The World Series Competition

Frederick Mosteller

Journal of the American Statistical Association, Vol. 47, No. 259. (Sep., 1952), pp. 355-380.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28195209%2947%3A259%3C355%3ATWSC%3E2.0.CO%3B2-U

Journal of the American Statistical Association is currently published by American Statistical Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/astata.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org/
Tue Aug 8 11:15:17 2006

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 259 SEPTEMBER 1952 *Volume 47*

THE WORLD SERIES COMPETITION*

FREDERICK MOSTELLER
Harvard University

Suppose a number of pairs of teams or products are compared on the basis of n binomial trials. Although we cannot know from the outcomes which teams or products were actually better, we wish to estimate the average probability that the better team or product wins a given trial, and thus to measure the discrimination provided by our test. World Series data provide an example of such comparisons. The National League has been outclassed by the American League teams in a half-century of World Series competition. The American League has won about 58 per cent of the games and 65 per cent of the Series. The probability that the better team wins the World Series is estimated as 0.80, and the American League is estimated to have had the better team in about 75 per cent of the Series.

INTRODUCTION

If we compare pairs of teams, products, drugs, or persons on the basis of a fixed number of binomial trials, and identify the member of the pair that wins the majority of trials as the better, we may be in error. For example, if the members of a pair are evenly matched, a decision on the basis of performance is equivalent to coin-flipping. On the other hand, if one of a pair is actually better, it is more likely that the better member will also be the winner. If we carry out such comparisons on many pairs under roughly comparable conditions, it is of

* This work was facilitated by support from the Laboratory of Social Relations, Harvard University.

It is a pleasure to acknowledge the numerous suggestions and criticisms made by various friends, though I have not availed myself of all their advice. Harry V. Roberts (University of Chicago) and Howard L. Jones (Illinois Bell Telephone Company) were so prodigal with suggestions that the paper has more than doubled in length since its first draft. I wish to thank K. A. Brownlee (University of Chicago), William G. Cochran (Johns Hopkins University), Herbert T. David (University of Chicago), Joseph L. Hodges (Universities of Chicago and California), and William H. Kruskal (University of Chicago) for their reading of the manuscript in early draft. My greatest debt is to Mrs. Doris Entwisle (Harvard University) who assisted with the computations, with gathering the data, and with the preparation of the manuscript.

interest to estimate the over-all effectiveness of the decision technique in the kinds of situations that have occurred in practice. Data from the World Series are available to illustrate many facets of this type of problem.

About World Series time each year most fans are wondering which team will win. The author is no exception, but he has also wondered about another question: Will the Series be very effective in identifying the better team?

By assuming that the probability that a particular team wins single games is fixed throughout a series, we can readily illustrate our point. Suppose team A and team B are matched in a series, and that the probability that A wins single games is $p=0.52$. Then team A is the better team, because it has the higher probability of winning single games, but in a short series team A may have little better than a 50-50 chance of winning. For example, in a 3-game series, team A can win by winning the first two games, the first and third, or the second and third. The probabilities of these three outcomes are $(0.52)^2$, $(0.52)(0.48)(0.52)$, and $(0.48)(0.52)^2$, respectively. These values add up to 0.529984, or about 0.530, which means that team A's chance to win the 3-game series is not a very big improvement over its chance to win a coin-flipping contest. The corresponding computation for a seven game series shows a small improvement in the probability that team A wins; the value is about 0.544. In principle, the longer the series, the better chance team A has to win. But practical considerations always limit the lengths of such comparison series.

In its simplest form, the question of the effectiveness of series in identifying the better team can be raised with respect to League play-offs. The American League uses a one-game play-off, while the National League uses a three-game play-off to settle first-place ties, and thus to choose their representative in the World Series. Intuitively one might suppose that the National League's three-game play-off would be more sensible, because the longer the series, the better chance the more skillful team has to emerge the victor, as the example in the previous paragraph suggests. On the other hand, if the teams have played 154 games to a dead heat, there is considerable evidence available to suggest that these teams' chances of winning single games are roughly equal, and therefore that the flip of a coin will be nearly as sensitive in deciding which team is better as the actual play of an additional one- or three-game series. An exception to this might occur when one of the tied teams had been improving as the season progressed, while the play of the other was falling off. Then the improving team would be expected to benefit from the longer series on the average, because its probability of winning single games would be larger than 0.5.

As a simple model for the League play-offs, we might suppose that one team has probability p of winning single games, and that the other has probability $1-p$ and further, that p remains the same for all games in the series under consideration. The reader may easily think of arguments against the assumption of the constancy of p from game to game (pitchers and ball parks are examples of variables that could contribute to variation in p). We will provide some evidence on this issue later. Corresponding to the probability p of winning a single game, we need the probability of winning an n-game series. We will call this probability $S(p, n)$. By an n-game series we mean a series that can last n games. Thus $n=7$ in recent World Series, though the number of games played may be as few as four. And, in general, an n-game series may be stopped considerably short of n games because play stops as soon as one team has won a majority. In acceptance sampling language, we would call this "truncated single sampling." Although the point is obvious, we note that $S(p, n)$ is the same whether the series is played to the full n games, or only played until a majority is won by one team. This is clear because no decisions about the name of the winner are changed once a particular team has a majority. This fact is useful because it means $S(p, n)$ can be computed directly from the binomial expansion as if all games had been played, instead of from the less familiar and less well tabulated distribution appropriate to truncated single sampling. With this "fixed p" model, slight deviations from equal probabilities of winning single games will lead to very little gain in the probability that the better team wins in a three-game play-off as compared with a one-game play-off. If the probability of the better team winning a single game is somewhat more than half, say $p=\frac{1}{2}+\epsilon$, $(\frac{1}{2}\geqq\epsilon>0)$, then the probability it wins a three-game series is

$$\begin{aligned} S(p, 3) &= p^3 + 3p^2(1-p) = p^2(3-2p) \\ &= (\tfrac{1}{2}+\epsilon)^2(2-2\epsilon) \\ &= \tfrac{1}{2} + \tfrac{3}{2}\epsilon - 2\epsilon^2, \end{aligned}$$

and the increase in the probability of correctly choosing the better team as we go from the one-game to the three-game play-off is

$$S(p, 3) - S(p, 1) = \frac{\epsilon}{2} - 2\epsilon^3.$$

If $\epsilon=0.01$ $(p=0.51)$, the gain in the probability S is essentially 0.005

for a three-game as compared with a one-game series. This means that in 200 one-game play-offs, the better team could expect to win 102 play-offs, but in 200 three-game play-offs, the better team could expect to win 103 play-offs, a scarcely noticeable improvement. There are few data for comparing play-off series within the major leagues, because ties are rare. Further, one-game play-offs would not provide any information on the point at hand. We will not therefore be able to pursue this question of differences between play-off teams to any conclusion, and so we proceed directly to consideration of World Series, where data are more plentiful. As a simple means of examining the power of a series for identifying the better team we provide Table 1. In Table 1 (suggested by K. A. Brownlee) we give the probability $S(p, n)$ that the better team wins an n-game series for $n=1, 3, 5, 7, 9$, and for various probabilities p. We also give the probability that the better team wins, or ties, in even-sized series $n=2, 4, 6, 8$.

COMPARISON OF MAJOR LEAGUES

One way to approach the question of equality of teams entering the World Series is to compare the two Leagues. Altogether there have been 44 seven-game Series from 1905 to 1951, and 4 nine-game Series (1903, 1919–1921, no Series in 1904), for a total of 275 games actually played. The American League has won 159 of these games, or 57.82 per cent. In any year i, the American League team is assumed to have had a probability p_i of winning single games. Then over the 48 years of World Series games there would be an average probability, say $p=\sum p_i/48$. If we adopt this view, then the average proportion of games won, 0.5782, is an estimate of p. Some might object to this estimate of p because the sampling is truncated (a seven-game Series is stopped when one team wins 4 games), or because we would have a better estimate if we averaged over the number of Series rather than the total number of games.

These objections are both reasonable, but it happens to turn out that the use of an estimate suited to truncated sampling gives almost exactly the same numerical result. We introduce this estimate because we plan to base a later argument on it. An unbiased estimate appropriate to truncated sampling can be obtained in the following manner:[1]

In a year when the American League wins, the estimate of p is taken to be

[1] M. A. Girshick, F. Mosteller, and L. J. Savage, "Unbiased estimates of certain binomial sampling problems with applications," *Annals of Mathematical Statistics*, 17 (1946), 20.

$$\frac{c-1}{c+x-1},$$

where c is the number of games it takes to win a Series and x is the

TABLE 1

PROBABILITY $S(p, n)$ THAT THE BETTER TEAM WINS AN n-GAME SERIES, WHEN ITS PROBABILITY OF WINNING SINGLE GAMES IS p

p	$S(p, 1)$	$S(p, 3)$	$S(p, 5)$	$S(p, 7)$	$S(p, 9)$
0.50	0.500	0.500	0.500	0.500	0.500
0.55	0.550	0.575	0.593	0.608	0.621
0.60	0.600	0.648	0.683	0.710	0.733
0.65	0.650	0.718	0.765	0.800	0.828
0.70	0.700	0.784	0.837	0.874	0.901
0.75	0.750	0.844	0.896	0.929	0.951
0.80	0.800	0.896	0.942	0.967	0.980
0.85	0.850	0.939	0.973	0.988	0.994
0.90	0.900	0.972	0.991	0.997	0.999
0.95	0.950	0.993	0.999	1.000	1.000
1.00	1.000	1.000	1.000	1.000	1.000

PROBABILITY THAT THE BETTER TEAM WINS (W) OR TIES (T) A SERIES OF AN EVEN NUMBER OF GAMES, WHEN ITS PROBABILITY OF WINNING SINGLE GAMES IS p

p	$n=2$		$n=4$		$n=6$		$n=8$	
	W	T	W	T	W	T	W	T
0.50	0.250	0.500	0.312	0.375	0.344	0.312	0.363	0.273
0.55	0.302	0.495	0.391	0.368	0.442	0.303	0.477	0.263
0.60	0.360	0.480	0.475	0.346	0.544	0.276	0.594	0.232
0.65	0.422	0.455	0.563	0.311	0.647	0.235	0.706	0.188
0.70	0.490	0.420	0.652	0.265	0.744	0.185	0.806	0.136
0.75	0.562	0.375	0.738	0.211	0.831	0.132	0.886	0.087
0.80	0.640	0.320	0.819	0.154	0.901	0.082	0.944	0.046
0.85	0.722	0.255	0.890	0.098	0.953	0.041	0.979	0.018
0.90	0.810	0.180	0.948	0.049	0.984	0.015	0.995	0.005
0.95	0.902	0.095	0.986	0.014	0.998	0.002	1.000	0.000
1.00	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000

number won by the National League; when the American League loses, the estimate is

$$\frac{y}{c+y-1}$$

where y is the number won by the American League. In our data the value of c is either 4 or 5 depending on whether a seven- or nine-game Series is used. The summary table (Table 2) shows the Series outcomes and the estimates corresponding to these outcomes. In addition, the four estimates of American League p_i for the nine-game Series are 4/7, 3/7, 4/6, 3/7. The average of the 48 estimates is 57.80 per cent, providing an estimate of 100 p that is scarcely different from the more naive per cent won. If we think of the team representing the American League as having a possibly different p_i in each Series, then 57.80 per

TABLE 2

OUTCOMES OF THE 44 SEVEN-GAME SERIES*

Games Won		Estimate of P for A.L.	Frequency
N.L.	A.L.		
4	0	0	3
4	1	1/4	4
4	2	2/5	1
4	3	3/6	7
3	4	3/6	4
2	4	3/5	10
1	4	3/4	9
0	4	1	6
		Total	44

* Data from *The World Almanac* 1952, New York *World-Telegram*, Harry Hansen (Ed.), p. 821. These data have also been checked in *The Official Encyclopedia of Baseball, Jubilee Edition*, Hy Turkin and S. C. Thompson (New York: A. S. Barnes Co., 1951).

cent is, in a reasonable sense, an estimate of the average probability of winning single games over the years.

Another question is whether the American League has done significantly better than the National League. We could check, merely on the basis of the number of Series won. The American League has won 31 of 48 Series, and under the null hypothesis of $p=\frac{1}{2}$, the probability[2] of 31 or more successes is 0.0297, or a two-sided probability of about 0.06.

Although it has little to do with the main discussion, another frequently-asked question is whether the American League has been improving through the years. To answer this question, at least partially, we break the 48 Series into four sets of 12, chronologically in Table 3. There is a slight but not statistically significant trend in the data. Of

[2] *Tables of the Binomial Probability Distribution*, National Bureau of Standards, Applied Mathematics Series 6 (U. S. Government Printing Office, 1950), p. 375.

TABLE 3

NUMBER OF SERIES WON BY THE AMERICAN LEAGUE IN 12 YEAR INTERVALS

Years	1903–15*	1916–27†	1928–39	1940–51	Total
Series Won by A.L.	7	7	9	8	31 of 48

* No Series in 1904.
† Includes National League victory in 1919, year of "Blacksox Scandal."

course there is one notable trend by the New York Yankees (A.L.) suggested by Table 4. Just what sort of significance test should be applied to a team chosen on the basis of its notable record, is an issue not at present settled by statistical theory, so we leave Table 4 without further analysis.

TABLE 4

TABLE OF WORLD SERIES WON AND LOST BY YANKEES BY YEARS

Years	Won	Lost	Totals
1903–27	2	3	5
1928–51	11	1	12
Totals	13	4	17

ESTIMATING THE PROBABILITY THAT THE BETTER TEAM WINS SINGLE GAMES

Since the American League has done rather well through the years, we will reject the idea that the league champions are equally matched ($p=\frac{1}{2}$) when they appear for the Series. How closely are they matched? Suppose each year we knew the single game probability p_i for the better team, then the average of such p's could be a measure of how well the teams were matched. We expect the average p to be greater than 0.5. Indeed in the present case our estimate for the average p for better teams should be higher than the 57.8 per cent we obtained for the American League, because we suspect that the American League team was in some years *not* the "better team." There should be no confusion here between the "winning team" and the "better team." The "winning team" is the team that wins the Series. The "better team" is the team with the higher probability of winning single games, whether or not it actually wins the Series. We anticipate that the better team sometimes loses a Series, just as the league champion loses single games to the last-place team within the league during the season.

Model A: The better team has the same p each year. The author has not discovered any good way of estimating the average p for better teams without making further unrealistic assumptions (the lack of reality may not be important, because the results may not be sensitive to the assumptions). What has been assumed in this section is that every year the better team has the *same* probability $p>\frac{1}{2}$ of winning single games. Of course, we cannot identify the better team in any particular Series, but we may by arithmetic manipulation derive an estimate from our half-century of data. For this purpose, we will neglect the four nine-game Series, because they cause considerable arithmetic trouble. Our data and Model A can be summarized in Table 5. The

TABLE 5

GAMES WON (SEVEN-GAME SERIES ONLY)

Winner	Loser	Frequency	Theoretical Proportion
4	0	9	p^4+q^4
4	1	13	$4p^4q+4pq^4$
4	2	11	$10p^4q^2+10p^2q^4$
4	3	11	$20p^4q^3+20p^3q^4$
		Total 44	1

algebraic expressions in the right-hand column are not the usual terms in the expansion of the binomial $(p+q)^7$ because we are working with truncated single sampling. On each line the first algebraic term represents the probability that the better team wins the Series in the number of games appropriate to the line, while the second term similarly represents the probability that the poorer team wins the Series. The sum of these two terms represents the total probability that the Series is won in the pattern of games given in the first two columns. Thus in the third line $10p^4q^2$ is the probability that the better team wins the Series in exactly six games.

If we represent a win by the better team as a B and a win by the poorer team as a W, the following 10 ways for the better team to win in exactly 6 games exhaust the possibilities:

BBBWWB	BWBWBB
BBWBWB	WBBWBB
BWBBWB	BWWBBB
WBBBWB	WBWBBB
BBWWBB	WWBBBB

These 10 ways correspond to the coefficient 10 in $10p^4q^2$. The factor p^4q^2 arises from the fact that we must have exactly 4 wins by the better team, whose single game probability is p, and exactly 2 wins by the poorer team whose single game probability is $q=1-p$. Similar computations account for the coefficients and powers of p and q corresponding to the other Series outcomes.

To estimate p for better teams we should not take merely the total number of games won by Series-winning teams and divide by the grand total of games. This will clearly overestimate the p-value, as an example with $p=\frac{1}{2}$ will show.

In Table 6 based on hypothetical data for equally matched teams, it

TABLE 6

EXPECTED RESULTS FOR 64 SERIES, $p=\frac{1}{2}$

Games	Won	Theoretical Frequency
4	0	8
4	1	16
4	2	20
4	3	20
	Total	64

turns out that the Series-winning teams won 256 of a total of 372 games, or 68.8 per cent of the games—but an estimate of $p=0.688$ would be rather far from the actual $p=0.500$. On the other hand, in the actual Series results, Table 5, the per cent of games won by the Series-winning team is only 72.1 (176 of a total of 244) which seems rather close to 68.8, so perhaps the previous assumption that the teams are unevenly matched is not in line with the facts. To investigate the facts, we need an estimation process. Three estimates seem reasonable:

1) Use the theoretical distribution to obtain a formula for the expected number of games won by the losing team in a 7-game series. This average will be a function of p. Equate this theoretical average to

the observed average and solve for p. This is the method of moments applied to the sample mean.[3]

2) Obtain the maximum likelihood estimate of p.

3) Obtain the minimum chi-square estimate of p.

Method 1) is by far the easiest computationally.

Method 1. We wish to obtain the average number of games won by the Series-losing team in terms of p. We multiply the theoretical proportions from Table 5 by the number of games won by the loser and add. This operation gives

$$\begin{aligned} A &= \text{Average number of games won per Series by the Series-loser} \\ &= 1(4p^4q + 4pq^4) + 2(10p^4q^2 + 10p^2q^4) + 3(20p^4q^3 + 20p^3q^4) \qquad (1) \\ &= 4pq[(p^3 + q^3) + 5pq(p^2 + q^2) + 15p^2q^2]. \end{aligned}$$

We note that

$$p^3 + q^3 = p^3 + 3p^2q + 3pq^2 + q^3 - 3p^2q - 3pq^2 = 1 - 3pq$$

and

$$p^2 + q^2 = p^2 + 2pq + q^2 - 2pq = 1 - 2pq.$$

Substituting these relations in (1) gives

$$\begin{aligned} A &= 4pq[1 - 3pq + 5pq(1 - 2pq) + 15p^2q^2] \\ A &= 4pq[1 + 2pq + 5p^2q^2]. \end{aligned} \qquad (2)$$

The value of A attains its maximum when $p = \frac{1}{2}$ as we would anticipate. In the 44 seven-game Series, the average number of games per Series won by the Series-loser was 1.5455. If we set this equal to A in equation (2) we can solve directly a cubic in pq and then a quadratic in p, or we might go directly to the 6th degree equation in p. Moderation and discretion suggest that we just substitute a few values of p in the expression, and see what values of p lead to outcomes close to the average wins of the Series-loser. We get

[3] It was suggested by William Kruskal (personal communication) that other estimates based on the method of moments seem equally plausible. He suggests as an example the statistic

$$\frac{\text{Number of games won by loser}}{\text{Number of games in Series}}.$$

Calculations similar to those in the text give the average value of this statistic as

$$B = 4pq\left[\frac{1}{5} + \frac{7}{30}pq + \frac{10}{21}p^2q^2\right].$$

When this is equated to its observed average (about 0.25), the estimate of p turns out slightly higher than 0.65.

p	A = Average Wins Expected by Loser
0.5	1.8125
0.6	1.6973
0.6500	1.5596
0.6667	1.5034
0.7	1.3780

Linear interpolation gives the estimate of p as 0.6542. Of course, the uncertainty of the estimate makes the use of this many decimal places quite misleading. On the basis of the evidence thus far available, then, the teams entering the World Series seem to be matched at about 65–35 for single games.

Method 2. If $P(0)$, $P(1)$, $P(2)$, $P(3)$ are the probabilities that the Series-losing team wins 0, 1, 2, or 3 games respectively in a Series, then the maximum likelihood approach involves finding the value of p that maximizes

$$[P(0)]^9[P(1)]^{13}[P(2)]^{11}[P(3)]^{11}.$$

The numbers 9, 13, 11 and 11 are the frequencies tabulated in Table 5 and the $P(x)$ are given in algebraic form in the Theoretical Proportions column in Table 5. Although tedious, this maximization was done, and the estimate obtained was 0.6551, encouragingly close to that obtained from Method 1.

Method 3. Finally the chi-square to be minimized was

$$\chi^2 = \frac{[9 - 44P(0)]^2}{44P(0)} + \frac{[13 - 44P(1)]^2}{44P(1)} + \frac{[11 - 44P(2)]^2}{44P(2)} + \frac{[11 - 44P(3)]^2}{44P(3)},$$

where $P(x)$, $x=0, 1, 2, 3$ has the same definition as before. Here the terms $44P(x)$ are the "expected numbers" for the usual chi-square formula. The p-value minimizing this chi-square turned out to be 0.6551.

The following table summarizes the results for the three methods.

Method	*Estimate*
Average Wins by Series Loser	0.6542
Maximum Likelihood	0.6551
Minimum Chi-square	0.6551

Presumably the close agreement between the minimum chi-square method and the maximum likelihood method is partly an accident of the particular empirical data, and partly owing to 44 being a fairly large number. For contingency problems equivalent to the present one, Cramér[4] points out that "the modified chi-square minimum method is . . . identical with the maximum likelihood method." However the "modified chi-square minimum method" neglects certain terms of the form:

$$\frac{[n_x - nP(x)]^2}{2n[P(x)]^2}, \qquad \begin{aligned} n &= \text{total number} \\ n_x &= \text{observed number of } x\text{'s} \end{aligned}$$

when the partial derivatives appropriate to minimizing chi-square are set equal to zero. We would not in general expect such a term to vanish, but the closeness of agreement suggests to the author that in the future he will usually prefer the easier maximum likelihood to the more tedious minimum chi-square when totals even as small as 44 are involved.

Using the estimate obtained by the minimum chi-square method, the observed value of χ^2 turns out to be 0.222, which for two degrees of freedom is fairly small, the probability of a larger value of chi-square being about 0.89, so the fit is rather good.

HOW OFTEN DOES THE BETTER TEAM WIN THE SERIES?

Based on past experience then, a reasonable estimate of the average probability of winning single games for the better team is about 0.65 according to Model A. Using this value, we can compute the probability of the better team winning a seven-game World Series as about $S(0.65, 7) = 0.80$ (see Table 1), so the better team would win about four out of five Series. If we push our assumptions to an extreme we might even estimate that the American League has had the better team about 75 per cent of the time. We can obtain this number by assuming that the American League had the better team a fraction of the time x, and recall that the American League won 31 of 48 Series. Then equating expected and observed proportions of Series won we have:

$$\begin{aligned} 0.80x + 0.20(1 - x) &= \frac{31}{48} = 0.646 \\ 0.60x &= 0.446 \\ x &= 0.743. \end{aligned}$$

Alternatively, we could use games won rather than Series won. Using our estimate of 0.65 as the single game probability for the better team,

[4] Harald Cramér, *Mathematical Methods of Statistics* (Princeton, N. J.: Princeton University Press, 1946), p. 426.

and recalling that the American League has won 0.578 of the games we have:

$$0.65x + 0.35(1 - x) = 0.578$$
$$0.30x = 0.228$$
$$x = 0.76.$$

These two methods both give estimates of about 75 per cent as the percentage of years in which the American League has had the better team. If the American League has had the better team 75 per cent of the time, or in about $0.75(48) = 36$ Series, these 36 better teams could expect to win about $0.80 \times 36 = 29$ Series and lose about $36 - 29 = 7$ Series. Similar computations suggest that the American League has had the poorer team 12 times, and that these poorer teams could expect to win two Series from their better National League opponents. The discussion just given shows why we do not use the per cent of Series won (65) as an estimate of the per cent of times the American League has had the better team. The side more often having the better team will suffer most in actual play due to lack of discrimination of the 7-game Series. If one League always had the better team, it would still lose a good many Series unless its single game p were quite high.

It might be supposed that the estimate of 80 per cent for the probability that the better team would win the Series would depend sensitively on the Model A assumption of a constant value of p for the better team in every year. The reader might be willing to accept the idea that our estimate of 0.65 as an average for better teams is reasonable, but feel that since there is surely a distribution of p values for better teams from year to year, the average of the $S(p_i, 7)$ may not be close to $S(p, 7)$. (It will be recalled that $S(p, 7)$ is the probability that a team with a single-game probability of p will win a 7-game Series.) For example, let $p_1 = 0.50$, $p_2 = 0.90$, and then the average $p = (0.50 + 0.90)/2 = 0.70$, while $S(0.50, 7) = 0.50$, $S(0.90, 7) = 1.00$, but $S(0.70, 7) = 0.87$ instead of $0.75 = \frac{1}{2}(0.50 + 1.00)$. This example shows that

$$\frac{\sum_{i=1}^{k} S(p_i, n)}{k} \neq S(p, n),$$

$$p = \frac{\sum_{i=1}^{k} p_i}{k}.$$

How important this objection is depends on the linearity of $S(p, 7)$ over the dense part of the distribution of p's from year to year. A graph of $S(p, 7)$ reveals that S is approximately linear in p over the range from $p=0.50$ to $p=0.75$. We would suppose that in World Series competition it would be relatively rare that single game p's exceeded 0.75, and therefore we feel that the lack-of-linearity argument is not a strong one against this estimate of 0.80 as the average probability that the better team wins the Series.

AN ALTERNATIVE METHOD OF ESTIMATION

Model B. Fixed p's within Series, but normally distributed p's from year to year. There is another way of estimating the number of times the American League has had the better team. We can take the view that each year the American League team has a true but unknown probability of winning single games p_i. For each of these yearly p_i we have an unbiased estimate $\hat{p}_i$; the distribution of these estimates was given in Table 2. If we let the observed mean of the estimates be $\bar{p}$, we can compute the sum of squares of deviations of the estimates $\hat{p}_i$ from $\bar{p}$. This sum of squares can be partitioned into two parts. One part has to do with $\sigma^2(\hat{p}_i)$, the variation of $\hat{p}_i$ around its true value p_i, and the other with $\sigma^2(p_i)$, the variation of the true p_i around their true mean p. Such a partition is standard practice in analysis of variance. We need to define

$$\begin{aligned} \sigma^2(\hat{p}_i) &= E(\hat{p}_i - p_i)^2, \qquad i = 1, 2, \cdots, n. \\ \sigma^2(p_i) &= \frac{\sum_{i=1}^{n} (p_i - p)^2}{n}, \end{aligned} \tag{3}$$

where E is the expected value operator. We need the expected value of $\sum(\hat{p}_i-\bar{p})^2$, it is

$$\begin{aligned} E\left[\sum_{i=1}^{n} (\hat{p}_i - \bar{p})^2\right] &= E\left[\sum_{i=1}^{n} \hat{p}_i^2 - n\bar{p}^2\right] \\ &= E\left[\sum_{i=1}^{n} \hat{p}_i^2 - \frac{(\sum \hat{p}_i)^2}{n}\right]. \end{aligned} \tag{4}$$

Recalling that

$$\begin{aligned} E(\hat{p}_i^2) &= \sigma^2(\hat{p}_i) + p_i^2, \qquad i = 1, 2, \cdots, n \\ E(\hat{p}_i\hat{p}_j) &= p_ip_j, \qquad i \neq j \end{aligned} \tag{5}$$

and using these results in (4) we have the well-known result expressed in words and symbols:

$$\text{Total Sum of Squares} = \text{Within Years Sum of Squares} + \text{Between Years Sum of Squares}$$

$$E\sum_{i=1}^{k} (\hat{p}_i - \bar{p})^2 = \frac{n-1}{n} \sum \sigma^2(\hat{p}_i) + n\sigma^2(p_i). \tag{6}$$

The value of $\sum(\hat{p}_i - \bar{p})^2$ can be computed from Table 2. We would like to know $\sigma^2(p_i)$ as an aid in estimating the average p for better teams. To get an estimate of $\sigma^2(p_i)$ we will have to estimate $\sum\sigma^2(\hat{p}_i)$. By a procedure like that used in deriving the average number of games won by the Series-loser (equation (2)) we can show that

$$\sigma^2(\hat{p}_i) = \frac{p_i q_i}{20} [5 - 3p_i q_i - 4p_i^2 q_i^2], \qquad q_i = 1 - p_i. \tag{7}$$

The derivation of (7) is lengthy, but is shown in the Appendix. Naturally this variance depends on the true p_i, but its value does not change rapidly in the neighborhood of $p_i = \frac{1}{2}$. Therefore we propose to estimate this error variance by evaluating $\sigma^2(\hat{p}_i)$ at the average value of the p_i's. For the 44 seven-game Series $\bar{p}$ is 0.583. Substituting this p-value in the formula for the variance (7) gives

$$\sigma^2(\hat{p}_i) = 0.0490 \text{ (error variance).}$$

We use this same value of $\sigma^2(\hat{p}_i)$ for all years. The total sum of squares $\sum(\hat{p}_i - \bar{p})^2$ for the 44 Series is 2.8674. The estimated between years sum of squares $\sum(p_i - p)^2$ is

$$2.8674 - 43(0.0490) = 0.7604.$$

Dividing this by 44 gives an estimate of the variance of the true p_i's from year to year as 0.0173, or an estimated standard deviation of p-values $\sigma(p_i) = 0.1315$. The departure of the observed average, 0.583, from 0.500 in standard deviation units is 0.63. If we assume that the distribution of true p's is normal, we estimate the percentage of times the American League had the better team to be 74 per cent. This result is close to our previous estimate of 76 per cent.

It has been suggested by Howard L. Jones (personal communication) that we might obtain an improved estimate $\sum\sigma^2(\hat{p}_i)/n$ by averaging the formula of equation (7) over the normal distribution with mean 0.583 and standard deviation 0.1315. When this was done the value

0.0455 was obtained instead of 0.0490. The new residual variance would be $2.8674-43(0.0455)=0.9109$. Dividing 0.9109 by 44 gives a corrected estimate of the variance of the true p_i's as 0.0207, or an estimated standard deviation $\hat{\sigma}(p_i)=0.1439$. This adjusted standard deviation can be used as before for estimating the proportion of p_i's higher than 0.500. We have the departure from the mean in standard deviation units as $(0.583-0.500)/0.1439$ or 0.58 units, which corresponds to a proportion of 72 per cent on a normal distribution. Thus a better approximation for the proportion of times the American League has had the better team using Model B is 72 per cent. Again the result is not far from our Model A estimate of 76 per cent.

If we make further use of this normality assumption, we can also estimate the average single-game probability for the better team. We break the assumed normal distribution of true p_i's for the American League into two parts. One part is the truncated normal for which $p_i>\frac{1}{2}$ (American League better), the other is the part for which $p_i<\frac{1}{2}$ (National League better). Then we obtain the average p_i for the American League when it is better, and for the National League when it is better, and weight each by the relative frequency it represents. The final result is an estimate of the average single-game p for the better team. The integration is shown in the Appendix. When $\sigma(p_i)$ is taken as 0.1315, the estimate is 0.626, but the improved estimate of $\sigma(p_i)$ as 0.1439 gives the final estimate as 0.634, which can be compared with our Model A estimate of 0.655.

TESTS OF THE BINOMIAL ASSUMPTIONS

We have emphasized the binomial aspects of the model. The twin assumptions needed by a binomial model are that throughout a World Series a given team has a fixed chance to win each game, and that the chance is not influenced by the outcome of other games. It seems worthwhile to examine these assumptions a little more carefully, because any fan can readily think of good reasons why they might be invalid. Of course, strictly speaking, all such mathematical assumptions are invalid when we deal with data from the real world. The question of interest is the degree of invalidity and its consequences. Obvious ways that the assumptions might be invalid are:

1) A team might be expected to do better "at home" than it does "away," and this would negate a constant probability because even the shortest Series may be played in two places. This possibility is strongly suggested both by intuition and by an examination of the results of regular season games in the major leagues. That it would hold for World Series games is not a foregone conclusion.

2) Winning a game might influence the chance of winning the next game, i.e., there may be serial correlation from game to game.

To examine the first of these issues, we collected the detailed results of four games in each Series. We chose four because that represents the least number played. Games were chosen as early as possible in each Series to provide two games played by each team in an "at-home" capacity. In a seven-game Series we ideally find the first two games played at the National League (American League) park, the next three at the American League (National League) park, and the last two in the National League (American League) park. When we observed this pattern, we used the first four games played. This ideal pattern was not actually used as often as one might suppose. Sometimes teams alternated parks after each game, when extensive travelling was not required. Sometimes both teams used the same park as did the New York Giants and the New York Yankees in the early days, or the St. Louis Cardinals and the St. Louis Browns; in such cases, we took the view that the home team was the second team to come to bat. In some Series there were ties that had to be thrown out. And sometimes the first four games could not be used because three would be played at one park followed by some number at the other park. Our final rule was to collect for each Series the first two games played with the National League team as the "home team," and the first two games played with the American League team as the home team.[5] One Series (1922, N. Y. Giants vs. N. Y. Yankees) had to be omitted because in the four non-tied games, all won by the Giants, one team was "at home" three times. Thus we were left with 47 sets of four games.

The plan of analysis is to compare the same team for two games "away" and two "at home." We arbitrarily chose the *first* "away" team for the comparison. We counted the number won by that team in its first two (non-tied) away games and subtracted this from the number it won in its first two at-home games. This difference is taken as a measure of the improvement of a World Series team playing at home over playing away. If this difference is strongly positive or negative on the average, we would have to reject the notion that the chance of winning a single game is constant throughout the Series. For example, in 1949 Brooklyn (NL) was the first away team in the Series with New York (AL). It won one of its away games, and none of its at-home games, for an improvement score of minus one. The average improvement score for the 47 sets of four games was $2/47=0.042$, and the standard error of this mean is approximately 0.14. So the improvement

[5] Hy Turkin and S. C. Thompson, *The Official Encyclopedia of Baseball, Jubilee Edition* (New York: A. S. Barnes Co., 1951).

score is only about a third of a standard error from the null hypothesis score of zero improvement. Thus far, then, we have no good evidence for rejecting the constant probability assumption and it has been shown that the probability of winning a game is not influenced very much by the "at-home" or "away" status of teams.

A possible rationale for explaining this at-home-away similarity observed in Series games and not observed in season games suggests itself. It may be that travelling fatigues the away team and thus tends to cut down proficiency. During the regular season, at-home teams remain stationary for long periods, and various opponent teams travel in to play them. In the Series, one team has to travel initially, but then both teams do equal amounts of travelling until the Series ends. If travelling is an important variable influencing outcomes of games, the Series tends to equalize this influence much more than regular season games.

Another possibility is that many teams are tailored to the home park because half the games are played at home: for example, the Boston Red Sox (A.L.) have a short left-field fence, and therefore hire a good many strong left-field hitters. But it may well be that League champions represent teams that are not much affected by change of park.

Another way to say that trials are not independent is to say that they are correlated serially, and obviously this means p changes from game to game depending on outcomes of previous games. To test for serial correlation, we examined the results of the first four games regardless of where played. Each of the 48 sets of four games was broken into two sets of two games, the first set consisting of game 1 and game 2, and the second set consisting of game 3 and game 4. If there is serial correlation, we might find that winning a game improved the chance of winning the next game. To test this we scored the American League team in each set of two games and thus constructed the 2×2 table shown in Table 7.

TABLE 7

PERFORMANCE OF AMERICAN LEAGUE TEAM IN 96 SETS OF TWO GAMES

		Second game		
		Win	Lose	Total
First game	Win	32	24	56
	Lose	24	16	40
	Total	56	40	96

It will be noted that the rows of this table are almost proportional to one another. There is a slight question about the interpretation of this result. To work this out, we must explore the situation when we have independence between the games. Suppose that the American League team in any particular series of 2 games has a probability p_i of winning each game. Then the expected values for the particular 2-game series would be shown by the following table:

		Second game		
		Win	Lose	Total
First game	Win	p_i^2	$p_i(1-p_i)$	p_i
	Lose	$p_i(1-p_i)$	$(1-p_i)^2$	$(1-p_i)$
	Total	p_i	$(1-p_i)$	1

The expected value for the total table for all the 2-game series would be represented by the following table in which we merely sum each entry of the previous table over the subscript i:

		Second game		
		Win	Lose	Total
First game	Win	$\sum p_i^2$	$\sum p_i(1-p_i)$	$\sum p_i$
	Lose	$\sum p_i(1-p_i)$	$\sum (1-p_i)^2$	$\sum (1-p_i)$
	Total	$\sum p_i$	$\sum (1-p_i)$	n

In the ordinary test for independence we estimate the Win-Win cell by multiplying the two Win margins together and dividing by the total number. In the present case we would have

$$(\sum p_i)^2/n$$

as this estimate. Clearly this is not identical to $\sum p_i^2$. But we will show that in our present problem it is very close. If each p_i is represented as the sum of a grand mean plus a departure from that mean in the form

$$p_i = p + e_i$$

where e is the departure, then $p_i^2 = p^2 + 2pe_i + e_i^2$ and $\sum p_i^2 = np^2 + \sum e_i^2$. Dividing both sides by n gives

$$\frac{\sum p_i^2}{n} = p^2 + \frac{\sum e_i^2}{n}$$

and the second term on the right-hand side is approximately the variance of the true p_i's. From our earlier work we have estimated the standard deviation of the p_i's to be about 0.13 or 0.14, so the variance of the p_i's would be estimated to be about 0.017 or 0.020. This would yield an expected discrepancy between $\sum p_i^2$ and $(\sum p_i)^2/n = np^2$ of about 96×0.017, which is roughly 1.6 or 2.0 units. So in our Table 7 the expected cell values should not be and are not very many cases away from the result predicted by the products of the margins. Therefore, approximate independence in the table of pairs of games is consonant with the notion of game-to-game independence.

To sum up: we have made tests of the reasonableness of the assumption of the constancy of p throughout the Series and of the independence of p from game to game, and we have found no reason to reject the hypothesis of binomiality, in spite of the fact that it disagrees with our intuition or with our knowledge of the facts of games within the regular season.

We have not, of course, completely tested binomiality. We have only checked two of the most obvious sources of disturbance that might be present. One could check on various additional conjectures for which data are available. But the final word on the assumptions would come from an analysis of replications of the games, and, of course, there are no replications of World Series games. The issue here is that though the average p does not change as we go from at-home to away or from first to second game, it is still possible that p itself changes from game to game, though in no systematic way. Furthermore, the fact that the principal assumptions are reasonable when using a model in connection with World Series data may not help much if one wishes to use the model with other kinds of data. The *methods* used to investigate the agreement of the assumptions with the facts may have value in other cases, though, especially when detailed information is available about orders-of-test and other pertinent facts.

ODDS QUOTED FOR SERIES

Something a good many fans are concerned about is the before-Series chances of the contenders. At the suggestion of Harry V. Roberts we have gathered the odds quoted in advance of the Series for 36 years—these odds are published by betting commissioners or sometimes are the odds being used generally by the public for small bets. One way to look at these odds is to consider them a group-judgment of the subjective probabilities associated with the two teams as they enter the Series.

Information on betting odds was found in articles in the *New*

York Times, and for all years represents information quoted on the morning of the opening day of the World Series. Naturally, information reported varies somewhat from year to year, and occasionally arbitrary procedures had to be adopted to get quantitative probabilities to associate with each team and to make the total probability unity. The betting fraternity may have been active in the early years, but reports for 1913, 1914, and 1915 were vague and seemed very hard to quantify. Therefore, these data start in 1916. We list the information for 1913, 1914, and 1915:

New York Times, Oct. 7, 1913	"No odds in the betting"
New York Times, Oct. 9, 1914	"Few backing Mac's team" . . . "American League are favorites but adherents . . . not loudly proclaiming powers"
New York Times, Oct. 8, 1915	"There is a little conservative betting . . . they are about evenly matched."

Starting in 1916, however, we observed remarks giving actual odds like: "Boston favored 10 to 8 (N. Y. *Times*, Oct. 7, 1916); "Giant partisans . . . will wager all the cash Chicago fans want at even money but they will not offer odds" (N. Y. *Times*, Oct. 6, 1917); " . . . oppressive silence among the fans who usually make wagers . . . no choice . . . a few Boston enthusiasts willing to wager 6 to 5 on the Red Sox but these are small bets" (N. Y. *Times*, Sept. 5, 1918). In some of the later years more detailed information is available giving the odds in both directions.

In the computation, we have averaged the odds in cases where more than one set was given (like the 1918 quote above), so that if the big money was being wagered at even money and smaller bettors were willing to wager at 6 to 5, we have arbitrarily decided that the over-all odds prevailing were 5.5 to 5. When the odds are given in both directions, the probabilities associated with the two teams do not add up to unity, because a percentage has been deducted so that the betting commissioners profit no matter who wins. Since we are not immediately interested in the breadwinning activities of the betting commissioners, we have divided the remaining probability (1 − sum of probabilities for the two teams) in half and added equal amounts of probability to the estimate for each team. A sample computation will clarify this procedure.

In 1931, the October 1 N. Y. *Times* gives one to two against the Athletics (A.L.) and 8 to 5 against the Cardinals (N.L.). To the fan this says: one dollar will get you two dollars if you bet against the Athletics and the Athletics lose the Series; if you choose to bet against the Cardinals, you must wager $8 to win $5 if the Cardinals lose. These figures lead us to the following preliminary assessment of the total probability:

$$\frac{1}{3}+\frac{8}{13}=\frac{13}{39}+\frac{24}{39}=\frac{37}{39}.$$

37/39 is 2/39 less than unity, so we arbitrarily add 1/39 to the fraction expressing the odds the bookmakers are giving for each team. Adding 1/39 to 24/39 gives a total probability for the Athletics' winning of 25/39, or 0.64.

Calculations like this were carried through for each year for the Series winner starting in 1916. Since the total probability is always unity, the probability for the loser is the complement of the winner's probability. (1924 material may not be worth including because of a scandal on the eve of the Series. The best information obtainable was "betting odds were not decisive for either contender." We arbitrarily decided to include this information and assessed it as a 50-50 situation.) It turns out that the average subjective probability associated with the winner is 55.33 per cent. This leads us to the opinion that the betting fraternity has some ability to pick the winner. Furthermore, by looking at the average probability for 1916–33 and comparing it with the average probability for 1934–51, we see that bettors are getting better at picking the winner. The average probability for the winner from 1916–33 is 0.5028, compared to 0.6039 for 1934–51. Of course, many will argue that such a powerful team as the Yankees have had on many occasions since 1934 makes prediction easy. Much more pertinent than average probabilities—at least from the bettor's point of view—is the number of times a probability greater than 1/2 was associated with the actual winner of the Series. These data are shown below, omitting years when the probability was 50-50:

	Probability greater than 0.5 published before Series	Probability less than 0.5 published before Series
Number of Winners	24	8

If this 24-8 split is tested against a 16-16 split (null hypothesis), the result is highly significant. Thus the favorite has won 75 per cent of the time. Recalling that we estimate that the better team wins only 80 per cent of the time, this represents rather good choosing. If we let x represent the fraction of times the better team is picked as the favorite, then the fraction of times the favorite would win on the average (using our previous 0.8 as the probability that better teams win) is $0.8x+0.2(1-x)$.

If we equate this to the observed fraction of the time that the favorite has won we find

$$0.8x + 0.2(1 - x) = 0.75$$

$$x = \frac{0.55}{0.60} = 0.917.$$

In other words, we estimate that the better team has been made the favorite 92 per cent of the time.

We have also computed the odds given before the Series for the American League teams, since in much of this paper attention has been directed to the American League. The average before-Series subjective probability associated with the American League team is 0.5802; for the same period (1916–51) the American League team has won 24 of 36 Series, or 66.67 per cent. We note a trend toward increasing favoritism toward the American League team when the available data are split in half. For 1916–33 the average probability given for the American League team is 0.5372 while for 1934–51 the corresponding figure is 0.6228.

The reader may wonder what the boundary for the probability is at the upper end—we can see that the favored team does not get odds better than 2/3 very often, and the lower limit is 1/2. For the 36-year period, the average value for the *larger* subjective probability is 0.6044. This means that the favorite gets odds of about 3 to 2 on the average.

These data on betting odds are presented for their own interest, rather than for any contribution they make to the problem of estimating the probability associated with the better team. The bettors seem to have ability to pick the winning team, and as time goes on the bettors are getting more confident of their judgment. The strategy (since 1935) seems to be to pick the American League team unless that team is the Browns, and if the Yankees are the American League team good odds are found as low as 1 to 2.

Finally, we have traced the financial successes and set-backs of dyed-in-the-wool American League bettors and dyed-in-the-wool National League bettors. It is assumed that each year $100 was wagered on, say, the American League team, at the odds prevailing, and then depending on the outcome of the Series, either a profit was made or a loss was taken. At the end of 36 years (1916–1951), a gambler betting only on the American League team would have been ahead $556. At the end of the same 36 years, another gambler betting only on the National League would have been behind $808.

SUMMARY

We have used two methods to estimate 1) the percentage of times the American League has had the better team, and 2) the average p-value for better teams. The first method (Model A) made the unrealistic assumption that the better team had the same chance of winning every year, and the second method (Model B) made the unrealistic assumption that the p-values for American League teams were normally distributed from year to year. Both methods have the assumption that the probability of winning single games is constant within a Series, and that the outcomes of games within a Series are independent. Two checks on the binomial assumption showed no reason to reject it—better, they showed good agreement with it. For the percentage of times the American League has had the better team the two methods of estimating led to results of 76 and 72, or in round numbers 75 per cent. For the average single-game p-value for better teams, the estimates are 0.655 and 0.634, or in round numbers about 0.65. The two methods yield fairly close agreement.

APPENDIX

Derivation of $\sigma^2(\hat{p})$

To derive the variance of $\hat{p}$ we write below all the outcomes (x, y), the estimates corresponding to these outcomes, the probabilities of these estimates, and then we compute the second raw moment of $\hat{p}$.

Outcomes	(4, 0)	(4, 1)	(4, 2)	(4, 3)	(3, 4)	(2, 4)	(1, 4)	(0, 4)
p	1	3/4	3/5	3/6	3/6	2/5	1/4	0
$P(\hat{p})$	p^4	$4p^4q$	$10p^4q^2$	$20p^4q^3$	$20p^3q^4$	$10p^2q^4$	$4pq^4$	q^4

$$\sum \hat{p}^2 P(\hat{p}) = p^4 + \frac{9}{4} p^4 q + \frac{18}{5} p^4 q^2 + 5p^4q^3 + 5p^3q^4 + \frac{8}{5} p^2q^4 + \frac{1}{4} pq^4 = E(\hat{p}^2)$$

To get the variance $\sigma^2(\hat{p})$ we subtract p^2, but in the form

$$p^2 = p^4 + 2p^4q + 3p^4q^2 + 4p^4q^3 + 4p^3q^4 + p^2q^4.$$

This gives

$$\sigma^2(\hat{p}) = pq\left[\tfrac{1}{4}p^3 + \tfrac{3}{5}p^3q + p^3q^2 + p^2q^3 + \tfrac{3}{5}pq^3 + \tfrac{1}{4}q^3\right].$$

Now the probabilities of the outcomes for a 5-game series have to add up to unity, so we write $\frac{1}{4}$ in the form

$$\tfrac{1}{4} = \tfrac{1}{4}(p^3 + 3p^3q + 6p^3q^2 + 6p^2q^3 + 3pq^3 + q^3).$$

This relation simplifies $\sigma^2(\hat{p})$ to

$$\sigma^2(\hat{p}) = \frac{pq}{20}\,[5 - pq(3p^2 + 10p^2q + 10pq^2 + 3q^2)].$$

Similarly a 3-game series has outcomes whose probabilities add to unity, and we write 3 in the form

$$3 = 3(p^2 + 2p^2q + 2pq^2 + q^2).$$

This relation simplifies $\sigma^2(\hat{p})$ to

$$\sigma^2(\hat{p}) = \frac{pq}{20}\,[5 - pq(3 + 4p^2q + 4pq^2)]$$

$$\sigma^2(\hat{p}) = \frac{pq}{20}\,[5 - 3pq - 4p^2q^2].$$

When $p=\frac{1}{2}$, $\sigma^2(\hat{p})=1/20$. A non-truncated binomial variance with $p=\frac{1}{2}$ would need $n=5$ to yield $pq/n=1/20$, so we might say that the effective number of observations for estimating p in a 7-game series is approximately 5. The effective number for estimating p is less than the average number of games, which is $5\frac{13}{16}$ when $p=\frac{1}{2}$.

Means of Truncated Normal Distributions

If $f(x)$ is a probability density function on the interval $(-\infty, \infty)$ and we wish to evaluate the mean value of x given that $x>a$, we can use the expression

$$E(x|\; x > a) = \frac{\int_a^\infty xf(x)dx}{\int_a^\infty f(x)dx}\cdot$$

In our problem we are interested in the average p for the better team. We have assumed that p is normally distributed for American League teams. The American League team is better when $p>\frac{1}{2}$, the National League team is better when $p<\frac{1}{2}$. We need to compute the mean p for the American League when it is better, and the mean p for the National League when it is better. These two means can then be weighted by their estimated frequency of occurrence to give a final weighted mean estimate of the average probability that the better team wins single games.

To obtain the truncated normal estimate for the average p for better teams, we first set up the normal distribution with mean $\bar{p}$ and standard deviation $\hat{\sigma}$. For printing convenience, we drop the bar and circumflex. When the American League has the better team, the mean p-value is

$$\text{(a)}\quad \frac{\dfrac{1}{\sigma\sqrt{2\pi}}\displaystyle\int_{0.5}^{\infty} x e^{-\frac{1}{2}[(x-p)/\sigma]^2} dx}{P(\text{A.L. better})} = \frac{\sigma \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[(0.5-p)/\sigma]^2} + pP(\text{A.L. better})}{P(\text{A.L. better})}.$$

When the National League has the better team the mean p is

$$\text{(b)}\quad \frac{\dfrac{1}{\sigma\sqrt{2\pi}}\displaystyle\int_{0.5}^{\infty} x e^{-\frac{1}{2}[x-(1-p)]^2/\sigma^2} dx}{P(\text{N.L. better})} = \frac{\sigma \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[(p-0.5)/\sigma]^2} + (1-p)P(\text{N.L. better})}{P(\text{N.L. better})}.$$

Weighting the contributions (a) and (b) by their probabilities of occurrence gives

$$\text{Final estimate} = 2\sigma(1/\sqrt{2\pi})e^{-\frac{1}{2}[(p-0.5)/\sigma]^2} + pP(\text{A.L. better}) + (1-p)P(\text{N.L. better}).$$

With $\sigma = 0.1315$, $p = 0.583$, $(p-0.5)/\sigma = 0.63$, $P(\text{A.L. better}) = 0.74$, we get

$$\text{Estimate} = 2(0.1315)(0.3271) + (0.583)(0.74) + 0.417(0.26) = 0.626,$$

as reported in the text. Using $\sigma = 0.1439$ from our second approximation, we have $p = 0.583$, $(p-0.5)/\sigma = 0.576$, $P(\text{A.L. better}) = 0.718$, and the improved approximation is 0.634, as the average single-game probability for the better team.