1,$$

$$-(f_2+f_4)\,\delta_1+\delta_2+\delta_5+\delta_8 = 0, \quad -(f_3+f_4)\,\delta_1+\delta_3+\delta_6+\delta_9 = 0, \qquad (2{\cdot}2)$$

$$-2f_1\,\delta_1+\delta_4 = 0, \quad -f_2\,\delta_1+\delta_7 = 0, \quad -f_3\,\delta_1+\delta_{10} = 0. \qquad (2{\cdot}3)$$

The positivity of δ_4^*, δ_7^* and δ_{10}^* shows that $\lambda_k(\delta^*) > 0$ ($k = 1,2,3$). Thus the constraints (2·1) are active. This geometric programming has degree of difficulty 4. From (1·1) with (2·1), (2·2) and (2·3), the optimum δ_i^* should satisfy

$$\frac{\delta_2^*}{\delta_2^*+\delta_3^*+2f_1} = \frac{\delta_5^*}{\delta_5^*+\delta_6^*+f_2} = \frac{\delta_8^*}{\delta_8^*+\delta_9^*+f_3} = t_1^*,$$

$$\frac{\delta_3^*}{\delta_2^*+\delta_3^*+2f_1} = \frac{\delta_6^*}{\delta_5^*+\delta_6^*+f_2} = \frac{\delta_9^*}{\delta_8^*+\delta_9^*+f_3} = t_2^*,$$

$$\delta_2^*+\delta_5^*+\delta_8^* = f_2+f_4, \quad \delta_3^*+\delta_6^*+\delta_9^* = f_3+f_4.$$

These are six equations in six unknows δ_2^*, δ_3^*, δ_5^*, δ_6^*, δ_8^* and δ_9^*, which can now be solved in an iterative manner (Rijckaert & Martens, 1976), requiring only solutions of sets of simultaneous linear equations.

2·2. *Estimates in the case of compounded probabilities*

Fisher (1942) considered maximum likelihood estimation of p and p', where the likelihood function is

$$L(p,p') = p^a(1-p)^{A-a}p'^b(1-p')^{B-b}(pp')^c(1-pp')^{C-c}.$$

This is equivalent to the following geometric programming problem. Minimize

$$t_1^{-(a+c)}\,t_2^{-(b+c)}\,t_3^{-(A-a)}\,t_4^{-(B-b)}\,t_5^{-(C-c)},$$

subject to

$$t_1+t_3 \leqslant 1, \quad t_2+t_4 \leqslant 1, \quad t_1t_2+t_5 \leqslant 1, \quad t_j > 0 \quad (j = 1,\ldots,5).$$

The dual problem is to maximize

$$\prod_{i=1}^{7}\delta_i^{-\delta_i}\prod_{k=1}^{3}\lambda_k^{\lambda_k},$$

where

$$\lambda_1 = \delta_2+\delta_3, \quad \lambda_2 = \delta_4+\delta_5, \quad \lambda_3 = \delta_6+\delta_7,$$

subject to

$$\delta_1 = 1, \quad -(a+c)\,\delta_1+\delta_2+\delta_6 = 0, \quad -(b+c)\,\delta_1+\delta_4+\delta_6 = 0,$$

$$-(A-a)\,\delta_1+\delta_3 = 0, \quad -(B-b)\,\delta_1+\delta_5 = 0, \quad -(C-c)\,\delta_1+\delta_7 = 0, \quad \delta_i \geqslant 0 \quad (i = 2, \ldots, 7).$$

From (1·1) we now obtain

$$\left(\frac{\delta_2^*}{\delta_2^*+\delta_3^*}\right)\left(\frac{\delta_4^*}{\delta_4^*+\delta_5^*}\right) = \frac{\delta_6^*}{\delta_6^*+\delta_7^*} = t_1^*\,t_2^*.$$

Thus, with $\delta_6^* = c-\lambda$,

$$(a+\lambda)\,(b+\lambda)\,(C-\lambda) = (A+\lambda)\,(B+\lambda)\,(c-\lambda), \tag{2·4}$$

a quadratic equation in λ. Also, because $\delta_i \geqslant 0$, since

$$\lambda = c-\delta_6^* = \delta_2^*-a = \delta_4^*-b,$$

the root of (2·4), which lies between $\max(-a, -b)$ and c, should be chosen to compute the maximum likelihood estimates of p and p', which will be equal to t_1^* and t_2^*, respectively.

Equation (2·4) was derived by Fisher by differentiation and algebraic manipulation of the resulting equations. We have shown here how this equation follows directly from a routine application of the geometric programming procedure.

2·3. *Success probabilities based on observations on sums of k Bernoulli random variables*

This example has been recently considered by Samaniego & Jones (1981). Let X_i $(i = 1, \ldots, k)$ be independent Bernoulli random variables with potentially different probabilities of success, p_i $(i = 1, \ldots, k)$. Let $Y = X_1 + \ldots + X_k$, and assume that it is not possible to observe the individual X_i's, but that a random sample $Y_1, \ldots, Y_n$ is available. This model arises in a reliability context.

Let n_i denote the observed frequency of the event $Y = i$ for $i = 0, 1, \ldots, k$ in the random sample of size n. Maximization of the likelihood function with respect to p turns out to be computationally difficult when conventional optimization algorithms are used. Samaniego & Jones, therefore, pursue an approach which cannot be guaranteed to produce the maximum likelihood estimate for a fixed sample size, but which can be shown to produce this estimate with limiting probability one when the components of p are distinct. The solution is obtained by first finding the k roots $\hat{\theta}_{(i)}$ of the kth degree polynomial equation

$$p(x) = \sum_{i=0}^{k} (-1)^k n_i x^{k-1} = 0 \tag{2·5}$$

ordered from the smallest to the largest modulus, and setting $\hat{p}_{(i)} = \hat{\theta}_{(i)}/(1+\hat{\theta}_{(i)})$, where $p_{(1)}, \ldots, p_{(k)}$ represent the ordered parameter vector, $p_{(i)} \leqslant p_{(i+1)}$.

We now show that a geometric program with reversed constraints can be used to obtain numerical values of maximum likelihood estimates in this case even when (2·5) does not have positive real roots. The problem can be solved for any k and n, but our discussion is illustrated by the case $k = 3$. The likelihood function is then, with

$$q_j = 1-p_j,$$

$$L(n,p) = (q_1 q_2 q_3)^{n_0}(q_1 q_2 p_3+q_2 q_3 p_1+q_3 q_1 p_2)^{n_1} \times (q_1 p_2 p_3+q_2 p_3 p_1+q_3 p_1 p_2)^{n_2}(p_1 p_2 p_3)^{n_3}.$$

This has the following formulation using reversed constraints, a signomial program. Minimize

$$t_1^{-n_0} t_2^{-n_0} t_3^{-n_0} t_4^{-n_3} t_5^{-n_3} t_6^{-n_3} t_7^{-n_1} t_8^{-n_2}$$

subject to

$$t_1+t_4 \leqslant 1, \quad t_2+t_5 \leqslant 1, \quad t_3+t_6 \leqslant 1,$$

$$t_7^{-1} t_1 t_2 t_6+t_7^{-1} t_2 t_3 t_4+t_7^{-1} t_3 t_1 t_5 \geqslant 1, \quad t_8^{-1} t_1 t_5 t_6+t_8^{-1} t_2 t_6 t_4+t_8^{-1} t_3 t_4 t_5 \geqslant 1,$$

$$t_j > 0 \quad (j = 1,\dots,8).$$

Since the above program is not convex, it is not possible to guarantee that a unique minimum exists. However, if it possesses a unique minimizing solution, it can be found by applying any one of the available geometric programming codes.

A solution to the optimization problem is now illustrated with reference to the hypothetical set of observations $n_0 = 7$, $n_1 = 4$, $n_2 = 3$, $n_3 = 6$. Here, (2·5) reduces to $7x^3-4x^2+3x-6 = 0$. This has roots

$$x_1 = 1{\cdot}0, \quad x_2 = \frac{-3+\surd(-159)}{14}, \quad x_3 = \frac{-3-\surd(-159)}{14}.$$

Thus, the maximum likelihood estimates cannot be obtained here by using the procedure of Samaniego & Jones. The numerical solution is found by applying the algorithm of Avriel, Dembo & Passy (1975), which yields

$$\hat{p}_{(1)} = \hat{p}_{(2)} = \hat{p}_{(3)} = 0{\cdot}4167.$$

This was verified to be a global maximum for the maximum likelihood problem.

REFERENCES

AVRIEL, M., DEMBO, R. & PASSY, U. (1975). Solution of generalized geometric programs. *Int. J. Numer. Methods Engrg.* **9**, 149–69.

DUFFIN, R. J. & PETERSON, E. L. (1972). Reversed geometric programming treated by harmonic means. *Indiana Univ. Math. J.* **22**, 531–50.

DUFFIN, R. J., PETERSON, E. L. & ZENER, C. (1967). *Geometric Programming.* New York: Wiley.

ECKER, J. G. (1980). Geometric programming: Methods, computations and applications. *S.I.A.M. Rev.* **22**, 338–62.

FISHER, R. A. (1942). The likelihood solution of a problem in compounded probabilities. *Ann. Eugen.* **9**, 306–7.

LUENBERGER, D. G. (1973). *Introduction to Linear and Non-linear Programming.* Reading, Mass: Addison-Wesley.

RAO, C. R., (1973). *Linear Statistical Inference and its Applications*, 2nd edition. New York: Wiley.

RIJCKAERT, M. J. & MARTENS, X. M. (1976). A condensation method for generalized geometric programming. *Math. Programming* **11**, 89–93.

SAMANIEGO, F. J. & JONES, L. E. (1981). Maximum likelihood estimation for a class of multinomial distributions arising in reliability. *J. R. Statist. Soc.* B **43**, 46–52.

[*Received December* 1981. *Revised July* 1982]