Jackson Campbell

Byron Hoy

Probability & Applied Statistics

Statistical Textbook Research Project

Data Set: Super Smash Bros. Melee Set Records

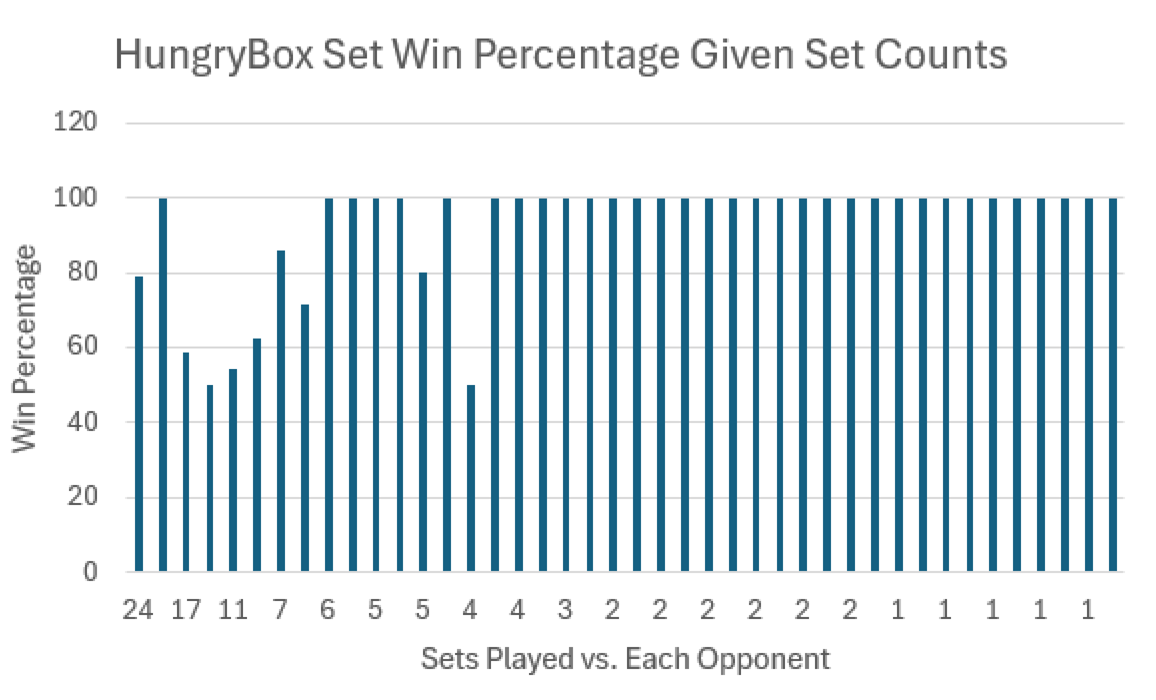
Abstract/Introduction:
Below is a statistical report on the Set History of Super Smash Bros. Melee top competitors, styled after each section covered throughout this past semester of probability and applied statistics with Byron Hoy. Each section covered is meant to give a breakdown or brief overview of the data within the excel spreadsheet and documentation.

Section 1.2: Graphical Methods of Statistics

Example 1) HungryBox

Since the SSBM leaderboards have made the switch to the newer-formed platform, HungryBox, a top-level competitor, has faced off against 42 different opponents. Given the win percentage, and match count, how could you construct a relative histogram to display the win percentage in decreasing set count order?

Answer:



What this entails within the given data is that while set counts steer higher based on which player HB is competing against, he has been a highly dominant player within the Melee competitive scene for years since entering. The lowest win percentage value recorded within the table for HB is 50%, which further proves how strong of a competitive player he is.

Section 1.3: Numerical Methods of Sets

Example 2) HungryBox Continued

Given the top 8 competitors against HungryBox in set count descending order, what is the mean of win percentages, and the standard deviation of the win percentage set? What is one standard deviation below the mean?

Answer: Mean – 70.27%, Standard Deviation – 17.12, One SD below the mean – 53.15%

What this shows further about HungryBox, continuing off of the previous example, is how strong his win percentage still is while only taking a small sample of his entire competition. While his average against the top 8 most played matchups is about 70%, his true win percentage sits at around 83%, while the true standard deviation is about 15.43. This set of numbers is similar to the smaller sample taken earlier, and continuing on and taking one standard deviation below the mean here yields a higher 67.57%, still 17 percent above the lowest recorded win percentage in the data spread.

Section 2.3: Set Notation

Example 3) Mew2King

Suppose Mew2King is to face off against competitors Mango and Leffen and has an equally likely chance to beat both competitors. Given all outcomes, the set of wins and losses after both matches is as follows:

$S = \{WW, WL, LW, LL\}$

Given the set above, create the following where A denotes the subset of no wins, B denotes the subset of at least one win, and C is the subset of two wins: A, B, C, A $\cup$ B, A $\cap$ C

$$A = \{LL\}, B = \{WW, WL, LW\}, C = \{LL\}, A \cup B = \{LL, WW, WL, LW\}, A \cap C = \{\emptyset\}$$

While these sets are not related to the given data, it is helpful to recall set notation and set operations as was performed during the beginning of the semester.

Section 2.4: Discrete Cases

Example 4) Mew2King Continued

Over the course of years of competition, Mew2King has faced off against several competitors, many of whom have shown up more than once. Given all matches played, M2K has a 1 in 7 match history frequency with HungryBox, 1 in 16 with Axe, 1 in 23 with ChuDat, and 1 in 29 with Armada. Given one competitor is to be randomly selected, what is the probability that Axe will not be Mew2King's matchup?

Answer: 15/16 or 93.75%.

While Mew2King has played against Axe in 7 total sets across just over 100 played, there are still several other competitors to face off against, all with varying frequencies themselves. This not only levels out the playing field, but gives a broader understanding of strengths and weaknesses when playing with other competitors in tournaments and local competitions.

Section 2.5: Sample-Point Methodology

Example 5) Listing Sample Spaces within Armada's Set History

Across the top competitors of Super Smash Bros. Melee, Armada, one of the top competitors, has an astounding 84.2 percent win percentage. Across the 76 sets collected, Armada has played games against Leffen, HungryBox, and Mew2King 9, 8 and 4 times respectively out of the 76 matches. Assign this data to a set, and include the other matches played. Then, assign probabilities to each point in the sample.

Answer: S = {AvL, AvHB, AvsM2K, AvsELSE}, P(S) = {.118, .105, .053, .724}

What this spread of possibilities shows within Armada's set history is while Armada has faced off against more when competitors are higher skilled such as HungryBox or Mew2King, he still has matches spread largely across the rest of the competitors listed in the set history. Even at the highest frequency of matches at 9 played against Leffen, the statistical frequency is only slightly larger than one tenth of the max value.

Section 2.6: Counting Sample Points

Example 6) Melee Tournament Setups

Given the tournament legal rules, players are allowed to choose from a selection of 5 different stages, 4 different rulesets during the match, 2 different versions of the game (patched or unpatched due to game-breaking bugs), and are allotted 8 different tournament legal characters at most tournaments. Given these conditions, how many unique stage matchups could be created for competitive play? Given a tournament of 16 players, where 2 players are to face off in a single match, how many possible match conditions can then be created for the entire tournament?

Answer: Part 1 = $5 * 4 * 2 * 8 * \ 8 = 2{,}560$ matchups, Part 2 = $\binom{16}{2} * 2560 = 307{,}200$ tournament match possibilities

From this panel of data, what can be gathered is that there are a vast array of possibilities for matchups in Super Smash Bros. Melee, including different stages, characters, and rulesets themselves for tournament play. Many players do not use a single character, so matches tend to vary widely based on which opponents are facing off within the bracket structure.

Section 2.7: Conditional Probability and Independence

Example 7) Mango and Sequential Wins

Given Mango's win percentage against HungryBox and Armada, with 5/11 and 2/9 respectively. Mango must not only face off against these two competitors, but then face off against Axe afterwards. Given the probabilities for the first two matches, how likely is Mango to win against Axe after beating the first two matches?

Answer: Since these matches are all independent of each other, the following can be performed:

Given: Mango vs. HungryBox = 5/11, Mango vs. Armada = 2/9, P(1)*P(2) = 10/99

$$P(B \mid A) = P(B) * P(A) = \frac{4}{6} * \frac{10}{99} = \frac{20}{297} \; or \; 6.73\%$$

Given that these events are independent of each other, to find our result is to simply multiply the given values together to find our highly unlikely outcome of Mango winning all three of these matches. Mango was only given a 10.1% chance to win against both HungryBox and Armada originally, and taking in the set history between Mango and Axe with a win percentage of 2/3, an even lower result of winning all three matches is found, showing that while Mango is still a high-level competitor for Super Smash Bros. Melee, the competition he faces is just as strong, if not stronger.

Section 2.8: Two Laws of Probability

Example 8) HungryBox and His Rival, Plup

Given HungryBox's strong overall winrate, with a staggering 157 wins out of 189 sets played, and his history against one of his main rivals, Plup, with a history of 19 wins out of 24 sets played, what is the probability of HungryBox winning against Plup again under either condition? What is the probability that Plup wins given these conditions?

Answer: $P(A \cup B) = P(A) + P(B) - P(A \cap B), \quad P(A \cap B) = \frac{157}{189} * \frac{19}{24} = .6576$

$$P(A \cup B) = .8306 + .7917 - .6576 = .9647 \; or \; 96.47\%$$

$$P(Plup\ wins\ instead) = 1 - .9647 = .0353 \; or \; 3.53\%$$

Given these outcomes, not is HungryBox a strong competitor in Melee, but nearly has a perfect chance to win against his rival Plup. Given both his overall win percentage and their match history, it is extremely unlikely for Plup to take a match off of HungryBox in any given set or tournament.

Section 2.9) Event-Composition Method

Example 9) SFAT Sequential Sets

Assuming independence from one another, if SFAT were to play a general 3 set series, what is the probability that he were to lose all three sets? What is the probability that all sets are won?

Answer: SFAT's overall win percentage: 77/123, or .626

$$P(All\ three\ won) = .626^3 = \ .2453\ or\ 25.53\%$$

$$P(All\ three\ lost) = (1 - .626)^3 = .0523\ or\ 5.23\%$$

Given these results, it shows that while SFAT has an overall positive win percentage as it lies above 50 percent, winning three sets sequentially will prove to be difficult with a probability of only 25.53 percent. However, it becomes even more unlikely to lose all three sets, as finding results leaves a 5.23 percentage rate to lose all three sets in a row. Thus, the most likely outcome is a mixture of wins and losses within these sets given the low rate for both outcomes.

Section 2.10: Baye's Rule and Total Probability

Example 10) Tournament Stakes

Envision a high-level tournament where both Mango and HungryBox are competing. We want to determine the probability that a randomly selected loss in the tournament belongs to Mango based on historical win rates and set history.

Setup: Mango's win percentage = 50/74 = .6757 = 65.57%
HungryBox's win percentage = 157/189 = .8307 = 83.07%
Mango's loss percentage = 1 - .6757 = .3243 = 32.57%
HungryBox's loss percentage = 1 - .8307 = .1693 = 16.93%
Total number of matches played = 189 + 74 = 263
Probability of Mango's matchup selected in this group = 74/263 = .2813 = 28.13%
Probability of HungryBox's matchup selected in this group = 189/263 = .7186 = 71.86%

Answer: Given the setup, total probability can now be obtained by computing the following:

$$P(Loss) = P(Loss|Mango) * P(Mango) + P(Loss|HBox) * P(HBox)$$

$$P(Loss) = \ .3243 * \ .2813 + \ .1693 * \ .7186 = \ .0912 + \ .1217 = \ .2129\ or\ 21.29\%$$

$$P(Mango|Loss) = \frac{P(Loss\ |Mango)P(Mango)}{P(Loss)} = \frac{.3243 * .2813}{.2129} = .4285\ or\ 42.85\%$$

From this, we can conclude that if a randomly selected loss from any tournament statistic were taken from these two, there is roughly a 43 percent chance that the match would be a loss on Mango's part. While there is a higher chance to select HungryBox's loss out of the bunch, this does not reflect on win percentage, as the match count itself is skewed towards HungryBox and thus holds a higher chance of selecting one of his 32 losses instead of Mango's 24.

Section 3.2: Distribution for Discrete Random Variables

Example 11) Reviewing Plup's Matches

When training for upcoming tournaments, it's crucial for players to watch their own matches to catch mistakes to learn how to improve their gameplay to increase their skills. While reviewing his own matches, Plup has a selection of matches to watch when scouting his gameplay against HungryBox and Cobol. Given the total match count, and wanting to review a total of 9 games, what is the probability that Plup reviews 4 games from HungryBox?

Answer:

$$p(4) = P(Y = 4) = \frac{\binom{24}{4}\binom{22}{5}}{\binom{46}{9}} = .254 \ or \ 25.4\%$$

Given this, when training and reviewing matches, there is only a 25.4 percent chance that within this set of 9 games to watch out 46, only 4 of them will be from HungryBox while 5 will be from Cobol. This gives a spread of data that can be input into different variable circumstances with this equation, demonstrating the outcome of training and who is more likely to be reviewed for tournament preparation.

Section 3.3: Expected Value of a Random Variable or Function

Example 12) Mango's Matchup History

Given the matchup history of Mango and the 29 competitors he has faced within the data set, calculate the mean and variance for the losses Mango faced for each opponent.
Setup:

$E(Y) = \sum_y yp(y)$ Given this formula for the expected values, p(y) is the probability of a certain amount of losses given a random competitor. Thus, a breakdown of matchup loss counts can be formed to find p(y) with the following data:

- 0 Losses: 21 out of 29 matchups
- 1 Loss: 3 out of 29 matchups
- 2 Losses: 1 out of 29 matchups
- 3 Losses: 2 out of 29 matchups

- 6 Losses: 1 out of 29 matchups
- 7 Losses: 1 out of 29 matchups

With this, the expected value of losses can now be calculated using the above summation:

$$E(Y) = \sum_{y} yp(y) = 0(.724) + 1(.103) + 2(.034) + 3(.069) + 6(.034) + 7(.034)$$
$$= 0 + .103 + .068 + .207 + .204 + .238 = .82$$

With this computed, the variance can now be computed using the following formula:

$$V(Y) = E[(Y - \mu)^2] = \sum_{y} (y - \mu)^2 p(y)$$

Following this formula, values can be added to complete the variance:

$$V(Y) = \sum_{y} (y - \mu)^2 p(y)$$
$$= (0 - .82)^2(.724) + (1 - .82)^2(.103) + (2 - .82)^2(.034)$$
$$+ (3 - .82)^2(.069) + (6 - .82)^2(.034) + (7 - .82)^2(.034)$$
$$= .487 + .003 + .047 + .328 + .912 + 1.299 = 3.076$$

With this, the expected and variance for Mango's set losses per opponent given the set history is .82 and 3.076 respectively. This shows that Mango's expected loss value is under 1, which is strong given the skew of 0 losses within competitors. Although, losses can veer as high as above this as 3 given the variance found above. While Mango is a strong competitor, the set history shows he's no stranger to losing in the history of sets, with just over a third of sets counting for a loss.

Section 3.4: Binomial Probability Distribution

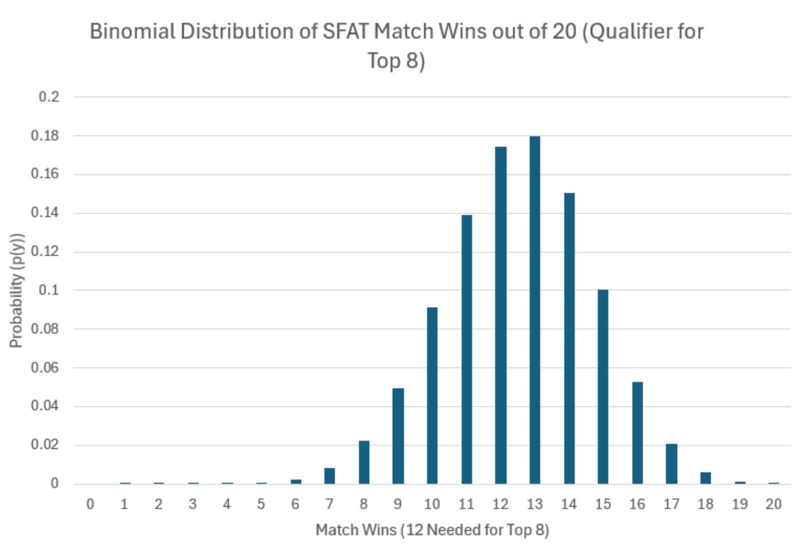
Example 13) Qualifying for Top 8

Suppose SFAT is competing for a chair in top 8 of a large scale, single elimination tournament. To qualify for top 8, 4 sets of best-of-5 must be won. Given these details, how likely is SFAT to qualify for a spot in top 8 of the tournament?

Answer: Given these details of the tournament, we can now formulate a binomial distribution solution to this problem. Given the prior win percentage of SFAT's set history, we can apply his winnings to this to find our success coefficient, p, with a value of .626, while our failure value, q, is simply 1 - .626, or .374. Now, an equation can be formed:

$$p(y) = \binom{20}{12} .626^{12} * .374^{20-12} = .1746 \; or \; 17.46\%$$

With this newfound data and formulation, we can see that qualification for top 8 within the tournament for SFAT is far unlikely to happen, as while above 50 percent, his win percentage is on the far lower end compared to some of his other competitors in the higher ranks of competitive Super Smash Bros. Melee. A player such as Armada or HungryBox has a higher chance of qualifying for top 8 rather than SFAT due to their higher win percentages given their set history. Furthermore, the expected value for this specific distribution falls at 12.52, demonstrating that even with a low possibility of winning 12 matches, it's the expected value of matches won. However, by looking at the variance, it shows that you could fall as low as almost 5 matches with a variance of 4.68, counteracting the 12.52 expected value set earlier and demonstrating that while although the outcome should be 12 wins, it could be far lower and SFAT misses the bar to qualify. Given the max number of 20 wins, a histogram for this specific binomial distribution can be created, showing that the data itself is in the form of a normal "bell" curve skewed towards 13 matches, which aligns with the expected value of 12.52.



Section 3.5: Geometric Distribution

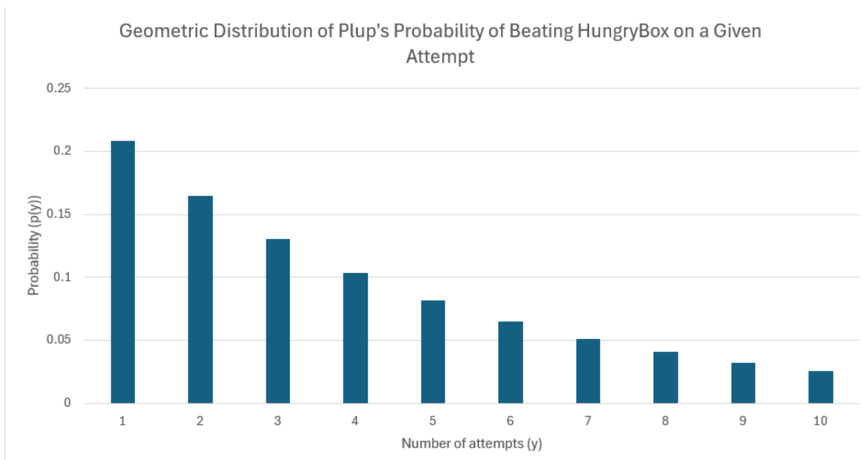
Example 14) Plup vs HungryBox

Plup, an experienced competitor, is rivals with one of the best players in the entire world, HungryBox. However, while facing each other several times throughout history, Plup has an abysmal win percentage, standing at only 20.8 percent. With these given odds, what is the probability that Plup wins against HungryBox after the 8^{th} attempt?

Answer: Given Plup's win percentage, standing at 20.8 percent or .208, a geometric probability distribution can be formed to find the answer. Given the amount of attempts, the equation is as follows:

$$p(y) = .792^{8-1} * .208 = .0406 \text{ or } 4.06\%$$

After plugging in the percentages above, along with our exponent of 8-1, a probability of only 4.06 percent arises on the 8^{th} attempt. While Plup has a much higher win percentage than just 4 percent, winning on this attempt for the first time is a much lower percentage. The likeliest attempt to win is actually the first attempt, as when plugging in 1 instead of 8, the same value for success, p, is found as the exponent becomes 0. The odds decrease as the attempts increase, as Plup is more likely to win for the first time on lower attempts instead of later on in matches. A histogram of probable outcomes can be created for this geometric distribution, and is pictured below this. Checking the expected value, which is simply 1 over the value of success, a value of 4.807 is found, denoting that the expected value of attempts before a success occurs is around 4.8 on average. While looking at the variance, we can see that this result varies as much as 18.3, which is a vast number of attempts, and after adding this onto the expected value, it could take Plup as many as about 23 attempts before he sees a win against HungryBox.



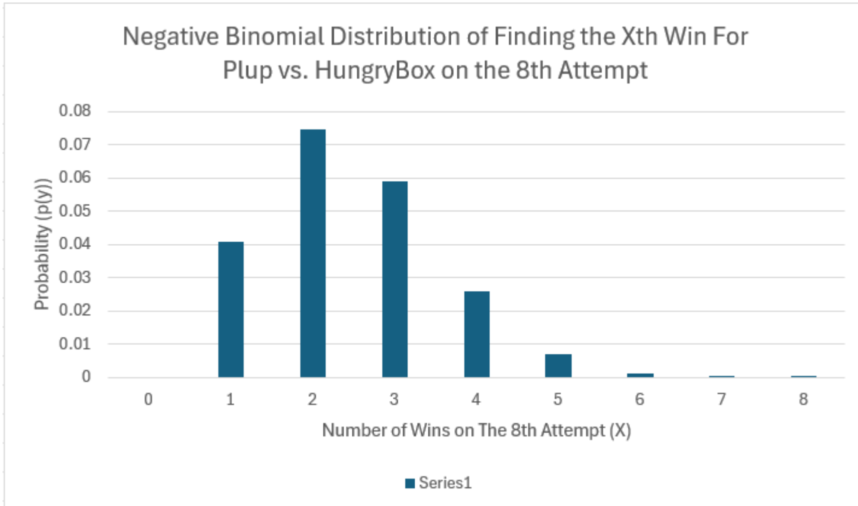
Section 3.6: Negative Binomial Probability

Example 15) Plup vs HungryBox Continued

Continuing on the above section, a negative binomial distribution can be used to determine the specific win number on a given attempt. Following this, and while still keeping 8 total attempts, a histogram can be created to display the probability of winning an nth time on a specific attempt, y. In this case, y would still be 8 as set above for geometric, but instead of looking for the first win, any number of wins can be looked for using the formula below:

$$\binom{8-1}{y-1}.208^{y} * .792^{8-y}$$

With said formula, numbers 0 through 8 can now be plugged into y, giving us the demonstrated outcome in the form of a histogram below:



With this newly created graph, a spread can be found from values 1 to 8, denoting the likelihood of finding the Xth success on the 8^{th} try, with 2 successes on the 8^{th} try being the likeliest value. Furthering this, the expected value found for 2 successes is 9.61, which is higher than the number used in the trial, 8. Following this, the variance is wildly higher than any other number at 36.61, showing that it could take as high as 37 matches to find exactly the 2^{nd} win. This data shows that the likeliest outcome of winning a singular match by the 8^{th} attempt is not only far lower than winning 2 matches by then, but 2 matches won by the 8^{th} attempt is just under the expected value, and instead on average it would take 8 attempts to attain 1.664 wins, which was discovered by plugging in values into a scientific calculator. As 1.664 is closer to 2 than 1, it shows the much higher likelihood of this result instead of the geometric result.

Section 3.7: Hypergeometric Distribution

Example 16) SFAT Qualifier Continued

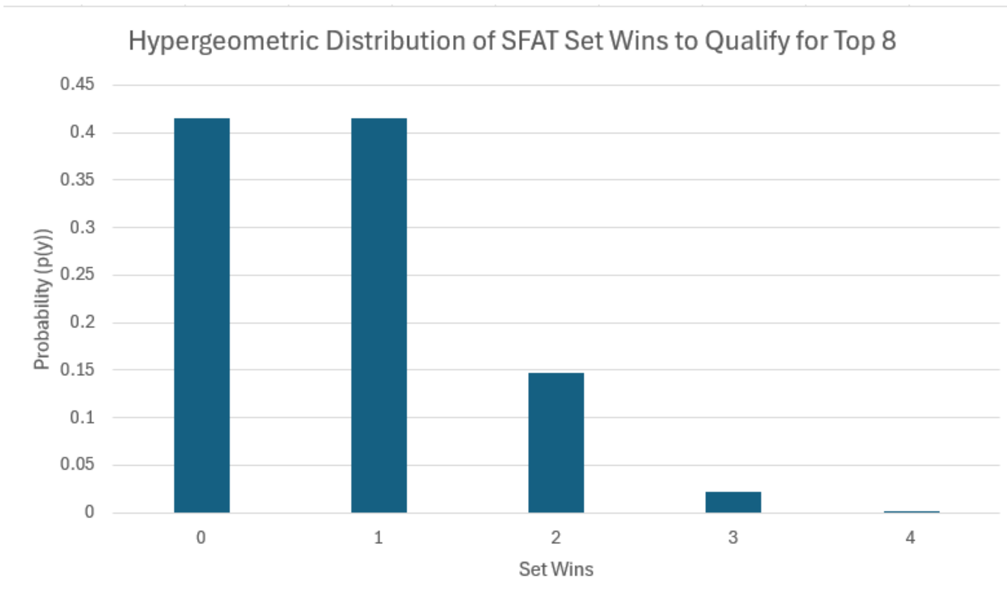
Backtracking to example 13, a hypergeometric probability distribution can be formulated for the large scale tournament given that 25 players hold win percentages above 65 percent in the tournament, and given that SFAT must win 4 sets to qualify for top 8, it can be assumed that there are 128 players within said tournament. Knowing this, a formula can now be created for hypergeometric distribution as shown below:

$$\frac{\binom{25}{4}\binom{128-25}{4-4}}{\binom{128}{4}} = .00118 \; or \; .118\%$$

Given these numbers, 128 is the total number of players in the tournament, 25 are the players with a high winning percentage, and both fours are the matches SFAT played and subsequently won to qualify for top 8. His very own winning percentage is not taken into account here, as it is unnecessary for the hypergeometric distribution function. Furthermore, the expected and variance values can now be calculated below alongside the formula above:

$$E(Y) = \frac{4 * 25}{128} = .78125, \quad V(Y) = 4\left(\frac{25}{128}\right)\left(\frac{128-25}{128}\right)\left(\frac{128-4}{128-1}\right) = .6138$$

With these values formulated, it can be shown that on average, SFAT would likely face around only 1 strong competitor within these 4 matches, but can vary just by only .6, meaning his chances of facing a strong player in these qualifying matches is still relatively low. Furthermore, to see a larger spread of outcomes, a histogram with the amount of wins SFAT receives can be created as shown below:



Given this, it can be shown that winning 0 or 1 set exactly are the most likely outcomes within this tournament, and odds decrease dramatically after 1, falling as low as below 1 percent as found earlier. With all data given, it shows that SFAT is highly unlikely to qualify for top 8 given the full breakdown of the tournament, no matter how likely he is to win 12 sets shown earlier in the binomial report section.

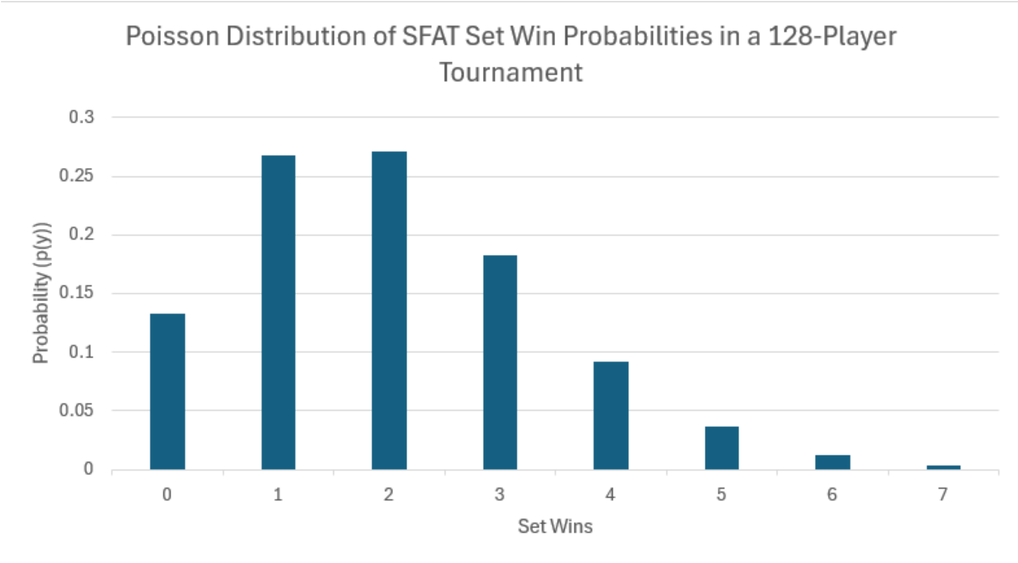
Section 3.8: Poisson Distribution

Example 17) SFAT Qualifier Continued pt. 2

Furthering the data above in sections 3.7 and 3.4, a Poisson distribution can now be created to see SFAT's expected results in this tournament. Given his prior tournament entries, around 38, and his total set wins, 77, the lambda for Poisson can now be found at 2.02 wins per tournament on average. Given this, a formula can now be set to discern how likely SFAT is to win a number of sets in a given tournament:

$$p(y) = \frac{2.02^y}{y!}e^{-2.02}$$

Given this formula and calculation, a histogram for the likely outcome of set wins in a 128-player tournament can now be found as shown below:



As shown, SFAT is likeliest to win 2 sets, as average set by the lambda above, he wins about 2.02 sets per tournament. The outcome itself holds the highest value of 27.06 percent, signifying that it's the likeliest outcome. Furthering this data, the expected and variance values can also be found, as both are equal to our lambda of 2.02. Given this, the expected value, or average, is equal to the lambda value as the lambda is also an average value for this function. Since the lambda value is also equal to the variance, it shows that results can vary as much as 2 set wins, giving SFAT a possibility of qualifying for a spot in top 8, as 2 wins higher than the average value is 4, guaranteeing a spot in top 8 as set in above conditions.

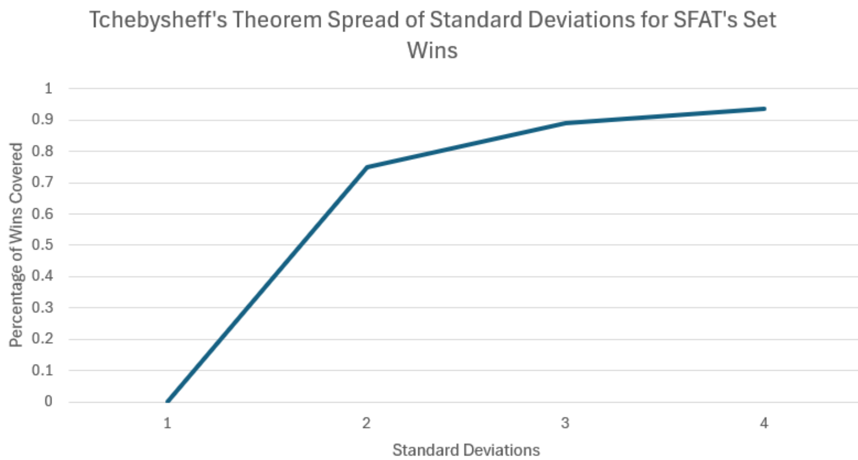
Section 3.11: Tchebysheff's Theorem

Example 18) SFAT, Final

On the topic of SFAT and set history furthermore, given his total win history and opponent numbering, an expected, variance, and standard deviation value can be found at 3.046, 8.23, and 2.87 respectively. With these data points, a spread using Tchebysheff's Theorem can now be created with the following formula:

$$P(|Y-\mu| < k\sigma) \geq 1 - \frac{1}{k^2} \;\; or \;\; P(|Y-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

By plugging in the expected and variance and computing these values, we now create a spread of how many standard deviation ranges cover SFAT's set wins:



With this graph, it's shown that by 4 standard deviations, almost all of SFAT's wins are covered in the spread. As the distribution itself is not normal, 1 standard deviation cannot be guaranteed any coverage here, and therefore starts at 2 standard deviations.

Note: Chapter 4 is structured around Continuous Random Variables, and Probability Density Functions and Cumulative Distribution Functions. All data will focus on HungryBox and the set history of all matches played against other competitors for this specific chapter.

Section 4.2: Distribution for Continuous Random Variables

Example 19) HungryBox's Set Distribution

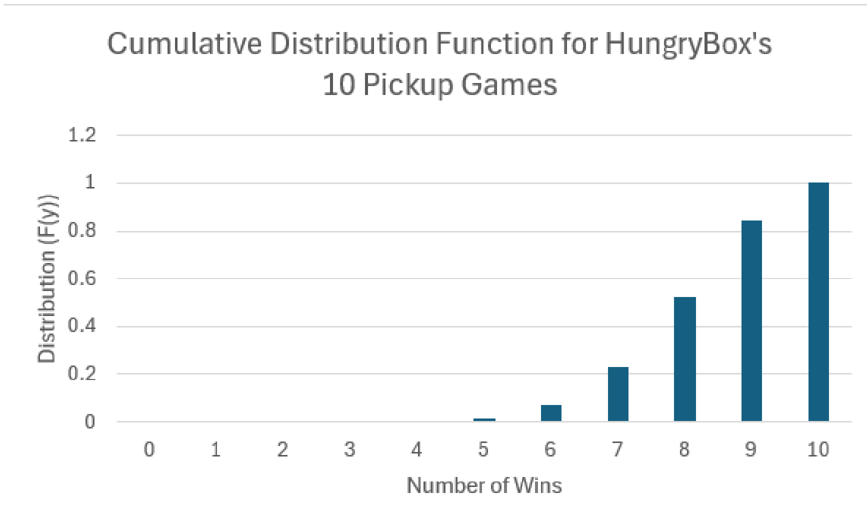
Suppose that HungryBox, one of the strongest competitors in Melee's history, has a win percentage overall of 83.06 percent, is to play 10 pickup sets against different opponents in any given order. Knowing this, a cumulative binomial distribution can be created to track where the totality of data lies on the spread from 0 to 10. The formula itself is as follows, similar to a discrete binomial function but with a continuous variable instead of a discrete variable:

$$p(y) = \binom{10}{y}.8306^{y} *.1694^{10-y}$$

With this function, a cumulative distribution can now be created with the data as such:

$$F(y) = \begin{cases} 0, & for\ y < 0 \\ 0.000000019004, & for\ 0 \leq y < 1 \\ 0.0000009535, & for\ 1 \leq y < 2 \\ 0.00002163, & for\ 2 \leq y < 3 \\ 0.00029277, & for\ 3 \leq y < 4 \\ 0.002626, & for\ 4 \leq y < 5 \\ 0.016392, & for\ 5 \leq y < 6 \\ 0.072805, & for\ 6 \leq y < 7 \\ 0.23131, & for\ 7 \leq y < 8 \\ 0.52358, & for\ 8 \leq y < 9 \\ 0.84296, & for\ 9 \leq y < 10 \\ 1, & for\ y \geq 10 \end{cases}$$

With this newly-formed function, a graph can now be formed to show the spread of the cumulative distribution, and where the function itself lies:



With this chart, the data shows that the most likely area of win, and the largest distribution of probability lies between 8 and 9, as the jump between the two variables is the largest compared to other numbers plugged into the y variable of the above function. Looking at a discrete standpoint, this is further proven by the result of y = 9, with a probability of .319, the highest probability compared to other numbers of wins in the function.

Section 4.3: Expected Values of Continuous Variables

Example 20) Set Distribution Continued

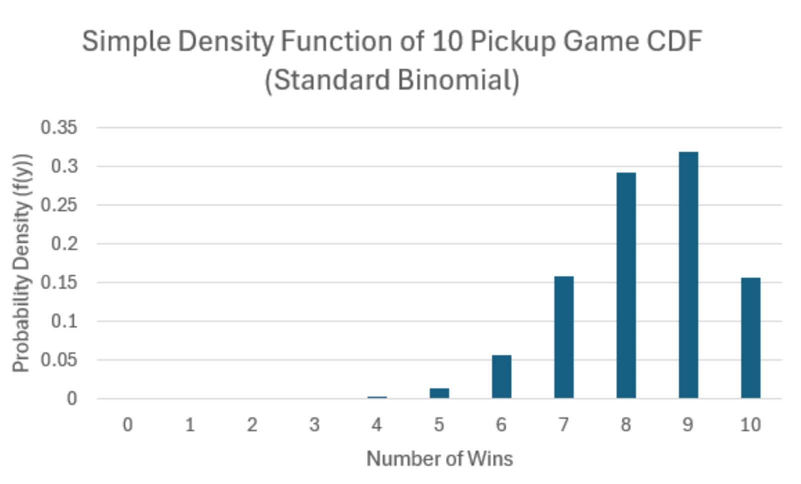
Given our foundational cumulative distribution function found in example 19, the expected values for this function can now be found by using the following formula set below:

$$E(Y) = \int_{-\infty}^{\infty} y f(y) dy$$

However, since the density function above involved discrete values, a standard integral will not evaluate the density function properly, and instead needs to be evaluated using a simple rate of change interval of F(y) – F(y-1) to yield each result. This then yields an answer for each variable for the probability density function:

$$f(y) = \begin{cases} 0 \\ 0.000000019004 \\ 0.0000009345 \\ 0.000020677 \\ 0.0002711 \\ 0.002333 \\ 0.013767 \\ 0.05641 \\ 0.158507 \\ 0.29227 \\ 0.31937 \\ 0.15704 \end{cases}$$

Given this, it is now visible that the density function of the above CDF is simply the values for the discrete binomial function, and can now be used to calculate the expected values for the distribution given the above formula. Since the function used was discrete and not continuous, an approximation summation instead of the integral above. Given this, the expected value is now 8.31, which falls in line with the standard binomial function expected value summation. A histogram for the PDF is also shown below as a reference:



Section 4.4: Uniform Distribution

Example 21) Uniform Win Distribution

Suppose that HungryBox's wins across all player matchups are uniformly distributed, with the underlying data in the table, it is clear that the amount of wins ranges from 1 to 19. Given this, a uniform density function can now be created to calculate any given range of wins in a randomly selected series with the following formula:

$$f(y) = \begin{cases} \dfrac{1}{19-1}, & if\ 1 \le y \le 19 \\ 0, & otherwise \end{cases}$$

Given this function, an integral to calculate any F(y) = P(Y) can be created, so long as y exists within the set range of wins. For example, if the probability that HungryBox were to win between 7 and 13 games were to be calculated, the following integral can be created and calculated:

$$P(7 \le y \le 13) = \int_{7}^{13} \frac{1}{19-1} dy = \frac{13-7}{19-1} = \frac{6}{18}\ or\ \frac{1}{3}$$

Given this spread of wins from 7 to 13, there is a 1/3 chance that the number of sets HungryBox wins will land between these numbers. The same calculations can be done for any set integral, and 13 and 7 can be replaced so long as they lie between the range set above in the density function. For this specific range and uniform function, the expected and variance values can also be calculated using the formulas below:

$$E(Y) = \frac{1+19}{2} = 10, V(Y) = \frac{(19-1)^2}{12} = 27$$

Under the uniform distribution, it is expected that HungryBox would win 10 sets per series of games on average given the function above. The variance stated shows that the possible set wins can vary largely, especially given how often certain matchups appear such as facing against Plup, with a total of 24 sets against each other.

Section 4.6: Gamma (Exponential) Distribution

Example 22) Time Before a Win

As found and stated much earlier in the report, HungryBox has an overall win percentage of .8306, and this number can be used to find the value of beta (β) in the exponential function. Using this, our value for beta simply is 1 / .8306, or 1.204. Using this value, an exponential function can be formed using the formula below:

$$f(y) = \begin{cases} \frac{1}{1.204} e^{-y/1.204}, & 0 \le y < \infty \\ 0, & elsewhere \end{cases}$$

Using this function, a cumulative distribution function can now be created with this formula to calculate percentages of finding a win within a given number of sets. For example, the question $P(Y \le 2)$ can now be solved, where Y is the amount of sets it would take to win. To solve this, simply plug 2 into y, and then the function itself can now be solved using the formula below:

$$P(Y \le 2) = \int_0^2 \frac{1}{1.204} e^{-y/1.204} dy = .81007 \text{ or } 81.01\%$$

Given this result, HungryBox has roughly an 81.01 percent chance of winning within 2 sets played. To calculate other probabilities, simply change the value of 2 in order to find how likely HungryBox is to win within a certain amount of sets played. While testing in Desmos, when calculating $P(Y \le 35)$, the result evaluated to 1, showing that at 35 sets played and beyond, HungryBox is guaranteed to win a set by said amount of games played, with the calculation shown below:

$$\int_0^{35} \frac{1}{1.204} e^{-\frac{y}{1.204}} dy$$

$$= 1$$

Note: Sections 5.2-5.3 Are condensed together for a final section below, as all sections relate to each other and therefore can be condensed into a larger, single example. This chapter section will also reintroduce Armada, as Armada's win percentage is highly similar to HungryBox, allowing for a joint probability function creation.

Sections 5.2, 5.3, 5.4: Bivariate/Multivariate Functions, Marginal/Conditional Probability Distributions, and Independent Random Variables

Example 23) HungryBox and Armada's Wins in Tournaments

Given that HungryBox's win percentage lies at 83.01, and Armada's lies at 84.21, and assuming Mew2King differs within +- 1.5 percent for the sake of setting up a joint probability distribution, a function can now be calculated with the notion that HungryBox is to play 3 matches. Let X denote the number of matches won by HungryBox, and Y to denote the number of matches Armada wins. With this, either player can receive 0, 1, 2, or all three wins. With this, a table can now be created to display the spread and likelihood of wins:

Armada's Wins\ HungryBox's Wins	*0*	*1*	*2*	*3*	*Total*
0	1.9308E-05	0.000283	0.00138268	0.00225184	0.0039195
1	0.00030891	0.0045278	0.02212201	0.03602802	0.0629867
2	0.00164744	0.02414733	0.1179794	0.19214185	0.33591602
3	0.00292867	0.04292689	0.20973285	0.34157199	0.60008907
Total	0.00488695	0.07188502	0.35121694	0.5719937	1

From the table above, independence can also be calculated, and shown with the following formula:

$$p(x, y) = p(x)p(y)$$

Following this formula, if p(0,0) is tested for, this simply means that p(x = 0) and p (y = 0) are multiplied, giving the following series of calculations:

$$p(0,0) = p(x = 0)p(y = 0)$$
$$p(0,0) = 0.000019, p(x = 0) = 0.00488695, p(y = 0) = .0039195, p(0)p(0) = .000019$$

With these calculations, it can be verified that x and y, or rather HungryBox and Armada's win totals, are independent from each other as well. With this, and other tournament-style competitions, most matches are independent from each other when overlaying this specific scenario on top of others. The only time variables become dependent on each other is if the calculation when one competitor plays another is performed, thus breaking the model above.

Conclusion + Extra Thoughts:

From this report on Super Smash Bros. Melee top competitor set history, varying points of data analytics could be performed to obtain a better understanding of how certain players' win percentages and match history lie compared to others. This data set was mainly a discrete data set, but still managed to translate well into the later half of the report, when continuous variables and calculus functions were introduced. Overall, performing calculations such as these when looking at statistical and historical competition data could allow someone to give themselves a competitive advantage when practicing and competing against other competitors.