NOTE: Cowles Commission Discussion Papers are preliminary materials circulated privately to stimulate private discussion and are not ready for critical comment or appraisal in publications. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

The Problem of Finding Optimal Decisions

Leo Törnqvist

December 15, 1952

We now assume that we have been able to construct an indicator of goodness \( g(a) \) to every plan of action considered \( a \in A \). The problem is then to invent an optimal method for finding such plans \( a \in A \) that

\[
g(a) = g(\hat{a}) = \max_{a \in A} g(a).
\]

The problem is thus to find a solution to a conditional optimum problem.

This problem can be mathematically attacked in many different ways:

I. The method of transforming the conditional optimum problem to an unconditional optimum problem for a function \( Q_c(a) \) which, when \( c \to c \), converges to \( g(a) \) for \( a \in A \) and to \(-\infty\) for \( a \notin A \).

\[
a = \lim_{c \to c} a_c
\]

\[
Q_c(a) = \lim_{c \to c} Q_c(a)
\]

Such functions \( Q_c(a) \) are for instance

\[
Q_c(a) = g(a) + \log \left[ \frac{1}{c} \right] \left( a \in A | c \right) \quad \text{where}
\]

\[
Q_c(a) = \log \left( a | A \right) \quad \text{for a point } a \text{ such that } a, a + c
\]

and \( c \) are all outside \( A \).

1. Research undertaken by the Cowles Commission for Research in Economics under contract Nonr-338(01), AR 037-095 with the Office of Naval Research.
1. If \( a \in A \mid \xi \) then \( 1 = |\xi| \) for a point \( a \) such that \( a, a + \xi \) and \( a + \xi \) are all inside \( A \).

II. The method of Lagrange multipliers

\[
G(\lambda, a) = g(a) + G(\lambda, a)
\]

where \( \lambda \) is a table containing so many numbers \( n_A \) that there always exists a finite number of tables \( (a, \lambda) \) such that \( (a, \lambda) \) is a solution to the system of equations

\[
\begin{align*}
G(\lambda, a) &= 0 \text{ for every } a \in A, \\
\frac{\partial G(\lambda, a)}{\partial a_2} &= \frac{\partial g(a)}{\partial a_2} + \frac{\partial G(\lambda, a)}{\partial a_2} = 0
\end{align*}
\]

where \( a_2 \) is one number in the position \( z \) of the set of numbers defining \( a_0 \). If \( n_A \) is reduced to \( n_A = 1 \) no solutions to II exist. The optimal solutions \( a_0 \) are all contained in some of the solutions to the system of equations II.

III. The method of moving from \( a_v \in A \) to \( a_{v+1} \in A \) whenever \( g(a_{v+1}) > g(a_v) \).

The numbers \( a_{v+1} \) are sought from a neighborhood \( a_v + e_v \Delta_v \) where \( \Delta_v \) is a table containing only one number \( \Delta_v \geq 0 \) and \( e_v \) is a table of numbers containing only the numbers \( \xi = 1 \) and \( 0 \) in different positions \( z \) of the tables \( A \). An upper limit \( \bar{\Delta}_v \) for \( \Delta_v \) is then defined for every \( \xi_v \) by the requirement that \( a_v + e_v \Delta_v \subset A \). The points \( a_v + e_v \Delta_v \) are studied in the order of the size of the sum \( N_v = \sum |e_v| \). When a direction \( e_v \) with \( \bar{\Delta}_v \geq 0 \) is found such that

\[
0 < \Delta_v < \bar{\Delta}_v, \quad g(a_{v+1} = a_v + e_v \Delta_v) > g(a_v),
\]

we proceed to find a new point \( a_{v+2} \) better than \( a_{v+1} \) and so on until the process stops, at an \( a_n = a_n(a_0) \). If \( N_v \) is restricted to a number \( n_v = n_v(N_v) \) (for instance, \( n_v = 6 \)), at a table \( a_n(a_0) \) we need necessarily have \( g(a_n) = g(1) \) but the difference \( g(a) - g(a_n) \) is usually relatively
small. By varying the "opening" \( a_0 \) of the game we can find a set of
tables \( a_n(a_0) \), which have the property that some kind of "probability"
that \( \max_{a_0} (g(\bar{a}) - g(a_n(a_0)) > \xi \) goes to zero very rapidly with an in-
creasing number of trials for openings \( a_0 \) and/or increasing numbers of
\( N_v \) and/or \( \xi \). For functions \( g(a) \) which have no uniquely defined partial
derivatives \( \frac{\partial g(a)}{\partial z} \) the method III to construct an increasing sequence \( g(a_v) \)
\( \rightarrow g(\bar{a}) \) seems to be the only promising method available. This method is
also the only one for finding the over all optimal solutions in the cases
when the method II picks out a denumerable set of extremal points (rela-
tive maxima or minima). In all complex cases we thus have to rely on a
method of successive improvements \( g(a_v) \rightarrow g(\bar{a}) \). Fortunately this method
seems to be also the most promising method for finding optimal decisions
in non trivial decision problems. In some cases it is also possible to
calculate a measure
\[
M(A \cap (g(a) > g)) = M(A, g)
\]
for the set of tables \( a_v \) belonging to the intersection of the sets \( A \),
and the set of \( a_v \), fulfilling \( (g(a) > g) \) as a function of \( g \). This measure
is zero for \( g \geq g(\bar{a}) \).

If we then determine the largest value of \( g \) for which \( M(A, g) \neq 0 \),
we have the possibility of determining \( g(\bar{a}) \). The limit sets \( (A, g) \subset (A, g)_{\xi} \)
such that
\[
\begin{align*}
0 & \quad \{ A \cap (g(a) > g + \xi) = 0 \\
\xi & \quad \{ A \cap (g(a) > g - \xi) \neq 0 \}
\end{align*}
\]
gives the solution to the optimal problem considered, but it is often
too complicated to use this straightforward way of finding an \( a \subset A \).

The problem of finding an optimal way for finding optimal decisions
is solvable only if we can construct an indicator of goodness for different methods of seeking ways to a whole set of approximative solutions in \( \varepsilon \) - neighbourhoods of the optimal solutions. A non-optimal solution to the decision problem to maximize \( g(a) \); \( a \in A \) can namely be a better solution to the problem of maximizing the indicator of goodness

\[
g(a_n, F_\lambda, \lambda); F_\lambda(\varepsilon_\lambda) = \mathbb{P}(g(a) = g(a_n) \leq \varepsilon_\lambda)
\]

for the whole activity to find a method \( \lambda \) to get good decisions \( a_n \in A \).

It seems to be clear that the variance of \( \varepsilon_\lambda \) is usually \( > 0 \) for the best method \( \lambda = \hat{\lambda} \).

For a full understanding of the principles expressed, they must be clarified by examples of applications to special optimum problems. This paper is only an introduction to this field of study.