University of Brasília

Institute of Exact Sciences

Department of Statistics

Master's Dissertation

Bayesian Optimization for Stake Allocation in Soccer Betting

by

Marcos Augusto Daza Barbosa

Brasília, November 2024

Bayesian Optimization for Stake Allocation in Soccer Betting

by

Marcos Augusto Daza Barbosa

Dissertation submitted to the Department of Statistics at the University of Brasília in fulfilment of the requirements for obtaining the Master Degree in Statistics.

Advisor: Prof. Dr. Guilherme Souza Rodrigues

Brasília, November 2024

To God, family and friends, whose support has been fundamental throughout this journey.

Acknowledgments

Thanks to the faculty of PPGEST/UnB, my advisor Dr. Guilherme Souza Rodrigues, all professors and colleagues who contributed to this work in any way. Also thanks to Prof. Eduardo Yoshio Nakano and Prof. João Batista de Morais Pereira (DME/UFRJ) for accepting the invitation to be part of this work's evaluation committee.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Resumo Expandido

OTIMIZAÇÃO BAYESIANA PARA ALOCAÇÃO DE APOSTAS EM FUTEBOL

Este trabalho apresenta métodos de otimização para alocar apostas em futebol, equipado com uma sólida modelagem probabilística dos eventos, para maximizar o retorno ajustado ao risco.

O crescimento das plataformas de apostas online facilitou o acesso ao mercado, mas as *odds* oferecidas incluem margens que garantem a lucratividade das casas, reduzindo o valor esperado para o apostador. Nesse cenário, o objetivo deste projeto é identificar e explorar ineficiências de mercado, utilizando uma abordagem de otimização baseada em probabilidades.

Diferente da maioria dos estudos, que se concentram em prever desfechos limitados, como vitória, empate ou derrota, este trabalho está equipado com a modelagem da superfície completa de probabilidades, representando todos os possíveis resultados de uma partida. Essa superfície, parametrizada a partir das *odds* oferecidas pelas casas de apostas, abrange uma variedade de mercados, como *Correct Score*, *Spread*, *Over/Under* e *Both Score*. Com essa modelagem, é possível explorar de maneira robusta diversas oportunidades de apostas, maximizando o uso das probabilidades implícitas nas *odds* e proporcionando uma diversificação mais eficiente.

A principal contribuição deste projeto é a otimização da alocação de apostas com base nessa superfície de probabilidades, por meio de diferentes funções objetivo e maneiras de obter a

distribuição dos retornos.

O processo de otimização utiliza métodos numéricos para determinar a melhor distribuição das apostas de acordo com a função objetivo predefinida. A alocação é feita não apenas para um jogo, mas simultaneamente para múltiplos jogos e mercados em um mesmo dia, para mÃºltiplas casas de apostas, diversificando a carteira e permitindo o gerenciamento do risco associado.

Uma das principais métricas utilizadas é a Sharpe Ratio, que mede o retorno ajustado ao risco. O Sharpe Ratio leva em conta o retorno esperado e a volatilidade dos retornos, permitindo que o sistema ajuste a alocação de apostas para maximizar o ganho potencial enquanto minimiza a volatilidade.

Outro critério é o Limite Inferior de Confiança (LCL), uma variação do Sharpe Ratio, que introduz o parâmetro λ para equilibrar expectativa e risco. A otimização LCL ajusta o peso dado ao risco, permitindo maior controle sobre a volatilidade; λ é sintonizado para que a alocação de apostas busque maximizar a expectativa de retorno com o controle do risco, priorizando maior estabilidade ou retornos mais altos, conforme o caso.

Para além das abordagens de objetivo único, foi implementada uma otimização multiobjetivo que considera simultaneamente a expectativa e o desvio padrão dos retornos. Essa abordagem utiliza Multi-Objective Bayesian Optimization (MOBO) para encontrar a fronteira de Pareto, um conjunto de soluções onde nenhum dos objetivos (expectativa de retorno e risco) pode ser melhorado sem sacrificar o outro.

Além disso, o sistema conta com um procedimento de alocação percentual variável para definir o percentual do orçamento a ser apostado. Esse percentual é tratado como uma variável adicional no processo de otimização, o que permite ajustar dinamicamente a exposição ao risco em função do desempenho do modelo e das condições do mercado de apostas.

Por fim, também foi incorporada uma técnica de otimização de retornos a longo prazo por meio de simulação de Monte Carlo, aplicando o Sharpe Ratio, para o retorno simulado em um horizonte de múltiplas rodadas de apostas, em um mesmo dia.

Os métodos empregados neste estudo utilizaram a Otimização Bayesiana para as tarefas

de otimização. Essa escolha se deu devido à eficácia da Otimização Bayesiana em explorar espaços de soluções complexos e encontrar pontos ideais de alocação de apostas, especialmente em situações de alta dimensionalidade e incerteza.

Os experimentos foram conduzidos com dados de apostas coletados entre 2019 e 2024, aplicando o método de otimização selecionado a partir da superfície de probabilidades. Os testes indicaram que a aplicação do Sharpe Ratio e de outras funções objetivo resultou em multiplicações significativas do capital inicial.

Assim, este estudo demonstra que uma combinação de modelagem probabilística e otimização multiobjetivo oferece uma abordagem promissora para desenvolver uma estratégia de apostas lucrativa e sustentável a longo prazo. Embora desafios práticos, como o monitoramento em tempo real e restrições de apostas, devam ser considerados, os resultados indicam que a abordagem pode maximizar o retorno ajustado ao risco, conceitualmente e teoricamente, aproveitando as ineficiências do mercado de apostas esportivas.

Palavras-chave: Otimização Bayesiana. Apostas. Futebol. Arbitragem. Sharpe Ratio. Multi-Objetivo.

Abstract

This study explores optimization methods for football betting allocation using probabilistic modeling to maximize risk-adjusted returns. As online betting markets have become increasingly accessible, bookmakers include profit margins in odds, reducing expected value for bettors. This project aims to detect and leverage market inefficiencies through a comprehensive probability surface covering multiple markets, such as Correct Score, Spread, Over/Under, and Both Teams to Score, parameterized from bookmaker odds. The primary contribution is optimizing betting allocation across these markets with various objective functions. Key metrics include the Sharpe Ratio and Lower Confidence Limit (LCL), enabling maximization of gains while managing risk. In one proposed method, a parameter representing the percentage of the budget to be bet is included directly within the optimization process, allowing dynamic adjustment of capital allocation based on model performance and market conditions. Additionally, a multi-objective Bayesian approach (MOBO) provides a Pareto frontier of solutions that balance return and risk simultaneously. We demonstrated significant gains, using Brazilian Serie A data, showing that combining probabilistic modeling with optimization methodsand led to high multiplication of initial capital, theoretically.

Keywords: Bayesian Optimization. Bets. Football. Arbitrage. Sharpe Ratio. Multi-Objective.

Contents

List of Tables

List of Figures

Abbreviations and Acronyms

MOBO	Multi-Objective Bayesian Optimization
MO	Multi-Objective
BO	Bayesian Optimization
MC	Monte Carlo
GP	Gaussian Process
SAA	Sample Average Approximation
COBYLA	Constrained Optimization BY Linear Approximations
h2h	Head to Head

Chapter 1

Introduction

Online betting platforms offer a wide range of options for different sports and games. They made it very easy for people to bet from anywhere in the world with just a few clicks.

Bookmakers set betting odds to manage their own risk and ensure profitability, reflecting not just the likelihood of an event occurring but also incorporating a built-in margin. This margin means that the odds offered are not fair representations of the true probabilities of each outcome. Essentially, when a bettor places a wager based on these odds, the potential payout is adjusted so that it is less than what would be expected if the odds were perfectly fair. This system enables most of the time for bookmakers to make a profit regardless of the outcome of the event, as they balance the books by adjusting odds to attract bets on all possible outcomes in a way that covers their liabilities. A final key factor in setting odds is market behavior: when bettors heavily favor one side, bookmakers adjust the odds in response, even if the fundamental probabilities stay constant.

Despite the sophisticated systems used by bookmakers, it remains possible for bettors to find profitability through strategic betting (Hubáček and Šír, 2023; Wheatcroft, 2020; Terawong and Cliff, 2024; Kaunitz, Zhong, and Kreiner, 2017).

Success hinges on adopting methods with a solid mathematical foundation, such as arbitrage

[1] or statistical modeling, which exploit odds discrepancies and inefficiencies. So informed bettors can identify overvalued odds and make calculated bets that defy the bookmakers' margins.

Inefficiencies within the soccer betting market have been identified, suggesting the potential to develop profitable betting strategies, as discussed by Angelini and De Angelis (2019). While a significant portion of existing research in sports betting is concentrated on analyzing a limited range of match outcomes, such as home or away wins in the US National Basketball League (Hubáček, Šourek, and Železnỳ, 2019; Matej et al., 2021) or the more traditional win, draw, or lose outcomes in soccer (Kaunitz, Zhong, and Kreiner, 2017), our study proposes to broaden the scope. Additionally, Mattera (2023) explores binary outcomes in *Over/Under* markets, but like others, does not consider the full spectrum of score possibilities in a soccer match, thus limiting potential betting opportunities.

Our paper extends this research by utilizing the surface of all outcomes in soccer matches. While studies such as those by Matej et al. (2021) consider multiple games simultaneously, they typically restrict their focus to simple outcomes. This paper seeks to expand the betting strategies by incorporating methods that allow for simultaneous consideration of multiple games.

Financial strategies in sports betting also typically employ methods like Modern Portfolio Theory (Markowits, 1952) and the Sharpe Ratio (Sharpe, 1994), alongside popular betting strategies such as the Kelly Criterion (Kelly, 1956) and its variations like the "fractional Kelly" (MacLean, Ziemba, and Blazenko, 1992). Building on these foundations, our study will mainly employ Bayesian Optimization (BO) to tackle the stake allocation problem.

The accuracy of predictions in bet exchanges, which often offer favorable conditions for discerning bettors, has been studied by Franck, Verbeek, and Nüesch (2010). He points out that selecting bets where the offered odds are more generous than those calculated by bookmakers can lead to profitable outcomes. This principle is supported by the findings of Mattera (2023),

[1] In betting, arbitrage refers to taking advantage of odds discrepancies across different bookmakers to secure a guaranteed profit, regardless of the outcome of the event. This is achieved by placing bets on all possible outcomes of an event across various bookmakers, where the odds are set such that the combined bets cover all outcomes profitably.

which advocate for a "value betting" strategy. According to this strategy, a bet is deemed worthwhile only if the probabilistic forecast (considered as the true likelihood of an event occurring) is greater than the probability implied by the bookmaker's odds. This approach aims to exploit the discrepancies between predicted probabilities and those implied by the odds, targeting long-term profitability. That principle is central to our project.

In this paper, we present a system for distributing stakes of bets for soccer for the Brazilian Serie A. Our novel approach has the following components:

First, we use the odds set by bookmakers for probabilistic modeling the outcomes for a given soccer match via a parametric model that describes the entire probability surface of the scores (Porfírio, 2023) [2]. While the majority of existing research in soccer betting focuses on constructing predictive models that leverage historical match data, including external variables such as location and weather conditions (Hubáček and Šír, 2023; Langseth, 2013; Maher, 1982; Dixon and Coles, 1997), this paper adopts a markedly different approach. Instead of relying on historical data, our methodology utilizes publicly available odds from bookmakers to estimate the true probabilities of match outcomes. Furthermore, the prevailing trend in soccer betting research centers on predicting the probabilities of a small set of outcomes, e.g., home team win, draw, or away team win. However, that approach overlooks the complexity associated to the multiplicity of possible scores. In contrast, this paper considers a surface of estimated probabilities associated to all outcomes of a soccer match.

This approach not only includes the basic match results but also explores a variety of other potential bets such as *Correct Score*, *Spread*, *Over/Under* and *Both Score* markets. This broader perspective allows us to develop more sophisticated betting strategies that are optimized to exploit inefficiencies across a wider array of betting markets.

Secondly, most studies involve the application of betting strategies on a game-by-game

[2] The work of Porfírio (2023) was developed as part of the same research project as this dissertation. The projects are complementary, conceived jointly, with partially overlapping development timelines. While Porfírio (2023) focuses on modeling the probabilities, this dissertation takes these probabilities as input and focuses on developing a strategy for bet allocation.

basis, typically in a sequential manner, i.e., focus on individual matches rather than considering a collection of games simultaneously. Our approach diverges by addressing the bet distribution that takes into account multiple matches concurrently. The exploration of all conceivable match scores and the consideration of multiple games provides ample opportunity because now we can have access to a vast set of betting market options.

In addition to considering multiple games and markets, it is also pertinent to address the use of multiple bookmakers. This approach facilitates the exploration of negatively correlated bets, enabling a more sophisticated form of soft arbitrage.

Lastly, given accurate probability estimates, the task now transitions to calculating the optimal stake allocation for a set of bets. This constitutes a numerical optimization problem, which is inherently complex due to its multi-dimensional nature and the dependency on the choice of the objective function. There are numerous potential objectives one could select, ranging from maximizing returns to minimizing risk, and this problem is further complicated when considering multiple objective functions simultaneously. It's important to note that employing a multi-objective is a complex strategy, capable of providing a more holistic view of the stakes allocation under varying criteria.

We selected the Sharpe Ratio (but not only) as our cornerstone measure to be maximized. This measure is particularly useful in contexts like betting where the objective is to maximize returns while considering the volatility of those returns. The Sharpe Ratio (Sharpe, 1994) inherently captures the trade-off between profit expectation and variance, providing a robust theoretical foundation to optimize our betting distribution.

To address the task of finding the optimal stake allocation, we will employ mainly Bayesian Optimization (BO). This method is particularly effective for complex optimization problems, like ours, that involve multiple dimensions. BO is well-suited for high-dimensional spaces because it efficiently navigates the search space by strategically choosing the next points to evaluate based on a probabilistic model. This model predicts the performance of different stake allocations without needing to exhaustively search every possible combination, which is crucial in multi-dimensional settings where the search space expands exponentially with each added dimension.

Our research is predicated on the hypothesis that consistent application of this strategy will not only validate an efficient allocation in soccer betting but also demonstrate a conceptually systematic way to achieve profit over extended periods. Nevertheless, it is important to note that the proposed method would face significant challenges in real-world operations. Technically, as it will be presented in following chapters, we are going to consider dozens of bookmakers, which would demand a sophisticated software system capable of handling real-time deposits and withdrawals across multiple games simultaneously. Also, operationally, bookmakers would impose limits on maximum bets and often block accounts with high winnings. Furthermore, from an accounting perspective, this study will not take into consideration any taxes, fees, or other costs that might be imposed by either the bookmakers or the government.

In summary, our approach is built upon two components: (1) statistical modeling and (2) optimization techniques to calculate the stake allocation for soccer bets with the goal to capitalize on market inefficiencies.

The rest of the paper is organized as follows. In the following section, we present background theory for modelling the probabilities (probability surface and the parametric model that describes it) and betting markets available for soccer. Chapter 3, we present the optimization methods, strategies and all hyperparameters that were critical for the tasks. Finally, in Chapter 4, we present the results of all experiments.

Chapter 2

Background

2.1 Odds

The term odds, in the context of probability, translates to "chance" and expresses how many times an outcome is more likely than its complement. According to Giolo, 2017, Definition 1 is provided as follows:

Definition 1. The chance of occurrence of an event "A" of interest (or chance of success) is given by

$$\mathcal{O}_A = \frac{P(A)}{1 - P(A)} = \frac{\text{Probability of event A occurring}}{\text{Probability of event A not occurring}}$$

For betting houses, the relationship between odds and probability is inverse to Definition 1. This point is very important for the betting house to protect its profit by paying out smaller amounts for more likely outcomes. Furthermore, the odds from betting houses are increased by 1 to ensure that the winning bettor receives at least the amount they bet.

In the context of football matches, the team with the higher probability of winning will present the lowest *odds* for victory, such that the event of interest for calculating this *odd* is that the team does not win. That is, let $Home$ be the event that describes the home team winning, when adapting Definition 1, it follows that the *odd* in favor of the home team is given by

$$\begin{aligned}\mathcal{O}_{Home} &= 1 + \frac{1 - P(Home)}{P(Home)} \\ &= \frac{P(Home) + 1 - P(Home)}{P(Home)} \\ &= \frac{1}{P(Home)}\end{aligned}$$

where $P(Home)$ is the probability of the home team winning the match. Thus, the odds offered by betting houses are given by Definition 2.

Definition 2. The odds (chance) for the occurrence of an event "A" of interest, as provided by betting houses, are given by

$$\mathcal{O}_A = \frac{1}{P(A)}$$

2.2 Betting Markets

2.2.1 Listing

In terms of sports betting, there are several markets to place bets on a match. Due to the scope of this study, the selected markets are only those that directly contribute to the reconstruction of the probability surface of the possible outcomes of each match, that is:

1. *Head to Head* (h2h): odds for the match result (home team win, away team win, or draw);

2. *Spread*: odds for the match result with a disadvantage for one of the teams. "spread +x/-x" means the fictitious scenario in which the home team starts the match with a “x” goal advantage. It is possible to bet on the home team win, draw, and away team win, always under this fictitious scenario;

3. *Over/Under*: odds for the sum of goals from the home and away teams, where you bet on a sum greater or less than the one set by the betting house;

4. *Correct Score*: odds for the exact result of the match. You bet on the match score like "1-1", "2-1", etc.

5. *Both Score*: odds for both teams to score at least one goal in the match or for at least one of the teams not to score any goals in the match.

2.2.2 Submarkets and Scenarios

Some of the presented markets are directly related to the odds for a given scenario of the match, that is, under the h2h market, betting houses present odds for the home team's victory scenario. This study nicknamed these cases as scenarios of betting markets.

Some of the presented markets are only related to the odds for a given scenario of the match through some kind of "submarket", for example, the Spread market first relates to the advantage or disadvantage for one of the teams, then betting houses present odds for the home team's victory scenario. This study nicknamed these cases as submarkets, so submarkets are the sum of goals from the *Over/Under* market and the advantages and disadvantages granted by the Spread market.

2.2.3 Probabilistic Properties

The events described by the *Head to Head* market cover all possible outcomes in a football match (home win, draw, and away win). Therefore, if the odds in this market have a probabilistic correspondence, then these probabilities should sum to exactly 100%. However, they need to be normalized due to the betting houses having a profit margin on the odds.

Conveniently, all markets share the property that their probabilities corresponding to the odds should sum to 100%, including the *Over/Under* market, because the set scores are always between two integers, such as 2.5, 3.5, etc. In other words, when studying a market individually, the odds of the market under study can be analyzed as a discrete probability distribution.

2.3 Normalization

Previous studies indicate that the probability estimates provided by betting houses through odds do not satisfy the condition that the sum of probabilities of all possible events equals 1, as the betting houses include their profit margins in the odds themselves (Koning and Zijm, 2023). Therefore, normalization techniques have been proposed to ensure that this sum equals 1.

The study by Koning and Zijm, 2023 points out two normalization methods: the multiplicative method, which involves dividing the probabilities found by their sum, and the Shin method proposed by Shin, 1992 and simplified by Clarke, Kovalchik, and Ingram, 2017 to an equivalent form. The technique adopted by this study is the Shin method.

The Shin method involves assuming a proportion of insiders, individuals who always make the correct bet, and applying normalization considering this proportion.

2.4 Probability Surface Construction Using Public Odds

In this paper, we explore and implement a structured betting strategy for soccer matches in Brazilian Serie A. The approach is built upon two fundamental components.

The cornerstone of our methodology involves leveraging publicly available betting odds to construct a probability surface. This surface describes the various possible outcomes of a soccer match. Betting odds, which reflect the aggregated insight of the betting market, are converted into probabilities that provide a detailed statistical representation of expected match outcomes. In our research, we will benefit from estimated outcome probabilities derived through statistical modeling conducted via the work of Porfírio, 2023.

We suppose that we have a $7 \times 7 = 49$ outcomes with unknown probabilities associated $p_{(i,j)}$ where $i = 0, \ldots, 6+$ represents the possible number of goals of the home team and $j =$

$0, \ldots, 6+$ for away team, where $6+$ represents the event 6 or more goals. Those probabilities are unknown, so we work with estimates $\hat{p}_{(i,j)}$.

Bookmakers offer odds that represent how much the stake is multiplicated if you actually win the bet. For example, if the bookmaker offers the odds of 2.2 to home team win, the bettor that puts 100 BRL for that outcome wins $100 \times 2.2 = 220$ BRL if that outcome occurs. If not, he will lost all the stake allocated.

2.4.1 Dataset

The dataset was organized into 2 json files, one file organizes information about the matches by league, and the other organizes the odds. Both the match information and the odds present data available for the years 2019, 2020, 2021, 2022 and 2023 and were collected through *web scraping* from the platform https://oddspedia.com/br.

2.4.1.1 Information

The file containing match information includes the names of the teams that played, the number of goals scored by each team during the match, the stage of the championship in terms of rounds, and the date and time of the match, such that this information can be accessed using unique identifiers for each match.

2.4.1.2 Odds

The odds file for the matches contains the same unique identifiers as the matches in the information file. However, the match identifiers lead to the respective betting houses, which in turn lead to their respective markets available for that match, and the markets lead to the odds.

2.4.1.3 Leagues

The leagues for which odds were collected for this study are some of the most popular in Brazil and around the world, but we will work only Brazilian Serie A.

2.4.2 Probabiliy Surface

2.4.2.1 Concept

For each match in the database, there is a special interest in the *Correct Score* odds, since Definition 2 allows for estimates of probabilities for each possible score of the match. Having estimates of the probabilities of the match outcome allows for an investigation of the consistency of the odds among different betting houses.

For probability surfaces where the X-axis represents the home team's goals and the Y-axis represents the away team's goals, with the values of the axes truncated at 6, generally, the sum of the estimated probabilities already exceeds 1, reflecting the betting house's profit margin, as the betting houses artificially insert a profit margin on the odds. Therefore, the surfaces will be analyzed with clustering, such that X and Y vary in $\{0, 1, 2, 3, 4, 5, 6+\}^2$, where $6+$ means "6 or more goals".

So we work with a family of models capable of providing a concise description of the entire probability surface. In other words, the task becomes the estimation of the model parameters instead of estimating each possible score for the match.

Since not all betting houses offer odds for exact scores, it is noted that the other markets listed in section 2.2 provide partitions of the probability surface, thus also contributing to the reconstruction of the surface.

2.4.2.2 Partitions

Each of the betting markets mentioned in section 2.2 was collected with the intention of providing estimates for the probabilities of each pair $(x, y) \in \{0, 1, 2, 3, 4, 5, 6+\}^2$ possible.

For example, the result of a draw describes pairs (x, y) such that $x = y$, therefore, the odds for a draw describe the sum of the probabilities along the diagonal of the surface, when $x = y$, as per Definition 2. Another case is the victory of the home team, which describes pairs (x, y) such that $x > y$, i.e., the odds for the home team victory describe the sum of all probabilities of the cells on the surface below the diagonal of the draw.

Similarly, for other betting markets, Figure 2.1 illustrates in blue which partitions of the surface are described by the odds for each available market. The total number of goals chosen was 2.5, and the *spread* was a 2-goal disadvantage for the home team.

Each betting market can be interpreted as one or more equations that help describe the probability surface. Under the notation where p_{ij} is the probability of the home team scoring i goals and the visiting team scoring j goals, and $\mathcal{O}_{Market}$ is the odd for a given betting market, it follows from Figure 2.1 that

$$\begin{cases}
\frac{1}{\mathcal{O}_{Draw}} & = \sum_{i=0}^{6} p_{ii} \\
\frac{1}{\mathcal{O}_{Home}} & = \sum_{i=0}^{6} \sum_{j=0}^{6} p_{ij} \cdot \mathbb{1}_{(i>j)} \\
\frac{1}{\mathcal{O}_{Visitor}} & = \sum_{i=0}^{6} \sum_{j=0}^{6} p_{ij} \cdot \mathbb{1}_{(j<i)} \\
\frac{1}{\mathcal{O}_{Over}} & = \sum_{i=0}^{6} \sum_{j=0}^{6} p_{ij} \cdot \mathbb{1}_{(j+i>T)} \\
\frac{1}{\mathcal{O}_{Under}} & = \sum_{i=0}^{6} \sum_{j=0}^{6} p_{ij} \cdot \mathbb{1}_{(j+i<T)} \\
\frac{1}{\mathcal{O}_{Spread_{Home}}} & = \sum_{i=0}^{6} \sum_{j=0}^{6} p_{ij} \cdot \mathbb{1}_{(i+S>j)} \\
\frac{1}{\mathcal{O}_{Spread_{Draw}}} & = \sum_{i=0}^{6} \sum_{j=0}^{6} p_{ij} \cdot \mathbb{1}_{(i+S=j)} \\
\frac{1}{\mathcal{O}_{Spread_{Visitor}}} & = \sum_{i=0}^{6} \sum_{j=0}^{6} p_{ij} \cdot \mathbb{1}_{(i+S<j)} \\
\frac{1}{\mathcal{O}_{(Yes)Both\ Score}} & = \sum_{i=1}^{6} \sum_{j=1}^{6} p_{ij} \\
\frac{1}{\mathcal{O}_{(No)Both\ Score}} & = \sum_{j=1}^{6} p_{0j} + \sum_{i=0}^{6} p_{i0}
\end{cases}$$

where T is the total number of goals set (2.5 for Figure 2.1), S is the spread advantage or

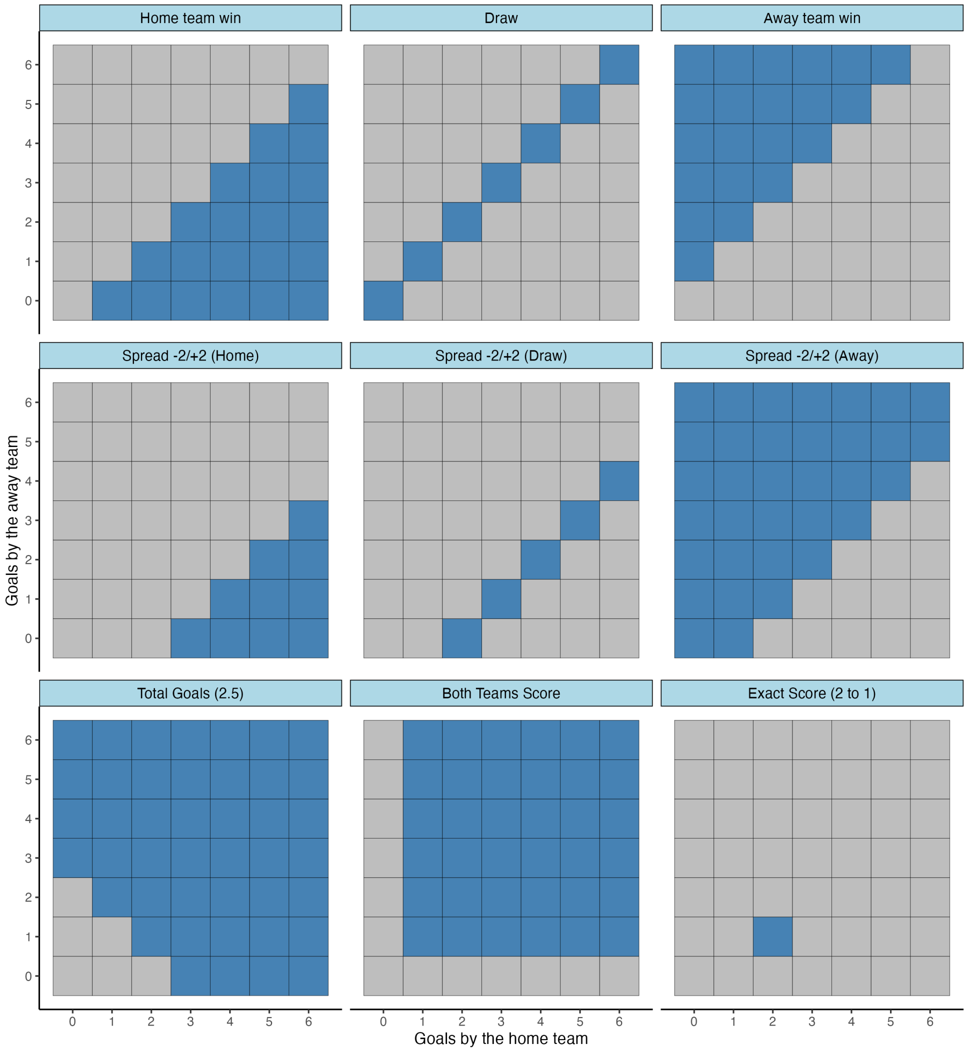


Figure 2.1: Set of outcomes whose probabilities are described by the odds of each betting market, in case the positive outcome materializes.

disadvantage for the home team (2 goals against the home team, for Figure 2.1) and $\mathbb{1}_{(condition)}$ is the Indicator function, presented in Equation (3.2). Combined with the probabilities provided by the odds for each cell by the *Correct Score* market, it follows that all equations contribute to the reconstruction of the surface.

$$\mathbb{1}_{(condition)} = \begin{cases} 1, & \text{if condition is satisfied} \\ 0, & \text{otherwise} \end{cases} \tag{2.1}$$

2.4.2.3 Vector Representation

Figure 2.1 presents the probability surfaces on cartesian planes, suggesting that the surface could be represented by a matrix. Each market can be represented by a matrix containing only values 0 or 1, 0 if the market does not pertain to a certain probability and 1 otherwise. For example, from Figure 2.1, the matrix $\boldsymbol{M}$ for the scenario "home team victory" is given by

$$\boldsymbol{M} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Consider the representation of the elements of a matrix M by M_{ij}, where $i, j \geq 0$. Then it would make sense that the index M_{10} represents the score 1 to 0 for the home team. Therefore, it follows that the matrix $\mathbb{M}$ for the scenario "home team victory" is given by

$$\mathbb{M} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

However, it is possible to eliminate the bidimensionality of the matrix representation and adopt a vector representation in which the scores follow 0 to 0, 0 to 1, 0 to 2,..., 1 to 0, 1 to 1, ..., 6+ to 5, 6+ to 6+. That is, the scenario "home team victory" can be represented by the vector $\mathbb{V}$ in the form

$$\mathbb{V} = \text{vec}\left(\mathbb{M}^T\right)$$

From this section, the vector corresponding to a given scenario of a given market will be denoted by the notation $\mathbb{V}_{[scenario]}$, such that the matrix representation of the market is composed of all its scenarios. For example, for the head-to-head (h2h) market, which includes the home team victory, visitor team victory, or draw, its matrix representation is given by

$$\mathbb{M}_{[h2h]} = \begin{bmatrix} \mathbb{V}^T_{[home]} \\ \mathbb{V}^T_{[draw]} \\ \mathbb{V}^T_{[visitor]} \end{bmatrix}$$

For markets with submarkets, such as *spread*, it is necessary to specify the submarket, such as *spread* $+2/-2$, then

$$\mathbb{M}_{[spread(+2/-2)]} = \begin{bmatrix} \mathbb{V}^T_{[home]} \\ \mathbb{V}^T_{[draw]} \\ \mathbb{V}^T_{[visitor]} \end{bmatrix}$$

in both cases, the vectors $\mathbb{V}$ are respective to the respective market to which the matrix $\mathbb{M}$ corresponds.

2.4.3 Parametric Model

2.4.3.1 Parameterization

Based on previous studies and available data, there is an understanding of a dependency structure between the number of goals scored by the home team and the visiting team. Therefore, this study proposes researching a model capable of fitting this dependency structure through the use of copulas to generate correlated observations, in conjunction with a mixture of the families of Negative Binomial distributions, responsible for the general model of the surface, and Poisson distributions, responsible for addressing the potential underestimation of probabilities for draw outcomes.

According to past studies (Mchale and Scarf, 2006), the choice of the Negative Binomial for the marginal distributions allows greater flexibility for the behavior of the model, as there are 2 parameters for each to give marginals, while for the inflation of the diagonal, the Poisson distribution seems to be flexible enough (Karlis and Ntzoufras, 2003).

The model is composed of the parameters:

1. r_X, p_X = Parameters of the Negative Binomial for the number of goals scored by the home team;

2. r_Y, p_Y = Parameters of the Negative Binomial for the number of goals scored by the visiting team;

3. ρ = Correlation parameter between the marginal distributions;

4. ϵ = Parameter regulating the mixture of distributions by including the distribution that inflates the diagonal of the surface.

5. α = Parameter of the Poisson distribution, describing the inflation of probabilities on the surface diagonal (draw outcomes);

2.4.3.2 Data Generating Process

Once the model parameters are known, the proposed model is obtained from the data-generating process:

Algorithm 1: Data Generating Process

```
Input:
• $\mathbf{\Theta} = (r_X, r_Y, p_X, p_Y, \rho, \epsilon, \alpha)$;
Output:
• Generated data instance.

Compute:
Generate a sample Z, such that Z $\sim$ Bernoulli($\epsilon$)
if $Z = 0$ then
    Generate W, such that W $\sim$ Poisson($\alpha$)
    Obtain $\boldsymbol{W} = (W, W)$
else
    if $Z = 1$ then
        Generate the vector $(X, Y)$, such that $(X, Y) \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$
        Obtain $(U, V) = (\Phi(X), \Phi(Y))$
        Obtain $\boldsymbol{W} = (W_1, W_2) = (F^{-1}(U|r_X, p_X), F^{-1}(V|r_Y, p_Y))$
$\boldsymbol{W} = (\min\{W_1, 6\}, \min\{W_2, 6\})$
```

where $\boldsymbol{W}$ is the vector obtained as a sample, and F^{-1} is the quantile function of the Negative Binomial. The proposed model is a joint probability distribution $P(W_1 = w_1, W_2 = w_2)$, where the support of this probability distribution is given by

$$(W_1, W_2) \in \{0, 1, 2, 3, 4, 5, 6+\}^2$$

In other words, the support of the distribution of match scores under the proposed model involves only 49 pairs with non-zero probability.

2.4.3.3 Score Distribution

It is known that the data generating process presented is a mixture of a Poisson distribution and another distribution whose form is not obvious to the researcher but will be referred to as the Score distribution. Analogously, the mixture with the Poisson distribution will be called the Adjusted Score distribution.

The data generating process proposed in Section 2.4.3.2 is computationally inefficient for the parameter fitting method to be discussed in Section 2.4.3.4.

To improve computational efficiency, we considered writing an analytical form for the Score distribution. The fact that the support of the Score distribution involves only 49 pairs suggests expressing its probability function in terms of a Bivariate Normal distribution.

Considering only the Score distribution, suppose that an iteration of the data generating process resulted in a vector (w_1, w_2). It is known that this vector was obtained from the quantile function of the Negative Binomial distribution applied to certain values (u, v). However, it is not possible to determine the values of the vector (u, v) because the quantile function of the Negative Binomial distribution is not injective. On the other hand, it is possible to determine the smallest value u_1 for which $F^{-1}(u_1|r_X, p_X) = w_1$ and, analogously, the largest value u_2 for which $F^{-1}(u_2|r_X, p_X) = w_1$. Similarly for w_2, for $(w_1, w_2) \in \{0, 1, 2, 3, 4, 5, 6+\}$, we have

$$\begin{aligned}
&P(W_1 = w_1, W_2 = w_2) = \\
&= P(u_1 < U < u_2, v_1 < V < v_2) \\
&= P(\Phi^{-1}(u_1) < \Phi^{-1}(U) < \Phi^{-1}(u_2), \Phi^{-1}(v_1) < \Phi^{-1}(V) < \Phi^{-1}(v_2)) \\
&= P(\Phi^{-1}(u_1) < X < \Phi^{-1}(u_2), \Phi^{-1}(v_1) < Y < \Phi^{-1}(v_2))
\end{aligned}$$

Since W was defined from the quantile function of the Negative Binomial distribution, it is possible to determine (u_1, u_2) from the inverse of the quantile function, that is, the cumulative distribution function applied in the form

$$\begin{cases} u_1 = P(W_1 \leq w_1 - 1|r_X, p_X) \\ u_2 = P(W_1 \leq w_1|r_X, p_X) \end{cases}$$

the process for determining (v_1, v_2) is analogous.

When any of the components of the vector (w_1, w_2) is 0 or 6, one must be careful with the cumulative distribution function of the Negative Binomial. Therefore, it is convenient to define the random variables G and H such that

$$H \sim \text{Negative Binomial}(r, p)$$

$$P(G \leq g|r, p) = \begin{cases} P(H \leq g|r, p) & , \text{ if } g < 6 \\ 1 & , \text{ if } g \geq 6 \end{cases}$$

That is, the task of writing the Score distribution in an analytical form was not concluded; nevertheless, an expression in terms of the Bivariate Normal distribution was reached, which is computationally more efficient than the data generating process described in Section 2.4.3.2.

Returning to the Adjusted Score distribution, for the Poisson distribution responsible for inflating the diagonal, it is convenient to define the random variables C and D, such that

$$C \sim \text{Poisson}(\alpha)$$

$$P(D = k|\alpha) = \begin{cases} P(C = k|\alpha), \text{ if } k \neq 6 \\ P(C \geq 6|\alpha), \text{ if } k = 6 \end{cases}$$

Finally, it is possible to obtain the probability function of the Adjusted Score distribution for each possible pair $(w_1, w_2) \in \{0, 1, 2, 3, 4, 5, 6+\}$ through Algorithm 2.

```
Algorithm 2: Calculation of the Adjusted Score Distribution
  Input:
  • $\Theta = (r_X, r_Y, p_X, p_Y, \rho, \epsilon, \alpha)$;
  • Maximum number of goals $M$, always used as 6 in this study.
  Output:
  • Probability vector with $M^2$ entries, incrementing the scores first for the visiting team
    and then for the home team.

  Compute:
  $goals = \{0, 1, 2, ..., M\}$
  $probabilities[(M+1)^2] = \{\}$
  for $w_1 \in goals$ do
      for $w_2 \in goals$ do
          $u_1, u_2 = (P(G \leq w_1 - 1 | r_X, p_X), P(G \leq w_1 | r_X, p_X))$
          $v_1, v_2 = (P(G \leq w_2 - 1 | r_Y, p_Y), P(G \leq w_2 | r_Y, p_Y))$
          $x_1, x_2 = (\Phi^{-1}(u_1), \Phi^{-1}(u_2))$
          $y_1, y_2 = (\Phi^{-1}(v_1), \Phi^{-1}(v_2))$
          $P_{XY} = P(x_1 < X < x_2,\ y_1 < Y < y_2 | \rho)$
          $P_D = P(D = w_1 | \alpha) \mathbb{1}_{(w_1 = w_2)}$
          $probabilities[7w_1 + w_2] = \epsilon P_{XY} + (1 - \epsilon) P_D$
  return $probabilities$
```

The Algorithm 2 highlights that the probabilities of the proposed model can be obtained from the mixture of a Bivariate Normal distribution and a Poisson distribution. The implementation of Algorithm 2 proved to be more than 1000 times computationally more efficient than calculating the probabilities based on the data generation process described in Section 2.4.3.2.

2.4.3.4 Model Parameter Estimation

Through the probabilistic property of betting markets described in Section 2.2.3 and the equations outlined in the system of equations 2.4.2.2, it follows that it is possible to use the equations of a market to obtain the corresponding probability provided by the proposed model. This results in a probability distribution provided by the betting house odds and another probability distribution provided by the proposed model.

Both obtained probability distributions provide probabilities for the same events of interest. Thus, it is possible to compare these probability distributions by calculating the Kullback-Leibler Divergence (D_{KL}). Through numerical derivatives, it is possible to determine how the model parameters should be adjusted to minimize D_{KL}. After multiple iterations, it is expected that the probabilities provided by the model will closely match those found through the odds.

The probability function resulting from Algorithm 2 describes probabilities for 49 possible cases, which can be allocated in a vector $\mathbb{T}$, similar to the case in Section 2.4.2.3. The advantage of this representation is that for a given market, the probabilities corresponding to each scenario are given by

$$\mathbb{M}_{[market]} \cdot \mathbb{T} = \begin{bmatrix} \text{Probability of scenario 1} \\ \text{Probability of scenario 2} \\ \vdots \\ \text{Probability of scenario n} \end{bmatrix} = \mathbb{Q}_{[market]}$$

Allocate the probabilities obtained from the odds, from a given betting house, for a given market, in the vector $\mathbb{P}_{[market]}$. Then,

$$D_{KL} = \sum_{i=1}^{n} \mathbb{P}_{[market]i} \log \left(\frac{\mathbb{P}_{[market]i}}{\mathbb{Q}_{[market]i}} \right),$$

where $\mathbb{V}_i$, $i = 1, 2, 3, ..., n$, is the i-th component of any vector $\mathbb{V}$.

Fixing the betting house, it is possible to calculate D_{KL} for all the markets that the betting house provided and obtain the average $\bar{D}_{KL}$ among all the divergences obtained. The process of estimating the model parameters follows similarly to the described single-market case, calculating numerical derivatives and minimizing the distance metric $\bar{D}_{KL}$ between the model distribution and the betting house distribution.

In matrix terms, let the support of the markets be given by

$$\mathbb{C} = \{h2h, spread(+1/-1), spread(+2/-2), ..., over/under(0.5), ..., both\, score, exact\, score\}$$

which may be reduced if a market is not available. Then,

$$\bar{D}_{KL} = \frac{1}{\|\mathbb{C}\|} \sum_{market \in \mathbb{C}} \sum_{i=1}^{n} \mathbb{P}_{[market]i} \log \left(\frac{\mathbb{P}_{[market]i}}{\mathbb{Q}_{[market]i}} \right)$$

This optimization process is described by Algorithms 3 and 4. The R language's optim function was specifically used to implement Algorithm 4, dealing with numerical derivatives and the process of minimizing the Kullback-Leibler Divergence in general.

Algorithm 3: Calculation of Mean Kullback-Leibler Divergence

```
Dependencies:
• Function $P(\mathbf{\Theta}, M)$, which executes Algorithm 2;
• Operator $\bigodot$, for matrix multiplication.
Input:
• $\mathbf{\Theta} = (r_X, r_Y, p_X, p_Y, \rho, \epsilon, \alpha)$;
• Maximum number of goals $M$, always used as 6 in this study;
• $shin_probs$, a dictionary (Python) where markets or submarkets map to probabilities
  normalized by the Shin method at a betting house;
• $mask$, a dictionary (Python) containing markets and submarkets as keys and a binary
  vector as value, as described in section 2.4.2.3;
Output:
• A scalar indicating the Mean Kullback-Leibler Divergence between the proposed
  Adjusted Score Distribution and the betting house.

Compute:
$probabilities = P(\mathbf{\Theta}, M)$
$D_{KL} = 0$
for $market \in shin\_probs$ do
    $market\_probs = probabilities \bigodot mask$
    for $i \in \{0, 1, ..., length(shin\_probs[market])\}$ do
        $D_{KL} \mathrel{+}= shin\_probs[market][i] \cdot \log\left(\frac{shin\_probs[market][i]}{market\_probs[i]}\right)$
$mean\_D_{KL} = \frac{D_{KL}}{length(odds)}$
return $mean\_D_{KL}$
```

Algorithm 4: Adjustment of the Adjusted Score Distribution

Dependencies:

- Function $P(\boldsymbol{\Theta}, M)$, which executes Algorithm 2;
- Function $MDKL(\boldsymbol{\Theta}, shin\_probs, mask)$, which returns the result of Algorithm 3 based on the probabilities already calculated by $P(\boldsymbol{\Theta}, M)$;
- Function $N(odds)$, which provides probabilities normalized by the Shin method from the odds of a market or submarket;
- Function $\nabla(\boldsymbol{\Theta}, Fun)$, where Fun is a function evaluated on the parameter set $\boldsymbol{\Theta}$. The function returns a numerical gradient of the function Fun;
- Function $Patch(\boldsymbol{\Theta}, Fun, grad\_Fun)$, which returns a set of parameters that minimizes the function Fun among the options provided by the gradient $grad\_Fun$;

Input:

- Any initial guess $\boldsymbol{\Theta} = (r_X, r_Y, p_X, p_Y, \rho, \epsilon, \alpha)$;
- Maximum number of goals M, always used as 6 in this study;
- $odds$, a dictionary (Python) in which markets or submarkets carry the odds from a betting house;
- $mask$, a dictionary (Python) containing as keys the markets and submarkets and as value a binary vector, as described in section 2.4.2.3;
- Convergence tolerance $\epsilon > 0$. • $maxit$, the stopping criterion of the algorithm, can be for example the number of updates without improvement.

Output:

- Vector of 7 parameters that describe the probability surface reconstructed through the Adjusted Score Distribution.

Compute:

```
$min\_D_{KL} = \infty$
$probabilities = P(\boldsymbol{\Theta}, M)$
$shin\_probs = N(odds)$
while $no\_improv < maxit$ do
    $mean\_D_{KL} = MDKL(probabilities, shin\_probs, mask)$
    if $mean\_D_{KL} < min\_D_{KL}$ then
        $min\_D_{KL} = mean\_D_{KL}$
        $no\_improv = 0$
    else
        $no\_improv += 1$
    $grad\_Fun = \nabla(\boldsymbol{\Theta}, MDKL)$
    $\boldsymbol{\Theta} = Patch(\boldsymbol{\Theta}, MDKL, grad\_Fun)$
return $\boldsymbol{\Theta}$
```

2.4.4 Average Model

When fitting models as in Section 2.4.3.3, each game, each betting house will present a set of 7 parameters that describe the probabilities according to their provided odds. However, it is expected that the parameters of each betting house are similar because they describe probabilities for the same game scores.

The study by Kaunitz, Zhong, and Kreiner, 2017 proposes, in a simpler scope, to take the average of the probabilities from each betting house for the same result. This average is called "consensus probability" and allows decision-making for bet allocation. Building on this idea of taking the average of measures for decision-making, this study proposes the construction of "consensus surfaces" from a "average model" among those obtained for the available betting houses.

If there are n betting houses available for a given game, then the estimated parameters are obtained as

$$\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_n$$

where θ represents any of the 7 parameters of the model. Assuming that there is a probability surface that truly describes the probabilities of the scores occurring, which can also be described by the model proposed in this study, and that the normalized surfaces obtained from the betting houses approach this real surface, then an idea to estimate the real parameter θ for a given game is to take the estimator

$$\bar{\hat{\theta}} = \frac{\sum_{i=1}^{n} \hat{\theta}_i}{n} \approx \theta$$

where θ represents any of the parameters of the model, that is, the set of estimated parameters is defined as

$$\hat{\Theta} = \left(\bar{\hat{r}}_X, \bar{\hat{r}}_Y, \bar{\hat{p}}_X, \bar{\hat{p}}_Y, \bar{\hat{\rho}}, \bar{\hat{\epsilon}}, \bar{\hat{\alpha}}\right)$$

and the probabilities of the consensus surface can be obtained as in Algorithm 2. For each possible pair $(w_1, w_2) \in \{0, 1, 2, 3, 4, 5, 6+\}$, calculate

$$P\left(W_1 = w_1, W_2 = w_2 | \hat{\Theta}\right)$$

To illustrate the estimation of probabilities across the surface of all outcomes, we can refer to the following figures.

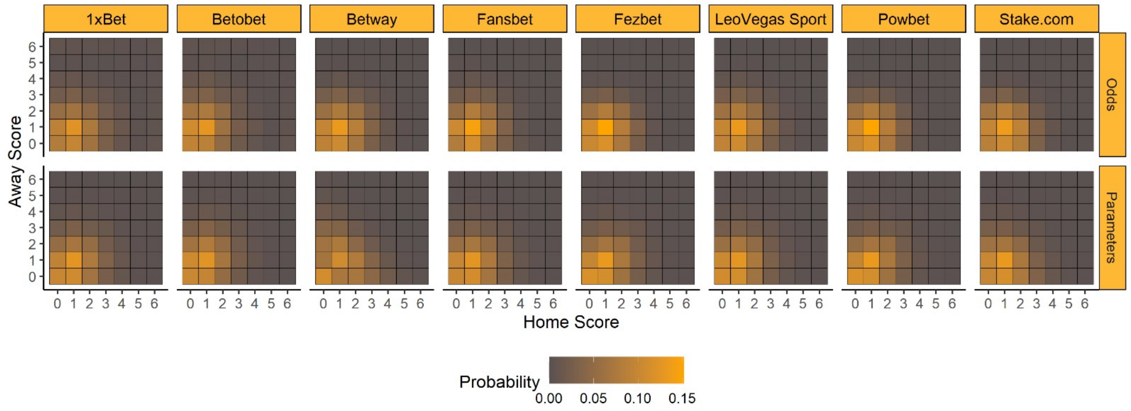


Figure 2.2: Comparisons between some surfaces obtained through the odds for *Exact Score* and surfaces obtained through the estimated parameters for the respective betting house for the game between Juventude and Coritiba for the 35th round of the Brazilian Serie A 2022.

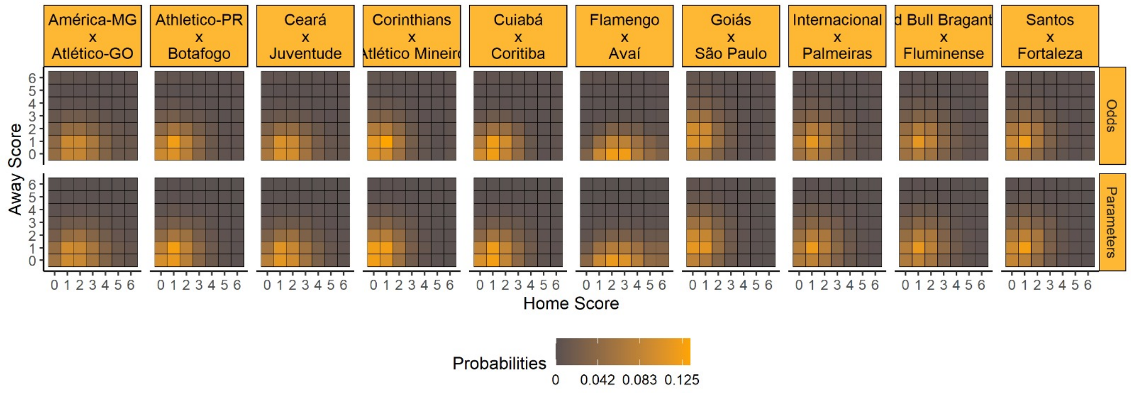


Figure 2.3: Comparisons between some surfaces obtained through the odds for *Exact Score* and surfaces obtained through the estimated parameters for the respective betting house for the last 10 games of the Brazilian Serie A 2022, only for the betting house 1xBet.

The Figures 2.2 and 2.3 show a great similarity between the surfaces from the betting houses and the adjusted surfaces, indicating that the model fitting was successful with respect to the ability to describe the probabilities corresponding to the odds provided by the betting houses through a probability distribution with the 7 proposed parameters.

2.5 Multi-Objective Bayesian Optimization

Multi-Objective Bayesian Optimization refers to using a Bayesian probabilistic approach for an optimization problem with multiple objectives.

Multi-objective (MO) optimization is a powerful approach employed in various fields to address problems with conflicting objectives, where the optimization of one criterion may come at the expense of another. We optimize an objective vector $\boldsymbol{f}(\boldsymbol{x}) : \mathbb{R}^d \rightarrow \mathbb{R}^M$ such that $\boldsymbol{f}(\boldsymbol{x}) = (f^{(1)}(\boldsymbol{x}), \ldots, f^{(M)}(\boldsymbol{x}))$ over a bounded set $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$. Essentially, there is no single solution, so a set of solutions is considered. At the core of this approach lies the concept of Pareto domination. This dominance principle asserts that a solution is superior to another if it achieves better performance in at least one objective without compromising the performance in any other. Consequently, Pareto domination defines a set of non-dominated solutions, forming the Pareto front, which represents the spectrum of trade-offs inherent in the optimization problem.

An objective vector $\boldsymbol{f}(\boldsymbol{x})$ Pareto-dominates $\boldsymbol{f}(\boldsymbol{x}')$, expressed as $\boldsymbol{f}(\boldsymbol{x}) \succ \boldsymbol{f}(\boldsymbol{x}')$, if $f^{(m)}(\boldsymbol{x}) \geq f^{(m)}(\boldsymbol{x}')$, $\forall\, m = 1, \ldots, M$ and exists ate least one $m' \in \{1, \ldots, M\}$ such that $f^{(m')}(\boldsymbol{x}) > f^{(m')}(\boldsymbol{x}')$.

We state that $\mathcal{P}^* = \{\boldsymbol{f}(\boldsymbol{x})\ s.t.\ \nexists\, \boldsymbol{x}' \in \mathcal{X} : \boldsymbol{f}(\boldsymbol{x}') \succ \boldsymbol{f}(\boldsymbol{x})\}$ defines the set of Pareto optimal solutions and $\mathcal{X}^* = \{\boldsymbol{x} \in \mathcal{X}\ s.t.\ \boldsymbol{f}(\boldsymbol{x}) \in \mathcal{P}^*\}$ as the associated Pareto optimal inputs. Equipped with that mathematical system, users have the option to choose a solution that involves a trade-off of objectives, tailored to their priorities.

In the realm of soccer betting, the optimization of a betting strategy presents a multifaceted challenge with the goal of maximizing returns while effectively managing risks. Several key objectives can be considered in the formulation of a multi-objective optimization problem tailored for a betting strategy. Bettors may prioritize risk-adjusted returns by maximizing metrics such as the Sharpe Ratio (Sharpe, 1994) or Sortino Ratio (Sortino and Price, 1994), which both balances return against volatility. The maximization of the probability of profitability could be another objective, emphasizing the likelihood of achieving positive returns. Furthermore, minimizing the maximum drawdown also stands as a posible objective, ensuring resilience against significant losses. The application of the Kelly Criterion (Kelly, 1956), aiming to maximize the expected logarithm of wealth, reflects a strategy that dynamically adjusts bet sizes based on perceived edges. Diversification objectives can be incorporated to maximize the distribution of bets across markets, promoting a well-rounded approach. Consideration of conditional metrics, such as minimizing Conditional Value-at-Risk (CVaR), provides insights into tail risks. Meanwhile, objectives like maximizing the probability of reaching a target wealth or the probabilistic Sharpe Ratio address specific wealth accumulation goals and statistical measures, respectively. The entropy of bet sizes and information entropy objectives introduce elements of diversity and unpredictability, fostering adaptability in decision-making. Combinations of those objectives collectively can be considered at the optimization task that we expect will lead to a robust and profitable betting strategy.

Bayesian Optimization (BO) is a method for optimizing expensive-to-evaluate black-box functions. BO uses a probabilistic surrogate model, typically a Gaussian Process (GP), to provide a posterior distribution $P(\boldsymbol{f} \mid \mathcal{D})$ over the true function values $\boldsymbol{f}$, given the observed data $\mathcal{D} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{n}$.

An acquisition function $\alpha : \mathcal{X}_{\text{cand}} \to \mathbb{R}$ uses the surrogate model to assign a utility value to a set of candidate points $\mathcal{X}_{\text{cand}} = \{\boldsymbol{x}_i\}_{i=1}^{q}$ to be evaluated on the true function. While the true $\boldsymbol{f}$ may be expensive to evaluate, the surrogate-based acquisition function is computationally efficient to optimize, yielding a set of candidates $\mathcal{X}_{\text{cand}}$ to be evaluated on f. The acquisition function delineates where to sample next considering a strategy for balancing the exploration-exploitation trade-off. If gradients of $\alpha(\mathcal{X}_{\text{cand}})$ are available, gradient-based optimization methods can be applied. Otherwise, gradients are either approximated.

Therefore, Bayesian Optimization (BO) has essentially two components: a probabilistic surrogate model and an acquistion function. The Algorithm 5 and Figure 2.4 outlines the steps of BO for a 1-D objective.

Algorithm 5: Basic Pseudo-code for Bayesian Optimization for 1-D objective function

```
Dependencies:
• Set N as the maximum number of iterations;

Place a Gaussian process prior on $f$.
Observe $f$ at $n_0$ initial points using a space-filling experimental design.
Set $n \leftarrow n_0$.
while $n \leq N$ do
    Update the posterior probability distribution of $f$ using all observed data.
    Find $x_n$, the optimizer (maximizer or minimizer) of the acquisition function $\alpha(x)$,
      computed using the current posterior distribution.
    Observe $y_n = f(x_n)$.
    Increment $n \leftarrow n + 1$.
return The solution $x^*$, either the point evaluated with the largest $f(x)$ or the point
  with the largest posterior mean.
```

BoTorch (Balandat et al., 2020) is one of the main python libraries for BO built on top of the popular PyTorch framework and we intensively use it for the optimization tasks. The idea of BO in BoTorch is associated to using Monte Carlo (MC) acquisition functions.

Here, we find it worthwile to provide the background of BO according to the BoTorch paper (Balandat et al., 2020), because that library is the main tool used in this study for the optimization tasks.

Given collected data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$, such that $x_i \in \mathcal{X}$ and $y_i = f_{true}(x_i) + v_i(x_i)$ with v_i representing a noise disturbing f_{true} (the BoTorch paper also extends for the case that f_{true} is multi-output, so $y_i, v_i \in \mathbb{R}^m$). Suppose that a bayesian statistical surrogate model is used to model the objective function f for any $\boldsymbol{x} = \{x_1, \ldots, x_q\}$ which produces a distribution over $f(\boldsymbol{x}) = (f(x_1), \ldots, f(x_q))$ and $y(\boldsymbol{x}) = (y(x_1), \ldots, y(x_q))$. Tipically, the model f is a Gaussian

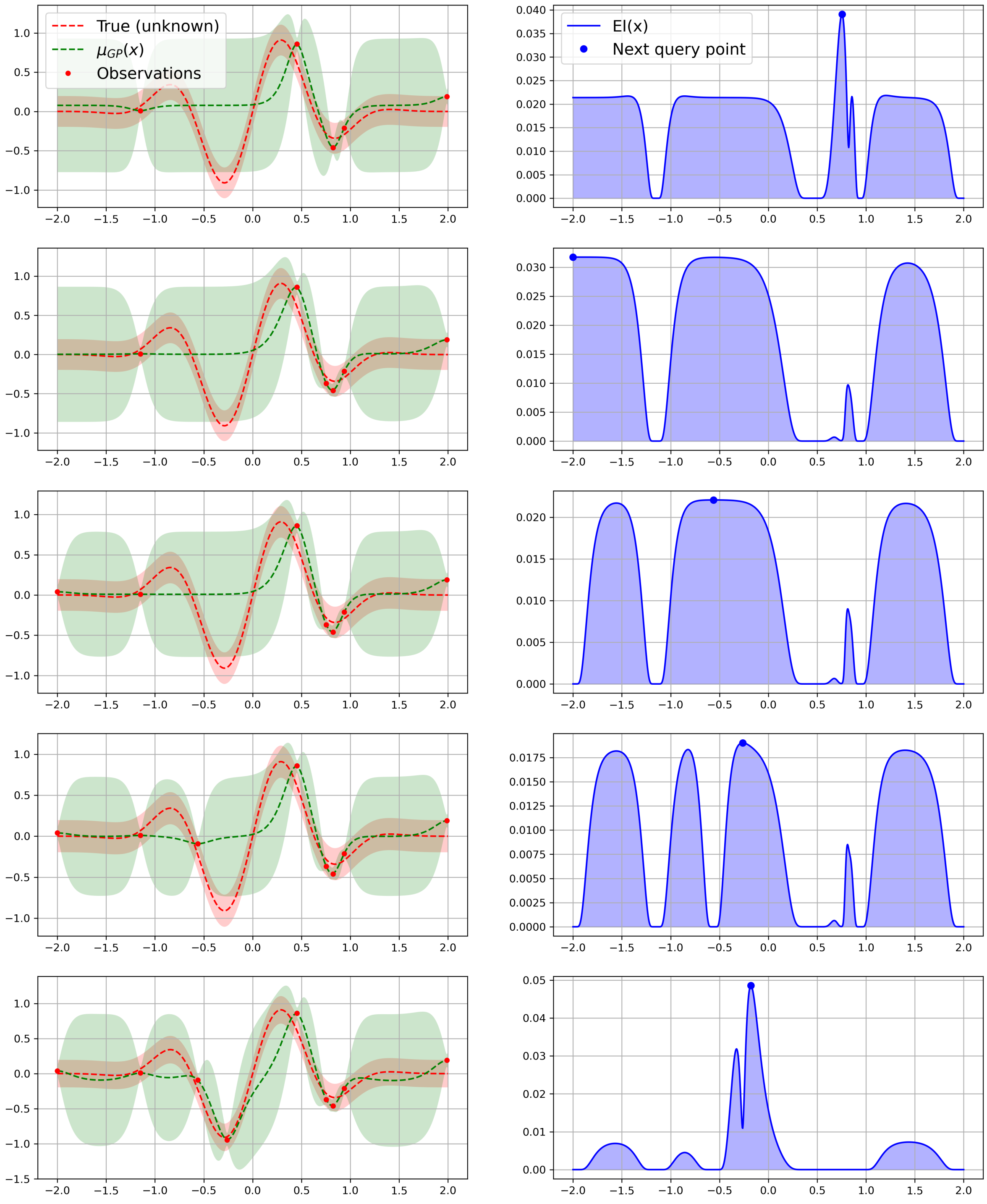


Figure 2.4: Bayesian Optimization with Gaussian Process (GP) for 1-D objective function across five iterations of minimization task. The first column shows the true objective function, the GP that approximates the true objective function and the uncertainty of the approximation. The second column shows the acquisition function Expected Improvement (EI), after every surrogate model fit, with the next point to be queried.

Process (GP) and v_i are independently and identically distributed (i.i.d). In that case, is possible to define the posterior distributions $f_{\mathcal{D}}(\boldsymbol{x})$ and $y_{\mathcal{D}}(\boldsymbol{x})$ conditioned on data $\mathcal{D}$, such that both are multivariate normal distributed. BoTorch relies on a MC approach for the acquisition functions, so there is no assumption regarding about the exact form of those posterior distributions.

Then the procedure optimizes the acquisition function evaluated on $f_{\mathcal{D}}(\boldsymbol{x})$ over the candidate $\boldsymbol{x}$. The acquisition function assesses the potential value resulting from evaluating the objective function at a new point $\boldsymbol{x}$, taking into account the existing posterior distribution over f and can be stated as

$$\alpha(\boldsymbol{x}, \Phi, \mathcal{D}) = E\left[a(g(f(\boldsymbol{x})), \Phi) \mid \mathcal{D}\right], \tag{2.2}$$

where $g : \mathbb{R}^{q\times m} \rightarrow \mathbb{R}^q$ is an objective function, $\Phi \in \boldsymbol{\Phi}$ are parameters independent of $\boldsymbol{x}$ in some set set $\boldsymbol{\Phi}$ and $a : \mathbb{R}^q \times \Phi \rightarrow \mathbb{R}$ is a utility function that defines the acquisition function.

MC integration can be employed to estimate the expectation 2.2 by utilizing samples drawn from the posterior distribution.

$$\hat{\alpha}_N(\boldsymbol{x}; \boldsymbol{\Phi}, \mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} a(g(\xi_D^i(\boldsymbol{x})), \boldsymbol{\Phi}). \tag{2.3}$$

suchat that we compute 2.3 via drawing i.i.d. samples $\xi_D^i\ (\boldsymbol{x})$ (Balandat et al., 2020). In order to create a new candidate set $\boldsymbol{x}$, it is necessary to optimize the acquisition function α. Accomplishing this effectively, particularly in higher dimensions, often involves utilizing gradient information $\nabla_x \alpha(x, \Phi, \mathcal{D})$.

The BO literature has previously explored the Monte Carlo approach for optimizing acquisition functions, often employing stochastic optimization methods. But in the paper Balandat et al., 2020, on the other hand, adopts a unique perspective by focusing on Sample Average Approximation (SAA).

Chapter 3

Method

Using the probability estimates derived from the odds, according to Porfírio (2023), as presented in Chapter 2, we employ numerical optimization methods to find the best distribution of stakes across various betting options. This process aims to optimize a pre-determined objective function. Our approach extends to simultaneous betting on multiple games and markets, which serves to diversify the betting portfolio. Multiple games are considered, because is very common to have soccer matches scheduled for the same day.

The methodology presented in this section examines bet allocation within a single-day and our system systematically replicates the method each day throughout the entire betting period.

Consider an allocation vector $\boldsymbol{\upsilon} = (\upsilon_1, \ldots, \upsilon_n)$, $\upsilon_i \in [0, 1]\ \forall i$, $\sum_{i=1}^{n} \upsilon_i \leq 1$ over a set of selected bets $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)$. Each component of $\boldsymbol{\upsilon}$ represents a percentage of the total budget allocated to a specific bet. As a risk management measure, we introduced certain times a safety parameter γ, where $0 < \gamma < 1$, such that $\sum_{i=1}^{n} \upsilon_i = \gamma$ which acts as an upper bound on the sum of all allocation percentages. When $\gamma < 1$, a portion of the budget is deliberately kept as a cash reserve. We can write the return as

$$R(\boldsymbol{w}, \boldsymbol{\upsilon}) = \sum_{i=1}^{n} \mathcal{O}_i \upsilon_i \mathbb{1}_{w_i}, \tag{3.1}$$

where $\mathcal{O}_i$ is i-th odd from a set of n bets considered and

$$\mathbb{1}_{w_i} = \begin{cases} 1, \text{ if } w_i \text{ bet succeeds} \\ 0, \text{ otherwise} \end{cases} . \tag{3.2}$$

With

$$E[R(\boldsymbol{w}, \boldsymbol{\upsilon})] = \sum_{i=1}^{n} \mathcal{O}_i \upsilon_i P(\omega_i) \tag{3.3}$$

$$Var[R(\boldsymbol{w}, \boldsymbol{\upsilon})] = \sum_{i=1}^{n} \mathcal{O}_i^2 \upsilon_i^2 + \sum_{i=1}^{n} \sum_{\substack{j=0 \\ j \neq i}}^{n} \mathcal{O}_i \upsilon_i \mathcal{O}_j \upsilon_j P(\omega_i \cap \omega_j) - \left[\sum_{i=1}^{n} \mathcal{O}_i \upsilon_i P(\omega_i) \right]^2 , \tag{3.4}$$

where $P(\omega_i)$ is the estimated probability of the bet event ω_i occurs, which is estimated by the probability surface built by the public odds, as described in Chapter 2.

In our setting, it is possible that opportunity ω_i is won simultaneously with opportunity ω_j, $i \neq j$ (e.g., consider bet of correct score (Home team: 2 and Away Team: 2) and bet that both teams scores a goal in that match).

Now, $\boldsymbol{\upsilon} = (\upsilon_1, \ldots, \upsilon_n)$ is derived via optimization of a pre-determined objective function $M(f(R))$, where f is the estimated probability distribution of the return. That is

$$\boldsymbol{\upsilon}^* = \arg\max_{\boldsymbol{\upsilon} \in \mathbb{R}^n} M(f(R(\boldsymbol{w}, \boldsymbol{\upsilon}))). \tag{3.5}$$

The specific form of M is a decision that must address the purpose of what we want to optimize, such as return, risk, or other statistical properties of the distribution.

Having that said, the Sharpe Ratio (but not only) will serve as a fundamental concept in the optimization tasks, as it provides a critical metric for evaluating the risk-adjusted returns. Variations of the Sharpe Ratio components and optimization strategies will be detailed in the subsequent sections, providing a comprehensive understanding of the methodologies applied.

3.1 Objective Functions

3.1.1 Sharpe Ratio

We first follow a setup of single objective function by optimizing the Sharpe Ratio, i.e,

$$\arg\max_{\boldsymbol{v}\in\mathbb{R}^n} \frac{E[R(\boldsymbol{w},\boldsymbol{v})]}{\sqrt{Var[R(\boldsymbol{w},\boldsymbol{v})]}} \tag{3.6}$$

$$\arg\max_{\boldsymbol{v}\in\mathbb{R}^n} \frac{\sum_{i=1}^{n} \mathcal{O}_i v_i P(\omega_i)}{\sqrt{\sum_{i=1}^{n} \mathcal{O}_i^2 v_i^2 + \sum_{i=1}^{n} \sum_{\substack{j=0\\ j\neq i}}^{n} \mathcal{O}_i v_i \mathcal{O}_j v_j P(\omega_i \cap \omega_j) - \left[\sum_{i=1}^{n} \mathcal{O}_i v_i P(\omega_i)\right]^2}} \quad . \tag{3.7}$$

Covariance in the variance equation captures how the outcomes of two or more bets relate to each other. If the outcomes are positively correlated (i.e., they tend to win or lose together), the overall risk increases because unfavorable results from one bet are likely to coincide with others. On the other hand, if the outcomes are negatively correlated or uncorrelated, the bets can act as a hedge, reducing overall volatility and thus lowering the portfolio's risk. Notice here that the covariance is being considered in the variance calculation.

So the presence of covariance allows the optimization to adjust the betting stakes v_i accordingly. If certain bets are strongly positively correlated, it might allocate less weight to them to avoid excessive concentration of risk. Conversely, uncorrelated or negatively correlated bets might be given higher weights, as they provide diversification benefits, stabilizing returns.

Therefore, the Sharpe Ratio, as a measure of risk-adjusted return, will thus reflect not only the potential for higher returns but also the risk minimization achieved by diversifying across more negatively correlated bets opportunities.

3.1.2 Lower Confidence Limit

In this variation, we modify the Sharpe Ratio by introducing a custom formulation that balances expectancy and risk, via a hyperparameter λ. Specifically, we define the objective function Lower Confidence Limit (LCL) as

$$LCL = E[R(\boldsymbol{w}, \boldsymbol{v})] - \lambda\sqrt{Var[R(\boldsymbol{w}, \boldsymbol{v})]}. \tag{3.8}$$

Here, λ is treated as hyperparameter to be tuned, which controls the trade-off between maximizing expectancy and minimizing the risk (standard deviation). The higher the λ, the more emphasis is placed on minimizing risk, while lower values of λ prioritize higher returns, potentially at the expense of greater volatility.

The objective is to optimize both the expectancy and the volatility of the portfolio, allowing the method to balance between risk and reward based on the optimal value of λ.

3.1.3 Multi-Objective with Sharpe Ratio components

In optimizing betting allocations for soccer games, it is essential to balance the trade-off between maximizing expected returns and minimizing risk. This balance can be effectively achieved through Multi-Objective Bayesian Optimization (MOBO).

Here, we consider both the expectancy and the standard deviation as the objectives. So MOBO seeks to optimize both $E[R]$ and $Var(R)$ simultaneously and separately. Precisely, here we maximize the expected return and minimize the standard deviation of the return, which are the two key components of the Sharpe Ratio.

By constructing surrogate models for these objectives, MOBO efficiently explores the space of possible betting strategies to identify the Pareto front-the set of non-dominated solutions where no objective can be improved without worsening another.

This approach differs significantly from optimizing the single objective function Sharpe Ratio, because (1) optimizing the Sharpe Ratio may obscure the individual contributions of expected return and risk, making it difficult to understand the trade-offs between them, (2) the Sharpe Ratio assumes a specific linear relationship between return and risk, which may not capture the complexities inherent in betting strategies and (3) it yields a single optimal solution, neglecting other potentially valuable strategies that could be more suitable depending on different risk preferences.

In contrast, MOBO provides a set of optimal solutions along the Pareto front, offering a spectrum of choices catering to various risk-return profiles.

Implementing MOBO involves the following steps:

- Initialization: begin with an initial set of betting strategies and evaluate their expected return and standard deviation.

- Surrogate Modeling: use Gaussian Processes or other suitable methods to create surrogate models for both objectives based on the initial data.

- Acquisition Function Optimization: select new betting strategies to evaluate by optimizing an acquisition function that balances exploration and exploitation.

- Iteration: update the surrogate models with new data and repeat the process until convergence criteria are met.

A deeper understanding of MOBO was presented at Section 2.5.

By focusing on both expectancy and standard deviation, MOBO acknowledges that maximizing returns often comes with increased risk. It provides a more nuanced optimization framework that accommodates the complex dynamics of betting markets.

In conclusion, MOBO offers a robust and flexible approach to optimizing betting allocations. It surpasses single-metric optimization methods like the Sharpe Ratio by providing a comprehensive view of the trade-offs between return and risk.

Here, after obtaining the Pareto front of optimal solutions, the strategy selects the allocation for the solution with the lowest standard deviation for implementation, which implies that the allocation is efficient, i.e., there is no other allocation with a higher expectancy for the same or lower risk.

3.2 Optimization Strategies

3.2.1 Filtering Bets Rule

Given the real probabilities estimates for all outcomes of soccer games derived by the statistical modeling, described in Chapter 2, we compare these with the odds offered by public bookmakers. A bet is considered favorable if the odds offered by the bookmaker are greater than the odds suggested by our model (real fair odds). This step is crucial as it identifies opportunities where the bookmaker's odds are mispriced relative to the true likelihood of an event. By betting only when the odds are in our favor, theoretically we increase the chances of achieving a positive return over time.

Once favorable bets are identified, they are ranked from the most favorable to the least. Ordering bets helps in prioritizing the stake allocation, ensuring that the bets with the greatest edge are given precedence. We assess the favorability of a bet by using a score, defined by

$$s_i = w\left(\frac{\mathcal{O}_i - \frac{1}{\hat{p}_i}}{\mathcal{O}_i}\right) + (1-w)\hat{p}_i, \tag{3.9}$$

accounting the public odd $\mathcal{O}_i$, the estimated real probability $\hat{p}_i$, and a weighting factor w. This score helps in deciding whether or not to place a bet. The expression $\left(\frac{\mathcal{O}_i - \frac{1}{\hat{p}_i}}{\mathcal{O}_i}\right)$ computes the relative difference between the public odds and the odds implied by our estimated real probability, $\frac{1}{\hat{p}_i}$ is the fair odds computed based on the probability estimate. If the model is correct, this is what the odds "should" be if they were fair (i.e., no bookmaker margin), $\mathcal{O}_i - \frac{1}{\hat{p}_i}$ gives the difference between the public odds and the calculated fair odds. A positive value indicates that the public odds are offering more payout than what is fair based on your model, suggesting a favorable bet. By dividing this difference by $\mathcal{O}_i$, we normalize it by the scale of the odds, effectively measuring how much more favorable the public odds are compared to the fair odds.

Furthermore, w is a weighting factor that determines how much emphasis to put on the relative difference between public odds and fair odds versus the real probability estimate itself, the term $w\left(\frac{\mathcal{O}_i - \frac{1}{\hat{p}_i}}{\mathcal{O}_i}\right)$ scales the normalized difference by w, focusing on the value offered by the bet relative to its price and the term $(1 - w)\hat{p}_i$ adds a component of the raw estimated probability, scaled down by $(1 - w)$. This part of the score reflects the inherent likelihood of the event as estimated by your model, regardless of the betting odds. The score s_i combines these two aspects (1) the attractiveness of the bet's odds relative to its fair value and (2) the underlying probability of the outcome occurring. A higher score indicates a more favorable bet, either because the odds are particularly good or because the event is likely to occur (or both).

After analysing and empirical testing, the scoring function equation 3.9 outlined in this study was selected over the straightforward expected value (EV) approach due to its superior performance in practice. While the traditional EV method focuses solely on the expected return by calculating

$$\mathrm{EV}_i = \hat{p}_i \mathcal{O}_i, \tag{3.10}$$

it does not account for the balance between the value of the odds and the probability of the outcome occurring.

Empirical results from backtesting showed that using our scoring function led to more accurate bet selections and higher overall returns compared to the EV method. The flexibility provided by the weighting factor w enabled optimization tailored to different betting scenarios, enhancing the model's adaptability and effectiveness. Consequently, the scoring function was adopted as the primary method for filtering and ranking bets.

3.2.2 Dynamic Percentage-Based Budget Allocation

In the above optimization tasks, an arbitrary percentage of the budget to bet must be fixed. However, in practice, the percentage of the budget allocated to betting is a critical parameter that impacts the performance of the strategy. To address this, in this technique, we insert the percentage of the budget to bet, denoted by γ, as an additional input to the vector allocation optimization task and then we have objective function called γ Sharpe Ratio.

The following equations consider both the preserved and allocated portions of the budget, unlike previous equations, which accounted only for the portion at risk. We can write the total return, considering the percentage of the budget to bet, as

$$T(\gamma, \boldsymbol{w}, \boldsymbol{v}) = B(1-\gamma) + \gamma B R(\boldsymbol{w}, \boldsymbol{v}) \tag{3.11}$$

such that B is the budget available for betting in the beginning and $\gamma \in [0, 1]$ is the percentage of the budget to be used at every betting day. The term $B(1-\gamma)$ represents the portion of the budget that remains uninvested, while $\gamma B R(\boldsymbol{w}, \boldsymbol{v})$ is the return from the bets placed.

The expected value, the second moment, and the variance of $T(\gamma, \boldsymbol{w}, \boldsymbol{v})$ are given by

$$E[T(\gamma, \boldsymbol{w}, \boldsymbol{v})] = B(1-\gamma) + \gamma B E[R(\boldsymbol{w}, \boldsymbol{v})], \tag{3.12}$$

$$E[(T(\gamma, \boldsymbol{w}, \boldsymbol{v}))^2] = B^2[(1-\gamma)^2 + 2\gamma(1-\gamma)E[R(\boldsymbol{w}, \boldsymbol{v})] + \gamma^2 E[R(\boldsymbol{w}, \boldsymbol{v})]^2], \tag{3.13}$$

$$Var[T(\gamma, \boldsymbol{w}, \boldsymbol{v})] = B^2\gamma^2 Var[R(\boldsymbol{w}, \boldsymbol{v})]. \tag{3.14}$$

With that, the γ Sharpe Ratio is defined as

$$\begin{aligned}\gamma\text{Sharpe Ratio} &= \frac{E[T(\gamma, \boldsymbol{w}, \boldsymbol{v})]}{\sqrt{Var[T(\gamma, \boldsymbol{w}, \boldsymbol{v})]}} \\ &= \frac{B(1-\gamma) + \gamma B E[R(\boldsymbol{w}, \boldsymbol{v})]}{B\gamma\sqrt{Var[R(\boldsymbol{w}, \boldsymbol{v})]}} \\ &= \frac{(1-\gamma) + \gamma E[R(\boldsymbol{w}, \boldsymbol{v})]}{\gamma\sqrt{Var[R(\boldsymbol{w}, \boldsymbol{v})]}} \end{aligned} \tag{3.15}$$

3.2.3 Joint Optimization of (γ, λ^*)

Finally, it is worth noticing that we can optimize both the percentage of the budget γ and λ from LCL method, from Section 3.1.2, simultaneously, to create the objective function Joint

(γ, λ), which is defined as

$$\text{Joint}\,(\gamma, \lambda) = E[T(\gamma, \boldsymbol{w}, \boldsymbol{v})] - \lambda\sqrt{Var[T(\gamma, \boldsymbol{w}, \boldsymbol{v})]}, \tag{3.16}$$

3.2.4 Long-Term Return Distribution

The long-term return optimization method in this study is built on Monte Carlo simulation to estimate the return in the long-term of a betting allocation in a day.

For each day, we bet on multiple games. Let r_{ji} denote the return for game i on day j, where $j \in \mathbb{Z}^+$ and $1 \leq j \leq d$ represents each betting day, d is the last day of a betting period,$i = 1, 2, \ldots, N_j$ represents each game within the day j. The total return for each day is the sum of returns across all games played that day:

$$R_j = \sum_{i=1}^{N_j} r_{ji}, \tag{3.17}$$

where N_j is the number of games on day j. This sum reflects the combined effect of all bets placed on that day.

The next step is to consider the long-term return over 100 instances, given a betting day. For each day, an observation of long-term return $\tilde{R}$ is defined as the product:

$$\tilde{R} = \prod_{j=1}^{100} R_j = \prod_{j=1}^{100} \left(\sum_{i=1}^{N_j} r_{ji} \right). \tag{3.18}$$

This equation captures the interaction of returns over all games and all rounds, representing a proxy for the long-term return of a given betting day. The key idea is that each r_{ij} contributes multiplicatively to the overall performance and compounding effects play a critical role in long-term.

This is a particularly complex procedure because we are not considering individual bets in isolation. Instead, multiple games are bet on each day, and the returns from these games interact over time, making the dynamics of the process highly non-linear.

To estimate robustly the long term return of a betting day, we apply Monte Carlo simulation. Specifically, for each simulation k, we compute:

$$\tilde{R}_k = \prod_{j=1}^{100} \left(\sum_{i=1}^{N_j} r_{ji,k} \right), \tag{3.19}$$

where $r_{ji,k}$ is the return for game i on day j in the k-th simulation. The index k indicates that different sets of returns are sampled in each iteration. We repeat this process 1000 times, generating 1000 synthetic observations of $\tilde{R}$.

By generating 1000 observations of the random variable $\tilde{R}$, we will obtain a distribution of returns observations over the long term, that resembles the true distribution of long-term returns f that we can then use to optimize the objective function M, as defined in Equation 3.5.

Here, in the end, we are interested still in the Sharpe Ratio, so the objective function to be optimized is the Sharpe Ratio of the long-term return and Monte Carlo simulation helps approximate both the expected value and standard deviation of returns, with the optimizer iteratively adjusting $\boldsymbol{v}$ to improve this risk-adjusted metric.

Using 1000 simulated outcomes, we approximate the Sharpe ratio as:

$$\text{Sharpe Ratio} \approx \frac{\frac{1}{1000}\sum_{k=1}^{1000} \tilde{R}_k}{\sqrt{\frac{1}{1000}\sum_{k=1}^{1000} (\tilde{R}_k - \bar{\tilde{R}})^2}}, \tag{3.20}$$

where $\bar{\tilde{R}}$ is the mean of the simulated outcomes. This process provides an estimate of the risk-adjusted performance of the betting strategy, accounting for both the multiplicative nature of returns and their variability across multiple games and rounds.

The complexity of this approach arises from the fact that we are dealing with multiple games each day, and the returns are multiplicative. Small changes in allocation or return outcomes can lead to large variations in our long term return proposition over time due to the compounding effect of daily returns. Additionally, different rounds may feature different numbers of games, and the variability in game outcomes further complicates the optimization process.

BO still could be used, here, but we adopted specifically the Constrained Optimization BY Linear Approximation (COBYLA) algorithm. We argue that BO optimization's model-fitting and acquisition steps are computationally demanding and COBYLA, which employs a direct search approach without needing to build a surrogate model, demonstrated faster convergence in this context. Additionally, with the non-deterministic nature of the objective function, COBYLA's simplicity and direct approach proved advantageous.

3.3 Bayesian Optimization Implementation

With the bets selected, an associated vector of allocation is used to place amounts on each bet. This involves setting up an optimization task for a pre-determined objective function. The allocation vector determines how funds are distributed among selected bets, which is critical for balancing potential returns against risk.

BO is effective in handling complex, noisy optimization problems typical.

In the initial phase of the optimization task, we generate quasi-random samples using the Sobol sequence for the starting point. Sobol sequences are a type of low-discrepancy sequence that are particularly effective in creating evenly distributed samples across multi-dimensional spaces. This method is crucial for efficiently exploring parameter spaces in BO, especially in high-dimensional scenarios. The use of Sobol sequences in sampling is rooted in their ability to provide a more uniform exploration of the parameter space with interesting properties and applications of Sobol sequences in numerical methods and optimization (Sobol, 1967).

The bounds for initial sampling and subsequent optimization phases are set differently, which is informed by specific operational needs. The initial Sobol sampling uses a narrow range of $[0.001, 0.1]$, targeted to explore a specific segment of the parameter space where promising solutions are hypothesized to exist. Conversely, during the acquisition function optimization,

Algorithm 6: Betting Procedure Daily

```
Dependencies:
• $7 \times 7$ probability surface estimated for each game $i$ of a day.
• List of all $k$ bet opportunities with the associated odds for each game of a day.
• Objective function $M$ to be optimized.

Initialize the list odds_dt to store the bets selected for all games of the day.
for $i$ in $1, \dots, N_j$ do
    Map each bet of the set of bet opportunities $\boldsymbol{\omega}^{(i)} = (\omega_1, \dots, \omega_k)$ of $i$-th game to the
      corresponding probability estimate.
    Filter the $n$ most favorable bets of $i$-th game ($n$ is a hyperparameter) and get
      $\boldsymbol{\omega}^{(i)} = (\omega_1, \dots, \omega_n)$, such that $(n < k)$, accordingly to Section 3.2.1.
    Store it to odds_dt.
Run optimization task for bets selected inside odds_dt, to optimize pre-determined
  objective function $M$ .
Apply Softmax function to the optimization result to obtain the stake allocation.
return Optimized allocation vector $\boldsymbol{v}$.
```

the bounds are much broader. This broader range accommodates the use of a softmax transformation applied during each evaluation iteration of the objective function, which maps the inputs of the vector allocation to a $[0, 1]$ range.

The softmax function is a common transformation used in machine learning and optimization tasks to convert a vector of arbitrary values into a vector that is defined on a simplex space. This transformation is particularly useful in the context of stake allocation, where the objective is to distribute funds among different bets ensuring that the sum of the allocations equals 1. The softmax function is defined as

$$\text{softmax}(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}. \tag{3.21}$$

However, despite this transformation, the results themselves are not confined within any strict bounds, implying a non-linear or complex relationship that these parameters have with the outcomes. Therefore, maintaining different bounds in each phase is justified by the necessity to align the parameters effectively with the characteristics of both the model and the softmax transformation.

```
Algorithm 7: Performance
  Dependencies:
  • Outcome data for all games of the $d$ betting days
  --------------------------------------------------------------------
  for $j$ in $1, \dots, d$ do
      for $i$ in $1, \dots, N_j$ do
          Compute financial return of game $i$ of the day $j$ using the stake allocation
            vector (according to Algorithm 6) and the outcome data.
      Compute financial return of the day.
  return Financial return of betting strategy for all the time period.
```

```
Algorithm 8: Optimization Task
  Dependencies:
  • Bets with associated public odds and probabilities estimates, given by statistical
    modeling and objective function to be optimized.
  • Maximum number of iterations max_iter.
  Output:
  • Optimized stake allocation vector.
  --------------------------------------------------------------------
  train_X ⟵ Random samples from Sobol sequence to initialize the optimization task.
  train_Y ⟵ Evaluation of these points using the objective function to obtain the
    corresponding outputs.
  best_value ⟵ Maximum value of objective function.
  best_candidate ⟵ Candidate that maximizes objective function.
  while number of iterations $\leq max\_iter$ do
      Standardize train_Y.
      Fit a GP model to the standardized data.
      Optimize the acquisition function to find new candidate.
      Apply softmax function to the new candidate to obtain the stake allocation vector.
      new_Y ⟵ objective function evaluated at new candidate.
      Append train_X and train_Y with the new candidate and its evaluation new_Y.
      if new_Y > best_value then
          best_value ⟵ new_Y
          best_candidate ⟵ new candidate
      number of iterations ⟵ number of iterations + 1
  return Optimized stake allocation vector.
```

The betting procedure on a day can is presented by following Algorithm 6 and the performance evaluation is described in Algorithm 7 and the optimization task is described in Algorithm 8..

3.4 Hyperparameters

The betting strategy optimization tasks involve several hyperparameters that can be tuned to achieve a better performance, they are parameters that are not learned during the optimization process, but rather set before the optimization begins. All the hyperparameters are detailed in Table 3.1.

Table 3.1: Hyperparameters

Hyperparameters	**Values**	**Description**
γ	(0, 1]	Percentage of the budget to bet
`min_games`	{0, 1, 2, 3, 4, 5, ...}	Minimum number of games to decide if betting day will be considered
`max_games`	{0, 1, 2, 3, 4, 5, ...}	Maximum number of games to decide if betting day will be considered
`max_bets`	{1, 2, 3, 4, 5, ...}	Maximum number of bets to consider
`bets_per_game`	{1, 2, 3, 4, 5, ...}	Number of bets per game
w	(0, 1)	Weight of the linear combination filter

For the sake of simplicity, we will not document exhaustively the configurations of hyperparameters for each optimization task, but we will present the relevant ones according to the methodology of each optimization task. All hyperparameter tuning jobs were done in a dataset that contains the period from 2019 to 2023, and the best hyperparameters were used to evaluate the performance of the betting strategy in the period from 2023 to 2024.

3.5 Evaluation

To evaluate the performance of the optimization tasks, the data has been split as usual:

- Training Set (2019-2023): we use this period to apply the methods and optimization techniques with multiple configurations of hyperparameters to, in the end, select the configuration for each optimization technique that yields the higher return in that period.
- Test Set (2023-2024): once the best combination of hyperparameters for each technique is determined, we apply the technique to the 2023-2024 period to validate their real-world performance on unseen data.

In the following section, we will present the optimization techniques (after the they being applied to the training set (2019-2023)), assessing their performance and validating them on the test set (2023-2024).

Chapter 4

Experiments

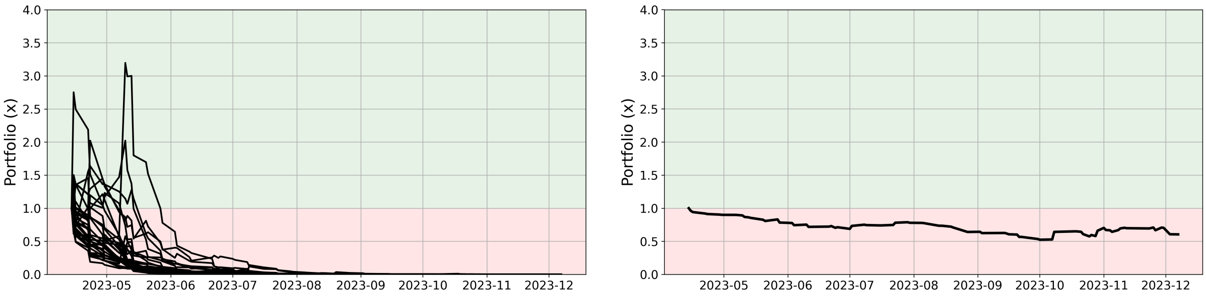
We present the results of the experiments using web-scraped data from the Brazilian Serie A soccer league. The bookmakers considered here were Bet365, LeoVegas Sport, 1xBet, Betway, 22Bet, ComeOn, LVBET, Betsson, Marathonbet, 888sport, Betfair, Vbet Sport, Betsafe, bet90, LSbet, NetBet, Bethard, Mr Green Sport, Betclic, Bodog, Bovada, Mobilebet, NordicBet, Rivalo, Pinnacle, Betano, Parimatch, 18Bet, Megapari Sport, Stake.com, GG.BET, Bwin, Betobet, Fansbet, Betwinner, 1xBit, Powbet, Sportaza, Suprabets, 4rabet, Amuletobet, Sportsbet.io, Dafabet LATAM, BC.Game Sport, Fezbet, Mystake, 31bet, Jackbit, Freshbet, Goldenbet, Weltbet, Dafabet, bet365, Midnite, Betibet, Roobet, 20Bet, TonyBet, Ivibet, Vave, SnatchCasino, Galera.Bet, Nine Casino, Mostbet. None bookmakers were deleted from the dataset at any point.

The data was collected from 2019 to 2024, and the experiments were conducted from 2019 to 2023. The data was divided into training and test sets, with the training set containing data from 2019 to 2022 and the test set containing data from 2023 to 2024.

Here, it is important to clarify how the charts should be interpreted. All graphs depicting portfolio performance are to be understood in terms of how many times the initial budget has been multiplied over the course of the betting strategy.

The portfolio axis on the charts represents this multiplier effect. Specifically, the first data point corresponds to a portfolio value of 1, indicating that no gains or losses have been realized at the start, meaning the initial budget remains unchanged. As the strategy progresses and results accumulate, subsequent data points will reflect how the portfolio grows or contracts. A portfolio value greater than 1 signifies that the initial budget has been increased, resulting in profit, whereas a value less than 1 indicates a loss.

4.1 Random and baseline



(a) Random strategies for typical bettor.

(b) Performance of baseline.

Figure 4.1: (a) Random strategies for typical bettor (b) Performance of baseline (Kaunitz, Zhong, and Kreiner, 2017).

The typical bettor, lacking access to our sophisticated system which bases probability estimates of bet events on rigorous statistical modeling, typically engages in betting through a simpler, more imprudent approach. We assume that such a bettor merely logs into a betting application and places daily a bet of 10% of the available budget on a single, arbitrarily selected soccer match. The performance of this conventional betting strategy, demonstrated in Figure 4.1a, presents results from 30 simulations. This visualization illustrates that for many participants, this approach will inevitably lead to financial ruin.

Furthermore, we established our initial baseline by implementing the straightforward strategy proposed in the paper "Beating the bookies with their own numbers" (Kaunitz, Zhong, and Kreiner, 2017). This approach suggests placing a fixed wager whenever the public odds exceed the true odds implied by the actual probability. We restricted our analysis to simple bets on home wins, draws, or away wins, as implemented in the original paper. The core idea behind this method is to exploit potential inefficiencies in the betting market by identifying instances where bookmakers may have overestimated the odds for a particular outcome. In theory, this should provide a statistical edge over time if the true probabilities are accurately estimated.

However, when applied to our dataset and under our specific conditions, this simple approach failed to yield profitable results, as shown in Figure 4.5b. The approach presented in exhibits several limitations. First, it overlooks the potential benefits of negative correlations between bets. Second, it lacks a proper estimation of the probability surface. Third, the strategy employs a fixed betting amount rather than adjusting the stake based on the total available budget, limiting its potential for optimal capital allocation. Additionally, here, we are using data from the Brazilian soccer league, which may present unique characteristics not considered in the original paper that analyzed from 818 leagues and divisions worldwide.

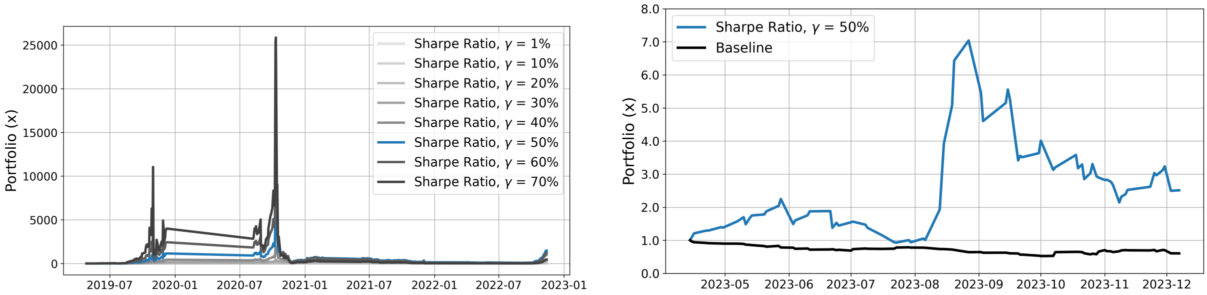
Finally, another probable explanation is that in the paper, they had a live system that monitored the odds and placed bets in real-time, while we are using a unique value to each betting opportunity to simulate the strategy. This lack of success with the baseline approach underscores the need for more sophisticated strategies.

4.2 Sharpe Ratio Optimization

Following the method presented in Section 3.1.1, we executed the single objective Sharpe Ratio optimization for values of γ (here, it is a fixed percentage of budget to bet), ranging from {1%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%}, in the training set. For the test set, we apply the method with the γ that led to the highest return in training set. The results are shown in Figure 4.2.

Although, Figure 4.2 does not show γ values from 70% to 100%, they were, in fact, utilized but led to ruin in the training set.

As shown in the plot, the portfolio dynamics vary significantly depending on the chosen



(a) Performance 2019-2023

(b) Performance 2023-2024

Figure 4.2: (a) Performance of BO for Sharpe Ratio maximization on training set. (b) Performance of BO for Sharpe Ratio maximization on test set.

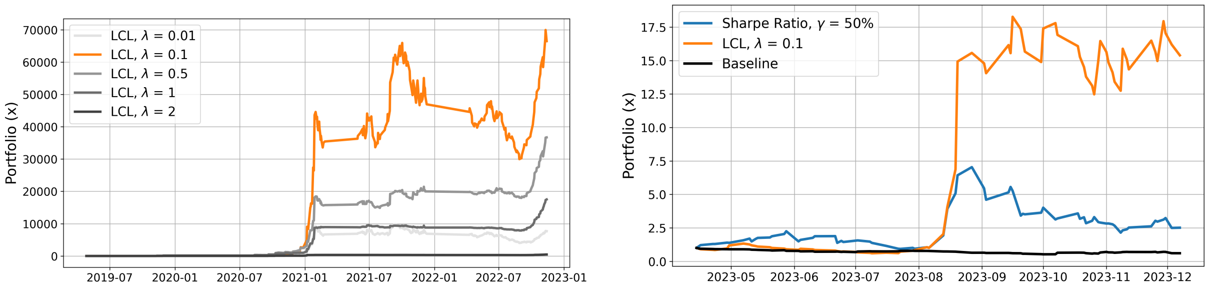
allocation percentage. The 50% allocation led to the highest final return, and this value will be used in the next strategies that adopt a fixed percentage of allocation to ensure fair performance comparisons.

4.3 LCL Optimization

The LCL optimization technique introduced a new hyperparameter, λ, as described in Section 3.1.2. With the goal to check how much the λ hyperparameter could increment or decrement the final return (compared to using the Sharpe Ratio criteria) for the experiment in Section 4.2 for the optimal $\gamma = 50\%$ value, we implemented the LCL optimization for the training set and tested the best configuration in the test set, but we noticed that $\gamma = 50\%$ was too aggressive for this method.

Therefore, we tested multiple values of λ for a small percentage of the budget to bet ($\gamma = 5\%$), and we found that $\lambda = 0.1$ led to the highest final return. The results of this optimization are shown in Figure 4.3.

We can see that $\lambda = 0.1$ led to the highest final return multiplying the initial budget by 15 times in the test set.



(a) Performance 2019-2023

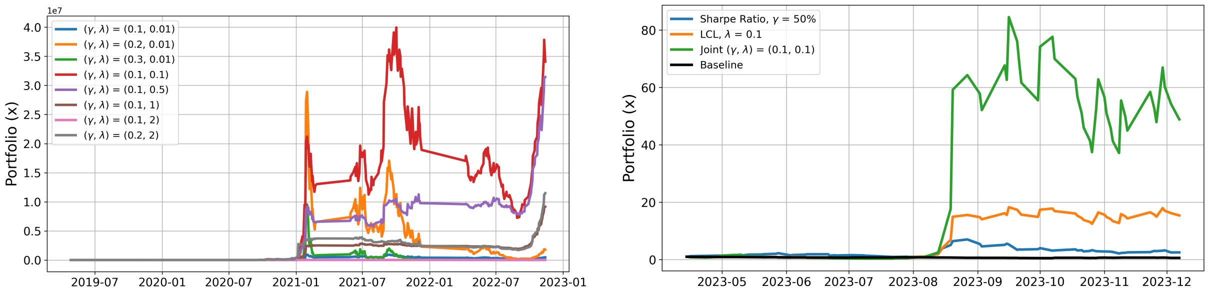
(b) Performance 2023-2024

Figure 4.3: (a) Performance of BO for LCL maximization on training set. (b) Performance of BO for LCL maximization on test set.

4.4 Joint Optimization of (γ, λ)

So, it is evident that λ and γ are crucial hyperparameters that can significantly impact the strategy's performance. To further explore the interaction between these two parameters, we conducted a joint optimization of the LCL method and the budget allocation percentage γ, as shown in Figure 4.4, as detailed in Section 3.2.3.

Here, we report only the runs that, for the training set, did not lead to ruin, and the best configuration was tested in the test set. We have considered values of γ of $\{1\%, 10\%, 20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%, 90\%, 100\%\}$ and λ values of $\{0.01, 0.1, 0.5, 1, 2\}$.



(a) Performance 2019-2023

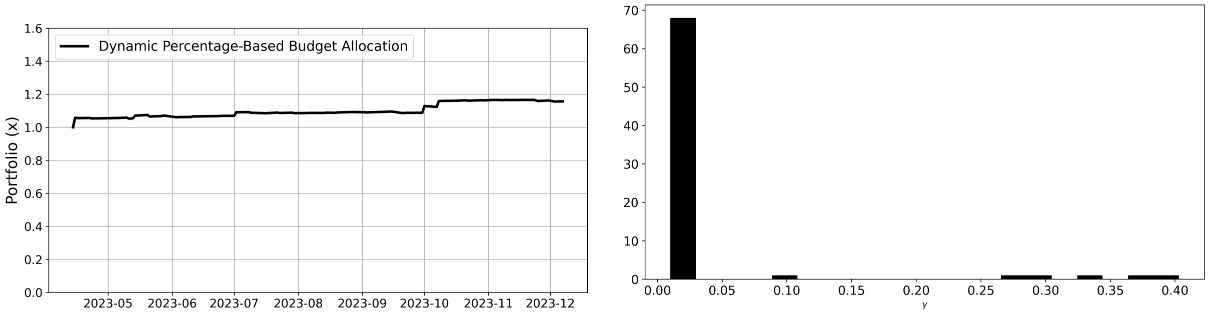
(b) Performance 2023-2024

Figure 4.4: (a) Performance of BO for joint maximization on training set. (b) Performance of BO for joint maximization on test set.

Therefore, the optimal values for $(\gamma = 0.1, \lambda = 0.1)$ led to a very high multiplication of the initial budget, here the initial budget was multiplied 50x in the test set.

4.5 Dynamic Percentage-Based Budget Allocation Optimization

The percentage of budget allocation γ is now a variable included in the optimization process, as proposed in Section 3.2.2 (γ is treated as part of the input vector to be optimized), unlike the previous methods, where an pre-determined arbitrary fixed percentage of the budget is set.



(a) Performance of BO for variable γ Sharpe Ratio maximization.

(b) Histogram of percentages of budget allocated set by the optimization.

Figure 4.5: (a) Performance of BO for variable γ Sharpe Ratio maximization. (b) Histogram of percentages of budget allocated set by the optimization.

Table 4.1: Mean and Median Returns for Different γ Buckets.

γ **Bucket**	**Mean Return**	**Median Return**
(0.0, 0.1125]	0.979	1.062
(0.225, 0.3375]	1.124	1.139
(0.3375, 0.45]	1.044	1.044

The BO optimization for this method with selected Sharpe Ratio optimization made the system too conservative, most of the percentages of the budget to bet were concentrated between 1% and 3%, but it is worth noticing that the cases where the percentage was set, by the optimization, in bucket (0.225, 0.3375] or (0.3375, 0.45], we obtained positive results, as shown in Table 4.1.

4.6 Long Term Return Distribution Optimization for Sharpe Ratio

Here, we present the results of the long-term optimization for Sharpe Ratio maximization using the Monte Carlo method, as detailed in Section 3.2.4. The results are shown in Figure 4.6.

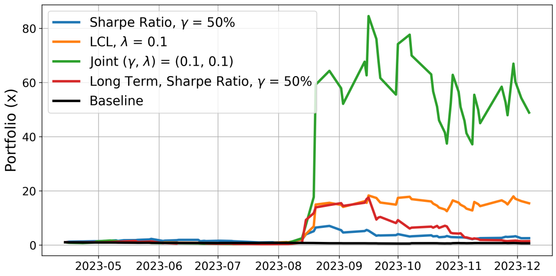


Figure 4.6: Performance of long-term COBYLA optimization for Sharpe Ratio maximization.

This method was the only one that used a not bayesian optimization method and the final return was not as high as the other methods. It led to a multiplication of the initial budget by almost 1.5x.

4.7 Multi-Objective Bayesian Optimization

Finally, we present the results of the multi-objective Bayesian optimization for the expectancy and standard deviation objectives, as detailed in Section 3.1.3. The results are shown in Figure 4.7

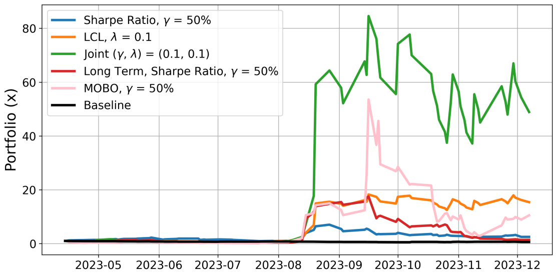
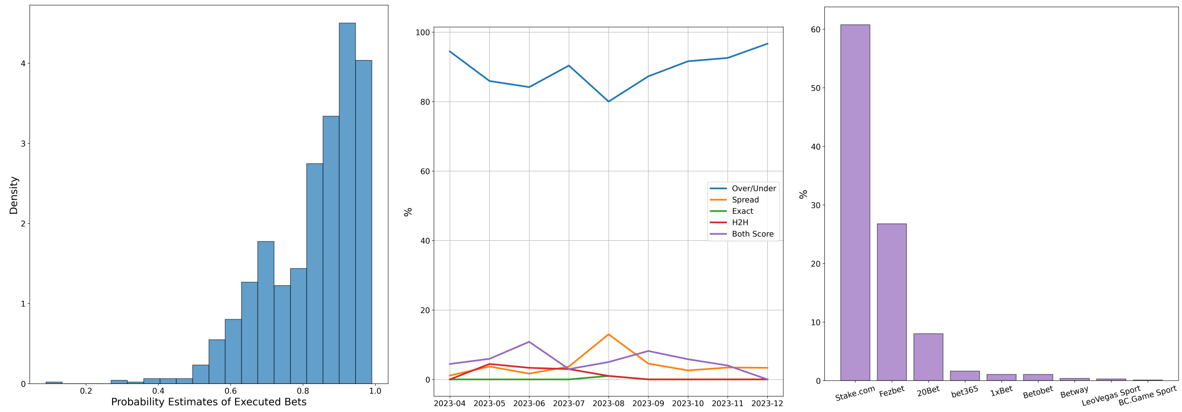


Figure 4.7: Performance of Multi-Objective BO with $E(R)$ and $Var(R)$ as objectives

This method led to a high multiplication of the initial budget of about 10x, falling behind the Joint Optimization of $(\gamma, \lambda) = (0.1, 0.1)$ optimization and LCL, $\lambda = 0.1$ methods.

4.8 Exploratory Data Analysis Study Case

We have presented the performance results of several optimization methods. Each one of them has its own characteristics, so if we dig into the associated data, we are able to understand the reasons behind the results. Here, we present an exploratory data analysis of the test set for the Joint Optimization of $(\gamma, \lambda) = (0.1, 0.1)$ method, as a study case. We will analyze the density of probabilities estimates of executed bets, the distribution of markets of executed bets, and the distribution of bookmakers of executed bets.



(a) Density of probabilities estimates of executed bets.

(b) Distribution of markets of executed bets.

(c) Distribution of bookmakers of executed bets.

Figure 4.8: (a) Density plot of the probabilities estimates of the executed bets. (b) Distribution of markets of the executed bets (the percentage of bets in each market is shown, such that the sum of all markets is 100% for each month). (c) Distribution of bookmakers of executed bets. All plots refer to the test set and Joint Optimization of $(\gamma, \lambda) = (0.1, 0.1)$ method.

We can see that the density of probabilities estimates of executed bets is right skewed, suggesting that the majority of bets have high probabilities estimates. Therefore, bets are being place in high likelihood events.

The *Over/Under* market has been the predominant betting choice for the majority of the period observed, consistently holding a significant share.

Finally, although we had dozens of bookmakers in the dataset, executed bets have consistently originated from a select few bookmakers. The consistent preference for these particular bookmakers suggests that they likely offer more attractive bets.

Chapter 5

Conclusion

In this work, we have proposed a series of optimization methods, using bayesian framework, to bet on soccer matches using Brazilian Serie A data, equipped by probability estimates of the bets that describe 7×7 matrix of possible outcomes of a soccer match. We have demonstrated that all proposed methods led to a positive return on investment for the test set that comprehends data from 2023 to 2024.

We highlighted our novel approach key characteristics. First, we explore negative correlations of the bets, during the optimization tasks. Second, we consider multiple games simultaneously every day. Third, we have the access to a mechanism that estimates the probabilities of the bets.

Considering those characteristics, we have proposed methods of optimization for the allocation vector that varied between changing the objective function, the distribution of the return itself and percentage of budget to be allocated.

Sharpe Ratio, $\gamma = 50\%$ led to a multiplication of 2.5x of initial portfolio, but did not reach the high levels of some other methods. LCL, $\lambda = 0.1$, by changing the objective function, has folded the initial portfolio by 15x. Now, the Joint $(\gamma = 0.1, \lambda = 0.1)$ was clearly the best method, with a multiplication of 50x of the initial portfolio. The Long Term, Sharpe Ratio, $\gamma = 50\%$ although it led to a positive return on investment, it performed considerably worse than the other methods. Finally, MOBO, our most complex procedure, that simultaneously and separately optimized the allocation vector for the objectives of expectancy and standard deviation, has performed well, with a multiplication of 10x of the initial portfolio.

The overall results show great multiplications of the initial portfolio, which is a strong indicator that the proposed methods are conceptually profitable. However, it is important to highlight that the methods would face critical issues in a real world operation. Technically, we are considering 64 bookmakers, which would require a complex software system to be implemented that could on the fly deposit and withdraw money from the bookmakers for multiple games simultaneously.

Operationally, the bookmakers have rules to limit the maximum bet, they block accounts that are winning too much. Accounting wise, we did not consider any taxes, fees, or costs that would be charged by the bookmakers or the government.

Finally, it is important to make it clear that this paper does not encourage gambling, but rather presents a series of mathematical optimizations tasks that are equipped by a mechanism that builds the probability estimates of the bets.

References

Angelini, Giovanni and Luca De Angelis (2019). “Efficiency of online football betting markets”. *International Journal of Forecasting* 35.2, pp. 712–721.

Balandat, Maximilian et al. (2020). “BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization”. *Advances in Neural Information Processing Systems 33*. URL: `http://arxiv.org/abs/1910.06403`.

Clarke, Stephen, Stephanie Kovalchik, and Martin Ingram (2017). “Adjusting Bookmaker’s Odds to Allow for Overround”. *American Journal of Sports Science* 5.6, p. 45. ISSN: 2330-8559. DOI: `10.11648/j.ajss.20170506.12`. URL: `http://www.sciencepublishinggroup.com/journal/paperinfo?journalid=155{\&}doi=10.11648/j.ajss.20170506.12`.

Dixon, Mark J and Stuart G Coles (1997). “Modelling association football scores and inefficiencies in the football betting market”. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 46.2, pp. 265–280.

Franck, Egon, Erwin Verbeek, and Stephan Nüesch (2010). “Prediction accuracy of different market structures-bookmakers versus a betting exchange”. *International Journal of Forecasting* 26.3, pp. 448–459.

Giolo, Suely Ruiz (2017). *Introdução á análise de dados categóricos com aplicações*. Blucher.

Hubáček, Ondřej and Gustav Šír (2023). “Beating the market with a bad predictive model”. *International Journal of Forecasting* 39.2, pp. 691–719.

Hubáček, Ondřej, Gustav Šourek, and Filip Železný (2019). “Exploiting sports-betting market using machine learning”. *International Journal of Forecasting* 35.2, pp. 783–796.

Karlis, Dimitris and Ioannis Ntzoufras (2003). “Analysis of sports data by using bivariate Poisson models”. *The Statistician*.

Kaunitz, Lisandro, Shenjun Zhong, and Javier Kreiner (2017). “Beating the bookies with their own numbers-and how the online sports betting market is rigged”. *arXiv preprint arXiv:1710.02824*.

Kelly, John L (1956). “A new interpretation of information rate”. *the bell system technical journal* 35.4, pp. 917–926.

Koning, Ruud H. and Renske Zijm (2023). “Betting market efficiency and prediction in binary choice models”. *Annals of Operations Research* 325.1, pp. 135–148. ISSN: 0254-5330. DOI: `10.1007/s10479-022-04722-3`. URL: `https://link.springer.com/10.1007/s10479-022-04722-3`.

Langseth, Helge (2013). “Beating the bookie: A look at statistical models for prediction of football matches.” *SCAI*, pp. 165–174.

MacLean, Leonard C, William T Ziemba, and George Blazenko (1992). “Growth versus security in dynamic investment analysis”. *Management Science* 38.11, pp. 1562–1585.

Maher, Michael J (1982). “Modelling association football scores”. *Statistica Neerlandica* 36.3, pp. 109–118.

Markowits, Harry M (1952). “Portfolio selection”. *Journal of finance* 7.1, pp. 71–91.

Matej, Uhrín et al. (2021). “Optimal sports betting strategies in practice: an experimental review”. *IMA Journal of Management Mathematics* 32.4, pp. 465–489.

Mattera, Raffaele (2023). “Forecasting binary outcomes in soccer”. *Annals of Operations Research* 325.1, pp. 115–134.

Mchale, I.G. and Phil Scarf (Jan. 2006). “Forecasting international soccer match results using bivariate discrete distributions”.

Porfírio, José Vítor Barreto (2023). “Modelagem probabilística em apostas esportivas . 2023. 60 f., il. Trabalho de Conclusão de Curso (Bacharelado em Estatística) - Universidade de Brasília, Brasília, 2023.” URL: https://bdm.unb.br/handle/10483/38514.

Sharpe, William F (1994). “The sharpe ratio, the journal of portfolio management”. *Stanfold University, Fall*.

Shin, Hyun Song (Mar. 1992). “Prices of State Contingent Claims with Insider Traders, and the Favourite-Longshot Bias”. *The Economic Journal* 102.411, pp. 426–435. ISSN: 0013-0133. DOI: 10.2307/2234526. eprint: https://academic.oup.com/ej/article-pdf/102/411/426/27038776/ej0426.pdf. URL: https://doi.org/10.2307/2234526.

Sobol, I (1967). “The distribution of points in a cube and the accurate evaluation of integrals (in Russian) Zh”. *Vychisl. Mat. i Mater. Phys* 7, pp. 784–802.

Sortino, Frank A and Lee N Price (1994). “Performance measurement in a downside risk framework”. *the Journal of Investing* 3.3, pp. 59–64.

Terawong, Chawin and Dave Cliff (2024). “XGBoost Learning of Dynamic Wager Placement for In-Play Betting on an Agent-Based Model of a Sports Betting Exchange”. *arXiv preprint arXiv:2401.06086*.

Wheatcroft, Edward (2020). “A profitable model for predicting the over/under market in football”. *International Journal of Forecasting* 36.3, pp. 916–932.