The Probability of a Field Goal; Rating Kickers

Donald A. Berry

and

Timothy D. Berry

University of Minnesota
School of Statistics
Technical Report No. 428

ABSTRACT

The probability that a field goal attempt will be successful is estimated. The estimate depends on the yardages of the kicker's previous successful and unsuccessful attempts. The estimate involves modeling both accuracy and distance. A method for ranking kickers based on this estimate is proposed. National Football League kickers are ranked on the basis of their 1983 regular season records.

1. Introduction

To indicate how likely it is that a kicker will make a particular 38-yard field goal, say, a football commentator sometimes gives the kicker's success proportion between 30 and 40 yards--9 of 12, say. But 38 yards may not be typical among the kicker's previous attempts between 30 and 40 yards. It might be, for example, that all of the 12 have been from closer than 38 yards.

More importantly, there is information concerning the kicker's ability to make a 38-yard field goal in his record for attempts from less than 30 and from greater than 40 yards. For example, he might have made fewer than 75% of his attempts from less than 30 yards, and presumably, a miss from 28 yards would also have been a miss from 38 yards. On the other hand, his success ratio from greater than 38 yards might be greater than 75%.

Our objective is to use all available information to estimate the probability of a successful field goal attempt of any given yardage for any kicker. Information about the kicker consists of the various yardages at which he was successful and the various yardages from which he was unsuccessful. The reason for a failure (too short, e.g.) and where the ball was placed between the hatch marks are not known. (This information is not ordinarily gathered by football leagues.) We develop a model to give this estimate in Section 2.

The estimate derived in Section 2 serves to rank various kickers for any given yardage. We propose a scoring method in Section 3 which is based on this estimate but does not depend on any particular yardage. This method does not have the obvious flaw that is possessed by a simple field goal percentage that ignores yardage: a kicker is not penalized unduly for missing a very long field goal attempt and is not rewarded unduly for making many very short field goals. In Section 4 we illustrate the method by ranking the field goal kickers in the National Football League according to their 1983 regular season records.

Morris and Rolph (1981, pp. 194-200) used logistic regression to estimate the probability of a field goal in both the NFL and AFL in 1969. They group the kickers in each league and also group attempts in ten-yard intervals. We are more interested in individual kickers. Since there is much less data for this case we prefer success-failure responses rather than success proportions for grouped yardages. But we feel that the assumed symmetry about the yardage at which the odds are even makes logistic regression inappropriate in any case: a kicker may reasonably have a 75% chance of success at 35 yards, say, a 50% chance at 50 yards, and be unable to kick a ball further than 55 yards --symmetry implies a 25% chance at 65 yards. One might decide to only use the logistic estimate up to a certain yardage, but the imposition of symmetry in an inherently asymmetric setting can make an estimate quite bad for practically all yardages.

2. The Model

We want a model which incorporates the two requirements for a successful field goal: accuracy and distance (we include height, incorporating it with the latter). However, since we are only modeling success and failure, and not, for example, that an attempt was wide or short, some confounding is necessary. Accuracy is handled in a formal way and maximum likelihood estimation is employed. That accuracy decreases with distance is incorporated explicitly in this development. The distance a kicker can kick a football is estimated from his record and incorporated directly with a very simple statistical model.

When a football is kicked, the directional deviation from the aim point depends only on the accuracy of the kicker--we assume he makes the same attempt at accuracy regardless of the distance. We assume further that the kicker always aims for the midpoint between the goal posts and the resultant deviation has a symmetric distribution. (We assume symmetry for expository reasons only; the assumption can be relaxed with no essential change in the later development.) Under these assumptions the standard deviation of the error distribution is proportional to the distance from the goal posts; see Figure 1.

We assume that the deviation x has a logistic density with standard deviation σ: for all real x,

$$f(x;\sigma) = \frac{(1/\sigma)e^{-x/\sigma}}{(1 + e^{-x/\sigma})^2}$$

The logistic distribution is convenient for this purpose since it is quite tractible. The probability of a successful field goal is then

$$\frac{2}{1 + e^{-D/\sigma}} - 1$$

where 2D is the distance between goal posts. Since σ is proportional to the field goal yardage y, we can write this probability as

$$\frac{2}{1 + e^{-h/y}} - 1$$

where h is a positive parameter that equals D divided by the multiple of y in the standard deviation. Note that this probability tends to 0 as $h \to 0$ and to 1 as $h \to 0$ for any y. The parameter h indexes the kicker's accuracy: the larger h is, the more accurate is the kicker. An interpretation of h is as follows: $h/\log 3 \doteq 0.91\,h$ is the distance at which there is a 50% chance of the line of the kick passing between the goal posts (independent of whether he can kick that far!) and $h/\log 7 \doteq 0.51\,h$ is the distance at which there is a 75% chance. We choose the latter multiple (arbitrarily) and call $h/\log 7$ the kicker's ACCURACY. In Figure 2 we set ACCURACY = 20 yards (which means $h = 38.92$) and illustrate the effect on the kicker's accuracy at 40 yards; at which he has a 45.1% chance of success.

The likelihood function of h depends on the kicker's record of success yardages and miss yardages; we include points-after-touchdown as yardage = 20. Suppose he had G successful and B unsuccessful attempts. Let $(y_1, y_2, \ldots, y_G)$ denote the yardages corresponding to successful attempts and $(y_1^*, y_2^*, \ldots, y_B^*)$ the yardages corresponding to unsuccessful attempts.

The likelihood function also depends on the probability that a kick will travel at least distances $y_1, \ldots, y_G$ and $y_1^*, \ldots, y_B^*$ (and be higher than the crossbar). Let $P(y)$ denote the probability that a kick travels at least distance y (regardless of accuracy). Then the likelihood of h is

$$L(h) = \prod_{i=1}^{G}\left[\frac{2}{1 + e^{-h/y_i}} - 1\right]P(y_i)\prod_{j=1}^{B}\left[2 - \frac{2}{1 + e^{-h/y_i^*}}\right](1 - P(y_i^*)).$$

We have worked with various models for P but found most of them quite unsatisfactory. One problem in estimating unknown parameters in such a model is that the data concerning distance are very sparse: field goals are not usually attempted near maximum range. We use the following function:

$$P(y) = \begin{cases} 1 & \text{if } y < m_* \\ 1 - \left(\frac{y-m_*}{m^*-m_*}\right)^2 & \text{if } m_* \leq y \leq m^* \\ 0 & \text{if } y > m^* \end{cases}$$

and estimate

$$m_* = \max_{1 \leq i \leq G} \{y_i\} - 5$$

$$m^* = \max_{1 \leq i \leq G} \{y_i\} + 5$$

So the distance that a kick travels is assumed to have a triangular density over an interval five yards on either side of the kicker's longest successful attempt. One implication of this assumption is that the probability of success for any y is not affected by a previously missed, very long attempt. This choice of P is somewhat arbitrary. But it has the desirable feature that all kickers are affected equally, depending on their longest field goal. If one were able to condition on information as to which attempts fall short then, of course, much better estimates of a kicker's distance distribution would be possible.

The likelihood function L is unimodal but the maximum likelihood estimate $\hat{h}$ requires a numerical procedure. We performed the necessary calculations with a personal computer (our BASIC program is available on request).

3. <u>An Example: Chris Bahr in 1983</u>

In the NFL 1983 regular season, Chris Bahr of the Los Angeles Raiders attempted 27 field goals and 21 were successful. The distances of these attempts (taken from newspaper accounts) are as follows, with failures underlined: 21, 22, 24, 26, 26, <u>26</u>, 27, 28, 28, <u>28</u>, 29, 32, 32, 36, <u>36</u>, 37, 38, 38, 39, 39, 41, <u>41</u>, <u>41</u>, 42, 42, 47, <u>52</u>. In addition, he attempted 53 points-after-touchdown, making 51. So G = 72 and B = 8.

We find $m_* = 42$, $m^* = 52$, and $\hat{h} = 76.34$. Setting aside the possibility that the kick falls short, Bahr's ACCURACY is $\hat{h}/\log 7 = 39.23$ yards. The estimated probability that Bahr makes a 35-yard field goal, for example, is

$$\frac{2}{1 + e^{-76.34/35}} - 1 = 79.71\%.$$

The estimated probability he makes a point-after-touchdown (or a 20-yard field goal) is

$$\frac{2}{1 + e^{-76.34/20}} - 1 = 95.70\%$$

(this is slightly smaller than his conversion ratio of 51/53 = 96.23% because his field goals indicate slightly less accuracy than this ratio suggests). The entire curve is shown in Figure 3.

4. <u>Ranking NFL Kickers in 1983</u>

Given a curve analogous to Figure 3 for each of a number of kickers, there are a variety of ways to compare them. One is to choose a particular yardage and rank by estimated probability. Another method is to rank on the basis of ACCURACY--the yardage at which the kicker is 75% accurate. This latter method is used in Table 1 to rank kickers in the National Football League according to their 1983 records.

A method of comparison which takes both accuracy and distance into account is to integrate the probability of a successful field goal from 0 to ∞ (or, equivalently, to m*). We define a kicker's

$$\text{OVERALL SCORE} = \sum_{y=1}^{\infty} \left[\frac{2}{1 + e^{-h/y}} - 1\right] P(y).$$

This is an index of average ability of the kicker in the sense that it is the average yardage for which his kicks are good. For example, Bahr's OVERALL SCORE of 42.41 indicates that the average distance his kicks travel before they would no longer go between the goal posts and over the crossbar is estimated to be 42.41 yards.

NFL kickers are also ranked by their 1983 OVERALL SCOREs in Table 1.

Some explanation of the way in which accuracy and distance affect OVERALL SCORE is in order. If a kicker is perfectly accurate and he kicks a field goal five yards longer than his previous longest, then his OVERALL SCORE is increased by five yards. More realistically, if a kicker's longest kick is from a distance at which he is about 60% accurate then his OVERALL SCORE would increase by about 60% of five yards--about three yards.

Acknowledgement: Professors Kinley Larntz and Sanford Weisberg pointed out the field goal example in the Morris-Rolph text.

REFERENCE

Morris, Carl N. and John E. Rolph (1981). Introduction to Data Analysis and Statistical Inference. Prentice-Hall, Inc., Englewood Chiffs, N.J.

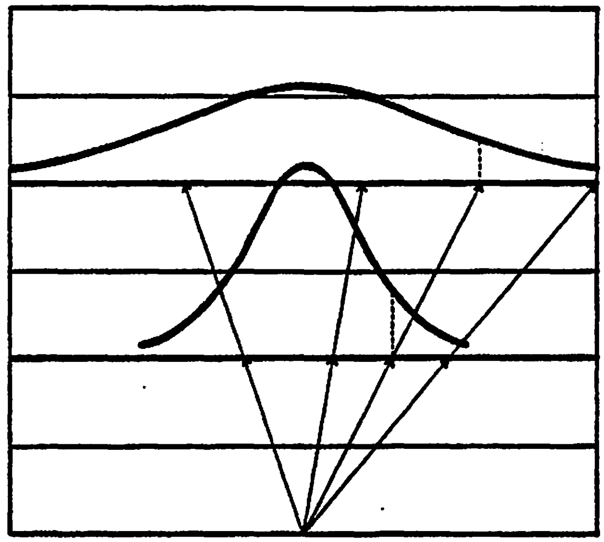

Figure 1. Deviation from the aim point is proportional to distance from goalposts.

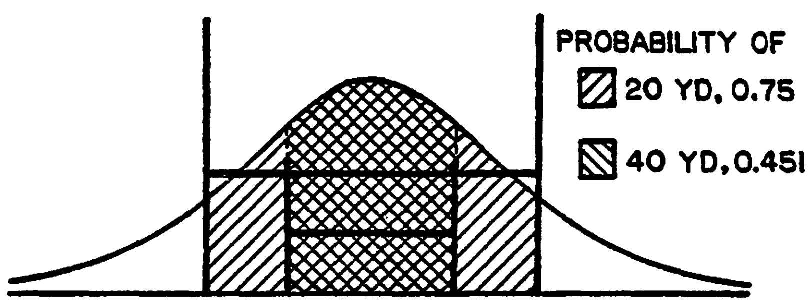

Figure 2. The logistic density for directional deviation; shaded areas indicate the probability of successful field goal from 20 and 40 yards for the case $h = 38.92$.

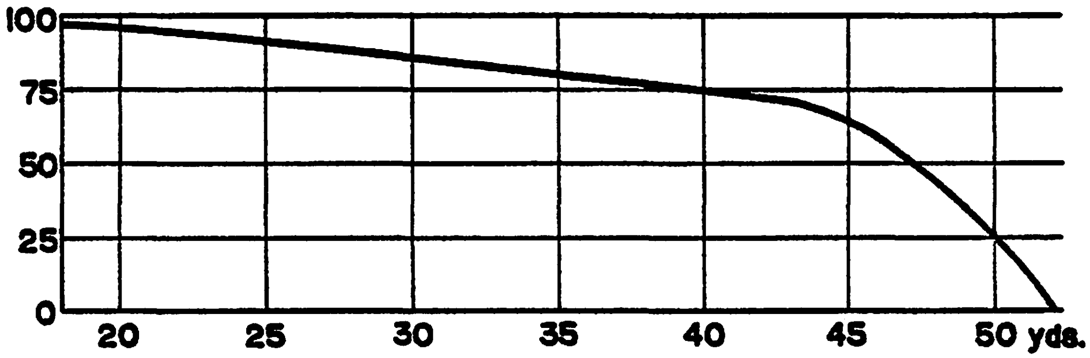

Figure 3. Estimated percent probability of successful field goal for Chris Bahr, 1983.

TABLE 1. ACCURACY and OVERALL SCORE for NFL kickers in the 1983 regular season

KICKER; TEAM	ACCURACY	KICKER	OVERALL SCORE
WERSHING, SAN FRANCISCO	58.80	LOWERY	52.25
STENERUD, GREEN BAY	51.61	HAJI-SKEIKH	51.62
HAJI-SKEIKH, N.Y. GIANTS	51.44	WERSCHING	49.97
G. ANDERSON, PITTSBURGH	51.02	ALLEGRE	48.92
LOWERY, KANSAS CITY	48.54	JOHNSON	48.26
KEMPF, HOUSTON	44.69	MURRAY	47.88
ALLEGRE, BALTIMORE	43.80	KEMPF	46.54
JOHNSON, SEATTLE	43.72	G. ANDERSON	46.45
M. BAHR, CLEVELAND	41.91	STENERUD	45.78
SEPTIEN, DALLAS	41.82	M. ANDERSON	45.69
MURRAY, DETROIT	40.64	MOSELEY	44.49
M. ANDERSON, NEW ORLEANS	39.57	VON SCHAMANN	43.85
C. BAHR, L.A. RAIDERS	39.23	KARLIS	43.72
LUCKHURST, ATLANTA	38.58	LUCKHURST	43.50
RICARDO, MINNESOTA	37.93	M. BAHR	43.12
MOSELEY, WASHINGTON	37.84	LEAHY	42.69
KARLIS, DENVER	37.36	BENIRSCKE	42.62
LEAHY, N.Y. JETS	36.15	C. BAHR	42.41
BREECH, CINCINNATI	35.30	SEPTIEN	42.41
VON SCHAMANN, MIAMI	34.64	BREECH	42.39
BENIRSCKE, SAN DIEGO	33.06	O'DONOGHUE	41.26
DANELO, BUFFALO	29.99	FRANKLIN	40.86
*STEINFORT, N.E. & BUFFALO	29.76	THOMAS	40.46
THOMAS, CHICAGO	29.57	RICARDO	40.07
O'DONOGHUE, ST. LOUIS	29.20	DANELO	39.63
FRANKLIN, PHILADELPHIA	28.48	CAPECE	38.42
NELSON, L.A. RAMS	28.34	NELSON	35.07
CAPECE, TAMPA BAY	26.64	*SMITH	32.79
*SMITH, NEW ENGLAND	21.79	*STEINFORT	31.87

*Fewer than 40 attempts, field goals plus p.a.t.'s.