Understanding Gambling Behavior and Risk Attitudes Using Massive Online Casino Data

DARTMOUTH

Jonathan Meng

Honors Thesis

Feng Fu, Advisor

Mathematics
Dartmouth College

April 24, 2018

Abstract

The statistical concept of Gamblers Ruin suggests that gambling has a large amount of risk. Nevertheless, gambling at casinos and gambling on the Internet are both hugely popular activities. In recent years, both prospect theory and lab-controlled experiments have been used to improve our understanding of risk attitudes associated with gambling. Despite theoretical progress, collecting real-life gambling data, which is essential to validate predictions and experimental findings, remains a challenge. To address this issue, we explore large amounts of publicly available online casino data collected from a customized web scraper. Next, we wish to analyze the dataset through the lens of current probability-theoretic models and discover empirical examples of gambling systems. For this purpose, we collected betting data from a DApp (decentralized application) on the Ethereum Blockchain, https://etheroll.com, which instantly publishes the outcome of every single bet (consisting of each bets Timestamp, Wager, Probability of Winning, UserID, and Profit). This data, which allows gamblers to tune their own probabilities, is well suited for studying gambling strategies and the complex dynamic of risk attitudes involved in betting decisions. These conclusions are of great interest to various entities, such as governments or casinos.

Contents

List of Figures

Acknowledgements

First, I would like to thank my Advisor, Feng Fu for being supportive and guiding me through this exploration of data. Next, I would like to thank my parents for supporting me through absolutely anything. I would also like to thank Herbert Cheng for providing an amazing LaTeX template and giving me great advice, Kai Yan and Alma Wang for proofreading and organization suggestions, and Tianhang Dong and James Detweiler for critiquing and fueling my ideas through insightful conversations. Lastly, I would like to thank Barry Yang for providing suggestions on aesthetics and Oliver Cheng for encouraging me even when it got hard.

Chapter 1

Introduction

1.1 Ethereum

The client server model, the most widely used computing network model in the world today, allows devices (clients) to request services or resources from other devices (servers). The client initiates a request to the server and receives a response, which usually gives the client the service or resource it requested. Some examples of this are the World Wide Web, or email. A major issue with this model is that if the server stops working, everything else also ceases functioning. Additionally, if hackers manage to break into the server, they could steal any client information (e.g. Social Security Numbers, Credit Card information) stored inside. This model inherently leads to centralization of computing power towards larger entities, such as government or multinational corporations. [1]

In contrast, a peer-to-peer network lets any of its members (nodes) share information or services on the network. All nodes have equal privilege, which means any node in the network can give another node in the network a desired resource or service. The most famous example of peer-to-peer networking is in torrenting, where an initial server, called a seed, uploads a file. Nodes of the torrent network (the swarm) divide up this file into pieces and request missing pieces from other computers in the network. Once pieces are obtained by a client, or downloading node, the pieces are constructed into the original file. In this way, computing power is not monopolized; it is shared. [2] This model is both fault-tolerant (i.e continues to work even if a single or multiple members fails), and decentralized.

Ethereum is a distributed, peer-to-peer computing network, released in 2015, that allows its nodes to conduct transactions and build applications. On the Ethereum network, the main currency, Ether, powers all peer-to-peer transactions for goods and services. [3]

One of the most important features of Ethereum is its usage of blockchain technology. The blockchain is a decentralized, publicly available chain of transactions. Anyone can download software (Geth, Parity) and turn their computer into a node, or a member of the Ethereum network. The peer-to-peer nature of the network allows computing power to be evenly distributed and accessible. Because all nodes contain a copy of the blockchain, each node has access to the same information. All nodes retain perfect information and verify transactions. Through the usage of blockchain technology, Ethereum aims to shift the current paradigm of computing from the client-server model to a decentralized, peer-to-peer model.

All nodes verify transactions in order to ensure that new transactions are not fraudulent. Once enough transactions are verified, these transactions are packaged together into a block. Certain nodes, called miners, then compete to compute a difficult cryptographic hashing problem, called ETHhash. [4] This system, which rewards miners for work done is referred to as a Proof of Work System. Once a miner solves the problem, the mined block is then added to the blockchain.

After successfully packaging a block, miners are awarded with currency that is used to pay for transactions, such as Ether, or Bitcoin. On the Ethereum network, this reward is up to 5 Ether. Because each transaction is verified by all the nodes in the network, blockchains are extremely resistant to attempts of fraudulent modification. If an attacker attempts to change the system, he or she would have to generate an alternate chain from scratch. According to the original white paper (specification) of Bitcoin, the block synchronization of these two parties is modeled as a binomial random walk. From this, we see that the effective probability of an attacker succeeding in creating a fraudulent blockchain approaches 0 if the attacker is more than 25 blocks behind the actual blockchain. [5]

Another important feature of blockchain technology is that it allows user to user transactions to be psuedoanonymous. This is due to a hashing of the transaction IDs and their corresponding wallet IDs. This is extremely important, as it allows for transparency of data.[6] Users do not have to worry about exposing their identity to the public.

Ethereum has also introduced the idea of programming blockchain operations through a technology called the smart contract. A smart contract is an automated script written in Ethereum's own scripting language, Solidity, that allows an individual to exchange a specified good or service. A popular comparison for smart contracts is the vending machine. If a user of the smart contract gives the vending machine a certain fee, and a product comes out. Accordingly, if a user inputs some cryptocurrency into a smart contract, it executes an exchange of goods or services. As smart contracts are also automated, they erase the need for a middleman. Smart contracts, if programmed properly, can be used for a variety of applications, such as vote automation or tax collection. Building an application on top of a smart contract creates a decentralized application (DApp). A Dapp is completely decentralized (no single owner) and automated by its associated smart contract. Currently, there are around 1,539 DApps on the Ethereum Blockchain. [7]

1.2 Application

We study the behavioral dynamics of gamblers on a DApp known as Etheroll. Etheroll simulates a virtual dice gambling game where all bets are made in Ether and published on the Ethereum blockchain. Etheroll has an associated smart contract on the Ethereum network which specifies house edges, payouts, and dividends to investors. [8] To begin the dice game, the gambler chooses a number between 2 and 99 (inclusive). The probability that the gambler wins is the number he or she chooses, minus 1, meaning that the gambler can choose between a 1% to 98% chance of winning. The payout (P') formula, if the house commission per bet is $e = 1\%$, probability of winning is p, and initial wager is W is:

$$P' = W \left(\frac{1 - p - e}{p} \right)$$

The smart contract then simulates a hundred-sided dice roll. If the result of the dice roll is any number smaller than the number the gambler chose, the gambler wins. After the transaction between the smart contract and the gambler processes, the gambler receives a payout (in Ether) directly to their Ethereum wallet which is inversely proportional to the probability they bet at. Naturally, lower probabilities of winning have higher payouts, and higher probabilities of winning have lower payouts.

These transactions are publicly available on the Ethereum blockchain. Due to the massive amount of verifying nodes on the Ethereum network, we can be sure about the validity of these transactions. We will explore this data for all four of Etheroll's smart contract updates from April 17th, 2017 to December 12th, 2017. Obtaining real life gambling data, especially data from gambles in casinos is very difficult, if not impossible to obtain. Because of this, mathematical models pertaining to gambling are almost entirely theoretically based. Every bet from Etheroll consists of the bets Timestamp, Wager, Probability of Winning, UserID, and Profit. With this data, we can attempt to empirically explore gambling behavior.

This dataset has many other interesting properties. Having access to timestamps allows us to identify possible changes in strategy influenced by gambling results over time, in their gambling patterns. The fact that gamblers are able to tune their own betting probabilities is also crucial. The ability to tune the effective odds in a wager allows us to evaluate probable risk profiles of certain gamblers. Additionally, we focus on characterizing the entire risk attitudes of the entire "gambling ecosystem" as a whole. We are also able to evaluate the existence and usage of staking gambling systems (path-dependent strategies). The unique completeness and continuity of this data also allows to us empirically evaluate some famous psychological frameworks, such as Cumulative Prospect Theory. We also wish to look at the effect of a gambler's cumulative "signal", or scaled cumulative profit on his probability distributions/strategies. This scaling allows us to model the lessened effect of losses and gains over time. This population of gamblers on the Ethereum blockchain allows us to empirically observe the tendencies of gamblers in a casino-like environment for the first time.

Chapter 2

Prior Models/Background

2.1 Cumulative Prospect Theory

One of the most popular models for evaluating how human beings behave under risk is Amos Tversky and Daniel Kahneman's Cumulative Prospect Theory model. [9] This model is an advancement on their original Prospect Theory model, which was based on a few findings: the framing effect, nonlinear preferences, source dependence, risk seeking behavior, and loss aversion. The "framing effect" is the idea that humans make decisions relative to a reference point, rather than the actual result. This model also incorporates the idea of nonlinear preferences, such that the difference in preferences between $P(0.99)$ vs. $P(1.0)$ is far different from $P(0.10)$ vs. $P(0.11)$. Another crucial aspect of this model is its assumption of source dependence, or the willingness to bet on a uncertain event based on its sourcing (e.g. a preference towards their own areas of expertise). People making decisions under risk also often exhibit a tendency towards risk seeking behavior, such as preferences towards low probability tail events and preferring substantial probabilities of a larger loss over sure losses.

In tandem, loss aversion is exhibited in experiments with losses and gains. Actors in general were found to be more affected by losses and gains, rather than final cumulative profit levels. Due to previous studies, Tversky and Kahneman noticed a distinct asymmetry in the preferences of gamblers towards gains over losses, too significant to be attributed to risk aversion or income effects. [10]

Cumulative Prospect Theory differs from normal prospect theory through its application of probability weighting to the entire cumulative distribution function, rather than individual probabilities. Again, we distinguish two phases in Prospect Theory: framing, and valuation. In the framing phase, the gambler creates the possible outcome space, Ω of the prospect and the actions required. Then, in the valuation section, the gambler assesses the value of the prospects and chooses the most favorable action.

The model that fits these findings involves a gambler that takes some prospect G, probability space $\mathbb{P}$, and outcome space Ω:

$$\mathbb{P} \times \Omega \supset G = \{(x_{-m}, p_{-m}), \ldots, (x_{-1}, p_{-1}), (x_0, p_0), \ldots, (x_1, p_1), \ldots, (x_n, p_n)\}$$

Where we denote each pair (x_i, p_i) as the outcome in which the gambler wins x_i with probability p_i, independent of other outcomes, such that $x_i < x_j$, $\forall i < j$, $x_0 = 0$, and $\sum_{i=-m}^{n} p_i = 1$. We refer to this as the framing phase, where the gambler constructs a representation of the outcomes relevant to his or her decision.

Next, our gambler evaluates the subjective value of G. Our gambler then takes this prospect, and assigns it the cumulative value:

$$V(G) = \sum_{i=-m}^{n} \pi_i v(x_i)$$

Where π_i is the decision weighting function:

$$\pi_i = \begin{cases} w\left(\sum_{k=i}^{n} p_k\right) - w\left(\sum_{k=i+1}^{n} p_k\right), \ \forall i \in [0, n] \\ w\left(\sum_{k=-m}^{i} p_k\right) - w\left(\sum_{k=-m}^{i-1} p_k\right), \ \forall i \in [0, n] \end{cases}$$

and $v : \Omega \to \mathbb{R}$ is some strictly increasing value function that varies between individual to individual, and $w_\pm : \mathbb{P} \to \mathbb{P}$ is a probability weighting function that takes the actual probability p_i of the prospect and maps it as some transformed probability $w_\pm(p_i)$. Let us refer to $w_+(p_i)$ as the probability weighting function for gains, and $w_-(p_i)$ as the probability weighting function for losses. These weighting functions are inverse S-shaped, while the value function is kinked, such that the region of losses is steeper than that of gains. We aim to maximize V.

Tversky and Kahneman then propose mappings:

$$v(x) = \begin{cases} x^\alpha & \text{for } x \geq 0 \\ -\lambda(-x)^\alpha & \text{for } x < -0 \end{cases}$$

$\forall x \in \Omega$, and

$$w_-(p) = \frac{p^{\delta_-}}{(p^\delta_- + (1-p)^\delta_-)^{1/\delta_-}}, \ w_+(p) = \frac{p^{\delta_+}}{(p^\delta_+ + (1-p)^\delta_+)^{1/\delta_+}}$$

Where $\alpha, \delta_\pm \in (0, 1)$, and $\lambda > 1$ are all shape parameters.

These functions conveniently capture all 5 assumptions that encompass the model. Values measured by v are not in terms of final, cumulative wealth, but instead in terms of loss and gains. This model also captures the loss aversion and risk seeking of gamblers through the concavity of v only over gains, compared to losses, where it is instead convex. This implies at medium probability losses, gamblers are risk seeking, but at similar probability gains, gamblers are risk taking. The formulation of v also captures the loss aversion sensitivity due to its kinked nature at its reference point 0, showing that losses affect gamblers more than gains. Lastly, the formulation of w (which is strictly increasing), overweights the extreme outcomes $x_{-m},\ x_n$ (overweighting of tail events). This captures the effect of gamblers overweighting the tails of their own probability distributions (nonlinear weighting). Lastly, it captures source dependence by being applicable to probabilistic and uncertain outcomes.

2.2 Barberis' Casino Model

An important application to our work is Nicholas Barberis' Casino Model. [11] This model defines a model that places gamblers with Cumulative Prospect Theory preferences in the context of a casino. We define casino gambling in this model as games such as blackjack, dice and slot machines. We will apply some of these ideas in our analysis of Etheroll, as Etheroll is a dice game.

Let us begin with the formulation of a casino that only offers expectation 0 bets (essentially a binomial tree). Let a casino offer a gambler with initial wealth $W = W_0 > h$ up to τ gambles, where τ is the gambler's forced exit time. The set of all times the gambler can gamble in the casino is denoted by the set of timesteps $T = \{0, \ldots, \tau - 1\}$. The gambler in this model must eventually exit by same time τ, either due to ruin ($W < 0$), fatigue, work, or other commitments. At every timestep

T_i, the casino offers a gambler a bet B:

$$B = \begin{cases} +h & \text{with } p = 0.5 \\ -h & \text{with } p = 0.5 \end{cases}$$

which has expectation $E(B) = 0$.

Let us model a possible sequence of gamblers our gambler can take in this model. Let us assume at time $T_0 = 0$ that our casino offers the bet B to our gambler, and he or she accepts the bet. This is referred to as entering the casino. At the next timestep, time $T_1 = 1$, the outcome of the gambler's bet is reported. We generalize this for any time $t \in [0, \tau - 2]$, such that at time t the gambler is offered a bet B, in which the result is reported in time $t+1$. He can exit at any time by declining the bet, and at time τ the gambler is forced to leave. This models a typical evening of play. We will assume that the forced exit time of a gambler on the Ethereum blockchain to be the last bet he or or she wagered during a day of play.

A simple way to visualize this kind of betting model is through a binomial tree representation. Starting with the root of the tree (which represents the initial accepted bet), the gambler travels either left or right of the root to a new node. The gambler travels left in a winning bet, but travels right in a loss. In a model like this, white nodes of the tree represent bets that the gambler will continue at regardless of outcome, and black nodes represent exit times. We will refer to each node as a tuple (i, j) where i represents the time the node corresponds to, and $j \in [i + 1]$ represents the leftwards shift of each node. This leftwards shift is represented by taking $j_i = j_{i-1}$ if the result of the gamble is a win, and $j_i = j_{i-1} + 1$ if the gamble is a loss (where j_i is the value of j conditioned on the result of the previous $i - 1$ gambles).

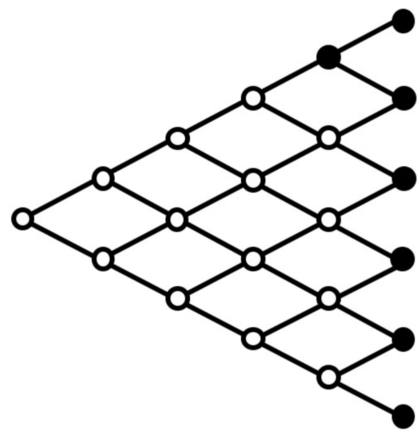


Figure 2.1: Binomial Decision Tree: $\tau = 5$

For reference, white nodes imply that the gambler plans to continue gambling at that node, and black nodes imply that the gambler plans to exit the casino at that node. We observe that using this notation, the gambler plans to only exit at the set $\omega_\tau = \{(4,1)\} \cup \{(i,j) : i = 5\}$

After this simple formulation, we present the behavioral assumptions that drive the analysis of this model. The first and most important assumption of this model is that for every time step $t \in T$, the gambler in this model aims to "maximize the Cumulative Prospect Theory value of his cumulative wins and losses at the moment he leaves the casino". In this aspect, we observe that we focus on the cumulative wealth obtained by this gambler, not the individual losses and wins he retains. This implies that the value function, $v : \Omega \to \mathbb{R}$ takes the cumulative profit of the gambler at a timestep as the input. This is based on the assumption that our gambler's reference point is the initial stake W_0. The reason we maintain this is because at the terminal time τ, we wish for the input of the value function to be the cumulative gains of the gambler, such that the valuation is framed as $v(\sum_{i=0}^{\tau} w_i - W_0)$, where $w_i \in \{-h, h\}$ is the result of each gamble at each timestep i. This assumption implies that gamblers in this model follow independent bets, and that bets in the beginning of the session matter far less than later bets.

The choice of prospect theory as a way to model the risk attitudes of these gamblers implies that there exists an interesting time inconsistency due to its probability distortion function w. We define a time inconsistency to be a change in the gambler's initial plan at some timestep t, influenced by the overweighting of small probabilities. For example, let us model a string of h bets, each winning some stake w_0 with a binomial tree of height h. Assume that the gambler's initial strategy is to gamble until the stopping condition that he or she reaches any node of depth d.

Observe that the probability that the gambler has made it to the $(t-2, 1)$ ($t-2$ wins in a row) is $P(E) = 1/2^{t-2}$. From the perspective of the gambler at $i = 0$, this gain has very low probability, but by the formulation of CPT, w overweights tail events, which causes our gambler to plan to gamble at $(t-2, 1)$. However, when the gambler actually reaches $(t-2, 1)$, the gambler does not continue to gamble (contrary to the initial plan). Observe that the gambler must make the valuation:

$$v(w_0 t - 2w_0) \geq v(w_0 t - w_0)w(0.5) + v(w_0 t - 3w_0)(1 - w(0.5))$$

to exit. We can frame this as the gambler only exits if his CPT value $v(w_0(t-2))$ at time $t-2$ exceeds the expectation of another bet B conditioned on his previous set of bets. Observe that this implies that:

$$v(w_0 t - 2w_0) - v(w_0 t - 3w_0) \geq w(0.5)(v(w_0 t - w_0) - v(w_0 t - 3w_0))$$

Which is true for all parameter values $\alpha,\ \delta \in (0, 1)$. This implies that the gambler will leave at time $t-2$, showcasing an interesting time inconsistency in his or her strategies.

In this simple model, we see the time inconsistency prevalent in many gamblers. This allows us to segment our gamblers into a variety of classes.

First among these classes is the naive gambler. This type of gambler does not realize that his or her probability weighting creates a time inconsistency. Let us model a possible path this gambler takes. At the initial time, $t = 0$, we define a plan as the mapping $s : T \times J \to \mathbb{Z}/2\mathbb{Z}$, where T is the set of all times, and J is the set of all possible leftwards shifts at every timestep. This choice of mapping represents the mapping of individual nodes to a decision, where 0 represents the gambler's decision to exit and 1 represents the gambler's decision to continue gambling. Denote the set of all possible plans as $S_{(0,1)}$. Let us also define the random variable W_s that represents the set of all cumulative profits and their associated probabilities of the gambler if he or she exits the casino at all the nodes defined by a plan s. This means this agent solves the problem at time 0:

$$\max_{s \in S_{(0,1)}} V(W_s)$$

where V is the value function that takes a prospect

$$G = \{(x_{-m}, p_{-m}), \ldots, (x_{-1}, p_{-1}), (x_0, p_0), \ldots, (x_1, p_1), \ldots, (x_n, p_n)\}$$

and maps it using:

$$V(G) = \sum_{i=-m}^{n} \pi_i v(x_i)$$

Our gambler only enters the casino if and only at time 0:

$$V^* = \max_{s \in S_{(0,1)}} V(W_s) > 0$$

We observe that these plans have a great deal of variability. First, we call a plan s a 'gain-exit' plan if the gambler's "expected length of time in the casino conditional on exiting with a gain is less than his expected length of time in the casino conditional on exiting with a loss". Simplified, this means that a gambler leaves at an earlier time t if he or she is winning, but stays until $t' > t$ if losing. This models gamblers who are more risk seeking, chase losses in an attempt to reach gains. Additionally, we observe that a gain-exit plan has a negatively skewed distribution. This distribution has a moderate probability of a small gain, but a low probability of a significant loss.

In contrast, a plan is defined as "Loss Exit" if," under the plan, the gambler's expected length of time in the casino conditional on exiting with a gain is greater than (the same as) his expected length of time in the casino conditional on exiting with a loss". Simplified, this means that a gambler leaves the casino earlier time t if he or she is losing, but stays at some time $t' > t$ if he or she is winning. This models more risk averse gamblers, who quickly exit based on losses. We observe that both of these plans are motivated by the overweighting aspect of the probability weighting function w. The loss-exit plan is naturally positively skewed, and thus attractive to gamblers. However, in certain parameter values, we observe that a gain-exit plan can also be attractive, given a low enough δ or α.

Assume that our gambler is satisfied with his choice of plan s. He enters the casino, and begins to gamble. Let us take some node (t, j) at $t \geq 1$. The gambler solves:

$$\max_{s \in S_{(t,j)}} V(W_s)$$

Assume there exists a solution s^*. Then the gambler gambles at (t, j) if and only if:

$$V(W_{s^*}) > U$$

Where U is the valuation mapped by v of leaving the casino. Interestingly, through numerical analysis, the naive gambler always changes from a loss-exit plan to a gain-exit plan.

Another category of gambler is the gambler who is sophisticated, or aware of the time inconsistency. However, this gambler is unable to commit to an initial plan s. This gambler uses backwards induction, working leftward from the right-most node of the binomial tree. Using the fact that he has a exit time τ, the gambler determines his actions at $\tau - 1, \tau - 2, \ldots$. Again, this gambler gambles at some node (t, j) if and only if

$$V(W_{s^*}) > U$$

as demonstrated in the naive case. However, in this case, the values of W_{s^*} are determined by backwards induction, implying time consistency. According to Barberis' numerical analysis, the gambler will overweight the tails of his distributions, leading the gambler to experience a negatively skewed distribution. This means he or she will be far pickier in which plans to enter with.

Last amongst these gamblers is the sophisticated gambler who is able to commit to his or initial plan. This gambler will commit to any $s \in S_{(0,1)}$. Similarly to the naive gambler, this gambler solves:

$$\max_{s \in S_{(0,1)}} V(W_s)$$

He searches all elements of $S_{(0,1)}$, and finds some solution s^*, such that

$$V^* = V(W_{s^*}) > 0$$

Interestingly, the naive gambler and this gambler both solve the same problem and if they chose the same s, would start with the same plan. It is very difficult for a gambler to exit when he or she has accumulated significant losses. In contrast, it is difficult for a gambler to continue when he or she has accumulated significant gains. In real life, a way to realize this forced restraint is for gamblers to enter a casino with a fixed amount of cash, and leave their ATM card at home. Often in the entire region of losses, these gamblers will be sorely tempted to gamble, following the time inconsistency.

2.3 Randomization of Strategies

In the paper "Path-Dependent and Randomized Strategies in Barberis' Casino Gambling Model", the authors aim to study Barberis' Casino Model with a focus on allowing randomized and path-dependent strategies. [?] We assume the same assumptions made in the casino gambling model of Barberis. We assume that the gambler takes random time τ to be his strategy. We observe that assuming CPT preferences, the gambler computes his relative value using the Choquet Integral (in continuous context):

$$V(X) = \int_0^{\infty} v(x) d[-w_+(1 - F_X(x))] + \int_{-\infty}^{0} v(x) d[w_-(F_X(x))] \tag{2.1}$$

Where X is the gain or loss relative to some initial reference point, and F_X is the CDF of X. w_+ is the probability weighting function associated with positive outcomes, and w_- is the probability weighting function associated with negative outcomes. $v(x)$ is the same as v in earlier sections.

Additionally, we observe that the actual cumulative wealth of the gambler after n bets (W_n) is modeled by S_n, which is the symmetric random walk on the integers. We take our initial wealth w_0 as a reference point for S_n. We derive for the exit time T a CPT preference:

$$\begin{aligned} V(S_T) = \sum_{n=1}^{\tau} v(n) \left(w_+(P(S_T \geq n)) - w_+(P(S_T > n)) \right) \\ + \sum_{n=1}^{\tau} v(-n) \left(w_-(P(S_T \leq -n)) - w_+(P(S_T < -n)) \right) \end{aligned} \tag{2.2}$$

Such that $+\infty - \infty = -\infty$. We wish to maximize $V(S_T)$.

Observe we take wish to compare three different types of strategic paradigms. First, we look at path-independent strategies, which are strategies in which $\forall t \geq 0$, $\{T = t\}$ is only determined by W_t.

Gamblers also take path-dependent strategies, in which for any time $t \geq 0$, $\{T = t\}$ is determined by the prior information set $\mathcal{F}_t = \sigma(S_u : u \leq t)$. This set is the entire gambling history of the gambler up to some time t.

Lastly, we take a strategy that is both path-independent and randomized. This strategy follows the same idea as the general path-independent strategy, with the exception of a few nodes in which the gambler flips a coin to decide his continue/exit plan. A result of heads means continue, where a result of tails means exit.

The authors then provide a mathematical formulation for these strategies. First, take some discrete-time Markov Chain $X = \{X_t\}_{t\in[0,T]}$. We also assume that the gambler chooses a optimal stopping time $\tau \leq T$ to maximize X_τ. Now, let us define $\mathcal{A}_M$, and $\mathcal{A}_D$ as the set of path-independent and path-dependent strategies respectively. Observe that using the information set $\mathcal{F}_t$ we can write $\mathcal{A}_D$ as the set of $\{\mathcal{F}_t\}_{t\geq 0}$ stopping times. Next, we observe that $\mathcal{A}_M$ can be represented as the set of T's in $\mathcal{A}_D$ such that $\forall t \geq 0$, conditioned on $\{T \geq t\}$, $\{T = t\}$ is dependent on only (t, S_t).

To define randomized, path-dependent strategies mathematically, we take a family of $0-1$ random variables $\zeta_{t,x}$, $t \in [0, \tau]$, $x \in \mathbb{Z}$ such that $\zeta_{t,x}$ is independent of $\{X_t\}$, and are mutually independent. These $0-1$ random variables represent the random coin flips are some timestep t, when $X_t = x$, with $\zeta_{t,x} = 0$ representing tails and $\zeta_{t,x} = 0$ representing heads. A possible strategy is stopping at the first time τ' when a coin toss turns up tails:

$$\tau' = \inf\{t \in [0, \tau] | \zeta_{t,X_t} = 0\} \tag{2.3}$$

Our information set $\mathcal{F}_t$ then becomes enlarged to $\mathcal{G}_t = \sigma(X_u, \zeta_{u,X_u}, u \leq t)$. Interestingly, τ' is path-independent, such that $\{\tau = t\}$ depends only on X_t and ζ_{t,X_t} (conditioned on $\{\tau \geq t\}$. We define our set of randomized, path-independent strategies as $\mathcal{A}_R$, where $\mathcal{A}_R$ is the set of all τ defined in (2.3). Observe that this set is a subset of $\mathcal{A}_C$.

2.4 Betting Systems

A major factor that we are interested in is to empirically find examples of betting systems, specifically staking betting systems. Let us define a betting system as any kind of structured strategy that attempts to produce a net profit. These systems promise to convert probabilistic house edges into player edges, a mathematical impossibility. Even more ridiculous is the fact that many of these betting systems are applied to games which consist of many random, independent trials in an attempt to alter their long term expectations. These systems are especially common in games like dice, roulette, and blackjack. However, it is possible for some of these systems to be slightly more profitable in the short term, at the cost of a large amount of risk. These systems are of empirical interest, as it is very difficult to find data of gamblers and their results in real life situations (such as casinos, or illegal configurations).

We can classify many of the most common staking gambling systems as negative-progression systems. A negative-progression gambling system conditions increasing bet sizes on losses. Many of the most famous and popular gambling systems are systems that follow this strategy.

2.4.1 Martingale

One of the systems we will evaluate is the Martingale gambling strategy. Assume that we are playing a game with two outcomes: win with probability p, and lose with probability q. The first step of implementing this strategy is to pick some initial stake $W = w_0$ at some initial starting point $\tau = 0$. Define the sequence $a_n = \{2^n\}_{n=0}^{\infty}$.At any timestep $i > 0$, our gambler bets some amount $a_i w_0$. At the next timestep, $i+1$, the gambler obtains the result of his or her gamble (either win or loss). We define his or her next bet size, w_{i+1}, to be:

$$w_{i+1} = \begin{cases} a_{i+1}w_0 & \text{if loss} \\ w_0 & \text{if win} \end{cases}$$

Clearly, bet sizes in this system grow exponentially fast, conditioned on continual losses. The logic behind betting in this system is that given a initial stake w_0, there exists some time τ in which the gambler will win with probability p. His losses will add up to $L = \sum_{i=0}^{\tau-1} 2^i w_0$, and we observe that $2^\tau w_0 - L = w_0$, meaning that the gambler will recover his or her initial stake. However, the gambler does not possess infinite wealth, so with exponentially increasing bet sizes, the probability of ruin approaches unity.

Assume that the gambler has a bankroll of $W = 2^n w_0$, and initial stake w_0. First, we observe that the probability of the gambler losing exactly n times in a row is q^n, and the probability he or she does not lose is $1 - q^n$. Then, we have the expected profit per round must be:

$$(1 - q^n)w_0 - (q^n)\sum_{i=1}^{n} 2^{i-1}w_0$$

As the gambler obtains w_0 if he or she does not lose. Additionally, observe that $\sum_{i=1}^{n} 2^{i-1}w_0 = w_0(2^n - 1)$. Observe this means our expression simplifies to:

$$w_0 - q^n w_0 - q^n(2^n w_0) + q^n w_0 = w_0(1 - (2q)^n)$$

Observe that in the case when $q > 0.5$ (unfair game), this expression is always negative. This implies that in an unfair game, the martingale is a losing proposition per bet (has negative expectation).

2.4.2 Large Proportion Betting

Another system of empirical interest is a system where the gambler bets everything per each wager. Let us assume that we are playing in a casino that offers positive expectation bets $p > 0.5$, with payouts per wager w of $w + qw$. Observe that we must have expectation per bet B:

$$E(B) = p(qW) - qW = qW(p - 1)$$

which is clearly less than 1 as $p - 1 < 0$.

2.4.3 Kelly Criterion

An empirical system similar to the bet everything system is the proportional betting system using the Kelly Criterion. Assume we are playing some biased game where $p > 0.5$. The Kelly Criterion tells us that given odds b (payout on a bet of 1 unit), probability of winning p, probability of losing q, the fixed fraction f^* of your bankroll W you should wager is:

$$f^* = \frac{bp - q}{b} = \frac{p(b+1) - 1}{b}$$

This fraction will maximize the logarithm of W.

Proof: Let us begin with initial wealth W_0. Let our total wins after n bets be represented by S_n. Assume that you bet the proportion f^* every time. If the gambler wins S_n times, but loses $n - S_n$. Then your cumulative wealth after n bets, W_n is modeled:

$$W_n = (1 + bf^*)^{S_n} (1 - f^*)^{n - S_n} W_0$$

Observe that our fractional gain is equivalent to:

$$r = \log \left(\frac{W_n}{W_0} \right)^{1/n} = \frac{1}{n} \left(S_n \log(1 + bf^*) + (n - S_n) \log(1 - f^*) \right)$$

Observe that:

$$E(r) = p \log(1 + bf^*) + q \log(1 - f^*)$$

We aim to maximize $E(r)$, so we take $E'(r) = 0$:

$$\frac{dE}{df^*} = \frac{pb}{(1 + bf^*)} - \frac{q}{1 - f^*} = 0$$

We simplify this, obtaining that:

$$\frac{pb}{(1 + bf^*)} - \frac{q}{1 - f^*} = 0 \implies f^* = \frac{bp - q}{b}$$

As desired. However, as detailed above (in the gambler's ruin section), even if the gambler follows this for an infinite amount of steps without decreasing his or her wager, his or her wealth will tend to ruin.

Chapter 3

Methods

3.1 Data Collection

This project entirely focuses on data collected from bets on the DApp (Decentralized Application) Etheroll. Data about these bets used to be hosted on a site https://www.cryptocurrencychart.com/etheroll-live-stats. To obtain this data, we used a customized screen scraper built in Python using Requests and Beautiful Soup. Requests is a user-friendly Python library designed to handle HTTP requests, and Beautiful Soup is a HTML parsing library. We then databased this data, which consisted of approximately 250 000 individual bets and 2600 gamblers in a mySQL database. These numbers are based off the four iterations of Etheroll's smart contract. Contract 1 ranges from 4/17/2017 to 4/24/2017, Contract 2 ranges from 5/4/2017 - 5/18/2017, Contract 3 ranges from 5/23/2017 - 10/25/2017, and Contract 4 ranges from 10/25/2017-12/12/2017. All of this data was collected when the minimum bet on Etheroll was still 0.1 Ether (a value which ranged roughly from 4.3 USD to 52 USD).

Taking some simple descriptive statistics for each contract:

Table 3.1: Environment Summary Statistics

Contract	Average Bet Size (ETH)	Unique Gamblers	Total Bets
1	1.595	201	3919
2	2.006	179	5671
3	1.527	1889	132826
4	0.987	516	43425

An raw individual bet consists of 7 fields: A Datetime stamp, Player identification (such as Professional #2924, All in #2922), Bet Size (in Ether), Chance (number chosen to roll under), Paid_ETH (Payout in Ether), and Paid_USD (Payout in USD, converted from Ether at the time). According to the maker of the website, the player identification names follow this pattern: Newbie - New address, All in - High value bets, Lucky - Wins against the odds, Play it safe - Multiple high chance bets, Against the odds - Losses with high win chance, One in a million - Won very low chance bet, Intermediate - more than 5 bets, Professional - more than 25 bets, Legend - more than 100 bets.

The Chance feature is useful for determining the general riskiness of the population. Additionally, having individual player identification codes allows us to subset our data to observe the habits of each individual gambler. Having access to the timescales of the gamblers also lets us observe some interesting time inconsistencies prevalent in the data. Lastly, knowing the Bet Sizes and Payouts per bet allows us to solve for the cumulative profit of the gambler at each timestep. This allows us to apply ideas from Cumulative Prospect Theory and Barberis' Casino Model.

In preprocessing our data, we aim to subset and clean the data in a sensible way. The first way to subset the dataset was to organized the data by individual gamblers. Next, we took these individual gamblers and subset them by the days they gambled in. Lastly, we computed the simple payout W for each gamble, which for some bet B, result $R \in \{\text{loss}, \text{win}\}$ was:

$$W = \begin{cases} R = \text{loss} & 0 \\ R = \text{win} & W(1-p-e)/p \end{cases}$$

Where e is the house commission, 0.01, and p is the probability of winning the gamble.

3.1.1 Methodology

We will explore this data using a variety of common data analysis techniques and qualitative analysis.

We will collect, analyze and visualize data using Python's pandas library. As covered earlier, the data was scraped from https://www.cryptocurrencychart.com/etheroll-live-stats. An alternate way to obtain this dataset would involve setting up a computer as a Geth node (Ethereum node member), and use the JSON-RPC API to repeatedly call for contract log events (where the smart contract specifies data).

First, we aim to characterize gamblers in Etheroll, through each of their lifetime bet, bet size, betting probability, and cumulative profit distributions. We will determine patterns in this data by analyzing histogram distributions. This will allow us to profile the average gambler's preference for probabilities of winning, and average bet frequency. Additionally, we will analyze the relationship between bet sizing and probabilities of winning.

Next, we aim to carefully characterize gamblers who follow a few types of gambling strategies. We first preselect a few path-independent (fixing of probabilities and wager size) and path-dependent strategies (martingale). To explain certain psychological behavior, such as loss function, we will explain the behavior in terms of prospect theory concepts such as the weighting function and value function. Additionally, we will try to qualitatively define the risk profile of gamblers, depending on strategies and gambling patterns. We do this through observing cumulative profits over time and changes in wager sizing. To characterize the local behavior near a gambler's exit time, we use the metric of scaled cumulative profit. Take some set of times $T = \{1, \dots, \tau\} \subset \mathbb{N}$, where τ is a forced exit time, and set of cumulative profits $W = \{w_0, w_1, \dots, w_\tau\}$, where for some $t \in T$, $w_t = \sum_{i=0}^{t} r_i$, where r_i is the profit at time i. We scale this cumulative profit, instead taking an alternate sequence $\{2^{-i} w_i\}_{i=0}^{\tau}$. This is an attempt to model the memory of a gambler, with the assumption that more recent bets are weighted more significantly than bets in the past. This follows a psychological phenomena known as the primacy and recency effect, where initial and final results are overweighted in terms of importance and memory. [12]

Chapter 4

Results

4.1 Characterizing the Population

4.1.1 Wager Sizing

This population of gamblers on the Ethereum blockchain allows us to empirically observe the tendencies of gamblers in a casino-like environment for the first time. The minimum bet-sizing of 0.1 Ether (4 - 53 USD in this dataset) simulates casino-like stakes. [13] In this section, we will characterize the types of gamblers in this online casino, the overall distribution of gamblers in the casino, and the paired cohort of winning gamblers and losing gamblers.

To first characterize this population of gamblers, we visualize the total bet frequency distributions of each gambler. Using histograms, we track each gambler's total gambles per smart contract, and the corresponding frequency of occurrence. In doing so, there is a very pronounced right skew in the distribution of the amount of gambles of each gambler. In fact, in each smart contract iteration, the gamblers who gamble only 1 to 10 times comprise approximately 60-65% of the entire population. This heavy right skew shows that most gamblers in this DApp are mainly recreational gamblers who place anywhere from 1 to 10 bets (See Table 4.1, Figure 4.1).

Table 4.1: Total Bet Distribution of Gamblers

Number of Total Bets	Contract 1(%)	Contract 2(%)	Contract 3(%)	Contract 4(%)
$1 - 10$	66.16	62.57	55.16	60.47
$10 - 100$	29.85	24.02	31.87	26.36
> 100	3.98	18.99	12.86	12.98

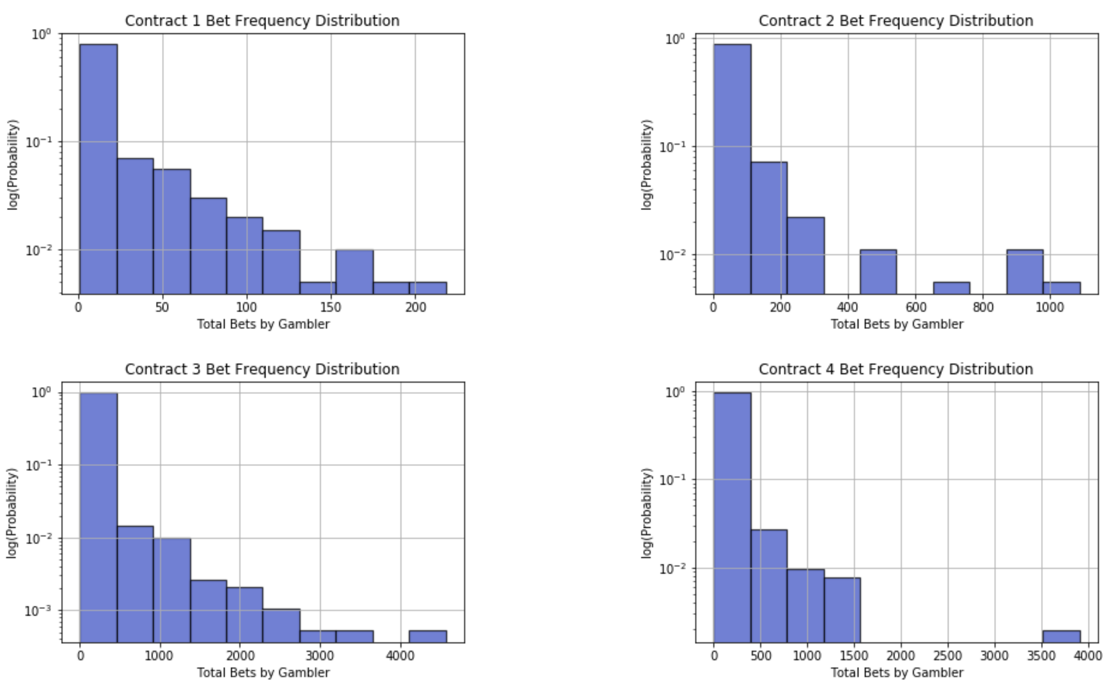


Figure 4.1: Bet Frequency Distributions

An interesting qualitative feature of these distributions is that throughout each contract iteration, the relative bet frequencies of these gamblers remained relatively constant. Another interesting feature of the data is the existence of a tail of gamblers who bet at high frequency. The "Whale Bettors", or bettors who bet more than a hundred bets and contribute most of the actual bets on the website comprise only a small fraction of the actual gamblers. Due to gambler's ruin, we see that these whale bettors, who frequently gamble, must be more risk-taking. In contrast, the gamblers who gamble less must be more risk-averse.

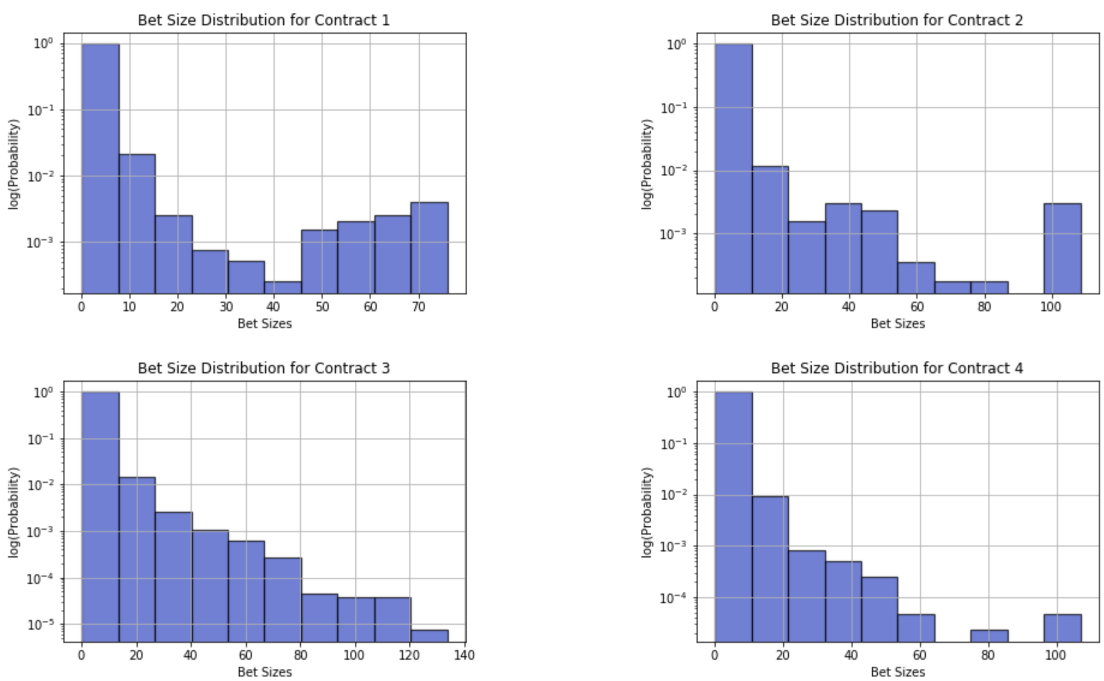


Figure 4.2: Bet Size Distributions

We see a very similar right skew in the distributions of Contracts 1, 2, 3 and 4. However, Contract 1 displays a surprising amount of gamblers that are willing to gamble at large bet sizes. Additionally, there are always a few gamblers willing to bet at significant sizings (> 80 ETH). Possible reasons for this were probably due to the relatively low price of Ethereum (approximately 1 ETH : 50 USD). Additionally, there were only 90,000 total transactions on the Ethereum network at the time. Many of these gamblers probably did not expect the prices to exponentially rise to 500

USD/ETH.

4.1.2 Probability Distributions

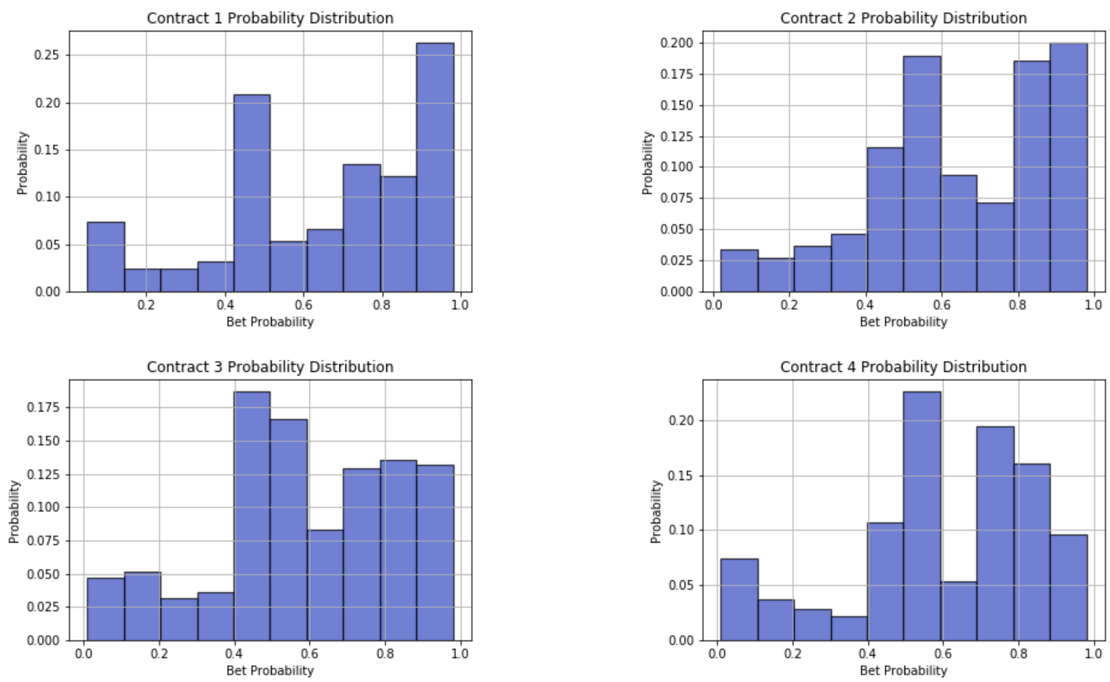


Figure 4.3: Total Population Distribution

In observing the overall distribution of the probabilities that the gamblers on Etheroll gamble at, we observe two interesting fixations. First, gamblers are extremely drawn to probabilities within the bound of $p = 0.4 - 0.6$. This is slightly different from what median Cumulative Prospect Theory preferences specify, as probabilities around $0.35 - 0.6$ are underweighted, rather than overweighted. Lower probabilities have the opposite pattern. Additionally, these gamblers also have a fixation towards probabilities with very high chances of winning, within $p = 0.8 - 0.99$. This showcases these gamblers are qualitatively more risk averse. This is an odd result. First, we observe that theoretically, gamblers in this casino are more likely to be a self-selecting, risk seeking group. First, these gamblers must have some interest in Ethereum, and are also forced to bet significant minimum bet sizings (5-53 USD).

Additionally, we wish to look at the probability distributions of two possible cohorts of the gambling population: gamblers who win, and gamblers who lose. To do this, we segment our data into gambles of gamblers who lose and those who win.

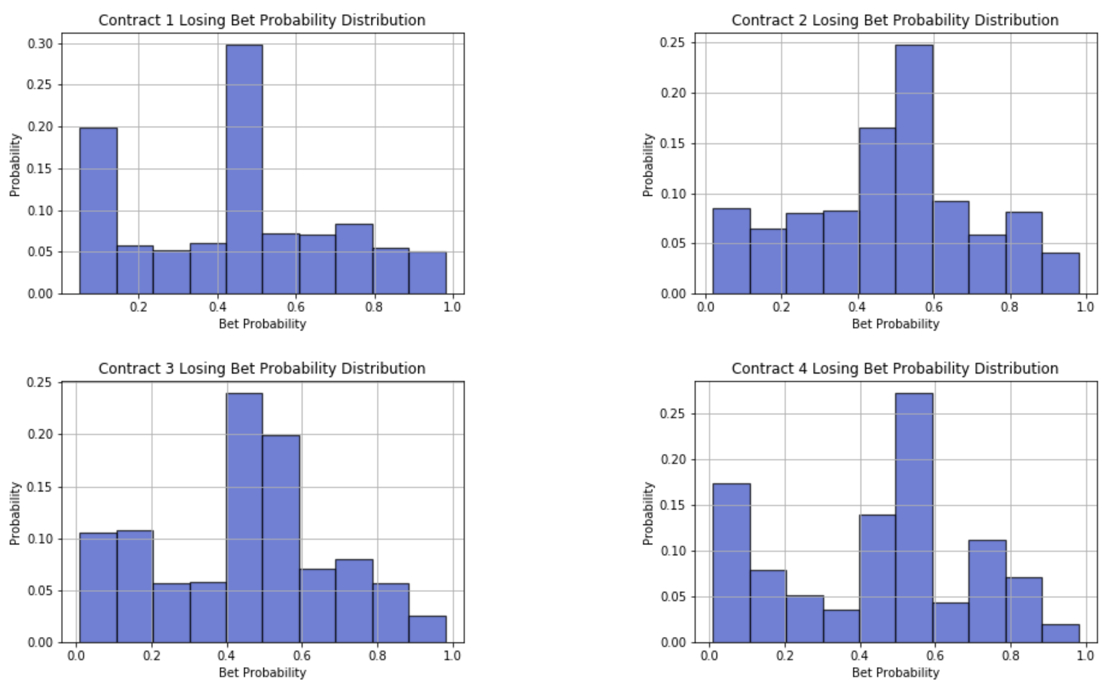


Figure 4.4: Losing Cohort Distributions

In all four contracts, the losing cohort of gamblers have very similar losing distributions (see Figure 4.3). In general, there is a large central mean at $p = 0.5$. In Contracts 2 and 3, there is a nearly normal distribution in their probabilities. We observe that in every contract, nearly 25% of bets are losing bets at around $p = 0.5$. Additionally, many of the extremely risky gamblers who bet at $p < 0.5$ are represented in this cohort. Naturally, gamblers who bet like this will tend to lose more often. The other tail end of the distribution comprises of the gamblers take $p > 0.5$

gamblers, tending to be less risky and loss averse. However, as $p \neq 1$, they are still bound to lose, and with a maximum probability of 0.99, they will still lose at least 1 out of 100 times on average.

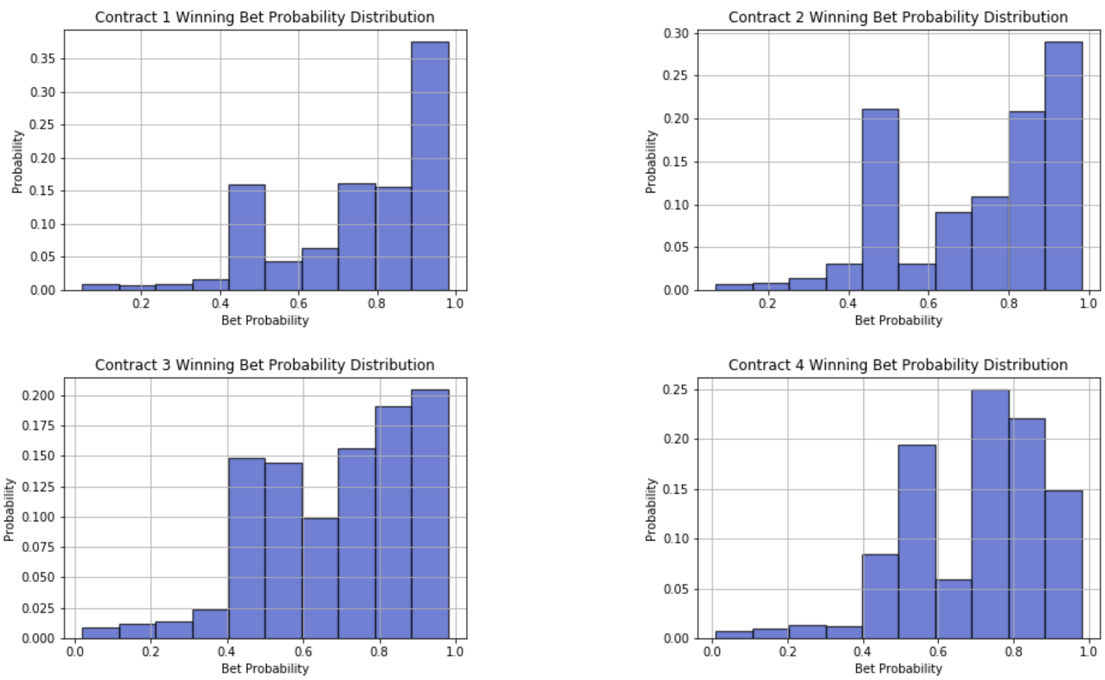


Figure 4.5: Winning Cohort Distributions

Lastly, we observe that the winning cohort also has relatively similar general distribution throughout the four contracts. This is due to the fact that with a large enough sample, individuality is mostly canceled out. At first glance, it is apparent that there is a distinct left skew in the distribution, with a large amount of bets being distributed at $p > 0.5$. However, there is still a noticeable fixation by these gamblers to bet at $p = 0.5$. We also notice that these winners probably tend to be more risk averse, as most of the data is accumulated at $p > 0.7$. Very few winners occur in the region of $p < 0.5$, which comprises less than 10% of the data on average per each contract iteration. Lastly, an interesting feature in almost every distribution is an aversion to $0.5 < p < 0.7$. This may be due to the shape of the probability weighting function, where probabilities that are higher than $p > 0.5$ are overweighted. An explanation for why $p > 0.7$ is so popular comes in the loss-aversion formulation of the value function. As gamblers generally wish to avoid losses, they tune their probabilities very high to avoid losses.

4.1.3 Probability vs Wager Sizing

Another way we can observe risk attitudes is in evaluating the relative bet sizing of gamblers versus their tuned probabilities of winning.

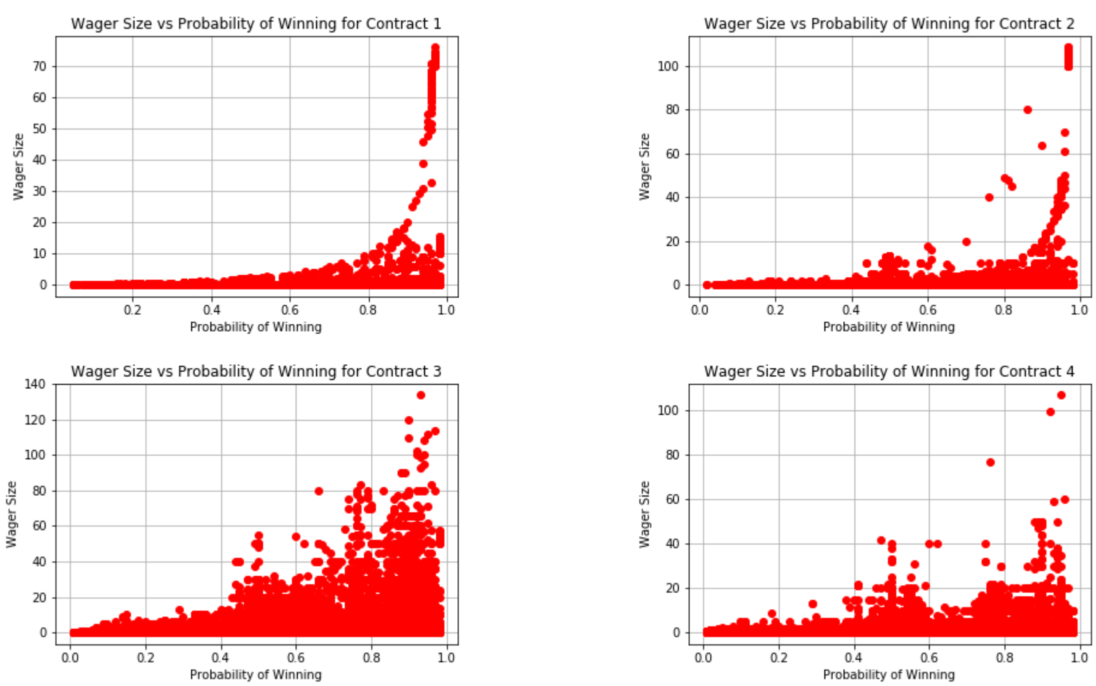


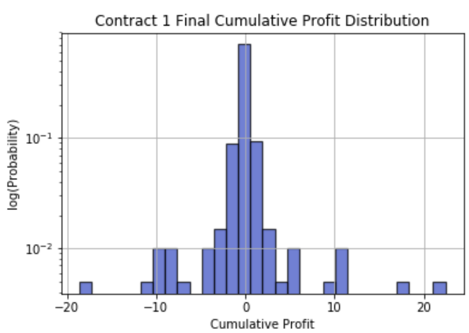
Figure 4.6: Wager Sizing versus Probability of Winning

These plots showcase the loss aversion of most gamblers. When staking very large bets, these gamblers exclusively bet at very high probabilities. The largest bets are always bet at extremely high probabilities. With smaller sizings, we see a whole range of probabilities of winning. In general, this is not an extremely surprising finding, as we expect most gamblers to be loss averse.

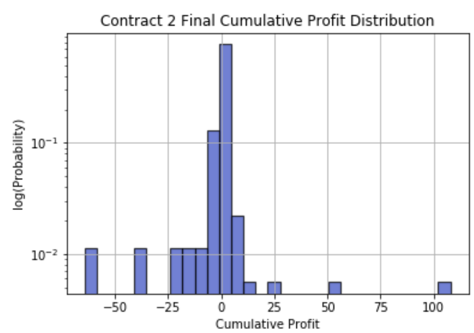
4.1.4 Cumulative Profit Distributions

Another interesting function of this is that these Bet Frequency and Probability distributions mostly shared similar shapes throughout all four contract iterations. However, the actual value of these bets greatly varied.

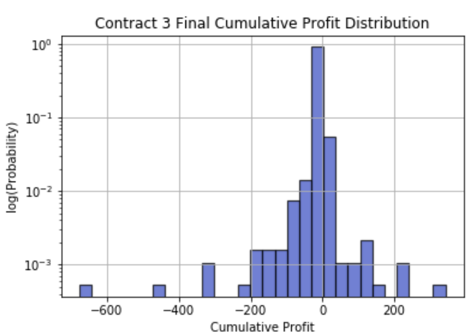
Additionally, it is interesting to see the distribution of the cumulative profits at the end of each gamblers gambling time (See Figure 4.6).



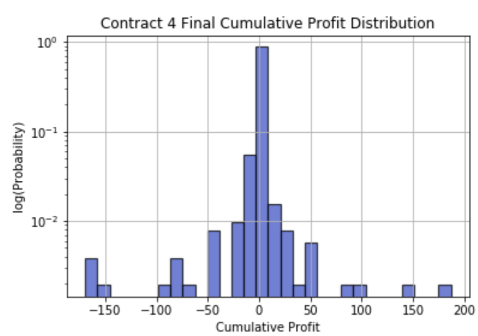
(a) $\mu = -0.006$



(b) $\mu = -0.537$



(c) $\mu = -2.880$



(d) $\mu = -0.658$

Figure 4.7: Distributions of Cumulative Profits

It appears that Contract 1, Contract 4 have a distinctly normal distribution with

$\mu < 0$. However, Contract 3 seems to have a left skewed distribution, and Contract 4 seems to have a right skewed distribution. We hypothesize that this distribution is most likely transient, meaning that the skew in Contracts 2 and 4 are determined by variance. We also can conclude that most gamblers do not really win anything (matching up with the fact that 60% of gamblers are recreational gamblers). We see that the mathematical edge of the casino has effectively shifted the normal distribution to the left, as expected. This implies that most gamblers are net losers, as expected from an edged game.

4.2 Gamblers of Empirical Interest

We are interested in looking at the existence of gambling strategies because it allows us to validate some of the ideas behind theoretical predictions. We search for strategies that are path independent and path dependent, allowing us to verify the types of gamblers in Barberis' Casino model. However, we will observe that it is impossible for us to observe sophisticated, committed gamblers gamblers from an empirical standpoint, as we do not know what devices .

Unfortunately, we will be unable to evaluate gamblers that follow proportional betting standards, as it is not possible to parse wallet data at the times they bet at. Because of this, we do not have access to their total wealth, and we cannot search for proportions they bet at. In response to this, we will evaluate gamblers that bet at a fixed wager in place of proportional bets. Another interesting system is the betting system in which a gambler continuously bets at a high probability (analogous to the bet everything system). This is one of the most common systems. Additionally, we will be able to evaluate gamblers who bet with negative progression, staking betting systems. We will evaluate the most common system: the martingale. This is because gamblers following the martingale always return to their original stake, making it easier for us to detect the usage of this system. We also aim to see if gamblers have a mixture of systems, such that they transition strategies over some timescale. The reason we choose these systems is because it allows us to possibly observe time inconsistencies in both path-independent (fixed wager/fixed probability/high probability) and path-dependent (mixed) strategies.

To find more simple systems, such as fixed probability, fixed wager or high probability betting systems, we will use Python's Pandas package for data analysis. To classify fixed systems, we convert our data, stored in csv files into Pandas Dataframe, and search for gamblers who have zero variance in their probability choice and wager sets. To classify high probability bettors, we choose any bettor who bets only at $p > 0.9$.

We also will only apply these methods to gamblers who have sufficient gambling data, e.g. whale bettors (bettors with over 100 total gambles). Additionally, we assume that the initial reference frame that these gamblers once they enter the online casino is the beginning of the day their betting session starts at. In this way, we will be able to see possible time inconsistencies in daily data. With this as our reference point, we found many gamblers that behave similar to the hypothesized path-independent betting, naive gamblers who are unaware of the time inconsistencies. However, we observe some interesting path dependent betting strategies (such as the martingales), or some mixture of betting strategies.

4.2.1 Fixed Probability

The most common and popular "strategy" found in this population is the fixed probability betting strategy. In this strategy, the gambler chooses some probability, usually $p > 0.5$ and continually bets. In evaluating gamblers who bet with fixed probabilities or wagers, we can observe what happens when a gambler has no strategic time inconsistency (irrespective to exit strategies).

An example of this strategy is found in the betting history of Professional #5619. Through the three date periods of 11/23/2017 3:17:43 PM - 11/23/2017 7:54:43 PM, 11/24/2017 8:22:40 AM - 11/24/2017 11:30:14 PM, and 11/25/2017 8:20:38 AM - 11/25/2017 10:04:12 PM, this gambler gambled 147 (whale) times, all at $p = 0.49$. This is an example of a simple, path-independent strategy. No matter what his results are, the gambler sticks to his initial strategy. Upon entering the casino, the gambler commits to a strategy, and even upon accumulating losses, he or she continues to gamble at a suboptimal probability. In gambling at $p < 0.5$, this gambler is taking significant risk - in the long run his or her expectation runs negative.

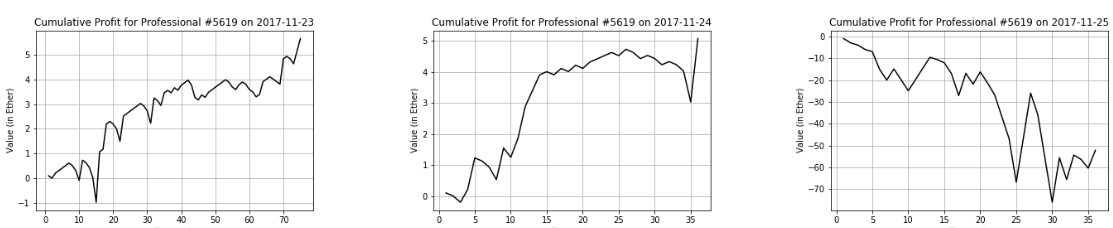


Figure 4.8: Raw Cumulative Profit vs. Timescale for 11/23/2017 - 11/25/2017

Figure 4.9: Wager Sizing vs. Timescale for 11/23/2017 - 11/25/2017

This gambler may have a time inconsistency in terms of exit strategies, but this is something we cannot empirically deduce. He or she follows the betting strategy, but exits once the losses reach some arbitrary stop-loss. We also observe loss-chasing behavior in the third day. This gambler begins by betting at an extremely large initial wager size of around 2 Ether (which is around the same size as the maximum bet he or she bet at in the past two days), and progressively increases bet sizing as losses accumulate. We observe a two-peak plateau in the bet sizing of this gambler. The first plateau is positive for the gambler. The gambler won two consecutive large bets (20 ETH), recovering a significant portion of his or her accumulated losses. Immediately, we observe a drastic decreasing in bet size, and a subsequent loss. The gambler continues to bet at high values, putting themselves more and more negative. After a string of consecutive losses, the gambler adjusts his or her bet size, eventually exiting a a loss of 41.44 Ether. Attempting to chase a few initial losses through increasing bet sizing only resulted in a larger loss. It seems that this gambler followed a gain exit pattern. Upon large wins at the end of his or her first sessions, the gambler stopped betting.

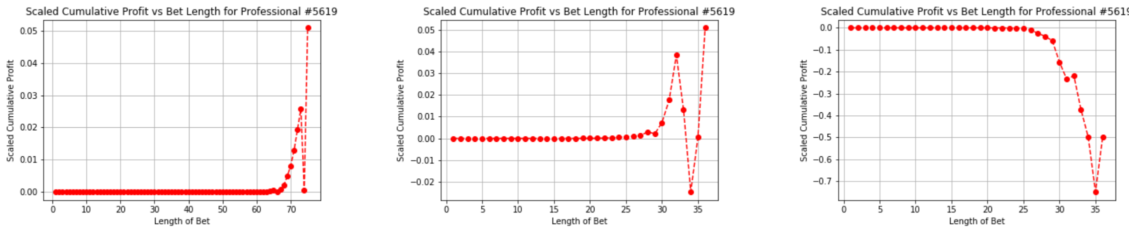


Figure 4.10: Wager Sizing vs. Timescale for 11/23/2017 - 11/25/2017

In a very heuristic way, we can label these kinds of gamblers as more risk seeking if $p < 0.5$ and less risk seeking if $p > 0.5$.

4.2.2 Fixed Wager

Rarer is the fixed wager strategy. This strategy is extremely simple. A gambler takes some initial stake W_0, and continually gambles, tuning his probability through wins and losses. Oddly, this strategy never the only strategy the gambler employed over the timescale of a day.

4.2.3 Martingale

As the Etheroll minimum bet size is quite large, we observe that a martingale system diverges very quickly. However, certain gamblers still follow this system. As observed, this gambler follows a martingale strategy. His bet size starts with an initial staking size $w_0 = 0.2$ Ether, and follows:

$$w_{i+1} = \begin{cases} 2^i w_0 & \text{if loss} \\ w_0 & \text{if win} \end{cases}$$

Where w_{i+1} is the $i+1$-th bet conditioned on i losses. In the wager sizing graphs (Figure 4.10), we observe a very clear return to the initial staking size after every win. These graphs showcase the nearly linear staking gains of the martingale for a gambler from the timespan of 5/6/17 – 5/7/17 (where he bet a total of 172 times).

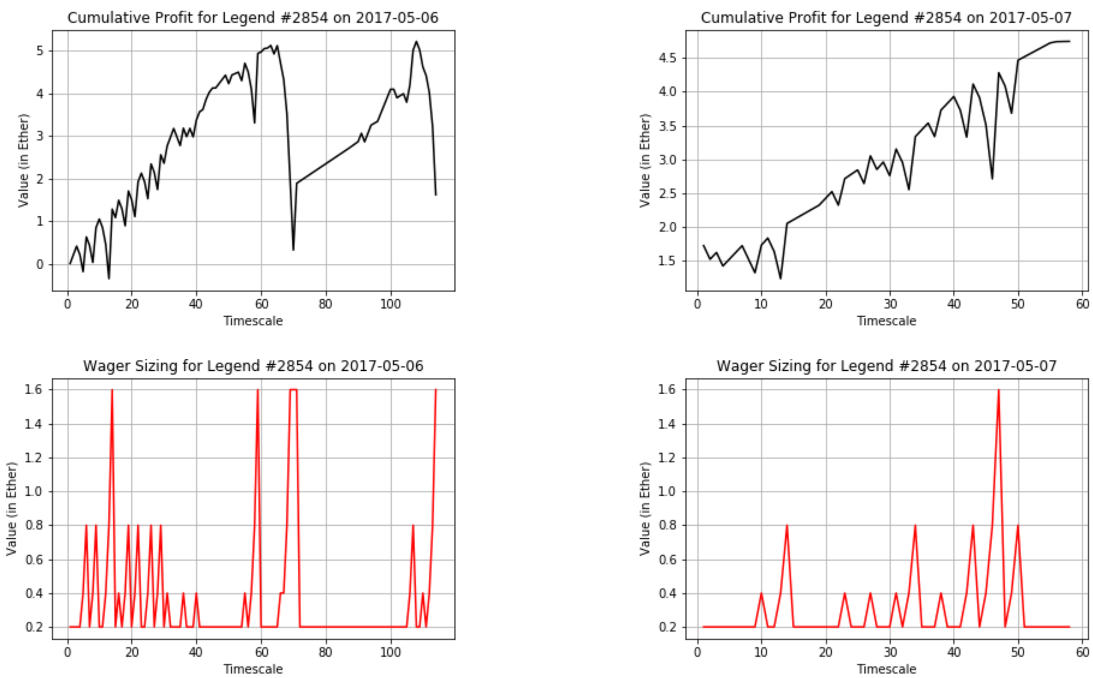
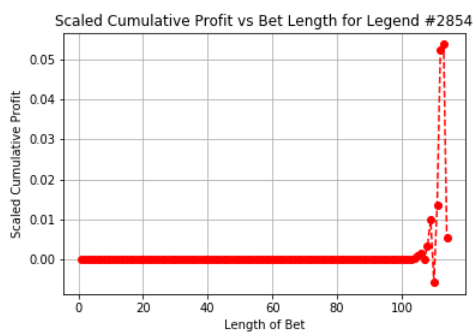


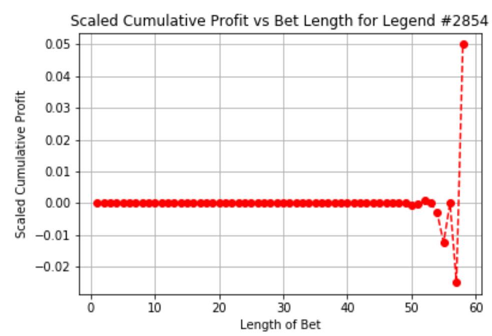
Figure 4.11: 5/6/2017 – 5/7/2017 Wager and Cumulative Profit Graphs

This is because of the guaranteed return of w_0 from this system. However, we notice on 5/6/17 there is a distinct drop in cumulative profits due the gambler having a stop loss after 6 consecutive lost bets. This showcases the dangerous fast divergence of the martingale system. We also see an interesting time inconsistency from this gambler. Looking at the data, we see the gambler's exact deviation from this strategy at his or her 69-th bet, where the gambler bets $2^3(0.2) = 1.6$ Ether at probability 0.5 and loses. If this gambler is following the martingale system, he or she should bet $2^4(0.2) = 3.2$ Ether at the same probability (to get a $1 : 1$ return). However, the gambler disregards this, and gambles the same gamble of 1.6 Ether, and loses again. This causes a sharp decrease in his or her cumulative gains. Interestingly, the gambler makes a third bet of the same amount, and same probability. Probabilistically, he or she will win this bet on average, but it comes with significant risk. This showcases the tendency of gamblers to chase losses - similar to the idea of gain-exit strategies. It also matches with the concept of the value function (Tversky and Kahneman's Prospect Theory). Theoretically, when making decisions under risk, gamblers become risk seeking when faced with losses. This helps to explain the "loss-chasing" phenomenon.

Other possible reasons for this sudden time inconsistency in strategy could involve the gambler's total, overall bankroll. From observing his or her wager sizing, we observe that the gambler never bets past 1.6 Ether. Hypothetically, the next bet in the martingale sequence (3.2) could simply be too much for the gambler to continue, forcing the gambler to deviate from the planned strategy.



(a) 5/6/17 Scaled Cumulative Profit



(b) 5/7/17 Scaled Cumulative Profit

The trajectory of the scaled cumulative profit of this gambler shows his or her valuation ("satisfaction") through the losses and gains, framed near the exit time. We observe that even though this gambler was a net positive (on 5/6/17), his or her scaled cumulative profit is very faintly positive, representing the effect of the severe loss. In contrast, we see the effect of ending with a large win on 5/7/17. The gambler's scaled cumulative profit is at a global maximum at his or her exit time, influencing his or her exit time. This gambler ended up as a net winner, winning 4.34 Ether.

Chapter 5

Conclusion

Through this initial data exploration, we discovered patterns of gambling behavior. One of these findings is that as expected, a large proportion (approximately 60%) of gamblers are recreational (0-10 lifetime bets). Additionally, we observe a nearly normal distribution ($\mu < 0$) of gamblers by cumulative profit, which is expected from Etheroll's 1% commission. Surprisingly, gamblers on Etheroll also follow a generally more loss-averse distribution of probabilities ($p > 0.5$). We also find some interesting gamblers who follow betting strategies (path-independent and path-dependent), and attempt to qualitatively explain their exit times, behavior and risk profiles through the lens of prospect theory and cumulative profit. In looking at a mixture of strategies, we see gamblers who display strategic time inconsistencies, gain-exit and loss-exit strategies, and loss-chasing behavior.

There are still many ways to explore this environment. As each individual has his or her unique utility function, it would be interesting to approximate this using a fitting of prospect theory's value function. Additionally, if we apply a similar parameter estimation as the one used by Tversky and Kahneman to approximate

w_+, w_-, we could approximate individual probability distortion functions. The issue with this method comes from the fact that many gamblers do not necessarily bet amongst the whole spectrum of probabilities (0,1), but instead bet within some chosen subset. Thus, it becomes almost impossible to estimate the whole distortion function. Having an estimate of the probability distortion or value function would allow us to quantify the riskiness of individuals. It could be interesting to see if certain gamblers have weighting functions which underweight low probabilities, but overweight high probabilities. Individuals with value functions that have steeper loss regions are more sensitive to losses, and thus more willing to take risks and chase losses.

Another item of interest could be using some kind of machine learning algorithm to search for other path-dependent strategies. Many other negative-progression, staking strategies exist, such as the D'alambert, Fibonacci, and Labouchere system. However, it is very hard to classify gamblers as gambling under these strategies, as it is rare for a gambler to perfectly follow a strategy. Being able to observe the deviation of gamblers from their ignitial strategies would be way to characterize the risk attitudes of the gamblers, and possible gain-exit vs. loss-exit patterns. It would also be of interest to find the existence of gamblers who are clearly loss-exiting against gain exiting.

References

[1] Ruffer SM, Yen D, and S Lee. Client/server computing technology: A framework for feasibility analysis and implementation. *International Journal of Information Management*, 15(2):135–150, 1995.

[2] What is torrenting? 4 things you need to know. https://www.cloudwards.net/what-is-torrenting/.

[3] Vitalik Buterin. Ethereum whitepaper: A next generation smart contract and decentralized application platform.

[4] Ethash. https://github.com/ethereum/wiki/wiki/Ethash.

[5] Satoshi Nakamoto. Bitcoin whitepaper. https://bitcoin.org/bitcoin.pdf.

[6] Jonas David Nick. Data-driven de-anonymization in bitcoin. https://nickler.ninja/papers/thesis.pdf, August 2015.

[7] State of the dapps: 1,539 projects built on ethereum. https://www.stateofthedapps.com/.

[8] Etheroll : Smart contract. https://etheroll.com/#/smart-contract.

[9] Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5:297–323, 1992.

[10] Amos Tversky and Daniel Kahneman. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

[11] Nicholas Barberis. A model of casino gambling. *Management Science*, 58:35 – 51, 20112.

[12] Alexandra B Morrison, Andrew RA Conway, and Jason Chein. Primacy and recency effects as indices of the focus of attention. *Front Hum Neurosci*, 8(6), 2014.

[13] Ethereum (eth) historical data — coinmarketcap. https://coinmarketcap.com/currencies/ethereum/historical-data/.