2019: DISSERTAÇÃO DE MESTRADO
Mestrado Profissional em Matemática - PROFMAT, Universidade Federal de São João Del-Rei - UFSJ
Sociedade Brasileira de Matemática - SBM

Modelagem Matemática para Previsão de Jogos de Futebol

Luiz Fernando Pinheiro Ramos[1]
Humberto Cesar Fernandes Lemos[2]
Ben Dêivide de Oliveira Batista[3]

Resumo: Este trabalho se propõe a construir um método para calcular probabilidades no futebol e comparar os resultados com o modelo do Departamento de Matemática da UFMG. Na metodologia apresentada, assumimos que os números de gols marcados pelos times em uma partida são independentes e seguem uma distribuição Poisson, em que a média reflete o fator ataque ou fator defesa. Aplicada à 250 jogos da Série A do Campeonato Brasileiro de 2018, a metodologia estudada apresentou resultados coerentes com o modelo da UFMG. Foi possível prever com boa precisão o campeão e obteve resultados satisfatórios para os classificados para a Copa Libertadores da América, como também para os times rebaixados. A média da medida de DeFinetti foi 0,568 e aproximadamente 57% dos resultados estão abaixo de 2/3. A metodologia proposta foi implementada no software R.

Palavras-chave: Probabilidade no Futebol, Teoria das Probabilidades, Lei dos Grandes Números, Distribuição de Poisson.

Abstract: This study proposes to construct a method to calculate soccer probabilities and compare the results with the UFMG Mathematics Department model. In the presented methodology, we assume that the number of goals scored by the teams in a match are independent and follow a Poisson distribution, where the average reflects the attack factor or defense factor. Applied in 250 games of Series A of the 2018 Brazilian Championship, the methodology studied presented results consistent with the UFMG model. It was possible to predict with good accuracy the champion and obtained satisfactory results for the qualifiers for the Copa Libertadores of America, as well as for the relegated teams. The DeFinetti measurement average was 0.568 and approximately 57 % of the results are below 2/3. The proposed methodology was implemented in the software R.

Keywords: Probability in Soccer, Probability Theory, Law of Large Numbers, Poisson distribution.

[1]Aluno do Mestrado Profissional em Matemática em Rede Nacional, Turma 2017
Instituição: Universidade Federal de São João Del-Rei - UFSJ
E-mail: luizpramos@yahoo.com.br
[2]Orientador do Trabalho de Conclusão de Curso
Departamento de Estatística, Física e Matemática - DEFIM, CAP-UFSJ
E-mail: humbertolemos@ufsj.edu.br
[3]Coorientador do Trabalho de Conclusão de Curso
Departamento de Estatística, Física e Matemática - DEFIM, CAP-UFSJ
E-mail: ben.deivide@ufsj.edu.br

1 Introdução

Uma das maiores paixões dos brasileiros, é sem dúvida, o futebol. Este esporte emocionante é uns dos assuntos mais frequentes nos meios de comunicação. Para [4], o futebol é tão encantador pela sua magia e dinamismo, que até aqueles que não são fãs declarados param para ver a seleção de seu país disputar uma Copa do Mundo.

O futebol se torna cada vez mais alvo de especialistas que buscam na matemática, através da probabilidade, meios para explicar e prever os resultados. Tal fato é tão impressionante que muitos especialistas da área esportiva dizem que o futebol é o esporte no qual a chance do mais fraco vencer o mais forte se maximiza em relação a qualquer outro esporte. Quantas vezes clubes que fizeram grandes investimentos foram derrotados por outros que, “em teoria”, eram mais fracos? Um elenco de ponta e um esquema tático infalível são primordiais nesse jogo, que às vezes parece um combate estratégico duelado pelos técnicos, mas nem sempre os melhores times vencem. É exatamente isso que torna o futebol tão impressionante [4].

Diante disso, o futebol sempre foi alvo de especulações sobre quem seria o campeão, o vice-campeão, quais os clubes seriam rebaixados, qual o time que marcaria mais gols, entre muitas outras perguntas relacionadas. Surgiram então os primeiros modelos estatísticos para tentar responder essas perguntas que muitos torcedores e admiradores do esporte faziam [6].

Em [10], o autor verificou por meio da porcentagem de vitórias obtidas como mandante que um time tem vantagem ao jogar em casa com o apoio da torcida. Em [11] propuseram uma metodologia bayesiana para calcular probabilidades das partidas de futebol, utilizando opiniões de especialistas e o ranking da FIFA como informações a priori. Também é possível encontrar trabalhos que utilizam a distribuição de Poisson, em [8] o autor utiliza esta distribuição para modelar os números de gols marcados pela Inglaterra, Irlanda, Escócia e País de Gales, no Campeonato Internacional Britânico 1883 à 1980. Em [9], o autor também considera a distribuição de Poisson em que as médias dependem dos efeitos ofensivos e defensivos de cada equipe.

Este trabalho propõe uma metodologia, baseada na distribuição de Poisson, para calcular probabilidades no futebol que será detalhado na seção 5. Na seção 2, apresentaremos um estudo sobre a Teoria da Probabilidade, trataremos as definições das probabilidades clássica e frequentista, destacaremos a probabilidade axiomática através dos axiomas de Kolmogorov, apresentaremos a probabilidade condicional e a independência de eventos. Na seção 3 introduzimos o conceito de variáveis aleatórias e quais são os seus tipos, dando destaque às variáveis discretas, e apresentamos a distribuição de Poisson. Na seção 4, apresentamos os tipos de convergência e a Lei dos Grandes Números, que é o conceito mais importante para o desenvolvimento deste trabalho. Na seção 5, detalharemos o método proposto por este trabalho, o qual será chamado de método do PROFMAT, e na seção 6 confrontaremos os resultados com o modelo do Departamento de Matemática da UFMG (modelo da UFMG), que é um modelo que corrobora com o presente trabalho, pois é de referência e respeitado nacionalmente. E por último, na seção 7, apresentaremos uma proposta envolvendo o cálculo de probabilidade em sala de aula.

2 Teoria das Probabilidades

Nesta seção introduzimos alguns dos conceitos mais elementares da teoria das probabilidades. Esta teoria matemática teve, como ponto de partida, resolver problemas relativos à apostas de jogos de azar. No ano de 1654, Antonio de Gombaud (1607-1684), conhecido como cavaleiro de Méré, era um dos apostadores e procurou Blaise Pascal (1623-1662) para solucionar um destes problemas. Pascal, por sua vez, consulta Pierre Fermat (1601-1665), com isso eles começam uma troca de cartas com o propósito de solucionar este e outros problemas semelhantes. Depois disso, em 1713, foi publicado a obra *Ars Conjectandi* o primeiro livro inteiramente dedicado à teoria das probabilidades de autoria de Jacob Bernoulli (1654-1705). Em 1718, na obra *The Doctrine of Chaucesme*, Abraham de Moivre (1667-1754) apresenta o conceito do Teorema Central do Limite e em 1812, na obra *Theorie Analytique des Probabilités*, Pierre Simon Laplace (1749-1827) apresentou a definição clássica de probabilidade. Em 1933, Andrei Kolmogorov (1903-1987) propôs certos axiomas que deram uma formalidade matemática para a teoria da probabilidade que prevalece até hoje. A teoria das probabilidades tem sido útil para modelar experimentos onde é necessário incorporar incerteza ou chance como um elemento essencial do modelo. Assim, a probabilidade pode ser definida como um ramo da matemática que é responsável pelo estudo dos experimentos aleatórios.

2.1 Experimentos Determinísticos e Aleatórios

Existem dois tipos de experimentos na natureza: determinísticos e aleatórios. Um experimento determinístico é aquele que produz o mesmo resultado quando é repetido nas mesmas condições, por exemplo, água aquecida a 100°C, sob pressão normal, entra em ebulição. Muitas outras leis da física são exemplos de experimento determinístico. Por outro lado, um experimento aleatório é aquele que, quando repetido nas mesmas condições, o resultado observado nem sempre é o mesmo e nem sempre é previsível. Lançar um dado e observar qual face está voltada para cima e resultados de jogos de futebol são exemplos de experimentos aleatórios.

2.2 Espaço amostral

Mencionamos que a probabilidade é o ramo da matemática que é responsável pelo estudo dos experimentos aleatórios. A princípio, não sabemos qual o resultado de um experimento aleatório, o que torna conveniente agrupar todos os resultados possíveis em um conjunto. Chamaremos este conjunto de espaço amostral.

Definição 2.1 (Espaço amostral) *O conjunto de todos os possíveis resultados de um experimento aleatório se chama espaço amostral e é representado pela letra* Ω *(ômega maiúscula). Um resultado particular do experimento é representado pela letra* ω *(ômega minúsculas).*

Exemplo 2.1 *Se um experimento aleatório consiste em observar o resultado de uma determinada equipe em um jogo de futebol, então claramente seu espaço amostral é* $\Omega = \{Vitória, Empate, Derrota\}$.

Exemplo 2.2 *Suponha que um experimento aleatório consiste em observar o tempo em que um time de futebol fica sem perder. Se considerarmos o tempo em escala contínua, então podemos adotar como espaço amostral o intervalo [0,* ∞*). Neste caso o espaço amostral é infinito e não enumerável. Se consideramos o tempo de forma discreta (rodadas, dias, semanas, etc), então um possível espaço amostral é* $\Omega = \{0, 1, 2, 3, ...\}$. *Neste caso o espaço amostral é infinito e enumerável.*

Muitas vezes, não queremos apenas saber qual a probabilidade de um resultado ω apenas, mas de uma coleção deles. No Exemplo 2.2, é relevante perguntar qual a probabilidade do time de futebol perder em até 3 rodadas, neste caso, não é suficiente saber quais são as chances do time perder somente na terceira rodada. Isto nos motiva seguinte definição.

Definição 2.2 (Evento) *Um evento é um subconjunto do espaço amostral* Ω *e é representado por uma letra maiúscula do alfabeto latino, por exemplo, A, B ou C.*

- Dado $\omega \in \Omega$, um evento ocorre se e somente se $\omega \in A$.
- O evento $A = \Omega$ é chamado de evento certo, e o evento $A = \emptyset$ é chamado de evento impossível. Estes são os eventos triviais.
- Um evento simples consta de um único resultado do experimento aleatório, e um evento composto conta de dois ou mais resultados.

Exemplo 2.3 *Voltando ao exemplo 2.2 e suponha que estamos interessados em calcular a probabilidade do time de futebol perder em até 3 rodadas, então podemos definir o evento* $A = \{\textit{Time de futebol perder em até 3 rodadas}\} = \{0, 1, 2, 3\}$.

Exemplo 2.4 *Se um experimento aleatório consiste em lançar um dado de 6 faces e observar a face voltada para cima, então o espaço amostral é* $\Omega = \{1, 2, 3, 4, 5, 6\}$. *Alguns eventos possíveis são.*

- $A = \{\textit{números pares}\} = \{2, 4, 6\}$.
- $B = \{\textit{número dois}\} = \{2\}$.
- $C = \{\textit{divisores de } 60\} = \{1, 2, 3, 4, 5, 6\}$.
- $D = \{\textit{números maiores que } 6\} = \emptyset$.

Observe que o evento A é composto, o evento B é simples, o evento C é certo e o evento D é impossível.

2.3 Probabilidade Clássica

Definição 2.3 (Probabilidade Clássica) *Seja A um subconjunto do espaço amostral Ω de cardinalidade finita. Definimos a probabilidade clássica do evento A como:*

$$P(A) = \frac{\#(A)}{\#(\Omega)}. \tag{1}$$

onde $\#(A)$ é a cardinalidade do conjunto A e $\#(\Omega)$ é a cardinalidade do espaço amostral.

Esta definição só é válida para espaços amostrais finitos, ou seja, precisamos assumir que o número de elementos em Ω é finito. Além disso, o espaço amostral deve ser equiprovável, ou seja, cada evento tem a mesma probabilidade de ocorrência. Assim para calcular a probabilidade de um evento A, só precisamos contar o número de elementos de A e dividir pelo número de elementos do conjunto total Ω.

Exemplo 2.5 *Voltado ao exemplo 2.4, temos um experimento aleatório que consiste em lançar um dado honesto de 6 faces e observar a face voltada para cima, então o espaço amostral é $\Omega = \{1, 2, 3, 4, 5, 6\}$. A probabilidade do evento $A = \{obter\ um\ número\ par\}$ é dada por:*

$$P(A) = \frac{\#(A)}{\#(\Omega)} = \frac{\#\{2, 4, 6\}}{\#\{1, 2, 3, 4, 5, 6\}} = \frac{3}{6} = \frac{1}{2}.$$

É possível verificar que a probabilidade clássica satisfaz as seguintes propriedades:

i) $P(\Omega) = 1$;

ii) $P(A) \geq 0$, para qualquer evento A;

iii) $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$, quando A e B são disjuntos, ou seja, quando a interseção é conjunto vazio.

2.4 Probabilidade Frequentista

Suponha que n repetições de um certo experimento aleatório sejam realizadas e que o número de vezes que um dado evento A ocorre é registrada. Então podemos definir a probabilidade de A como:

Definição 2.4 (Probabilidade Frequentista) *Seja n_A o número de ocorrências de um evento A em n repetições de um experimento aleatório. A probabilidade frequentista do evento A é definida por:*

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}. \tag{2}$$

Neste caso, nem sempre é possível realizar o experimento aleatório uma infinidade de vezes e, por enquanto, não podemos garantir a existência do limite definido acima. Portanto, através da definição anterior, não é possível calcular a probabilidade exata de qualquer evento, mas é possível obter uma aproximação empírica do valor $P(A)$, ou seja:

$$P(A) \approx \frac{n_A}{n}. \tag{3}$$

Exemplo 2.6 *Suponha que estamos interessados em calcular a probabilidade de vitória do Atlético - MG em um determinado jogo do campeonato brasileiro de futebol. Para isso vamos simular o jogo milhões de vezes e contar em quantas o Atlético foi vencedor, então:*

$$P(Vitória\ Atlético) \approx \frac{n_{VitóriaAtlético}}{n_{simulações}}.$$

Nas próximas seções esse resultado será formalizado através da Lei dos Grandes Números, que garante, em particular, que $P(A)$ converge para a probabilidade do evento A.

A probabilidade frequentista também satisfaz as propriedades que mencionamos no fim da seção 2.3.

2.5 Probabilidade Axiomática

Ao contrário das probabilidades definidas nas seções anteriores, a definição axiomática de probabilidade não estabelece uma forma direta de calcular as probabilidades, mas sim, regras que os cálculos das probabilidades devem satisfazer. A seguir apresentaremos três axiomas propostos por Andrei Kolmogorov em 1933.

Axiomas de Kolmogorov: Os três axiomas de Kolmogorov, também conhecido como axiomas da probabilidade são:

1) $P(A) \geq 0$, para qualquer evento A;

2) $P(\Omega) = 1$;

3)$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$, quando $A_1,\ A_2, \cdots$ são disjuntos dois a dois.

Um axioma é uma proposição aceita como verdade e sobre a qual se baseia uma teoria, no nosso caso, a teoria da probabilidade. É possível verificar que as definições anteriores de probabilidade satisfazem os três axiomas. A seguir apresentamos a definição axiomática de probabilidade.

Definição 2.5 (Probabilidade Axiomática) *Uma medida de probabilidade é qualquer função P definida para uma coleção de eventos, que satisfaz os três axiomas de Kolmogorov.*

Os axiomas acima fornecem todos os elementos necessários para que possamos de fato calcular a probabilidade de um determinado time ser campeão brasileiro. Ou seja, o primeiro axioma nos diz que a probabilidade de um time ser campeão é pelo menos 0, o segundo que a probabilidade de um dos times da série A ser campeão é 1 e o terceiro que a probabilidade de um time mineiro ser campeão é a soma das probabilidades de América, Atlético-MG e Cruzeiro o serem. Então dos Axiomas de Kolmogorov podemos deduzir a Lei dos Grandes Números, e esta sim fundamenta um método de cálculo de probabilidades bem preciso, i.e., o método de simulações [2].

Em outras palavras, tomando como ponto de partida que a probabilidade oscila entre 0 e 1 e que união disjunta é a soma das probabilidades (Axiomas de Kolmogorov) segue um resultado importantíssimo: a probabilidade de um evento é a razão entre o número de ocorrências do mesmo e o número de ensaios quando este tende a infinito (Lei dos Grandes Números). Ou mais precisamente

$$probabilidade = \lim_{n\to\infty} \frac{\text{n}^{\text{o}} \text{ de ocorrências em } n \text{ ensaios}}{n} \tag{4}$$

Portanto, a definição 2.4 está bem fundamentada, e se quisermos realmente saber qual a probabilidade de um dado time ser campeão, o que temos de fazer é repetir vários campeonatos e ver em quantos deles o clube leva a melhor. Quanto mais edições forem realizadas, mais preciso o cálculo. A probabilidade (4) será apresentada com rigor no teorema 4.1.

2.6 Sigma Álgebra

Nesta seção vamos definir uma coleção de subconjuntos do espaço amostral fechada sobre operações contáveis de união, interseção e complemento, chamada σ-álgebra (sigma-álgebra).

Definição 2.6 (σ-álgebra) *Uma coleção $\mathbb{A}$ de subconjuntos de um espaço amostral Ω não vazio, é uma σ-álgebra (sigma-álgebra) de subconjuntos de Ω se ela satisfizer as seguintes condições:*

1) $\Omega \in \mathbb{A}$;

2) *Se* $A \in \mathbb{A}$, *então* $A^c \in \mathbb{A}$;

3) *Se* $A_1, A_2, A_3, \ldots \in \mathbb{A}$, *então* $\bigcup_{k=1}^{\infty} A_k \in \mathbb{A}$.

Os elementos de $\mathbb{A}$ são chamados de eventos.

A partir de agora chamaremos de evento os elementos que pertencem a uma σ-álgebra associada a um espaço amostral.

Voltando na definição 2.5, uma medida de probabilidade é qualquer função P definida para uma coleção de eventos que satisfaz os três axiomas. Essa coleção de eventos é a σ-álgebra e corresponde ao domínio da função.

Exemplo 2.7 *Seja Ω um espaço amostral de um experimento aleatório, então os seguintes subconjuntos de Ω são σ-álgebras.*
a) $\mathbb{A} = \{\emptyset, \ \Omega\}$.

b) $\mathbb{A} = \{A, A^c, \emptyset, \ \Omega\}$, *onde* $A \subset \Omega$.

2.7 Espaço de Probabilidade

Com os elementos definidos nas seções anteriores, podemos agora definir a estrutura matemática para modelar um experimento aleatório.

Definição 2.7 (Espaço de Probabilidade) *Um espaço de probabilidade é uma terna* $(\Omega, \mathbb{A}, P)$ *onde* Ω *é um conjunto não-vazio,* $\mathbb{A}$ *é uma* σ*-álgebra dos eventos de* Ω *e* P *é uma medida de probabilidade definida sobre* $\mathbb{A}$.

Exemplo 2.8 *Suponha que um experimento aleatório consiste em lançar uma moeda honesta e observar a face voltada para cima. Então temos:*

- $\Omega = \{cara, \ coroa\}$, *com* $\#\Omega = 2$.

- $\mathbb{A} = \{\emptyset, \ cara, \ coroa, \ \Omega\}$.

- $P(cara) = P(coroa) = \dfrac{1}{\#(\Omega)} = \dfrac{1}{2}$.

  Logo $(\Omega, \mathbb{A}, P)$ *é um espaço de probabilidade.*

2.8 Probabilidade Condicional

Muitas vezes nos deparamos com situações em que, antes da realização de algum experimento descrito em um espaço de probabilidade (Ω, $\mathbb{A}$, P), temos alguma informação adicional. Queremos saber o quanto que essa informação pode afetar a medida de probabilidade dos eventos de $\mathbb{A}$.

Definição 2.8 (Probabilidade Condicional) *Seja* $(\Omega, \mathbb{A}, P)$ *um espaço de probabilidade e sejam* A *e* B *eventos quaisquer com* $P(B) \neq 0$*, definimos a probabilidade condicional de* A *dado* B *como:*

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{5}$$

Exemplo 2.9 *Suponha que um experimento aleatório consiste em lançar um dado honesto e observar a face voltada para cima. Definimos os seguintes eventos:*

$A = \{\textit{número três}\} = \{3\}$.
$B = \{\textit{números ímpares}\} = \{1, 3, 5\}$.

Então a probabilidade de A dado B é dada por:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\#\{3\}}{\#\{1, 3, 5\}} = \frac{1}{3}.$$

No futebol, a probabilidade condicional explica a variação das probabilidades dos times ao longo do campeonato, por exemplo, se um time vence uma partida, é natural que, para a próxima rodada, sua probabilidade de vitória aumente.

2.9 Independência de eventos

Quando a ocorrência de um evento não afeta a probabilidade de ocorrência de outro evento, dizemos que eles são independentes.

Definição 2.9 (Independência de eventos) *Dizemos que dois eventos A e B são independentes se:*

$$P(A \cap B) = P(A)P(B). \tag{6}$$

Exemplo 2.10 *Um experimento aleatório consiste em lançar um dado honesto de 6 faces e observar a face voltada para cima, então o espaço amostral é $\Omega = \{1, 2, 3, 4, 5, 6\}$.*

a) Os eventos $A = \{1, 2, 3\}$ e $B = \{3, 5\}$ são independentes pois, $P(A \cap B) = P(A)P(B) = \frac{1}{6}$.

b) Os eventos $A = \{1, 2, 3\}$ e $B = \{1, 3\}$ não são independentes pois, $P(A \cap B) = \frac{1}{3}$ e $P(A)P(B) = \frac{1}{6}$.

3 Variáveis Aleatórias

Nesta seção definiremos uma variável aleatória como uma função cujo o domínio é o espaço amostral e a imagem é o conjunto dos números reais, ou seja, o resultado do experimento aleatório é um número real obtido pela variável aleatória.

3.1 Variáveis Aleatórias

Vamos considerar um espaço de probabilidade (Ω, $\mathbb{A}$, P) associado a um experimento aleatório qualquer.

Definição 3.1 (Variáveis Aleatórias) *Uma variável aleatória X é uma função mensurável $X : \Omega \to \mathbb{R}$ tal que para qualquer número real x,*

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathbb{A}. \tag{7}$$

Dado um valor x, o conjunto de todos os elementos de Ω para os quais X atribui um valor menor ou igual a x deve estar contido na σ-álgebra definida em Ω (vide Definição 2.6).

Definição 3.2 (Função de Distribuição) *A função de distribuição de uma variável aleatória X, denotada por $F_X(X)$, é definida por*

$$F_X(X) = P(X \leq x), x \in \mathbb{R}. \tag{8}$$

3.2 Tipos de Variáveis Aleatórias

Existem três tipos de variáveis aleatórias: as contínuas, que assumem valores em qualquer intervalo dos números reais, as discretas, que assumem somente um conjunto enumerável (finito ou infinito) de valores e as mistas, que são contínuas e discretas ao mesmo tempo. Como estamos interessados em modelar os números de gols feitos e sofridos, vamos focar apenas nas variáveis discretas.

Definição 3.3 (Variáveis Aleatórias Discretas) *Dizemos que uma variável aleatória X é discreta se toma um número finito ou enumerável de valores, isto é, se existe um conjunto finito ou enumerável $\{x_1, x_2, ...\} \subset \mathbb{R}$ tal que $X(\omega) \in \{x_1, x_2, ...\} \ \forall_{\ (t)} \in \Omega$. A função $p(x_i)$ definida por $p(x_i) = P(X = x_i), i = 1, 2, ...$, é chamada função de probabilidade de X.*

Definição 3.4 (Função de Probabilidade) *A função de probabilidade, $P_x : \mathbb{R} \to [0, 1]$, de uma variável aleatória discreta é definida por*

$$P_x(X) = P_x(X = x) = P(\{\omega \in \Omega : X(\omega) = x\}), sendo \sum_x P_x(X) = 1. \tag{9}$$

3.3 Esperança de variáveis aleatórias

Definição 3.5 (Esperança de variáveis aleatórias) *Se X é uma variável aleatória com função de probabilidade $p(x)$, então a esperança de X, denotada por $E[x]$, é dada por*

$$E[x] = \sum_x xp(x), onde \sum_x xp(x) < \infty. \tag{10}$$

3.4 Independência de variáveis aleatórias

O conceito de independência de eventos que foi mostrado em seções anteriores pode ser estendido para o caso de variáveis aleatórias como definido a seguir.

Definição 3.6 (Independência de variáveis aleatórias) *Dizemos que duas variáveis aleatórias X e Y são independentes se os eventos $(X \leq x)$ e $(Y \leq y)$ são independentes para quaisquer valores reais de x e y, isto é, se a igualdade for satisfeita*

$$P[(X \leq x) \cap (Y \leq y)] = P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y). \tag{11}$$

3.5 Variáveis Aleatórias com Distribuição de Poisson

A distribuição de probabilidade de Poisson é utilizada quando queremos observar o número de ocorrências de um certo evento dentro de um intervalo de tempo, como por exemplo, o número de chamadas de uma central telefônica ou o número de gols feitos por um time de futebol durante uma partida. Para modelar este tipo de situação vamos definir a variável aleatória X como o número de ocorrência de um evento dentro de um intervalo específico de tempo, ou seja, X pode assumir os valores $0, 1, 2, ...$ Agora suponhamos que conhecemos a taxa média de ocorrência do evento de interesse, que denotaremos pela letra grega λ (lambda). O parâmetro λ é não negativo, e é interpretado como número esperado de ocorrências do evento dentro do intervalo de tempo. Definimos a distribuição de Poisson como:

Definição 3.7 (Distribuição de Poisson) *Uma variável aleatória X tem distribuição de Poisson com parâmetro $\lambda > 0$, denotamos $X \sim$ $Poisson(\lambda)$, quando sua função de probabilidade é*

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, ... \quad . \tag{12}$$

Seja X uma variável aleatória que conta a quantidade de gols feitos ou sofridos por um time de futebol, pela natureza da variável, dado as características apresentadas no início da seção 3.5, assumiremos que $X \sim$ $Poisson(\lambda)$, onde λ é a média de gols feitos ou sofridos por este time.

Exemplo 3.1 *Suponha que a quantidade de gols marcados como mandante pelo Atlético - MG tenha uma distribuição de Poisson com parâmetro $\lambda = 2, 13$. A probabilidade do Atlético - MG fazer dois gols no próximo jogo como mandante é dada por:*

Se X representa o número de gols marcados como mandante, temos

$$P\{X = 2\} = e^{-2,13}\frac{2, 13^2}{2!} = 0, 270.$$

Teorema 3.1 (Somas de duas variáveis aleatórias de Poisson independentes) *Se X e Y são variáveis aleatórias de Poisson independentes com respectivos parâmetros λ_1 e λ_2, então a distribuição de $X+Y$ é Poisson com parâmetro $\lambda = \lambda_1 + \lambda_2$*

Demonstração: Como o evento $\{X+Y=n\}$ pode ser escrito como a união dos eventos disjuntos $\{X=k, Y=n-k\}$, $0 \leq k \leq n$, temos

$$P\{X+Y=n\} = \sum_{k=0}^{n} P\{X=k, X+Y=n\} \tag{13}$$

$$= \sum_{k=0}^{n} P\{X=k, Y=n-k\} \tag{14}$$

$$= \sum_{k=0}^{n} P\{X=k\}P\{Y=n-k\}. \tag{15}$$

Como X e Y são Poisson, então:

$$P\{X+Y=n\} = \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k!)} \tag{16}$$

$$= e^{-(\lambda_1+\lambda_1)}\sum_{k=0}^{n} \frac{\lambda_1^k \lambda_2^{n-k}}{k!(n-k)!}. \tag{17}$$

Multiplicando por $\frac{n!}{n!}, temos$:

$$P\{X+Y=n\} = \frac{e^{-(\lambda_1+\lambda_1)}}{n!}\sum_{k=0}^{n} \frac{n!}{k!(n-k)!}\lambda_1^k \lambda_2^{n-k}. \tag{18}$$

Pelo Binómio de Newton, $(\lambda_1+\lambda_2)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!}\lambda_1^k \lambda_2^{n-k}$, temos que:

$$P\{X+Y=n\} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n. \tag{19}$$

Assim, $X+Y$ tem distribuição de Poisson com parâmetros $\lambda_1+\lambda_2$. $\square$

Uma outra estatística de interesse específico para este trabalho é $W = (X+Y)/2$, isto é, a média de duas variáveis aleatórias de Poisson independentes, isto é, $X \sim Poisson(\lambda_1)$ e $Y \sim Poisson(\lambda_2)$. Podemos pensar por analogia ao Teorema 3.1 que W também teria distribuição Poisson. Contudo, podemos mostrar que a função de probabilidade de W dada por:

$$P(W = w) = \frac{e^{-(\lambda_1+\lambda_2)}}{w!} \sum_{y=0}^{w} \frac{w!}{(2w-y)!y!} \lambda_1^{w-y} \times \lambda_2^{y}, \tag{20}$$

não é uma função de probabilidade de Poisson. Isso ocorre, simplesmente devido ao termo que quantifica o número de variáveis envolvidas em W, no nosso caso é o valor 2. Caso, não houvesse esse termo, o somatório em (20) se resumiria ao termo $(\lambda_1 + \lambda_2)^w$, daí teríamos uma distribuição Poisson.

Assim, com estudos preliminares e simulações realizadas no R, percebemos que a expressão em (20) se aproxima de uma distribuição de Poisson de parâmetro igual a $(\lambda_1 + \lambda_2)/2$, isto é,

$$P(W) \approx \frac{e^{\lambda^*} \times (\lambda^*)^w}{w!}, \tag{21}$$

em que $\lambda^* = \frac{\lambda_1+\lambda_2}{2}$. Uma vantagem de se utilizar a aproximação em (21) é que esta função já está implementada em softwares como o R, e os resultados são satisfatórios dentro de problemas relacionados a jogos de futebol.

Será baseado em (21) a proposta do método para calcular probabilidades no futebol que será discutido na subseção 5.1.

Suponha que estamos simulando o jogo entre o time A e o time B, e que o time A faz em média 3 gols e o time B sofre em média 1 gol, é mais prudente esperar que o time A faça 2 gols do que ele faça 4 gols, então, neste trabalho, iremos somar a média de gols feitos como mandante de um time A com a média de gols sofridos como visitante do time B e tirar a média, que terá distribuição de Poisson e será denotada de fator mandante. Já a média de gols feitos como visitante com a média de gols sofridos como mandante será denotada de fator visitante, e também terá distribuição de Poisson. O exemplo 3.2 ilustra esta situação e na seção 5 definiremos com mais detalhes esses fatores.

Exemplo 3.2 *Suponha que em uma partida de futebol entre Atlético - MG e Chapecoense, a quantidade de gols marcados como mandante pelo Atlético - MG tenha uma distribuição de Poisson com parâmetro* $\lambda_1 = 2,13$ *e a quantidade de gols sofridos como visitante pela Chapecoense tenha uma distribuição de Poisson com parâmetro* $\lambda_2 = 2,27$. *Então o fator mandante do Atlético - MG, para essa partida, será aproximado por uma distribuição de Poisson com parâmetro* $\lambda = 2,20 = \dfrac{2,13 + 2,27}{2}$

4 Teoremas Limite

Nesta seção apresentaremos um dos resultados teóricos mais importantes da teoria da probabilidade, que é a Lei dos Grandes Números. Este resultado tem papel fundamental na execução deste trabalho.

Sejam X e $(X_n)_{n\in\mathbb{N}}$ variáveis aleatórias definidas num mesmo espaço de probabilidade $(\Omega,\ \mathbb{A},\ \mathrm{P})$.

Definição 4.1 (Convergência em probabilidade) *Dizemos que X_n converge em probabilidade para X, denotado $X_n \xrightarrow{P} X$, se para todo $\varepsilon > 0$*

$$P\left(|(X_n - X)| \geq \varepsilon\right) \to 0 \quad \textit{quando} \quad n \to \infty \tag{22}$$

A convergência em probabilidade garante que para valores grandes de n as variáveis X_n e X são aproximadamente iguais com probabilidade alta. A seguir, apresentaremos a Lei dos Grandes Números, que é a primeira aproximação para a soma de muitas variáveis aleatórias.

Sejam $X_1, X_2, ...X_n$ variáveis aleatórias integráveis em $(\Omega,\ \mathbb{A},\ \mathrm{P})$ e $S_1, S_2, ...S_n$ suas somas parciais dadas por

$$S_n = X_1 + X_2 + ... + X_n. \tag{23}$$

Definição 4.2 (Lei Fraca dos Grandes Números) *Dizemos que a sequência $(X_1, X_2, ...)$ satisfaz a Lei Fraca dos Grandes Números se, para todo $\varepsilon > 0$, vale*

$$P\left(\left|\frac{S_n - E(S_n)}{n}\right| \geq \varepsilon\right) \to 0, \quad \textit{quando} \quad n \to \infty. \tag{24}$$

ou seja, se

$$\frac{S_n - E(S_n)}{n} \xrightarrow{P} 0. \tag{25}$$

Teorema 4.1 *Seja $(\Omega,\ \mathbb{A},\ P)$ um espaço de probabilidade e $A \in \Omega$ um evento com probabilidade p. Seja $S_n(A)$ a variável aleatória que conta o número de ocorrencias de A em n ensaios independentes. Então, para todo $\varepsilon > 0$,*

$$P\left(\left|\frac{S_n(A)}{n} - p\right| \geq \varepsilon\right) \to 0, \quad \textit{quando} \quad n \to \infty \tag{26}$$

O teorema 4.1 é a uma prova da definição 2.4 e ele nos diz que se realizarmos um grande número de observações independentes sobre um determinado evento, então a razão entre o número de ocorrências do mesmo e o número de observações converge para a probabilidade do evento acontecer. Esta afirmação, antes dos Axiomas de Kolmogorov, era a base da teoria frequencista, que dava à teoria de probabilidades um caráter, a princípio, mais empírico que matemático. A dedução da Lei dos Grandes Números a partir dos Axiomas de Kolmogorov nos mostra que o "empirismo" em questão já não guarda mais eventuais resquícios pejorativos que o termo possa ter (teste, tentativa, suposição, apriorismo, etc). Muito pelo contrário, o "frequencismo" passa a ser então a nossa grande ferramenta para se estimar probabilidades que (sempre) desconhecemos. Mais ainda, está muito bem fundamentado do ponto de vista teórico, com todo o rigor matemático [2].

5 Metodologia

Nesta seção, iremos apresentar o método proposto por este trabalho e uma medida para verificar a qualidade dos resultados.

5.1 Método Proposto

Para [2], o modo de se estimar probabilidades em futebol é simples e de certa forma comum aos diferentes modelos. A ideia é partir da situação atual do campeonato e simular os jogos restantes. Ao final, o computador gera a classificação dos times e a registra. Depois, volta a fazer o mesmo inúmeras vezes, e ao término do processo já está apto a divulgar as probabilidades que nos interessam. Se, por exemplo, queremos saber a probabilidade de um dado time ser campeão, basta ver a proporção entre as simulações em que terminou líder sobre o total de simulações.

O método do PROFMAT atribui, a cada time participante do campeonato, dois vetores de médias

$$Mandante = (GFM; GSM) \quad e \quad Visitante = (GFV; GSV), \tag{27}$$

onde GFM é a média de gols feitos como mandante, GSM é a média de gols sofridos como mandante, GFV é a média de gols feitos como visitante, GSV é a média de gols sofridos como visitante e essas médias são calculadas da seguinte maneira:

$$GFM = \frac{\sum_{i=0}^{r}(\text{gols feitos como mandante})}{r+1}, \tag{28}$$

$$GSM = \frac{\sum_{i=0}^{r}(\text{gols sofridos como mandante})}{r+1}, \tag{29}$$

$$GFV = \frac{\sum_{i=0}^{r}(\text{gols feitos como visitante})}{r+1}, \tag{30}$$

$$GSV = \frac{\sum_{i=0}^{r}(\text{gols sofridos como visitante})}{r+1}, \tag{31}$$

onde r é quantidade de rodadas do campeonato já realizadas e i é o índice que percorre essas rodadas – veja que o somatório inicia em zero, pois na primeira rodada do campeonato não temos nenhuma rodada realizada.

A cada rodada, os vetores são realimentados. Se o time jogou como mandante, muda-se o seu vetor mandante, e o mesmo vale para seu vetor visitante, caso o jogo tenha sido fora de casa. Para [2], a ideia de se tratar um mesmo time como dois diferentes – um em casa, e outro fora dela – não é nenhum tipo de esquizofrenia futebolística. Ao contrário, reflete apenas o modo como a torcida, ou o próprio campo onde a equipe habitualmente joga, podem influenciar no desempenho do clube. Em suma, jogos iguais em campos distintos, são distintos.

De acordo com [2], a questão chave é como atualizar os vetores de médias de duas equipes que se enfrentam, após a realização do jogo. Pode-se dizer que isto é o que distingue os diferentes métodos de calcular probabilidades. A premissa da qual partimos é a de que, na primeira rodada do campeonato, quando $r = 0$, todos os times recebem as mesmas médias GFM, GSM, GFV e GSV, que são baseadas no último campeonato. Em 2017, por exemplo, essas médias foram GFM = GSV = 1,38 e GFV = GSM = 1,04, então na primeira rodada do campeonato de 2018, todos os times recebem esses dois vetores de médias

$$Mandante = (1,38; 1,04) \quad e \quad Visitante = (1,04; 1,38). \tag{32}$$

Na primeira rodada do campeonato brasileiro de 2018, o Grêmio venceu o Cruzeiro por 1 a 0 no Mineirão, então para simular a segunda rodada deste campeonato os vetores de médias dessas equipes são atualizados da seguinte forma

$$CRU_{mandante} = \left(\frac{1.38+0}{2}; \frac{1.04+1}{2}\right) \quad e \quad CRU_{vistante} = (1.38; 1.04). \tag{33}$$

$$GRE_{mandante} = (1.38; 1.04) \quad e \quad GRE_{vistante} = \left(\frac{1.04+1}{2}; \frac{1.38+0}{2}\right). \tag{34}$$

Como o Cruzeiro jogou em casa e o Grêmio fora, os vetores $CRU_{vistante}$ e $GRE_{mandante}$ não são alterados. Perceba que estamos dividindo por 2, pois temos apenas uma rodada do campeonato realizada (r = 1), ou seja, estamos calculando a média entre os resultados da primeira rodada com o parâmetro do campeonato passado, se por exemplo, estivermos simulando a 21ª rodada (quando r = 20), a soma do parâmetro do campeonato passado com os resultados da 20 rodadas passadas será dividida por 21. Seguindo esse raciocínio, a partir de uma rodada qualquer, simulamos todo o campeonato e armazenamos os resultados.

Com as médias GFM, GSM, GFV e GSV bem definidas, criamos o fator mandante ($\frac{GFM+GSV}{2}$) e o fator visitante ($\frac{GFV+GSM}{2}$). Como foi dito anteriormente, o fator mandante segue a Poisson com parâmetro $\frac{GFM+GSV}{2}$ e o fator visitante segue a Poisson com parâmetro $\frac{GFV+GSM}{2}$. Com esses dois fatores podemos formar o vetor média do jogo que é

$$VetorJogo = \left(\frac{GFM+GSV}{2}; \frac{GFV+GSM}{2}\right), \tag{35}$$

em que a primeira coordenada é o fator mandante e a segunda o fator visitante.

Voltando ao exemplo 3.2, onde o fator mandante é Poisson(2,20) e suponha que o fator visitante seja Poisson(1,50), então o vetor jogo será

$$V_{CAM\text{x}CHA} = (2,20; 1,50). \tag{36}$$

Com esse vetor o computador calcula as probabilidades através da função de probabilidade da Poisson e divide o intervalo $[0,1]$ em seis partes de acordo com as probabilidades acumuladas. As Tabelas 1 e 2 mostram as probabilidades e suas acumuladas.

Tabela 1: Probabilidade de gols para o fator mandante.

Gols	**Probabilidade**	**Acumulada**
0	0,111	0,111
1	0,244	0,355
2	0,268	0,623
3	0,197	0,819
4	0,108	0,928
5 ou mais	0,072	1,000

Fonte: Próprio autor.

Tabela 2: Probabilidade de gols para o fator visitante.

Gols	**Probabilidade**	**Acumulada**
0	0,223	0,223
1	0,335	0,558
2	0,251	0,809
3	0,126	0,935
4	0,047	0,982
5 ou mais	0,018	1,000

Fonte: Próprio autor.

Então para o fator mandante as seis partes do intervalo são: [0, 0,111], (0,111, 0,355], (0,355, 0,623], (0,623, 0,819], (0,819, 0,928] e (0,928, 1].

Para o fator visitante as seis partes do intervalo são: [0, 0,223], (0,223, 0,558], (0,558, 0,809], (0,809, 0,935], (0,935, 0,982] e (0,982, 1].

Logo após, o computador sorteia dois números aleatórios entre 0 e 1, um para o fator mandante e o outro para o fator visitante e vê quais subintervalos contém o número sorteado caiu. Se, por exemplo, o número sorteado para o fator mandante é $0,553$ e para o fator visitante é $0,448$, então o time mandante faz 2 gols e o time visitante faz 1 gol, assim o resultado do jogo é 2x1 para o time mandante.

Depois de simular uma rodada, atualizam-se os vetores médias de todos os times, de acordo com os resultados da rodada, e simula-se a rodada seguinte. Ao final da simulação do campeonato, o time que ficar em primeiro lugar será o campeão, os seis primeiros vão para libertadores, do sétimo ao décimo segundo vão para sulamericana e os quatro últimos são rebaixados. Assim, este processo é repetido milhares de vezes, para este trabalho fizemos 500 mil simulações, e a probabidade de um time ser campeão é quantidade de vezes que ele terminou em primeiro lugar dividido pelo número de simulações. Caso estejamos interessados nas probabilidades do resultados de um jogo, ao invés de todo um campeonato, o raciocínio é análogo: a probabilidade de vitória de um time é a quantidade de vezes que ele venceu dividido pelo número de simulações da partida.

A seguir, apresentamos um método para medir a qualidade dos resultados, as probabilidades encontradas pelo método apresentado e seu percentual de acerto comparado ao modelo do Departamento de Matemática da UFMG.

5.2 Qualidade dos Resultados

Para verificar a qualidade dos resultados do método proposto na subseção anterior, iremos utilizar a medida de DeFinetti: este método foi proposto por [7] e consiste na consideração de um plano em $\mathbb{R}^3$ como representação geométrica do conjunto de possíveis previsões probabilísticas. Os vértices desse plano correspondem às ocorrências dos resultados e os demais pontos a todas as outras possíveis previsões, ou seja,

$$S = \{(PVM, PEM, PVV) \in \mathbb{R}^3 : PVM + PEM + PVV = 1, PVM \geq 0, PEM \geq 0, PVV \geq 0\}. \tag{37}$$

onde PVM denota a probabilidade de vitória do mandante, PEM a probabilidade de empate e PVV a probabilidade de vitória do visitante.

A medida de distância de DeFinetti corresponde à distância euclidiana quadrática entre o ponto correspondente à (distribuição) probabilidade prevista e o vértice de uma previsão [5], um índice dado pela média aritmética das distâncias de DeFinetti, é chamado de medida DeFinetti. Para o futebol, associam-se os vértices $(1, 0, 0)$, $(0, 1, 0)$ e $(0, 0, 1)$ respectivamente à vitória da equipe mandante, ao empate, e à vitória da equipe vistante. Ao vetor de probabilidades atribuídas para uma determinada partida associa-se o ponto $(PVM, PEM, PVV) \in S$.

Assim, a distância de DeFinetti será igual a:

$$(PVM-1)^2+(PEM-0)^2+(PVV-0)^2 \text{ se a equipe mandante vencer a partida;} \tag{38}$$

$$(PVM-0)^2+(PEM-1)^2+(PVV-0)^2 \text{ se a partida terminar empatada;} \tag{39}$$

$$(PVM-0)^2+(PEM-0)^2+(PVV-1)^2 \text{ se a equipe visitante ganhar a partida.} \tag{40}$$

Por exemplo, se a previsão for (0.45, 0.30, 0.25) e o resultado for vitória da equipe mandante (1, 0, 0), então a distância de DeFinetti é $(0.45-1)^2+(0.30-0)^2+(0.25-0)^2 = 0.3025$. Nas previsões temos que um padrão comumente utilizado é a atribuição equiprovável de probabilidades ($PVM = PEM = PVV = 1/3$), ou seja, atribuir chances iguais a cada resultado em cada jogo. Para essa atribuição a medida de DeFinetti é igual $(1/3-1)^2+(1/3-0)^2+(1/3-0)^2 = 2/3$.

Assim, podemos considerar métodos de previsões de qualidade minimamente aceitável, aqueles que apresentam medidas de DeFinetti menores que 2/3 e de má qualidade, aqueles que apresentam medidas maiores que 2/3 [3].

6 Resultados

Nesta seção, apresentamos os resultados obtidos da aplicação do método no Campeonato Brasileiro de 2018. Primeiramente, realizamos uma breve análise dos resultados dos jogos. Em seguida, apresentamos os resultados para as probabilidades dos jogos de uma determinada rodada e para o campeonato inteiro. Se quisermos saber as probabilidades do campeonato para uma determinada rodada, utilizamos os resultados obtidos até a rodada anterior. Por exemplo, para as probabilidades da rodada 28, foram utilizados os resultados do campeonato até a rodada 27.

6.1 Análise Descritiva

Realizamos uma análise descritiva dos dados do Campeonato Brasileiro de 2018. A Tabela 3 indica o número de gols marcados que cada equipe fez como mandante, visitante e no geral. Como podemos ver, o Atlético-PR teve o melhor ataque como mandante (44 gols), o Flamengo teve o melhor ataque como visitante (29 gols) e o Palmeiras foi o melhor ataque do campeonato (64 gols). O Brasileirão de 2018 teve um total de 827 gols em 380 jogos, uma média de 2,18 gols por jogo. Desses 380 jogos, 202 (53%) terminaram com a vitória do mandante, 68 (18%) com a vitória do visitante e 110 (29%) com empate. O resultado mais comum foi a vitória do mandante por 1x0 (68 jogos), 238 jogos (63%) tiveram menos de três gols e 142 (37%) tiveram três ou mais gols.

Tabela 3: Número de gols marcados pelos times no campeonato.

Times	**1º Turno**		**2º Turno**		**Geral**	
	Mandante	**Visitante**	**Mandante**	**Visitante**	**Mandante**	**Visitante**
América-MG	13	5	9	3	22	8
Atlético-MG	19	14	12	11	31	25
Atlético-PR	16	2	28	8	44	10
Bahia	17	4	10	8	27	12
Botafogo	9	9	13	7	22	16
Ceará	6	4	12	10	18	14
Chapecoense	13	6	11	4	24	10
Corinthians	10	12	9	3	19	15
Cruzeiro	10	4	15	5	25	9
Flamengo	17	12	13	17	30	29
Fluminense	9	10	8	5	17	15
Grêmio	16	6	20	6	36	12
Internacional	14	12	18	7	32	19
Palmeiras	17	9	25	13	42	22
Paraná	7	2	6	3	13	5
Santos	14	4	14	14	28	18
São Paulo	15	15	10	6	25	21
Sport	11	8	8	8	19	16
Vasco da Gama	16	6	13	6	29	12
Vitória	11	9	11	5	22	14

Fonte: Próprio autor.

6.2 Probabilidades para uma partida

Ao todo, foram simulados 250 jogos e as probabilidade de vitória, empate e derrota foram calculadas da 14ª até a 38ª rodada. No método do PROFMAT, a medida DeFinetti associadas as probabilidades desses jogos obteve uma média igual a 0,5680, enquanto que no modelo da UFMG a média foi de 0,6273.

Por exemplo, na Tabela 4 apresentamos as probabilidades para cada uma das partidas da rodada 24 juntamente com o placar observado e a medida de DeFinetti. Como já era esperado, as probabilidades de vitória são maiores quando o time é mandante.

Observe que o método do PROFMAT estima as probabilidades de vitória do mandante maiores que as estimadas no modelo da UFMG. Com relação ao empate, as probabilidades estimadas pelo modelo da UFMG são maiores que as do método do PROFMAT, fato que pode ser observado na Figura 1.

Tabela 4: Probabilidades e medida de DeFinetti para a rodada 24.

Jogo	**Mandante**	**Visitante**	**Probabilidades**						**Placar**	**DeFinetti**	
			Vitória		**Empate**		**Derrota**				
			1	**2**	**1**	**2**	**1**	**2**		**1**	**2**
Jogo 1	Sport	Cruzeiro	0,398	0,218	0,321	0,510	0,186	0,272	0x0	0,698	0,362
Jogo 2	São Paulo	Bahia	0,624	0,462	0,244	0,395	0,132	0,143	1x0	0,218	0,466
Jogo 3	Flamengo	Chapecoense	0,653	0,487	0,233	0,341	0,115	0,172	2x0	0,188	0,409
Jogo 4	América	Ceará	0,474	0,306	0,292	0,431	0,234	0,263	0x0	0,780	0,486
Jogo 5	Fluminense	Botafogo	0,525	0,247	0,267	0,471	0,209	0,282	1x0	0,341	0,868
Jogo 6	Internancional	Grêmio	0,482	0,302	0,332	0,453	0,186	0,255	1x0	0,414	0,772
Jogo 7	Palmeiras	Corinthians	0,538	0,367	0,248	0,434	0,214	0,199	1x0	0,321	0,628
Jogo 8	Paraná	Santos	0,386	0,231	0,316	0,524	0,299	0,245	0x2	0,740	0,898
Jogo 9	Vitória	Vasco	0,470	0,455	0,282	0,309	0,248	0,236	1x0	0,423	0,448
Jogo 10	Atlético - MG	Atlético - PR	0,574	0,474	0,274	0,366	0,152	0,160	3x1	0,279	0,437

Legenda: 1 = PROFMAT - UFSJ e 2 = DEMAT - UFMG.

Fonte: Próprio autor e http://www.mat.ufmg.br/futebol.

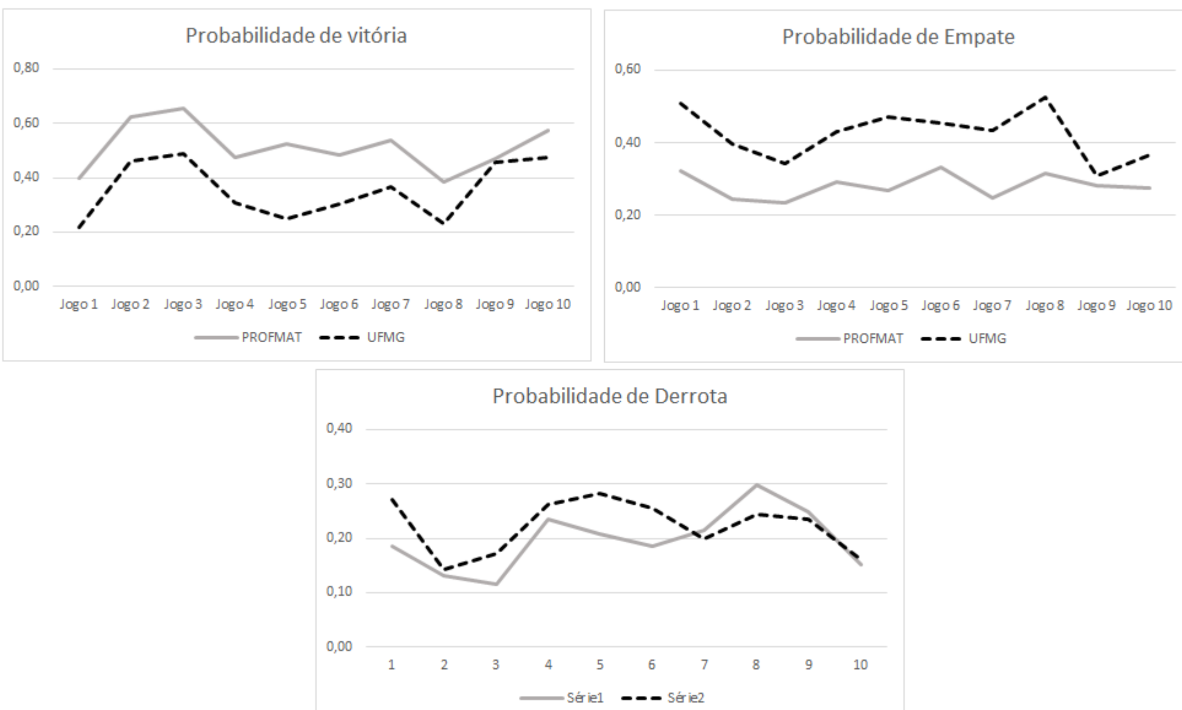

Figura 1: Probabilidades da rodada 24.
Fonte: Próprio autor e http://www.mat.ufmg.br/futebol.

Na Tabela 5 apresentamos o acerto de cada modelo, consideramos como acerto a ocorrência do evento com maior probabilidade. Dos 250 jogos simulados, o mandante venceu em 136, dos quais o método do PROFMAT acertou 131 e o modelo da UFMG 68. Quanto ao empate, 70 jogos terminaram com o placar igual, o método do PROFMAT não acertou nenhum resultado e o da UFMG acertou 38. Já o visitante venceu em 44 jogos, o método do PROFMAT acertou 6 e o da UFMG acertou apenas 1.

Aqui já detectamos algumas falhas do método do PROFMAT, que são a falta de acerto de empate e o pouco acerto de vitória do visitante. O ponto positivo do método é o bom acerto de vitória do mandante. Já o modelo da UFMG, acerta bem os empates e vitórias do mandantes, mas também deixa a desejar nas vitórias dos visitantes.

Na Figura 2 é possível ver a medida de DeFinetti para cada um dos jogos da rodada 24, tanto o modelo da UFMG quanto o método do PROFMAT apresentaram 3 resultados acima de 2/3.

Tabela 5: Percentual de acerto por modelo.

Modelo	**Vitória (136)**	**Empate (70)**	**Derrota (44)**
PROFMAT	131 (96,32%)	0 (0,00%)	6 (13,64%)
UFMG	68 (50,00%)	38 (54,28%)	1 (2,27%)

Fonte: Próprio autor e http://www.mat.ufmg.br/futebol.

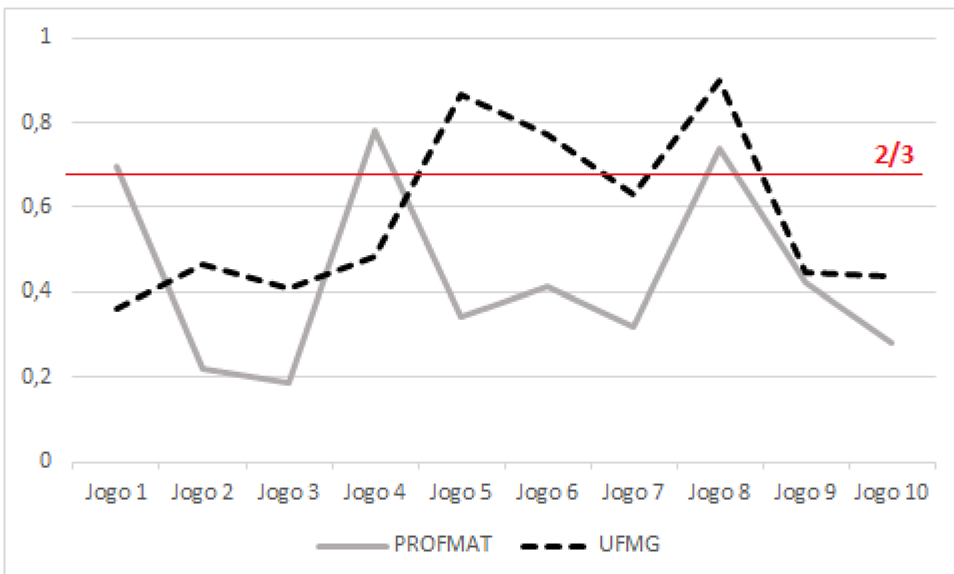

Figura 2: Medida de DeFinetti para a rodada 24.
Fonte: Próprio autor e http://www.mat.ufmg.br/futebol.

Na Tabela 6 apresentamos os percentis da medida de DeFinetti, no modelo da UFMG, aproximadamente 60% dos resultados estão abaixo de 2/3, enquanto que no modelo proposto por esse trabalho esse percentual é próximo de 57%.

Tabela 6: Percentis da medida de DeFinetti.

Percentis	**DeFinetti**	
	PROFMAT	**UFMG**
0.56	0.648	0.631
0.57	**0.670**	0.646
0.58	0.682	0.655
0.59	0.693	0.664
0.60	0.702	**0.668**

Fonte: Próprio autor e http://www.mat.ufmg.br/futebol.

6.3 Probabilidades para a Classificação Final do Campeonato

Para obter as probabilidades de um determinado time ser campeão, de ir para Libertadores ou de ser rebaixado, realizamos 500 mil simulações para cada rodada a partir da 14ª rodada. Na Tabela 7, apresentamos as probabilidades para os seis melhores times do campeonato, tomadas ao início de cada rodada do segundo turno, de se tornarem campeões. Observe que a partir da rodada 29 o Palmeiras passou a ser o time favorito ao título, 0,569 no método do PROFMAT e 0,553 no modelo da UFMG, reflexo de uma sequência de 13 jogos sem derrota, com 3 empates e 10 vitórias.

Tabela 7: Probabilidade de conquista do título para cada um dos seis melhores times do campeonato.

Rodada	Palmeiras		Flamengo		Grêmio		Internacional		São Paulo		Atlético-MG	
	1	2	1	2	1	2	1	2	1	2	1	2
20	0,049	0,028	0,142	0,142	0,095	0,127	0,224	0,197	0,447	0,464	0,041	0,044
21	0,065	0,045	0,156	0,187	0,058	0,083	0,336	0,299	0,360	0,361	0,023	0,023
22	0,061	0,036	0,150	0,159	0,040	0,049	0,277	0,205	0,459	0,510	0,011	0,038
23	0,126	0,092	0,088	0,090	0,071	0,100	0,289	0,204	0,413	0,473	0,012	0,039
24	0,168	0,145	0,057	0,063	0,060	0,083	0,385	0,349	0,310	0,338	0,019	0,021
25	0,173	0,144	0,055	0,059	0,024	0,023	0,425	0,409	0,304	0,345	0,019	0,019
26	0,189	0,161	0,055	0,045	0,040	0,047	0,369	0,376	0,329	0,354	0,018	0,015
27	0,289	0,260	0,076	0,079	0,054	0,088	0,338	0,322	0,233	0,244	0,008	0,007
28	0,341	0,330	0,050	0,049	0,083	0,129	0,339	0,342	0,178	0,144	0,009	0,006
29	0,569	0,553	0,114	0,109	0,042	0,078	0,198	0,186	0,074	0,078	0,003	0,002
30	0,628	0,619	0,121	0,125	0,010	0,018	0,216	0,212	0,024	0,026	0,000	0,000
31	0,707	0,735	0,180	0,169	0,004	0,006	0,102	0,085	0,006	0,006	0,000	0,000
32	0,776	0,793	0,121	0,109	0,001	0,001	0,089	0,085	0,014	0,012	0,000	0,000
33	0,862	0,897	0,059	0,037	0,001	0,001	0,076	0,064	0,003	0,002	0,000	0,000
34	0,920	0,933	0,022	0,010	0,001	0,003	0,055	0,053	0,001	0,005	0,000	0,000
35	0,952	0,967	0,013	0,005	0,000	0,000	0,035	0,028	0,000	0,000	0,000	0,000
36	0,946	0,966	0,044	0,023	0,000	0,000	0,010	0,011	0,000	0,000	0,000	0,000
37	0,973	0,984	0,023	0,016	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
38	1,000	1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Legenda: 1 = PROFMAT - UFSJ e 2 = DEMAT - UFMG.

Fonte: Próprio autor e http://www.mat.ufmg.br/futebol.

O São Paulo iniciou o segundo turno como o time com maior probabilidade de título, 0,447 no método do PROFMAT e 0,464 no modelo da UFMG, mas, nos 10 primeiros jogos do returno, conquistou apenas 11 pontos, aproveitamento de 36,67%, com isso sua probabilidade de título foi caindo a cada rodada.

O Internacional no inicio do returno era o segundo time com maior probabilidade de título, 0,224 no método do PROFMAT e 0,197 no modelo da UFMG mas, assim como o São Paulo, não teve um bom início de segundo turno, conquistando 18 pontos em 10 jogos, aproveitamento de 60%.

Já o Palmeiras, nestes mesmos 10 jogos, venceu 8 e empatou 2, conquistando 26 pontos, um aproveitamento de 86,67%, fato que justifica o crescimento da probabilidade de título a cada rodada.

As probabilidades de título do método do PROFMAT estão coerentes com as do modelo da UFMG: na Figura 3, é possível ver um comparativo dessas probabilidades para cada um desses seis melhores times. Observe que de modo geral, o compartamento da curva de probabilidades do método do PROFMAT acompanha a curva do modelo da UFMG.

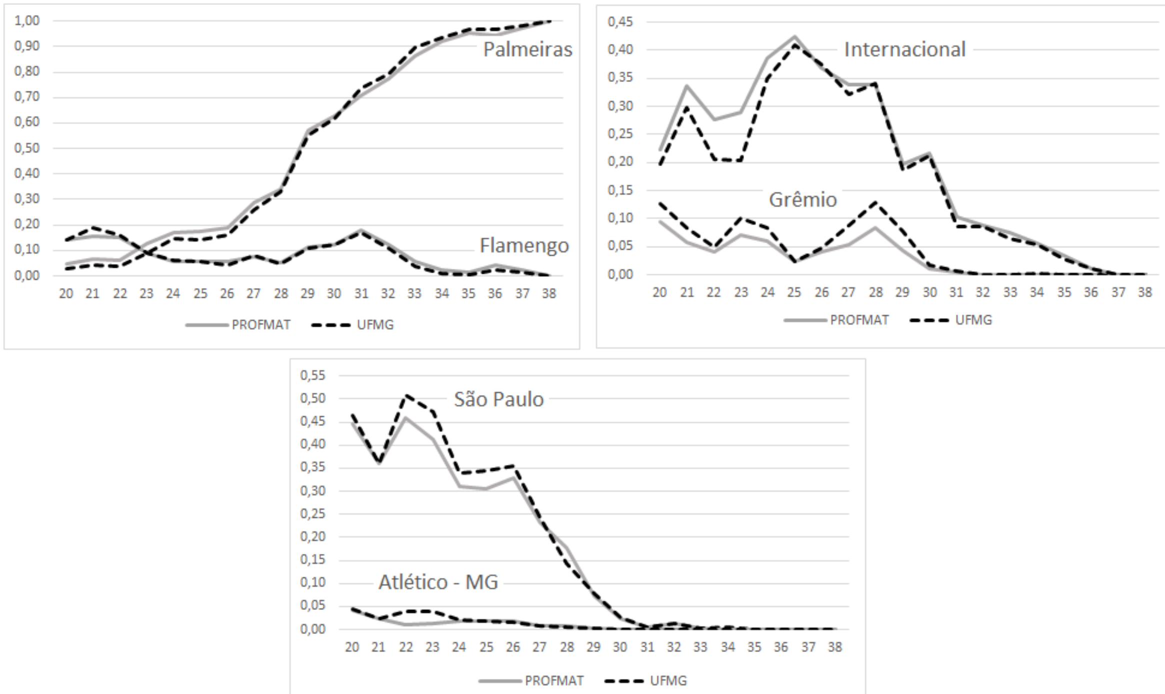

Figura 3: Probabilidade de conquista do título para cada um dos seis melhores times do campeonato.

Fonte: Próprio autor e http://www.mat.ufmg.br/futebol.

Na Tabela 8, apresentamos as probabilidades de conquista de uma vaga para a Libertadores de cada uma das seis melhores equipes. Todas as seis equipes iniciaram o segundo turno com probabilidade acima de 0,700. O São Paulo, que era o líder do campeonato, tinha 0,988 no método do PROFMAT e 0,989 no modelo da UFMG, mas como teve um início ruim de segundo turno, conquistou a vaga só na 32ª rodada.

O Internacional que era o vice líder, iniciou o returno com probabilidade de conquista de uma vaga na Libertadores de 0,962 no método do PROFMAT e 0,950 no modelo da UFMG e conquistou a vaga em definitivo na 29ª rodada.

Já o Palmeiras, conquistou a vaga na 28ª rodada, reflexo da sua campanha de 13 jogos sem derrota. O Flamengo conquistou a vaga na Libertadores na 30ª rodada e o Grêmio na 33ª rodada.

O último time a conquistar a vaga na Libertadores foi o Atlético-MG, que entrou na última rodada disputando com o Atlético-PR, as probabilidades desta disputa podem ser vistas na Figura 4.

Tabela 8: Probabilidade de conquista de uma vaga na libertadores para cada um dos seis melhores times do campeonato.

Rodada	Palmeiras		Flamengo		Grêmio		Internacional		São Paulo		Atlético-MG	
	1	2	1	2	1	2	1	2	1	2	1	2
20	0,806	0,735	0,934	0,920	0,904	0,913	0,962	0,950	0,988	0,989	0,758	0,735
21	0,879	0,810	0,956	0,957	0,882	0,892	0,986	0,982	0,987	0,986	0,712	0,666
22	0,890	0,777	0,961	0,946	0,854	0,817	0,984	0,969	0,994	0,993	0,621	0,756
23	0,946	0,887	0,918	0,882	0,920	0,898	0,985	0,969	0,992	0,992	0,638	0,747
24	0,973	0,953	0,906	0,867	0,930	0,910	0,995	0,992	0,991	0,989	0,771	0,710
25	0,986	0,979	0,936	0,925	0,906	0,842	0,998	0,998	0,995	0,996	0,837	0,806
26	0,989	0,984	0,941	0,916	0,944	0,920	0,997	0,997	0,997	0,997	0,850	0,794
27	0,996	0,993	0,967	0,958	0,965	0,963	0,997	0,996	0,994	0,994	0,792	0,676
28	0,999	0,999	0,973	0,970	0,991	0,992	0,999	0,999	0,996	0,996	0,880	0,805
29	1,000	1,000	0,995	0,992	0,988	0,985	0,998	0,988	0,993	0,992	0,833	0,681
30	1,000	1,000	0,999	0,999	0,984	0,983	1,000	1,000	0,991	0,990	0,769	0,633
31	1,000	1,000	1,000	1,000	0,988	0,991	1,000	1,000	0,989	0,994	0,691	0,555
32	1,000	1,000	1,000	1,000	0,958	0,966	1,000	1,000	0,998	0,999	0,542	0,364
33	1,000	1,000	1,000	1,000	0,998	0,999	1,000	1,000	1,000	1,000	0,442	0,287
34	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,479	0,321
35	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,685	0,554
36	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,682	0,581
37	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,871	0,839
38	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,952	0,889

Legenda: 1 = PROFMAT - UFSJ e 2 = DEMAT - UFMG

Fonte: Próprio autor e http://www.mat.ufmg.br/futebol.

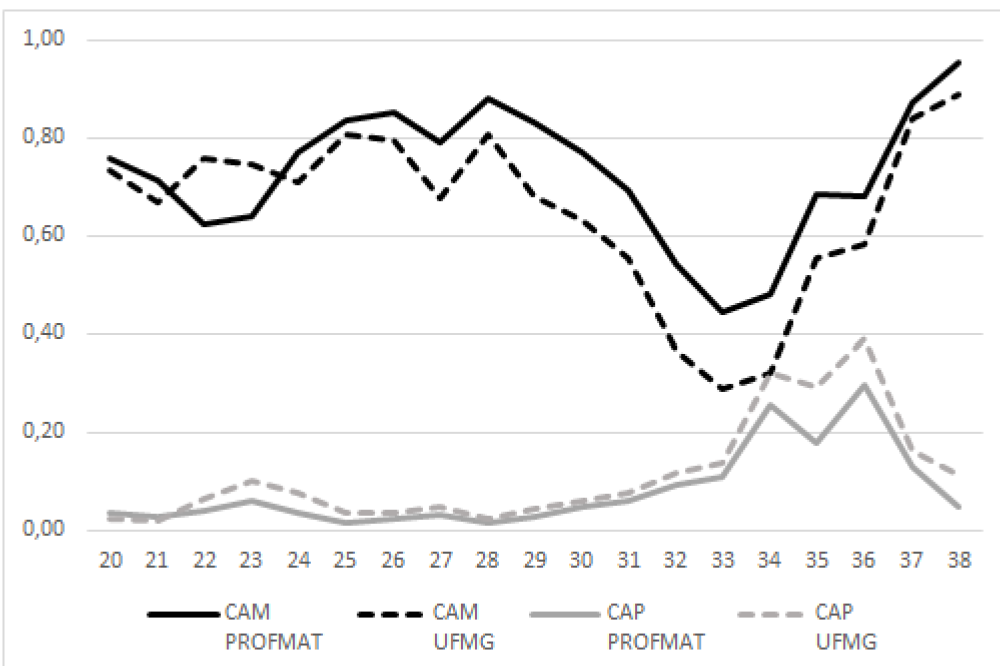

Figura 4: Probabilidade de conquista de uma vaga na Libertadores Atlético - MG (CAM) versus Atlético - PR (CAP).

Fonte: Próprio autor e http://www.mat.ufmg.br/futebol.

Na Figura 5, é possível ver um comparativo das probabilidades de conquista de uma vaga para a Libertadores para cada um desses seis melhores times. Observe que de modo geral, o compartamento da curva de probabilidades do método do PROFMAT acompanha a curva do modelo da UFMG.

Uma observação importante é que o Cruzeiro foi campeão da Copa do Brasil antes da 31ª rodada, mas ele não esteve entre os seis melhores e o Atlético-PR foi campeão da Sulamericana após o fim do campeonato brasileiro.

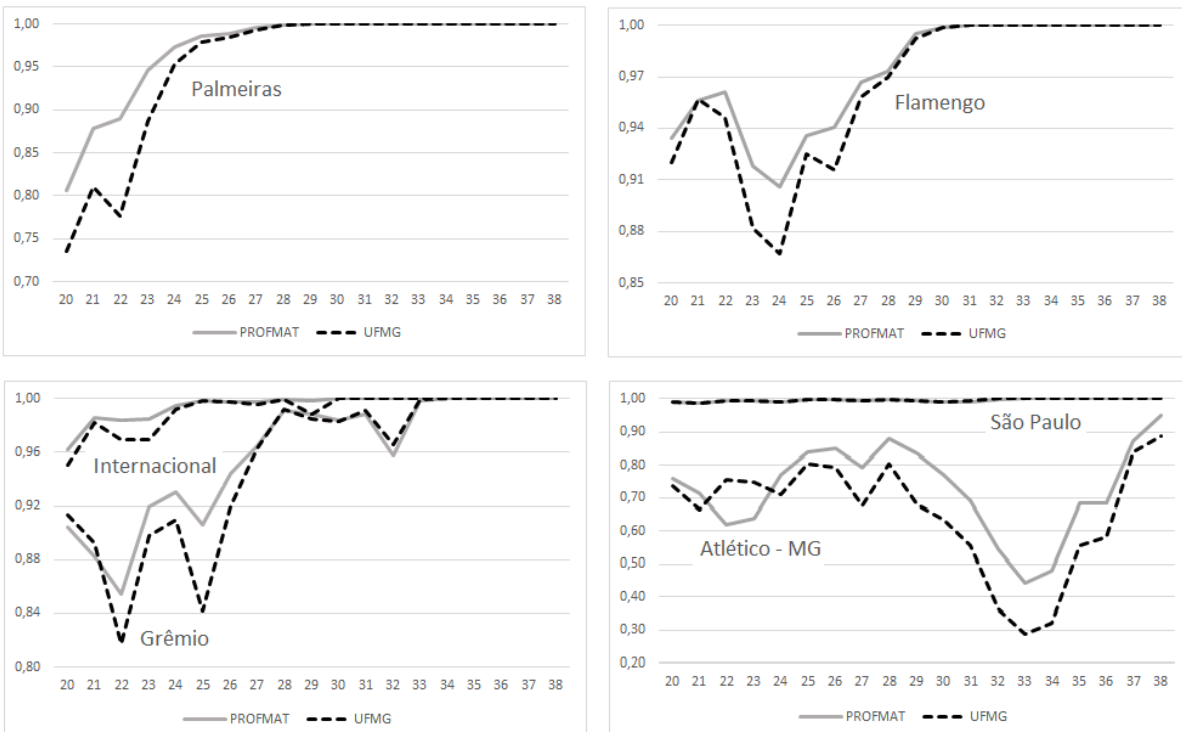

Figura 5: Probabilidade de conquista de uma vaga na Libertadores para cada um dos seis melhores times do campeonato.

Fonte: Próprio autor e http://www.mat.ufmg.br/futebol.

Na Tabela 9, apresentamos as probabilidades de rebaixamento das seis piores equipes, ao fim do campeonato os quatro últimos colocados caem para a segunda divisão. O Paraná, que era o lanterna do campeonato com 14 pontos ganhos e aproveitamento de 24,6%, iniciou o returno com probabilidade de rebaixamento de 0,874 no método do PROFMAT e 0,807 no modelo da UFMG e foi matematicamente rebaixado na 29ª rodada.

O Ceará, vice lanterna com 16 pontos ganhos e aproveitamento de 28,1%, foi o time que apresentou maior discrepância nas probabilidades, iniciou o segundo turno com probabilidade de rebaixamento de 0,723 no método do PROFMAT e 0,491 no modelo da UFMG, mas venceu 5 dos 10 primeiros jogos do returno, (inclusive venceu o Flamengo no Maracanã e Cruzeiro no Mineirão), empatou 3 e perdeu 2, aproveitamento de 60% e conquistou 18 pontos. Com esses resultados o Ceará saiu da zona de rebaixamento na 26ª rodada e foi até o fim do campeonato sem voltar para ela.

O América iniciou o returno com probabilidade de rebaixamento de 0,223 no método do PROFMAT e 0,216 no modelo da UFMG, ele estava na 12ª colocação com 22 pontos, mas nas últimas 10 rodadas do campeonato perdeu 6 jogos, inclusive perdeu em casa para o Paraná que já estava rebaixado, e acabou rebaixado.

O vitória foi rebaixado na 37ª rodada e cinco times entraram na última rodada com chances de serem rebaixados, Vasco, Chapecoence, Fluminense, América e Sport, os dois últimos foram rebaixados.

Na Figura 6, é possível ver a probabilidades de rebaixamento para as seis piores equipes do campeonato e novamente o compartamento da curva de probabilidades do método do PROFMAT acompanha a curva do modelo da UFMG.

Tabela 9: Probabilidade de rebaixamento para os seis piores times do campeonato.

Rodada	**Paraná**		**Vitória**		**América**		**Sport**		**Vasco**		**Ceará**	
	1	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
20	0,874	0,807	0,510	0,452	0,223	0,216	0,334	0,371	0,261	0,232	0,723	0,491
21	0,879	0,800	0,530	0,498	0,099	0,097	0,493	0,493	0,231	0,202	0,756	0,497
22	0,914	0,867	0,413	0,618	0,127	0,102	0,606	0,601	0,171	0,152	0,817	0,696
23	0,952	0,930	0,310	0,489	0,181	0,173	0,545	0,521	0,315	0,267	0,672	0,511
24	0,973	0,960	0,295	0,231	0,113	0,100	0,626	0,654	0,379	0,398	0,548	0,409
25	0,989	0,976	0,176	0,114	0,130	0,097	0,675	0,651	0,436	0,443	0,534	0,399
26	0,994	0,991	0,249	0,210	0,184	0,194	0,744	0,797	0,511	0,520	0,435	0,300
27	0,997	0,996	0,387	0,329	0,142	0,150	0,819	0,850	0,374	0,343	0,466	0,398
28	0,999	0,998	0,407	0,390	0,158	0,179	0,856	0,887	0,364	0,315	0,313	0,217
29	1,000	1,000	0,615	0,581	0,248	0,278	0,801	0,788	0,392	0,350	0,159	0,224
30	1,000	1,000	0,385	0,360	0,235	0,293	0,868	0,856	0,253	0,169	0,227	0,302
31	1,000	1,000	0,398	0,420	0,246	0,328	0,775	0,761	0,313	0,269	0,250	0,155
32	1,000	1,000	0,598	0,601	0,351	0,484	0,537	0,525	0,395	0,344	0,127	0,066
33	1,000	1,000	0,655	0,732	0,622	0,713	0,367	0,322	0,174	0,147	0,214	0,139
34	1,000	1,000	0,773	0,772	0,878	0,915	0,341	0,350	0,244	0,267	0,258	0,180
35	1,000	1,000	0,716	0,782	0,882	0,919	0,349	0,300	0,240	0,261	0,248	0,192
36	1,000	1,000	0,888	0,905	0,676	0,662	0,478	0,398	0,271	0,324	0,175	0,150
37	1,000	1,000	0,980	0,990	0,858	0,843	0,729	0,778	0,036	0,030	0,037	0,023
38	1,000	1,000	1,000	1,000	0,781	0,671	0,715	0,764	0,109	0,105	0,000	0,000

Legenda: 1 = PROFMAT - UFSJ e 2 = DEMAT - UFMG.

Fonte: Próprio autor e http://www.mat.ufmg.br/futebol.

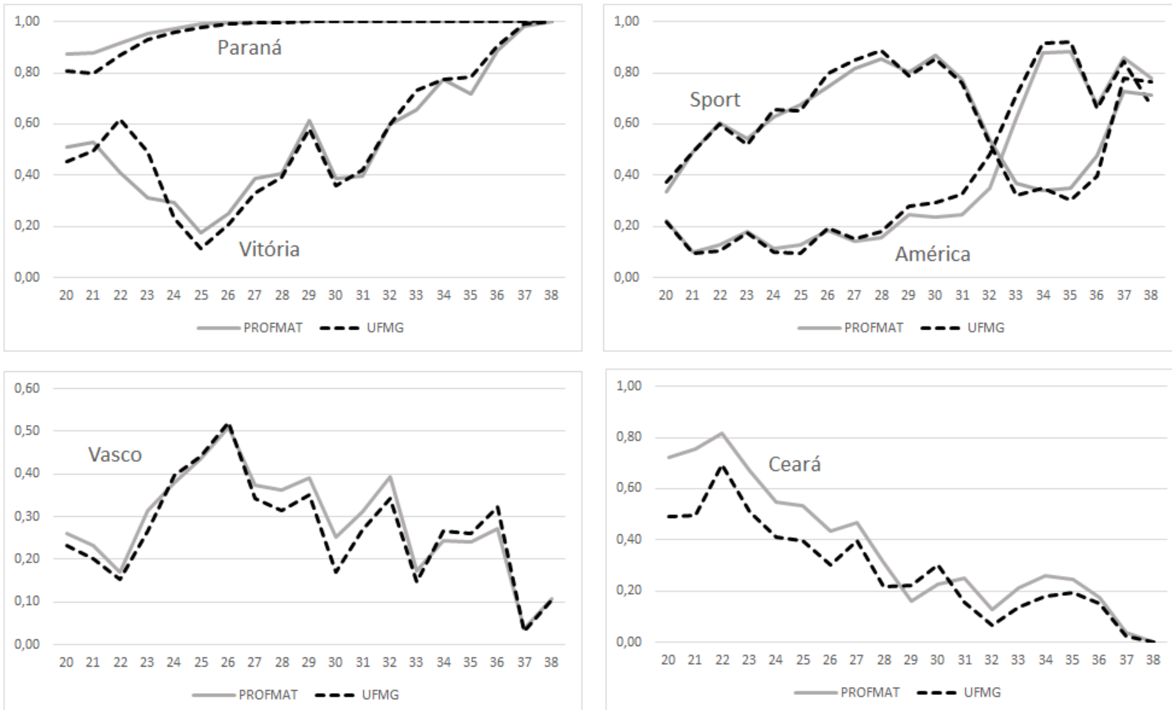

Figura 6: Probabilidade de rebaixamento para os seis piores times do campeonato
Fonte: Próprio autor e http://www.mat.ufmg.br/futebol/

7 Propostas de Atividades em Sala de Aula

Nesta seção iremos propor uma atividade que pode ser utilizada em sala de aula para a aplicação da distribuição de Poisson no cálculo de probabilidade no futebol. A atividade visa apresentar, com auxílio do Excel, um método de cálculo de probabilidade bem dinâmico para que os alunos possam discutir os resultados e apresentar novas ideias.

Primeiramente, o professor deverá detalhar a distribuição de Poisson, explicando como obter as probabilidades da partida, se possível, da rodada que está para acontecer.

A próxima etapa o professor deverá solicitar que os alunos formem cinco grupos e cada grupo ficará responsável pelo cálculo de dois jogos, visto que são 10 jogos por rodada. De acordo com [14], o professor poderá aproveitar este momento para conversar sobre os conceitos de probabilidades e as Leis dos Grandes Números, comentando sobre as simulações que são realizadas através de um programa e que determinam as probabilidades dos times para serem campeões ou para outras situações.

Em seguida, com o suporte do Excel, o professor deverá auxiliar os grupos na construção das Tabelas 1 e 2 para cada jogo da rodada. Com os dados destas tabelas, o grupo de alunos construirá as Tabelas 10, 11 e 12 que serviram de base para o cálculo das probabilidades da rodada.

A escolha da próxima rodada é ideal para que, ao final das tabelas construídas e apresentadas, assim que as partidas ocorrerem na realidade, os alunos possam comparar os resultados obtidos e estes servindo como pauta de discussão em sala de aula criando oportunidade para o professor esclarecer a diferença entre “chance de ocorrer” e a “ocorrência real” [14].

Tabela 10: Probabilidade de vitória do mandante.

Resultado	Cálculo	Probabilidade
1x0	0,244 x 0,223	0,055
2x0	0,268 x 0,223	0,060
2x1	0,268 x 0,335	0,090
3x0	0,197 x 0,223	0,043
3x1	0,197 x 0,335	0,065
3x2	0,197 x 0,251	0,048
4x0	0,108 x 0,223	0,023
4x1	0,108 x 0,335	0,035
4x2	0,108 x 0,251	0,026
4x3	0,108 x 0,126	0,013
5x0	0,072 x 0,223	0,015
5x1	0,072 x 0,335	0,023
5x2	0,072 x 0,251	0,017
5x3	0,072 x 0,126	0,008
5x4	0,072 x 0,047	0,003
Probabilidade de vitória do mandante		**0.525**

Fonte: Próprio autor.

Tabela 11: Probabilidade de empate.

Resultado	Cálculo	Probabilidade
0x0	0,111 x 0,223	0,026
1x1	0,244 x 0,335	0,084
2x2	0,268 x 0,251	0,068
3x3	0,197 x 0,126	0,024
4x4	0,108 x 0,047	0,005
5x5	0,072 x 0,018	0,001
Probabilidade de empate		**0.208**

Fonte: Próprio autor.

Tabela 12: Probabilidade de vitória do visitante.

Resultado	Cálculo	Probabilidade
0x1	0,111 x 0,335	0,039
0x2	0,111 x 0,251	0,029
1x2	0,244 x 0,251	0,063
0x3	0,111 x 0,126	0,015
1x3	0,244 x 0,126	0,031
2x3	0,268 x 0,126	0,034
0x4	0,111 x 0,047	0,006
1x4	0,244 x 0,047	0,012
2x4	0,268 x 0,047	0,013
3x4	0,197 x 0,047	0,009
0x5	0,111 x 0,018	0,002
1x5	0,244 x 0,018	0,005
2x5	0,268 x 0,018	0,005
3x5	0,197 x 0,018	0,004
4x5	0,108 x 0,018	0,002
Probabilidade de vitória do visitante		**0.267**

Fonte: Próprio autor.

Considerações Finais e Pesquisas Futuras

Neste trabalho realizamos, através da distribuição de Poisson, uma aplicação de uma metodologia bem simples para calcular probabilidades no futebol. Aplicando à 250 jogos da Série A do Campeonato Brasileiro de 2018, a metodologia estudada apresentou bons resultados preditivos, a média da medida de DeFinetti foi 0,568 e aproximadamente 57% dos resultados estão abaixo de 2/3.

Com a simulação foi possível prever com boa precisão o campeão (Palmeiras), e obteve resultados satisfatórios para os classificados para a Copa Libertadores da América, como também para os times rebaixados. Apesar de 250 jogos ser uma amostra pequena e da falta de acerto de empate e de vitória do visitante, os resultados obtidas pelo o método do PROFMAT estão coerentes com o modelo do Departamento de Matemática da UFMG, que é um modelo de referência e respeitado nacionalmente.

Para trabalhos futuros é necessário melhor a qualidade do ajuste para prever empates e vitórias do visitante, além disso, uma linha de pesquisa futura é aplicar a média de gols para os últimos 5 ou 10 jogos e não para o campeonato todo. Também seria interessante colocar um peso para os gols, por exemplo, fazer gol no Palmeiras, que teve a melhor defesa em 2018, deferia ter um peso maior do que fazer gol no Paraná, que foi a pior defesa. Em caso de clássicos, os dois times podem entrar como mandante, já que geralmente clássicos não tem favoritos.

A metodologia apresentada pode ser usada para outros campeonatos de pontos corridos, como por exemplo a Série B do Campeonato Brasileiro ou o Campeonato Inglês. É importante ressaltar que o modelo utilizado neste trabalho não leva em consideração vários fatores que podem influenciar o resultado de um determinado jogo, tais como a presença de torcida, suspensões, entre outros. Todas as implementações computacionais foram realizadas utilizando o Software R que é um ambiente computacional de livre acesso.

Agradecimentos

Agradeço em primeiro lugar a Deus, por me propocinar essa oportunidade, nada é possível sem a presença Dele. Agradeço à minha esposa Isabella por todo apoio, paciência e compreensão. À minha mãe Rosane, que nunca mediu esforços para a minha educação. Aos irmãos Pedro e Marcos por todo apoio. Ao meu primo Victor pela ajuda durante esse tempo de estudo. Agradeço aos amigos da turma de 2017 do PROFMAT, em especial aos amigos Bruna, Edson, Jonilson e Wendel. Aos amigos Renato Faria e Guilherme Furts pela ajuda com a parte computacional. Ao amigo Cézar Augusto pela ajuda com o LaTeX. Agradeço à professora Mariana Garabini, por todo carinho e dedicação para o bom andamento do PROFMAT no CAP. Agradeço à CAPES pelo apoio financeiro. Agradeço ao professor Gilcione, do Departamento de Matemática da UFMG, pela contribuição ao trabalho. Agradeço ao CEFET-MG, sem essa instituição esse sonho não seria possível. Por fim, um agradecimento especial aos meus orientadores, Dr. Humberto Cesar Fernandes Lemos e Dr. Ben Dêivide de Oliveira Batista, pela paciência, dedicaçãao e ensinamentos em prol da conclusão deste trabalho final.

Referências

[1] B. R. James, *Probabilidade: um curso em nível intermediário.* IMPA, 2015.

[2] B. N. B. de Lima, F. Brochero, G. N. Costa, G. M. Zeferino, M. T. Cunha e R. V. Martins: *Probabilidades no futebol.* Revista Matemática Universitária, ns.48/49, artigo 02, junho/dezembro de 2010.

[3] A. K. Suzuki: *Modelagem Estatística para Previsão Esportiva: Uma aplicação no Futebol.* Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos para obtenção do grau de Mestre em Estatística, 2007.

[4] F. F. Farias: *Análise e Previsão de Resultados de Partidas de Futebol.* Dissertação apresentada ao Instituto de Matemática da Universidade Federal do Rio de Janeiro para obtenção do grau de Mestre em Estatística, novembro de 2008.

[5] L. Arruda: *Poisson, Bayes, Futebol e DeFinetti.* Dissertação apresentada ao Instituto de Matemática e Estatística da Universidade de São Paulo para a obtenção do grau de mestre em Estatística, 2000.

[6] A. Suzuki e L. Tavares: *Modelagem Estatística para Previsão Esportiva: Uma aplicação no Futebol.* Revista Eletrônica Matemática e Estatística em Foco, Volume 3 - Número 1, Páginas: 32 a 47, Maio de 2015.

[7] B. DeFinetti: *Probability, Induction and Statistics.* John Wiley, London, 1972.

[8] J.B. Keller: *A Characterization of the Poisson Distribution and the Probability of Winning a Game.* The American Statistician, 48(4):294–298, 1994.

[9] A. Lee: *Modeling Scores in the Premier League: Is Manchester United Really the Best?* Chance, 10(1):15–19, 1997.

[10] R. Pollard: *Home advantage in soccer: a retrospective analysis.* Journal of Sports Sciences, pp. 237 – 248, 1986.

[11] A. Suzuki, L. Salasar, F. Louzada-Neto e J. Leite: *A bayesian approach for predicting match outcomes: The 2006 (Association) Football World Cup.* Journal of the Operational Research Society, 61:1530–1539 (October 2010), 2009.

[12] S. Ross: *Probabilidade : um curso moderno com aplicações*; tradutor: Alberto Resende De Conti. - 8. ed. - Porto Alegre : Bookman, 2010.

[13] C. A. B. Dantas: *Probabilidade: Um Curso Introdutório.* Edusp, 2008

[14] C. A. G. Carneiro: *Probabilidades no Futebol.* Dissertação apresentada ao Departamento de Estatística, Física e Matemática da Universidade Federal de São João Del-Rei para obtenção do grau de Mestre em Matemática, 2017.