Powerball, Expected Value, and the Law of (very) Large Numbers

Christopher J. Mecklin
Murray State University

Robert G. Donnelly
Murray State University

Journal of Statistics Education Volume 13, Number 2 (2005),
ww2.amstat.org/publications/jse/v13n2/mecklin.html

Copyright © 2005 by Christopher J. Mecklin and Robert G. Donnelly, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Coefficient of variation; Expectation; Lottery; Probability.

Abstract

In this paper, we consider some combinatorial and statistical aspects of the popular “Powerball” lottery game. It is not difficult for students in an introductory statistics course to compute the probabilities of winning various prizes, including the “jackpot” in the Powerball game. Assuming a unique jackpot winner, it is not difficult to find the expected value and the variance of the probability distribution for the dollar prize amount. In certain circumstances, the expected value is positive, which might suggest that it would be desirable to buy Powerball tickets. However, due to the extremely high coefficient of variation in this problem, we use the law of large numbers to show that we would need to buy an untenable number of tickets to be reasonably confident of making a profit. We also consider the impact of sharing the jackpot with other winners.

1. Introduction

Our motivation for writing this paper was the popular Powerball Lottery game, which is played in 23 states including our home state of Kentucky. For $1, a player chooses five different numbers between 1 and 53 (inclusive). In addition, the player chooses a powerball number between 1 and 42 (inclusive). The player wins the jackpot if his or her five numbers match the five numbers between 1 and 53 chosen without replacement in a random drawing and if the player’s powerball number also matches the powerball number between 1 and 42, also drawn randomly. The jackpot is a varying large amount of money which will add its value from one drawing to the next if there is no winner. There are also other ways to win smaller prizes, as will be indicated later. The jackpot prize is typically several million dollars. According to the Multi-State Lottery Association (2003), Andrew Whittaker of West Virginia
won the largest undivided lottery prize in U.S. history on Christmas Day, 2002. He won a jackpot of $314.9 million; by choosing the “cash” option, he actually won $170.5 million before taxes.

There exist a large number of books on lotteries. The vast majority of these books, such as Howard (1997), are so full of inane mathematical and statistical errors as to be not worth the paper they are printed on. A rare exception is Henze and Riedwyl (1998). This book is one of the few mathematically accurate and honest books that exists on lotteries.

An r/s lottery consists of choosing r distinct numbers from S = {1, 2, ..., s}. To win the jackpot, one must match all r numbers drawn without replacement from S. The probability of winning the jackpot is $1/\binom{s}{r}$. Many lotteries in various U.S. states and European countries are r/s lotteries; it is very common for r = 6 and s to be an integer between 40 and 60. For example, in the state of Kentucky, the Lotto South game is a 6/49 lottery. The number of possible tickets is $\binom{49}{6} = 13,983,816$.

A phenomenon of lottery games is that the number of tickets purchased increases greatly as the size of the jackpot increases. If several drawings go by without a jackpot winner, the jackpot goes to many millions of dollars, which encourages regular lottery players to buy more tickets than usual and infrequent players to participate. This is a desirable situation for lottery commissions, who keep a large percentage of the money raised in ticket sales to fund various governmental programs.

However, with millions of players per drawing, most r/s lotteries do not go very long without a jackpot winner. In order to build up a larger jackpot, which encourages more players, there needs to be even more possible combinations, making the game harder to win. Now consider a $r_1/s_1 + r_2/s_2$ lottery, which involves choosing $r_1$ distinct numbers from $S_1 = \{1, 2, ..., s_1\}$ and $r_2$ distinct numbers from $S_2 = \{1, 2, ..., s_2\}$. Typically, $r_2 = 1$ and that number is referred to as a supernumber or a powerball. The popular Powerball lottery game is a 5/53 + 1/42 lottery. Mega Millions is a similar game played in 11 states; it is a 5/52 + 1/52 lottery. The calculations throughout this paper will focus on Powerball, but could be replicated for Mega Millions.

### 2. Powerball Odds

In addition to the multi-million dollar prize won by matching all 5 regular numbers and the powerball, there are several other scenarios where one can win a lesser prize. For instance, matching the first five numbers but failing to match the powerball still wins a prize of $100,000. There are a total of 9 winning scenarios; these scenarios and their associated winnings and probabilities are given in the table below, where $n$ is the number of the first five numbers that are matched. Also in the table is the $10^{th}$ and most common scenario, in which the $1$ is lost.

| Scenario | n | Powerball? | Prize | Probability |
|----------|---|-----------|------|-------------|

Table 1. Probabilities and Prizes of Powerball Scenarios.
We will use combinatorial reasoning to verify the probabilities above; this is an exercise accessible to a student in a first course in statistics and/or probability. It is easy enough to count the total number of possibilities for the random drawing. There are five numbers chosen without replacement from the set \( S_1 = \{1, 2, ..., 52, 53\} \) and one number chosen from the set \( S_2 = \{1, 2, ..., 41, 42\} \). This gives us

\[
{53 \choose 3} \cdot 42 = 120,526,770
\]

possibilities.

Now consider the number of possible ways a player could get exactly \( n \) of the first numbers right (where \( 0 \leq n \leq 5 \)) and also get the powerball correct. There are five correct numbers, of which the player gets \( n \), and there are 48 incorrect numbers, of which the player gets 5 - \( n \). There is only one way to get the powerball number right. This gives

\[
{5 \choose n} \cdot {48 \choose 5 - n} \cdot 1
\]

possibilities.

Finally, consider the number of possible ways a player could get exactly \( n \) of the first five numbers right (where \( 0 \leq n \leq 5 \)) but get the powerball number wrong. There are five correct numbers, of which the player gets \( n \), and there are 48 incorrect numbers, of which the player gets 5 - \( n \). There are 41 ways to get the powerball number wrong. This gives

|   |   |   |   |   |
|---|---|---|---|---|
| #1 | 5 | Y | Jackpot | 1:120,526,770 |
| #2 | 5 | N | $100,000 | 41:120,526,770 |
| #3 | 4 | Y | $5,000 | 240:120,526,770 |
| #4 | 4 | N | $100 | 9,840:120,526,770 |
| #5 | 3 | Y | $100 | 11,280:120,526,770 |
| #6 | 3 | N | $7 | 462,480:120,526,770 |
| #7 | 2 | Y | $7 | 172,960:120,526,770 |
| #8 | 1 | Y | $4 | 972,900:120,526,770 |
| #9 | 0 | Y | $3 | 1,712,304:120,526,770 |
| #10 | 0, 1, 2 | N | None | 117,184,724:120,526,770 |
possibilities.

3. Expected Value and Variance

It is not difficult to find the expected value and the variance of a discrete probability distribution. If \( X \) is a discrete random variable and \( f(x) \) is the value of its probability distribution at \( x \), then the expected value is

\[
\mu_X = E(X) = \sum_x x f(x)
\]

the variance is

\[
\sigma_X^2 = \sum_x (x - E(X))^2 f(x)
\]

and the standard deviation is \( \sigma_X = \sqrt{\sigma_X^2} \).

Suppose that \( X \) represents the number of dollars we win (or lose) when we purchase one Powerball ticket. If our ticket wins no prize, then \( X = -1 \) (this is scenario #10); if we are in scenario #9, then \( X = 3 - 1 = 2 \) and so on through scenario #2. One difficulty is that the value of \( X \) when scenario #1 (i.e. winning the jackpot) occurs is not known. Let \( j \) represent the value of the jackpot. For now, we will ignore the possibility that there could be \( i > 1 \) winners of the jackpot, where our winnings would be \( \$j/i \). We are interested in finding when \( E(X) > 0 \), since a player might have the notion that it is beneficial to buy lottery tickets when the expectation is positive. We will show later that, from a probabilistic viewpoint, this notion is naive.

Let us compute \( E(X) \) and \( Var(X) \), ignoring realities like federal and state taxes and the option of taking the jackpot in one cash payment, rather than an annuity paid over several years. Let \( n \) be the number of the first five numbers matched, where \( 0 \leq n \leq 5 \). Referring to Table 1, we summed the probabilities for the winning scenarios and determined that the chance of winning some prize was 3,342,046 out of 120,526,770. This is only about 2.8%. Therefore, the other 117,184,724 ticket combinations (about 97.2%) will result in the loss of one dollar.

The expected value is

\[
E(X) = \frac{j - 99,638,177}{120,526,770}
\]

For \( E(X) > 0 \), the jackpot needs to be \( j > 99,638,177 \). If the jackpot is exactly \( j = 99,638,177 \) and there is \( i = 1 \) winner, the standard deviation is \( \sigma_X = 9075.97 \).
So, at first glance, we may think that it is beneficial to play Powerball whenever \( j \geq 99,638,177 \). Unfortunately, the following issues all work to reduce the value of the jackpot:

- Federal and state taxes
- The fact that the jackpot winning is much less than the advertised value if one chooses to take the prize in a single lump-sum cash payment rather than a multi-year annuity
- The chance of multiple jackpot winners

Let us also assume that the winner will choose the one time cash payment if he/she hits the jackpot. This will reduce the prize considerably from the advertised jackpot value. As an example, recall the case of Andrew Whittaker of West Virginia. According to the Multi-State Lottery Association (2003), he was the sole winner of the $314.9 million Powerball drawing on 12/25/2002. Mr. Whittaker chose the lump-sum payout, which was $170.5 million. After taxes, he was left with $113.4 million. The lump-sum payout is \( \frac{170.5}{314.9} \) or about 54% of the advertised jackpot. After taxes, Whittaker kept \( \frac{113.4}{170.5} \) or about 66.5% of the lump-sum payout. So Whittaker only cleared about \( \frac{113.4}{314.9} \) or about 36% of the advertised jackpot.

Let us assume that these percentages are typical for a large Powerball jackpot. If the advertised jackpot is \( j \) dollars, a single jackpot winner will only win \( 0.36j \) dollars. Taxes will also need to be paid if one wins a Scenario #2 or #3 prize. We will assume that we will keep 70% of our winnings after taxes in those situations. In Scenarios #4-9, let us assume that the small prizes are countered by a sufficient number of losing tickets such that the player has a yearly loss and thus does not need to pay taxes on the winnings. The implication is that the value of \( j \) such that \( E(X) > 0 \) will increase.

| Scenario | n | Powerball? | x   | f(x)               |
|----------|---|------------|-----|-------------------|
| #1       | 5 | Y          | $0.36j | 1:120,526,770     |
| #2       | 5 | N          | $0.7(99,999) | 41:120,526,770 |
| #3       | 4 | Y          | $0.7(4,999) | 240:120,526,770 |
| #4       | 4 | N          | $99   | 9,840:120,526,770 |
| #5       | 3 | Y          | $99   | 11,280:120,526,770 |
| #6       | 3 | N          | $6    | 462,480:120,526,770 |
| #7       | 2 | Y          | $6    | 172,960:120,526,770 |
| #8       | 1 | Y          | $3    | 972,900:120,526,770 |
| #9       | 0 | Y          | $2    | 1,712,304:120,526,770 |
| #10      | 0, 1, 2 | N       | -1   | 117,184,724:120,526,770 |
The expected value is

\[ E(X) = \frac{0.36j - 101,228,092.7}{120,526,770} \]

For \( E(X) > 0 \), the jackpot needs to be \( j > 281,189,146.40 \). If the jackpot is exactly \( j = 281,189,146.40 \) and there is \( i = 1 \) winner, the standard deviation is \( \sigma_X = 9,220.69 \).

If the advertised jackpot is \( j = 315 \) million (about what it was for Whittaker’s win) and we assume only \( i = 1 \) jackpot winner, then the expected value per dollar ticket is

\[ E(X) \approx 0.101 \]

with standard deviation \( \sigma_X = 10,329.39 \).

So it appears that it might be advisable to purchase lottery tickets in the relatively rare situations when the advertised jackpot exceeds approximately \( 281.2 \) million due to the positive expectation. However, we will show that this is not the case.

Let us compare large jackpot Powerball to the common and simple casino game Roulette. In American casinos, the Roulette wheel has 38 slots: 36 of which are numbered 1, 2, ..., 36 and are colored either red or black, and 2 of which are numbered 0, 00 and are colored green. A common bet is to select one integer from \( W = 1, 2, ..., 36 \). If your choice comes up on the next spin of the wheel, then you are paid out at odds of 35:1; otherwise, you lose.

It is simple to compute the expected value and standard deviation of this game. Let \( Y \) be the amount won or lost on a $1 bet.

| Result of Spin | \( Y \) | \( f(y) \) |
|----------------|--------|---------|
| Win            | $35    | 1/38    |
| Lose           | -$1    | 37/38   |

The expected value is

\[ E(Y) = -\$ \frac{2}{38} \approx -\$0.053 \]

per spin, with \( \sigma_Y = 5.763 \). The expectation is only negative from the players’ standpoint; it is positive
$0.053 per spin from the casino’s perspective. Notice that this expectation is approximately one-half the expected value of Powerball with a $315 million jackpot and that the standard deviation is much smaller (by several orders of magnitude) in Roulette than Powerball.

4. The Law of (very) Large Numbers

Most students realize that games of chance such as Roulette are profitable (even after expenses of doing business) for a casino in the long-run; the best a player can hope for is to be “lucky” in the short-term. Shouldn’t the same hold true for playing a game of chance such as Powerball if we are disciplined enough to only play when the expected value is positive? After all, the weak law of large numbers tells us that if a random sample is drawn from a probability distribution, the probability that the sample mean $\bar{X}$ will deviate from $\mu$ by less than some arbitrary quantity $\varepsilon$ approaches 1 for a sufficiently large sample size $n$.

Theorem: Weak Law of Large Numbers (Mood, Graybill, & Boes, 1974).

Let $f(x)$ be a probability density function with mean $\mu$ and finite variance $\sigma^2$. Let $\bar{X}_n$ be the sample mean of a random sample of size $n$ from $f(x)$. Choose constants $\alpha$ and $\varepsilon$ such that $\varepsilon > 0$ and $0 < \alpha < 1$. If $n$ is an integer where

$$n \geq \frac{\sigma^2}{\varepsilon^2 \alpha} \quad (1)$$

then

$$P\left(-\varepsilon < \bar{X}_n - \mu < \varepsilon\right) \geq 1 - \alpha \quad (2)$$

Typically the constant $\varepsilon$ is chosen to be very close to zero and the theorem is used to demonstrate that as $n \rightarrow \infty$, the sample mean will eventually get $\varepsilon$ - close to $\mu$. For example, suppose we have a distribution with mean $\mu = 0.5$ and $\sigma^2 = 0.25$. An example of a distribution with these values for the mean and the variance is the Bernoulli distribution. It mathematically models the flipping of a fair coin, letting $X = 1$ when we obtain heads and $X = 0$ when we obtain tails.

We will flip a fair coin $n$ times, count the number of heads obtained, and divide by $n$. Suppose we want to be at least 95% sure that the sample mean $\bar{X}_n$ will be between .499 and .501. We have $\sigma^2 = 0.25$, $\alpha = 1 - 0.95 = 0.05$, and $\varepsilon = 0.01$; therefore

$$n \geq \frac{\sigma^2}{\varepsilon^2 \alpha} = \frac{0.25}{(0.01)^2 (0.05)} = 50,000$$

To be only at least 50% sure, $\alpha = 1 - 0.5 = 0.5$ and the required sample size would be $n \geq 5000$.

Figure 1 shows the weak law of large numbers in action for a simulation of 50000 coin tosses implemented with the R statistical computing package.
While it is customary to take the constant $\mathcal{E}$ to be very close to zero, we will instead take $\mathcal{E}$ to be equal to the mean of our probability distribution; that is, $\mathcal{E} = \mu$. Substituting $\mu$ for $\mathcal{E}$ in equation (2) yields

$$
\mathcal{P}(-\mu < \overline{X}_n - \mu < \mu) > 1 - \alpha
$$

$$
\mathcal{P}(0 < \overline{X}_n < 2\mu) \geq 1 - \alpha
$$

This inequality tells us that if $n$ is sufficiently large as defined in inequality (1), then there is at least a 100(1 - $\alpha$)% chance that the sample mean $\overline{X}_n$ will be between 0 and $2\mu$. For large $n$, this essentially is the probability that $\overline{X}_n > 0$, since the probability that $\overline{X}_n > 2\mu$ is virtually zero for large $n$. In the context of a game of chance, this is the probability that our winnings are greater than our losses after $n$ trials of the game.

To determine how large $n$ needs for there to be a probability of at least 1 - $\alpha$ of having positive winnings, substitute $\mu = \mathcal{E}$ into inequality (1):
In statistics, we define the coefficient of variation, or $CV$, as the ratio of the standard deviation to the mean of a distribution. That is,

$$CV = \frac{\sigma}{\mu}$$

Substituting $CV$ into inequality (3) yields

$$n \geq \left( CV \right)^2 \frac{1}{\alpha}$$

Thus, we see that the sample size $n$ required to be confident of positive winnings is proportional to the square of the coefficient of variation.

Now reconsider random variables $X$ and $Y$, which correspond to Powerball and Roulette, respectively. We will compute the coefficients of variation for both $X$ and $Y$. The computation for Powerball will assume a jackpot of $315$ million (i.e. Whittaker’s Christmas Day win) and a single jackpot winner.

$$CV_X = \frac{10,329.39}{0.101} = 102,271.2$$

$$CV_Y = \frac{5.763}{0.053} = 108.7$$

Notice that the coefficient of variation for Powerball (even in a very large jackpot) is larger than the $CV$ for Roulette by about 3 orders of magnitude. This means, of course, that the squared coefficients of variation for the two games will differ by about 6 orders of magnitude. To put it another way, the sample size required to be confident of positive winnings as a Powerball player will be approximately 1,000,000 larger than that of a casino offering Roulette. To be more precise, if we set $\alpha = 0.05$ to be at least 95% confident that $\bar{X}_x$ is positive (for the casino), we need the following sample sizes:

$$n_x = \frac{(102271.2)^2}{0.05} \approx 2.09 \times 10^{11}$$

$$n_y = \frac{(108.7)^2}{0.05} \approx 236,314$$

Even if we relax the level of confidence for having $\bar{X}_x \geq 0$ in Powerball, huge samples are still required. For 50% confidence, we need
and for 10% confidence, we need

\[ n_X = \frac{(102271.2)^2}{0.9} \approx 1.16 \times 10^9 \]

While \( n \) is fairly large, it is not unreasonable that over time, there will be hundreds of thousands and possibly even millions of bets placed at a roulette wheel. Even taking into account the casino’s expenses, over the long run games such as Roulette will be profitable.

However, \( n_X \) is approximately 200 billion, which is ludicrously large. This indicates that to be at least 95% confident of winning money on the Powerball, even if we are disciplined enough to only play when the expected value is positive and fortunate enough to be the unique winner when we hit the jackpot, we will have to play hundreds of billions of times. This would require hundreds of billions of dollars (which we don’t have) and hundreds of billions of opportunities to play Powerball when the jackpot is high. Even if we buy hundreds or thousands or even millions of tickets when the expectation is positive, we will probably die long before we are ahead. If we lowered the confidence level to 50% or even 10%, we would still expect to need to play 10 to 20 billion times to realize a profit.

Even if we did have hundreds of billions of dollars at our disposal to play Powerball, we probably wouldn’t want to. Risking $200 billion for the chance to win a prize of even $200 million would be proportional to a person with a yearly income of $50,000 risking that entire salary on a game of chance for the opportunity to win a $50 prize.

To give a graphical sense of the typical fortunes one would have either playing Powerball or running a Roulette wheel, we have simulated each game 50000 times. We have assumed each Powerball ticket or Roulette bet is $1 and that we are playing Powerball with a $315 million jackpot with a unique winner. Figure 2 and Figure 3 show the results of the simulations for Powerball and roulette, respectively.
The unfortunate Powerball player, after 50000 plays, had lost $42,892, or about 85.8 cents per $1 ticket. Notice the graph closely resembles a straight line with slope -1 and $y$-intercept 0. The line has occasional “jumps” (virtually imperceptible to the naked eye) where the player won one of the minor prizes.
In contrast, notice there are many more ups and downs with Roulette. In this simulation, the casino ended up $4496 ahead after 50000 spins, an average of about 8.9 cents per spin, which is somewhat higher than the theoretical mean of $\mu = 5.3$ cents per spin. Over thousands of more plays, the sample mean would fall back to $\mu$. There are periods over thousands of plays where the casino (or you) has both lost money and won money. In the end, the weak law of large numbers guarantees the casino will make money (and the player will lose money) at a rate of $\mu$ per dollar bet. But the much smaller coefficient of variation gives the player a reasonable chance of being ahead in the short-term.

We certainly would encourage instructors and students to design their own simulations. One could choose to reproduce our simulations of Powerball and roulette, or choose to reproduce different lotteries or different casino games. It would be interesting to compute the mean, variance, and coefficient of variation for other games of chance to compare with our values for Powerball and roulette.

It is doubtful that anyone would play casino games such as roulette, blackjack, slot machines, craps, etc. if the coefficient of variation was of order of magnitude 6. The average player would never experience a win and would eventually quit playing, since the prizes available in this game are not monumentally large. However, people seem perfectly happy to risk a dollar (or many dollars) per week on the lottery although the $CV$ is monstrously large; the faint glimmer of hope of becoming an instant multi-millionaire is enough to keep millions of players interested.
5. Multiple Jackpot Winners

So far in our discussion, we have made the simplifying assumption that there is a unique jackpot winner. This is not a terrible assumption, since the empirical evidence gathered from 597 Powerball drawings from 11/5/1997 until 7/26/2003 yielded the following results:

| Number of Jackpot Winners | Number of Occurrences |
|---------------------------|-----------------------|
| 0                         | 538                   |
| 1                         | 49                    |
| 2                         | 9                     |
| 3                         | 0                     |
| 4                         | 1                     |
| 5+                        | 0                     |

Table 4. Number of Jackpot Winners, Powerball 1997-2003.

So about 90% of Powerball drawings fail to yield a jackpot winner, which leads to a larger jackpot available in the next drawing. This serves the interests of lottery commissions well. Since the Powerball game is very difficult to win, there is often an opportunity to build a large jackpot and stimulate a frenzy of ticket buying. Earlier, we assumed that all jackpot winners were unique to simplify the computation of the expected value and variance of Powerball. Empirically, we see that multiple jackpot winners is a relatively uncommon event. Only 10 of 597 drawings (about 1.7%) and 10 of 59 jackpot wins (about 17%) featured multiple winners.

We have already seen that the necessary jackpot for a positive expectation is rather large, about $281 million, for Powerball. This fact, coupled with the large variance and huge coefficient of variation, led us to conclude that one would need to play Powerball hundreds of billions of times to reasonably expect to make a profit. Practically, this is not possible.

A more advanced problem is to determine the probability of having exactly $i$ jackpot winners from among $k$ players who each randomly choose one of $n$ possible ticket combinations. For Powerball, $n = 120,526,770$. Equivalently, we can think of this as a problem where $k$ players each randomly choose one number (with repetition allowed) from the set $S = \{1, 2, ..., n - 1, n\}$.

We will name the players in this game $P_1, P_2, ..., P_{k-1}, P_k$, where you are player $P_k$. What is the probability that there will be exactly $i$ players including you who pick the winning number? First, since we are assuming you are a winner, we need to decide how many ways there are to have $i - 1$ winners out of the first $k - 1$ players. This number is $\binom{k-1}{i-1}$. Next, consider how many choices are possible for the remaining $k - i$ players who do not win. Each such player can choose from among $n - 1$ numbers (i.e. all
numbers except the winning number), which gives a total of \((n - 1)^{k - i}\) possibilities. Finally, there are a grand total of \(n^k\) choices that the \(k\) players in this game can make. Thus the probability that there are exactly \(i\) winners including you is:

\[
\frac{\binom{k-1}{i-1} (n-1)^{k-1}}{n^k}
\]

Now say the payoff is \(j\) split evenly among the winners. Then your expected winnings in this game would be (we are ignoring, for now, taxes and other factors that serve to reduce \(j\) for now):

\[
\sum_{i=1}^{k} \frac{\binom{k-1}{i-1} (n-1)^{k-1}}{n^k} \cdot \frac{j}{i}
\]

(4)

If \(k\) and \(n\) are large, as they will be for Powerball, then numbers like \(n^k\), \((n - 1)^{k - i}\), and \(\binom{k-1}{i-1}\) could be difficult to compute. In what follows, we suggest one approach for analyzing this expectation (students and instructors are encouraged to consider other approaches). After a bit of algebraic manipulation, (4) can be written as:

\[
\frac{j}{k} \sum_{i=1}^{k} \binom{k}{i} \cdot \left(1 - \frac{1}{n}\right)^{k-i} \cdot \left(\frac{1}{n}\right)^i
\]

(5)

Now we can use the binomial theorem to obtain a simplified exact expression of (5).

**Binomial Theorem** For positive integer \(k\),

\[
(a + b)^k = \sum_{i=0}^{k} \binom{k}{i} a^{k-i} b^i
\]

(6)

We will let \(a = 1 - 1/n\), \(b = 1/n\), and subtract off the \(0^{th}\) term, obtaining

\[
\frac{j}{k} \sum_{i=1}^{k} \binom{k}{i} \cdot \left(1 - \frac{1}{n}\right)^{k-i} \cdot \left(\frac{1}{n}\right)^i = \frac{j}{k} \left[\left(1 - \frac{1}{n}\right)^k - \binom{k}{0} \left(1 - \frac{1}{n}\right)^k \cdot \left(\frac{1}{n}\right)^0\right]
\]

\[
= \frac{j}{k} \left[1 - \left(1 - \frac{1}{n}\right)^k\right]
\]

Now let us introduce our Powerball particulars into the computations. The expected value of \(X\), the
average win/loss per Powerball ticket purchase, will be approximately equal to the expected winnings when we hit the jackpot, which we will call $W$, plus the expected winnings of the smaller prizes and no prize given in Scenarios #2 - #10, which we will call $V$. That is, $E(X) = E(V) + E(W)$, where

$$E(V) = -\frac{101,228,092.70}{120,526,770} \approx -0.84$$

and

$$E(W) = \frac{0.36j}{k} \cdot \left[ 1 - \left( 1 - \frac{1}{n} \right)^k \right]$$

We use $0.36j$ instead of $j$ since we expect to only take home 36% of the advertised jackpot value when we take the lump-sum payment and factor in taxes.

So in order to have a positive expected value, we need to solve the following inequality for $j$:

$$\frac{0.36j}{k} \cdot \left[ 1 - \left( 1 - \frac{1}{n} \right)^k \right] \geq 0.84 \quad (7)$$

In inequality (7), we know $n = 120,526,770$. The table below finds the minimum necessary jackpot $j$ for positive expected value for various values of $k$, assuming the possibility of multiple winners.

Table 5. Minimum Size of Powerball Jackpot Needed for a Positive Expected Value.

| Number of Players $k$ | Minimum Jackpot $j$       |
|-----------------------|---------------------------|
| 10,000,000            | $293,057,107$             |
| 20,000,000            | $305,207,483$             |
| 30,000,000            | $317,679,592$             |
| 40,000,000            | $330,472,332$             |
| 50,000,000            | $343,584,163$             |
| 60,000,000            | $357,013,121$             |
| 70,000,000            | $370,756,823$             |
| 80,000,000            | $384,812,481$             |
| 90,000,000            | $399,176,916$             |
Recall that the minimum jackpot necessary for a positive expectation, assuming a unique winner, was

\[ j = \$281,189,146.40 \]

It is typical for the Powerball lottery to have about 10 to 20 million players for the smaller jackpots (i.e. drawings held after wins) but at least 50 million players when the jackpot has been built up after several successive drawings without a jackpot winner. When we factor in the possibility of multiple jackpot winners, the advertised jackpot value needs to be very large (usually over $300 million!) to have a positive expectation on the purchase of a ticket. Coupling this with the application of the law of (very) large numbers from the previous section is quite sobering. It is just not rational to play Powerball with any expectations of winning money, even if we limit our play to large jackpot situations.

6. Factors in Ticket Purchase

Our calculations in the previous sections implicitly made a very large assumption which is false. We were assuming that we select a ticket combination at random, which we could very well do by selecting “Quick-Pick” or using our own random number generator. But we assumed that all of our fellow Powerball players are picking their combinations at random, which is not the case. Henze and Riedwyl (1998) state that quick-pick sales make up approximately 62% of ticket sales, and about that same percentage of jackpot winners used quick-pick. This means that about 38% of tickets purchased involved the person manually choosing their numbers. According to the FAQ at www.powerball.com, the figures are about 70%/30% in favor of quick-pick for powerball. Again, about 70% of jackpot winner in Powerball (41 of the first 59) used quick-pick.

Kadell and Ylvisaker (1991) were granted access from a lottery commission to data showing the number of purchases of each combination. They noted that certain combinations are purchased much more often that would be expected by chance. In fact, some combinations have been observed to have been purchased hundreds or even thousands of times in a single lottery. For example, the most popular combination in the October 29, 1988 drawing of the California Lotto 6/49 was 7-14-21-28-35-42, which was purchased 16,771 times. We could speculate the incredible popularity of this combination was due to the fact that we start with the “lucky” number 7, and then choose the first 6 multiples of 7.

Chapter 5 of Henze and Riedwyl (1998) discussed many popular strategies of players, which are therefore foolish since buying popular combinations increases the likelihood of splitting a jackpot with several others. Popular strategies include (not a comprehensive list):

- Choosing arithmetic progressions (e.g. 1-2-3-4-5-6 or 2-5-8-11-14-17)
- Choosing winning combinations from previous draws
- Modifying previous winning combinations (e.g. adding 1 to each number in a previous winning combination)
- Choosing “hot” or “cold” numbers (a statistically nonsensical strategy suggested in many of the lay books about lotteries)
- Choosing powers of 2 (e.g. 1-2-4-8-16-32)
- Choosing perfect squares (e.g. 1-4-9-16-25-36)
- Choosing all prime numbers (e.g. 2-3-5-7-11-13)
- Choosing Fibonacci numbers (e.g. 1-2-3-5-8-13)
- Choosing only numbers that are less than or equal to 31; many people choose numbers based on
birthdays, anniversaries, etc.

In general, using any simple rule to choose your numbers is foolish, since it is likely that others will use the same rule. Instead, you want the combination(s) that you purchase to be purchased only by you, in order to avoid splitting the jackpot. Henze and Riedwyl (1998) suggest that “quick-pick” is a simple way to make it more probable that you will avoid buying one of the popular combinations and also discuss some more sophisticated ideas for selecting a combination that is likely to be “unpopular”. However, their discussion is mostly focused on 6/s lotteries without a supernumber. Since Powerball has a supernumber and the chance of hitting the jackpot is so miniscule, it hardly seems worth the trouble to go beyond “quick-pick” in your quest for a unique combination.

The good news is that if we avoid choosing some of the particularly popular combinations (as listed above), we should be able to somewhat reduce the probability of sharing a jackpot. Between 11/5/1997 and 7/26/2003, the largest number of distinct winners in Powerball was 4, which occurred in the drawing of August 15, 2001. It would be very interesting to a neutral spectator to observe a lottery with dozens or even hundreds of jackpot winners if one of the very popular combinations were to occur. Of course, the winners would be annoyed that their lucky moment was not as rewarding as it could have been!

7. Conclusion

We have investigated various combinatorial and statistical aspects of the popular lottery game known as Powerball. These aspects can be used as motivating examples in statistics and probability courses at varying levels of mathematical sophistication. The simple counting arguments made in section 2 are used to develop a reasonable probability distribution or model describing one’s chances in Powerball. These results are accessible to even a non-calculus based introductory course in statistics. These students are also equipped to follow through the results of section 3, where the expected value and variance for the probability distribution are found. Even if one stops at this point, these students can see that the jackpot needs to be very large before there is any hope of Powerball having a positive expectation from the player’s standpoint. This would also be a great opportunity to introduce the underrated coefficient of variation, which is used in section 4.

The subsequent sections are mathematically more sophisticated and one may wish to avoid or at least “handwave” these results in the non-calculus course. However, students enrolled in a standard post-calculus course in probability and statistics, such as an engineering statistics or a typical undergraduate sequence in mathematical statistics should be able to follow section 4, where it is shown that the law of large numbers often requires “large” to be very large. One who follows the lead of Rossman, Chance, and Ballman (2000) and uses an activity-oriented approach might have the students re-do my simulations of Roulette, Powerball, or other games of chance for themselves. Section 5 removes the assumption that there is a unique jackpot winner.

Section 6 is a non-technical section that is accessible to any level of student. The instructor might wish to point out that all possible tickets are equally likely to be chosen in the drawing, despite the poor advice of self-proclaimed lottery “experts” like Howard (1997). However, as is thoroughly discussed by Henze and Riedwyl (1998), it is desirable to avoid certain combinations. The instructor might also wish to discuss why using randomization via the “Quick-Pick” option is an easy way for the player to avoid choosing a popular combination. The class could discuss why Henze and Riedwyl are correct in advocating “Quick-Pick” for those who must play and why Howard is wrong in her condemnation of “Quick-Pick” and advocacy of finding “hot” and “cold” numbers.
Acknowledgements

The authors would like to thank the editor and the referees for their comments on an earlier version of this paper that led to an improved final version. In particular, the mathematics in Section 5 is now much simpler and more elegant due to a referee.

References

Henze, N. and Riedwyl, H. (1998), How to Win More: Strategies for Increasing a Lottery Win, Natick, MA: A.K. Peters.

Howard, G. (1997), Lottery Master Guide, (3rd ed.), Las Vegas, NV: Smart Luck Publishers.

Kadell, D. and Ylvisaker, D. (1991), “Lotto Play: The Good, the Gair, and the Truly Awful,” Chance, 4, 22-25.

Mood, A., Graybill, F., and Boes, D. (1974), Introduction to the Theory of Statistics (3rd ed.), New York: McGraw-Hill.

Multi-State Lottery Association (2003). Home page found at: www.powerball.com

Rossman, A., Chance, B., and Ballman, K. (2000). “A Data-Oriented, Active Learning, Post-Calculus Introduction to Statistical Concepts, Methods, and Theory,” Proceedings of the American Statistical Association, Section on Statistical Education.

Christopher J. Mecklin
Department of Mathematics and Statistics
Murray State University
Murray, KY 42071
U.S.A.
christopher.mecklin@murraystate.edu

Robert G. Donnelly
Department of Mathematics and Statistics
Murray State University
Murray, KY 42071
U.S.A.
rob.donnelly@murraystate.edu