RUTCOR
RESEARCH
REPORT

Estimation of Cause-Effect Relationship under Noise

András Prékopa[a]

RRR 3-93, March 1993

RUTCOR • Rutgers Center for Operations Research • Rutgers University • P.O. Box 5062 • New Brunswick New Jersey • 08903-5062
Telephone: 908-932-3804
Telefax: 908-932-5472
Email: rrr@rutcor.rutgers.edu

[a]RUTCOR, Rutgers University, New Brunswick, New Jersey 08903

Rutcor Research Report
RRR 3-93, March 1993

Estimation of Cause-Effect Relationship under Noise

András Prékopa

Abstract. Events, that occur subsequently or simultaneously, cause some other event as effect. The latter can be observed with noise and the problem is to estimate the weights of the causes in the realization of the effect.

Acknowledgements: Research supported by the Air Force Office of Scientific Research, Grant numbers: AFOSR-89-0512B and F49620-93-1-0041.

1 Introduction

The paper deals with the following problem. Suppose that n events occur simultaneously or subsequently and their occurencies are represented by the realizations of the Boolean variable n-tuple: $x_1, ..., x_n$. If at a time the ith event occurs, then $x_i = 1$ in that realization. The occurence of the ith event gives rise to a secondary action that has a magnitude equal to a_i. The total magnitude of the secondary action eqals $a_1x_1 + \cdots + a_nx_n$. The numbers $a_1, ..., a_n$, however, are unknown.

Suppose we are given a further variable y such that one realization of y corresponds to one realization of the variables $x_1, ..., x_n$. We may not be able to measure the value of y but we can observe if $y > y_0$ or $y \leq y_0$, where y_0 is some known threshold value. We assume that the following relationship exist

$$y = a_1x_1 + \cdots + a_nx_n + u, \tag{1}$$

where u is a random variable that we term noise.

A sample of size N is taken, i.e., we have the $N \times (n+1)$ array of $0, 1$ values:

$$\begin{array}{lll} x_{11}, ..., x_{n1} & (1 \text{ if } y_1 > y_0, & 0 \text{ if } y_1 \leq y_0) \\ \cdots & \cdots & \cdots \\ x_{1N}, ..., x_{nN} & (1 \text{ if } y_N > y_0, & 0 \text{ if } y_N \leq y_0) \end{array}$$

and our aim is to estimate the effect of the event $x_i = 1$, relative to the occurence of the inequality $y > y_0$. This is equivalent to estimate $a_1, ..., a_n$.

The present paper was inspired by the paper of Crama, Hammer and Ibaraki (1988) in which a nonstochastic cause-effect relationship problem is formulated. In everyday terms the problem is the following: somebody registers each day which foods consumed and if he or she had headache; if there are altogether n foods then the food consumption is characterized by the Boolean variables $x_1, ..., x_n$ so that $x_i = 1$ if the ith food has been consumed on the given day, otherwise $x_i = 0$. One further Boolean variable equals 1 or 0, if the person had headache on that day or not. The problem is to estimate the weight of the event $x_i = 1$ in creating headache.

The difference between the problem of Crama, Hammer and Ibaraki and the problem of this paper is that here we attribute the "headache" to some additive effects of the foods consumed and assumed the presence of some noise. If this value surpasses the threshold y_0 then "headache" in fact occurs, otherwise it does not.

2 Formulation of the Estimation Problem

We may assume that there exists a known value d such that $a_1 + \cdots + a_n \leq d$. There may also exists some other known relationships regarding the numbers $a_1, ..., a_n$ e.g., we may be informed that $a_1 \geq a_2 \geq \cdots \geq a_n$. In the original problem we do not assume that $a_1 \geq 0, ..., a_n \geq 0$. However, since $a_1, ..., a_n$ will be decision variables in a linear programming problem and, using a well-known technique, we can replace each variable that is not restricted by nonnegativity constrainet by the difference of two nonnegative variables, we may assume that all $a_1, ..., a_n$ are nonnegative. If all restrictions for $a_1, ..., a_n$ are linear then we may assume that these are of the form

$$\begin{aligned} Aa &= b \\ a &\geq 0, \end{aligned} \tag{2}$$

where $a = (a_1, ..., a_n)^T$, A is an $m \times n$ matrix and b is an m-component vector. In this formulation some of the a_i are slack variables, i.e., they are introduced simply to convert inequalities into equalities.

Considering the random variable u, let $F(z)$ designate its probability distribution function:

$$F(z) = P(u \leq z), \quad -\infty < z < \infty.$$

Using this, we may write

$$\begin{aligned} P(y > y_0) &= P(u > y_0 - a_1x_1 - \cdots - a_nx_n) \\ &= 1 - F(y_0 - a_1x_1 - \cdots - a_nx_n), \end{aligned} \tag{3}$$

$$\begin{aligned} P(y \leq y_0) &= P(u \leq y_0 - a_1x_1 - \cdots - a_nx_n) \\ &= F(y_0 - a_1x_1 - \cdots - a_nx_n). \end{aligned} \tag{4}$$

In case of the ith experiment there is a random variable u_i which is distributed as u so that the ith value of y, designated by y_i, equals

$$y_i = a_1x_{1i} + \cdots + a_nx_{ni} + u_i.$$

The random variables $u_1, ..., u_N$ are assumed to be independent.

Let I_1 and I_2 designate those subsets of $\{1, ..., N\}$ for which $y_i > y_0$ or $y_i \leq y_0$, respectively. Using maximum likelihood estimation, the likelihood function can be written as

$$L = \prod_{i \in I_1} [1 - F(y_0 - a_1 x_{1i} - \cdots - a_n x_{ni})] \prod_{i \in I_2} F(y_0 - a_1 x_{1i} - \cdots - a_n x_{ni}). \tag{5}$$

Our estimation problem can be formulated as the following nonlinear programming problem:

$$\text{Min} \left\{ -\sum_{i \in I_1} ln[1 - F(z_i) - \sum_{i \in I_2} ln[F(z_i) \right\}$$

subject to

$$\begin{aligned} Aa &= b \\ a_1 x_{1i} - \cdots - a_n x_{ni} + z_i &= y_0, \qquad i = 1, ..., N \\ a_j &\geq 0, \qquad j = 1, ..., n. \end{aligned} \tag{6}$$

3 Numerical Solution of the Estimation Problem

Problem (6) is a linearly constrained nonlinear programming problem with separable objective function (i.e. it is the sum of single variable functions). If both $F(z)$ and $1 - F(z)$ are logconcave functions then each term in the objective function is convex. Now, logconcavity is quite a frequently occuring property of univariate and multivariate probability distribution functions.

A nonegative real functions $f(z)$, $z \in R^k$ is said to be logconcave if for every $z_1, z_2 \in R^k$ and $0 < \lambda < 1$, we have the inequality

$$f(\lambda z_1 + (1 - \lambda) z_2) \geq [f(z_1)]^{\lambda} [f(z_2)]^{1-\lambda}.$$

If f is stricly positive then this is equivalent to the concavity of $ln\, f(z)$.

A sequence $\{p_k\}_{-\infty}^{\infty}$ is said to be logconcave if for every integer k we have the inequality

$$p_k^2 \geq p_{k-1} p_{k+1}.$$

If $f(z)$, $z \in R^k$ is a logconcave probability density function then the corresponding probability distribution function $F(z)$ is logconcave and if $k = 1$ then also $1 - F(z)$ is logconcave (for $k = 1$ see, e.g., Barlow and Proschan (1965) and for the general case see Davidovich, Korenblum and Hacet (1969), Prékopa (1971, 1973)).

For logconcave sequences the key theorem is the one of Fekete (1912) which says that the convolution of two logconcave sequences is also logconcave. This implies that $F(k) = \sum_{i \leq k} p_i$ and $1 - F(k)$ are also logconcave sequences.

In what follows we will use the notation $\bar{F}(z) = 1 - F(z)$.

In practice, problem (6) may have a large number of variables and constraints. The proposed method, however, exploits the special structure of prblem (6) and solves it efficiently. We will assume that the random variable y has discrete logconcave distribution, i.e., the sequence of its probability distribution is logconcave. If y has a continous logconcave distribution, i.e., its probability density function is logconcave then we discretize it by the use of a sufficiently dense sequence.

Let $v_1, ..., v_k$ be the possible values of the random variable y. Picking further values v_0 and v_{k+1} (the necessity of which will be explained later) such that $v_0 < v_1$ and $v_{k+1} > v_k$, we approximate the functions $-\,ln\,F(k)$, $-\,ln\,\bar{F}(z)$ by piecwise linear convex functions. This is done so that we take some large positive numbers as values of $-\,ln\,F(v_0)$, $-\,ln\,\bar{F}(v_k)$ and $-\,ln\,\bar{F}(v_{k+1})$ respectively, furthermore, connect by line segments the planar points

$$(v_0, -\,ln\,F(v_0)), ..., (v_{k+1}, -\,ln\,F(v_{k+1}))$$

as well as the points

$$(v_0, -\,ln\,\bar{F}(v_0)), ..., (v_{k+1}, -\,ln\,\bar{F}(v_{k+1})).$$

Since both $\bar{F}(v_k)$ and $\bar{F}(v_{k+1})$ are 0, theoretically, their large values have to be chosen so that the polygon, obtained from the latter sequence of points should produce a convex function.

The obtained functions will be designated by $G(z)$ and $H(z)$, respectively.

The next step is to represent the function values $G(z)$ and $H(z)$ as optimum values of some linear programming problems. This is done by the so called λ-representations. This means that we introduce the λ variables $\lambda_0, ..., \lambda_{k+1}$ and obtain

$$
\begin{aligned}
\text{Min} \sum_{j=0}^{k+1} G(v_j)\lambda_j &= G(z) \\
\text{subject to} & \\
\sum_{j=0}^{k+1} v_j\lambda_j &= z \\
\sum_{j=0}^{k+1} \lambda_j &= 1 \\
\lambda_j &\geq 0, \qquad j = 1, ..., k+1,
\end{aligned}
\tag{7}
$$

further

$$
\begin{aligned}
\text{Min} \sum_{j=0}^{k+1} H(v_j)\lambda_j &= H(z) \\
\text{subject to} & \\
\sum_{j=0}^{k+1} v_j\lambda_j &= z \\
\sum_{j=0}^{k+1} \lambda_j &= 1 \\
\lambda_j &\geq 0, \qquad j = 1, ..., k+1.
\end{aligned}
\tag{8}
$$

Replacing $G(z)$ and $H(z)$ for $-\,ln\,F(z)$ and $-\,ln\,\bar{F}(z)$, respectively, in problem (6), further, applying the λ-representation for each $G(z_i)$ and $H(z_i)$, we obtain the optimization problem

$$\text{Min} \quad \left\{ \sum_{i \in I_1} \sum_{j=0}^{k+1} G(v_j)\lambda_{ij} + \sum_{i \in I_2} \sum_{j=0}^{k+1} H(v_j)\lambda_{ij} \right\}$$

subject to

$$\begin{array}{rcrcll} & & Aa & = & b & \\ a_1x_{1i} + \cdots + a_nx_{ni} & + & \sum_{j=0}^{k+1} v_j\lambda_j & = & y_0, & i \in I_1 \\ a_1x_{1i} + \cdots + a_nx_{ni} & + & \sum_{j=0}^{k+1} v_j\lambda_j & = & y_0, & i \in I_2 \\ & & \sum_{j=0}^{k+1} \lambda_j & = & 1, & i \in I_1 \cup I_2 \\ a_1 \geq 0, ..., a_n \geq 0 & , & \lambda_j & \geq & 0, & \text{all } i, j. \end{array} \tag{9}$$

Here we see why we had to introduce v_0, v_{k+1} and how we need to choose them. Without v_0, v_{k+1} the second set of equality constraints may introduce some undesired restrictions on $a_1, ..., a_n$. If we choose v_0, v_{k+1} so that

$$v_0 \leq y_0 - a_1x_{1i} - \cdots - a_nx_{ni} \leq v_{k+1}, \quad i = 1, ..., N$$

for all possibile $a = (a_1, ..., a_n)^T$ satisfying $Aa = b$, $a \geq 0$ then the above mentioned difficulty is avoided. Note that sometimes we do not need one or both of the values v_0, v_{k+1}.

The coefficient matrix of problem (9) is shown in Figure 1 (assuming that the elements of I, are the first among the numbers $1, ..., N$).

Problem (9) can be solved efficiently by a method proposed by Prékopa (1990). This method is particularly efficient if the matrix A has a small number of rows. In this case k may be very large (1000 or larger). This means that we can use this method also in the case when we are given a continously distributed y and we intend choose a large number of dividing points in order to discretize the distribution.

A	0
$\begin{matrix} x_{11} & \cdots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1N} & \cdots & x_{nN} \end{matrix}$	$\begin{matrix} v_0 & \cdots & v_{k+1} & & & & & \\ & & & \ddots & & & & \\ & & & & v_0 & \cdots & v_{k+1} \end{matrix}$
0	$\begin{matrix} 1 & \cdots & 1 & & & & & \\ & & & \ddots & & & & \\ & & & & 1 & \cdots & 1 \end{matrix}$

Figure 1: The coefficient matrix of the equality constraint in problem (9).

4 Numerical Example

The numerical example presented below serves for illustration. Assume we are given the Boolean variables x_1, x_2, x_3, x_4, x_5, x_6 and we have the following sample of size 4:

x_1	x_2	x_3	x_4	x_5	x_6	(1 if $y_1 > y_0$ 0 if $y_1 \leq y_0$)
0	1	1	0	0	1	0
1	0	1	1	0	1	0
0	1	0	1	0	1	1
1	0	1	1	1	0	1

Let $v_0 = -2$ and $v_1 = -1$, $v_2 = 0$, $v_3 = 1$ (there is no need to introduce v_4) and assume that

$$P(y = -1) = 0.3, \quad P(y = 0) = 0.5, \quad P(y = 1) = 0.2.$$

This implies that

$$F(-2) = 0, \quad F(-1) = 0.3, \quad F(0) = 0.8, \quad F(1) = 1$$

$$\bar{F}(-2) = 1, \quad \bar{F}(-1) = 0.7, \quad \bar{F}(0) = 0.2, \quad \bar{F}(1) = 0.$$

The values of the functions $G(z)$ and $H(z)$ at the points -2, -1, 0, 1 are given in the table below

z	-2	-1	0	1
$G(z)$	1000	1.2	0.22	0
$H(z)$	0	0.36	1.6	1000

As regards a_1, a_2, a_3, a_4, a_5, a_6 the only constraints are $a_1 + a_2 + a_3 + a_4 + a_5 + a_6 \leq 2$ and $a_1 \geq 0$, $a_2 \geq 0$, $a_3 \geq 0$, $a_4 \geq 0$, $a_5 \geq 0$, $a_6 \geq 0$. Introducing the slack variable $a_7 \geq 0$ we convert the first inequality into equality.

Assume $d = 2$ and $= 0.5$. Then we have the following linear programming problem

$$\text{Min} \quad \left\{ \sum_{i=1}^{2} (1000\lambda_{i0} + 1.2\lambda_{i1} + 0.22\lambda_{i2} + 0\lambda_{i3}) \right.$$
$$\left. + \sum_{i=3}^{4} (0\lambda_{i0} + 0.36\lambda_{i1} + 1.6\lambda_{i2} + 1000\lambda_{i3}) \right\}$$

subject to

$$\begin{array}{rrrrrrrrrrrrcl}
a_1 & +a_2 & +a_3 & +a_4 & +a_5 & +a_6 & +a_7 & & & & & & = & 2 \\
 & a_2 & +a_3 & & & +a_6 & & -2\lambda_{10} & -\lambda_{11} & +0\lambda_{12} & +\lambda_{13} & & = & 0.5 \\
a_1 & & +a_3 & +a_4 & & +a_6 & & -2\lambda_{20} & -\lambda_{21} & +0\lambda_{22} & +\lambda_{23} & & = & 0.5 \\
 & a_2 & & +a_4 & & +a_6 & & -2\lambda_{30} & -\lambda_{31} & +0\lambda_{32} & +\lambda_{33} & & = & 0.5 \\
a_1 & & +a_3 & +a_4 & +a_5 & & & -2\lambda_{40} & -\lambda_{41} & +0\lambda_{42} & +\lambda_{43} & & = & 0.5 \\
 & & & & & & & \lambda_{i0} & +\lambda_{i1} & +\ \lambda_{i2} & +\lambda_{i3} & & = & 1, \quad i = 1,2,3,4
\end{array}$$

$$a_1 \geq 0,\ a_2 \geq 0,\ a_3 \geq 0,\ a_4 \geq 0,\ a_5 \geq 0,\ a_6 \geq 0,\ a_7 \geq 0, \qquad \lambda_{ij} \geq 0 \quad \text{all } i, j.$$

The basic components of the optimal solution are:

$$a_2 = 0.5, \quad a_4 = 1, \quad a_5 = 0.5$$

$$\lambda_{12} = 1, \quad \lambda_{13} = 0, \quad \lambda_{21} = 0.5, \quad \lambda_{22} = 0.5, \quad \lambda_{31} = 1, \quad \lambda_{41} = 1$$

All the other components of the optimal solution are 0.

We only need the optimal a_i, $i = 1, 2, 3, 4, 5, 6$ values. The result tells us that the variables x_2, x_4, x_5 represent the principal causes and these share 25%, 50% and 25%, respectively, in the realization of the effect.

References

[1] Barlow, R.E. and Proschan, F. (1965). *Mathematical Theory of Reliability.* Wiley, New York.

[2] Crama, Y., Hammer, P.L. and Ibaraki, T. (1988). Cause-Effect Relationship and Partially Defined Boolean Functions. *Annals of Operations Research* **16**, 299–326.

[3] Davidovich, Yu.S., Korenblum, B.J. and Hacet, B.J. (1969). On a Property of Logarithmic Concave Functions. *Dokl. Akad. Nauk.* **185**, 1215–1218.

[4] Fekete, M. and Pólya, G. (1912). Über ein Problem von Laguerre. *Rendiconti del Circolo Matematico di Palermo* **23**, 89–120.

[5] Prékopa, A. (1971). Logarithmic Concave Functions with Applications to Stochastic Programming. *Acta Sci. Math.* (Szeged) **32**, 301-316.

[6] Prékopa, A. (1973). On Logarithmic Concave Measures and Functions. *Acta Sci. Math.* (Szeged) **34**, 335-343.

[7] Prékopa, A. (1990). Dual Method for a One-Stage Stochastic Programming Problem with Random RHS Obeying a Discrete Probability Distribution. *ZOR-Methods and Models of Operations Research* **34**, 441-461.