On the Probability of Winning a Football Game

HAL STERN*

Based on the results of the 1981, 1983, and 1984 National Football League seasons, the distribution of the margin of victory over the point spread (defined as the number of points scored by the favorite minus the number of points scored by the underdog minus the point spread) is not significantly different from the normal distribution with mean zero and standard deviation slightly less than fourteen points. The probability that a team favored by p points wins the game can be computed from a table of the standard normal distribution. This result is applied to estimate the probability distribution of the number of games won by a team. A simulation is used to estimate the probability that a team qualifies for the championship playoffs.

KEY WORDS: Goodness-of-fit tests; Normal distribution.

1. INTRODUCTION

The perceived difference between two football teams is measured by the point spread. For example, New York may be a three-point favorite to defeat Washington. Bets can be placed at fair odds (there is a small fee to the person handling the bet) on the event that the favorite defeats the underdog by more than the point spread. In our example, if New York wins by more than three points, then those who bet on New York would win their bets. If New York wins by less than three points (or loses the game), then those who bet on New York would lose their bets. If New York wins by exactly three points then no money is won or lost. The point spread is set so that the amount bet on the favorite is approximately the same as the amount bet against the favorite. This limits the risk of the people who handle the bets.

The point spread is of natural interest as a predictor of the outcome of a game. Although it is not necessarily an estimate of the difference in scores, the point spread has often been used in this capacity. Pankoff (1968), Vergin and Scriabin (1978), Tryfos, Casey, Cook, Leger, and Pylypiak (1984), Amoako-Adu, Marmer, and Yagil (1985), and Zuber, Gandar, and Bowers (1985) considered statistical tests of the relationship between the point spread and the outcome of the game. Due to the large variance in football scores, they typically found that significant results (either proving or disproving a strong relationship) are difficult to obtain. Several of these authors then searched for profitable wagering strategies based on the point spread. The large variance makes such strategies difficult to find. Other authors (Thompson 1975; Stefani 1977, 1980; Harville 1980) attempted to predict game outcomes or rank football teams using information other than the point spread.

The results of National Football League (NFL) games seem to indicate that the true outcome of a game can be modeled as a normal random variable with mean equal to the point spread. This approximation is developed in some detail, and two applications of this approach are described.

2. DATA ANALYSIS

The data set consists of the point spread and the score of each NFL game during the 1981, 1983, and 1984 seasons. Many newspapers list the point spread each day under the heading "the betting line." The sources of the point spread for this data set are the *New York Post* (1981) and the *San Francisco Chronicle* (1983, 1984). There is some variability in the published point spreads (from day to day and from newspaper to newspaper), however, that variability is small (typically less than one point) and should not have a large impact on the results described here. An attempt was made to use point spreads from late in the week since these incorporate more information (e.g., injuries) than point spreads from early in the week. For reasons of convenience, the day on which the data were collected varied between Friday and Saturday. The 1982 results are not included because of a player's strike that occurred that year. The total number of games in the data set is 672. More recent data (from 1985 and 1986) are used later to validate the results of this section. For each game the number of points scored by the favorite (F), the number of points scored by the underdog (U), and the point spread (P) are recorded. The margin of victory over the point spread (M) is defined by

$$M = F - U - P$$

for each game. The distribution of M is concentrated on multiples of one-half since F and U are integers, while P is a multiple of one-half.

A histogram of the margin of victory over the point spread appears in Figure 1. Each bin of the histogram has a width of 4.5 points. The chi-squared goodness-of-fit test indicates that the distribution of M is not significantly different from a Gaussian distribution with mean zero and standard deviation 13.86 (computed from the data). The sample mean of M is .07. This has been rounded to zero because it simplifies the interpretation of the formula for the probability of winning in the next section. All observations of M larger than 33.75 in magnitude are grouped together, leading to a chi-squared test on 17 bins. The chi-squared test statistic is 15.05, between the .5 and .75 quantile of the limiting chi-squared

*Hal Stern is Assistant Professor, Department of Statistics, Harvard University, Cambridge, MA 02138. The author thanks Thomas M. Cover and a referee for helpful comments on the development and presentation of this article.

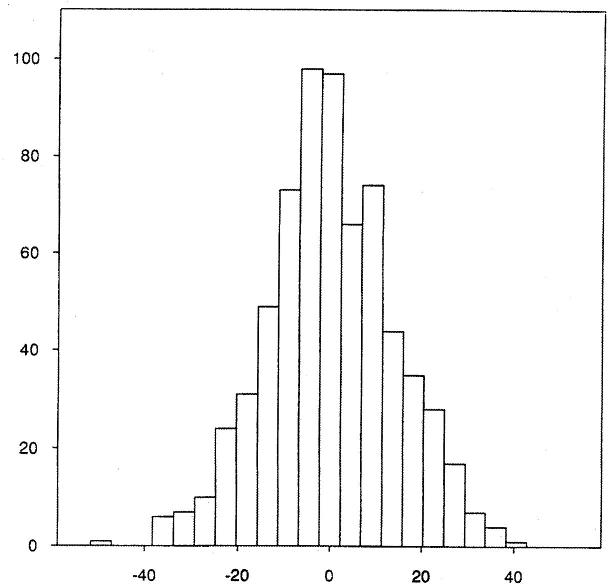

Figure 1. Histogram of the Margin of Victory Over the Point Spread (M). Goodness-of-fit tests indicate that the distribution of M is approximately Gaussian.

distribution (14 degrees of freedom—17 bins with two estimated parameters). The hypothesis of normality is also consistent with histograms having larger and smaller bin widths. Naturally the normal distribution is just an approximation. The variable M is concentrated on multiples of one-half, and integer values occur twice as often as noninteger values. This would not be the case if normality provided a more exact fit.

The Kolmogorov–Smirnov test is more powerful than the chi-squared test. The value of this test statistic is .913. Since the parameters of the normal distribution have been estimated from the data, the usual Kolmogorov–Smirnov significance levels do not apply. Using tables computed by Lilliefors (1967), we reject normality at the .05 significance level but not at the .01 significance level. This test is sensitive to the fact that the mode of the data does not match the mode of the normal distribution. We continue with the normal approximation despite this difference.

The results of the 1985 and 1986 seasons, collected after the initial analysis, provide additional evidence in favor of the normal approximation. The chi-squared statistic, using the parameters estimated from the 1981–1984 data, is 16.96. The p value for the chi-squared test is larger than .25. The Kolmogorov–Smirnov test statistic is .810, indicating a better fit than the original data set (the p value is approximately .10). More recent data may be used to verify that the approximation continues to hold.

3. THE PROBABILITY OF WINNING A GAME

What is the probability that a p-point favorite wins a football game? The natural estimate is the proportion of p-point favorites in the sample that have won their game. This procedure leads to estimates with large standard errors because of the small number of games with any particular point spread. The normal approximation of the previous section can be used to avoid this problem.

The probability that a team favored by p points wins the game is

$$\Pr(F > U \mid P = p) = \Pr(F - U - P > -P \mid P = p).$$

The argument in Section 2 shows that $M = F - U - P$ is approximately normal. A more detailed analysis indicates that normality appears to be a valid approximation for $F - U - P$ conditional on each value of P. This is difficult to demonstrate since there are few games with any particular value of P. A series of chi-squared tests were performed for games with similar point spreads. The smallest sample size was 69 games; the largest was 112 games. Larger bins were used in the chi-squared test (a bin width of 10.5 points instead of the 4.5 points used in Fig. 1) because of the size of the samples. Neighboring bins were combined so that each bin had an expected count of at least five. None of the eight tests was significant; the smallest p value was greater than .10. These tests seem to indicate that normality is an adequate approximation for each range of point spreads. If we apply normality for a particular point spread, p, then $F - U$ is approximately normal with mean p and standard deviation 13.861. The probability of winning a game is then computed as

$$\Pr(F > U \mid P = p) = 1 - \Phi\left(-\frac{p}{13.861}\right) = \Phi\left(\frac{p}{13.861}\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable.

The normal approximation for the probability of victory is given for some sample point spreads (the odd numbers) in Table 1. The observed proportion of p-point favorites that won their game, $\hat{P}$, and an estimated standard error are also computed. The estimates from the normal formula are consistent with the estimates made directly from the data. In addition, they are monotone increasing in the point spread. This is consistent with the interpretation of the point spread as a measure of the difference between two teams. The empirical estimates do not have this property. A linear approximation to the probability of winning is

$$\Pr(F > U \mid P = p) \approx .50 + .03p.$$

This formula is accurate to within .0175 for $|p| < 6$.

Table 1. The Normal Approximation and the Empirical Probability of Winning

Point spread	$\Pr(F > U \mid P)$	$\hat{P}$	Standard error
1	.529	.571	.071
3	.586	.582	.055
5	.641	.615	.095
7	.693	.750	.065
9	.742	.650	.107

4. APPLICATIONS

Conditional on the value of the point spread, the outcome of each game (measured by $F - U$) can be thought of as the sum of the point spread and a zero-mean Gaussian random variable. This is a consequence of the normal distribution of M. We assume that the zero-mean Gaussian random variables associated with different games are independent. Although successive football games are almost certainly not independent, it seems plausible that the random components (performance above or below the point spread) may be independent. The probability of a sequence of events is computed as the product of the individual event probabilities.

For example, the New York Giants were favored by two points in their first game and were a five-point underdog in their second game. The probability of winning both games is $\Phi(2/13.861)\Phi(-5/13.861) = .226$. Adding the probabilities for all $\binom{16}{k}$ sequences of game outcomes that have k wins leads to the probability distribution in Table 2. The point spreads used to generate Table 2 are:

$$2, -5, -6, 6, -3, -3.5, -5, 0, \\ -6, -7, 3, -1, 7, 3.5, -4, 9.$$

The Giants actually won nine games. This is slightly higher than the mean of the distribution, which is 7.7. Since this is only one observation, it is difficult to test the fit of the estimated distribution.

We use the results of all 28 teams over three years to assess the fit of the estimated distribution. Let

$$p_{ij}(x) = \text{probability that team } i \text{ wins } x \text{ games during season } j \text{ for } x = 0, \ldots, 16$$

$$F_{ij}(x) = \text{estimated cdf for the number of wins by team } i \text{ during season } j$$
$$= \sum_{t \le x} p_{ij}(t),$$

Table 2. Distribution of the Number of Wins by the 1984 New York Giants

Number of wins	Probability
0	.0000
1	.0002
2	.0020
3	.0099
4	.0329
5	.0791
6	.1415
7	.1928
8	.2024
9	.1642
10	.1028
11	.0491
12	.0176
13	.0046
14	.0008
15	.0001
16	.0000

and

$$X_{ij} = \text{observed number of wins for team } i \text{ during season } j$$

for $i = 1, \ldots, 28$ and $j = 1, 2, 3$. The index i represents the team and j the season (1981, 1983, or 1984). The distribution $p(\cdot)$ and the cdf $F(\cdot)$ represent the distribution of the number of wins when the normal approximation to the distribution of M is applied. Also, let U_{ij} be independent random variables uniformly distributed on (0, 1). According to a discrete version of the probability integral transform, if $X_{ij} \sim F_{ij}$, then $F_{ij}(X_{ij}) - U_{ij}p_{ij}(X_{ij})$ has the uniform distribution on the interval (0, 1). The U_{ij} represent auxiliary randomization needed to attain the uniform distribution. A chi-squared test is used to determine whether the transformed X_{ij} are consistent with the uniform distribution and therefore determine whether the X_{ij} are consistent with the distribution $F_{ij}(\cdot)$. The chi-squared statistic is computed from 84 observations grouped into 10 bins between 0 and 1. Four different sets of uniform variates were used, and in each case the data were found to be consistent with the uniform distribution. The maximum observed chi-squared statistic in the four trials was 13.1, between the .75 and the .90 quantiles of the limiting distribution. The actual records of NFL teams are consistent with predictions made using the normal approximation for the probability of winning a game.

Using the point spreads of the games for an entire season as input, it is possible to determine the probability of a particular outcome of the season. This type of analysis is necessarily retrospective since the point spreads for the entire season are not available until the season has been completed. To find the probability that a particular team qualifies for the postseason playoffs, we could consider all possible outcomes of the season. This would involve extensive computations. Instead, the probability of qualifying for the playoffs is estimated by simulating the NFL season many times. In a simulated season, the outcome of each game is determined by generating a Bernoulli random variable with probability of success determined by the point spread of that game. For each simulated season, the 10 playoff teams are determined. Six playoff teams are determined by selecting the teams that have won each of the six divisions (a division is a collection of four or five teams). The winning team in a division is the team that has won the most games. If two or more teams in a division are tied, then the winner is selected according to the following criteria: results of games between tied teams, results of games within the division, results of games within the conference (a collection of divisions), and finally random selection. It is not possible to use the scores of games, since scores are not simulated. Among the teams in each conference that have not won a division, the two teams with the most wins enter the playoffs as "wildcard" teams (recently increased to three teams). Tie-breaking procedures for wildcard teams are similar to those mentioned above.

The 1984 NFL season has been simulated 10,000 times. For each team, the probability of winning its division

Table 3. Results of 10,000 Simulations of the 1984 NFL Season

Team	Pr(win division)	Pr(qualify for playoffs)	1984 actual result
National Conference—Eastern Division			
Washington	.5602	.8157	division winner
Dallas	.2343	.5669	
St. Louis	.1142	.3576	
New York	.0657	.2291	wildcard playoff team
Philadelphia	.0256	.1209	
National Conference—Central Division			
Chicago	.3562	.4493	division winner
Green Bay	.3236	.4170	
Detroit	.1514	.2159	
Tampa Bay	.1237	.1748	
Minnesota	.0451	.0660	
National Conference—Western Division			
San Francisco	.7551	.8771	division winner
Los Angeles	.1232	.3306	wildcard playoff team
New Orleans	.0767	.2291	
Atlanta	.0450	.1500	
American Conference—Eastern Division			
Miami	.7608	.9122	division winner
New England	.1692	.4835	
New York	.0607	.2290	
Buffalo	.0051	.0248	
Indianapolis	.0042	.0205	
American Conference—Central Division			
Pittsburgh	.4781	.5774	division winner
Cincinnati	.3490	.4574	
Cleveland	.1550	.2339	
Houston	.0179	.0268	
American Conference—Western Division			
Los Angeles	.4555	.7130	wildcard playoff team
Seattle	.2551	.5254	wildcard playoff team
Denver	.1311	.3405	division winner
San Diego	.1072	.2870	
Kansas City	.0511	.1686	

has been computed. The probability of being selected for the playoffs has also been determined. The results appear in Table 3. Each estimated probability has a standard error that is approximately .005. Notice that over many repetitions of the season, the eventual Super Bowl champion San Francisco would not participate in the playoffs approximately 12% of the time.

5. SUMMARY

What is the probability that a team favored to win a football game by p points does win the game? It turns out that the margin of victory for the favorite is approximated by a Gaussian random variable with mean equal to the point spread and standard deviation estimated at 13.86. The normal cumulative distribution function can be used to compute the probability that the favored team wins a football game. This approximation can also be used to estimate the distribution of games won by a team or the probability that a team makes the playoffs. These results are based on a careful analysis of the results of the 1981, 1983, and 1984 National Football League seasons. More recent data (1985 and 1986) indicate that the normal approximation is valid outside of the original data set.

[*Received June 1989. Revised December 1989.*]

REFERENCES

Amoako-Adu, B., Marmer, H., and Yagil, J. (1985), "The Efficiency of Certain Speculative Markets and Gambler Behavior," *Journal of Economics and Business*, 37, 365–378.

Harville, D. (1980), "Predictions for National Football League Games via Linear-Model Methodology," *Journal of the American Statistical Association*, 75, 516–524.

Lilliefors, H. W. (1967), "On the Kolmogorov–Smirnov Test for Normality With Mean and Variance Unknown," *Journal of the American Statistical Association*, 62, 399–402.

Pankoff, L. D. (1968), "Market Efficiency and Football Betting," *Journal of Business*, 41, 203–214.

Stefani, R. T. (1977), "Football and Basketball Predictions Using Least Squares," *IEEE Transactions on Systems, Man, and Cybernetics*, 7, 117–121.

——— (1980), "Improved Least Squares Football, Basketball, and Soccer Predictions," *IEEE Transactions on Systems, Man, and Cybernetics*, 10, 116–123.

Thompson, M. L. (1975), "On Any Given Sunday: Fair Competitor Orderings With Maximum Likelihood Methods," *Journal of the American Statistical Association,* 70, 536–541.
Tryfos, P., Casey, S., Cook, S., Leger, G., and Pylypiak, B. (1984), "The Profitability of Wagering on NFL Games," *Management Science,* 30, 123–132.
Vergin, R. C., and Scriabin, M. (1978), "Winning Strategies for Wagering on Football Games," *Management Science,* 24, 809–818.
Zuber, R. A., Gandar, J. M., and Bowers, B. D. (1985), "Beating the Spread: Testing the Efficiency of the Gambling Market for National Football League Games," *Journal of Political Economy,* 93, 800–806.