© Journal of Applied Mathematics & Decision Sciences, 1(1), 53–65 (1997)
Reprints Available directly from the Editor. Printed in New Zealand.

Maximum Likelihood Estimates with Order Restrictions on Probabilities and Odds Ratios: A Geometric Programming Approach

D. L. BRICKER
Dept. of Industrial Engineering, The University of Iowa, Iowa City, IA 52242, USA

K. O. KORTANEK
Dept. of Management Science, The University of Iowa, Iowa City, IA 52242, USA

L. XU
Milliman and Robertson, Actuaries and Consultants, St. Louis, MO, USA

Abstract. The problem of assigning cell probabilities to maximize a multinomial likelihood with order restrictions on the probabilies and/or restrictions on the local odds ratios is modeled as a posynomial geometric program (GP), a class of nonlinear optimization problems with a well-developed duality theory and collection of algorithms. (Local odds ratios provide a measure of association between categorical random variables.) A constrained multinomial MLE example from the literature is solved, and the quality of the solution is compared with that obtained by the iterative method of El Barmi and Dykstra, which is based upon Fenchel duality. Exploiting the proximity of the GP model of MLE problems to linear programming (LP) problems, we also describe as an alternative, in the absence of special-purpose GP software, an easily implemented successive LP approximation method for solving this class of MLE problems using one of the readily available LP solvers.

Keywords: Maximum Likelihood Estimates(MLE), Local Odds Ratios, Geometric Programming(GP).

1. Introduction

Suppose that we have observed frequencies $(f_1, f_2, ..., f_r)$ of a discrete random variable Y which we assume to have multinomial distribution with parameters n and $p = (p_1, p_2, ..., p_r)$. In estimating p, it would be appropriate to use any prior knowledge of the relationships among the probabilities. For example, one might require an estimating distribution for which $p_{i+1} \leq p_i$, i.e., the probabilities are nonincreasing (or more generally, $p_{i+1} \leq \alpha_i p_i$) or else $p_{i+1} \geq p_i$, i.e., the probabilities are nondecreasing(or more generally, $p_{i+1} \geq \alpha_i p_i$ for positive α_i). Hence, finding the maximum likelihood estimator will require the solution of

$$\max_{p} \prod_{i=1}^{r} p_i^{f_i} \tag{1.1}$$

subject to either the restrictions

$$\frac{p_{i+1}}{\alpha_i p_i} \leq 1 \;\; i = 1, 2, ..., r-1 \tag{1.2}$$

in the case of nonincreasing probabilities, or, in the case of nondecreasing probabilities, the restriction

$$\frac{\alpha_i p_i}{p_{i+1}} \leq 1 \;\; i = 1, 2, ..., r-1 \tag{1.3}$$

and, of course,

$$\sum_{i=1}^{r} p_i \;=\; 1 \tag{1.4}$$

$$p_i \;\geq\; 0 \;\; i = 1, 2, ..., r \tag{1.5}$$

A special case of the constraints arise when all the $\alpha_i' s$ are one. This important case is solved by Robertson, Wright, and Dykstra [14]. Among early appearances of the more general constraints with non–unitary $\alpha_i' s$ was the thesis of Willy Gochet, [11], where the purpose was to identify these problems as geometric programs [10].

In this paper we review a simple modification for expressing the above and other more complex MLE problems with constraints on odds ratios as posynomial geometric programs. Section 2 of this paper summarizes some of the properties of geometric programming (GP), and Section 3 explains how these MLE problems can be formulated as a GP problem, for which special-purpose optimization software is available. Section 4 describes the proximity of the GP model to a log-linear system, and explains how, by approximating the sum of the probabilities by a monomial, complete log-linearity is achieved, so that, in the absence of specialized GP software, the problem might be solved using commonly-available linear programming (LP) software. Section 5 presents a small example to illustrate the LP approach described in Section 4. In Section 6, we consider maximum likelihood estimation of joint probability distributions of two or three discrete ordinal random variables, when it is possible to specify positive or negative association between two random variables, or local 3-factor interaction in the case of three random variables. Section 7 solves a GP model of a MLE problem from the literature, using a state-of-the-art GP solver. The final section presents a summary of the paper.

2. Geometric Programming Model

The general primal problem of geometric programming (GP) (Duffin et al.[10]) is to

$$\begin{aligned} &\text{Minimize} \;\; g_0(x) \\ &\text{subject to} \;\; g_i(x) \leq 1, i = 1, 2, ..., m \;, \\ &x > 0 \end{aligned} \tag{2.1}$$

where $x \epsilon R^m$ and each function $g_i(x)$ is a *posynomial* with T_i terms, i.e.,

$$g_i(x) = \sum_{j=1}^{T_i} c_{ij} \prod_{k=1}^{N} x_k^{a_{ijk}} \tag{2.2}$$

The exponents a_{ijk} are arbitrary real numbers, but the coefficients c_{ij} are assumed to be positive constants and the decision variables x_n are required to be strictly positive.

The corresponding posynomial GP dual problem is to

$$\text{Maximize} \quad v(\delta, \lambda) = \prod_{i=0}^{m} \prod_{j=1}^{T_i} (\frac{c_{ij}\lambda_i}{\delta_{ij}})^{\delta_{ij}} \tag{2.3}$$

$$\text{subject to} \quad \sum_{i=0}^{m} \sum_{j=1}^{T_i} a_{ijk}\delta_{ij} = 0 \ , \quad k = 1, 2, ..., N \tag{2.4}$$

$$\lambda_i = \sum_{j=1}^{T_i} \delta_{ij} \ , \quad i = 1, 2, ..., m \tag{2.5}$$

$$\lambda_0 = 1 \tag{2.6}$$

$$\delta_{ij} \geq 0, \quad j = 1, 2, ..., T_i, \quad i = 0, 1, 2, ..., m \tag{2.7}$$

Note that there are two sets of dual variables: dual variable λ_i corresponds to posynomial i (i.e., the primal objective if index $i = 0$, or constraint if $i \geq 1$); dual variable δ_{ij} is associated with term j of posynomial i. This dual problem offers several computational advantages: after using (2.5) to eliminate λ, the logarithm of the objective (2.3) is a concave function to be maximized over a linear system. This linear system has T variables, where $T = T_0 + T_1 + ... + T_m$, and (N+1) equations, and hence the difference T-(N+1) is referred to as its *degree of difficulty*. In case the degree of difficulty is negative, i.e., $T < N + 1$, the dual feasible region is overdetermined and may be empty; this "pathological" case rarely occurs for well-formulated problems. (Cf. Bricker et al., [7].) The GP dual problem does possess some undesirable properties which make it not amenable to solution by a general purpose nonlinear optimization algorithm, most notably the fact that the objective function is nondifferentiable when one or more variables are zero.)

If an optimal dual solution (δ^*, λ^*) is known, then the following relationships may be used to compute a primal solution x^* under suitable conditions, [10], Theorem 1, p80:

$$\delta_{ij}^* g_i(x^*) = \lambda_i^* c_{ij} \prod_{n=1}^{N} x_n^{*a_{ijn}}, \quad j = 1, 2, ..., T_i, \quad i = 0, 1, ..., m \tag{2.8}$$

where $g_0(x^*) = v(\delta^*, \lambda^*)$ and, for $i > 0$, $g_i(x^*) = 1$ if $\lambda_i^* \neq 0$. Note that, from these relationships, one may obtain a system of equations linear in the logarithms of the optimal values of the primal variables:

$$\sum_{k=1}^{N} a_{ijk} \ln x_k = \ln\left(\frac{\delta_{ij}^* g_i(x^*)}{\lambda_i^* c_{ij}}\right), \quad j = 1, 2, ..., T_j, \tag{2.9}$$
$$\text{for each } i = 0, 1, ..., m, \text{ such that } \lambda_i^* \neq 0$$

Typically, but not always, the system of linear equations (2.9) uniquely determines the optimal x^*. See Dembo [8] for a discussion of the recovery of primal solutions from the dual solution in general.

3. Formulation of the Constrained MLE Problem as a GP

Replace the maximization objective (1.1) with the equivalent objective to minimize its reciprocal,

$$\min_{p} \prod_{i=1}^{r} p_i^{-f_i} \tag{3.1}$$

and, since the GP primal constraints must be inequalities, replace the equation (1.4) by the relaxation

$$\sum_{i=1}^{r} p_i \leq 1. \tag{3.2}$$

(Because the objective (3.1) is strictly nonincreasing, the inequality constraint (3.2) will certainly be "tight" at an optimal p.) And so (3.1), (3.2),(1.5), and either ordering constraint (1.2) or (1.3) constitute an instance of a posynomial geometric programming problem with $N = r$ variables and $T = 2r$ terms, having $T - (N + 1) = 2r - (r + 1) = r - 1$ degrees of difficulty. The utility of geometric programming (GP) for solving another particular MLE problem (an overlapping sample frame problem) was long ago noted by Alldredge and Armstrong [2] (cf. also Mazumdar and Jefferson [13]). Wong [15] has used the GP model to establish the relationship between the maximum likelihood estimation problem and the entropy maximization formulation of the trip distribution problem in traffic studies. To the best of our knowledge, however, no discussion of the use of GP models for MLE with restrictions on odds ratios (cf. Section 6).

4. Approximation by a Log-Linear System

Of course, if it were not for constraint (3.2), the above MLE problem would be log-linear, i.e.,

$$\min_{p} \left[-\sum_{i=1}^{r} f_i \ln p_i\right] \tag{4.1}$$

subject to

$$\ln p_{i+1} \ - \ \ln p_i \ \le \ -\ln \alpha_i \ , \ \ i = 1, 2, ..., r-1 \tag{4.2}$$

which, after a logarithmic transformation $z_i = \ln p_i$ is a *linear programming* problem (in which the variables are unrestricted in sign.) Unfortunately, because (3.2) contains multiple terms, the logarithmic transformation of (3.2) would yield a nonseparable and nonlinear constraint.

It is possible, however, to approximate (3.2) with a monomial constraint by means of the *arithmetic-geometric mean inequality* [10], p110. This inequality might be stated in the form:

Given $\mu_i > 0$*,i=1,2,...,r, such that* $\sum_{i=1}^{r} \mu_i \ = \ 1$*, then for all* $p_i > 0$*,*

$$\sum_{i=1}^{r} p_i \ge \prod_{i=1}^{r} (\frac{p_i}{\mu_i})^{\mu_i} \tag{4.3}$$

with equality if and only if

$$\frac{p_1}{\mu_1} = \frac{p_2}{\mu_2} = ... = \frac{p_r}{\mu_r}. \tag{4.4}$$

(For example,

$$\frac{a}{2} + \frac{b}{2} \ge a^{\frac{1}{2}} b^{\frac{1}{2}} \tag{4.5}$$

i.e., the arithmetic mean of two positive numbers is at least as great as their geometric mean, with equality if and only if a=b.)

Therefore, the monomial constraint

$$\prod_{i=1}^{r} (\frac{p_i}{\mu_i})^{\mu_i} \le 1 \tag{4.6}$$

is a relaxation of constraint (3.2), i.e., p_i satisfying (4.6) have a sum exceeding 1, unless, of course, equality is attained in (4.3). The conditions for this equality occurring for our choice of μ is, according to (4.4), that μ be identical to the distribution p. The monomial constraint (4.6) is equivalent to the log-linear constraint

$$\sum_{i=1}^{r} \mu_i \ln p_i \le \sum_{i=1}^{r} \mu_i \ln \mu_i \tag{4.7}$$

so that (4.1), (4.2) and (4.7) constitute a linear programming problem in r variables (ln p_i) and r inequality constraints, and is easily solved by widely available LP software. This suggests an iterative method, *successive linear programming*, in which the initial LP is constructed using μ equal to the relative empirical frequency vector. After solving and recovering an estimate of the optimal vector p by the exponential transformation, the feasibility of constraint (3.2) is tested; if the sum of the estimated "probabilities" exceeds one, another linear approximation (4.6) is constructed using μ equal to the current (normalized) approximation of p. (Using the arithmetic/geometric mean inequality to approximate a multiterm generalized polynomial by a monomial function is commonly referred to as *condensation* (Duffin [9]). Several popular geometric programming algorithms using condensation in essence follow just the iterative approach outlined above!)

5. Example I

One hundred observations of the ordinal random variable Y were measured, with Y=1, 2, 3, and 4 with frequencies 32, 35, 15, and 18, respectively. We wish to find the MLE of the distribution of Y, subject to the restriction that $P\{Y = i+1\} \leq P\{Y = i\}$, i=1,2,3, i.e., probabilities are nonincreasing. This problem may be formulated as the GP problem in which we wish to minimize the reciprocal of the likelihood function, namely

$$\text{Minimize } p_i^{-.32} p_2^{-.35} p_3^{-.15} p_4^{-.18} \tag{5.1}$$

subject to

$$0 \leq p_4 \leq p_3 \leq p_2 \leq p_1 \tag{5.2}$$

i.e.,

$$p_1^{-1} p_2 \leq 1, \;\; p_2^{-1} p_3 \leq 1, \;\; p_3^{-1} p_4 \leq 1 \tag{5.3}$$

and

$$p_1 + p_2 + p_3 + p_4 \leq 1 \tag{5.4}$$

as well as nonnegativity restrictions on the probabilities.

Letting μ initially be the vector of relative frequencies, i.e., $\mu = [0.32, 0.35, 0.15, 0.18]$, constraint (5.4) is approximated, according to (4.6), by

$$3.763270635 p_1^{0.32} p_2^{0.35} p_3^{0.15} p_4^{0.18} \leq 1 \tag{5.5}$$

We adjoin the constraints

$$p_i \leq 1 \;,\; i = 1, 2, 3, 4 \tag{5.6}$$

in order to guarantee boundedness of the solution. (Constraints (5.6) are no longer redundant when we relax (5.4).) Thus we obtain the following linear program (after employing the logarithmic transformation $z_i = \ln p_i$):

$$\begin{array}{lll} \text{Minimize} & -.32z_1 - .35z_2 - .15z_3 - .18z_4 & \\ \text{subject to} & -z_1 + z_2 & \leq \; 0 \\ & \quad -z_2 + z_3 & \leq \; 0 \\ & \qquad -z_3 + z_4 & \leq \; 0 \\ & 0.32z_1 + 0.35z_2 + 0.15z_3 + 0.18z_4 & \leq \; -\ln 3.763270635 \\ & z_i \leq \; 0 \;\; for \;\; i = 1, 2, 3, 4 & \end{array} \tag{5.7}$$

where the nonpositivity constraints derive from $p_i \leq 1$. (Since LP codes permit (and in many instances require!) nonnegativity constraints to be handled implicitly, we would be well-advised to perform still another variable transformation, $z_i' = -z_i$.) The solution of this LP (using the exponential transformation) provides us the estimate $p_1 = p_2 = p_3 = p_4 = 0.2657263049$. Obviously this solution is infeasible in (5.4), i.e., the sum of the probabilities is 1.062905219, and so we compute another monomial approximation of (5.4), using μ equal to the normalized value of p just computed, i.e., $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0.25$, i.e., the uniform distribution. According to (4.6), this yields the linear constraint

$$0.25z_1 + 0.25z_2 + 0.25z_3 + 0.25z_4 \leq -\ln 4 = -1.325288429 \tag{5.8}$$

The previous LP with this added constraint has the solution $p_1 = 0.4806291165, p_2 = p_3 = p_4 = 0.2010558216$, which sum to 1.083796581, and so yet another linear constraint may be constructed, using $\mu_1 = \frac{0.480629}{1.08379} = 0.443468$ and $\mu_2 = \mu_3 = \mu_4 = \frac{0.201056}{1.08379} = 0.185511$, and added to the LP. The sum of p_i which solves this next LP ($p_1 = 0.3518276341, p_2 = 0.2611114982$, $p_3 = p_4 = .2062060617$) is reduced to 1.025352. Continuing in this manner, the violation in constraint (5.4) is reduced to 0.000267823 by the tenth LP solution, and less than 10^{-10} by the twentieth LP solution ($p_1 = p_2 = 0.334997511, p_3 = p_4 = 0.165000249$), very near to the optimal solution $p_1^* = p_2^* = 0.3350000000$, $p_3^* = p_4^* = 0.1650000000$.

The solution p^* obtained above is the same as the maximum likelihood estimator restricted to antitonic functions on Y (i.e. satisfying the constraints $p_1 \geq p_2 \geq p_3 \geq p_4$) as described in Section 2 of Chapter 7 in Robertson, Wright, and Dykstra [14], 1988. Robertson et al. demonstrate that, geometrically, p_i^* is the slope of the linear segment to the left of y_i in the least concave majorant of F_4 (where $F_4(y)$ is the Empirical Distribution Function (EDF) of p). Thus, the specific MLE example problem (with $\alpha = 1.0$) considered in this section might be solved more easily without the use of a mathematical programming model. Our introduction of this GP model here was to provide a simple illustration of the successive LP approach, which is applicable to MLE with more general constraints, e.g. order restrictions with $\alpha < 1$, or the MLE problems in the following section requiring estimates of joint distributions with restrictions on local odds ratios. The above successive LP approach (which may also be referred to as a cutting-plane approach) has the advantage that it utilizes widely-available LP software; as one might expect, however, it suffers from the fact that the added linear constraints become nearly colinear as the algorithm progresses, limiting the level of precision which is attainable. The example MLE problem in Section 7, with restrictions on odds ratios, which will be introduced in the next section, will be solved by a much superior interior-point algorithm which does not exhibit this undesirable behavior.

6. Cross Classifications

Consider two ordinal discrete random variables Y_1 and Y_2, and an $r \times c$ rectangular array in which the $(i,j)^{th}$ cell contains the relative frequency f_{ij} with which the event " Y_1 = i and Y_2 = j" was observed. Let p_{ij} be the true probability of this event. The *local odds ratios*

$$\theta_{ij} = \frac{p_{ij}p_{i+1,j+1}}{p_{i+1,j}p_{i,j+1}} \tag{6.1}$$

can be used to specify properties of association between Y_1 and Y_2. For example, θ_{ij}=1 for all i and j implies independence, while $\theta_{ij} \geq (\leq)1$ implies a positive (negative) association between the two random variables (cf. Agresti [1]).

Just as the order restriction (1.2) discussed earlier may be expressed as a log-linear constraint, so also the constraints

$$\frac{p_{ij}p_{i+1,j+1}}{p_{i+1,j}p_{i,j+1}} \leq \alpha_{ij} \tag{6.2}$$

and

$$\frac{p_{ij}p_{i+1,j+1}}{p_{i+1,j}p_{i,j+1}} \geq \alpha_{ij} \text{ or equivalently , } \frac{p_{i+1,j}p_{i,j+1}}{\alpha_{ij}p_{ij}p_{i+1,j+1}} \leq 1 \tag{6.3}$$

where $\alpha_{ij} > 0$, may both be expressed as log-linear constraints. Thus, the maximum likelihood estimation problem

$$\text{Maximize} \prod_{i=1}^{r} \prod_{j=1}^{c} p_{ij}^{f_{ij}} \tag{6.4}$$

subject to $p_{ij} \geq 0$, either (6.2) or (6.3) and

$$\sum_{i=1}^{r} \sum_{j=1}^{c} p_{ij} = 1 \tag{6.5}$$

might be solved by the successive linear programming approach which was applied above. El Barmi and Dykstra [6] (cf. also [5]) have also developed a convergent iterative algorithm for this problem, based upon Fenchel duality theory, for the case $\alpha = 1$.

In order to formulate this problem as a geometric program, the equality (6.5) must be relaxed to the inequality

$$\sum_{i=1}^{r} \sum_{j=1}^{c} p_{ij} \leq 1 \tag{6.6}$$

which will surely remain "tight" at the optimum, and the objective restated as a minimization (of the reciprocal of the likelihood):

$$\min_{p} \prod_{i=1}^{r} \prod_{j=1}^{c} p_{ij}^{-f_{ij}} \tag{6.7}$$

So in this case also, the restricted maximum likelihood estimator is the optimum of a posynomial geometric programming problem.

Although an order restriction on marginal probabilities, e.g., $\sum_j p_{i-1,j} \leq \sum_j p_{ij}$, cannot be expressed as a posynomial GP constraint, by applying condensation to the denominator of the equivalent constraint

$$\frac{\sum_j p_{i-1,j}}{\sum_j p_{ij}} \leq 1$$

one does obtain a posynomial GP constraint. Hence, a MLE with order restrictions on the marginal distributions might be computed by solving a succession of approximating GP problems. (Cf. Avriel et al. [4]).

Consider further the case of a *three-dimensional* $r \times c \times l$ array where the third dimension (level) corresponds to a third ordinal factor, and let p_{ijk} denote the corresponding joint probability. Within each level k $(1 \leq k \leq l)$ we define the odds ratios

$$\theta_{ij(k)} = \frac{p_{ijk} p_{i+1,j+1,k}}{p_{i+1,j,k} p_{i,j+1,k}} \tag{6.8}$$

The *ratio of odds ratios*

$$\theta_{ijk} = \frac{\theta_{ij(k+1)}}{\theta_{ij(k)}} \tag{6.9}$$

may be used to specify local three-factor interaction: a value of 1.0 for all i, j, and k specifies no such interaction, while $\theta_{ijk} \geq 1$, i.e., $\theta_{ij(k+1)} \geq \theta_{ij(k)}$, for all i, j, and k, specifies nondecreasing local conditional association between Y_1 and Y_2 as Y_3 increases. (In a similar fashion, nonincreasing local conditional association may be specified.) In the next section, we present an example estimation problem illustrating this type of order restriction on the odds ratios.

7. Example II

We illustrate the geometric programming approach using data found in J. R. Ashford and R. Snowden [3]. Medical examinations were performed on a sample of coalminers in the United Kingdom. Each of 18,000 subjects was classified according to age (the third factor), as well as to whether he exhibits each of two symptoms, namely "breathlessness" and "wheeze" (factors 1 and 2, respectively). The frequencies f_{ijk} of these classifications are presented in Table 7.1.

It is hypothesized that the two factors "breathlessness" and "wheeze" have a positive association within each age level, i.e., $\theta_{ij(k)} \geq 1$ but, however, that this positive association become weaker as the third factor, age, increases, i.e.,

$$\theta_{11(k)} \geq \theta_{11(k+1)} \text{ for k=1,...8} \tag{7.1}$$

Table 7.1. Coalminers Classified by Breathlessness, Wheeze, and Age

	Breathlessness			
	Yes		No	
	Wheeze		Wheeze	
age	yes	no	yes	no
20-24	9	7	95	1841
25-29	23	9	105	1654
30-34	54	19	177	1863
35-39	121	48	257	2357
40-44	169	54	273	1778
45-49	269	88	324	1712
50-54	404	117	245	1324
55-59	406	152	225	967
60-64	372	106	132	526

Note that, since the first and second factors have only a single pair of levels each, only a single odds ratio $\theta_{11(k)}$ is defined for level k of the third factor (age).) The maximum likelihood estimator is to be computed so as to satisfy this hypothesis, that is, we wish to

$$\min_{p} \prod_{i=1}^{2} \prod_{j=1}^{2} \prod_{k=1}^{9} p_{ijk}^{-f_{ijk}} \tag{7.2}$$

subject to

$$\sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{9} p_{ijk} \; = \; 1, \tag{7.3}$$

and

$$1 \leq \frac{p_{1,1,k+1}p_{2,2,k+1}}{p_{2,1,k+1}p_{1,2,k+1}} \leq \frac{p_{11k}p_{22k}}{p_{21k}p_{12k}} \; , \; k = 1, 2, ..., 8 \tag{7.4}$$

i.e.,

$$p_{119}^{-1}p_{229}^{-1}p_{219}p_{129} \leq 1 \tag{7.5}$$

$$p_{11k}^{-1}p_{22k}^{-1}p_{21k}p_{12k}p_{1,1,k+1}p_{2,2,k+1}p_{2,1,k+1}^{-1}p_{1,2,k+1}^{-1} \leq 1 \;\; k = 1, 2, ..., 8 \tag{7.6}$$

and $p_{ijk} \geq 0$ for all i, j, and k. This geometric programming model has $N = 36$ primal variables and 10 constraints (7.3), (7.5) and (7.6). With a total of $T = 46$ terms in the objective and constraints, the dual geometric program has 46 dual variables δ_{ij} with $N+1 = 37$ constraints, so that the degree of difficulty is $T - (N+1) = 9$.

The solution was easily found using a code written by Xiaojie Xu, an implimentation of the algorithm described in Kortanek, Xu, and Ye [12]. The optimal joint probabilities found by this algorithm are shown in Table 7.2.

Table 7.2. Joint Probabilities (percent) by Breathlessness and Wheeze by Age, provided by GP

	Breathlessness			
age	Yes		No	
	Wheeze		Wheeze	
	yes	no	yes	no
20-24	0.51811	0.30156	4.80976	94.37056
25-29	1.22203	0.56468	5.92482	92.28847
30-34	2.55561	0.89919	8.37672	88.16848
35-39	4.34783	1.72476	9.23464	84.69278
40-44	7.44822	2.37990	12.03173	78.14015
45-49	11.41490	3.50361	13.36571	71.71578
50-54	19.13117	5.79706	11.92146	63.15031
55-59	23.51403	8.37169	12.54312	55.57117
60-64	32.26272	9.81474	12.10348	45.81906

For example, the probability that a coalminer between 20-24 years of age exhibits both wheeze and breathlessness is 0.51811 percent, while for a coalminer in the next age category (25-29 years), this probability has increased to 1.22203 percent. The logarithm of the objective function (7.2) at this solution is 1.28427×10^4. On the other hand, the solution provided by El Barmi and Dykstra [6], shown in Table 7.3, evaluated by the same objective function (7.2), has a somewhat larger value, 1.31267×10^4. (Recall that the objective function is the reciprocal of the likelihood, and is to be *minimized.*)

Table 7.3. Joint Probabilities (percent) of Breathlessness and Wheeze by Age from El Barmi and Dykstra

	Breathlessness			
age	Yes		No	
	Wheeze		Wheeze	
	yes	no	yes	no
20-24	0.51793	0.30174	4.80945	94.37087
25-29	1.22222	0.56449	5.92462	92.28867
30-34	2.55561	0.89920	8.37672	88.16848
35-39	4.34783	1.72476	9.23464	84.69278
40-44	7.43184	2.37467	12.00528	78.18821
45-49	11.41286	13.36319	3.50291	71.72104
50-54	19.13110	5.79713	11.92153	63.15024
55-59	23.51429	8.27143	12.54286	55.57143
60-64	32.26232	9.81514	12.10387	45.81866

Table 7.4 shows the local odds ratios evaluated for the two solutions.

Table 7.4. Local Odds Ratios

Age	GP Solution	El Barmi/Dykstra Solution
20-24	33.70939	33.68048
25-29	33.70939	33.72735
30-34	29.91436	29.91436
35-39	23.11908	23.11908
40-44	20.32540	20.38272
45-49	17.48152	17.48647
50-54	17.48152	17.48115
55-59	12.44395	12.44478
60-64	12.44395	12.44277

Note that, for both estimates, the odds ratios all exceed 1.0 as required, but that in the case of the solution given by El Barmi and Dykstra, the odds ratio for the age category 25-29 exceeds slightly that for age category 20-24, violating the restriction that the association be nonincreasing with respect to age. The geometric programming solution, on the other hand, satisfies the restrictions that the positive association be weaker as the age increases, and has an objective value which is approximately 2.2 percent better than the solution of El Barmi and Dykstra. Aside from the (perhaps insignificant) improvement in the quality of the solution, an advantage to using GP instead of the El Barmi and Dykstra approach is the availability of software for solving posynomial GP problems.

8. Summary

It has long been known that the likelihood function fits well into the posynomial geometric programming framework. The purpose of this paper has been to review the applicability of GP to MLE problems in which restrictions have been placed on the relative magnitudes of the probability estimates and also to demonstrate that restrictions on the odds ratios fit well into the GP model. By doing so, one is able to utilize GP software for the computation.

The fact that only one constraint of these GP models has multiple terms (viz., the restriction that the sum of the probabilities cannot exceed 1) makes them especially amenable to solution by a successive LP method, since after a monomial approximation to that constraint, the problem becomes log-linear. While inferior to state-of-the-art geometric programming algorithms, this approach may give satisfactory estimates if high precision is not required.

References

1. A. Agresti. A survey of strategies for modeling cross classifications having ordinal variables. *Journal of the American Statistical Association*, 78(3):184–198, 1983.
2. J.R. Alldredge and D.W. Armstrong. Maximum likelihood estimation for multinomial distribution using geometric programming. *Technometrics*, 16(4):585–587, 1974.

3. J.R. Ashford and R.D. Snowden. Multi-variate probit analysis. *Biometrics*, 26:535–546, 1970.
4. M. Avriel, R. Dembo, and U. Passy. Solution of generalized geometric programs. *International Journal for Numerical Methods in Engineering*, 9:149–168, 1975.
5. H. El Barmi and R.L. Dykstra. Restricted multinomial maximum likelihood estimation based upon fenchel duality. *Statistics and Probability Letters*, 21:121–130, 1994.
6. H. El Barmi and R.L. Dykstra. Maximum likelihood estimates via duality for log-convex models when cell probabilities are subject to convex constraints. The University of Iowa, Dept. of Statistics and Actuarial Science, Technical Report No. 265, 1997.
7. D. Bricker, J.C. Choi, and J. Rajgopal. On geometric programming problems having negative degree of difficulty. *European Journal of Operational Research*, 63:102–113, 1993.
8. R.S. Dembo. Dual to primal conversion in geometric programming. *Advances in Geometric Programming*, pages 333–342, 1990.
9. R.J. Duffin. Linearizing geometric programs. *SIAM Review*, 21:1–227, 1970.
10. R.J. Duffin, E.L. Peterson, and C.L. Zener. *Geometric Programming – Theory and Applications*. John Wiley and Sons, New Year, 1967.
11. W. Gochet. *Computational Treatment of Some Linear Programming Optimization Problems with Application to Geometric and Probabilistic Programming*. PhD thesis, Carnegie–Mellon University, Graduate School of Industrial Administration, Pittsburgh, Pennsylvania, 1973.
12. K.O. Kortanek, X. Xu, and Y. Ye. An infeasible interior–point algorithm for solving primal and dual geometric programs. *Mathematical Programming*, 76:155–182, 1997.
13. M. Mazumdar and T. R. Jefferson. Maximum likelihood estimates for multinomial probabilities via geometric programming. *Biometrika*, 70(1):257–261, 1983.
14. T. Robertson, F.T. Wright, and R.L. Dykstra. *Order Restricted Statistical Inference*. John Wiley and Sons, New York, 1988.
15. Danny S. Wong. Maximum likelihood, entropy maximization, and the geometric programming approaches to the calibration of trip distribution models. *Transportation Research–B*, 15B(5):329–343, 1981.