Lottery Principles:

History, Statistics, and Economics

Over the past fifty years State lotteries have become a significant source of revenues as well as a popular game. We all know that lottery tickets are statistically "lost money"—they must be or the State wouldn't offer them! What we don't all know is that there are some lottery formats that have generated systematic profits for savvy players, and that there are some ways to improve the chances of a win.

What are the mathematical and statistical mechanics underlying lotteries? Why do people buy these almost certain losers? Are there ways to improve the odds, perhaps even coming out ahead?

These are among the questions addressed in this blog and, hopefully, answered to your satisfaction. We begin with the rudiments of lottery analysis: the mathematics of counting the chances of winning, and of computing the associated probabilities. Then we consider the application of those principles to the analysis of typical lotteries including Powerball games. Finally we turn to lotteries that might be "winnable" because of the specific format, like roll-downs. The blog ends with a discussion of the value of statistical strategies for success, and with some insights into the prospect of having to share your Jackpot with another player.

Only the basic mathematics is used in the text; more arcane math is relegated to footnotes or to the *Addendum*.

Some pages are left blank for formatting purposes.

This is a draft of a work-in-progress. I'd appreciate constructive comments.
Please email them to webmaster @ fortunearchive.com (ignore spaces)

2018

Page Intentionally Left Blank

1. Lottery Basics

Like all games of chance, lotteries demand a simple basic skill—counting. But the counting methods used are of a special type called Combinatorial Mathematics. The reason for this special class of arithmetic is not that the counting is esoteric in principle; it is simply that the numbers involved can be so large that a compact method is necessary.

Combinatorial mathematics addresses questions that can be phrased in the form, *How many ways can N things be placed into groups of size n?* Often we phrase the question more casually as, *How many ways can N things be taken n at a time.* An example is rolling dice. Suppose you roll a six-sided die six times. How many numbers can come up *if* the order counts and there is no replacement (a number can't be used twice). The answer is simple: there are 6 numbers that can come up on the first roll, five different numbers on the second roll, four on the third, and so on. This gives 6•5•4•3•2•1 = 720 ways; we call these *permutations*. The 720 permutations will include repetitions of the same 6 numbers; for example 65321, 132465, 123456 are three of the 720 permutations, each repeating the same 6 numbers but in a different order. We can eliminate those repetitions to derive the number of *combinations*, each combination a unique collection of the first six integers. We will find only one *combination* of 6 integers in those 720 permutations.

Ok, that was easy! Now consider a real lottery. The player buys a ticket with 46 numbers from 1 to 46; he must select any 6 of those numbers with no number selected twice (that is, with no replacement); this is called a "46/6 lottery." The Jackpot winners (there can be more than one) must have selected the identical six numbers that the Lottery randomly selects in a public drawing; the order of the numbers is irrelevant. This is a question of combinations because the winner doesn't have to select the six numbers in any particular order—all that's required is that the six numbers match the winning six numbers.[1]

So how many distinct six-number groups can be selected from 46 numbers? You can list out all the possibilities, but with 6,744,109,680 possible permutations of 46 items taken 6 at a time, ultimately reducing to 9,336,819 combinations when repetitions are eliminated, you'll spend years at the task. You clearly need a quick-and dirty counting method.

And that is why we start with Combinatorial Mathematics.

[1] There is nothing to prevent lotteries from being based on permutations—but there is also nothing to be gained.

1.1 Permutations

We first need to understand a mathematical operation called a *factorial*, denoted by an integer followed by an exclamation mark. If we take the first N integers and multiply them together to get $N \cdot (N\text{-}1) \cdot (N\text{-}2) \cdots 3 \cdot 2 \cdot 1$, the result is called *N-factorial* and denoted as *N!*. Thus, multiplying the first 5 integers to get $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$ is the calculation of 5-factorial (5*!*).

The number of permutations of N things taken n at a time is calculated as

$$P_n^N = \frac{N!}{(N-n)!}$$

so, as we know, the number of permutations of 46 numbers taken 6 at a time is $P_6^{46} = \frac{46!}{40!} = 46 \cdot 45 \cdot 44 \cdot 43 \cdot 42 \cdot 41 = 6{,}744{,}109{,}680$.[2]

If the lottery required that you pick not only the correct six numbers but also in a particular order, the odds would be determined by the number of possible permutations. But Lottery Commissions don't care about the order in which you pick the correct numbers—they only care that you picked the correct numbers *in any order*.

1.2 Combinations

In general, for any n permutations there are n! identical ways of writing one combination: for example, P_6^{46} permutations include 6! = 720 possible repetitions of the same 6 numbers. To eliminate those repetitions we divide the number of *permutations* of 46 things into 6-thing groups (6,744,109,680) by the number of repetitions (720) to derive the number of *combinations* (9,366,819) of 46 things into 6-thing groups.

A more concise notation defining the number of combinations of N things in n-thing groups is

$$C_n^N = \frac{P_n^N}{n!} = \frac{N!}{(N-n)!n!}$$

so with 6,744,109,680 permutations of 46 numbers into six-number groups, the number of combinations is $C_6^{46} = \frac{46!}{(46-6)!6!} = 9{,}366{,}819$. Of those 9,366,819 combinations, only one is a match to the lottery's 6-number selection, so the *chance* that a ticket wins the Jackpot is 1-in-9,366,819, and the *probability* is $\frac{1}{9{,}366{,}819}$; this is a measly .00001068% probability of a Jackpot win.

---

[2] The 46! can be written $46 \cdot 45 \cdot 44 \cdot 43 \cdot 42 \cdot 41 \cdot 40!$ Dividing by 40! (the number of repetitions) leaves $46 \cdot 45 \cdot 44 \cdot 43 \cdot 42 \cdot 41$.

That is the essentials of combinatorial mathematics. It is the vehicle by which statistical analysis is done.

1.3 Statistical Analysis

You all hated statistics in college, but if you don't get the basics you are short-changing yourself at the lottery. Statistical analysis is the evaluation of prospects with risky outcomes.[3] How should you evaluate an opportunity that can have many—perhaps infinitely many— results, assuming that you know (or can accurately estimate) the "chances" of each result?

Consider a way-too-simple example of a risky decision. An oil wildcatter is contemplating drilling a new well. He knows neither the precise results in sales revenue and costs, nor the precise chances attached to each possible outcome. But he can estimate them from looking at records for drilling wells in similar geographical, geological, and topographical situations. This estimation process is called *statistical inference*.

Suppose he has done his homework and the results are distilled into Table 1, below. There is a 65 percent probability that the well will do "poorly" by losing \$25,000, but there is also a 35 percent probability it will earn a profit of \$55,000. Table 1 summarizes the information assuming a \$75 price of a barrel.

Table 1
Analysis of Prospects

Probability	Barrels	Sales *	Costs	Win/Loss
.65	500	\$50,000	\$75,000	-\$25,000
.35	1,500	\$150,000	\$95,000	+\$55,000

*Assumes a market price of \$75 per barrel.

How should our wildcatter assess the prospects of the well? The standard first step is to calculate the *expected value* of its profit. In general, an *expected value* of a random variable (often called a *mean* or an *average*) is the sum of all possible outcomes, each multiplied by its probability. For our wildcatter this is

$$E = p \cdot x_W + (1 - p) \cdot x_L = \$3{,}000$$

[3] Lotteries are risky prospects. Formally, there is a difference between a *risky* prospect and an *uncertain* prospect. When both the odds of success (or failure) and the payoff of each outcome can be accurately estimated, the prospect can be rationally evaluated and is called "risky." If neither the odds nor the outcomes can be calculated with any precision, rational analysis is impossible and the prospect is called "uncertain." This fine distinction is often ignored, as it should be.

where p is the probability of a win (.35), (1-p) is the probability of a loss (.65), x_L is the value of the loss (-$25,000), and x_W is the value of a win (+$55,000). The *expected profit* is $3,000. Thus, wells like this one are, on average, profitable so it meets the profitability test.

The wildcatter turns to the next, and more difficult, question: in light of the risk involved, is the $3,000 profit sufficient to justify drilling the well? The answer depends on two things. First, what measure of risk does he use? Second, what is his psychological propensity to take risks? First, let's consider the risk measure. While there are many measures of risk in the financial and statistical literature, the most commonly used is the *standard deviation* of the outcomes, defined as the square root of the *variance* of outcomes. This is a measure of the average spread around the expected value: a low standard deviation (spread) describes low risk; a high spread is high risk.[4]

The variance of a random variable, denoted σ^2, is the expected *squared* deviation of each result from the expected value, or

$$\sigma^2 = p \cdot (x_W - E)^2 + (1-p) \cdot (x_L - E)^2$$

For our wildcatter the variance is $\$^2 1,492,895,044$, an incomprehensively large number that occurs because the variance is in the meaningless units of *squared dollars*. To put the variance on the same scale as the expected value, we use the square root of the variance, called the *standard deviation*.

$$\sigma = \sqrt{\sigma^2} = \$38,636$$

So, the wildcatter muses, the project has an expected $3,000 profit but the variability of profit around that average is a high $38,636. In a large number of projects like this he would earn between -$35,636 and +$41, 636 on 68% of the wells.[5]

Is there enough reward to compensate for the risk? That is a question of the *risk premium* the wildcatter requires to invest in the well. When added to the explicit costs of the project, this risk premium determines the c*ost of capital*—the minimum increment over expected costs required to accept the risks.

---

[4] To many this seems strange because their notion of risk is, "How much can I lose?" or "What is the chance that I'll lose, and by how much?" But the standard deviation measures variability, not loss, and to the statistician "risk" means "variability."

[5] A statistical problem with two outcomes (win and loss), and constant probabilities of win and loss in repeated independent trials, can be assessed using a *binomial distribution*. For a large number of trials a binomially distributed random variable will experience 68 percent of the outcomes in the interval $E \pm 1$, that is, within a one standard deviation band around the mean.

Suppose, for example, our wildcatter requires a 20 percent return on his expected investment. His expected investment is the expected costs, $82,000. The minimum profit he must expect to get with drilling is, therefore, $98,400. The $3,000 expected profit is woefully short, so the wildcatter passes up the opportunity.

But how did our wildcatter come up with the 20 percent cost of capital in the example? The answer is subjective—the wildcatter must consider his personal taste or distaste for risk. This will lead him to choosing a risk premium—the minimum excess of expected return over expected cost that he requires to proceed. We will not exploit this area extensively, but we now turn to the question of what subjective elements might be in the wildcatter's head.

1.4 The Risk Premium

The notion of a risk premium is tied in with the question, "Why would anyone buy lottery tickets?" Clearly, ticket buyers will, on average, lose most of their $2 so the average outcome can't he a profit. If, like our wildcatter, they had a positive risk premium—needed to expect returns in excess of costs—they wouldn't buy tickets. So do they have a negative risk premium? Or is the notion of a risk premium simply irrelevant in the context of a lottery?

Reasons have been advanced to explain lottery participation without reference to a risk premium. First, the lottery might simply be a *form of entertainment* that generates sufficient endorphins at each drawing to compensate for the expected loss; in this case we might call frequent players "gambling addicts" for whom the expected loss is simply the price of playing at the casino. Second, players might be *overconfident*, believing that, in spite of the published odds, they will be the ones to get the winning tickets and earn a positive net payoff. We all know people who have this trait—they think everyone else will lose, but they will win! Third, the player might simply be *overly competitive*: he or she enjoys beating the other side and values a win far more than he dislikes a loss because he can bask in the win, perhaps crowing about it, and forget the loss.

These are all matters for psychologists to explore. The financial and economic literature gives a fourth reason: the player's subjective response to the lottery's risk. This is different from the motives above because it is a rational calculation based on the player's enjoyment derived from financial rewards and risks. Some wonkish detail is relegated to parts A2 and A3 of the Addendum.

Lottery players know that they will lose on average; those who think they don't are seriously underestimating the full costs that they've paid for their play. But there is a rational foundation for playing a losing game: the loss on every ticket is small—$2 for a Powerball ticket—and while the chance of a win is quite small, the potential value of the win can be huge. If the player is a *risk lover*, defined as one whose happiness is elevated more by a win than it is depressed by a loss—he will buy into some losing opportunities, as long as they don't lose too much, because the rare but large win more than compensates for the frequent but small losses. The risk lover will have a negative risk premium.

A more common attitude toward risk is *risk aversion*. A risk averter is more sensitive to the losses than he is to the wins, so he will require a positive risk premium: he must, on average, make a sufficient profit to reward him for accepting risk. A risk averter will come out ahead in the long run because he requires that the odds be in his favor. He will not be seen in a casino unless he is the overconfident or competitive type, or he is paying for entertainment.

While risk lovers dominate in casinos, risk averters dominate in financial markets like common stocks and corporate bonds. The evidence for this clear—on average, the return on financial investments increases with the risk: average stock returns exceed average corporate bond returns, corporate bond returns exceed government bond returns, and government bond returns exceed the returns on bank deposits and other cash items.

We'll not consider the complex matter of why some people invest in stocks and bonds *and* play the lottery: they act as both risk lovers *and* as risk averters. This is briefly discussed in part A3 of the Addendum.

Finally, and solely for the sake of completeness, there is an intermediate state of risk assessment called *risk neutrality.* A risk neutral person simply ignores risk and bases his decisions on the expected value—his risk premium is zero; he will take on risk even if, on average, he just breaks even. If you know one, let me know the name and address.

2. Standard Lotteries

2.1 Public vs. Private Lotteries

If you're of a certain age you'll recall when "numbers rackets" were in the news. These private lotteries were condemned as both immoral and illegal. Now state lotteries abound that differ from the numbers rackets in only two ways—they are legal (a matter of politics), and the winnings are taxed. The moral objection to gambling has apparently been overridden by the needs of the public purse, as has the aversion to bilking the poor.[6] Perhaps now the chief objection to private lotteries is that they compete with the state lotteries! This competition could be devastating—the odds of winning in a private numbers game are much better than in a similarly structure State lottery, and the numbers racket has less administrative cost.

The typical numbers game is based on three random digits, giving 1-in-1,000 (.001) probability of winning; this is far greater than the chance of winning in any state lottery, and the relatively high probability of a private lottery win means frequent winners, which means active participation. The three winning numbers are selected in a variety of interesting ways: one is to use the last three numbers in the daily betting pool at a designated racetrack; another is to use the last three numbers in the daily trading volume on the New York Stock Exchange. Private lotteries eschew the expensive and visually exciting random-number generating machines used by state lotteries in televised drawings—they are low budget betting pools.

Anecdotal evidence suggests that a winning number in 3-digit numbers ticket costing $1 is in the $500-$600 range; this implies an expected win of $.50-$.60 and an expected net loss of $.40-$.50 per $1 ticket. In contrast, as we will see, players in State lotteries experience an average loss greatly exceeding 50 percent on an after-tax basis.

2.2 Roll-**Over** Lotteries

The typical lottery has a roll-***over*** format: in the event that there is no Jackpot winner, the fixed-value payouts for the lower-tier winners are paid but the Jackpot is retained and rolled over to the Jackpot at the next drawing. This continues until there is a drawing with a Jackpot winner. The Jackpot is then paid out and the next drawing has a minimum starting Jackpot that can then be built up with future roll-overs.

---

[6] State lottery players are surprisingly well educated and affluent by the standards of the numbers racket players.

Consider the following example from the now-defunct Massachusetts *Cash Winfall* lottery, which we'll discuss in detail later; the data are shown in Table 2. There are 46 numbers from which a player selects 6 on each ticket. The Jackpot is $1,600,00—a realistic number. For each drawing the fixed prizes are paid out and if there is no Jackpot winner the Jackpot is rolled over to the next drawing. The expected payoffs for a $2 ticket are shown in the table below.

Table 2
Hypothetical Roll-Over Lottery

Matches	**Prize**	**Chance**[7]	**Probability** *Single Ticket*	**Expected Value** *cents*
□□□□□□	$1,600,000	1 in 9,366,819	.0000001068	16.96
□□□□□	$4,000	1 in 39,028.41	.00002562	10.25
□□□□	$150	1in 800.58	.001249	18.74
□□□	$5	1 in 47.40	.02111	10.55
□□ (Free Bet)	$2	1 in 6.83	.1464	29.28
□	0	----	----	
All Above				**85.78**

* The number of *n*-number combinations with exactly *x* matching numbers and *n-x* non-matching numbers is $C_x^N C_{n-x}^{N-x}$ with *N*=46, *n*=6, and *x*=4. See part A1 of the Addendum on the hypergeometric distribution.

The player in Table 2 expects to get about 86 cents from his $2 ticket, for an expected net loss of $1.14 on each ticket.

Roll-overs can continue indefinitely—indeed, the probability is that they will continue for quite a while—in 2003 the Massachusetts *MegaMillions* game went for the entire year with no Jackpot winners. The Jackpot builds until a Jackpot-winning ticket arrives. Throughout this period of accumulation the expected value of a Jackpot-winning ticket increases as the Jackpot increases (remember, the chance of winning is always 1-in-9,366,819), and two things happen: first, new players enter the game, attracted by the improved Jackpot payout and bringing in even more

---

[7] To repeat the information in the Addendum, suppose that we are to select *n* numbers from a list of *N* numbers. We note that that because the total number of possible combinations is C_n^N, and only one combination of n numbers can win the Jackpot, the probability of winning is 1-in-C_n^N. But what if we want to compute the chance of getting a lower-tier prize. To be specific, suppose that *N* =46, *n*= 6 and we want to calculate the chance of winning a four-number prize (*x* = 4). How many possible tickets can win with four correct numbers out of 6?

A common error is simply use the same procedure as for the Jackpot probability and compute C_4^{46} possible combinations as the answer. But those 163,185 combinations are not just those with exactly four matching balls—they also include five- and six-ball matches, which also have four correct balls. What you seek is the number of combinations with *exactly* four correct numbers and two incorrect numbers.

This probability is described by the *hypergeometric distribution*, a variant of the geometric distribution. The number of combinations with exactly four "good" and two "bad" numbers is $C_4^{46} C_2^{46-4}$ and the probability of that occurring is $(C_4^{46} C_2^{46-4})\ /C_6^{46}$.

revenues to be distributed; and players engage in bulk purchases of ticket to increase their chance of taking the jackpot. Later we'll investigate the relationship between the Jackpot size and the number of tickets in play.

One way to assess a roll-over is to determine how large the Jackpot must be before each ticket becomes a breakeven prospect? Stated differently, what is the minimum Jackpot that will give each ticket a $2 expected value? In Table 2 the lower-tier winning tickets have an expected value of 68.82 cents, so the break-even occurs when the expected value of the Jackpot is $1.32. This happens when the Jackpot reaches $12, 375,472.

A second approach is to ask how many drawings you would expect to play before a Jackpot win. The answer (see part A1 of the Addendum regarding the geometric distribution) is given by the geometric distribution—it is the probability of *not* winning for n-1 drawings, then winning in the n^{th} drawing; this is $(1-p)^{n-1}p$. The expected number of drawings before a win is $E(n) = \frac{1}{p}$. We've seen that the probability of a Jackpot win on any single drawing is $p = \frac{1}{9{,}366{,}819}$ so the expected number of drawings before a first win is 9,366,819—*you would expect to play all of the possible combinations before a win*, and your ticket cost would be $18, 733,624 (less any accumulated lower-tier prizes). At one drawing a month you'd never live to see the win—you'd have to task your descendants with the job.

To achieve a Jackpot win during your lifetime you'd have to bulk-buy tickets for each drawing. For example, if you bought 300,000 tickets per drawing the probability of a Jackpot win at any drawing is .0318 and you'd expect to play 31.5 times before a Jackpot win.[8] Again, the expected ticket cost would be $18, 733,624 (l*ess* any accumulated lower-tier prizes), but you've got a shot at winning in your lifetime; still, you might well have to play far longer than the expected number of drawings.

2.3 Bulk-Buying in a Roll-Over Lottery

There is an old joke. Two businessmen are at a bar and one says to the other, "My company loses $5 on every widget we make." The other replies, "Gee, how do you stay in business?" The first answers, "We make it up on volume!"

No, it's not funny; but it illustrates a point. We will see later that for some lotteries bulk buying is both common and can be profitable. But are there any bulk-buying advantages for the typical rollover lottery?

The answer is: maybe, but only after a sequence of rollovers. For the example in Table 2, the probability of having a Jackpot-winning ticket if you buy 300,000 tickets is .0318. But you also multiply the ticket cost in the same proportion and

---

[8] The geometric distribution applies with p = .0318. The expected number of trials before a win is $\frac{1}{p}$.

300,000 times an expected loss is still an expected loss. In order to come out ahead you want to wait for Jackpot accumulation to make the payoff sufficiently high, *then* buy in.

2.4 Powerball Lotteries

Powerball lotteries, now run by 47 states, have become a great source of state revenue. In Powerball the ticket buyer selects five numbers from a card with 69 numbers, plus a sixth *Powerball* number. The lottery concludes with a public drawing in which five numbered "white balls" are chosen from 69 numbered white balls, followed by selection of one red "powerball" from 26 numbered red balls. The result is six-numbers that define a Jackpot winner.

Table 3
Pennsylvania Powerball Game

Ticket	Prize	Chance	Probability	Expected Payout
□□□□□+■	Jackpot	1 in 292,201,338	.000000003422	----
□□□□□	$1,000,000	1 in 11,688,054	.00000008556	$ .0856
□□□□+■	$50,000	1 in 913,129	.0000010951	$ .0548
□□□□	$100	1 in 36,525	.000027385	$ .0027
□□□+■	$100	1 in 14,494	.00006899	$ .0069
□□□	$7	1 in 579.76	.0017250	$ .0117
□□+■	$7	1 in 701.33	.001426	$ .0010
□+■	$4	1 in 91.98	.010872	$ .0435
■	$4	1 in 38.32	.0261	$ .1044
				$ .3106

□= White ball number ■= Powerball number

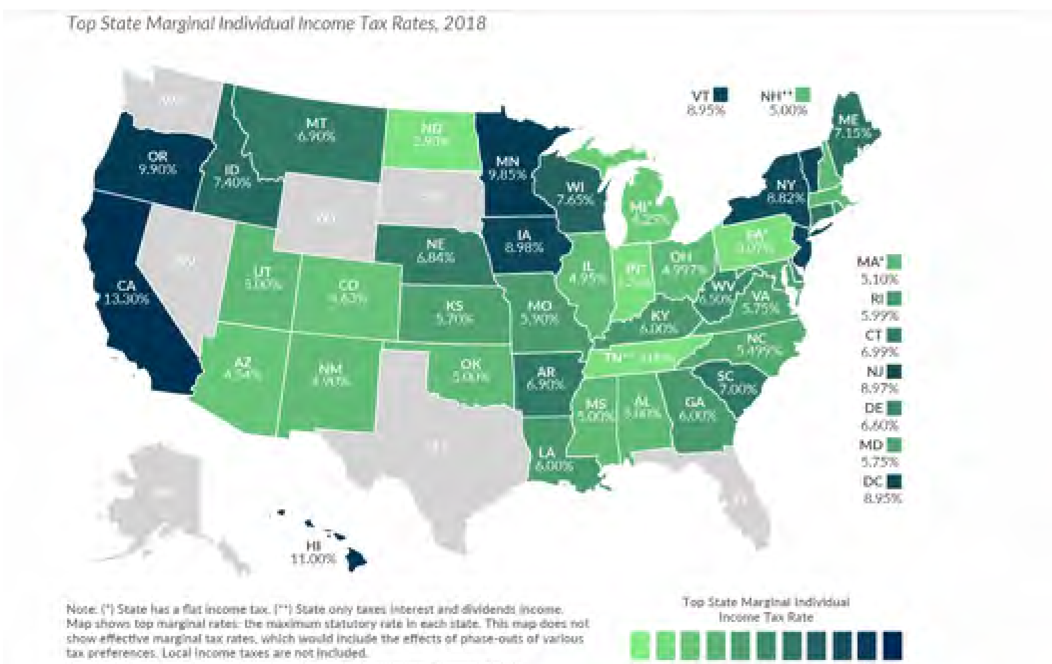
The Powerball game also has a "multiplier" option: for an extra dollar the player can buy this option and increase the payoff by a multiple of 2, 3, 4, 5, or 10. The multiplier is randomly selected at the public drawing just before the full drawing; it affects only the sub-Jackpot prizes and the 10x multiplier is not in effect if the Jackpot annuity is more than $150 million. Thus, if the randomly selected multiplier is 3, you win triple the standard payoffs for all lower-tier prizes.

The Powerball lottery has changed several times since its introduction in the 1990s. In 2012 the Powerball minimum Jackpot increased to $40 million; in 2015 the number of white balls increased from 59 to 69 and the red balls were cut from 35 to 26; this change reduced the probability of a Jackpot win from 1-in-175,293,510 to the current 1-in-292,201,338.

In order to win the Jackpot the holder must have correctly chosen the five non-Powerball (white) numbers, of which there are $C_5^{69} = \frac{69!}{(69-5)!5!} = 11{,}238{,}513$ possibilities *plus* the correct Powerball (red) option from $C_1^{26} = \frac{26!}{(26-1)!1!} = 26$ combinations. That is, to win the Jackpot he must chose the correct five numbers from 11,238,513 combinations multiplied by the 26 powerball number: a staggering 292,201,338 possibilities.

2.5 Taxes on Lottery Winnings

As noted above, State lottery winning are subject to both Federal and State income taxes. The federal government automatically collects a 25% withholding tax at the time the prize is awarded, and a winner's federal tax liability can be as high as 40.8%.[9] In addition, the state in which the prize ticket was sold will also collect an income tax unless the state has no income tax, as in Florida. State income tax rates can be as high as 13.3% for ordinary income (California) and some states have higher rates for investment income than for ordinary income. The winner of a large Jackpot can expect to pay taxes at the maximum rates for investment income, making the total tax rate on winnings around 41% in no-income-tax states and 51% in California.



Income Tax Rates By State, 2018

[9] The maximum federal ordinary income tax rate is 37%. On top of that a 3.8% rate is paid on "investment income."

Page Intentionally Left Blank

3. "Winnable" Lotteries: Roll-Downs

We've seen that in roll-***over*** lotteries the Jackpot rolls over to the next drawing if there is no Jackpot-winning ticket. The only way to improve one's prospects is to wait until the Jackpot is large enough, then buy tickets in bulk. But there are lotteries—and strategies—that can improve the odds. Perhaps the best known is a roll-***down*** format. Regrettably (for the player) these are no longer available.

<u>Roll-*Down* Lotteries</u>

A roll-down lottery treats an uncollected Jackpot not as a deferred opportunity to be transferred to the next drawing but as a current opportunity for the winning tickets at lower tiers in the current drawing. An example is the Massachusetts *Cash Winfall* lottery. As noted above, there were no jackpot winners in the Massachusetts *MegaMillions* roll-over lottery in 2003. As a result, interest in the lottery—and ticket sales—fell sharply. The Massachusetts Lottery Commission looked for a new format and found it. The Massachusetts *Cash Winfall* game, modeled on a Michigan game, had its first drawing in September of 2004.

The *Cash Winfall* game required selection of 6 numbers from a list of 46 possible numbers. The Jackpot had a starting minimum of $500,000 and rolled over until the Jackpot exceeded $2,000,000.[10] If the Jackpot exceeded $2,000,000 the drawing converted to a roll-*down* format.

The roll-over lottery in Table 2 is, in fact, a description of Mass *Cash Winfall* game with a Jackpot is less than $2 million. In that discussion we found that the expected prize from a single $2 ticket is 86 cents, so the expected *net* value of a ticket is -$1.14. Now we consider the prospects of the same lottery if there is no Jackpot winner and, the Jackpot exceeds $2 million, converting the lottery format to a roll-down.

Savvy players must predict a roll-down so that they can be prepared for a roll-down. While players will make their own estimates, the Massachusetts Lottery Commission made regular public reports to assist players in estimating the Jackpot size at the next drawing. Experience suggested that if a Mass *Cash Winfall* game's Jackpot exceeded about $1.6 million the probability of a roll-down at the next drawing was quite high. Players understood that the size of the pot at the next

---

[10] The $500,000 minimum Jackpot is the "starting new" prize for the first drawing after a winner has taken the Jackpot.

drawing was uncertain, but they began to move in when they smelled a roll-down in the next few drawings.

The Massachusetts Lottery Commission reported that at a roll-down drawing it typically held $1.15 to be paid out for each $1.00 of tickets sold. If this was paid out it was a baked-in-the-cake 15 percent profit! Of course, this 15 percent profit applied to the "average" player and individual players are subject to risk and can experience total loss. Still . . . , where else could you get an average fifteen percent for just a few days of "investment."

Bulk buying of tickets became common as savvy players entered the game en *masse* in anticipation of a roll-down. They had two motives: first, bulk buying increases the ticket revenues and pushes the lottery into a roll-down; second, bulk buying allows players to get a wide distribution of ticket numbers, thereby enhancing the prospects of a win.

Consider a specific Mass *Cash Winfall* drawing on February 8, 2010, shown below in Table 4. The Jackpot had reached a bit more than $2.4 million so the drawing converted to a roll-down. The Commission paid off the fixed prizes for the five-, four-, three-, and two-match tickets (the last received 209,000 free tickets); it also distributed the Jackpot among the lower-tier winners according to Lottery rules.

Table 4
Cash Distributions
Massachusetts *Cash Winfall* Game
February 8, 2010 Roll-Down
Holder of 200,000 Tickets

		Expected	Cash Prize Distributions		
Match	**Chance**	**Win Tickets**	**Standard**	**Roll-Down**	**Total**
□□□□□	1-in 39,028.41	5	$20,000	$ 90,480	$110,480
□□□□	1-in-858.58	250	$37,500	$164,380	$201,880
□□□	1-in-47.4	4219	$21,095	$ 92,185	$113,280
□□*		29,282	-----	-----	-----
Totals		***33,726***	***$78,595***	***$347,045***	***$425,640***

* The free ticket for □□—is a noncash prize and not included above. There were 29,283 free-ticket prizes, worth a total of $58, 566 in future drawings..

Table 4 reports the ;Massachusetts Lottery Commission's estimates of the expected cash winnings for a player who bought 200,000 tickets in the February 8, 2010 roll-down. The expected *cash* winnings for this bulk buyer would be about $425,000; expected total winning. Adding the free tickets at $2 each brings this to about $484,000.[11] After netting out the $400,000 cost of tickets, the expected net *cash* winning yielded a profit of $25,640 in cash and $84,206 in cash and noncash prizes. This means a cash profit rate of 6.4% and an overall (cash + noncash) 21.0% profit rate.

Of course, the Lottery books have to balance—an expected net win for players was an expected net loss for the Lottery Commission and for the State (less the taxes it collected on winnings). The Commission was well aware of this, but its mandate was to sell tickets and make money in the long run. It (perhaps correctly) perceived that ticket sales improved enough at a roll-down drawing that the extra revenues covered the occasional loss. For whatever reason, it kept roll-down losses outside the public's eye and treated them as a cost of doing business.

In 2005 an MIT undergraduate looking for a senior thesis topic noticed that roll-downs could create winning opportunities if a large number of tickets was bought. The enterprising student organized Random Strategies LLC, a company owned and operated by his family and friends; at its peak the company employed seven people, most full-time.[12] Over the LLC's seven-year life there were 769 *Cash Winfall* drawings, of which 44 were roll-downs worthy of its attention. The Lottery Commission estimated that Random Strategies netted $3.5 million over seven years by bulk-buying as many as 312,000 tickets for a single roll-down drawing.[13]

A $3.5 million profit is nothing to sneer at, but it assumes that ticket expenses are the only cost. The hidden labor and other capital costs associated with bulk-buys are not considered. It also neglects to distribute the winnings across several full-time employees and probable part-timers; if the annual profit per employee was known, we might be able to judge whether Random Strategies generated income in excess of the opportunity cost of full-time "gainful" employment.

The non-cash "opportunity costs" of bulk buying are probably substantial if one imputes market prices to them. Bulk buying is a time-intensive activity not done simply by making a phone call or ordering online. Under Massachusetts law at the time, lottery tickets must be hand completed—machine-generated tickets were not legal and any machine-generated ticket was invalid. That restriction was not rigidly

[11] See the July 27, 2012 letter to then-State Treasurer Steven Grossman from the State Inspector General Gregory Sullivan. The letter exonerates the lottery officials, tells the history of the Mass *Cash Winfall* program, and gives useful examples of how it benefitted savvy players. It can be downloaded from www.mass.gov/ig/publications/reports-and-recommendations/2012/lottery-cash-winfall-letter-july-2012.pdf.

[12] Random Strategies was only one of several "syndicates" operating in Massachusetts. Students at Boston University and Northeastern University formed another, and the Selbee group in Georgia formed a syndicate to buy Michigan roll-down lottery tickets. There were undoubtedly more such groups around the country.

[13] Random Strategies considered 300,000 tickets the "sweet spot" of bulk buying to achieve their two goals: get a broad distribution of numbers, and push the lottery into a roll-down.

enforced, but if hand-generated tickets were standard practice, and could be completed at a rate of 2 per minute, the purchase of 200,000 tickets would take 1,667 hours, equivalent to 0.83 man-*years* of labor.[14] The Lottery Commission estimated that machine-generated tickets could be cut at a rate of 100 per minute, requiring only 33 man-*hours*!

There are additional sources of labor costs beyond cutting the tickets. In order to improve the odds by bulk buying, a player should select a wide distribution of numbers that conforms to the statistical properties of the Lottery machine. This requires understanding the Lottery machine's workings, i.e., statistics, and it also requires that the agents sent off to buy tickets be armed with lists of ticket numbers to select, thus avoiding both duplicate tickets and the agent's internal biases. Working from numbers sheets will slow down the ticket generation process and add man-hours to the process. And even though the numbers lists can be easily produced on a computer using a random number-generator, that itself involves costs.

Finally, after the drawing the tickets from each bulk purchase must be sorted to pick out the winners and losers—winning tickets need to be claimed and losing tickets need to be held for tax purposes. This sorting might be done several times to ensure that no opportunities are missed, and losing tickets needed to be kept for tax purposes. One bulk-buying group in the Michigan *Cash Windfall* game reported that a search for winners from a single roll-down drawing might require 20 man-hours, and would have to be done more than once.

So the man-hours required to run a bulk purchase operation can be considerable, and those are not the only costs. The money needed to buy tickets in bulk involved either an opportunity cost or explicit interest and associated expenses. Storage of the massive amount of information—or scanning it into a computer—was an additional cost.

Perhaps the real payoff for Random Strategies LLC was the kick from beating the system!

During the early 2000s *Cash Winfall* games were introduced in a number of states. Not until 2011 did the *Boston Globe* bring roll-down losses to public attention. This led to an investigation into whether lottery officials had rigged the game to make money through syndicates of friends. The investigation revealed no such thing—the players had just seen a quirk in the payoffs. Even so, the Mass *Cash Winfall* game was terminated in 2012, and soon all other *Cash Winfall* games were closed down.

[14] A man-year is 2,000 man-hours, computed at an average 0f 250 eight-hour workdays.

4. Do Winning Strategies Exist?

Why do many people believe that there are strategies that make lotteries winnable? One reason is that some players have won multiple times using their personal "system." These people are poster children for hopeful players.

Consider Stefan Mandel who reportedly won fourteen lotteries; his earlier prizes were in Australia, which eventually barred him from future lotteries. His most profitable victory was a 1992 Virginia lottery that paid him at least $25,000,000. Mandel's strategy is available to all: he found a small lottery (in terms of tickets sold) with a large Jackpot, and he bought all 7.1 million possible combinations. No sleepless nights for Mr. Mandel.

Or think of Richard Lustig, whose first of seven wins won him $10,000 in 1992, whose maximum win was $842,000 in 2002, and whose last win was $99,000 in 2010. He reports that his lifetime winnings were over $1 million. Mr. Lustig does not tell us what his costs for this win of roughly $1.2 million over 28 years cost in time and money. Mr. Lustig showed remarkable self-awareness when he reported that,

> *I didn't even realize that I had a method until my fourth win*.[15]

Apparently, after his fourth win Mr. Lustig reverse-engineered himself to uncover the method of which he was unaware. Such introspection usually takes years on an analyst's couch!

These two noted players have one thing in common: their wins began in the 1990s before lotteries became a very big and complex business with standardized products and very many number combinations to select. Back then you could shop for a weak lottery format with less combinations, then bulk-buy in moderate quantities—and clinch the win, as did Mr. Mandel.

With the very large numbers of players today and the standardization of lotteries (as noted above, Powerball is played in 47 states) it should be no surprise that others have won multiple lotteries. But sometimes a winner has some unique talents.

For example, Joan Ginther, a Stanford Ph.D. in statistics, collected almost $21 million in four wins. Her primary strategy—beyond knowing statistics—was to stifle multiple winners by playing in small lotteries with relatively little competition.

---

[15] See www.huffingtonpost.com/2013/11/08

She also adopted the Sam Ervin "I'm just a country lawyer" approach to masking her identity as a statistician. Presumably to prevent copycats from following her around to her lottery venues. Only recently was she unmasked.[16]

And in 2016 Nicholas Kapoor, a statistics professor at Fairfield University, won $100,000—not yet a multiple winner, he appears to be on that track.

Do statisticians know something that most players don't? The answer is—Yes! *They know statistics*. They know how to think about matters of pure chance, they know how the Lottery machine "thinks." Games of pure chance, like lotteries, are sometimes called games against "nature." The opponent is a random number-generating machine without facial tics, ulterior motives, bluffability, or any of the other human failings that work in games like poker. You can't read your opponents in a lottery, and machines coded by statisticians pick modern lottery numbers.

I'm reminded of a powerful scene in the 1970 movie *Patton*. George C. Scott, the great actor playing General George S. Patton, has just decimated General Erwin Rommel's tanks at the Battle of El Guettar in North Africa. Patton looks over the destruction, smugly smiles, and announces,

Rommel! You glorious bastard. I read your book.

Statisticians have read the book. They don't look for meaningless patterns—they know how to separate the wheat from the chaff. Most of us can't do that. We mortals need to believe that there is a way to ferret out the truth even when the truth is that there is no truth; we seek patterns even though nature is random. So instead of mimicking the Lottery's random number-generator, we try to look into the past or into ourselves and find patterns for the future.

Let's look first at some things that don't work, though there have been winners who used them. I call them *Personal Strategies* because they come from within the player, not from within the game.

Personal Strategies

In this category are strategies based on the player's preferences, personal experiences, or personal relationships. They have no basis in statistics or in how the Lottery machine operates. I suspect that they are not really strategies to win, rather they are mnemonics that make it easier to pick numbers.

One such method is the use of *birthdays* of family and friends. That these occasionally win is no evidence of their success—with enough players using the same strategy, we'd expect to hear of winners with any strategy. The problem with the birthday method is that it reduces your chances of winning because it violates

[16] Sam Ervin was the chairman of the Watergate Commission in early 1972.

the statistical rules used by the Lottery machine. For example, consider the birthday method in a 6/46 game of choosing 6 numbers from 46. Using the months and days of birth limits you to the first 31 integers (at least, until a 32nd day of a month comes around). In effect, the player is limiting himself to a 6/31 game, for which there are 736,281 possible combinations. He has reduced the available combinations by 92 percent! But the Lottery machine hasn't—it might spit out any of the other 8,630,538 combinations.

A similar practice is *hot/cold number strategies*. Lotteries keep track of which numbers come up and report the "most frequent" numbers. An example from a popular lottery website is shown below. Eighteen numbers are reported as "most common" from the possible 69 white-ball numbers in a Powerball game. Each has roughly the same frequency—some numbers not included in the image have far lower frequencies.

Figure 2
Example of Frequent White Ball Numbers in Powerball

Source: www.lottoball.com

Some players might conclude that the Lottery machine's frequent picks are hot numbers and should be included in their selection at the next drawing; others argue that they are *formerly* hot tickets and will be cold at the next drawing. Both would be wrong—the 69 white ball numbers all have equal probabilities of selection. Pure chance determines the Lottery machine's pick, and pure chance can give runs of frequently selected numbers as well as runs of *not* frequently selected numbers.

Once again, those who use hot/cold number methods will reduce their chances to win by focusing their selection on a subset of randomly selected number. They might win, but they are not improving their chances.

Statistics-Based Strategies

These are strategies that rely on and conform to the Lottery machine's coding. First up is the *odd-even* strategy: choose a balanced mix of odd and even numbers. The second column of Table 5 presents the actual frequency distribution of odd and even numbers for the 5 white balls that the Powerball lottery machine spits out at drawings.

Clearly, extreme odd-even mixes are more rare than are balanced mixes. This alone does not mean that you should choose a balanced mix—perhaps the data from lottoball.com are aberrant, containing unusual runs in the mix, or perhaps the lottery machine isn't really a random number-generator. But the third column reports the theoretical, i.e. statistical, frequencies of each mix if the lottery machine is a true random number-generator.

Table 5[17]
Even-Odd Frequencies
Powerball White Balls
Odd-Even Frequencies in Five Powerball "White" Numbers

Odd-Even Mix	**Actual Frequency**	**Theoretical Frequency**
0 even - 5 odd	2.54%	3.13%
1 even - 4 odd	15.85%	15.62%
2 even - 3 odd	32.75%	31.25%
3 even - 2 odd	31.37%	31.25%
4 even - 1 odd	15.01%	15.62%
5 even - 0 odd	2.39%	3.13%

Source: www.lottoball.com and author's calculation.

The theoretical frequency distribution is an extremely close match to the actual distribution. This suggests that the Lottery machine *is* random, and that playing a balanced even-odd mix of numbers is a good strategy. After all, that's just the way the statistics book reads.

A similar strategy stresses *high-low number patterns*. A simple application is to assign half of the 69 white ball numbers in Powerball (say, balls 1-35) to the low-number category and the remaining 34 balls to the high-number group. Label the two groups "L" and "H" and assign them equal probabilities. Then calculate the probabilities of a mix from 0 L's and 5 H's to 5 L's and 0 H's to determine how you

---

[17] There are five "white ball" numbers to be selected. There are two possible outcomes for each ball—odd or even. The probability of *x* odd numbers and 5-x even numbers is $C_x^5(\frac{1}{2})^5$, x = 1, 2, . . .,5. For the High-Low mix we have $C_x^{69}(\frac{1}{2})^{69}$, x = 1, 2, . . .,69.

should fill out your ticket. This is precisely the same exercise as the Odd-Even exercise, and it also will tell you to balance the High and Low numbers. In general, when playing the lottery you should adopt balanced strategies because that's the way the Lottery machine will do it.

The Multiple-Winner Problem

In 2005 one Powerball drawing had 110 winners of the five-number prize. This once-in-a-millennium result had a strange reason: a fortune cookie company had always included a recommended five-number sequence to be played in any lottery, and for one run of cookies it inadvertently guessed correctly. The result was a stampede of lottery players selecting that number.

Perhaps the surest way to better payoffs is to reduce the chances that other winners will share the Jackpot. For this reason, multiple winners like Stefan Mandel, Richard Lustig, and Joan Ginther shopped for lotteries with better odds, fewer players, and smaller Jackpots. They also avoided the personal strategies like birthday-based numbers, knowing that they both reduce the probability of winning and increase the chances of sharing prizes.

Though marketed by state entities, today's lotteries are national arrangements. For example, Powerball games in Pennsylvania and Florida are not based on tickets sold in those states; rather, all Powerball players, regardless of state, are in the same national Powerball game. With 47 participating states, Powerball is a single game but with 47 state agencies managing, marketing, and monitoring activity in their states.

This decentralization of information makes it difficult to get information on multiple winners: Florida has had 12 residents win the Powerball Jackpot since 2009, and there were no multiple *Florida* winners. But to obtain information on multiple winners in Powerball drawings one would have to go through the reports for the other 46 lottery agencies.

So what can we say about the frequency of multiple winners at Powerball drawings? In the absence of data, we can estimate the probability of more than one Jackpot winner for a Powerball drawing. Because the probability of a winning ticket is the same for all tickets, one ticket's chances are the same as every other ticket's and the binomial distribution applies. Thus, the probability of a Jackpot win by x tickets out of N is,

$$P(x; N, p) = C_x^N p^x (1-p)^{N-x}$$

Suppose that there are N tickets sold for a Powerball drawing. The probability of a single ticket winning the Jackpot is $p = \frac{1}{292{,}201{,}338}$. Table 6 summarizes probabilities of multiple winners for 1, 10, and 100-million tickets sold.

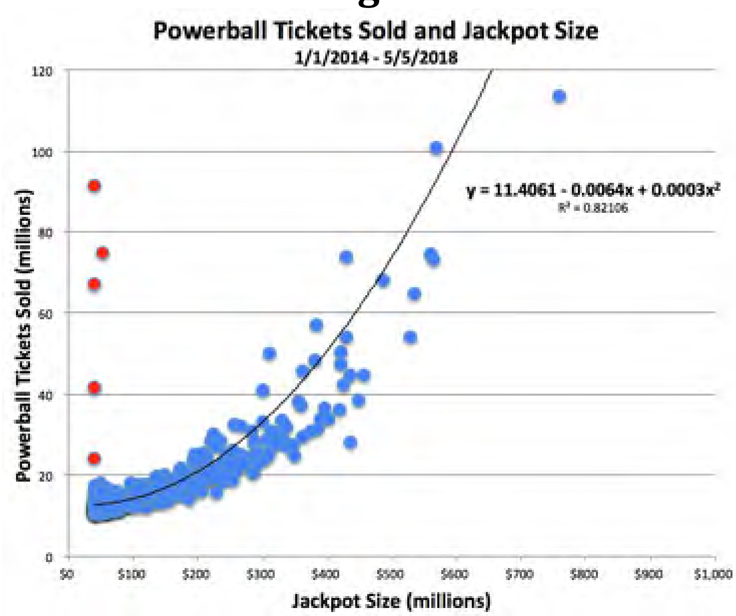
Table 6
Probability of x Powerball Jackpot Winning Tickets

# of Wins	Number of Tickets Sold		
x	*1 million*	*10 million*	*100 million*
0	99.66%	96.63%	71.02%
1	.34%	3.42%	34.22%
2	.00%+	.06%	5.86%
3	.00%+	.00%+	.69%

Not surprisingly, the probability of zero wins is quite high, but it falls as the number of tickets increases. The probability of one winner rises to 34% at 100 million tickets. The probability of two or more winners also increases significantly to almost 6 percent. The message: to minimize sharing your prize choose lotteries with small numbers of tickets sold. This dependence between the size of a lottery in number of tickets and the probability of sharing the Jackpot is a reason that the multiple lottery winners—at least those who write books—play in smaller lotteries. Joan Ginther—the multiple-winning statistician—is reported to scout out lotteries in small western towns; Stefan Mandel found smaller lotteries to hit his Jackpots; Robert Lustig has won 7 times, but none of them were huge payoffs and in 25 years he's reaped less than $1.5 million.

Of course, the probability of sharing a prize is dependent on the size of the Jackpot because larger Jackpots induce more ticket sales. This is investigated in Figure 3 above, showing the combination of size of the Jackpot (x-axis) and number of tickets sold (y-axis) for each of the 454 drawings from January, 2014 through May 5, 2018.

Figure 3

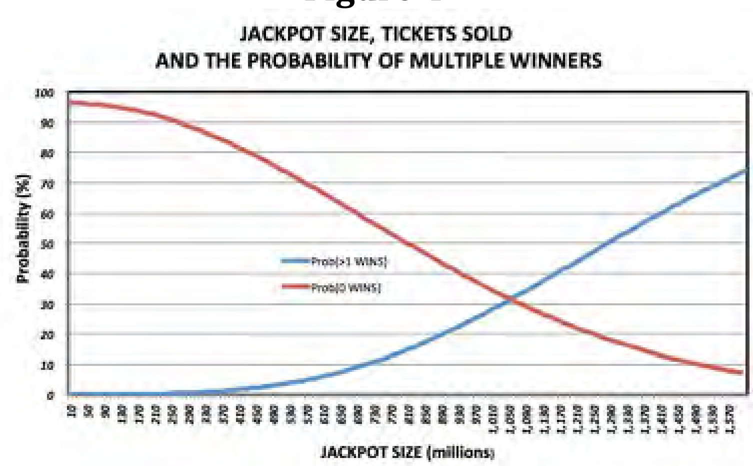


Source: Author's calculations using Excel and EasyStat.

The scatter points marked in red are outliers with very high ticket numbers at very low Jackpot sizes; these were excluded from the regression analysis. Clearly tickets sold and Jackpot size are positively *and* nonlinearly related—greater Jackpots elicit increases in the number of tickets sold, and at increasing rates.**[18]** It appears that the relationship's nonlinearity begins when the Jackpot size reaches about $250-$300 million, what might be called the lower end of "super-Jackpots."

This nonlinearity is important when estimating the expected value of a Powerball ticket. The binomial calculations in Table 6 assume no relationship between tickets sold and Jackpot prize, but Figure 3 rejects this assumption. To capture this nonlinearity we modified the probability model underlying Table 6 by having the number of tickets sold increase with Jackpot size according to the regression reported in Figure 3. The result is summarized in Figure 4, below.

Figure 4



The probability of *more than one* winning ticket (blue line) rises as the Jackpot size—hence the number of tickets sold—increases: at a $1.6 billion Jackpot (the highest experienced) there are so many tickets sold that there is a greater than 76 percent probability of more than one winner, and less than a 10 percent probability of no winners!

[18] This updates an analysis titled "According to Math, Here's When You Should Play a Powerball" by Walter Hickey in www.businessinsider.com, September 16, 2013

5. Summary

This assessment of lottery games lays out the essential statistical foundations for assessing the games and playing smarter. Of course, smarter play might not be winning play, but your chances will improve if you think like a random-number generating machine. Methods that don't do this will reduce the chances of success.

We have identified several strategies for smarter play:

- Focus on lotteries that are smaller in both number of ticket and jackpot size. Each win will be smaller but the probability of sharing is lower.

- Think like a random number-generator—do as the machine would do.

- Balance the numbers you choose so there are roughly equal odds and evens, or highs and lows.

- Wait till a large Jackpot appears before buying in, and then buy as many tickets as money and time allow.

- Avoid personal strategies—birthdays, license numbers, numbers baked into Fortune cookies, ages of family pets, etc. They will often limit your choices of numbers that the machine will select.

- Use Quik-Pics to select numbers—they are a reflection of the soul of the machine.

When you think a lottery strategy works, remember the story of the two economists walking along a sidewalk headed for lunch. One looks down and sees a $20 bill and says to his colleague, "Hey, that's $20 bill lying on the sidewalk!" The colleague walks on and the puzzled first economist rushes up to him and says, "Hey, that's a free lunch, Why pass it by?" The second economist replies, "If it was really there, it would have already been picked up."

The moral of the story—beyond the silliness of economists—is that in the real world good opportunities don't last long. Usually someone has already picked up the $20 bill. Keep the following in mind:

A Prime Rule of Lottery Play

When there is an indication of a successful lottery strategy, the opportunity will quickly disappear as both the players and the lotteries change their behavior: players will crowd into the successful strategy, creating multiple winners who milk the innovator. Lotteries will change their formats to reduce the winning odds of the strategy. After all, the Virginia lottery changed format after Stefan Mandel's coup ,and winnable lottery formats are simply eliminated (the roll-down format). By the time those on the outside know about a successful strategy, it's too late!

Still, there is one sure way to make a bit of cash in the lotteries.

Don't buy the books.

Addendum

A1. Lottery-Relevant Probability Distributions

The following definitions are used in these notes:

P(x; a, b) is the probability of x given the parameter values a and b.
p : the constant probability of an event occurring at each drawing (e.g, win)
q : the constant probability of an event not occurring at each drawing, q = 1-p
μ : the expected value of a random variable, μ = E(x)
σ^2: the variance of the values of a random variable
σ : the standard deviation of the value of a random variable, $\sigma = \sqrt{\sigma^2}$

The probability distributions we discuss all relate to dichotomous events that can be cast as "good" or "bad," "right" or "wrong, "success" or "failure;" each outcome can be described as a "0" or a "1." The purpose of the distributions are all to evaluate the number of "successes" that occur in N independent trials. For example, if John randomly asks *N* women for dates, he can use these distributions to assess the probability that *x* of the requests receive a positive response, i.e. are "1"s.

<u>The Geometric Distribution</u>

The geometric distribution gives the probability that in a sequence of *N* independent trials, each with the same probability of success, there will be *x* successes and *N* – *x* failures. You are indifferent to the exact order of wins and losses. The probability is

Geometric Distribution for Lottery Analysis

$$P(x; N, p) = (1 - p)^{N-x}\, p^x \quad x = 1, 2, \ldots, N$$

for which $\mu = \frac{1}{p}$

$$\sigma^2 = \frac{1-p}{p^2}$$

The geometric distribution can also be used for "waiting time" problems. In this case you *do* care about the order. For example, the probability of having the first success occur on the *x*th trial is

$$P(x; p) = (1 - p)^{x-1} p$$

for which $\mu = \frac{1}{p}$ and $\sigma^2 = \frac{1-p}{p^2}$

If a fair coin is flipped repeatedly, how many times will one face come up before the other face shows? With p = ½, the expected number of flips is 2. The variance is also 2.

The Binomial Distribution

Like the geometric distribution, from which it derives, the binomial distribution addresses situations in which a random event can have two outcomes: 1 or 0, win or lose, take a nap or no nap. There are N independent trials, in each of which the probability of a 1 ("win") is the same. Noting that there are C_x^N orders in which you can have x wins, the probability of x "wins" in N trials is

Binomial Distribution for Lottery Analysis

$$P(x;\, N,\, p) = C_x^N (1 - p)^{N-x}\, p^x$$

for which $\mu = Np$ *and* $\sigma^2 = Np(1\text{-}p)$

Note that tis is simply C_x^N times the geometric distribution probability of x wins in N trials; that reflects the number of orders in which N trials can yield x successes and N-x failures.

The Hypergeometric Distribution

Suppose, again, that you have two possible outcomes that can be labeled as 1 or 0 with 1 = success and 0 = failure. Suppose there is a box holding N items of which K are "bad" and N-K are "good." Randomly draw a sample of n balls from the box without replacement . What is the probability that *exactly* x of them are successes and n - k are failures?

The hypergeometric distribution starts with a population of N items in which N - X are "good" and X are "bad." A random sample of size n is taken from the population of N items, without replacement. What is the probability that in that sample x are "bad and the other n - x items are "good?" The answer is given by

The Generalized Hypergeometric Distribution

$$P(x;\, N, n, K) = \frac{C_x^X C_{n-x}^{N-X}}{C_n^N}, \quad x = 1, 2, 3, \ldots, n\text{-}1, n$$

for which $\mu = n(\frac{X}{N})$ where $(\frac{X}{N})$ = p is the probability of bad results

and $\sigma^2 = = \frac{N-n}{N-1}[np(1\text{-}p)]$

For example, a box contains N = 100 marbles; X = 30 of them are red ("bad") and (N – X) = 70 are green ("good"). A sample of size n = 12 is taken from the population of 100. What is the probability that exactly k = 2 are "bad".

The answer is $\frac{C_2^{30} C_{10}^{70}}{C_{12}^{100}}$ = 0.1637, a 16.73% probability of having 2 red ("bad") balls in the sample of 12 balls. The expected number of "bads" in the trial is 3, the variance is 2.24 and the standard deviation is 1.5 bad balls.

The description above of the hypergeometric distribution does not quite fit our lottery selection problem because we don't know the number of "bad" balls in the population. Fortunately—as so often happens in the wonderful world of mathematics—there is a symmetry in the numbers that allows a transformation of the problem that eliminates explicit reference to X.[19] The transformed statement of the probability of x "bad" balls in a sample of n is

Hypergeometric Distribution for Lottery Analysis

$$P(x; N, n) = \frac{C_x^n C_{n-x}^{N-n}}{C_n^N} \qquad x = 1, 2, 3, \ldots, n\text{-}1, n$$

where μ = and σ^2 =

Consider a 6/46 lottery. We want to calculate the probability that our 6 selected numbers out of 46 have exactly 4 numbers that are "good," that is, among the 6 numbers selected by the lottery machine. By "exact" we mean that four numbers of our 6 match *and* two numbers don't match.

Using the description of hypergeometric probability just given, we calculate the probability that *exactly* 4 items in a 6-item selection will match the lottery –selected six numbers

$$P(4; 46, 6) = \frac{C_4^6 C_{6-4}^{46-6}}{C_6^{46}} = \frac{C_4^6 C_2^{40}}{C_6^{46}} = .0000125 \text{ or } .125\%$$

The Poisson Distribution

The Poisson distribution is another probability distribution in the dichotomous (zero/one) class. One of its many uses has been in investigating reports of "hot zones" for cancers, as

---

[19] The road to this transformation begins with the easily proved observation that $C_b^a = C_{a-b}^a$. Note also that $\frac{C_x^X C_{n-x}^{N-X}}{C_n^N} = \frac{C_x^n C_{X-x}^{N-n}}{C_X^N}$.

when a community experiences a high rate of occurrence of an event. The question is, "Is this rate of occurrence unusually high?"

The Poisson distribution has one parameter u, the *average* rate of occurrence of an event. It is used to predict the number of random events

The Poisson Distribution

$$P(x; u) = \frac{e^{-u}u^{x}}{x!} \quad x = 1, 2, 3, \ldots$$

where $\mu = u$ and $\sigma^2 = u$

Note that the mean and variance are both u, and the standard deviation is $\sqrt{\mu}$.

An example: the Marine Patrol in an area investigates 10 boating accidents in a normal month. Last month there were 15 accidents and the question is, "Has something happened to increase the chances of accidents, or is thus just a blip due to chance? With $u = 10$ the probability of exactly 15 accidents is 3.47% and the probability of 15 *or more* accidents is 8.36%. That is not outside of the experience from pure chance.[20]

A lottery-related use of the Poisson is in predicting the number of multiple winners in a similar lottery. From past data in similar lotteries the average number of multiple winners at drawings can be computed; this becomes the u for the Poisson.

If, say, one drawing in every 10 has 2 or more winners, the per-drawing average is 0.1. The Poisson distribution tells us that if $u = .1$ the probability of two winners is tells us that the probability of 2 winners is 4.52%, and the probability of 3 winners is 0.02%. Four or more winners are extremely rare.

[20] The common standard of "unusual" in statistics is when the probability of a bad event or worse is less that 5%. In this case the probability of 15 or more accidents if 10 is the normal accident rate is over 8%. This is within the industry-standard "bounds of chance."

A2. The Economics of Risk-Bearing

A tenet of the economic of risk-taking is that people don't make choices to maximize the expected vale of their wealth. Rather, they attempt to maximize the expected satisfaction they derive from wealth. Satisfaction, called *utility* because of the utilitarian philosophy that once dominated the field, is a function of wealth that has the property that it increases with the level of one's wealth. It is represented by the function U(W), which just says that "utility" (U), depends on wealth (W). There is only one restriction on the shape of this relationship: the *marginal utility of wealth* is positive, that is $U'(W) > 0$. In summary,

$$\text{(A2.1)} \qquad U = U(W) \text{ where } U' > 0$$

This is, of course, a *hypothetical construct*: nobody knows the degree of their satisfaction— the hypothetical unit is called a "util," but there really is no unit of measurement. You can't one person's satisfaction with another's.

Still, (A2.1) is summarizes what is commonly believed: that people like wealth and can never have too much.

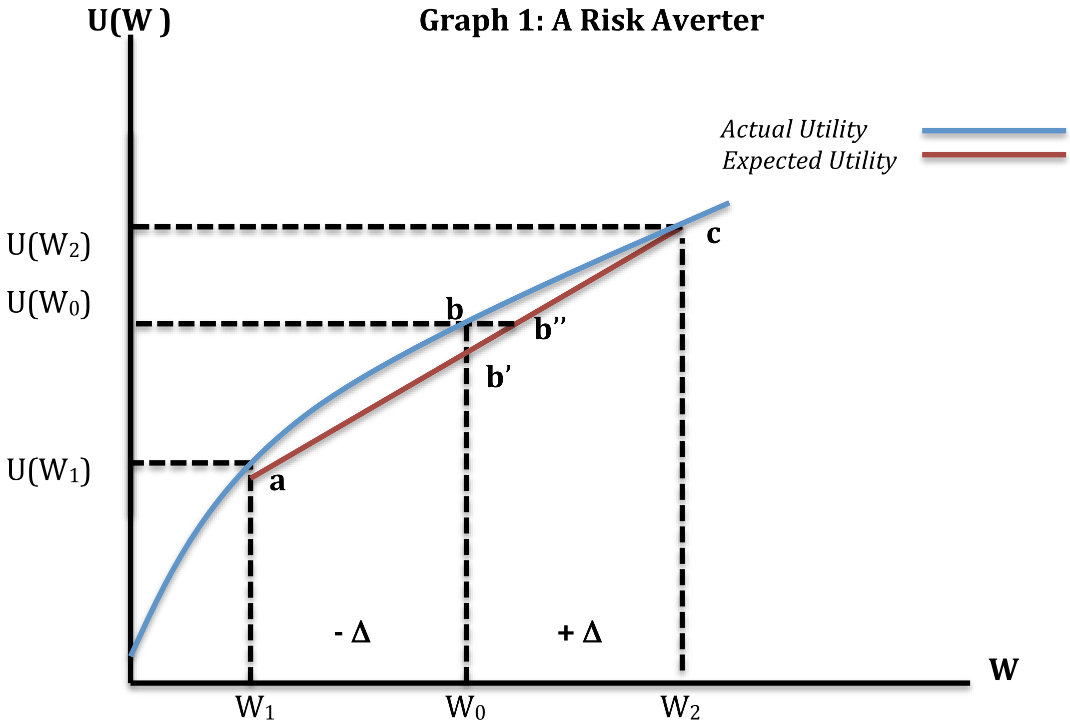
<u>Risk Aversion</u>

Suppose also that we adopt another piece of information: utility is, as before, an increasing function of wealth ($U'>0$), but the wealthier you are, the less additional satisfaction you get from a further increase in wealth, i.e., the rate of increase in satisfaction (marginal utility) declines as wealth increases. Thus is a statement about $U''(W)$—the second derivative. It says that while the first derivative U' is always positive, the second derivative is always negative. Together these pieces of information say that I can never be too rich (I always like more) but my increased satisfaction as wealth rises wealth diminishes as I get more affluent. This reflects is a common household belief—as people accumulate wealth, the kick they get from an additional dollar diminishes.

Diminishing marginal utility gives a shape to the relationship between utility and wealth. I will assess whether the increase in utility from a win exceeds the decrease in utility from a loss. With diminishing marginal utility I am "loss averse," making me a "risk averter."

The graph below exhibits a risk averter's utility-wealth relationship. Suppose he or she has W_0 of wealth in a riskless asset called "cash;" this gives him utility of $U(W_0)$. Now comes a land developer offering an investment opportunity that will either increase his wealth by $+\Delta$, to W_2 and his utility to $U(W_2)$ or will reduce his wealth by the same amount, $-\Delta$, to W_1, giving him utility of $U(W_1)$.

Graph 1: A Risk Averter



The concave blue curve shows actual utility and the straight rust line shows *expected* utility, the utility expects to enjoy if he buys into the risky proposal. Expected utility and expected wealth are defined as:

$$E(U) = pU(W_2) + (1-p)U(W_1)$$

$$E(W) = pW_2 + (1-p)W_1$$

with p being the probability of a win and $(1-p)$ the probability of a loss. If $p = 0$ the player will lose for sure, putting him at wealth W_1 and utility $U(W_1)$; this is at the southwest end of the expected utility line (point **a**). If, on the other hand, $p = 1$ so he is certain of winning, he will be at point **c**, with W_2 wealth and utility $U(W_2)$. If $p = ½$ so the odds are even, the expected wealth will be W_0—the project is a break even opportunity—and he will be at point **b** for actual utility with cash, or **b'** for expected utility with risk. Note that at **b** his expected wealth is W_0, the initial certain wealth, so he experiences no expected wealth change, his expected utility is (**b**-**b'**) less than the utility he'd get with the certain cash position.

The important point is that anywhere between **a** and **c** the risk averter will experience an expected utility less than the actual utility he would have at that level of wealth if he'd avoided the risky opportunity. For example, if $p = ½$ the development opportunity is a wash because the expected wealth is W_0, the wealth he had in cash: he will have exchanged a certain wealth of W_0 for a risky opportunity that doesn't improve his expected wealth.

There are two equivalent ways to describe our risk averter. First, he will need a minimum increase in expected wealth in order to exchange a riskless cash position for a risky asset. That minimum increase—the *risk premium*—is **b – b'** in the above graph; it will be greater the faster marginal utility of wealth diminishes as wealth increases, i.e. the more risk averse he is.

Second, and equivalent, the risk averter requires the odds to be in his favor by at least enough to provide the risk premium. He will accept no outcome to the southwest of **b"** on the rust-colored line.

This analysis leads to several propositions:

- A risk averter will never take a fair bet—a bet with an expected value equal to it's cost—because fair bets expose him to risk without a reward in the form of an expected increase in wealth.

- A risk averter must expect some improvement in his situation to entice his to buy a risky asset. That improvement can be in he form of an expected increase in wealth that exceeds his risk premium, or in a probability of a win exceeding that which makes the risky asset a fair bet. This increase in probability is exactly that required to meet the player's risk premium.

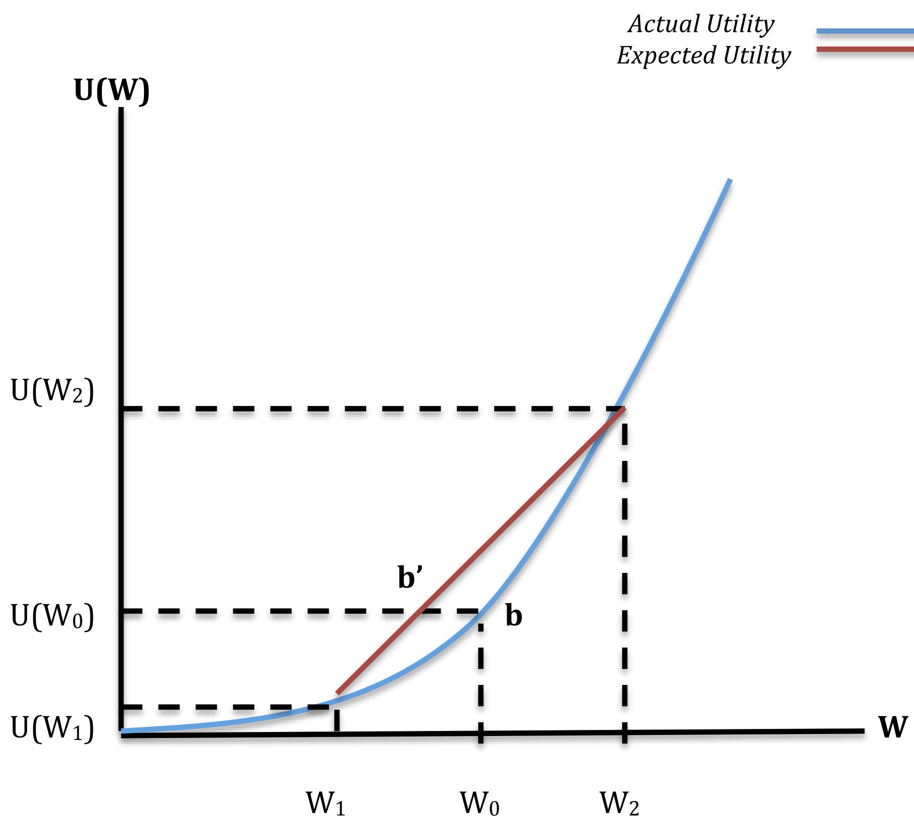
- A risk averter will not play a lottery for financial gains because lotteries offer too low a probability of winning; in fact, lottery tickets are over-whelming losers on average. He might, however, play a lottery or non-financial reasons like entertainment (enjoying the game) or over-confidence (overestimating his odds).

- A risk averter will confine his "gambling" to traded securities like stocks and bonds because they earn a risk premium—the average return exceeds the returns on safe assets like cash or government bonds. The reason for that is that the other players are also risk averse, so a risk premium must be earned for all. Economists would say that the marginal player must be risk averse.

Risk Loving

Having slogged through risk aversion, risk loving should be easy going. The only behavioral difference is that he has *increasing marginal utility*, predisposing him to taking risks because the satisfaction of a win exceeds the disappointment of a loss.

Graph 2: A Risk Lover



Increasing marginal utility means that the risk lover will take an unfair bet—one that has an expected loss. He has a *negative* risk premium—he is willing to exchange a safe position for a risky position even with an expected loss, so long as the expected loss is not too much.

The chart above shows a risk lover's utility function with his actual position at point **b** with wealth W_0 and utility $U(W_0)$. Notice that at **b'** his expected utility matches his actual at **b**—he would take the risky prospect so long as he expects to *lose* no more than is **b' – b**; his risk premium is **b' – b**, which is negative. Stated equivalently, he will take a bet even if the probability of winning is lower than that giving a fair bet. You can see him at the casino *and* in the stock market (where he would be called the "inframarginal investor," that is, an investor who is among the most eager to enter a market dominated by risk averting "marginal investors.")

A3. Prospect Theory

There are library bookshelves groaning with the weight of books on how investors assess and respond to risk. But arguably the best work—certainly the most influential—is from two psychologists who earned the Nobel Prize in Economics.[21] Amos Tversky and Daniel Kahneman spent their careers as a team investigating how people respond to gains and losses in a variety of decisions, both financial and non-financial. Their work has become the foundation of a blend of psychology and economics called *Behavioral Economics*. Among the concepts they developed is *Prospect Theory*, an approach to how humans really assess risks. A well known component of Prospect theory is familiar to us all—*Loss Aversion*: people are far more sensitive to losses, even small losses, than they are to gains: we are risk averse, at least in matters of large losses and gains, though there may be some small ranges that elicit risk loving.

The mainstream economic approach to risk-taking is rationalist, stressing the logical consistency of choices. But psychological experiments have shown that people are not logically consistent and rational. A simple example suffices: inform a group of people that a new medical treatment will improve the health of 98 percent of the sufferers; inform another group that 2 percent of takers will not be helped. You will find that the first group is more favorably inclined toward the treatment than the second. Yet both statements are identical!

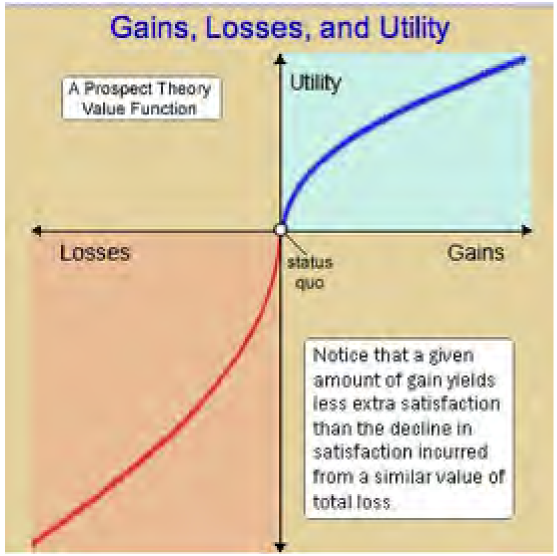
Another example is that some companies tell a new employee that they are automatically enrolled in a pension plan unless they *opt out*. Other companies tell new employees that they are not enrolled in a pension plan unless they *opt in*. It turns out that considerably more permanent pension plan enrollments with the automatic opt-out plan than with the automatic opt-in approach. even though those are identical options—in each case you can choose to enroll or not; all that's different is the way you express an option.

Psychologists Amos Tversky and Daniel Kahneman spent their careers as a team uncovering and investigating similar inconsistencies. The first example just given they called *Framing Bias*: the answer to a question depends on how the question is framed—people will tend to ally with the most positive frame. The second example is the *Default Bias*: people stay with the approach that is automatic—they tend to not express their preferences.

As part of their research they also created experiments involving choices about risky opportunities. They summarized their results in what they called *Prospect Theory*. According to Prospect Theory, the representative individual had a utility-of-wealth function like that below.

---

[21] Tversky died before the Nobel Prize decision. Under Nobel Economics Prize rules he was not eligible for the prize. But that award clearly is not Kahneman's alone.



Typical "Value Function"

This image shows several notable features:

- *Risky prospects are evaluated not by levels of wealth but by changes in wealth*. People choose zero (no change) as their reference point and think in terms of gains and losses. This is consistent with the common association of "risk" with loss rather than with variability.

- *People are Loss Averse*. A small loss from the reference point is resisted more than a small win is sought.

- *In the region of losses, people act as risk lovers*. Once they know they have a loss position, they take bets that will increase their expected loss. Thus, if they enter the loss region at a casino game, they will continue to play in the hope of a win to recoup the loss. They are like the classic embezzler who doubles up on bets to regain his losses and repay the embezzled funds before the embezzlement is discovered.

- *In the region of gains people act as risk averters*, requiring a positive risk premium to take on a prospect with a variable return. Thus, they will invest in financial securities because of their expected gain (risk premium).