# Statistics and Gambling

Kyle Siegrist
Department of Mathematical Sciences
University of Alabama in Huntsville
siegrist@math.uah.edu

January 13, 2010

## 1 Introduction

Statistics can broadly be defined as the science of decision-making in the face of (random) uncertainty. Gambling has the same definition, except in the narrower domain of a gambler making decisions that affect his fortune in games of chance. It is hardly surprising, then, that the two subjects are closely related. Indeed, if the definitions of "game", "decision", and "fortune" in the context of gambling are sufficiently broadened, the two subjects become almost indistinguishable.

Let's review a bit of the history of the influence of gambling on the development of probability and statistics. First, of course, gambling is one of the oldest of human activities. The use of a certain type of animal heel bone (called the *astragalus*) as a crude die dates to about 3500 BCE (and possibly much earlier). The modern six-sided die dates to about 2000 BCE.

The early development of probability as a mathematical theory is intimately related to gambling. Indeed, the first probability problems to be analyzed mathematically were gambling problems:

1. *De Mere's problem* (1654), named for Chevalier De Mere and analyzed by Blaise Pascal and Pierre de Fermat, asks whether it is more likely to get at least one six with 4 throws of a fair die or at least one double six in 24 throws of two fair dice.

2. The *problem of points* (1654), also posed by De Mere and analyzed by Pascal and Fermat, asks for the fair division of stakes when a sequence of games between two players (Bernoulli trials in modern parlance) is interrupted before its conlcusion.

3. *Pepys' Problem* (1693), named for Samuel Pepys and analyzed by Isaac Newton, asks whether it is more likely to get at least one six in 6 rolls of a fair die or at least two sixes in 12 rolls of the die.

4. *The matching problem* (1708), analyzed by Pierre-Redmond de Montmort, is to find the probability that in a sequence of card draws, the value of a card is the same as the draw number.

5. *St. Petersburg Paradox* (1713), analyzed by Daniel Bernoulli, deals with a gambler betting on a sequence of coin tosses who doubles his bet each time he loses (and leads to a random variable with infinite expected value).

Similarly, the first books on probability were written by mathematician-gamblers to analyze games of chance: *Liber de Ludo Aleae* written sometime in the 1500's by the colorful Girolamo Cardano and published posthumously in 1663, and *Essay d'Analyse sur les Jeux de Hazard* by Montmort, published in 1708. See [2] and [4] for more on the influence of gambling on the early development of probability and statistics.

In more modern times, the interplay between statistics and game theory has been enormously fruitful. Hypothesis testing, developed by Ronald Fisher and Karl Pearson and formalized by Jerzy Neyman and Egon Pearson is one of the cornerstones of modern statistics, and has a game-theory flavor. The basic problem is choosing between a presumed null hypothesis and a conjectured alternative hypothesis, with the decision based on the data at hand and the probability of a type 1 error (rejecting the null hypothesis when it's true). Influenced by the seminal work of John von Neumann and Oscar Morgenstern on game theory and economics [7], the Neyman-Pearson hypothesis-testing framework was extended by Abraham Wald in 1940's to *statistical decsion theory* [8]. In this completely game-theoretic framework, the statistician (much like the gambler) chooses among a set of possible decisions, based on the data at hand according to some sort of value function. Statistical decision theory remains one of the fundamental paradigms of statistical inference to this day.

## 2 Bold Play in Red and Black

Gambling continue to be a source of interesting and deep problems in probability and statistics. In this section, we briefly describe a particularly beautiful problem analyzed by Dubins and Savage [3]. A gambler bets, at even stakes, on a sequence of Bernoulli trials (independent, identically distributed trials) with success parameter $p \in (0, 1)$. The gambler starts with an initial fortune and must continue playing until he is ruined or reaches a fixed target fortune. (The last two sentences form the mathematical definition of *red and black*.) On each trial, the gambler can bet any proportion of his current fortune, so it's convenient to normalize the target fortune to 1; thus the space of fortunes is the interval $[0, 1]$.

The gambler's goal is to maximize the probability $F(x)$ of reaching the target fortune 1, starting with an initial fortune $x$ (thus, $F$ is the value function in the context of statistical decision theory). The gambler's strategy consists of decisions on how much to bet on each trial. Since the trials are independent, the only information of use to the gambler on a given trial is his current fortune.

Thus, we need only consider *stationary, deterministic strategies*. Such a strategy is defined by a *betting function* $S(x)$ that gives the amount bet on a trial as a function of the current fortune $x$.

Dubins and Savage showed that in the sub-fair case ($p \leq \frac{1}{2}$), an optimal strategy is *bold play*, whereby the gambler, on each trial, bets his entire fortune or the amount needed to reach the target (whichever is smaller). That is, the betting function for bold play is

$$S(x) = \begin{cases} x, & 0 \leq x \leq \frac{1}{2} \\ 1 - x, & \frac{1}{2} \leq x \leq 1 \end{cases}$$

Conditioning on the first trial shows that the value function $F$ for bold play satisfies the functional equation

$$F(x) = \begin{cases} pF(2x), & x \in [0, \frac{1}{2}] \\ p + (1-p)F(2x-1), & x \in [\frac{1}{2}, 1] \end{cases} \tag{1}$$

with boundary conditions $F(0) = 0$, $F(1) = 1$. Moreover, $F$ is the unique bounded solution of (1) satisfying the boundary conditions. This functional equation is one of the keys in the analysis of bold play. In particular, the proof of optimality involves showing that if the gambler starts with some other strategy on the first trial, and then plays boldly thereafter, the new value function is no better than the value function with bold play.

Interestingly, as Dubins and Savage also showed, bold play is not the unique optimal strategy. Consider the following strategy: Starting with fortune $x \in [0, \frac{1}{2})$, the gambler plays boldly, but with the goal of reaching $\frac{1}{2}$. Starting with fortune $x \in (\frac{1}{2}, 1]$, the gambler plays boldly, but with the goal of not falling below $\frac{1}{2}$. In either case, if the gambler's fortune reaches $\frac{1}{2}$, he plays boldly and bets $\frac{1}{2}$. Thus, the betting function $S_2$ for this new strategy is related to the betting function $S$ of bold play by

$$S_2(x) = \begin{cases} \frac{1}{2}S(2x), & 0 \leq x < \frac{1}{2} \\ \frac{1}{2}S(2x-1), & \frac{1}{2} < x \leq 1 \\ \frac{1}{2}, & x = \frac{1}{2} \end{cases}$$

By taking the three cases $x \in [0, \frac{1}{2})$, $x = \frac{1}{2}$, and $x \in (\frac{1}{2}, 1]$, it's easy to see that that the value function $F_2$ for strategy $S_2$ satisfies the functional equation (1). Trivially the boundary conditions are also satisfied, so by uniqueness, $F_2 = F$ and thus $S_2$ is also optimal.

Once one sees that this new strategy is also optimal, it's easy to construct an entire sequence of optimal strategies. Specifically, let $S_1 = S$ denote the betting function for ordinary bold play and then define $S_n$ recursively by

$$S_{n+1}(x) = \begin{cases} \frac{1}{2}S_n(2x), & 0 \leq x < \frac{1}{2} \\ \frac{1}{2}S_n(2x-1), & \frac{1}{2} < x \leq 1 \\ \frac{1}{2}, & x = \frac{1}{2} \end{cases}$$

Then $S_n$ has the same value function $F$ as bold play and so is optimal for each $n$. Moreover, if $x \in (0, 1)$ is not a binary rational (that is, does not have the form $\frac{k}{2^n}$ for some $k$ and $n$), then there exist optimal strategies that place arbitrarily small bets when the fortune is $x$. This is a surprising result that seems to run counter to a naive interpretation of the law of large numbers.

Bold play in red and black leads to some exotic functions of the type that are not usually associated with a simple, applied problem. The value function $F$ can be interpreted as the distribution function of a random variable $X$ (the variable whose binary digits are the complements of the trial outcomes). Thus $F$ is continuous, but has derivative 0 almost everywhere if $p \neq \frac{1}{2}$ (singular continuous). If $p = \frac{1}{2}$, $X$ is uniformly distributed on $[0, 1]$ and $F(x) = x$. If $G(x)$ denotes the expected number of trials under bold play, starting with fortune $x$, then $G$ is discontinuous at the binary rationals and continuous at the binary irrationals.

Finally, note that when the gambler plays boldly, his fortune process follows the deterministic map $x \mapsto 2x \mod 1$, until the trial that ends the game (with fortune 0 or 1). Thus, bold play is intimately connected with a discrete dynamical system. This connection leads to other interesting avenues of research (see [5]).

# References

[1] David Blackwell and M.A. Girshick. *Theory of Games and Statistical Decisions*, Dover Publications, 1979.

[2] F.N. David. *Games, Gods and Gambling, A History of Probability and Statistical Ideas*, Dover Publications, 1998.

[3] Lester E. Dubins and Leonard J. Savage. *Inequalities for Stochastic Processes (How to Gamble If You Must)*. Dover Publications, 1976.

[4] Richard A. Epstein. *The Theory of Gambling and Statistical Logic*, Academic Press, 1977.

[5] Marcus Pendergrass and Kyle Siegrist. "Generalizations of Bold Play in Red and Black" *Stochastic Processes and Their Applications*, 92, 2001

[6] Leonard J. Savage. *The Foundations of Statistics*, Dover Publications, 1972.

[7] John von Neumann and Oscar Morgenstern *Theory of Games and Economic Behavior*, Princeton, 1944

[8] Abraham Wald. *Statistical Decision Functions*, Wiley, 1950.