Technical Notes

OPERATIONS RESEARCH ON FOOTBALL

Virgil Carter

Cincinnati Bengals Football Team, Cincinnati, Ohio

and

Robert E. Machol

Northwestern University, Evanston, Illinois

(Received December 2, 1970)

Based on a census of 8,373 plays from the first 56 games of the 1969 schedule of the National Football League, this note calculates, for various field positions, the expected point values of possession of the football with first down and ten yards to go, and discusses the strategic implications of these values.

SEVERAL operations-research studies have been performed[1] on strategy and tactics for baseball. To the best of our knowledge, no comparable coherent effort has been published concerning professional football, although a brief note appeared in 1954.[2]

In order to perform such a study, we obtained data on the 56 games played during the first half of the 1969 National Football League schedule. Each of the 8,373 individual plays in these games was coded, punched, and entered into a computer, and all analyses were made on this data base. Comparative analyses of results from the total sample with results from subsamples, as well as routine computations of standard errors, indicated that, for most of the questions we wished to ask, this sample was sufficiently large.

A number of statistical analyses were performed, and these will be published elsewhere. Most of them tended to be negative in their results, or to indicate that optimum strategies appeared to have been evolved, apparently by intuitive methods. However, for the one item we report on here, we have deduced what we consider to be interesting numbers and an indication of a change that should be made in present football strategies.

This analysis concerns the expected value of possession of the football, with first down and ten yards to go, at any particular point on the playing field. The basic formula for expected value is, of course, $E(X) = \sum X_i P(X_i)$. The number of possible outcomes is 103, the first four being touchdown ($X = +7$), field goal ($X = +3$), safety ($X = -2$), and an opponent's touchdown (due to a fumble or an interception) ($X = -7$). Perhaps 6.98 might have been a more accurate reward for a touchdown, but 7 seemed an adequate approximation. The remaining possible outcomes consist of eventually turning over the ball to the opponents at one of the 99 possible points on the field. It is then assumed that the appropriate value of X here is $-E(X)$. This leads to a system of 99 equations in 99 unknowns. Since not enough data were available to determine these 99 probabilities with adequate accuracy, the field was divided into ten strips, namely, 99 to 91 yards to go, 90 to 81 yards to go, 80 to 71 yards to go, etc. These data sets are identified by their mid-points in Table I. The smallest number of data points in any set was 57 (this set centered on 95 yards to go), and the largest was 601 (this set centered on 75 yards to go). This condensation led to a system of ten equations in 10 unknowns.

The results are presented in Table I. The analysis was based on a study of 2,852 first-and-ten plays. An independent calculation was performed for the subset of 1,258 first downs immediately following turnover (i.e., the start of series), and the

TABLE I

The Expected Point Values of Possession of the Football with First Down and Ten Yards to Go for Various Ten-Yard Strips

Center of the ten-yard strip (yards from the target goal line): X	Expected point value: $E(X)$
95	−1.245
85	−0.637
75	+0.236
65	0.923
55	1.538
45	2.392
35	3.167
25	3.681
15	4.572
5	6.041

average absolute difference was less than a quarter of a point. Psychological differences as well as statistical fluctuations may have affected this difference.

There are some obvious inadequacies in the analysis. In the first place, the several hundred situations starting on the 20-yard line should probably have been separated into an eleventh category. Apparently, first and ten on one's own 20 has an expected value very close to zero, which indicates wisdom on the part of the rules-makers. It would have been nice to check to see how exactly this comes out. More significantly, the value of a kick following a score was ignored, although this is presumably negative. The free kick from the 20-yard line following a safety obviously has negative expected value, but this occurs with sufficiently low frequency as to have little effect on the analysis. More important, the kickoff from the 40-yard line following a field goal or touchdown is a very frequent occurrence and appears to have a negative expected value. This is considered qualitatively in the following discussion.

The numbers in Table I, which are given to three decimal places, are obviously in error by perhaps several tenths of a point. Nonetheless, they have some obvious qualitative value. In basketball, a good deal of strategy is built about the accepted fact that the rebound is worth about one point. There is a similar qualitative value here in having an approximate knowledge of the value of having the ball at a particular point on the field. More significantly, it appears to us that one technical decision is regularly being made incorrectly in professional football. Specifically, the negative value of having the ball with first and ten very close to one's own goal line has been ignored. We are referring to the type of situation where a team has the ball, fourth and goal, in which case a field goal is routinely attempted. Even in desperate situations (say six points behind in the fourth quarter, when a field goal is considered rather worthless), the usual tactic under a fourth and goal situation would be an attempt to pass. The rules of professional football, if a pass is incomplete in the end zone, require that the ball be brought out to the 20. We would recommend that the ball be rushed under these circumstances. The formal analysis is as follows: If the decision is to kick, there can be two possible outcomes. Three points will be made with probability varying from about 0.98 if the ball is well centered and the team has a good kicker to perhaps 0.6 if the angle is very bad and the kicker is poor. From this expected value must be subtracted the negative expected value of the ensuing kickoff. If the kick is missed, the ball is brought out to the 20 where the expected value is approximately zero. On the other hand if the ball is run, there is a probability of perhaps 0.2 or 0.3 of making virtually 7 points; and if this run fails to make a touchdown, the ball is turned over so close to the enemy goal line that there is something between one and two points of negative value for the opposition. Unfortunately, there is insufficient data to pin down the exact variation in this negative value within the last few yards of the field. But even if the negative value is only the one and a quarter points, which appears to be the average for the one- to nine-yard lines, it seems clear that the analysis on a fourth-and-goal situation is very different from the situation where one has fourth and four on, say, the 20-yard line. In the latter case, the kick seems well justified.

A similar analysis is pertinent in evaluating the 'coffin-corner' punt.

We would like to take this opportunity to suggest one other area in which most professional football teams seem to be making an incorrect choice of strategy. No statistics have been gathered to support this, but the error (as we assume it to be) does derive from a naive misunderstanding of decision theory. We refer to the strategy of calling time-outs during the last two minutes of either half.

If one is considering calling a time out, the Type I error is in calling it when one should not, and the penalty arises when the ball is turned over to the opponents before the time expires. This may happen through a fumble or interception, through running out of downs followed by a punt, through a missed field-goal attempt, or through an actual score after which the opponents again get the ball. This Type I error appears to be a very frequent occurrence. As a well-known example, we cite the Cleveland-Oakland game of November 8, 1970, in which Cleveland, attempting to break a tie, called a time out and subsequently lost the ball by an interception; then BLANDA had enough time to win the game by kicking a field goal.

The Type II error consists of not calling a time out when one should. The naive assumption appears to have been that the penalty for this is running out of time. This is wrong. The penalty for a Type II error is running out of time when one still has time outs left to call, and we believe this to be an extremely rare occurrence, if indeed it has ever occurred.

That is, we assert that the Type I error, calling a time out when one should not, and being penalized for it, has been made many times; whereas, the Type II error, failing to call a time out when one should, is extremely rare. Specifically, therefore, we recommend that a team that is behind seven points or less should never call a time out when there are more than 30 seconds to play if it has the ball, or more than one minute to play if the opponents have the ball. We are referring here only to the time outs that are called for the exclusive purpose of stopping the clock.

REFERENCES

1. GEORGE LINDSEY, review of EARNSHAW COOK, *Percentage Baseball*, Second Edition, The MIT Press, Cambridge, Mass., 1966, in *Opns. Res.* **16**, 1088–1089 (1968). This review lists previous research literature.
2. CHARLES M. MOTTLEY, "The Application of Operations-Research Methods to Athletic Games," *Opns. Res.* **2**, 335–338 (1954).

OPTIMIZATION OF THE CAPACITY IN A STORAGE SYSTEM

S. G. Loo and A. Ghosal

Monash University, Clayton, Victoria, Australia

(Received June 7, 1970)

This paper presents the problem of determining the optimal capacity of a storage system with respect to some specified criteria. It assumes that the storage system is subject to an input X and a release Y, at least one of which is a random variable following a known distribution function, so that the storage function Z is a stochastic process. The optimal capacity over a time horizon $(0, T)$ is determined by maximizing the expected profit.

CONSIDER the operation of a storage system of finite capacity K over a time horizon $(0, T)$: let $Z(t)$ represent the storage level or units of occupied storage space at time t, and suppose that the revenue is $r(t, K)$ per unit storage per unit time, and that the continuous interest rate is δ.

PRESENT VALUE OF NET PROFIT

THE PRESENT value of total revenue over a time horizon $(0, T)$ is given by (cf. GHOSAL[2])

$$r_0(T, K)=\int_0^T e^{-\delta t}Z(t)r(t, K)\,dt. \tag{1}$$

When time is a discrete variable, $t=1, 2, 3, \cdots, T$, we have the corresponding relation

$$r_0(T, K)=\sum_{t=1}^{t=T} (1+i)^{-t}Z(t)r(t, K),$$