Engineering&Science
DECEMBER 1980

Gambling with Statistics

by GARY A. LORDEN

The mathematical theory of probability originated with attempts to analyze games of chance involving cards and dice. Even modern textbooks on probability devote substantial space to examples and calculations related to gambling games. What is less well known is the extent to which certain methods of statistical inference are also intimately related to the study of gambling.

One such area is my own research specialty, sequential analysis, which is based upon the idea of sampling data "one point at a time," the total amount of sampling being determined by carefully specified rules. These rules call for larger samples when the data are inconclusive, and smaller samples when early results are dramatic. Methods of sequential analysis have been applied to a broad range of practical problems, and have been particularly successful in areas such as clinical trials, reliability testing, and quality control, where the taking of data is necessarily spread out over time or is very costly. A great many problems of statistical inference can be solved by sequential methods, usually with much greater efficiency than is possible using classical techniques that take a sample of fixed size.

Because of the one-at-a-time nature of sequential sampling, the performance of these statistical procedures depends on the same sort of "chance fluctuation" phenomena that are of interest to gamblers. It is therefore not surprising that many of the methods of probability theory that have proved useful in the theory of sequential analysis are also quite illuminating when applied to questions about gambling such as: What are the chances of winning a lot of money in a particular game? What makes some bets and betting schemes successful more often than others?

Let's analyze the following situation. Suppose you find yourself in a gambling casino with $100 in your pocket and feel a desperate need to win $2000. Which of the following bets would be the best way to start?

(A) Put $10 on "8 the hard way" at craps.
(B) Risk $40 on "red" at roulette.
(C) Wager $10 on the "pass line" at craps.

A typical Caltech student would check the odds and percentages:

A pays $90 profit, but the odds against it are 10 to 1, and the percentage in favor of the house is 9.1 percent.

B pays $40 with odds of 10 to 9 against, and the house percentage is 5.3 percent.

C pays $10 with unfavorable odds of 36 to 35 (approximately), and the house percentage is 1.4 percent.

To make the choice more definite, let's assume that, whichever you choose, you will continue with the same proposition and will always bet the same fraction of your money, no matter how much you have, until you win your bundle (if ever). For example, if you choose A and get up to $500, you will wager $50 on the next turn.

Are you ready for the answer? It appears at the end of the next paragraph, so to avoid prejudicial information you should make your decision now.

The most popular choice is the third, based on the low house percentage. Some prefer the second, because it in-

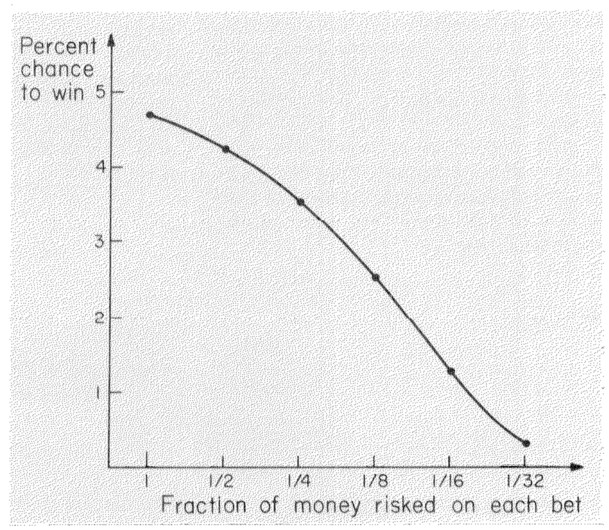


Figure 1. Probability of 20-fold win on the pass line.

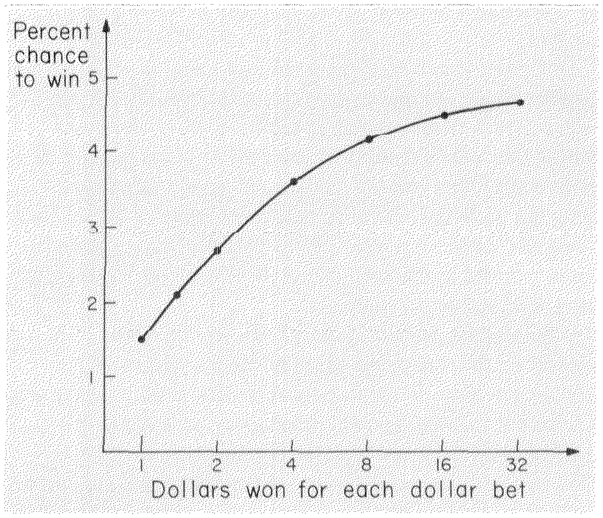


Figure 2. Probability of 20-fold win with various payoff ratios. All calculations assume 1) the fraction of money risked is 10 percent, and 2) the house advantage is 2 percent; that is, if R is the payoff ratio, then .98/(1 + R) is the probability of winning on each bet.

volves betting a larger fraction of your capital (which helps). Only an occasional optimist or devil's advocate picks the first, but that is actually the best bet of the three (and the third is the worst).

In fact, all three bets give the player roughly the same chance of success — small. Your probability of winning is 2.32 percent with A, 2.31 percent with B, and 2.14 percent with C. In a perfectly fair game (with no advantage for the casino) your chances of turning \$100 into \$2000 would be 1 in 20, or 5 percent. Why are the probabilities of A, B, and C so much lower? And how can it be that A is the best?

Figures 1 and 2 provide some insight into the answer. It is true that lower house percentages, like the one in C, help the player's chances. But there are two other important factors. One is the fraction of your money that you risk on each bet. The other is the payoff ratio — the number of dollars profit for each dollar bet when the bet wins. Figure 1 shows how dramatically the chances of a big win are reduced by risking small fractions of your money. Option B beats option C by virtue of risking 40 percent rather than 10 percent. Figure 2 illustrates the improvement in the player's chances resulting from a high payoff ratio. Bet A risks the same fraction as bet C and faces a much tougher house percentage, but its 9-to-1 payoff ratio more than compensates for this disadvantage.

Gambles like these can be analyzed easily by the methods of modern probability theory. One of these methods uses the notion of a martingale, the mathematical model of a fair game, which is used extensively in sequential analysis. Fair games have very nice mathematical properties. As mentioned before, the chances of turning \$100 into \$2000 are 1 in 20, and any other ratio works the same way. For example, the chances of turning \$56.30 into \$347.50 are 563 out of 3475.

A gambler using system A is playing an unfair game, of course, so his chances of winning can't be calculated that easily. But there is a trick we can use — a way of "keeping score" so that the game becomes fair. For example, in system A, if the player has \$100, his score is $(100)^{1.256}$, the number 100 raised to the power 1.256 (which is easy to compute on a scientific calculator). If he gets to \$2000, his score is $(2000)^{1.256}$, and so on. Why 1.256? Because this is the power (which I like to call the "casino power") that makes the scores behave like a fair game. Starting with $(100)^{1.256} = 325.1$, the chances of making it to $(2000)^{1.256} = 13{,}998.9$ are 3251 out of 139,989, which is the 2.32 percent mentioned earlier.

How is the casino power of 1.256 for bet A calculated? The player starting with \$100 has one chance in 11 to win his \$10 bet and jump up to \$190, and 10 chances in 11 to lose and drop down to \$90. If the power p is used for "scoring," then the player has a score of 100^p initially and after the first bet has either 190^p or 90^p. Taking account of the respective probabilities, his score after the first bet is, on the average,

$$\frac{1}{11} \times 190^p + \frac{10}{11} \times 90^p.$$

If this equals 100^p — his score *before* betting — then the game is fair. And this requires p = 1.256. Similar reasoning applied to bets B and C yields p = 1.257 and p = 1.283, respectively. Note that the *smaller* the casino power p, the *better* are your chances of winning, which are 100^p out of 2000^p, or $(.05)^p$. If p = 1, then the original bet is fair. Unfavorable bets will always have a casino power p bigger than 1. (The only exception is a bet with literally *no* chance to win, for which the casino power is undefined and, no doubt, of limited practical interest.) One of the nicest features of the notion of casino power is the fairly obvious principle that the chances of winning in a sequence of different types of bets are no greater than they would be if *all* of the bets had the same casino power as the best bet in the sequence. This simple principle makes it easy to refute a lot of the intricate betting schemes (particularly for craps) that are studied at length in the pseudo-scientific gambling literature.

The best casino power I know of for an unfavorable game is p = 1.0081 for craps in casinos that offer "double odds" side bets. This is obtainable by betting one-third of your current capital on the pass line and the remaining two-thirds behind the line. (For explanations of these terms, consult your local croupier.) Since casino chips don't come in fractions, you can only approximate this scheme in actual play, but casino powers of about 1.009 can be achieved in this way by players who are willing to risk everything at once. (Of course, the game of blackjack is favorable under proper conditions, and more limited bet sizes are advantageous in favorable games, where casino power is an irrelevant notion.)

What does all this have to do with statistics? Surprisingly enough, the same kind of calculations are needed to determine the chances of error in a statistical test. What's more, the most efficient methods for testing statistical hypotheses, those of sequential analysis, involve some of the same mathematical problems that a gambler is concerned with.

To understand these problems, consider the following hypothetical problem in statistics. You have a very large box containing millions of marbles, some of them white, the rest red. You want to pick at random a sample of a small number of marbles and infer which color the majority of marbles are. The classical methods of statistics lead to procedures like this: Determine in advance a sample size, take a sample of that size, and see which color most of the marbles in the sample are.

The chances of error in a test like this depend on the size of the sample and on the *actual* fraction of red marbles in the box. If 45 percent of the marbles are red, and we sample 121 (a convenient size for later comparison), then the probability of getting 61 or more red marbles in the sample is 13.5 percent, and this is the probability of error; if 48 percent of the marbles are red, the probability of error is much larger, namely, 33 percent. We can reduce the chances of error by taking a larger sample size. For example, sampling 241 marbles will yield error probabilities of 5.9 percent for 45 percent red and 26.7 percent for 48 percent red.

The main goal of statistical theory is to find methods of sampling and drawing inferences that are the most efficient — that is, have the smallest possible error probabilities for a particular sample size. Said another way, if we want specified error probabilities, then the most efficient tests are those that require the smallest possible sample size.

Going back to our problem with the marbles, it's hard to see how one could possibly improve upon the efficiency of the test that draws a sample and picks the color of the majority of that sample. In fact, there is a famous theorem in statistics that guarantees that no test taking a sample of 121 marbles can possibly improve upon the error probabilities of that test. Nevertheless, there *are* more efficient tests. These tests do not fix the sample size in advance, but instead sample marbles one at a time sequentially, following a previously specified rule for stopping the test and deciding which color predominates.

This approach to statistical inference, called sequential analysis, was developed in the mid-1940s by a mathematical statistician named Abraham Wald, whose book on the subject opened up new fields of research and applications. Wald's principal method, called the Sequential Probability Ratio Test (SPRT), gives recipes like the following for the marble problem: Sample one at a time until you have 11 more marbles of one color than the other. This could happen after 11 marbles, if all are the same color, or, perhaps, after 41 marbles are sampled, 26 white and 15 red. Whenever this situation occurs, the SPRT stops and decides the color that's ahead is in the majority in the box of marbles.

This process can be considered as a game in which a gambler bets even money that the next color drawn will be red. Think of the gambler starting with 11 chips and betting one at a time until he loses all 11 chips or wins 11 more chips. That corresponds exactly to the rule for carrying out an SPRT, but it is also an example of the famous "gambler's ruin problem." In fact, if the ratio of red to white marbles in the box is 9 to 10, this is just like betting on red in Las Vegas roulette.

What are the probabilities of error for the SPRT? They depend on the true percentages of the two colors in the box, just as they do for fixed sample size tests. If there are 45 percent red marbles, say, then an error is made if the gambler gets 11 ahead without previously getting 11 behind in a game where he has a 45 percent chance to win on each bet. To solve this "gambler's ruin problem," we can use the same idea of inventing a scoring system to make a fair game. Start off with a score of 1 and multiply the score by 55/45 every time he wins, and by 45/55 every time he loses. Since the chances of these are 45 out of 100 and 55 out of 100, respectively, *on the average* each bet multiplies his score by

$$\frac{45}{100} \times \frac{55}{45} + \frac{55}{100} \times \frac{45}{55} = \frac{55}{100} + \frac{45}{100} = 1.$$

So on the average this score goes neither up nor down, and the game is fair. If he wins the 11 chips, his score at that time will be $(55/45)^{11}$, whereas if he loses, it will be $(45/55)^{11}$. If p and $1-p$ are his probabilities of winning and losing, then his average score when the game stops is

$$p\,(55/45)^{11} + (1-p)\,(45/55)^{11}.$$

Since the scoring system makes the game fair, this average score at the end must equal his score at the beginning, which was 1. That fact gives a simple equation to solve for p, and the answer is $p = .099$. Thus the probability of error is 9.9 percent for the SPRT when there are 45 percent red marbles in the box. For 48 percent red marbles, the same calculations with 45 and 55 replaced by 48 and 52 show that the error probability is 29.3 percent.

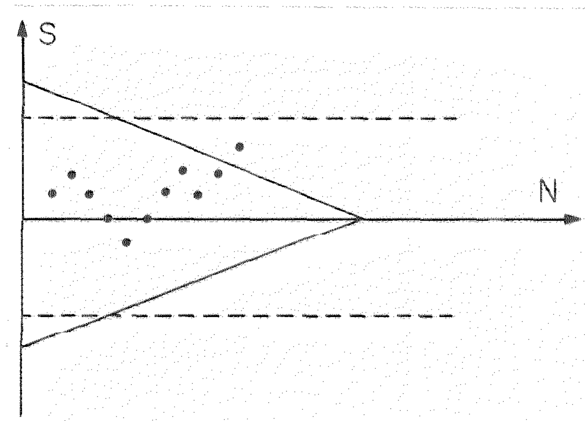
Why should a statistician want to use a test like the SPRT that imitates a gambling game? Because it is more efficient than fixed sample size tests. It turns out that the average number of marbles sampled by the SPRT depends on the proportions of red and white marbles in the box, but the worst case is the one where the two colors are equally numerous. In this case the sample size of the SPRT is 121, on the average — exactly the same as the fixed sample size test discussed earlier. So it makes sense to compare their error probabilities.

In the table below, which compares the error probabilities and average sample sizes of the SPRT with those of the fixed sample size test discussed earlier, notice that the SPRT yields improvements both in error probabilities and in average sample sizes. Of course, one can make other comparisons. For example, suppose we specify that the probability of error should be no more than 9.9 percent if 45 percent of the marbles in the box are of one color and 55 percent are of the other color. Then the smallest fixed sample size that accomplishes this is 164 — almost twice as large as the average sample size of the SPRT for these percentages of the two colors. In fact, Wald and Jacob Wolfowitz proved over 30 years ago that the SPRT is the most efficient possible test in a certain sense. For example, every other test needs a larger average sample size than 88.2 to obtain 9.9 percent error probabilities in the case of 45 percent or 55 percent red marbles.

	Probability of Test Error		Average Sample Size	
Fraction of red marbles in box	SPRT	Fixed sample size	SPRT	Fixed sample size
45%	9.9%	13.5%	88.2	121
48%	29.3%	33.0%	113.8	121
50%	—	—	121	121
52%	29.3%	33.0%	113.8	121
55%	9.9%	13.5%	88.2	121

Other efficiency criteria are important, however, and it turns out that the average sample size of 121 in the worst case can, in fact, be made smaller by choosing a different sequential test having the same 9.9 percent error probabilities. Finding tests that have the smallest possible average sample size in the worst case has been recognized as an important problem in statistical theory since it was first studied over 20 years ago by Jack Kiefer and Lionel Weiss. Among the difficulties associated with the problem is the fact that, except for situations like our marble problem that have a symmetrical character, it is not possible to identify the worst case in advance. It depends on the sequential test that is used. What we really want is the test whose worst performance is better than the worst performance of all other tests with the same error probabilities.

The problem of finding such tests, called "Kiefer-Weiss solutions," remains unsolved. However, work done in the last few years at Caltech has produced practical methods for constructing sequential tests that come close to solving it, and are, therefore, nearly as efficient as possible in this sense. The kind of sequential test that accomplishes this can be described most easily by a picture. Suppose that after we sample each marble, we plot a point on a graph of S versus N (above, right), where N is the number of marbles we've sampled so far, and S is the number of red marbles so far, minus the number of white marbles. Then an SPRT can be pictured by drawing a pair of parallel lines (dotted lines in the picture) that correspond to particular values of S, like +11 and −11. Whenever one of the plotted points reaches one of the dotted lines, the SPRT stops sampling. The solid lines in the picture, on the



other hand, represent another sequential test, which stops sampling when one of the plotted points reaches or crosses over one of these lines.

My own research on the Kiefer-Weiss problem identified a class of sequential tests that look like the solid-line picture, which I called 2-SPRTs because they are constructed from pieces of two different SPRTs in a certain way. It turned out that one could prove that these sequential tests are very nearly the best possible in a sense closely related to the Kiefer-Weiss problem. My student, Michael Huffman, recently showed in his PhD thesis how to choose from the class of all 2-SPRTs those tests that will come the closest to solving the Kiefer-Weiss problem. Though all of this theory is what statisticians call "asymptotic" (that is, based on what would happen if the sample sizes were very large), it is possible to make very detailed calculations on a computer that show in particular situations just how close these methods come to solving the Kiefer-Weiss problem. In typical applications, the answer seems to be that one achieves about 97 or 98 percent of the best possible efficiency.

Current research in sequential analysis is expanding the scope of both theory and applications. Recent work has led to the development of useful sequential procedures in many classical areas of statistics, such as the analysis of variance, nonparametric statistics, and the design of experiments. A growing stockpile of techniques has taken the theory of sequential analysis far beyond the study of the Sequential Probability Ratio Test and its refinements. Some of the problems now being solved by sequential analysis methods, such as reactions to trends and changes in distribution, and assignment of confidence intervals with fixed precision regardless of the variance of the data, cannot be dealt with at all by fixed sample size techniques. Along with all of this usefulness and respectability comes the continuing realization that sequential analysis, perhaps more than any other branch of statistics, is firmly rooted in the phenomena that govern one of mankind's most cherished disreputable pastimes: gambling. □