Sampling; Random Walk

1 What the Mean Means

1.1 Bernoulli's "Stupidest Man"

We have focused on the expectation of a random variable because it indicates the "average value" the random variable will take. But what precisely does this mean?

We know a random variable may never actually equal its expectation. We also know, for example, that if we flip a fair coin 100 times, the chance that we actually flip *exactly* 50 heads is only about 8%. In fact, it gets less and less likely as we continue flipping that the number of heads will exactly equal the expected number, *e.g.*, the chance of exactly 500 heads in 1000 flips is less than 3%, in 1,000,000 flips less than 0.1%, ...

But what is true is that the fraction of heads flipped is likely to be *close* to half of the flips, and the more flips, the closer the fraction is likely to be to 1/2. For example, the chance that the fraction of heads is within 5% of 1/2 is

- more than 24% in 10 flips,
- more than 38% in 100 flips,
- more than 56% in 200 flips, and
- more than 89% in 1000 flips.

These numbers illustrate the single most important phenomenon of probability: the average value from repeated experiments is likely to be close to the expected value of one experiment. And it gets more likely to be closer as the number of experiments increases. This result was first formulated and proved by Jacob D. Bernoulli in his book *Ars Conjectandi* (The Art of Guessing) published posthumously in 1713. In his Introduction, Bernoulli comments that[1]

> even the stupidest man—by some instinct of nature *per se* and by no previous instruction (this is truly amazing)—knows for sure that the more observations . . . that are taken, the less the danger will be of straying from the mark.

But he goes on to argue that this instinct should not be taken for granted:

[1] These quotes are taken from Grinstead & Snell, *Introduction to Probability*, American Mathematical Society, p. 310.

> Something further must be contemplated here which perhaps no one has thought about until now. It certainly remains to be inquired whether after the number of observations has been increased, the probability... of obtaining the true ratio... finally exceeds any given degree of certainty; or whether the problem has, so to speak, its own asymptote—that is, whether some degree of certainty is given which one can never exceed.

Here's how to give a technical formulation of the question Bernoulli wants us to contemplate. Repeatedly performing some random experiment corresponds to defining n random variables equal to the results of n trials of the experiment. That is, we let $G_1, \ldots, G_n$ be independent random variables with the same distribution and the same expectation, μ. Now let A_n be the average of the results, that is,

$$A_n ::= \frac{\sum_{i=1}^{n} G_i}{n}.$$

How sure we can be that the average value, A_n, will be close to μ? By letting n grow large enough, can we be as certain as we want that the average will be close, or is there is some irreducible degree of uncertainty that remains no matter how many trials we perform? More precisely, given any positive tolerance, ϵ, how sure can we be that the average, A_n, will be within the tolerance of μ as n grows? In other words, we are asking about the limit

$$\lim_{n \to \infty} \Pr\{|A_n - \mu| < \epsilon\}.$$

Bernuolli asks if we can be sure this limit approaches certainty, that is, equals one, or whether it approaches some number slightly less than one that cannot be increased to one no matter how many times we repeat the experiment. His answer is that the limit is indeed one. This result is now known as the Weak Law of Large Numbers. Bernoulli says of it:

> Therefore, this is the problem which I now set forth and make known after I have pondered over it for twenty years. Both its novelty and its very great usefulness, coupled with its just as great difficulty, can exceed in weight and value all the remaining chapters of this thesis.

With the benefit of three centuries of mathematical development since Bernoulli, it will be a lot easier for us to resolve Bernoulli's questions than it originally was for him.

1.2 The Weak Law of Large Numbers

The Weak Law of Large Numbers crystallizes, and confirms, the intuition of Bernoulli's "stupidest man" that the average of a large number of independent trials is less and less likely to be outside a smaller and smaller tolerance around the expectation as the number of trials grows.

Theorem 1.1. *[Weak Law of Large Numbers] Let $G_1, \ldots, G_n, \ldots$ be independent variables with the same distribution and the same expectation, μ. For any $\epsilon > 0$,*

$$\lim_{n \to \infty} \Pr\left\{\left|\frac{\sum_{i=1}^{n} G_i}{n} - \mu\right| \geq \epsilon\right\} = 0.$$

This Law gives a high-level description of a fundamental probabilistic phenomenon, but as it stands it does not give enough information to be of practical use. The main problem is that it does not say anything about the *rate* at which the limit is approached. That is, how big must n be to be within a given tolerance of the expected value with a specific desired probability? This information is essential in applications. For example:

- Suppose we want to estimate the number of voters who are registered Republicans. Exactly *how many* randomly selected voters should we poll in order to be sure that 99% of the time, the average number of Republicans in our poll is within 1/2%, that is, within 0.005, of the actual percentage in the whole country?

- Suppose we want to estimate the number of fish in a lake. Our procedure will be to catch, tag and release 500 fish caught in randomly selected locations in the lake at random times of day. Then we wait a few days, and catch another 100 random fish. Suppose we discover that 10 of the 100 were previously tagged. Assuming that our 500 tagged fish represent the same proportion of the whole fish population as the ones in our sample of 100, we would estimate that the total fish population was 5000. But how confident can we be of this? Specifically, *how confident* should we be that our estimate of 5000 is within 20% of the actual fish population?

- Suppose we want to estimate the average size of fish in the lake by taking the average of the sizes of the 500 in our initial catch. How confident can we be that this average is within 2% of the average size of all the fish in the lake?

In these Notes we first prove a basic result we call the Pairwise Independent Sampling Theorem. It provides the additional information about rate of convergence we need to calculate numerical answers to questions such as those above. It simply summarizes in very slightly more general form, the reasoning we used in Notes 13 to determine the sample size needed to estimate voter preference.

The Sampling Theorem follows from Chebyshev's Theorem and properties of the variance of a sum of independent variables. A version of the Weak Law of Large Numbers will then be an easy corollary of the Pairwise Independent Sampling Theorem.

1.3 Pairwise Independent Sampling

Theorem 1.2. *[Pairwise Independent Sampling] Let*

$$S_n ::= \sum_{i=1}^{n} G_i \tag{1}$$

where $G_1, \ldots, G_n$ are pairwise independent variables with the same mean, μ, and deviation, σ. Then

$$\Pr\left\{\left|\frac{S_n}{n} - \mu\right| \geq x\right\} \leq \frac{1}{n}\left(\frac{\sigma}{x}\right)^2. \tag{2}$$

Proof. We observe first that the expectation of S_n/n is μ:

$$\begin{aligned}
\mathrm{E}\left[\frac{S_n}{n}\right] &= \mathrm{E}\left[\frac{\sum_{i=1}^n G_i}{n}\right] && \text{(by def. of } S_n\text{)} \\
&= \frac{\sum_{i=1}^n \mathrm{E}[G_i]}{n} && \text{(linearity of expectation)} \\
&= \frac{\sum_{i=1}^n \mu}{n} \\
&= \frac{n\mu}{n} = \mu. && (3)
\end{aligned}$$

The second important property of S_n/n is that its variance is the variance of G_i divided by n:

$$\begin{aligned}
\mathrm{Var}\left[\frac{S_n}{n}\right] &= \frac{1}{n^2}\mathrm{Var}[S_n] && (\mathrm{Var}[aR] = a^2\,\mathrm{Var}[R]) \\
&= \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^n G_i\right] && \text{(def. of } S_n\text{)} \\
&= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}[G_i] && \text{(pairwise independent additivity)} \\
&= \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}. && (4)
\end{aligned}$$

This is enough to apply Chebyshev's Bound and conclude:

$$\begin{aligned}
\Pr\left\{\left|\frac{S_n}{n} - \mu\right| \geq x\right\} &\leq \frac{\mathrm{Var}[S_n/n]}{x^2}. && \text{(Chebyshev's bound)} \\
&= \frac{\sigma^2/n}{x^2} && \text{(by (4))} \\
&= \frac{1}{n}\left(\frac{\sigma}{x}\right)^2.
\end{aligned}$$

□

Theorem 1.2 provides a precise general statement about how the average of independent samples of a random variable approaches the mean. It generalizes to many cases when S_n is the sum of independent variables whose mean and deviation are not necessarily all the same, though we shall not develop such generalizations here.

1.4 Proof of the Weak Law

Theorem 1.3. *[Pairwise Independent Weak Law] Let*

$$S_n ::= \sum_{i=1}^n G_i,$$

where $G_1, \ldots, G_n, \ldots$ are pairwise independent variables with the same expectation, μ, and standard deviation, σ. For any $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\left\{\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right\} = 0.$$

Proof. Choose x in Theorem 1.2 to be ϵ. Then, given any $\delta > 0$, choose n large enough to make $(\sigma/x)^2/n < \delta$. By Theorem 1.2,

$$\Pr\left\{\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right\} < \delta.$$

So the limiting probability must equal zero. □

Notice that for the sake of simplicity and motivation, we chose to state at the outset a Weak Law, namely Theorem 1.1, that does not mention variance, but only requires finite expectation. On the other hand, Theorem 1.3 requires finite variance, and so does not directly imply Theorem 1.1.

But notice also that Theorem 1.1 requires that the variables have the *same distribution* and are the *mutually* independent. On the other hand, Theorem 1.3 requires only pairwise independence and does not require the distributions to be the same, only that expectations and variances are the same. The pairwise independent version, Theorem 1.3, proved above is enough for nearly all applications. Theorem 1.1 only comes into play when the variance is unbounded; this case is not important to us, so we omit a proof of Theorem 1.1.

1.4.1 Oscillations in the Average

A weakness of both the Weak Law as well as our Pairwise Independence Sampling Theorem 1.2 is that neither provides any information about the way the average value of the observations may be expected to *oscillate* in the course of repeated experiments. In later Notes we will briefly consider a *Strong* Law of Large Numbers which deals with the oscillations. Such oscillations may not be important in our example of polling about Gore's popularity or of birthday matches, but they are critical in gambling situations, where large oscillations can bankrupt a player, even though the player's average winnings are assured in the long run. As the famous economist Keynes is alleged to have remarked, the problem is that "In the long run, we are all dead."

1.5 Confidence Levels

Now suppose as in Notes 13, that a pollster uses the Pairwise Independence Sampling Theorem, or even better, an estimation of the binomial distribution, to conclude that a sample of 662 random voters will yield an estimate of the fraction, p, of voters who prefer Gore that is in the interval $p \pm 0.04$ with probability at least 0.95.

The pollster takes his sample and finds that 364 prefer Gore. It's tempting, **but sloppy**, to say that this means

> *"With probability 0.95, the fraction, p, of voters who prefer Gore is* $364/662 \pm 0.04$. *But* $364/662 - 0.04 > 0.50$, *so there is a 95% chance that more than half the voters prefer Gore."*

What's objectionable about this statement is that it talks about the probability or "chance" that a real world fact is true, namely that the actual fraction, p, of voters favoring Gore is more than 0.50. But p is what it is, and it simply makes no sense to talk about the probability that it is something else. For example, suppose p is actually 0.49; then it's nonsense to ask about the probability that it is within 0.04 of 365/662 – it simply isn't.

A more careful summary of what we have accomplished goes this way:

> We have described a probabilistic procedure for estimating the value of the actual fraction, p. The probability that *our estimation procedure* will yield a value within 0.04 of p is 0.95.

This is a bit of a mouthful, so special phrasing closer to the sloppy language is commonly used. The pollster would describe his conclusion by saying that

> At the 95% *confidence level*, the fraction of voters who prefer Gore is $364/662 \pm 0.04$.

It's important to remember that confidence levels refer to the results of estimation procedures for real-world quantities. The real-world quantity being estimated is typically unknown, but fixed; it is not a random variable, so it makes no sense to talk about the probability that it has some property.

2 Random Walks

2.1 Gamblers' Ruin

Random Walks nicely model many natural phenomena in which a person, or particle, or process takes steps in a randomly chosen sequence of directions. For example in Physics, three-dimensional random walks are used to model Brownian motion and gas diffusion. In Computer Science, the Google search engine uses random walks through the graph of world-wide web links to determine the relative importance of websites. In Finance Theory, there is continuing debate about the degree to which one-dimensional random walks can explain the moment-to-moment or day-to-day fluctuations of market prices. In these Notes we consider 1-dimensional random walks: walks along a straight line. Our knowledge of expectation and deviation will make 1-dimensional walks easy to analyze, but even these simple walks exhibit probabilistic behavior that can be astonishing.

In the Mathematical literature, random walks are for some reason traditionally discussed in the context of some social vice. A one-dimensional random walk is often described as the path of a drunkard who randomly staggers left or right at each step. In the rest of these Notes, we examine one-dimensional random walks using the language of gambling. In this case, a position during the walk is a gambler's cash-on-hand or *capital*, and steps on the walk are bets whose random outcomes increase or decrease his capital. We will be interested in two main questions:

1. What is the probability that the gambler wins?
2. How long must the gambler expect to wait for the walk to end?

In particular, we suppose a gambler starts with n dollars. He makes a sequence of \$1 bets. If he wins an individual bet, he gets his money back plus another \$1. If he loses, he loses the \$1. In each bet, he wins with probability $p > 0$ and loses with probability $q ::= 1 - p > 0$. The gambler plays until either he is bankrupt or increases his capital to a goal amount of T dollars. If he reaches his goal, then he is called an overall *winner*, and his *profit* will be $m ::= T - n$ dollars. If his capital reaches zero dollars before reaching his goal, then we say that he is "ruined" or *goes broke*.

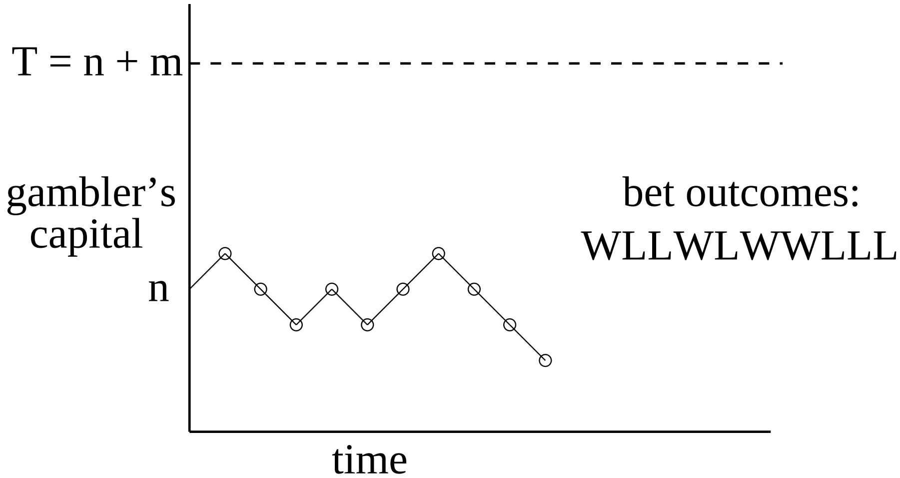

Figure 1: *This is a graph of the gambler's capital versus time for one possible sequence of bet outcomes. At each time step, the graph goes up with probability* p *and down with probability* $1-p$*. The gambler continues betting until the graph reaches either 0 or* $T = n + m$*.*

The gambler's situation as he proceeds with his $1 bets is illustrated in Figure 1. The random walk has boundaries at 0 and T. If the random walk ever reaches either of these boundary values, then it terminates. We want to determine the probability, w, that the walk terminates at boundary T, namely, the probability that the gambler is a winner.

In a fair game, $p = q = 1/2$. The corresponding random walk is called *unbiased*. The gambler is more likely to win if $p > 1/2$ and less likely to win if $p < 1/2$; the corresponding random walks are called *biased*.

Example 2.1. Suppose that the gambler is flipping a coin, winning $1 on Heads and losing $1 on Tails. Also, the gambler's starting capital is $n = 500$ dollars, and he wants to make $m = 100$ dollars. That is, he plays until he goes broke or reaches a goal of $T = n + m = \$600$. What is the probability that he is a winner? We will show that in this case the probability $w = 5/6$. So his chances of winning are really very good, namely, 5 chances out of 6.

Now suppose instead, that the gambler chooses to play roulette in an American casino, always betting $1 on red. A roulette wheel has 18 black numbers, 18 red numbers, and 2 green numbers. In this game, the probability of winning a single bet is $p = 18/38 \approx 0.47$. It's the two green numbers that slightly bias the bets and give the casino an edge. Still, the bets are almost fair, and you might expect that the gambler has a reasonable chance of reaching his goal—the 5/6 probability of winning in the unbiased game surely gets reduced, but perhaps not too drastically. Not so! His odds of winning against the "slightly" unfair roulette wheel are less than 1 in 37,000. If that seems surprising, listen to this: *no matter how much money* the gambler has to start, *e.g.*, $5000, $50,000, $\$5 \cdot 10^{12}$, his odds are still less than 1 in 37,000 of winning a mere 100 dollars!

Moral: Don't play!

The theory of random walks is filled with such fascinating and counter-intuitive conclusions.

2.2 The Probability Space

Each random-walk game corresponds to a path like the one in Figure 1 that starts at the point $(n, 0)$. A winning path never touches the x axis and ends when it first touches the line $y = T$. Likewise, a losing path never touches the line $y = T$ and ends when it first touches the x axis.

Any length k path can be characterized by the history of wins and losses on individual $1 bets, so we use a length k string of W's and L's to model a path, and assign probability $p^r q^{k-r}$ to a string that contains r W's. The *outcomes* in our sample space will be precisely those string corresponding to winning or losing walks.

What about the infinite walks in which the gambler plays forever, neither reaching his goal nor going bankrupt? A recitation problem will show the probability of playing forever is zero, so we don't need to include any such outcomes in our sample space.

As a sanity check on this definition of the probability space, we should verify that the sum of the outcome probabilities is one, but we omit this calculation.

2.3 The Probability of Winning

2.4 The Unbiased Game

Let's begin by considering the case of a fair coin, that is, $p = 1/2$, and determine the probability, w, that the gambler wins. We can handle this case by considering the expectation of the random variable G equal to the gambler's dollar gain. That is, $G = T - n$ if the gambler wins, and $G = -n$ if the gambler loses, so

$$\mathrm{E}[G] = w(T - n) - (1 - w)n = wT - n.$$

Notice that we're using the fact that the only outcomes are those in which the gambler wins or loses—there are no infinite games—so the probability of losing is $1 - w$.

Now let G_i be the amount the gambler gains on the ith flip: $G_i = 1$ if the gambler wins the flip, $G_i = -1$ if the gambler loses the flip, and $G_i = 0$ if the game has ended before the ith flip. Since the coin is fair, $\mathrm{E}[G_i] = 0$.

The random variable G is the sum of all the G_i's, so by linearity of expectation[2]

$$wT - n = E(G) = \sum_{i=1}^{\infty} E(G_i) = 0,$$

which proves

[2] We've been stung by paradoxes in this kind of situation, so we should be careful to check that the condition for infinite linearity of expectation is satisfied. Namely, we have to check that $\sum_{i=1}^{\infty} \mathrm{E}[|G_i|]$ converges.

In this case, $|G_i| = 1$ iff the walk is of length at least i, and $|G_i| = 0$ otherwise. So

$$\mathrm{E}[|G_i|] = \Pr\{\text{the walk is of length } \geq i\}.$$

But a recitation problem will show that there is a constant $r < 1$ such that

$$\Pr\{\text{the walk is of length } \geq i\} = O(r^i).$$

So the $\sum_{i=1}^{\infty} \mathrm{E}[|G_i|]$ is bounded term-by-term by a convergent geometric series, and therefore it also converges.

Theorem 2.2. *In the unbiased Gambler's Ruin game with probability $p = 1/2$ of winning each individual bet, with initial capital, n, and goal, T,*

$$\Pr\{\textit{the gambler is a winner}\} = \frac{n}{T}. \tag{5}$$

Example 2.3. Suppose we have \$100 and we start flipping a fair coin, betting \$1 with the aim of winning \$100. Then the probability of reaching the \$200 goal is $100/200 = 1/2$—the same as the probability of going bankrupt. In general, if $T = 2n$, then the probability of doubling your money or losing all your money is the same. This is about what we would expect.

Example 2.4. Suppose we have \$500 and we start flipping a fair coin, betting \$1 with the aim of winning \$100. So $n = 500, T = 600$, and $\Pr\{\text{win}\} = 500/600 = 5/6$, as we claimed at the outset.

Example 2.5. Suppose Albert starts with \$100, and Eric starts with \$10. They flip a fair coin, and every time a Head appears, Albert wins \$1 from Eric, and vice versa for Tails. They play this game until one person goes bankrupt. What is the probability of Albert winning?

This problem is identical to the Gambler's Ruin problem with $n = 100$ and $T = 100+10 = 110$. The probability of Albert winning is $100/110 = 10/11$, namely, the ratio of his wealth to the combined wealth. Eric's chances of winnning are $1/11$.

Note that although Albert will win most of the time, the game is still fair. When Albert wins, he only wins \$10; when he loses, he loses big: \$100. Albert's—and Eric's—expected win is zero dollars.

Another intuitive idea is confirmed by this analysis: the larger the gambler's initial stake, the larger the probability that he will win a fixed amount.

Example 2.6. If the gambler started with one million dollars instead of 500, but aimed to win the same 100 dollars as in the Example 2.4, the probability of winning would increase to $1M/(1M + 100) > .9999$.

2.5 A Recurrence for the Probability of Winning

To handle the case of a biased game we need a more general approach. We consider the probability of the gambler winning as a function of his initial capital. That is, let p and T be fixed, and let w_n be the gambler's probabiliity of winning when his initial capital is n dollars. For example, w_0 is the probability that the gambler will win given that he starts off broke; clearly, $w_0 = 0$. Likewise, $w_T = 1$.

Otherwise, the gambler starts with n dollars, where $0 < n < T$. Consider the outcome of his first bet. The gambler wins the first bet with probability p. In this case, he is left with $n + 1$ dollars and becomes a winner with probability w_{n+1}. On the other hand, he loses the first bet with probability $q ::= 1 - p$. Now he is left with $n - 1$ dollars and becomes a winner with probability w_{n-1}. Overall, he is a winner with probability $w_n = pw_{n+1} + qw_{n-1}$. Solving for w_{n+1} we have

$$w_{n+1} = \frac{w_n}{p} - w_{n-1}\frac{q}{p}. \tag{6}$$

This kind of inductive definition of a quantity w_{n+1} in terms of a linear combination of values w_k for $k < n + 1$ is called a *homogeneous linear recurrence*. There is a simple general method for solving such recurrences which we now illustrate. The method is based on a guess that the form of the solution is $w_n = c^n$ for some $c \neq 0$. It's not obvious why this is a good guess, but we now show how to find the constant c and verify the guess.

Namely, from (6) we have

$$w_{n+1} - \frac{w_n}{p} + w_{n-1}\frac{q}{p} = 0. \tag{7}$$

If our guess is right, then this is equivalent to

$$c^{n+1} - \frac{c^n}{p} + c^{n-1}\frac{q}{p} = 0.$$

Now factoring out c^{n-1} gives

$$c^2 - \frac{c}{p} + \frac{q}{p} = 0.$$

Solving this quadratic equation in c yields two roots, $(1-p)/p$ and 1. So if we define $w_n ::= ((1-p)/p)^n = (q/p)^n$, then (7), and hence (6) is satisifed. We can also define $w_n ::= 1^n$ and satisfy (7). Since the lefthand side of (7) is zero using either definition, it follows that any definition of the form

$$w_n ::= A\left(\frac{q}{p}\right)^n + B \cdot 1^n$$

will also satisfy (7). Now our boundary conditions, namely the values of w_0 and w_T, let us solve for A and B:

$$\begin{aligned} 0 &= w_0 = A + B, \\ 1 &= w_T = A\left(\frac{q}{p}\right)^T + B, \end{aligned}$$

so

$$A = \frac{1}{(q/p)^T - 1}, \qquad B = -A, \tag{8}$$

and therefore

$$w_n = \frac{(q/p)^n - 1}{(q/p)^T - 1}. \tag{9}$$

Our derivation of (9) ensures that it gives a formula for w_n which satisfies (6) and has the right values at $n = 0$ and $n = T$. Moreover, the values determined by (9) are the *only ones* that satisfy (6) and the boundary conditions at 0 and T, though we won't prove this. This implies that the Gambler's probability of winning is indeed given by (9).

The solution (9) only applies to biased walks, since we require $p \neq q$ so the denominator is not zero. That's ok, since we already worked out that the case when $p = q$ in Theorem 2.2. So we have shown:

Theorem 2.7. *In the biased Gambler's Ruin game with probability, $p \neq 1/2$, of winning each bet, with initial capital, n, and goal, T,*

$$\Pr\{\textit{the gambler is a winner}\} = \frac{(q/p)^n - 1}{(q/p)^T - 1}. \tag{10}$$

The expression (10) for the probability that the Gambler wins in the biased game is a little hard to interpret. There is a simpler upper bound which is nearly tight when the gambler's starting capital is large.

Suppose that $p < 1/2$; that is, the game is biased *against* the gambler. Then both the numerator and denominator in the quotient in (10) are positive, and the quotient is less than one. So adding 1 to both the numerator and denominator increases the quotient[3], and the bound (10) simplifies to $(q/p)^n/(q/p)^T = (p/q)^{T-n}$, which proves

Corollary 2.8. *In the Gambler's Ruin game biased against the Gambler, that is, with probability $p < 1/2$ of winning each bet, with initial capital, n, and goal, T,*

$$\Pr\{\textit{the gambler is a winner}\} < \left(\frac{p}{q}\right)^m, \tag{11}$$

where $m ::= T - n$.

The amount $m = T - n$ is called the Gambler's *intended profit*. So the gambler gains his intended profit, m, before going broke with probability at most $(p/q)^m$. Notice that this upper bound does not depend on the gambler's starting capital, but only on his intended profit. The consequences of this are amazing:

Example 2.9. Suppose that the gambler starts with \$500 aiming to profit \$100, this time by making \$1 bets on red in roulette. By (11), the probability, w_n, that he is a winner is less than

$$\left(\frac{18/38}{20/38}\right)^{100} = \left(\frac{9}{10}\right)^{100} < \frac{1}{37,648}.$$

This is a dramatic contrast to the unbiased game, where we saw in Example 2.4 that his probability of winning was 5/6.

Example 2.10. We also observed that with \$1,000,000 to start in the unbiased game, he was almost certain to win \$100. But betting against the "slightly" unfair roulette wheel, even starting with \$1,000,000, his chance of winning \$100 remains less than 1 in 37,648! He will almost surely lose all his \$1,000,000. In fact, because the bound (11) depends only on his intended profit, his chance of going up a mere \$100 is less than 1 in 37,648 *no matter how much money he starts with*!

The bound (11) is exponential in m. So, for example, doubling his intended profit will square his probability of winning.

Example 2.11. The probability that the gambler's stake goes up 200 dollars before he goes broke playing roulette is at most

$$(9/10)^{200} = ((9/10)^{100})^2 = \left(\frac{1}{37,648}\right)^2,$$

which is about 1 in 70 billion.

---

[3] If $0 < a < b$, then

$$\frac{a}{b} < \frac{a+1}{b+1},$$

because

$$\frac{a}{b} = \frac{a(1+1/b)}{b(1+1/b)} = \frac{a+a/b}{b+1} < \frac{a+1}{b+1}.$$

The odds of winning a little money are not so bad.

Example 2.12. Applying the exact formula (10), we find that the probability of winning \$10 before losing \$10 is

$$\frac{\left(\frac{20/38}{18/38}\right)^{10} - 1}{\left(\frac{20/38}{18/38}\right)^{20} - 1} = 0.2585\ldots$$

This is somewhat worse than the 1 in 2 chance in the fair game, but not dramatically so.

Thus, in the fair case, it helps a lot to have a large bankroll, whereas in the unfair case, it doesn't help much.

2.6 Intuition

Why is the gambler so unlikely to make money when the game is slightly biased against him? Intuitively, there are two forces at work. First, the gambler's capital has random upward and downward *swings* due to runs of good and bad luck. Second, the gambler's capital will have a steady, downward *drift*, because he has a small, negative expected return on every bet. The situation is shown in Figure 2.

For example, in roulette the gambler wins a dollar with probability $9/19$ and loses a dollar with probability $10/19$. Therefore, his expected return on each bet is $9/10 - 10/19 = -1/19 \approx -0.053$ dollars. That is, on each bet his capital is expect to drift downward by a little over 5 cents.

Our intuition is that if the gambler starts with a trillion dollars, then he will play for a very long time, so at some point there should be a lucky, upward swing that puts him \$100 ahead. The problem is that his capital is steadily drifting downward. If the gambler does not have a lucky, upward swing early on, then he is doomed. After his capital drifts downward a few hundred dollars, he needs a huge upward swing to save himself. And such a huge swing is extremely improbable. As a rule of thumb, *drift dominates swings* in the long term.

We can quantify these drifts and swings. After k rounds for $k \leq \min(m, n)$, the number of wins by our player has a binomial distribution with parameters $p < 1/2$ and k. His expected win on any single bet is $p - q = 2p - 1$ dollars, so his expected capital is $n - k(1 - 2p)$. Now to be a winner, his actual number of wins must exceed the expected number by $m + k(1 - 2p)$. But we saw before that the binomial distribution has a standard deviation of only $\sqrt{kp(1-p)}$. So for the gambler to win, he needs his number of wins to deviate by

$$\frac{m + k(1 - 2p)}{\sqrt{kp(1 - 2p)}} = \Theta(\sqrt{k})$$

times its standard deviation. In our study of binomial tails we saw that this was extremely unlikely.

In a fair game, there is no drift; swings are the only effect. In the absence of downward drift, our earlier intuition is correct. If the gambler starts with a trillion dollars then almost certainly there will eventually be a lucky swing that puts him \$100 ahead.

If we start with \$10 and play to win only \$10 more, then the difference between the fair and unfair games is relatively small. We saw that the probability of winning is $1/2$ versus about $1/4$. Since swings of \$10 are relatively common, the game usually ends before the gambler's capital can drift very far. That is, the game does not last long enough for drift to dominate the swings.

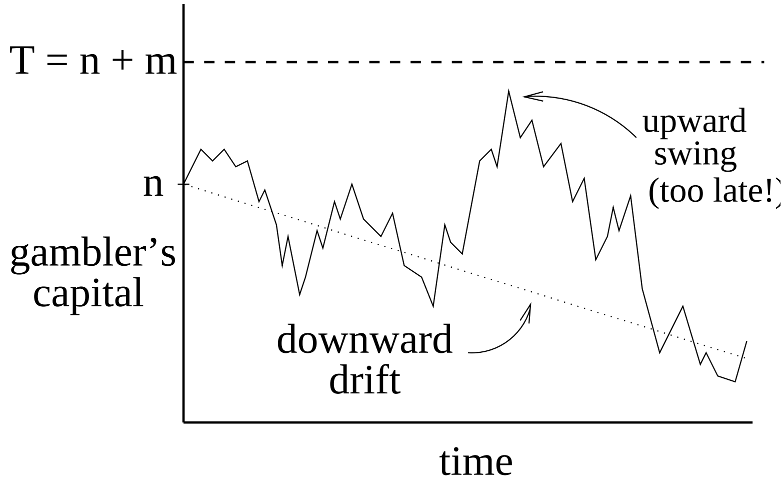

Figure 2: *In an unfair game, the gambler's capital swings randomly up and down, but steadily drifts downward. If the gambler does not have a winning swing early on, then his capital drifts downward, and later upward swings are insufficient to make him a winner.*

2.7 How Long a Walk?

Now that we know the probability, w_n, that the gambler is a winner in both fair and unfair games, we consider how many bets he needs on average to either win or go broke.

2.8 Duration of a Biased Walk

Let Q be the number of bets the gambler makes until the game ends. Since the gambler's expected win on any bet is $2p - 1$, Wald's Theorem should tell us that his game winnings, G, will have expectation $\mathrm{E}[Q](2p-1)$. That is,

$$\mathrm{E}[G] = (2p-1)\,\mathrm{E}[Q], \tag{12}$$

In an unbiased game (12) is trivially true because both $2p - 1$ and the expected overall winnings, $\mathrm{E}[G]$, are zero. On the other hand, in the unfair case, $2p - 1 \neq 0$. Also, we know that

$$\mathrm{E}[G] = w_n(T-n) - (1-w_n)n = w_n T - n.$$

So assuming (12), we conclude

Theorem 2.13. *In the biased Gambler's Ruin game with initial capital, n, goal, T, and probability, $p \neq 1/2$, of winning each bet,*

$$\mathrm{E}[\textit{number of bets till game ends}] = \frac{\Pr\{\textit{gambler is a winner}\}\,T - n}{2p-1}. \tag{13}$$

The only problem is that (12) is not a special case of Wald's Theorem because $G = \sum_{i=1}^{Q} G_i$ is not a sum of *nonnegative* variables: when the gambler loses the ith bet, the random variable G_i equals

-1. However, this is easily dealt with.[4]

Example 2.14. If the gambler aims to profit \$100 playing roulette with n dollars to start, he can expect to make $((n+100)/37,648-n)/(2(18/38)-1) \approx 19n$ bets before the game ends. So he can enjoy playing for a good while before almost surely going broke.

2.9 Duration of an Unbiased Walk

This time, we need the more general approach of recurrences to handle the unbiased case. We consider the expected number of bets as a function of the gambler's initial capital. That is, for fixed p and T, let e_n be the expected number of bets until the game ends when the gambler's initial capital is n dollars. Since the game is over in no steps if $n = 0$ or T, the boundary conditions this time are $e_0 = e_T = 0$.

Otherwise, the gambler starts with n dollars, where $0 < n < T$. Now by the conditional expectation rule, the expected number of steps can be broken down into the expected number of steps given the outcome of the first bet weighted by the probability of that outcome. That is,

$$e_n = p\,\mathrm{E}\left[Q \mid \text{gambler wins first bet}\right] + q\,\mathrm{E}\left[Q \mid \text{gambler loses first bet}\right].$$

But after the gambler wins the first bet, his capital is $n+1$, so he can expect to make another e_{n+1} bets. That is,

$$\mathrm{E}\left[Q \mid \text{gambler wins first bet}\right] = 1 + e_{n+1},$$

and similarly,

$$\mathrm{E}\left[Q \mid \text{gambler loses first bet}\right] = 1 + e_{n-1}.$$

So we have

$$e_n = p(1+e_{n+1}) + q(1+e_{n-1}) = pe_{n+1} + qe_{n-1} + 1,$$

which yields the linear recurrence

$$e_{n+1} = \frac{e_n}{p} - \frac{q}{p}e_{n-1} - \frac{1}{p}.$$

[4] The random variable $G_i + 1$ is nonnegative, and $\mathrm{E}\left[G_i + 1 \mid Q \ge i\right] = \mathrm{E}\left[G_i \mid Q \ge i\right] + 1 = 2p$, so by Wald's Theorem

$$\mathrm{E}\left[\sum_{i=1}^{Q}(G_i+1)\right] = 2p\,\mathrm{E}\left[Q\right]. \tag{14}$$

But

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{Q}(G_i+1)\right] &= \mathrm{E}\left[\sum_{i=1}^{Q}G_i + \sum_{i=1}^{Q}1\right] \\
&= \mathrm{E}\left[(\sum_{i=1}^{Q}G_i) + Q\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{Q}G_i\right] + \mathrm{E}\left[Q\right] \\
&= \mathrm{E}\left[G\right] + \mathrm{E}\left[Q\right]. \qquad (15)
\end{aligned}$$

Now combining (14) and (15) confirms the truth of our assumption (12).

For $p = q = 1/2$, this equation simplifies to

$$e_{n+1} = 2e_n - e_{n-1} - 2. \tag{16}$$

There is a general theory for solving linear recurrences like (16) in which the value at $n+1$ is a linear combination of values at some arguments $k < n+1$ plus another simple term—in this case plus the constant -2. This theory implies that

$$e_n = (T - n)n. \tag{17}$$

Fortunately, we don't need the general theory to *verify* this solution. Equation (17) can be verified routinely from the boundary conditions and (16) using strong induction on n.

So we have shown

Theorem 2.15. *In the unbiased Gambler's Ruin game with initial capital, n, and goal, T, and probability, $p = 1/2$, of winning each bet,*

$$\mathrm{E}\left[\textit{number of bets till game ends}\right] = n(T - n). \tag{18}$$

Another way to phrase Theorem 2.15 is

$$\mathrm{E}\left[\text{number of bets till game ends}\right] = \text{initial capital} \cdot \text{intended profit}. \tag{19}$$

Now for example, we can conclude that if the gambler starts with \$10 dollars and plays until he is broke or ahead \$10, then $10 \cdot 10 = 100$ bets are required on average. If he starts with \$500 and plays until he is broke or ahead \$100, then the expected number of bets until the game is over is $500 \times 100 = 50,000$.

Notice that (19) is a very simple answer that cries out for an intuitive proof, but we have not found one.

2.10 Quit While You Are Ahead

Suppose that the gambler never quits while he is ahead. That is, he starts with $n > 0$ dollars, ignores any goal T, but plays until he is flat broke. Then it turns out that if the game is not favorable, *i.e.*, $p \leq 1/2$, the gambler is sure to go broke. In particular, he is even sure to go broke in a "fair" game with $p = 1/2$. [5]

Lemma 2.16. *If the gambler starts with one or more dollars and plays a fair game until he is broke, then he will go broke with probability 1.*

Proof. If the gambler has initial capital n and goes broke in a game without reaching a goal T, then he would also go broke if he were playing and ignored the goal. So the probability that he will lose if he keeps playing without stopping at any goal T must be at least as large as the probability that he loses when he has a goal $T > n$.

But we know that in a fair game, the probability that he loses is $1 - n/T$. This number can be made arbitrarily close to 1 by choosing a sufficiently large value of T. Hence, the probability of his losing while playing without any goal has a lower bound arbitrarily close to 1, which means it must in fact be 1. □

[5] If the game is favorable to the gambler, *i.e.*, $p > 1/2$, then we could show that there is a positive probability that the gambler will play forever, but we won't examine this case in these Notes.

So even if the gambler starts with a million dollars and plays a perfectly fair game, he will eventually lose it all with probability 1. In fact, if the game is unfavorable, then Theorem 2.13 and Corollary 2.8 imply that his expected time to go broke is essentially proportional to his initial capital, *i.e.*, $\Theta(n)$.

But there is good news: if the game is fair, he can "expect" to play for a very long time before going broke; in fact, he can expect to play forever!

Lemma 2.17. *If the gambler starts with one or more dollars and plays a fair game until he goes broke, then his expected number of plays is infinite.*

Proof. Consider the gambler's ruin game where the gambler starts with initial capital n, and let u_n be the expected number of bets for the *unbounded* game to end. Also, choose any $T \geq n$, and as above, let e_n be the expected number of bets for the game to end when the gambler's goal is T.

The unbounded game will have a larger expected number of bets compared to the bounded game because, in addition to the possibility that the gambler goes broke, in the bounded game there is also the possibility that the game will end when the gambler reaches his goal, T. That is,

$$u_n \geq e_n.$$

So by (17),

$$u_n \geq n(T - n).$$

But $n \geq 1$, and T can be any number greater than or equal to n, so this lower bound on u_n can be arbitrarily large. This implies that u_n must be infinite.

Now by Lemma 2.16, with probability 1, the unbounded game ends when the gambler goes broke. So the expected time for the unbounded game to *end* is the *same* as the expected time for the gambler to *go broke*. Therefore, the expected time to go broke is infinite. □

In particular, even if the gambler starts with just one dollar, his expected number of plays before going broke is infinite! Of course, this does not mean that it is likely he will play for long. For example, there is a 50% chance he will lose the very first bet and go broke right away.

Lemma 2.17 says that the gambler can "expect" to play forever, while Lemma 2.16 says that with probability 1 he will go broke. These Lemmas sound contradictory, but our analysis showed that they are not. A moral is that naive intuition about "expectation" is misleading when we consider limiting behavior according to the technical mathematical definition of expectation.

3 Infinite Expectation

So what are we to make of such a random variable with infinite expectation? For example, suppose we repeated the experiment of having the gambler make fair bets with initial stake one dollar until he went broke, and we kept a record of the average number of bets per experiment. Our theorems about deviation from the mean only apply to random variables with finite expectation, so they don't seem relevant to this situation. But in fact they are.

For example, let Q be the number of bets required for the gambler to go broke in a fair game starting with one dollar. We could use some of our combinatorial techniques to show that

$$\Pr\{Q = m\} = \Theta(m^{-3/2}). \tag{20}$$

This implies that

$$\mathrm{E}[Q] = \Theta\left(\sum_{m=1}^{\infty} m \cdot m^{-3/2}\right) = \Theta\left(\sum_{m=1}^{\infty} m^{-1/2}\right).$$

We know this last series is divergent, so we have another proof that Q has infinite expectation.

But suppose we let $R ::= Q^{1/5}$. Then the estimate (20) also lets us conclude that

$$\mathrm{E}[R] = \Theta\left(\sum_{m=1}^{\infty} m^{-13/10}\right)$$

and

$$\mathrm{E}\left[R^2\right] = \Theta\left(\sum_{m=1}^{\infty} m^{-11/10}\right).$$

Since both these series are convergent, we can conclude that $\mathrm{Var}[R]$ is finite. Now our theorems about deviation can be applied to tell us that the average *fifth root* of the number of bets to go broke is very likely to converge to a finite expected value.

We won't go further into the details, but the moral of this discussion is that our results about deviation from a finite mean can still be applied to natural models like random walks where variables with infinite expectation may play an important role.

4 Review of Markov, Chebyshev and Binomial bounds

Let us review the methods we have for bounding deviation from the mean via the following example. Assume that I.Q. is made up of thinkatons; each thinkaton fires independently with a 10% chance. We have 1000 thinkatons in all, and I.Q. is the number of thinkatons that fire. What is the probability of having a truly extraordinary IQ[6] of 228?

So the I.Q. is a Binomial distribution with $n = 1000, p = 0.1$. Hence, $\mathrm{E}[\text{I.Q.}] = 100, \sigma_{\text{I.Q.}} = \sqrt{0.09 \times 1000} = 9.48$.

An I.Q. of 228 is $128/9.48 > 13.5$ standard deviations away.

Let us compare the methods we have for bounding the probability of this I.Q..

1. Markov:
$$\Pr\{\text{I.Q.} \geq 228\} \leq \frac{100}{228} < 0.44$$

2. Chebyshev:
$$\Pr\{\text{I.Q.} - 100 \geq 128\} \leq \frac{1}{13.5^2 + 1} < \frac{1}{183}$$

[6] Marilyn vos Savant, author of the "Ask Marilyn" newspaper column mentioned in Notes 10 is alleged to have an I.Q. of 228.

3. Binomial tails:

$$\Pr\{\text{I.Q. } \geq 228\} = \Pr\{1000 - \text{I.Q. } \leq 772\} = F_{0.9,1000}(772) < 10^{-31}$$

   Here we used the formula for the binomial tail from Notes 12 with $p = 0.9, n = 1000, \alpha = 0.772$.

Most applications involve probability distributions for which the Markov and Chebyshev bounds are very weak. So in general, the more we know about the distribution, the better bounds we can obtain on probability of deviation. Of course there are distributions where the Markov and Chebyshev bounds are best possible; in those cases, knowledge of the distribution obviously does not lead to better bounds.