STA 440 Final Project

Dhruv Dadhania

April 2021

1 Introduction

Poker is well known as a game of imperfect information. As a result, it has received considerable attention in fields related to artificial intelligence [5]. Models for games such as chess, checkers and Scrabble can often win against top human players by continuously making game theory optimal (GTO) decisions. In poker, unlike many similar games, GTO play does not guarantee maximal reward [5]. Player modeling is necessary to determine the frequencies individual players perform certain actions. However, modelling player behavior is an extremely difficult challenge.

Many different attempts have been made to address this problem, including binning opponent play styles [2], using neural networks [3] and Bayesian approaches [4]. Despite this, the problem is largely unsolved, with many of these approaches having weak results in true game play. Given a multitude of different decisions occur in each poker match, and it would be infeasible to model each, a few different specific questions are chosen below.

1.1 Understanding Poker

Poker is a card game generally focused on producing the best five card hand. In this case, "best" is defined according to a ranking system that is the standard across many different variations of the game. In this particular analysis, we focus on the variation known as Texas Hold 'em. In this variation, players are provided two cards that are unique to them and hidden from all other players (known as the pocket cards). Four rounds of betting occur: pre-flop, post-flop, post-turn, and post-river. The flop refers to the first three community cards which any player may use to create their best five card hand. The turn and the river are the fourth and fifth community cards, respectively. During a round of betting, players choose to fold, check/call or raise. A fold indicates that players do not wish to continue playing that particular hand, and are not required to put in any more money for this hand. A check or call forces the player to put the minimum amount of money to continue playing. A raise increases the amount necessary to continue for all players at the table. If a player does not fold before the flop, they are said to have opened their hand. At the conclusion of the last round of betting, if at least two players remain, a showdown occurs where players show their cards and the player with the best five card hand wins all the money that was bet in that round (the pot). Many factors influence betting decisions. One factor is position. Given that poker is a game of imperfect information, the order in which players act plays a critical role. Making bet decisions after other players allows players to incorporate more information, the decisions of those who acted before them. Position refers to the order in which players act, with 1 being the best position, also known as the dealer. Another important concept is hand strength. Before the flop is shown, it is impossible to know which hand will win, but some hands are more likely than others. Hand strength refers to a rating system for the strength of the pocket cards a player has, before any community cards are shown. Greater information about the game of Texas Hold 'em is readily available online and in past literature [2, 3].

1.2 Data

The University of Alberta maintained a dataset of Internet Relay Chat (IRC) poker games in the late 90s. This dataset, containing more than 10 million hands, records all actions taken in a specific game (betting choices, community cards and cards shown during the showdown, for example). Additionally, player usernames are consistent across the game, allowing for player modeling of individuals. Making use of publicly available data cleaning scripts[1], we can simplify this dataset and extract only the Hold 'em hands that have complete information (Hold 'em hands with all features provided). This dataset is referred to as the condensed dataset throughout this analysis. Depending on the research task, complete information may not be necessary (e.g. if we are interested in the length of time a player stays in the game we do not need to know their pocket cards), and a modification of these scripts can be used to extract all Hold 'em hands instead. Throughout this work, such a dataset is referred to as the full dataset. The choice of using exclusively Hold 'em hands is somewhat arbitrary. However, the decision to down sample in this way was made because: (1) different poker variants have different rules, often making player modeling across all games unfeasible, (2) Hold 'em is the most common variant in the dataset, and (3) this choice is consistent with prior literature [2, 3].

Data on player opening hand strength was also included. Hands were binned into one of four tiers. Tier 1 corresponds to hands that should be opened in all positions, Tier 2 to those which should be opened mid or late, Tier 3 for those only to be opened late, and Tier 4 for hands that are not recommended to be opened with. The decisions of what cards go in what tier are fairly similar between different sources, though, as hard cutoffs were necessary, a single source had to be chosen[2].

1.3 Research Question / Hypothesis

The task of player modeling is a broad space. In this analysis, we focus on three related questions that help capture player decisions.

- Do individual players have significant and capturable differences in their likelihood of raising pre-flop?

  We are interested in seeing if we can model differences in the probability of raising before the flop occurs when controlling for the strength of the players hand and position in the condensed dataset. Doing so is an important aspect of player modeling to determine differences in play style between individuals. We hypothesize that the differences between players is a capturable quantity.

- Does the range of hands that players reach the showdown with vary based on experience?

  When a player reaches the showdown, they are likely to believe their hand will win the game. We want to compare the hand strength of individuals who have logged a large number of hands in these IRC games to those who have not in this condensed dataset. Doing so would allow us to better understand if future player modeling algorithms should consider experience, and how experience impacts hand strength. We hypothesize that more experienced players will reach the showdown with better cards than less experienced players.

- Do players change their play style based on past performance?

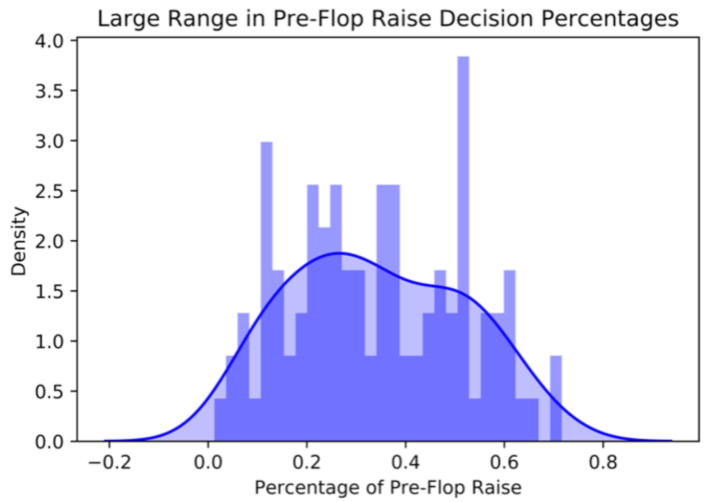
  Modeling play style is difficult enough when we consider a static player (one who never changes their decision rules). However, true players are more complex and we aim to capture some of this complexity by determining if the length of time a player stays in on a specific hand is influenced by their success in the previous hand, given the full dataset. Interestingly, a similar analysis appears to have been performed in the past, on a different data set [1]. We hypothesize that players who had large wins in previous hands will be more likely to play out their next hand.

[1] https://github.com/allenfrostline/PokerHandsDataset
[2] https://www.cardschat.com/poker-starting-hands.php

1.4 Exploratory Data Analysis

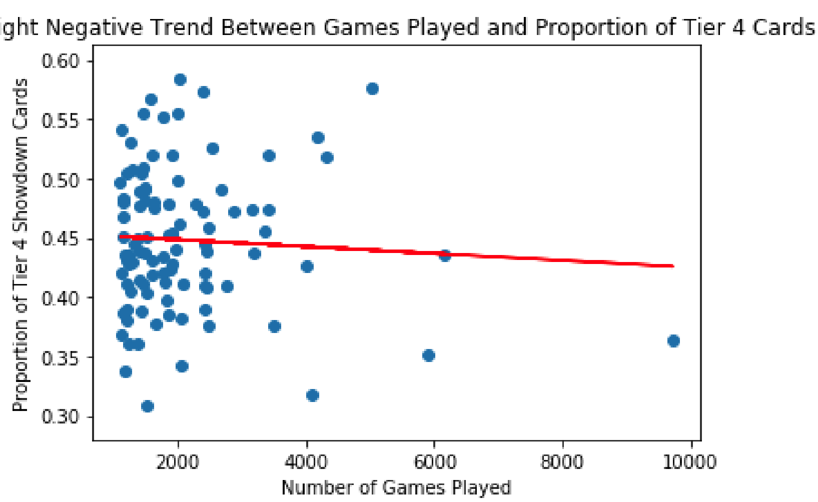
1.4.1 Pre-Flop Raise

As a preliminary investigation into the difference between players' raise likelihood pre-flop, we investigate the frequency that players perform this action in the histogram below. A kernel density estimate is included to demonstrate the relative uniformity of raise percentages. As shown, players have drastically different raise percentages, and we hypothesize that knowing which player is involved in a decision may be a significant element to a model that predicts if a pre-flop raise occurs.



1.4.2 Showdown Cards

Similar preliminary analysis of the showdown card question demonstrates a slight negative relationship between the number of games played and proportion of Tier 4 cards brought to the showdown. This indicates that a greater number of games played is associated with a decreased likelihood of bringing the worst cards to showdown. Each data point in the scatter plot represents a single player. This relationship is marginal, and consistent with the hypothesis proposed prior to performing exploratory data analysis. A rigorous comparison may better determine if a relationship exists and if, controlling for other factors, this negative relationship is maintained.



1.4.3 Past Performance

When considering the impact of past performance, a survival analysis framework is considered. In such a framework, the event of interest is a fold. Censoring occurs when either the showdown occurs or a player wins the hand before the showdown occurs (indicating that the true length of time they would have stayed in the hand is unknown). The following table represents the percent of players that have not folded after a win and after a loss.

	Pre-Flop	Post-Flop	Post-Turn	Post-River
Post Win	85%	82%	80%	77%
Post Loss	67%	64%	61%	59%

Though a more rigorous analysis is necessary, to consider the effects of potentially impactful external factors (such as position and stack size), the preliminary results clearly show that observations after a win are associated with decreased folding likelihoods.

2 Methodology

A variety of factors may influence player decision in poker. To encapsulate this, a number of variables are controlled for in this report's analysis. These variables include: strength of hand, player position and stack size. The strength of a player's hand was evaluated based on the criteria discussed in the Data section of the Introduction (chosen to be tier 1-4). Better hands are, unsurprisingly, more likely to win, and players are expected to raise and play more often with these hands. Another variable that was controlled for is the player's position which refers to their location at the table, relative the dealer. Two players in the game are required to put in a bet, regardless of their cards (the small and big blind, players next to the dealer). Further, players who act last in a hand know the decisions of the other players, prior to making their decision. As a result, players are likely to play differently based on their position at the table. In this analysis, the stack size was also considered for its effect on game play. Stack size refers to the amount of money that a player has remaining at the point when they started the hand. In this analysis, the stack size variable represents tens of thousands of "dollars". The money has no real value equivalent but is referenced as "dollars" throughout gameplay. The scaling factor was chosen as the average amount of "dollars" is slightly above seventy thousand. Players may change their play style based on how much money they have available to them (e.g. they may call more bets when they have a larger stack).

Other variables were also considered, including the relative card strength of opponents at the table and the interaction between stack size and position. The former was ruled out as the data was insufficient to create this variable - only players who reach the showdown have their pocket cards revealed in the dataset. The latter was also ruled out as a deeper investigation indicated that this relationship would not be meaningful. Optimal poker strategy provides no justification for such an interaction effect, as players operating under this strategy are meant to only consider the the raw value of their position or stack size, and not the interaction between them. We suspect that the poker enthusiasts captured in this dataset would both know and execute optimal poker strategy. However, an interesting extension to this work may investigate if optimal strategy is followed, or if variables such as stack size impact decisions that are meant to be purely based on position, or vice versa.

2.1 Pre-Flop Raise

To determine if there are statistically significant differences in pre-flop raise probabilities between players, a logistic model is fit with player, strength of hand, position, and stack size. The model is specified below:

$$logit(p_i) = \beta_0 + \beta_1 I(HandStrength_i = 2) + \beta_2 I(HandStrength_i = 3) + \beta_3 I(HandStrength_i = 3)$$

$$+\beta_4 Position_i + \beta_5 StackSize_i + \beta_k player_{ik}$$

In this formulation, i indexes each hand in the dataset. Here, p_i refers to the probability that a raise occurred pre-flop. The β terms represent the coefficients associated with a given variable. $player_{ik}$ is an indicator that represents if the ith observation comes from player k for k ranging from 1 to the total number of players in the dataset. Player and hand strength are interpreted as categorical variables. Stack size and position are interpreted as continuous variables. A hand strength of 1 is treated as the baseline. Stack size is naturally a continuous variable, but this treatment of position requires some justification. We believe that treating this variable as continuous allows us to capture the ordinal nature of the variable. Additionally, we believe that there is an approximately linear effect from position. The difference between any two positions at the table is precisely equal to the number of people between the two players at these positions. As a result, we suspect that there will be an approximately linear "position effect" and treating this variable as continuous is justified. The choice of a logistic model is a natural choice for modeling the binary decision of if a raise occurred. Other binary models are feasible in this situation, though less common in the poker literature.

To determine if the addition of player is significant, a likelihood ratio test is performed to compare the model above with a model that does not include the player variable. The test is performed at the 0.05 alpha level. Model diagnostics and assumption verification - including residual plots, independence discussion and multicollinearity - are included in the Appendix.

2.2 Showdown Cards

As with all evaluations of player hand strength in this analysis, the strength of hands at the showdown is evaluated on a 1-4 scale, with 1 being the best. An ordinal model is considered. However, an ordinal requires the proportional odds assumption. The data violate this assumption, as the relationships between different categories are not always equivalent. This problem is well defined by the stack size variable. In such a model, the probability of playing better cards are likely to have little to no relationship with stack size (good hands are always recommended to be played). However, worse hands (Tier 4 cards) may have a much more pronounced relationship with stack size as players will only play these hands they are not worried about going bankrupt. A multinomial logistic regression model is chosen instead to compensate for this issue. The variables considered in this model are position, stack size and amount of games played. The model is specified below:

$$log(\frac{P(Tier1)}{P(Tier4)}) = \beta_{10} + \beta_{11}Position_i + \beta_{12}StackSize_i + \beta_{13}GamesPlayed_i$$

$$log(\frac{P(Tier2)}{P(Tier4)}) = \beta_{20} + \beta_{21}Position_i + \beta_{22}StackSize_i + \beta_{23}GamesPlayed_i$$

$$log(\frac{P(Tier3)}{P(Tier4)}) = \beta_{30} + \beta_{31}Position_i + \beta_{32}StackSize_i + \beta_{33}GamesPlayed_i$$

In this model, we are evaluating the probability of reaching the showdown with a specific tier hand relative the baseline of Tier 4 cards. Position and stack size are defined as previously, as is i and β. However, the β terms in this model are indexed both by their relative position in the model and the tier that they are predicting (meaning 15 different β terms are estimated in this model). The number of games played is a continuous variable that reflects the total number of games that the player has played, in the dataset, up until this point. The number of games played is used as a representation for the frequency that the player plays poker. To determine if a player who plays more often reaches the showdown with different cards, a model without games played is also created. The two models are compared through a likelihood ratio test at the 0.05 alpha level. Discussion of assumptions (notably independence and multicollinearity) are available in the Appendix.

2.3 Past Performance

To determine if there is an association between past performance and how a player plays the next hand, a discrete time survival analysis framework is incorporated. The event being modeled is when the player chooses to fold. As previously noted, censoring occurs when the showdown occurs or if everyone else at the table folds before the showdown, indicating that the true length of time that the player of interest would play, before folding, is unknown. First, a binary variable is created that indicates if a player won the last hand. To quantify how long a player stays in on a hand, four time steps are considered: pre-flop, post-flop, post-turn, and post-river. A hand-period data set (similar to person-period data) is created where each row represents one of these time steps and the event corresponds to if a fold occurred. Both stack size and position are incorporated as variables to control for. The final model uses the model formulation put forth in [6]. The model is written as:

$$logit(h_{ij}) = (\alpha_1 D_{1j} + \alpha_2 D_{2j} + \alpha_3 D_{3j} + \alpha_4 D_{4j}) + (\beta_1 StackSize_i + \beta_2 Position_i + \beta_3 PriorWin_i)$$

In this model, i corresponds to an individual player and j corresponds to a time period (ranging from 1 to 4 as discussed above). α refers to the weights for a given time step, while β maintains its meaning as the coefficient for a particular variable. D refers to different time steps in the model. h is the hazard at a specific time for a specific hand. Other survival models were considered, including the more common Cox model, however the discrete nature of the lifetimes requires a discrete time survival model.

A likelihood ratio test is again conducted at the 0.05 alpha level to determine the significance of PriorWin. Model diagnostics (including independence and multicollinearity) are discussed in the Appendix.

3 Results

Results for each of the research questions are demonstrated below. Further analysis is provided in the discussion section.

3.1 Pre-Flop Raise

After fitting a logistic model to the data, the following coefficient estimates, p-values, and 95% confidence intervals were produced. The estimates provided are coefficient estimates for the logit of the probability of a pre-flop raise occurring.

Name	Estimate	P-Value	95% CI - Low	95% CI - High
Intercept	1.485	<0.001	1.453	1.516
Tier 2 Hand Strength	-0.974	<0.001	-1.003	-0.945
Tier 3 Hand Strength	-1.225	<0.001	-1.255	-1.195
Tier 4 Hand Strength	-1.812	<0.001	-1.837	-1.788
Position	-0.7524	<0.001	-0.770	-0.735
Stack Size	0.004	0.063	0.000	0.008

An additional model is fit with the player variable and results are produced below. Player coefficient estimates were not included in the below model output in the interest of space.

Name	Estimate	P-Value	95% CI - Low	95% CI - High
Intercept	2.428	<0.001	2.336	2.519
Tier 2 Hand Strength	-1.114	<0.001	-1.147	-1.082
Tier 3 Hand Strength	-1.423	<0.001	-1.455	-1.390
Tier 4 Hand Strength	-2.077	<0.001	-2.105	-2.049
Position	-0.892	<0.001	-0.911	-0.872
Stack Size	0.007	0.010	0.002	0.012

The likelihood ratio test comparing a model without player variables and with player variables yielded a p-value of <0.001, statistically significant at the 0.05 level. Verification of model assumptions - including independence discussion, multicollinearity tests, and residuals plots - are available in the Appendix. No major assumption violations were noted.

3.2 Showdown Cards

After fitting a multinomial logistic regression model, without number of games played, the following output was created. The estimates represent coefficient estimates for the log of the probability of a particular card tier occurring over the baseline, tier 4.

Tier 1 vs baseline (Tier 4):

Tier (Baseline 4)	Name	Estimate	P-Value	95% CI - Low	95% CI - High
1	Intercept	-0.412	<0.001	-0.537	-0.458
	Position	-0.041	0.001	0.017	0.065
	Stack Size	0.009	0.014	0.002	0.017

Tier 2 vs. baseline (Tier 4):

Tier (Baseline 4)	Name	Estimate	P-Value	95% CI - Low	95% CI - High
2	Intercept	-0.454	<0.001	-0.494	-0.416
	Position	-0.020	0.108	-0.004	0.043
	Stack Size	0.012	0.001	0.005	0.019

Tier 3 vs baseline (Tier 4):

Tier (Baseline 4)	Name	Estimate	P-Value	95% CI - Low	95% CI - High
3	Intercept	0.228	<0.001	0.197	0.259
	Position	0.245	<0.001	0.227	0.263
	Stack Size	0.007	0.025	0.000	0.014

Adding in a variable for the number of games played produces the following results:

Tier 1 vs baseline (Tier 4):

Tier (Baseline 4)	Name	Estimate	P-Value	95% CI - Low	95% CI - High
1	Intercept	-0.459	<0.001	-0.503	-0.415
	Position	-0.042	0.001	0.018	0.066
	Stack Size	0.013	0.001	0.005	0.021
	Games Played	-1.414×10^{-5}	<0.001	-2.100×10^{-5}	-7.286×10^{-6}

Tier 2 vs. baseline (Tier 4):

Tier (Baseline 4)	Name	Estimate	P-Value	95% CI - Low	95% CI - High
2	Intercept	-0.403	<0.001	-0.447	-0.359
	Position	0.020	0.096	-0.004	0.044
	Stack Size	0.016	0.001	0.009	0.023
	Games Played	-1.929×10^{-5}	<0.001	-2.617×10^{-5}	-1.242×10^{-5}

Tier 3 vs baseline (Tier 4):

Tier (Baseline 4)	Name	Estimate	P-Value	95% CI - Low	95% CI - High
3	Intercept	0.338	<0.001	0.304	0.372
	Position	0.246	<0.001	0.228	0.265
	Stack Size	0.013	<0.001	0.006	0.020
	Games Played	-4.030×10^{-5}	<0.001	-4.573×10^{-5}	-3.487×10^{-5}

The likelihood ratio test comparing the two models yielded a p-value of <0.001, statistically significant at the 0.05 level, unsurprising given the statistical significance of games played in each tier. Verification of model assumptions - including independence discussion and multicollinearity tests - are available in the Appendix. No major assumption violations were noted.

3.3 Past Performance

The discrete time survival analysis model was first fit without the boolean variable indicating if the player won the last game. It had the following model output. The estimates in question correspond to coefficients representing the logit of hazard.

Name	Estimate	P-Value	95% CI - Low	95% CI - High
Intercept	-2.136	<0.001	-2.161	-2.110
Position	0.430	<0.001	0.425	0.435
Stack Size	-0.156	<0.001	-0.176	-0.138
D_1	-2.013	<0.001	-2.049	-1.976
D_2	-2.534	<0.001	-2.581	-2.490
D_3	-2.601	<0.001	-2.649	-2.552

After fitting the same model with a boolean variable indicating if the player won the last game, the following model output was generated:

Name	Estimate	P-Value	95% CI - Low	95% CI - High
Intercept	-2.023	<0.001	-2.050	-1.996
Position	0.419	<0.001	0.414	0.425
Stack Size	-0.140	<0.001	-0.159	-0.122
Prior Win	-0.467	<0.001	-0.501	-0.433
D_1	-2.001	<0.001	-2.038	-1.965
D_2	-2.525	<0.001	-2.570	-2.479
D_3	-2.591	<0.001	-2.639	-2.543

A likelihood ratio test comparing the two models yielded a p-value of <0.001, statistically significant at the 0.05 level, also unsurprising given the statistical significance of prior win. Verification of independence and multicollinearity tests are available in the Appendix. Some patterning of residuals appeared to exist, but we maintain that the underlying assumptions are not completely violated. We do, however, recommend future analyses consider alternative model specifications to verify that the results of this analysis do not change with model choice, and are consistent in models with more optimal residuals.

4 Discussion

4.1 Pre-Flop Raise

When interpreting the results of the pre-flop raise analysis, the most important aspect is the significance of adding the player variables to the model at a 0.05 level. A model that included individual coefficient estimates for each player in the dataset was found to have a statistically significant difference from a model without, through a likelihood ratio test. Interpretation of these variables is difficult, because the dataset had many players, and because these coefficient estimates are based on an arbitrary baseline player. However, this significance provides evidence that players have significant and capturable differences in their likelihood of raising pre-flop. Of interest, the control variables in our analysis were almost entirely statistically significant at the 0.05 level, with the exception of stack size in the model without individual player coefficients. The hand strength variable in both models was found to have a negative coefficient for each tier, compared to a tier 1 baseline, indicating that non-tier 1 pocket cards are associated with a lower pre-flop raise probability, in general. These variables have increasingly large negative coefficients, for tier 2, 3, and 4. These results match intuition as players are more likely to raise with cards that are more likely to win, and we suspect that the probability of raising is positively associated with the probability of winning. The coefficient on position is similarly negative in both models, indicating that the estimated probability of a pre-flop raise occurring decreases as position increases (or, equivalently, as position gets worse). This is also intuitive as players in worse positions are less likely to win, and are therefore recommended to avoid raising. The stack size variable had a positive coefficient in both models. The coefficient was interestingly insignificant in the model without players. The positive coefficient indicates that the estimated probability of a pre-flop raise occurring increases with stack size, which we might suspect as players with few available chips are unlikely to perform a pre-flop raise. The results of this analysis are meaningful as they provide evidence that: (1) the difference between players probability of raising pre-flop is significantly different and be captured in a logistic regression model, when accounting for other factors, and (2) the associations between control variables and the probability of raising pre-flop in this real world dataset are consistent with intuition.

4.2 Showdown Cards

With respect to the showdown cards analysis, the model trained with games played had a statistically significant difference from the model trained without it, and the games played coefficient in each tier was found to be significant at the 0.05 level. This provides evidence that the cards that players reach the showdown with is associated with games played, a proxy for experience, when controlling for other factors. Interestingly, the games played coefficient was negative for each tier, relative to the baseline of tier 4. This indicates that a larger number of games played is associated with a decreased log probability of any tier divided by the probability of the tier 4 baseline. Equivalently, higher games played is associated with a decreased logit of any tier relative to the tier 4 baseline, holding all else constant. Additionally, the magnitude of this negative value increases with tier, as tier 1 cards have the smallest coefficient estimate on games played and tier 3 have the largest. We hypothesized that more experienced players, players who had logged more games played, would reach the showdown with better cards. However, the opposite associated appears to be present, potentially indicating that increased number of games played is associated with a "looser" play style. Once again, stack size had a significant and positive coefficient for each tier. This indicates that the logit of any tier cards relative to the baseline tier 4 increases as stack size increases. This may indicate that players are more likely to bluff when they have a small stack, but further investigation is necessary to determine if this trend is evident. The position variable is interestingly negative for tier 1 cards, becomes insignificant for tier 2 cards and is positive for tier 3 cards. This is a particularly interesting trend considering the baseline is tier 4 and we suspected that a more linear trend would be present. The negative coefficient for tier 1, relative to baseline, indicates that, holding all else constant, the logit of tier 1 relative to the baseline decreases as position increases (or, equivalently, as position gets worse). The opposite trend is true for tier 3. Further investigation is necessary to determine the cause for such a change. The results of this analysis are meaningful as they provide evidence that experience is an important element of player modeling, and this association may be important to include in future models.

4.3 Past Performance

The analysis of past performance demonstrated a statistically significant difference at the 0.05 level between a discrete time survival analysis model trained with a boolean variable indicating if a prior win occurred and one trained without it. Additionally, this boolean variable was significant at the same level. This indicates that if a player won the last hand has a significant association with the length of time that they remain in the next hand. The negative coefficient on this variable indicates that having a prior win was associated with a decreased logit of hazard, and therefore hazard. Such a finding is consistent with the literature presented in the introduction, despite using a different model and dataset. The key takeaway is that a player winning the last hand is associated with a decreased hazard, where the event of interest is a fold. This has practical applications as it provides evidence for a, potentially exploitable, association. If players are less likely to fold after winning, other players who are aware of this could take this into account and update their own decisions accordingly. Further analysis is necessary to determine if there is a causal relationship at play, and if this relationship can be well exploited. The other variables in this analysis were shown to have the expected direction, with hazard increasing with position and decreasing with stack size. We expect that worse positions will have higher hazards given that the event of interest is a fold and these positions are far more likely to fold at each time point. We also suspect that players with a larger amount of money will be less likely to fold, as they see less "risk" in each bet. That is, each bet they have to call represents a smaller portion of their total money and they have the ability to play more loose.

4.4 Limitation and Next Steps

Though this analysis has many meaningful and important takeaways, these takeaways are somewhat limited and many opportunities for further research exist. The primary limitation of this analysis is the age of the data. Though this dataset remains a standard, even within the literature, professional poker play has changed drastically in the past decade. Though the rules themselves have not changed, the decisions players make, in general, and the factors that are incorporated into these decisions have changed. As a result, the takeaways from this analysis may not be applicable to modern gameplay. Further investigation is necessary to determine if these results generalize, though modern data, at equivalent size to this dataset, is extremely difficult to find.

Another limitation of this work is the variety of players represented in this dataset. In general, analysis of poker gameplay tends to focus on professional play. The underlying IRC poker games were not limited based on skill level, and it is possible that novice and intermediate players are also represented. The players in this dataset are almost entirely anonymous, though some professional players have come forward to claim their anonymous handle - indicating that at least some subset of the dataset represents professional players. Depending on the intended application of the results of this analysis, the diversity of underlying players may or may not be beneficial. These results are likely more generalizable to the broad class of poker players, of all skill levels, than they are to exclusively professional players.

Many opportunities for extensions to this work exist. The two main ones are: (1) considering more modern data, and (2) considering a broader number of questions. The benefits of modern data have been discussed, but amassing such data is incredibly difficult. We suspect that professional poker players, organizations, or casinos may have access to this data and may be interested in investigating if these associations and trends are maintained. On the second point, the research questions discussed in this analysis have only scratched the surface of all of the questions that may be interesting in this field. Many more questions are necessary for the development of poker automation than can be reasonably addressed in any one analysis. We hope that this analysis acts as a starting point for future work.

4.5 Conclusion

Player modeling in poker is an extremely challenging problem, but one with clear benefits both for professional poker play and for the advancement of poker automation. Breaking up this complex space into smaller, investigatable sub-questions has provided interesting and useful relationships. The results of this analysis demonstrate that clear, and potentially actionable, associations are present in the play data from this specific dataset. As extensions to this work are created, as previously discussed, we hope to see these conclusions extended and meaningfully applied.

References

[1] David Eil and Jaimie W. Lien. Staying Ahead and Getting Even: Risk Attitudes of Experienced Poker Players. *Games and Economic Behavior*, 2014.

[2] Billings et al. Opponent Modeling in Poker. *AAAI-98 Proceedings*, 1998.

[3] Davidson et al. Improved Opponent Modeling in Poker. *AAAI Press*, 2000.

[4] Southey et al. Bayes' Bluff: Opponent Modelling in Poker. *UAI*, 2005.

[5] Jonathan Rubin and Ian Watson. Computer Poker: A Review. *Artificial Intelligence*, 2011.

[6] Judith D. Singler and John B. Willet. It's about Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events. *Journal of Educational Statistics*, 1993.

5 Appendix

5.1 Pre-Flop Raise Model Assumption Verification

5.1.1 Independence

Pre-flop raise decisions are independent of one another. It is unlikely that any one observation in the dataset meaningfully impacts the other observations. As a result, we determine that the independence assumption is not violated.

5.1.2 Multicollinearity

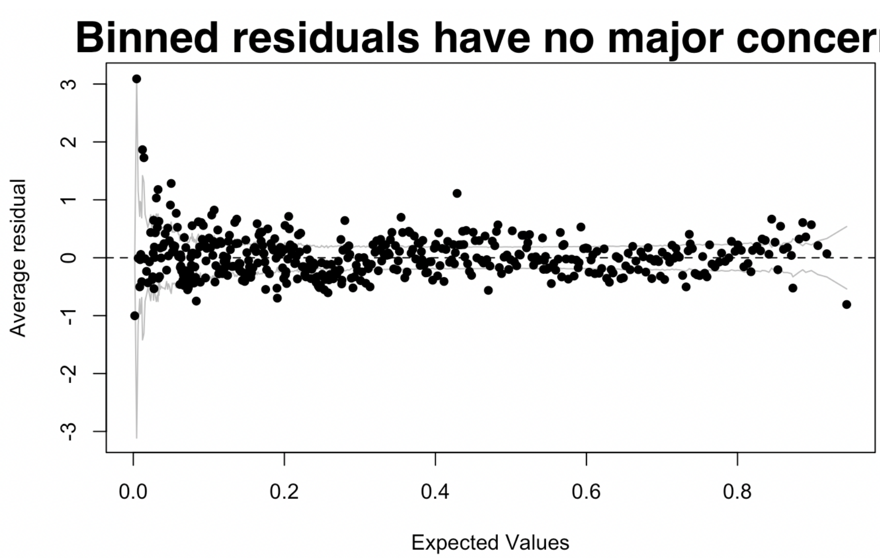
Variance inflation factor values are demonstrated below for each of the non-player variables. The player values did not exceed any reasonable benchmark of detecting multicollinearity (<2).

Name	VIF
Tier 2 Hand Strength	1.662
Tier 3 Hand Strength	1.651
Tier 4 Hand Strength	2.491
Position	8.233
Stack Size	1.321

Though the VIF for position is rather high (8.233), it is lower than our benchmark of 10 and, consequently, we do not believe that there is strong evidence of multicollinearity.

5.1.3 Residual Plots

Residual plots for the generated model (with player) are included below:



One residual was removed from the binned residual plot. This was done as the size of the residual was too large and the resulting graph was unable to differentiate between any other points. Despite the size of this residual, the leverage of this point was low enough that no further action was determined to be necessary. Overall, no major assumption violations appear to be present.

5.2 Showdown Cards Model Assumption Verification

5.2.1 Independence

Showdown cards are independent of one another. Players do not know the showdown cards of other players until the end of the hand so the cards are independent in this sense. There is a slight relation to each other, as one player having one card makes it impossible for another player to have that same card, but this is believed to have only a minor impact. As a result, we determine that the independence assumption is not violated.

5.2.2 Multicollinearity

Variance inflation factor values are demonstrated below for each of the variables.

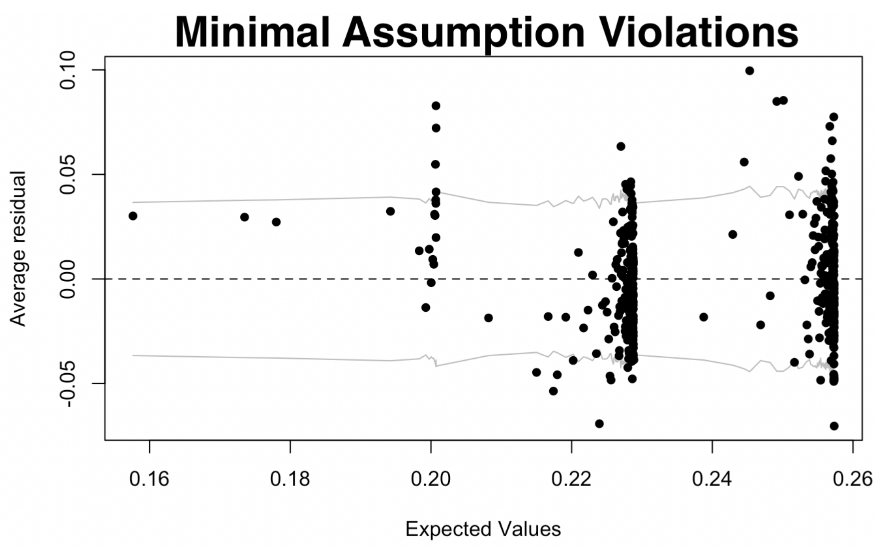
Name	VIF
Position	2.453
Stack Size	1.017
Games Played	2.453

There is no evidence of multicollinearity.

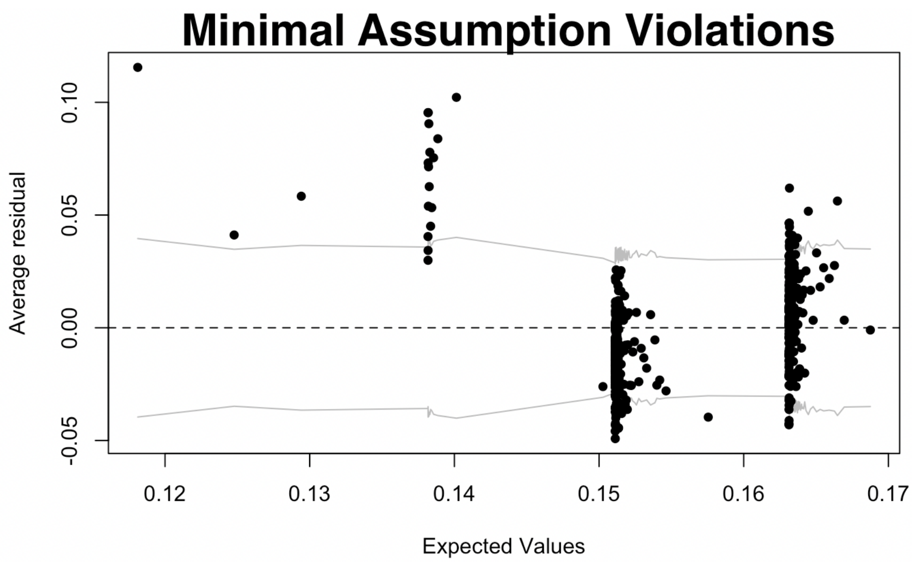
5.3 Residual Plots

A separate binned residual plot was created for each tier, relative to the tier 4 baseline, for the model with games played. Though the binned residuals are not ideal, we believe that no major assumption violations are noted in these residuals.

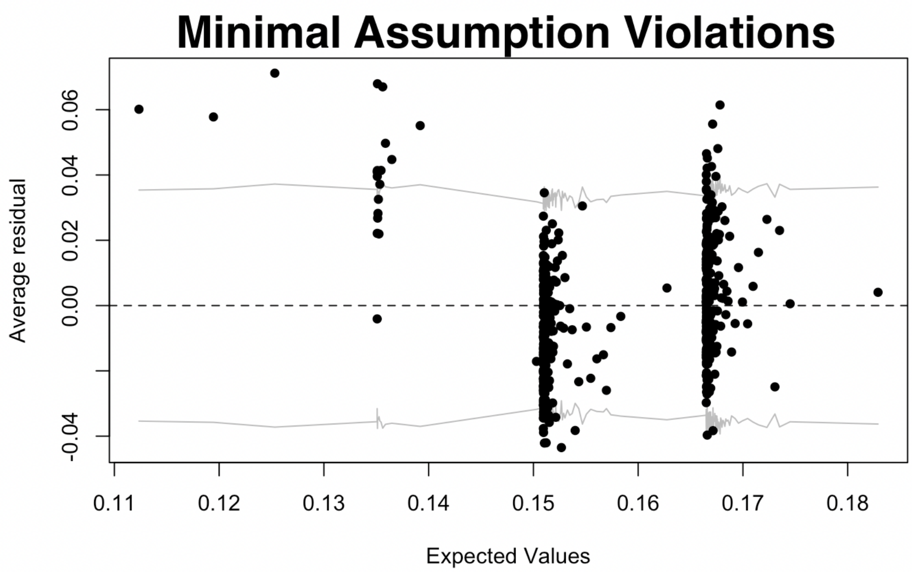
5.3.1 Tier 1



5.3.2 Tier 2



5.3.3 Tier 3



5.4 Past Performance Assumption Verification

5.4.1 Independence

Similarly to prior discussion, we expect that the time until a fold between players is independent between observations. There is some dependence, as if everyone else at the table folds the hand in question is counted as a censored observation, but this does not substantially violate our assumption. As a result, we do not believe there are major independence violations.

5.4.2 Multicollinearity

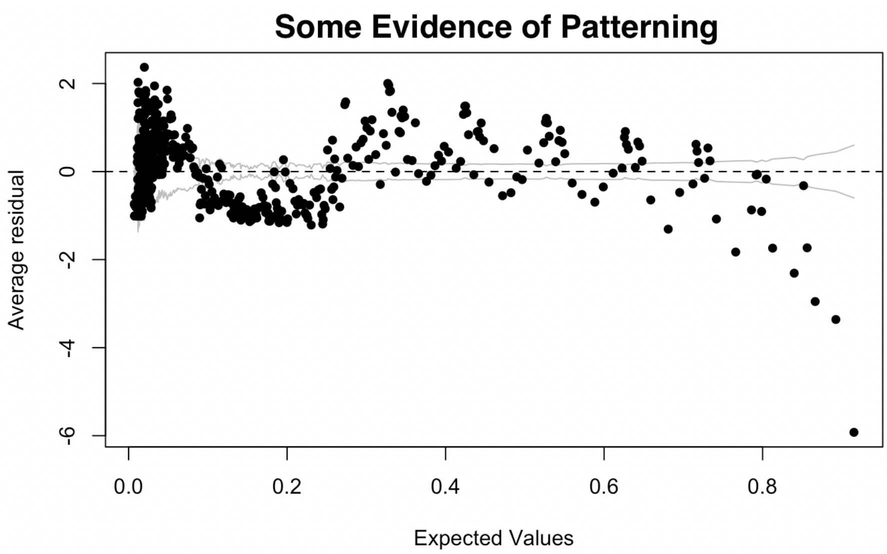
Variance inflation factor values are demonstrated below for each of the variables.

Name	VIF
Position	1.570
Stack Size	1.354
Prior Win	1.253
D_1	1.272
D_2	1.243
D_3	1.226

There is no evidence of multicollinearity.

5.4.3 Residual Plots

Residual plots for the generated model (with prior win) are included below:



Unfortunately, there appears to be some evidence of patterning. We believe this may indicate an assumption violation or at minimum a cause for concern. In this analysis, we treat it as a cause for concern and recommend that future analysis extend this work by considering alternative discrete time survival models, especially those with potentially less evidence of assumption violation.