HAVE THE WORLD SERIES BEEN FIXED?

R. L. Helmbold

August 1970

P-4447

HAVE THE WORLD SERIES BEEN FIXED?

R. L. Helmbold

The RAND Corporation,* Santa Monica, California

I. INTRODUCTION

In 1921 certain members of the Chicago White Sox admitted to accepting bribes to throw the 1919 series to the Cincinnati Reds, creating the famous "Black Sox" scandal. Because of the potential influence of organized crime it would be desirable to maintain careful surveillance over sports events. However, the law enforcement forces are limited in the amount of effort they can devote to that task. As a result, there is some interest in attempting to discover the extent to which the presence of external influence in sporting events can be detected by such evidence as may be provided by the pattern of outcomes of the sporting events themselves. To the extent such "pattern" evidence is effective in detecting the presence of external influence, the available law enforcement effort can be concentrated on the more suspicious looking cases.

In this paper, we will try to see whether or not aids can be provided to assist law enforcement management in concentrating its effort on the more suspicious areas. To make the discussion concrete, the case of the World Series Baseball competition will be considered, and some ways in which external influences on the pattern of series outcomes might be detected are developed and discussed.

---

*Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of The RAND Corporation or the official opinion or policy of any of its governmental or private research sponsors. Papers are reproduced by The RAND Corporation as a courtesy to members of its staff.

II. HISTORY OF THE WORLD SERIES

The history of the series is an old story to baseball fans. However, for those who do not follow the game closely the following historical notes may hold some interest. The first World Series was held in 1903, but was played under a rule where the "best of 9" games was required to win the series. However, no generally-accepted set of regulations governed the play in this series. Partly as a result of conflicting opinions regarding the rules governing series competition, the New York Giants in 1904 refused to play a post-season series, although challenged to do so by the Boston Red Sox (who consequently claimed the World Championship by default). A formal set of rules was soon adopted and were used to govern the 1905 series play. A World Series has been held each year since that time. Although the "best of 9" rule was reinstated for a brief span of three years (1919-1921), all other series have been played under the "best of 7" games wins the series rule.

As of the end of the 1969 season, 66 World Series had been played, 62 of which were governed by the "best of 7" rule and 4 of which were governed by the "best of 9" rule. The American League team won 39 (59%), and the National League team 27 (41%) of the 66 series played. Through 1969, 386 World Series games had been played, the average series length being 5.8 games. The American League team won 211 (54.7%), and the National League team won 175 (45.3%) of these games. When only the series played under the "best of 7" rules are considered, the corresponding figures are 355 games played in 62 series for an average series length of 5.7 games. American League wins amount to 195 games (54.9%), and National League wins to 160 games (45.1%). In 1919, the "Black Sox" scandal was created by eight players of the Chicago White Sox who were later (1921) indicted on a charge of accepting bribes to throw that year's series to the Cincinnati Reds. They were subsequently cleared by a jury, but were nevertheless banned by Commissioner Landis from ever again holding any position in professional baseball. Table 1 gives the number of World Series games won by League and decade. (Neither the above figures nor Table 1 include tie games. One tie game took place in each of the following years: 1907, 1912, 1922. There are no

Table 1

WORLD SERIES GAMES WON BY DECADE

Decade	Games Won: American League	Games Won: National League	Total
1903-1909	14	21	35
1910-1919	35	22	57
1920-1929	33	26	59
1930-1939	35	19	54
1940-1949	33	27	60
1950-1959	32	29	61
1960-1969	29	31	60
TOTAL	211	175	386

other tie games recorded.) Data used here and elsewhere in this paper are based on Reference 1, unless specifically indicated otherwise.

Tables 2 and 3 give the distribution of the lengths of series played, grouped by decade, for all series (Table 2) and separately for just the series played under the "best of 7" rules (Table 3). Since the series involving the "Black Sox" scandal occurred in 1919, a year for which the series was played under a "best of 9" rule, it is included in Table 2, but not in Table 3.

III. MOTIVES

The trend in average receipts per game played since 1948 are shown on Figure 1. An inflation factor of about 3 percent, represented by the solid curve, accounts adequately for the general trend during this period. The spread about the trend line can probably be accounted for by attendance variations, which amounted by-and-large to about $\pm$ 20 percent from year to year. Income from radio and television coverage is not included in the receipts shown in Figure 1.

In 1905, in the course of a five-game series, 91,723 attended and receipts amounted to $68,435 (Reference 2), or about $13,687 per game.

Table 2

DISTRIBUTION OF SERIES DURATION BY DECADE*

Decade	Series Duration, Number of Games Played					Total Series
	4	5	6	7	8	
1903-1909	1	2	1	1	1*	6*
1910-1919	1	4	3	1	1*	10*
1920-1929	3	1	1	4*	1*	10**
1930-1939	3	2	3	2	0	10
1940-1949	0	4	2	4	0	10
1950-1959	2	0	3	5	0	10
1960-1969	2	2	0	6	0	10
TOTAL	12	15	13	23*	3***	66****

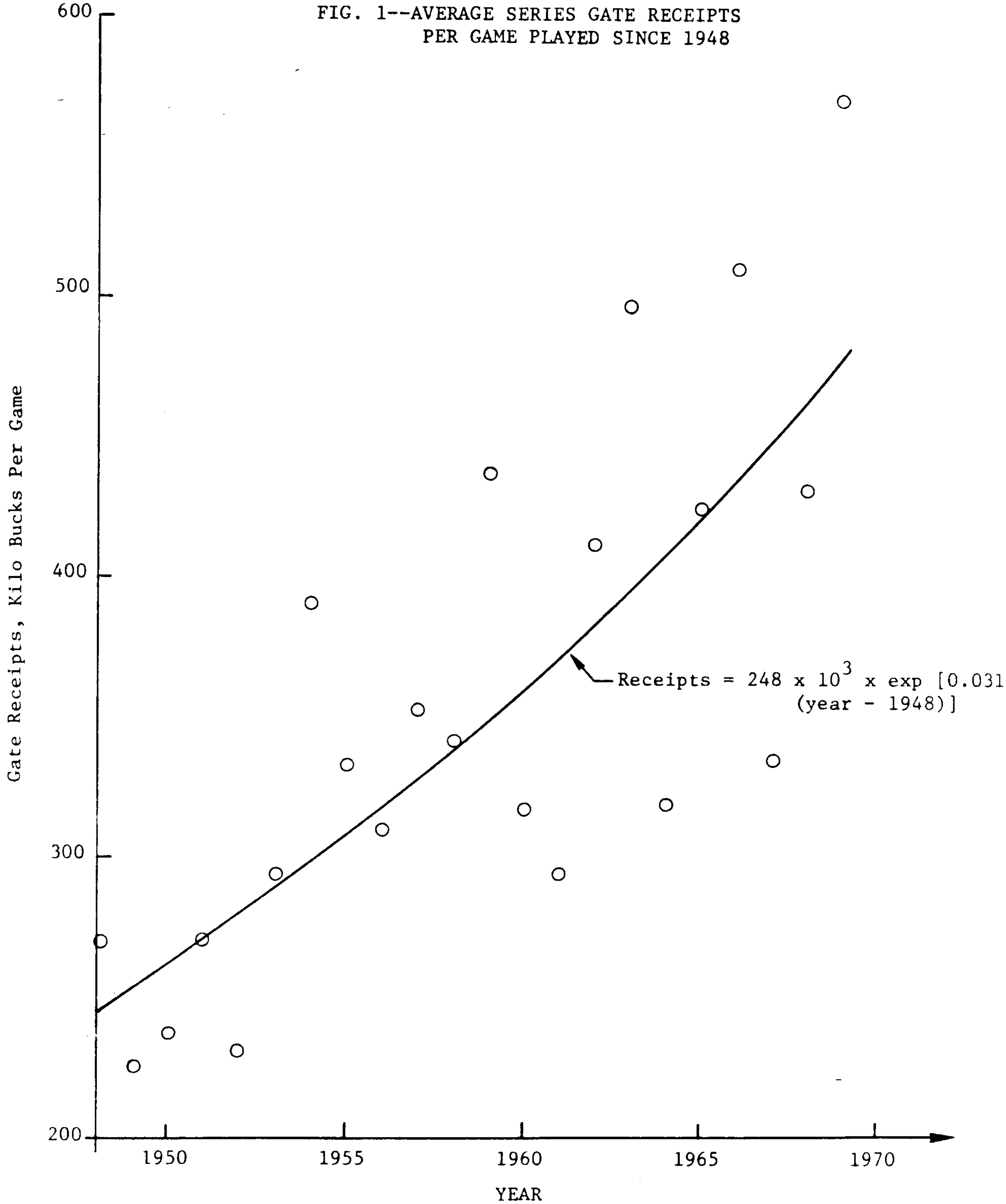
*The number of asterisks in a cell of the table indicates the number of series in that cell played under "best of 9" rules.

Table 3

DISTRIBUTION OF SERIES DURATION BY DECADE
FOR SERIES PLAYED UNDER "BEST OF 7" RULES

Decade	Series Duration, Number of Games Played				Total Series
	4	5	6	7	
1903-1909	1	2	1	1	5
1910-1919	1	4	3	1	9
1920-1929	3	1	1	3	8
1930-1939	3	2	3	2	10
1940-1949	0	4	2	4	10
1950-1959	2	0	3	5	10
1960-1969	2	2	0	6	10
TOTAL	12	15	13	22	62

FIG. 1--AVERAGE SERIES GATE RECEIPTS PER GAME PLAYED SINCE 1948



This is only about one-fifth of what would be estimated using the fitted exponential curve shown in Figure 1. Apparently the more recent trend cannot be satisfactorily extrapolated backward in time over a period of some 43 years. I'm sure that failure of such an extrapolation will hardly surprise anyone.

Ball park seating capacities have changed over the years. Reference 1 gives 1969 seating capacities ranging from a low of 25,420 (Sicks Stadium, Seattle) to a high of 76,977 (Cleveland Municipal Stadium, Ohio), with a median of 44,500 (Houston Astrodome, Texas). The World Series attendance during the period 1948-1969 has ranged from a low of 43,441 per game to a high of 70,131 per game, with a median of 52,047 per game. The maximum per game World Series attendance was attained in 1959 when the Los Angeles Dodgers met the Chicago White Sox. This was the Dodger's first year in Los Angeles and may be atypical. The next highest per game World Series attendance is 62,877. Thus, except for the 1959 series, attendance has been about 52,000 per game, ± 20 percent.

Distribution of World Series receipts is largely governed by policies of long-standing, and conducted under fairly public view. For example, earnings from radio and television coverage customarily goes to the Players' Pension Fund. Fifteen percent of all World Series gate receipts go to the Office of the Commissioner of baseball. Sixty percent of the gate receipts from the first four games is distributed among the players of the World Series teams, the players of the runner-up teams in each league, and the players of the teams finishing second and third in the divisional competition in their respective leagues. For example, in 1969, the New York Mets team members (as National League and World Series winners) each received $18,338.18. The Baltimore Orioles players each received $14,904.21 as World Series competitors and as American League Pennant winners. The Atlanta Braves players each received $6,944.10 and the Minnesota Twins players each received $6,460.12 since these two teams were runners-up in their respective leagues. The balance of the gate receipts after payments have been made to the Commissioner's office (15%) and to the player's pool (60% of the first four games) are distributed equally to the two series clubs and their respective leagues.

In addition to the monies received through gate receipts and mass media coverage, which are publicized and accounted for, there are the sums wagered on the outcome of the series. Mosteller (Reference 3) gives a good account of the odds published in the New York Times newspaper over a period of some 36 years, and of their relation to the actual outcome of the series (they are a much better predictor of series outcome than a random selection would be). It would be more difficult to estimate the total amount of money involved in wagers.

However, it seems evident that nearly all the parties involved--the mass media, public, club owners, advertisers, players, and gamblers--benefit to some extent from a long series.

IV. A PROBABILITY MODEL OF THE WORLD SERIES

The model we will develop will actually be somewhat more general than is required to deal with the World Series competition. However, the added generality does not create any significant extra difficulty in deriving the results. The basis for the model can be described as follows. Suppose that two sides engage in a series of games under the following rules:

(1) Each game is scored as a win for one side and a loss for the other. No tie games are permitted.

(2) At most, $k = 2m - 1$ games will be played.

(3) The first side to win $m = \left[\frac{k}{2} \right] + 1$ games (i.e., a clear majority) is counted as the series winner, the other side as the series loser.

(4) Game outcomes are mutually independent random variables with $P = \text{prob} \left\{ \text{side 1 wins} \right\}$ constant from one game to the next.

A few comments on these assumptions may be in order here. First, if tie games are permitted in principle, but the outcomes of tied games are ignored in determining the number of games played and the winner of the series, then the situation is no different in actuality than that described in the above. Second, the restriction of k to odd values is quite natural. If k = 2m, and the first side to win m games takes the series, then the series will have been decided at the end of at most (2m - 1) games, so that no (2m) th game will need to be played. However, if k = 2m and (m + 1) games must be won to take the series, then it is possible at the conclusion of (2m) games to have a tied series, thus requiring a playoff game, which would, of course, be game (2 (m + 1) -1), an odd number. The probability, under these rules, that the series is forced into a playoff is, from (3),

$$P_{2m} = 2 \binom{2m-1}{m-1} p^m q^m \qquad (1)$$

The assumptions of independence made in Rule (4) above are of an entirely different kind than those discussed in the previous paragraph. They are made partly because intuitively more satisfactory assumptions are nearly as arbitrary, without the saving simplicity of (4), and partly because of the results reported in the literature. Mosteller (Reference 3) has examined the possibility that the probability of winning a game might depend on whether the game was played at "home" or "away". He has also made a study of the possibility of serial dependence in which winning a game might influence the probability of winning the next game. In each case, Mosteller concludes that there is no objective statistical basis for rejecting (4). This does not mean that there is no effect of "home" versus "away" games, nor does it mean that there is no serial correlation. But it does mean that if these effects are present, they do not by themselves exert a decisive influence on the outcomes; instead any effect they may have is relatively minor.

The derivation of implications from the assumptions proceeds as follows. Let p_i be the probability that the series ends at the conclusion of game i with side 1 the winner. We will have

$$p_{m+n} = \binom{m+n-1}{m-1} p^m q^n \qquad (2)$$

where $n = 0,1,2, \ldots, k-m,$

$q = 1 - p,$

and, of course, p_i equal to zero for other values of i. To prove (2), note that in order for side 1 to win the series with game (m + n), side 1 must first win any (m - 1) out of the first (m + n - 1) games, while losing n of them, and then win the (m + n) th game. The probability that the series lasts for exactly (m + n) games is

$$P_{m+n} = \binom{m+n-1}{m-1} \left(p^m q^n + p^n q^m \right) \qquad (3)$$

where $n = 0,1,2, \ldots, k - m,$

since the series may end with either side the winner.

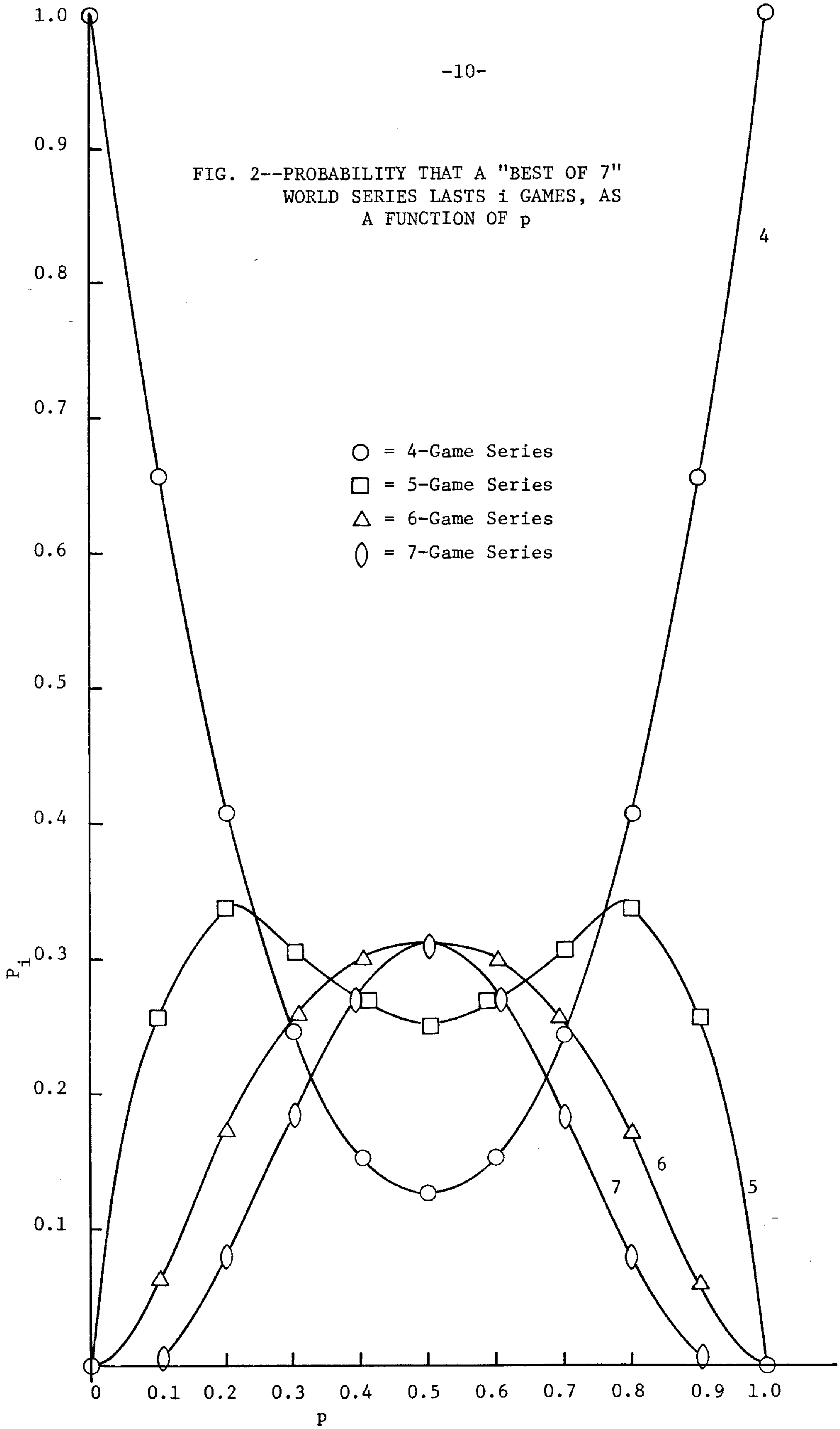
An equivalent setup is to give each side m counters at the start of the series, and to require the loser of a game to surrender (to an impartial umpire) one counter. The first side to go broke loses the series. In this form, we have a kind of "gambler's ruin" type problem, with the probability of side 1 surviving given by

$$p(1) = \sum_{n=0}^{k-m} p_{m+n} \qquad (4)$$

and the probability of termination after (m + n) games given by (3).

For application to World Series played under the "best of 7" rule the appropriate values of the parameters are k = 7 and m = 4. For World Series played under the "best of 9" rule, the values should be taken as k = 9 and m = 5. The probabilities P_i given by (3) are plotted as a function of p on Figure 2 for series played under "best of 7" rules. Table 4 gives the values of P_i together with the expected series length and standard deviation in series length for series played under the "best of 7" rules as a function of p. Since the probabilities are symmetric with respect to p, the values of $p > 0.5$ can be obtained

FIG. 2--PROBABILITY THAT A "BEST OF 7" WORLD SERIES LASTS i GAMES, AS A FUNCTION OF p



from Table 4 by using the symmetry relation

$$T(p) = T(1 - p) \tag{5}$$

where T(p) is any table entry in Table 4. Table 5 is like Table 4, but is for series played under the "best of 9" rules. The symmetry relation (5) applies to Table 5 as well as to Table 4.

Table 4

PROBABILITIES, EXPECTATIONS, AND STANDARD DEVIATIONS OF "BEST OF 7" SERIES LENGTHS FOR SELECTED VALUES OF P

P	P_4	P_5	P_6	P_7	E(n)	$\sigma(n)$
0.0	1.00000	0.00000	0.00000	0.00000	4.00000	0.00000
0.1	0.65620	0.26280	0.06642	0.01458	4.43938	0.68312
0.2	0.41120	0.33280	0.17408	0.08192	4.92672	0.95268
0.3	0.24820	0.31080	0.25578	0.18522	5.37802	1.04974
0.4	0.15520	0.26880	0.29952	0.27648	5.69728	1.03655
0.5	0.12500	0.2500	0.31250	0.31250	5.81250	1.01357

Table 5

PROBABILITIES, EXPECTED VALUE, AND STANDARD DEVIATION OF "BEST OF 9" SERIES LENGTHS FOR SELECTED VALUES OF P

P	P_5	P_6	P_7	P_8	P_9	E(n)	$\sigma(n)$
0.0	1.0000000	0.0000000	0.0000000	0.0000000	0.0000000	5.0000000	0.0000000
0.1	0.5905000	0.2952900	0.0886950	0.0209223	0.0045927	5.5538177	0.7779074
0.2	0.3280000	0.3289600	0.1996800	0.0974848	0.0458752	6.2042752	1.1352390
0.3	0.1705000	0.2606100	0.2447550	0.1879983	0.1361367	6.8586617	1.2865374
0.4	0.0880000	0.1862400	0.241920	0.2515968	0.2322432	7.3538432	1.2623805
0.5	0.0625000	0.1562500	0.234375	0.2734375	0.2734375	7.5390625	1.2177229

V. SOME STATISTICAL ANALYSES OF THE DATA

A Problem Area

Now, from Figure 2 and Table 4, it can be seen that the maximum probability of a 7-game series under "best of 7" rules occurs when $p = 1/2$. This is also what one would expect on the basis of intuition, for the best chance of a long series should be obtained when the competing teams are evenly matched. The maximum probability of a 7-game series is then about 0.31, and if in any year the actual probability of winning a game is different from 1/2, then the probability of a 7-game series in that year is less than 0.31. On the basis of these calculations, then, one would be encouraged to think that betting even odds that the series ends before the seventh game would be a good bet. In fact, it would appear that a gambler offering to give odds of 2-to-1 against a 7-game series would still hold a (slightly) favorable position. Nevertheless, the way the series have been turning out, as shown in Table 3, betting at 2-to-1 odds against the series lasting 7 games is a bad bet. The net gain/loss over the entire series for a 1-unit bet each year (where no bets are placed on series played under the "best of 9" rules) would be a net loss of 4 units, and a net loss since 1940 of 15 units. An actual gambler faced with this history of events might very well consider whether or not some external influence was in evidence, especially during the 1940-1969 period, tending to stretch out the series beyond their normal length. This is the question we will study in the remainder of this paper, i.e., "Have the series been artificially lengthened beyond their normal length, especially during the 1940-1969 period?"

Does the Distribution of Series Lengths Change From Decade-to-Decade?

One of the first questions that needs study is whether the series length changes over time. To answer this question, the data of Table 3 for series of 4 and 5 games were merged into one group (the short series group), and the data for series of 6 and 7 games were merged into another group (the long series group). The 1903-1909 data were merged with those for 1910-1919, thus making the period 1903-1919 one "long" decade. The resulting tabulation was then treated as a contingency table, and the chi-square test was used to test for independence of the (series duration) x (decade) factors (see Reference 4). A chi-square value of 3.715 was obtained with 5 degrees of freedom, which is not significant at the 50 percent level. Accordingly, there is no evidence that the distribution of series lengths is significantly different from decade to decade.

Are the Series Lengths Distributed as Predicted by the Probability Model?

If the distribution of series lengths are not significantly different from decade to decade, are they in accord with the distribution predicted by the probability model of the preceding section? Some prior information on the probability of p of winning a particular game is helpful, although we could estimate a value of p from the data themselves. Two different values of p will be used. One value is $p = 1/2$, which maximizes the probability of a long series, and which represents a perfect symmetry between the competing teams in each series played. The second value used for p is $p = 0.655$, which is the value determined by Mosteller (Reference 3) as an estimate of the probability that the best team will win a particular game. Mosteller based his estimate on World Series games played up to and including the 1951 series. The observed distribution used in the chi-square "goodness of fit" test (see Reference 4) is determined by the marginal totals from the last line of Table 3. It is appropriate to use these values since, as we have seen, the distribution does not significantly depend on the decade.

With $p = 1/2$, the chi-square value is 4.800 with 3 degrees of freedom. A larger deviation from the probability model could be expected on the basis of chance alone about 18.7 percent of the time.

With $p = 0.655$, the chi-squared value is 5.747. At three degrees of freedom, this value of chi-square would be significant at about the 12.5 percent level. However, there is a question about the proper value to take for the degrees of freedom, since this value of p was estimated by Mosteller using some of the same data as are given in Table 3. If the number of degrees of freedom is taken to be 2 instead of 3, then the significance of the 5.747 value of chi-square increases to the point where such a deviation from the probability model would be exceeded by chance alone only about 5.655 percent of the time.

The proper interpretation of these significance levels is a controversial matter. In the case we are studying, it seems appropriate to ask how "significant" the results have to get before a legal investigation is in order. And this decision seems likely to depend on one's predisposition to favor the view that some sort of hanky-panky is going on. Thus, if the investigator is already half-convinced that the series are being improperly influenced, then a significance level of 6 percent, or even 19 percent, might very well convince him that his intuitive suspicions are correct. The informal sort of argument used here is enticing, but not particularly clear. In order to make the reasoning more transparent, we will use a model based on Bayesian probability techniques.

A Bayes Rule Application

The usual Bayes rule formula is as follows:

$$P(B/A) = \frac{P(A/B)P(B)}{P(A/B)P(B) + P(A/\bar{B})P(\bar{B})}$$

where $P(A/B)$ is the probability of the observation A on the hypothesis B, and $P(B/A)$ is interpreted as the a posteriori probability of the hypothesis B, given that the outcome A has been observed. To apply this rule to our present case, let us suppose that A is the observation that the chi-square value is "too large". This is not particularly explicit just yet. We will show how it can be made more explicit later on, provided we are willing to adopt a certain very simple stylization of the situation.

Suppose we take B to be the hypothesis that the World Series games are shaded toward the 7-game series. In particular, we suppose that B is the hypothesis that a random 20 percent of the series have been artificially stretched to 7 games. We will suppose that $\bar{B}$ is the hypothesis that no series have been stretched. Now, the probability of various series lengths, on the basis of the value of p being 0.655, is as follows: $P_4 = 0.198229$, $P_5 = 0.291123$, $P_6 = 0.279860$, $P_7 = 0.230788$. When the series is artificially stretched to 7 games 20 percent of the time, the corresponding probabilities of various series lengths are distorted as follows: $P'_4 = 0.158583$, $P'_5 = 0.232898$, $P'_6 = 0.223888$, $P'_7 = 0.384630$. The value of chi-square on the basis of hypothesis $\bar{B}$ was computed earlier. If the degrees of freedom are taken as 3, $P(A/\bar{B}) = 0.125$. If the degrees of freedom are taken as 2, $P(A/\bar{B}) = 0.057$. We rather cavalierly average these two values, and use as the estimate of $P(A/\bar{B}) = 0.091$.

On the basis of the hypothesis, B, the chi-square value comes out to be almost exactly 0.7. At 3 degrees of freedom, this yields $P(A/B) = 0.873$. At 2 degrees of freedom, $P(A/B) = 0.705$. Again simply averaging the two yields as our estimate of $P(A/B) = 0.789$.

Substituting these values into Bayes' rule yields

$$P(B/A) = \frac{P(B)}{0.115 + 0.885P(B)}$$

a graph of which is shown as the solid line on Figure 3.

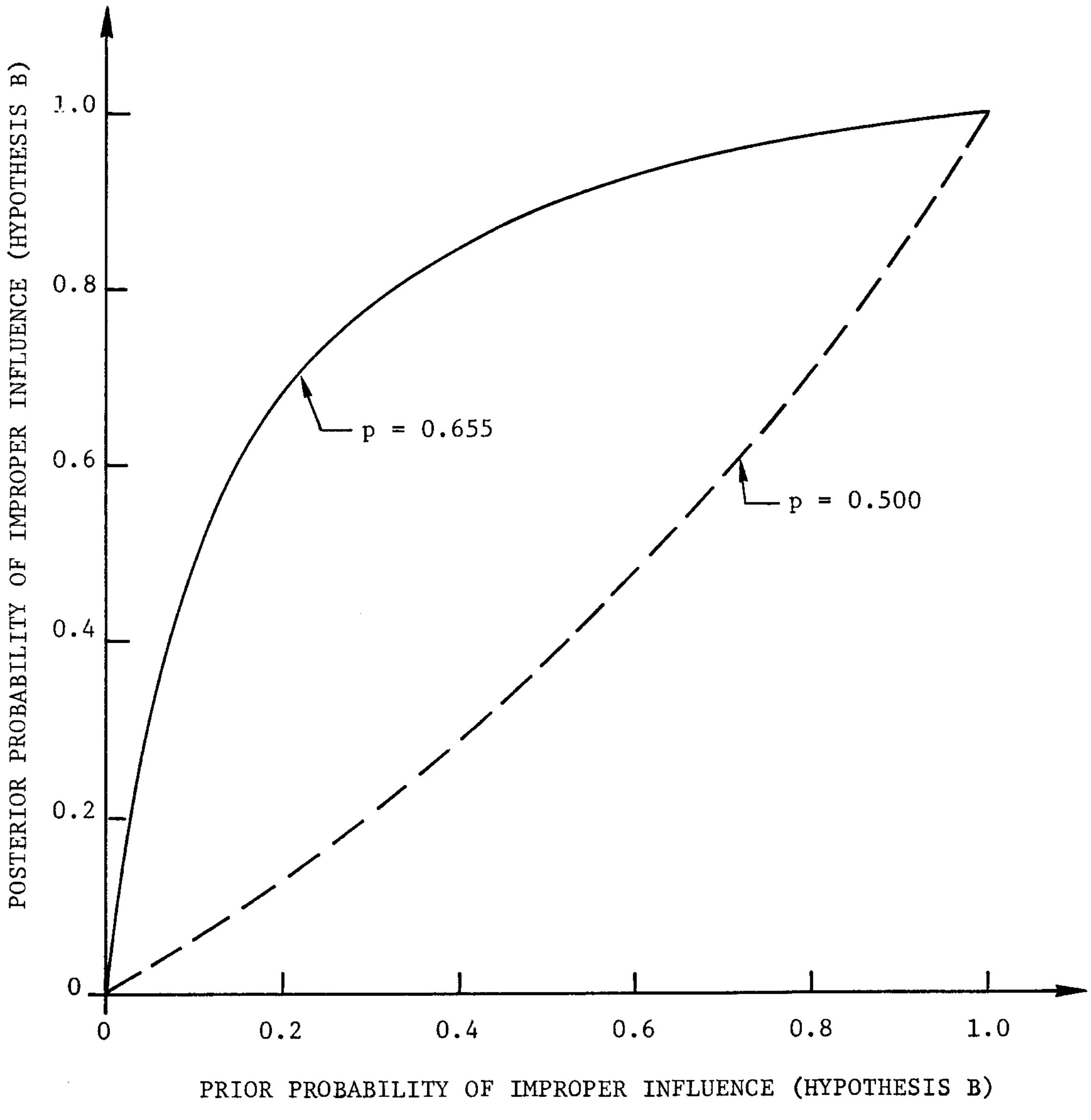
If the corresponding steps are taken with $p = 1/2$ instead of $p = 0.655$ (for $p = 1/2$ the degrees of freedom are taken as 3) then one obtains

$$P(B/A) = \frac{P(B)}{1.654 - 0.654P(B)}$$

which is shown as the dashed curve in Figure 3.

The curves on Figure 3 indicate the extent to which the posterior estimate of the likelihood of external influence of the World Series depends on one's prior assessment of that likelihood, as conditioned by the observed extent to which the distribution of observed series lengths differs from the theoretical one derived from the probability model. For example, if $p = 0.655$, and if one is suspicious that the series might be externally influenced and selects 0.2 as his prior subjective probability that hypothesis B holds, then on the basis of Bayes' rule, his posterior subjective probability in favor of B is almost 70

FIG. 3--POSTERIOR SUBJECTIVE PROBABILITIES OF HYPOTHESIS B
(SERIES SHADED TO 7-GAME LENGTH)
AS A FUNCTION OF ITS PRIOR PROBABILITY



percent. However, if $p = 1/2$, then his posterior subjective probability in favor of B declines to about 13 percent.

The highly stylized version of a Bayesian procedure presented here is susceptible in principle to almost unlimited refinement. Since our purpose is to illustrate and to illuminate possibilities through relatively elementary examples, the refinements and their associated complexities will not be discussed. Instead, the interested reader is referred to References 5 and 6 where careful descriptions of the theoretical basis for such extensions can be found.

A Sequential Test Program

To introduce the idea of a sequential monitoring procedure, we will indulge in a small fairy-tale.* Suppose that in the year 1903 a law enforcement agency, exerting great foresight, decided that it should adopt a firm policy of monitoring the World Series baseball contests in an effort to obtain prompt warning of any wrongdoing. Let us also suppose that they had on their staff a brilliant statistician--one who was about 40 years ahead of his time. Between the head of the agency and the statistician it is decided after some discussion and debate that the probability model developed in an earlier section of this paper is adequate for their purposes. These two also decide that $p = 0.655$ is the "true" value of the probability of winning a game. With this value of p, the probability model yields the probability of a 7-game series as $P_7 = 0.230788$. On the basis of these beginnings, the two decide that a critical upper bound on the probability of a 7-game series is $P_1 = 0.350$, while a critical lower bound on the probability of a 7-game series is $P_0 = 0.231$. The meaning of these critical bounds is as follows. If "enough" evidence should ever be accumulated that the probability of a 7-game series is at the upper bound, then this will be considered sufficient evidence of foul play to initiate a thorough investigation of the situation. On the other hand, if "enough" evidence is accumulated that the probability of a 7-game series is at

---

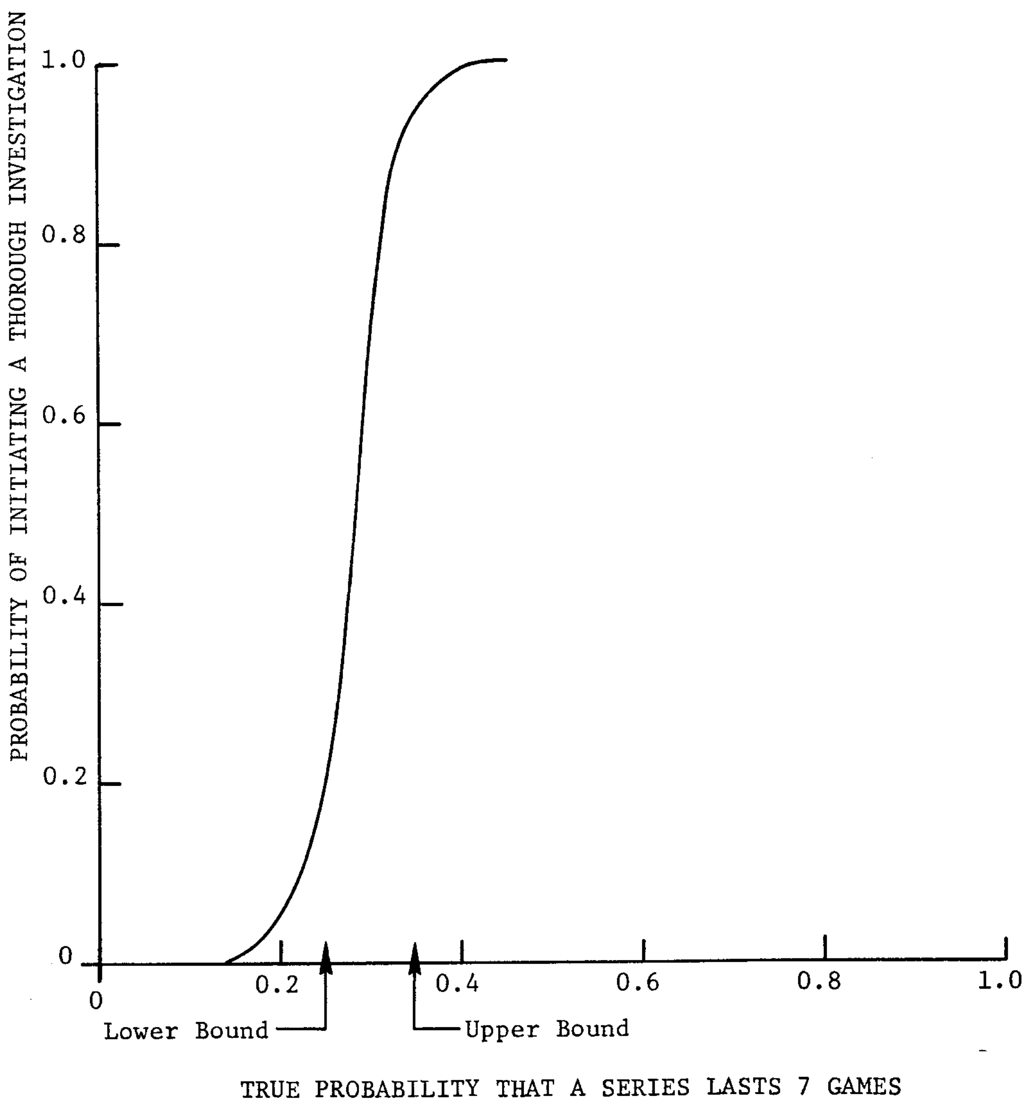
*The sequential statistical test used in this fictional example is described in Reference 7.

the lower bound, then this will be considered sufficient evidence that the series are behaving normally and the monitoring program will be discontinued.

After some further discussion the two men decided to determine how much evidence is "enough" through the following considerations. The head of the agency is willing to initiate a thorough investigation of the situation even if there is a 10 percent risk that the upper bound has been reached by a random statistical fluctuation. However, he insists that he must be 90 percent sure that the investigation is justifiable (i.e., that the upper bound has been exceeded by some systematically-operating factor and is not due to chance). The brilliant statistician immediately recognizes that the appropriate action is to take the probability of a Type I error as 10 percent. The agency head also declares that he wants to catch foul play 95 percent of the time where there really is some improper influence on the series competition, and that he is willing to misjudge the situation (by not investigating even though the probability of a 7-game series actually is at the upper bound) as much as 5 percent of the time. Our brilliant statistician immediately translates this judgment into a probability of a Type II error of 5 percent. The statistician then translates this whole set of decisions into a graph of the kind shown in Figure 4, and obtains the agency head's agreement that it represents the sense of his wishes on the matter. And so the monitoring program is initiated. The statistician goes off by himself with the information he has helped assemble and draws a control chart with a major trend line and two lines parallel to it. As each year rolls by and a series is played, another point is plotted on the control chart. The work is turned over to a clerk, who plots the points each year, and has instructions to notify the agency head and the statistician if the points at any time fall outside the "Continue Monitoring" band marked on the chart. When series are played under the "best of 9" rules, they are ignored and no points are plotted on the chart for those years. The chart is faithfully kept by the clerk year after year, with no decisive outcome.

Suddenly, at the end of the 1967 series (see series number 60 on Figure 5), the clerk notifies the agency head and the statistician that

FIG. 4--PROBABILITY OF INITIATING THOROUGH INVESTIGATION



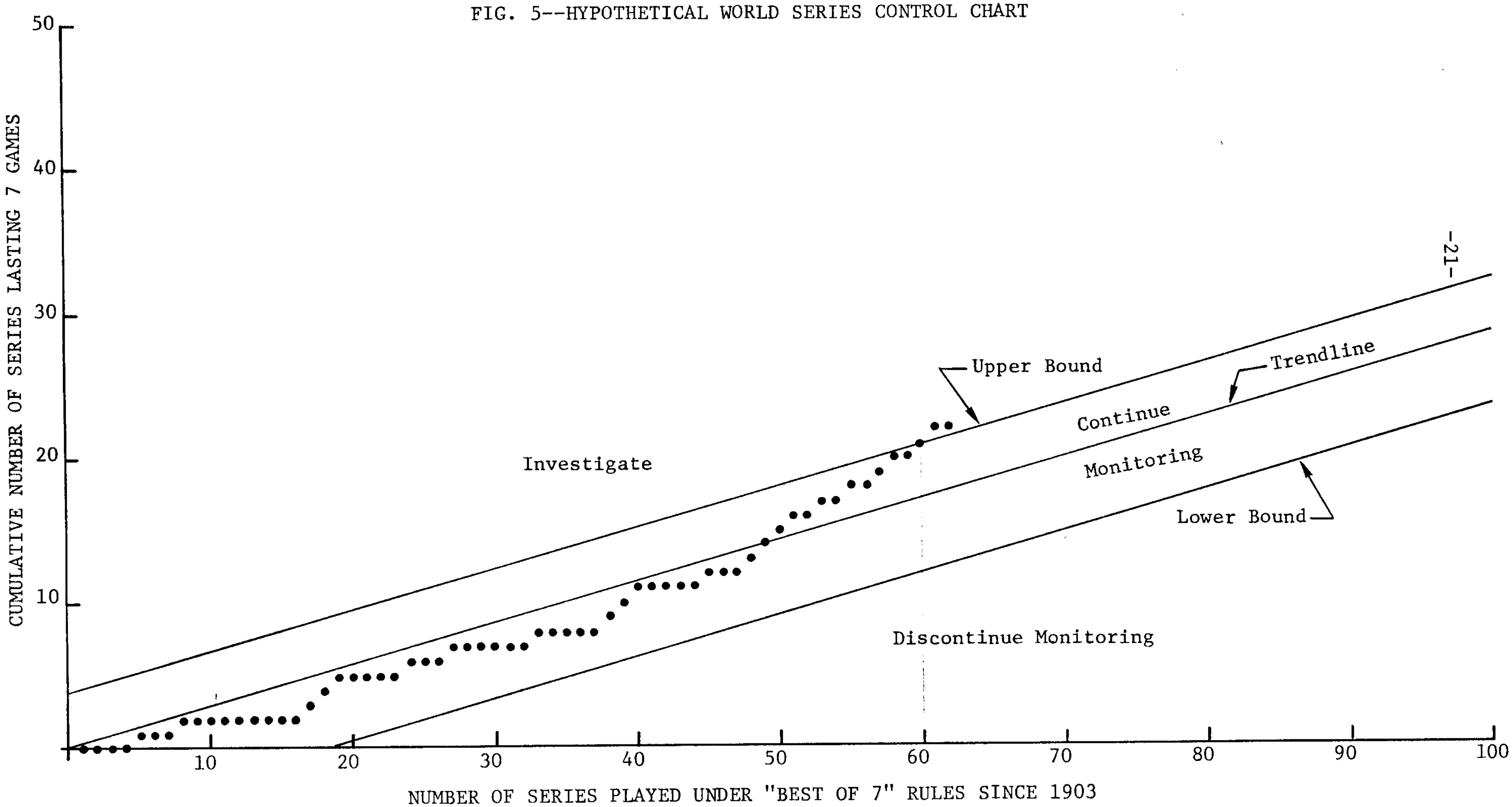
plotted points seem to have reached the "Investigate" boundary on the control chart. However, a careful check on the computation shows that the boundary lies ever-so-slightly above the plotted point. The agency head and the statistician pay very close attention to the 1968 series (series number 61 on Figure 5), which lasts for 7 games and puts the plotted points definitely above the "Investigate" upper bound. The agency head immediately assigns ten of his best men, with full staff support, to investigate the situation. Every effort is made to keep the investigation secret so as to not give advance warning to potential suspects. As of this writing, this hypothetical investigation would have been in progress for nearly two years with no results yet made public.

I will leave it to the reader to decide whether this fictional and hypothetical investigation would find evidence of improper influence, whether it would be a wild goose chase triggered by an unusual random fluctuation, or whether a more exacting statistical treatment* would indicate that the originally-proposed technique was too unsophisticated to justify the investigative effort.

---

*For instance, if p had been taken as $p = 0.500$ instead of $p = 0.655$, then the evidence to date would have remained in the "Continue Monitoring" band of the corresponding control chart.

FIG. 5--HYPOTHETICAL WORLD SERIES CONTROL CHART



REFERENCES

1. *The 1970 World Almanac and Book of Facts,* L. H. Long, ed., Newspaper Enterprise Association, Inc., New York, 1969.
2. Encyclopedia Americana, Article on *Baseball.*
3. Mosteller, F., *The World Series Competition,* Journal of the American Statistical Association, Vol. 47, No. 259, September 1952, pp. 355-380.
4. Crow, E. L., F. A. Davis, and M. W. Maxfield, *Statistics Manual,* NAVORD Report 3369, NOTS 948, U.S. Naval Ordnance Test Station. Republished by Dover Publications, Inc., New York, 1960.
5. Fishburn, P., *Decision and Value Theory,* John Wiley & Sons, Inc., New York, 1964.
6. Raiffa, H., *Decision Analysis,* Addison-Wesley, Reading, Mass., 1968.
7. Mood, A. M., *Introduction to the Theory of Statistics,* McGraw-Hill Book Co., New York, 1950.

P- 4447 HAVE THE WORLD SERIES BEEN FIXED? Helmbold